

層別因子を含む探索的な回帰分析入門

説明変数に質的変数と量的変数の両方を含み、反応が量的変数として得られた場合の代表的な統計モデルとして共分散分析が知られている。一般的に共分散分析は、1 因子実験に際して結果に影響をおよぼすことが明らかな量的な共変量を含む解析法として定式化されている。観察研究においても、層別因子を考慮した回帰分析も共分散分析として認識されている。一定期間に何らかの処置を行い、量的な主反応を得た場合に、副反応が間接的に主反応に影響を与える交絡変数を含めた解析も共分散分析として認識されている。さらに、直線ではなくシグモイド曲線のあてはめる場合についても取りあげる。このように多くの分野で共分散分析が活用されているのであるが、総合的に論じている成書は見当たらない。そこで、関連する成書のデータを引用しつつ、Excel の回帰分析を主体に、行列関数を活用した新たな解析方法を提言し、統計ソフト JMP と SAS による解析結果も合わせて示すことにより、層別因子を含む探索的な回帰分析入門とする。

層別因子を含む探索的な回帰分析入門 目次

0. はじめに-----	1
1. 層別因子を含む各種の回帰分析の実際 -----	7
2. デザイン行列を活用した 1 因子実験データの解析 -----	53
3. 繰り返しが不揃いの 2 因子実験データの解析-----	89
4. 欠測値がある直交表の線形モデルによる解析-----	121
5. デザイン行列を用いた回帰分析の基礎-----	155
6. 伝統的な共分散分析からの脱却-----	195
7. 共変量を含む 3 因子実験データの探索的解析 -----	219
8. 交絡変数と共変量を含む 2 群比較 -----	243
9. 前後差の前値は常に共変量 -----	281
10. 層別因子を含むロジスティック曲線のあてはめ -----	327
11. 各種のシグモイド曲線を用いた逆推定 -----	367
12. ミカエリス・メンテン式をめぐる新たな統計解析 -----	405
13. ロジスティック曲線のさらなる活用 -----	455
文献, 文献索引, 索引, 解析用ファイル一覧 -----	489

続・高橋セミナー <https://www.yukms.com/biostat/takahasi2/index.htm>

高橋セミナーは、生物検定法の再構築をテーマとし 1999 年 12 月に第 1 回目を開催し、2007 年 3 月の第 27 回で終了した。本年 2011 年から統計モデルの活用をテーマにした「続・高橋セミナー」を再開する。テーマは、経時データの解析、繰返しが不揃いの枝分かれ実験データの解析、生存時間データの解析、反復を伴う各種実験データの解析などを対象にする。これらのデータ解析には、最小 2 乗法ではなく最尤法あるいは制限付最尤法(REML)を用いることになる。線形モデルに対する最小 2 乗法について書かれた日本語の教科書は多数あるが、最尤法あるいは制限付最尤法を用いた統計解析の方法を対象にした日本語の教科書は乏しい。英文の専門書は数多くあるものの、実際の解析に際しては SAS などの統計ソフトを使いことを前提にした記述となり、統計解析を行なう当事者たちにとって最尤法あるいは制限付最尤法を用いる統計ソフトを「ブラックボックス」として使わざるを得ない状況となっている。「続・高橋セミナー」では、新たなテーマのみならず、これまでの高橋セミナーの内容をすべての見直し、再編集を行い、追加・訂正を適宜加えて順次公表して行く。目指すのは「用量反応関係の統計モデルの基礎と応用」である。基礎概念の理解のために Excel による計算を充実させる。第 1 回は、「JMP による各種分割実験入門 – 変量効果モデルの基礎 –」としたが、第 7 章は「Excel ソルバーによる REML 法入門」である。 2011 年 12 月 27 日 高橋 行雄

第 9 回 「最尤法によるポアソン回帰分析入門」公開の経緯

第 9 回に先だって 2019 年 10 月に第 8 回「最尤法による探索的ポアソン回帰」を公開した。これは第 9 回の第 7 章として準備していた内容を、先行して取りまとめたものである。全ての章の原稿が揃い推敲を重ね、2020 年 4 月に第 1 章から順次公開し、7 月の第 13 章で章ごとの公開を完結した。さらに推敲を重ね全章を 2021 年 1 月に一括公開した。ただし、全章をプリントしようとする PDF で全 500 ページとなるため、利用者の適宜を図るために 2021 年 3 月 Amazon が提供しているプリント・オン・デマンドによる出版を行い、紙での提供も合わせて行うことにした。

第 12 回 「層別因子を含む探索的な回帰分析入門」公開の経緯

第 12 回に先だって、2022 年 4 月に第 10 回「線形モデルによる欠測値がある直交表の解析 – 謎めいた最小 2 乗平均と 95%信頼区間の活用 –」を公開した。これは第 12 回の第 4 章として準備していた内容を、先行して取りまとめたものである。また、2023 年 3 月に第 11 回「第 11 回「JMP で繰返しが不揃いの 2 元配置データの解析ができるの? – 平方和の分解ではなくデザイン行列と最小 2 乗平均の活用 –」」を公開した。これは第 12 回の第 3 章として準備していた内容を、先行して取りまとめたものである。全ての章の原稿が揃い推敲を重ね、2023 年 12 月に第 1 章から順次公開し 2024 年 3 月に第 13 章で章ごとの公開を完結した。さらに、推敲を重ね全章を一括公開し、第 9 回「最尤法によるポアソン回帰分析入門」と同様に Amazon が提供しているプリント・オン・デマンドによる出版による提供も合わせて行う予定である。

非売品、無断複製を禁ずる

第 12 回 続高橋セミナー 層別因子を含む探索的な回帰分析入門

2024 年 6 月 高橋 行雄

BioStat 研究所(株)

〒105-0014 東京都港区芝 1-12-3 の 1005

takahashi.stat@nifty.com , FAX : 03-3452-8035

層別因子を含む探索的な回帰分析入門 目次

0.	はじめに -----	1
1.	層別因子を含む各種の回帰分析の実際 -----	7
1.1.	年収の比較調査データ -----	7
	別々の回帰直線のあてはめ, デザイン行列 X を用いた単回帰分析, デザイン行列 X を用いた 3 本の回帰直線の同時あてはめ, 共通の傾きを 持つ 3 本の回帰直線, 交互作用の検討, 回帰直線上の推定値にする 95% 信頼区間, 会社間の推定値の差の 95%信頼区間, デザイン行列とは何か	
1.2.	季節による洗浄用水の温度と回収液の濃度の関連 -----	19
	季節による層別, JMP の「グラフ・ビルダー」による層別散布図, JMP の 「モデルのあてはめ」による解析, JMP の「予測プロファイル」, 対比による 水準間の差の推定, Excel による (1, -1) 対比型ダミー変数を用いた解析, (1, -1) 対比型ダミー変数における回帰直線の推定式, Excel による予測 プロファイル, 水準間の差の予測プロファイル	
1.3.	前値を共変量とする交互作用がある抗うつ剤の評価 -----	30
	非平行な 2 本の直線, 推定されたパラメータによる交互作用の検討, 分散分析表による交互作用の検討, 非平行な 2 本の回帰直線の差, 2 本の直線間の差に対する 95%信頼区間および t 検定, 2 本の回帰 直線が有意な差となる前値 x_0 , JMP による検証	
1.4.	発電プラント周辺でのカキの成長 -----	38
	前値を共変量とする対照群を含む 2 元配置, 共変量を含む 1 元配置として モデル化, 1 元配置の水準の組み合わせ平均, 1 元配置を 2 元配置として 再構築, SAS による検証	
1.5.	除草剤の散布による雑草の成長抑制 -----	47
	MCP 剤に対する共通の切片を持つ回帰直線, PCP 剤に対する原点を 共通とする回帰直線, R による検証(原点共通の場合)	
2.	デザイン行列を活用した 1 因子実験データの解析 -----	55
2.1.	繰り返しが等しい 1 因子実験データ -----	55
	各種の平方和の分解による分散分析表, ダミー変数を用いた線形モデル, デザイン行列 X を用いた計算の実際, データの構造式における効果の推定,	

目次

水準平均の 95%信頼区間, 水準間の差と 95%信頼区間, 水準平均に対する 伝統的な分散の推定, 水準平均の差とその分散の推定, 水準効果の分散 の推定, 分散の加法性が成り立たない効果の差, セル平均モデル	
2.2. 繰り返しが不揃いな 1 因子実験データ -----	69
平方和の分解, デザイン行列 X を用いた回帰分析, データの構造式に おける効果, 水準平均に対する 95%信頼区間, 水準平均の差に対する 95%信頼区間, 分散の加法性での対応と限界	
2.3. (0, 1)型ダミー変数を用いた 1 因子実験 -----	77
(0, 1)型ダミー変数のデザイン行列 X を用いた場合, (0, 1)型ダミー変数 場合の 95%信頼区間	
2.4. 1 因子実験の量的変数に対する多項式回帰 -----	80
デザイン行列 X を用いた多項式回帰, 3 次式のあてはめ, 2 次式 のあてはめ, デザイン行列 X を用いた単回帰分析, あてはまりの悪さ LOF 解析, LOF 解析に代わる逐次平方和(タイプ I の平方和)	
3. 繰り返しが不揃いの 2 因子実験データの解析 -----	89
3.1. 繰り返しが等しい 2 因子実験 -----	89
平方和の分解による分散分析表の作成, 修正項 CT を用いた平方和 の計算, 推定値に対する 95%信頼区間のグラフ表示, 推定値の差に 関する 95%信頼区間のグラフ表示, (1, -1) 対比型ダミー変数, デザイン行列 X を用いた回帰分析の適用, 要因効果に対する推定と 95%信頼区間, 効果 α などに対する有効反復数 n_e を用いた分散の推定, Excel の回帰分析を用いた分散分析表の作成	
3.2. 繰り返しが不揃いの 2 因子実験 -----	102
繰り返しが不揃いの 2 因子実験データ, きれいな層別散布図の作成手順, データの構造式, 偏差平方和の計算, デザイン行列 X を用いた回帰分析, 分散分析表の作成, 最適水準と現行水準の差の推定と 95%信頼区間, 啓蒙活動	
3.3. タイプ I, タイプ II, タイプ III の平方和 -----	111
主効果を主体にしたタイプ II の平方和, 主効果と交互作用も平等に扱う タイプ III の平方和, SAS の GLM プロシジャによるタイプ I, タイプ II, タイプ III の平方和, 対比によるタイプ III の計算, タイプ III の平方和の使い方, 現行水準と最適水準の差の 95%信頼区間	

4.	欠測値がある直交表の線形モデルによる解析 -----	121
4.1.	繰り返しのない 2 因子実験データ -----	121
	構造モデル・回帰モデル・線形モデル, 線形モデルに対する正規方程式, デザイン行列 X を用いた正規方程式, 正規方程式の解, パラメータの 共分散行列の活用, 線形和の推定と 95%信頼区間, 組合せ水準の 95%信頼区間, 因子の水準間の差の推定および 95%信頼区間, 現行水準と最適水準の差および 95%信頼区間	
4.2.	直交表に対する線形モデル -----	135
	直交表を用いた実験データの解析, 直交表における構造 (DE) モデルと線形モデル	
4.3.	欠測値がある直交表に対する線形モデルの活用 -----	138
	直交表による実験で欠測値が生じた場合, Excel の回帰分析を用いた パラメータの推定, 欠測値がある場合の各種の推定	
4.4.	JMP の「モデルのあてはめ」による解析 -----	144
	(1, -1) 対比型ダミー変数, 予測プロファイルを用いた推定, カスタム 検定による最適水準と現行水準の差の推定, 最小 2 乗平均, 各種の ダミー変数の歴史的背景, JMP の使い勝手	
4.5.	SAS/GLM による欠測値がある直交表の解析 -----	150
	GLM プロシジャによる解析プログラム, GLM プロシジャの出力, Lsmeans ステートメントによる最小 2 乗平均, Estimate ステートメントによる 最小 2 乗平均, (1, 0) 型および (0, 1) 型ダミー変数の扱いにくさ	
5.	デザイン行列を用いた回帰分析の基礎 -----	155
5.1.	回帰分析における各種の推定 -----	155
5.2.	デザイン行列を用いた行列計算の基礎 -----	157
	Excel シート上にデザイン行列 X を設定, 行列計算の実際, デザイン行列の 転置と積和, シグマを用いた積和の計算, デザイン行列 X の積和, デザイン 行列 X と反応ベクトル Y との積和, 英語版 Wikipedia に見るデザイン行列の定義	
5.3.	偏差平方和による回帰パラメータの推定 -----	164
	回帰パラメータの推定, 正規方程式, 正規方程式の解による パラメータの推定, 偏差平方を用いたパラメータの推定の実際	
5.4.	デザイン行列 vs. 偏差平方和を用いた回帰分析 -----	169
	行列計算によるパラメータ推定, デザイン行列 X と偏差平方和での 推定式の相違, 偏差平方和ベースのパラメータの分散の推定, シグマを用いた回帰分析の功罪	

5.5.	Excel の行列計算による回帰分析の実際-----	174
	Excel の行列計算による回帰パラメータの推定, 回帰パラメータの 分散および共分散の推定, Excel の関数を用いた分散分析表の作成, パラメータの共分散行列の活用, 回帰直線の 95%信頼区間, 伝統的な方法, 現実的な対応, 平方和の分解に対する補足	
5.6.	逆推定値に対する各種の 95%信頼区間の推定 -----	183
	逆推定とは何か, デルタ法による逆推定値に対する近似 95%信頼区間, 逆推定値に対する正確な 95%信頼区間, 個別データに対する逆推定値の 正確な 95%信頼区間, Excel ソルバーを用いた逆推定の正確な 95%信頼区間, 面積の標準偏差を推定したい	
5.7.	JMP の回帰分析による逆推定 -----	190
	JMP の「二変量の関係」による回帰分析, 回帰直線の 95%信頼区間の 計算式, 「モデルのあてはめ」による逆推定値の 95%信頼区間, 逆推定値の 95%信頼区間の直接推定	
6.	伝統的な共分散分析からの脱却-----	195
6.1.	伝統的な共分散分析の解析手順 -----	195
	季節ごとの回帰直線, 重層的な回帰分析の適用, 原点を通る回帰直線, 交互作用を含む分散分析表の作成, 平行な回帰直線のあてはめ, 平行な直線間の差の 95%信頼区間, 共分散分析は難解なのか?	
6.2.	Excel の回帰分析を活用した共分散分析 -----	204
	(0, 1)型ダミー変数を用いたデザイン行列 \mathbf{X} , 逐次平方和による分散分析表, 共通の傾きを持つ場合の各種の推定, 水準間の差の予測プロファイル, 洗浄水の温度 x に関する予測プロファイル	
6.3.	SAS の GLM プロシジャによる共分散分析-----	211
	SAS でのテキストデータの取り込み, GLM プロシジャによる交互作用 $x \times A$ の 評価, GLM プロシジャが内部で生成するデザイン行列 \mathbf{X} , 交互作用を 含まない解析, Lsmeans ステートメントによる最小 2 乗平均の推定, Estimate ステートメントによる最小 2 乗平均の差の推定	
7.	共変量を含む 3 因子実験データの探索的解析 -----	219
7.1.	Excel の散布図を主体にした問題解決-----	219
	(2×2)格子状の層別散布図に回帰直線の重ね書き, 共変量 X の分布の 均一性の確認, (2×2)格子状の層別散布図の併合	

7.2.	回帰直線に対する個別データの95%信頼区間 -----	224
	交互作用を含む線形モデル, 個別データの95%信頼区間, 色B別の個別データの95%信頼区間	
7.3.	Excelの回帰分析による交互作用を含む変数減少法 -----	230
	ダミー変数, 交互作用を含む線形モデル, 変数選択の手順, 推定されたパラメータを用いた層別回帰直線の推定	
7.4.	JMPのステップワイズ法による変数選択 -----	236
	手作業での変数選択, 自動的な変数選択, 質的変数と量的変数の 交互作用を含む変数選択, 変数選択の結果をさらに深掘する	
8.	交絡変数と共変量を含む2群比較 -----	243
8.1.	体重の増加抑制は総摂餌量の影響だけか -----	243
	体重の増加抑制, 2群間のt検定を使った検討, 体重の増加量 および増加率, 前後差か増加率かの判断基準, 交絡変数・総摂餌量 の影響, 共変量・前値の影響, 共変量・胎仔数の影響	
8.2.	Excelによる複数の共変量を含む探索的な回帰分析 -----	250
	前後差に対する共変量としての前値. 複数の共変量との交互作用, 複数の共変量を考慮した(2×2)格子状の散布図	
8.3.	JMPによる複数の共変量を含む探索的な回帰分析 -----	255
	ステップワイズ法による変数選択, 「標準最小2乗法」による予測プロファイルの活用	
8.4.	交絡変数の多彩な影響への対応 -----	260
	投与終了時の体重は交絡変数, 解析手法の選択のための決定樹, 1) XとYに相関があるか, 2) 2群の回帰直線が同一線上か, 3) 2本の直線の傾きは同じか, 4) 回帰直線が原点を通るか, 5) 2本の直線が原点を通らず傾きが同じでない場合	
8.5.	体重を交絡変数とする回帰分析 -----	268
	群により回帰直線の傾きが異なる場合, 体重の総平均での臓器 重量の比較, 予測プロファイル作成, 回帰直線が同一直線上と 見なせる場合, 2本の回帰直線が原点を通る場合, 交絡変数を常に念頭にした現実的な対応	

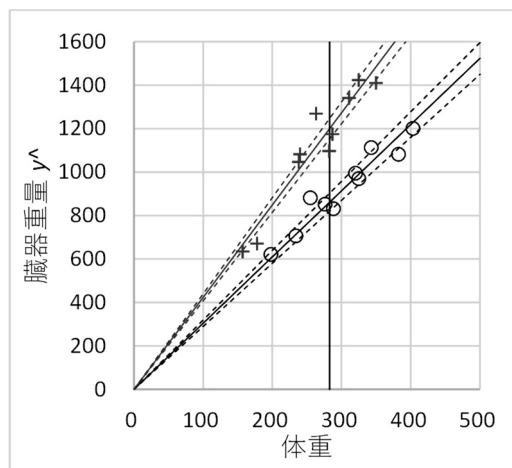
9.	前後差の前値は常に共変量-----	281
9.1.	「平均への回帰」をめぐる葛藤-----	281
	多様な揺らぎ, 前後差での群間比較, 癖のある共変量	
9.2.	前後差 D に潜む「平均への回帰」の可視化-----	283
	Excel による正規乱数の発生, 前値 X の揺らぎの可視化, 個体内の 揺らぎに起因する時点間の相関, 前後差 D の分散のインフレーション, 時点間の相関の期待値が $\rho=0.3$ の場合	
9.3.	後値 Y および前後差 D に対する群間比較-----	290
	「平均への回帰」現象, 1) 前後差 D にのみ偶然に有意な差が出る場合, 2) 後値 Y にのみ偶然に有意な差が出る場合, 3) 後値 Y および前後差 D 共に有意な差が出る場合	
9.4.	前値 X を共変量とした後値 Y および前後差 D の解析-----	294
	前値を共変量とする解析, 1) 前後差 D にのみ偶然に有意な差が出る場合, 共分散分析における最小 2 乗平均, 最小 2 乗平均の 95%信頼区間 2) 後値 Y にのみ偶然に有意な差が出る場合, 交互作用を含む共分散分析, 交互作用がある場合の最小 2 乗平均, 3) 後値 Y および前後差 D 共に有意な差が出る場合, JMP による主効果モデル, JMP による交互作用モデル, 中心化された場合の最小 2 乗平均, JMP の解析結果を Excel でグラフ化, 交互作用を含める解析の妥当性	
9.5.	前値 X と後値 Y をめぐる各種の群間比較の第 1 種の過誤 (α エラー)-----	310
	1: ($\bar{X}_{A_1} > \bar{X}_{A_2}$) 場合の α エラー, 2: ($\bar{X}_{A_1} \gg \bar{X}_{A_2}$; 相関が大) 場合の α エラー, 3: ($\bar{X}_{A_1} \gg \bar{X}_{A_2}$; 相関が小) 場合の α エラー, 4: ($\bar{X}_{A_2} \approx \bar{X}_{A_1}$) 場合の α エラー, 5: ($\bar{X}_{A_1} < \bar{X}_{A_2}$) 場合の α エラー, SAS によるシミュレーションによる第 1 種の過誤の検討	
9.6.	前値 X と後値 Y めぐる各種の群間比較の検出力 ($1-\beta$ エラー)-----	318
	1: 前値 X が ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合の検出力, 相関の期待値: $\rho=0.90$, 相関の期待値: $\rho=0.10$, 2: 前値 X が ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) の場合の検出力, 3: 前値 X が ($\bar{X}_{A_1} < \bar{X}_{A_2}$) と逆に大きい場合の検出力, 相関の期待値: $\rho=0.90$, 相関の期待値: $\rho=0.10$, SAS によるシミュレーションによる検出力 ($1-\beta$) の検討	

10. 層別因子を含むロジスティック曲線のあてはめ -----	327
10.1. カドミウムガス曝露後の肺活量の減少-----	327
デザイン行列を用いた3本の回帰直線の同時あてはめ, 3本の回帰直線の傾きの差に対する t 検定, 残差平方和の差から交互作用の平方和の算出, 回帰直線の傾きの差を用いた検討, 回帰直線の差の95%信頼区間, スプライン曲線による直線のあてはめの妥当性チェック	
10.2. 複数のロジスティック曲線の同時あてはめ-----	335
ロジスティック曲線のあてはめの基礎, ソルバーによる残差平方和の最小化, パラメータの初期値の探索的な推定, パラメータの初期値の設定の正攻法, Excelのソルバーの活用法	
10.3. ロジスティック曲線の95%信頼区間-----	342
ロジスティック曲線式のパラメータに関する偏微分, パラメータに関する95%信頼区間, ロジスティック曲線の95%信頼区間, ロジスティック曲線の差の95%信頼区間	
10.4. JMPによる3本のロジスティック曲線の同時あてはめ-----	349
JMPによるロジスティック曲線のあてはめ, JMPによる非線形回帰, ロジスティック曲線の95%信頼区間, JMPによるロジスティック曲線式のパラメータに関する偏微分	
10.5. SASの非線形NLINプロシジャによる解析-----	358
SAS/NLINプロシジャによる解析, NLINプロシジャの出力結果をExcelの折れ線グラフで	
10.6. 残差線形化法によるロジスティック曲線のパラメータ推定-----	363
残差を反応とした回帰分析の反復, Wikipediaのガウス・ニュートン法	
11. 各種のシグモイド曲線を用いた逆推定 -----	367
11.1. 累積分布関数のシグモイド曲線としての活用-----	367
正規分布, ロジスティック分布, パラメータの線形化, 累積ロジスティック分布関数の量的な反応への拡張, 酵素反応速度論のヒルの式, ゴンペルツ・最大極値分布, ワイブル・最小極値分布	
11.2. ロジスティック・ゴンペルツ・ワイブル曲線のあてはめ-----	379
3種のシグモイド曲線のあてはめ, 10パーセント反応の用量 D_{10} の逆推定, 50パーセント反応の用量 D_{50} の推定, 正規分布 vs. ロジスティック分布	
11.3. ゴンペルツ・最大極値曲線に対する各種の95%信頼区間-----	385
残差線形化法によるパラメータ推定, ゴンペルツ・最大極値曲線の	

	95%信頼区間, 10パーセント反応の逆推定値の95%信頼区間, Excel および統計ソフトの脆弱性	
11.4.	JMP によるゴンペルツ・最大極値曲線の推定および逆推定 -----	392
	JMP の「曲線のあてはめ」, 「非線形回帰」によるゴンペルツ・最大極値曲線の あてはめと逆推定, シグモイド曲線の95%信頼区間	
11.5.	オフセットを活用したシグモイド曲線に対する逆推定 -----	397
	オフセットを活用した逆推定, SAS の非線形 NLIN プロシジャによる解析, ラットの子宮重量の SAS データセット化, オフセット無しの解析・オフセットを 用いた逆推定, ゴンペルツ・最大極値曲線の95%信頼区間, SAS の結果は正しいのか?	
12.	ミカエリス・メンテン式をめぐる新たな統計解析 -----	405
12.1.	ミカエリス・メンテン式によるパラメータ推定 -----	405
	線形化によるパラメータの推定精度, a) 非線形回帰によるミカエリス・メンテン式 のあてはめ, b) 両逆数・Lineweaver-Burk 変換, c) Hanes-Woolf 変換, d) Scatchard・Eadie-Hofstee 変換, 非線形回帰に対する線形化のパラメータの 推定精度の比較, 文献の推定値との差異	
12.2.	統計ソフト SAS によるミカエリス・メンテン式のパラメータ推定 -----	421
	非線形回帰のパラメータ推定値の SAS による検証, パラメータの相関行列と 共分散行列の相互関係, ミカエリス・メンテン曲線の95%信頼区間, 非線形 回帰による Hanes-Woolf 変換によるパラメータの直接, 誤差の伝播式を 用いた傾き β_1 の逆数の標準誤差	
12.3.	ミカエリス・メンテン式に基づく阻害定数 K_i の推定 -----	428
	阻害定数 K_i の推定方法による不一致, 酵素阻害実験データ, 阻害定数 K_i の推定方法, 両逆数・Lineweaver-Burk 変換によるミカエリス定数 $(K_m)_i$ の推定, 阻害定数 K_i の揺らぎの実際, 傾き $(K_m)_i / (V_{max})_i$ を用いた 阻害定数 K_i の推定	
12.4.	最大反応 V_{max} を共通とした場合の阻害定数 K_i の推定 -----	439
	(1, 1) 標示型ダミー変数の設定, Excel のソルバーを用いた非線形回帰, 阻害定数 K_i の推定方法による揺らぎ	
12.5.	非線形回帰による阻害定数 K_i の直接推定 -----	443
	阻害定数 K_i を直接推定するための式の導出, NLIN プロシジャによる 阻害定数 K_i の直接推定, 両逆数・Lineweaver-Burk 変換による阻害定数 K_i の 推定結果との比較, Excel の行列関数を用いた阻害定数 K_i の直接推定, Excel のソルバーを用いた阻害定数 K_i の直接推定	

12.6. 阻害様式の確認のための手順 -----	451
Excel ソルバーを用いた個別パラメータの一括推定, NLIN プロシジャによるパラメータの一括推定, 阻害様式の判別, 線形変換で悩むのは終わりにしませんか	
13. ロジスティック曲線のさらなる活用 -----	455
13.1. 陰性対照と陽性対照を含むロジスティック曲線のあてはめ -----	455
内臓痛試験データ, 非線形モデルを用いたパラメータの同時推定, ダミー変数を用いた非線形の解析モデル, JMP の「非線形回帰」, Excel のソルバーを用いたパラメータの推定, SAS の NLIN プロシジャによる解析, NLIN プロシジャによるパラメータに関する偏微分, 残差線形化法によるロジスティック曲線のあてはめ	
13.2. 平行でない直線のあてはめによる効力比の推定 -----	469
血圧の上昇作用の抑制, 用量別の比較, 同一の反応となる用量の比較, 効力比 ρ の 95%信頼区間,	
13.3. 平行な直線のあてはめによる効力比の推定 -----	475
切片が異なる平行な直線のあてはめ, SAS の NLIN プロシジャによる効力比の直接推定, 平行でない回帰直線 vs. 平行な回帰直線	
13.4. 平行なロジスティック曲線による効力比の推定 -----	481
直線のあてはめは便宜的, ロジスティック曲線による効力比の直接推定, NLIN プロシジャによるパラメータの標準誤差と 95%信頼区間, Excel による効力比の 95%信頼区間の推定, 線形モデルと非線形モデルの解法の対比	
文献, 文献索引, 索引, 解析ファイル一覧 -----	489
文献 -----	491
文献索引 -----	495
索引 -----	497
解析ファイル一覧 -----	523

原点を通る2本の回帰直線の95%信頼区間



0. はじめに

量的な説明変数と反応の関係について、直線をあてはめる回帰分析がよく知られている。反応に影響を与える質的な説明変数も考慮したい場合には、それぞれの水準（カテゴリ）ごとに直線をあてはめ、反応に対する関係を仔細に検討したい。複数の直線が平行とみなせる場合、傾きが異なるが直線が互いに交わらない場合、直線がある一点を通るように見える場合、直線が互いに交差する場合など、千差万別である。このような場合について、どのような解析を行なったら良いのであろうか。

得られたデータから回帰直線のパラメータ（切片と傾き）を求め、それらの標準誤差を推定し、統計的に評価する方法は、良く知られており、複数の回帰直線を別々に求めることも容易である。さて、切片は異なるが傾きは共通の平行な直線をあてはめて、それらのパラメータと標準誤差を求めたいが、どのようにしたら良いのであろうか。また、切片あるいは原点を共通とする傾きが異なる 2 本の回帰直線をあてはめ、それらのパラメータと標準誤差を求めたいが、どのようにしたら良いのであろうか。

説明変数に質的な変数と量的な変数の両方を含み、反応が量的変数として得られた場合の統計解析法とし、共分散分析が連想される。だが、共分散分析について解説されている成書は、なかなか見出しにくい。共分散分析について記述がある成書の多くは、共変量がある 2 群比較を主体にし、解析方法も計算手段が乏しい時代に定式化された単回帰分析を重層的に用いる煩雑な解析手順が示されている。共分散分析は、反応に影響を与える量的な変数（共変量）が存在し、実験の開始直前に測定値を得ることはできるが、制御因子としてコントロールできない変数として定義されている。共分散分析は、このような共変量の影響を統計的に取り除き、制御因子としての質的な変数についての精度を高めた群間比較をするための方法として定式化されている。そのために、共変量 X と質的な変数 A との交互作用 $X \times A$ が無いことが大前提とされている。これは、交互作用が有ると共変量の影響を統計的に除いた群間比較ができなくなるためである。

量的な説明変数と質的な説明変数が混在するようなデータに対し、共分散分析を適用すれば、と思った瞬間に、伝統的な共分散分析の枠組みに絡め取られることになる。その呪縛は、“共変量”にある。共分散分析の目的は、質的な変数の水準（カテゴリ）間の比較であり、邪魔になる“共変量”の影響を統計的に取り除きたいのである。量的な反応に対する回帰分析を行ない、質的な変数で層別（層別因子）したときに、どのような変化が回帰直線に起きる

かを検討したいとの思いは、“共分散分析”を適用すればいいのではないかと思った瞬間に打ち砕かれる。本書の「層別因子を含む探索的な回帰分析入門」は、“共分散分析”をめぐる呪縛にからめとらないような解析法を提示することにある。

単回帰分析は、手計算による偏差平方を主体にした解析方法が手順化され、広く普及している。重回帰分析については、行列計算が前提となり、それを避けるために統計ソフトの使用が推奨され、その結果として推定された回帰パラメータの解釈が主体になりがちである。層別因子を含む回帰分析の基本は、層別因子を複数のダミー変数に置き換え、説明変数に対する回帰分析（重回帰分析）の適用である。身近な Excel の「回帰分析」も 1 変数のみならず複数の変数が扱えるようになっており、「単・重」の区別なく「回帰分析」と言うことにする。

質的変数をダミー変数とし、量的変数との交互作用も扱う場合は、種々の推定を伴うので、「回帰モデル」よりも「線形モデル」として区別する方が適している。「線形モデル」は、行列計算を前提にしており、回帰分析、各種の分散分析、もちろん共分散分析などを統一的に解析できることが知られている。ただし、行列計算を前提にしているので、統計の入門書で取り扱うことが避けられてきた。幸いなことに、身近な Excel で「線形モデル」の解析に必要な行列関数が使用できるようになり、Excel の分析ツールの「回帰分析」と組み合わせて使い、結果のグラフ化につなげることにより、統計ソフトによる解析に勝るとも劣らないことを実感してもらいたいと願っている。

「層別因子を含む探索的な回帰分析入門」と題したのは、回帰分析、分散分析、共分散分析を統合した解析を目指すためである。そのために、取り上げる解析事例については、様々な成書の中から事例をよりすぐり、線形モデルによる解析手順で統一することにした。線形モデルは、行列計算を基本としているので、「シグマを見ただけで毛嫌いされ、ましてや行列の計算式があると見向きもされない」のだと聞かされ続けた先人たちの教訓を念頭にしつつ、「行列計算は簡単でパワフルで簡単」と思ってもらえるように工夫を重ねている。

取り上げた適用事例のデータは、Excel シート上に取り込み、形を整えて Microsoft の Word に「図（拡張メタファイル）」に張り込んでいる。読者は、別途 Web 上に公開している Excel シートを開くと、寸分たがわない解析用データを目にすることができる。さらに、同寸大の層別散布図を Excel シート上に見いだすこともできる。本書に示してあるほとんどの表および図は、Excel で作成したものを Word に張り付けたものである。また、様々な計算を Excel シート上で行っており、数式タブでの（数式の表示、参照元のトレース、参照先のトレース）などを用いて、どのような計算が行われているのかを表示することができる。

本書は、読者が偏差平方による単回帰分析について、統計ソフトなどを用いた回帰分析を行なった経験があることを前提としている。だが、線形モデルの解析におけるデザイン行列については、まったく知らないし、行列の式を見るのも嫌だ、との思いを持っている読者を念頭にしている。さらに、回帰分析により、説明変数 x と反応 y を使い、切片と傾きを推定する計算式は見たことがあるが、それらの標準誤差は統計ソフトの出力を用いているので、計算方法はよく知らない、と思われている読者を想定している。

第 1 章は、層別因子を含む回帰分析の典型的な事例についての解析の手順、解析結果の解釈のために必要となる推定法を示している。第 1.1 節の「年収の比較調査データ」では、1 本の回帰直線の切片と傾きを求めるためのデザイン行列の作成から始め、3 本の回帰直線の同時あてはめに必要なダミー変数の作成方法を説明し、様々な線形モデルをデザイン行列に反映する考え方を示している。さらに、平行な 3 本の回帰直線をあてはめるためのデザイン行列の作成方法を示し、線形モデルによる解析方法を導入している。第 1.2 節以後、様々な事例を取り上げ線形モデルによる解析方法を展開している。第 1 章は、想定している読者層に難しいと思われるが、数式などは読み飛ばし、事例の場面設定に対する結果の解釈に慣れ親しんでもらいたい。

Excel での行列計算になじめない場合は、第 5 章から始めてもらいたい。偏差平方和を積み上げる単回帰分析の解析方法の復習から始め、デザイン行列を用いた行列計算による単回帰分析を導入する。そのために、行列計算の基礎の基礎から始める。Excel シート上の長方形（矩形）を太線で囲んだの数値の集まりを“行列”とし、行列の和、行列の積などの基本演算の練習から始め、行列計算を自ら行なう切っ掛け作りを意図している。「行列の計算式があると見向きもされない」のは、抽象的で具体性に欠けたものに興味を持たないと思われている人達を念頭にしている。第 1 章でいきなりデザイン行列から始めたのは、多彩な「層別因子を含む回帰分析」の解析に際し、行列の計算式を使う線形モデルの解析法の力強さを実感してもらいたいためである。

第 2 章から第 5 章は、基本的な線形モデルによる解析法のコツと着眼点を習得してもらいたいとの願いを込めている。そのために、従来の解析方法で簡単に解ける事例を先に提示し、線形モデルを用いた解析法を導入している。次いで、従来の解析方法では解くことができないが、線形モデルを適用すれば、容易に解析できる事例を示している。

第 2 章では、デザイン行列に慣れ親しんでもらうことを意図している。1 因子実験データの平方和の分解による解析法に引き続き、ダミー変数を用いたデザイン行列を用い、Excel の回帰分析を用いた線形モデルによる解析法を導入する。さらに、量的変数に対する 1 因子実

験データに多項式をあてはめる事例を取り上げる。何次式が統計的に妥当なのかの探索方法を示すことにより、デザイン行列を用いた解析法の有用性を感じてもらいたい。

第3章では、繰り返しが不揃いの2因子実験データを取り上げる。2因子実験は、実験計画法の代表例であるが、繰り返しが揃っていることを前提にしている。繰り返しが不揃いの場因に、平方和の分解による分散分析表が作成できないことを例示し、デザイン行列を用いた線形モデルによる解析方法を例示する。繰り返しが不揃いの場合には、解析の考え方により分散分析表の平方和に違いが生ずるので、それらの使い分けについて示す。

第4章は、欠測値がある直交表と題し、線形モデルの基本である正規方程式を取り上げている。まず、繰り返しが無い2因子実験におけるデザイン行列を用いた正規方程式を導入し、行列計算による正規方程式の解の求め方を示している。次いで、直交表における構造モデルを線形モデルに展開し、(1, -1)対比型ダミー変数によるデザイン行列の作成法を示している。さらに、線形モデルによる解析が必要不可欠となることを、欠測値がある直交表の事例を用いて例示している。

第5章では、単回帰分析を例にしたExcelによる行列計算の入門に引き続き、パラメータの共分散行列を活用した逆推定を取り上げている。逆推定は、単回帰で推定された切片と傾きの比となるのであるが、その95%信頼区間を求めるためにパラメータの共分散行列を活用した誤差の伝播式(デルタ法)について詳しく示している。また、逆推定が標準装備されているJMPによる解析事例を示し、Excelによる結果と対比することにより、推定結果の相互検証を行なっている。

第6章では、先人たちの努力の結晶とも言うべき共分散分析の解析手順を、Excelの力を借りて再現する。この手順は、単回帰分析を重層的に用い「共分散分析」を扱っている成書に共通しているが、線形モデルによる解析法とはなっていない。そこで、生成されたデザイン行列に対し、Excelの回帰分析による(回帰モデル、主効果モデル、交互作用モデルを順次適用することにより共分散分析の分散分析表の作成方法を示す。さらに、共変量の影響を統計的に取り除いた推定値に関する群間比較のために必要となるパラメータの共分散行列の活用実現方法を丁寧に示す。Excelでの解析結果の検証のため、SASのGLMプロシジャを用いた解析法方法も合わせて示す。

第7章は、共変量を含む3因子実験データの探索的解析である。共分散分析は、共変量と実験因子との間で交互作用がないことを前提にし、実験因子の水準間の群間比較の精度を上げる方法として定着しているので、交互作用ある場合への対応方法が未整備となっている。

この事例は、3 因子の内の 1 因子と共変量間に交互作用が出てしまう事例であり、Excel の散布図を主体にした探索的な解析法に対応し、多くの交互作用を含むデザイン行列に対し、探索的な変数選択の実施例となっている。さらに、交互作用を含む変数増減法（ステップワイズ法）が行える JMP による解析事例も取り扱っている。

第 8 章は、共変量だけではなく交絡変数を含む解析事例を取り上げている。交絡変数を含む解析は、共分散分析を適用すればいいのではないかと短絡的に考えられがちである。実験に取り上げた因子の影響を全く受けないのが共変量の定義である。交絡変数は、実験因子の影響を受け、その結果として、反応に対し影響を間接的に与える。そのために、2 本の直線がクロスする場合のみならず、原点を通る 2 本の直線となる場合、2 本の直線が平行だが完全に分離してしまう場合など、共分散分析で想定外の事例がおきてしまう。このような交絡変数に対する解析の考え方と解析方法を総合的に示す。

第 9 章は、経時測定された前値と後値に関連する統計的な諸問題を取り扱う。実験前に 2 群間の前値の平均値を揃える群分けがなされない場合、2 群間の前値の平均値に有意な差が出てしまう場合もあることを Excel によるシミュレーションにより実感してもらうことから始める。次いで、平均への回帰現象が引き起こす不可解な現象について Excel によるシミュレーションを用いて可視化する。前後差による群間比較、後値のみによる群間比較の危うさを時点間の相関に注目し、 α エラーと $(1-\beta)$ 検出力を用いて浮き彫りにする。

第 10 章から第 13 章は、回帰直線のあてはめに無理があるような事例に焦点をあて、直線ではなく層別因子を含むシグモイド状のロジスティック曲線などのあてはめを扱う。シグモイド曲線状となる量的反応に対し、非線形回帰によりパラメータの推定が行えるのことは知られてはいるが、非線形回帰そのものがブラックボックス的であり、一般的には敬遠され続けているのが現状である。そのために、本書では、各種の事例に対する非線形回帰の解析結果を示すだけでなく、誰にでも手軽に量的反応に対しシグモイド曲線のあてはめができるように Excel のソルバーを活用した解析法も提示する。

第 10 章は、層別因子を含む回帰分析を行なったところ無視できない交互作用が存在し、結果の解釈を試みるが、隔靴搔痒的な解釈となる事例を取り上げる。この事例は、直線のあてはめ自体が便宜的であり、ロジスティック曲線のあてはめが本質的な事例である。ロジスティック曲線のあてはめは、非線形モデルの統計ソフトになじみがある人にとっては、取っつき易いのであるが、そうではない人たちにとっては、敷居がとっても高く感じられる。ここで取り上げるのは、最大値が共通で、異なる最小値を持つ 3 本のロジスティック曲線の同時あてはめである。Excel のソルバーを用いることにより、誰にでも自由自在に様々な曲線のあては

めを可能にすることを例示する。理解を深めるために、JMP および SAS の非線形回帰を用いた解析結果も合わせて示し、Excel の行列関数を用いた非線形回帰についても例示する。

第 11 章は、量的な反応データにシグモイド曲線をあてはめ、ある反応となる用量を逆推定する各種の方法を示す。各種の累積分布関数の (0~1) の範囲でのシグモイド曲線を、(最小値~最大値) に拡張する。量的な反応に対する 3 種のシグモイド曲線 (ロジスティック曲線, ゴンペルツ曲線, ワイブル曲線) をあてはめ、Excel のソルバーを用いて残差平方和によるあてはまりの良さを評価する。あてはまりの良かったゴンペルツ・最大極値曲線に対し、Excel による非線形回帰を線形化 (ガウス・ニュートン法) により行い、求められた曲線の 95% 信頼区間から、10% 反応量に対する逆推定値の 95% 信頼区間を求める方法を示した。さらに、JMP の「曲線のあてはめ」、「非線形回帰」を用いた逆推定について紹介し、SAS の NLIN プロシジャによる逆推定値をとその 95% 信頼区間を直接求めるためのオフセットの適用方法を示す。

第 12 章は、非線形モデルの典型例であるミカエリス・メンテン式をめぐる諸問題を扱う。ミカエリス・メンテン式に対して、様々な線形式への変換方が提案され実用的な方法として定着している。回帰係数の比などで計算されるミカエリス定数の標準誤差を、誤差の伝播式を適用することにより求め、非線形回帰の場合と比較検討した結果を示す。さらに、阻害薬を加えた場合にミカエリス定数が変化する。この変化分から阻害定数を算出する場合の誤差の伝播についても明らかにし、非線形回帰の適用の有用性について示す。

第 13 章の最初に取り上げるのは、陰性対象群と陽性対照群を含む実験データにロジスティック曲線をあてはめる非線形回帰の事例である。まず、JMP の非線形回帰、Excel のソルバーを使った簡便法、SAS の NLIN プロシジャによるパラメータに関する偏微分式を活用した Excel による正攻法による解析法を示し、ブラックボックス的な非線形回帰の計算方法を可視化している。最後に取りあげる事例は、平行線検定法による効力比の推定法として知られている典型的なデータを用い、平行線があてはまらないとした場合の効力比とその 95% 信頼区間の求め方、平行とみなした場合の効力比を対比する。さらに、得られたデータにシグモイド曲線状の弱い関連を手掛かりに、2 本の平行なシグモイド曲線をあてはめ、効力比とその 95% 信頼区間をパラメータとして直接推定する方法を SAS の NLIN プロシジャに加え、Excel による非線形回帰も合わせて示す。

1. 層別因子を含む各種の回帰分析の実際

ほとんどの統計の教科書で回帰分析が取り上げられている。ただし、単回帰分析に特化した解析法が示されているだけで、層別因子を含む複数の回帰直線の同時あてはめについて取り上げている成書はごく限られている。そのほとんどは、手計算時代の偏差平方和を主体にした解析法が示されており、技巧的かつ難解であり推奨できない。多くの統計ソフトは、デザイン行列を主体にした行列計算による解析を行なっているが、その解析法はブラックボックス的である。そこで、特徴的な事例が示されている成書の事例について、デザイン行列を用いた回帰分析を主体にし、各種のパラメータの95%信頼区間の推定に、Excelの行列関数を活用することにより、見通しの良い汎用的な解析方法を例示する。

1.1. 年収の比較調査データ

表 1.1 に示すのは、芳賀 (2014), 「医薬品開発のための統計解析 第2部 実験計画法 改訂版, 第4章 共分散分析」に示されているデータで、就職先を検討するための会社 A (A_1, A_2, A_3) の従業員をランダムに 10 人を選び、年齢 X および年収 Y を調査した結果である。

表 1.1 年収の比較調査の結果 [芳賀 (2014), 表示 4.1.1]

	会社A ₁		会社A ₂		会社A ₃	
	年齢	年収	年齢	年収	年齢	年収
No	x	y	x	y	x	y
1	34	684	34	692	34	723
2	33	788	34	712	39	762
3	34	764	37	700	37	883
4	37	836	43	843	29	678
5	29	606	37	748	33	699
6	26	696	32	580	36	830
7	37	766	46	837	41	900
8	38	862	36	667	37	835
9	32	606	41	805	45	905
10	31	708	35	689	39	793
平均	33.10	731.60	37.50	727.30	37.00	800.80
		総平均	35.87	753.23		

別々の回帰直線のあてはめ

年収の平均を比較すると、高い順に (A_3 : 800.8 万円, A_1 : 731.6 万円, A_2 : 727.3 万円) となる。平均年齢が3社で微妙に異なるので、年齢を加味して年収を比較したい。図 1.1 に示すのは、会社 A (A_1, A_2, A_3) 別の Excel で作成した層別散布図に、近似曲線の線形近似を使って回帰直線を重ね書きした結果である。

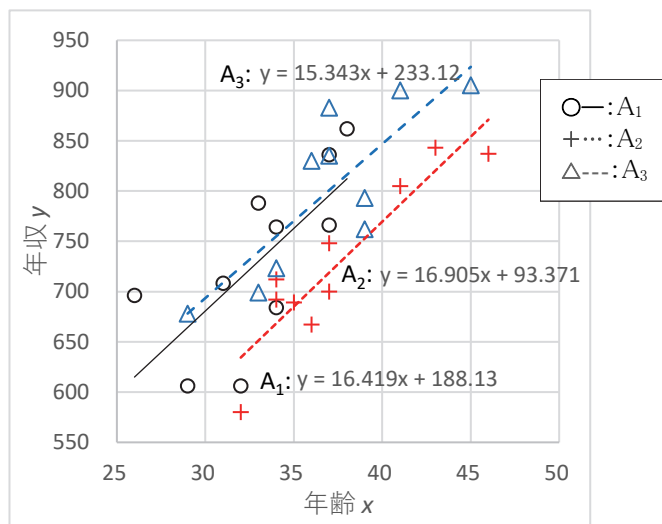


図 1.1 層別散布図による会社別の年齢と年収の関連

図 1.1 から A_3 社の平均年収が 800.8 万円と高いのは、平均年齢が 37.0 歳と A_1 社に比べ高いためと推測される。 A_1 社の平均年収 731.6 万円は、平均年齢が 33.1 歳と低いためであり、同一年齢で比較をすれば A_3 社と同程度の年収であることが、回帰直線が重なっていることから読み取れる。 A_2 社の年収が 727.3 万円と A_1 社と同程度であったが、平均年齢が 37.5 歳と高いためであり、同一年齢に固定して年収を比較すれば、3 社の中で最も低いことが読み取れる。

一般的に同業の会社であっても年齢構成は異なるし、企業規模により同一年齢であっても年収に違いがあることも知られている。図 1.2 に示すように模式的に表せば、会社の違いが年収に影響を与えていることもあれば、会社の違いが年齢構成にも影響を与えることにより、年齢が年収に影響を与えていることもあり、会社間の年収の比較をするためには、年齢の違いを統計的に補正する必要がある。

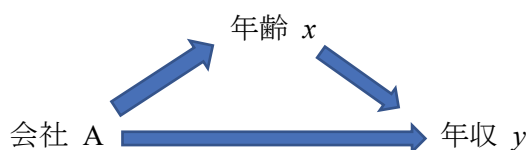


図 1.2 年収に影響をおよぼす要因

デザイン行列 X を用いた単回帰分析

A_1 社についてデザイン行列 X を用いた単回帰分析を行なってみよう. 表 1.2 に示すようにデザイン行列 X は, (切片 x_0 , 年齢 x_1) のように, 切片に対応する変数 $x_0 = 1$ を含んでいる. 回帰分析を行なうための統計ソフトでは, 説明変数として年齢 x_1 のみを指定するのが一般的であり, 切片 $x_0 = 1$ を内部で自動的に付加している. 原点を通る回帰直線を推定したい場合には, 切片を含まない解析が必要となるので, 統計ソフトでは, 「切片の自動付与を抑制」オプションが備わっている.

表 1.2 左に示す (10 行×2 列) のデザイン行列 X をそのまま用いて回帰分析をしたい. 表 1.2 右に示す Excel の回帰分析では, 「定数に 0 を使用」をオン として切片を自動的に付与することを抑制し, 「入力 X の範囲」で変数 x_0 と x_1 の (10 行×2 列) のデータの範囲を選択している. 表 1.2 右上段に示す「分散分析表」の「回帰」の自由度が 2 となっているのは, 変数 x_0 と x_1 の 2 変数に対応し, 「合計」の自由度 10 は, データ数に, 「残差」の自由度は, それらの差 $df = 10 - 2 = 8$ に対応している. 表 1.2 右上段に示す回帰パラメータ (係数) の推定で, 「切片」の係数が「0」となっているのは, 「定数に 0 を使用」をオンとしたためで, まったく意味のない出力である.

表 1.2 A_1 社についてデザイン行列 X を用いた単回帰分析

デザイン行列 X									
	切片	年齢	年収						
No	x_0	x_1	y	分散分析表, 「定数に 0 を使用」を on					
1	1	34	684	自由度	変動	分散	分散比		
2	1	33	788	回帰	2	5387135	2693567	629.9015	
3	1	34	764	残差	8	34209	4276.17		
4	1	37	836	合計	10	5421344			
5	1	29	606						
6	1	26	696	係数	標準誤差	t	P -値		
7	1	37	766	切片	0	#N/A	#N/A	#N/A	
8	1	38	862	$\hat{\beta}_0$	x_0	188.1334	191.7648	0.9811	0.3553
9	1	32	606	$\hat{\beta}_1$	x_1	16.4189	5.7597	2.8506	0.0215
10	1	31	708	回帰の自由度が 2 なのは, x_0 と x_1 の 2 変数に対応					

推定されたパラメータ (係数) から回帰式

$$\left. \begin{aligned} \hat{y} &= \hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 \\ &= 188.1334 x_0 + 16.4189 x_1 \end{aligned} \right\} \quad (1.1)$$

が得られ, 図 1.1 に示した A_1 社に対する近似曲線の重ね書きした回帰式に一致する. パラメータ (係数) の標準誤差 SE は,

$$\left. \begin{aligned} SE(\hat{\beta}_0) &= 191.7648 \\ SE(\hat{\beta}_1) &= 5.7597 \end{aligned} \right\} \quad (1.2)$$

であり, t 値は,

$$\left. \begin{aligned} t(\hat{\beta}_0) &= \frac{\hat{\beta}_0}{SE(\hat{\beta}_0)} = \frac{188.1334}{191.7648} = 0.9811 \\ t(\hat{\beta}_1) &= \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{16.4189}{5.7597} = 2.8506 \end{aligned} \right\} \quad (1.3)$$

として計算されている。これらは、残差の自由度 8 の t 分布に従うことから p 値は、

$$\left. \begin{aligned} p(\hat{\beta}_0) &= 0.3553 \\ p(\hat{\beta}_1) &= 0.0215 \end{aligned} \right\} \quad (1.4)$$

となり、切片 $\hat{\beta}_0 = 188.1334$ は、統計的に有意ではなく、傾き $\hat{\beta}_1 = 16.4189$ は、 $p < 0.05$ と統計的に有意な差となっている。

デザイン行列 X を用いた 3 本の回帰直線の同時あてはめ

単回帰分析を拡張して A_2 社、 A_3 社についても一括して回帰分析によりパラメータを推定する方法はあるのだろうか。図 1.1 に示した 3 本の回帰直線の回帰式を同時推定するために、ダミー変数を用いたデザイン行列 X を生成する。ダミー変数は、表 1.3 に示すように質的変数を何らかの量的変数に置き換えた変数の総称であり、どのような量的変数に置き換えるかは自由であり、解析の目的に合ったダミー変数を生成する必要がある。図 1.1 に示した 3 本の回帰直線の切片と傾きを同時に推定するために表 1.3 に示す 4 種のダミー変数の内から、1) 標示型ダミー変数を選択する。

表 1.3 各種のダミー変数

(1, 1) 標示型				(1, -1) 対比型				(0, 1) 型				(1, 0) 型			
会社A	a_1	a_2	a_3	会社A	a_1	a_2	a_3	会社A	a_1	a_2	a_3	会社A	a_1	a_2	a_3
A_1	1	0	0	A_1	1	0	.	A_1	.	0	0	A_1	1	0	.
A_2	0	1	0	A_2	0	1	.	A_2	.	1	0	A_2	0	1	.
A_3	0	0	1	A_3	-1	-1	.	A_3	.	0	1	A_3	0	0	.

(1, 1) 標示型ダミー変数は、 A_1 社に対し $a_1 = 1$ 、 A_1 社以外は 0、同様に A_2 社に対し $a_2 = 1$ 、 A_2 社以外は 0、 A_3 社に対し $a_3 = 1$ 、 A_3 社以外は 0 とする。表 1.2 では、切片に対する変数として x_0 、年齢を x_1 としたのであるが、会社 A ごとに異なる変数とするために、(1, 1) 標示型ダミー変数 (a_1, a_2, a_3) を会社ごとの切片、会社 A ごとの年齢を (x_1, x_2, x_3) とする。

会社	切片	年齢
A_1 社	a_1	x_1
A_2 社	a_2	x_2
A_3 社	a_3	x_3

表 1.4 左に示すように、これらの 6 変数を (30 行×6 列) のデザイン行列 X とし、Excel の回帰分析で、「定数に 0 を使用」をオンとすることにより、3 本の回帰直線の同時のあてはめを行なうことができる。

表 1.4 会社 A (A₁, A₂, A₃) 別の回帰直線の同時あてはめ

----- デザイン行列 X -----													
No	a ₁	x ₁	a ₂	x ₂	a ₃	x ₃	y	分散分析表, 「定数に 0 を使用」を on					
1	1	34	0	0	0	0	684		自由度	変動	分散	分散比	
2	1	33	0	0	0	0	788	回帰	6	17182504.95	2863750.82	1061.47	
:								残差	24	64750.05	2697.92		
9	1	32	0	0	0	0	606	合計	30	17247255.00			
10	1	31	0	0	0	0	708						
11	0	0	1	34	0	0	692		係数	標準誤差	t	P-値	
12	0	0	1	34	0	0	712	切片	0	#N/A	#N/A	#N/A	
:								$\beta_0^{A_1}$	a ₁	188.1334	152.3196	1.2351	0.2287
19	0	0	1	41	0	0	805	$\beta_1^{A_1}$	x ₁	16.4189	4.5750	3.5889	0.0015
20	0	0	1	35	0	0	689	$\beta_0^{A_2}$	a ₂	93.3714	146.7120	0.6364	0.5305
21	0	0	0	0	1	34	723	$\beta_1^{A_2}$	x ₂	16.9048	3.8877	4.3482	0.0002
22	0	0	0	0	1	39	762	$\beta_0^{A_3}$	a ₃	233.1202	144.9811	1.6079	0.1209
:								$\beta_1^{A_3}$	x ₃	15.3427	3.8932	3.9409	0.0006
29	0	0	0	0	1	45	905	Excelシート上では, 有効数字15桁である					
30	0	0	0	0	1	39	793	「セルの書式設定」で「数値」とし, 小数点以下 4桁で設定					

A₁社に対する変数 (a₁, x₁) は, 他の A₂, A₃社の場合の変数を全て 0 としている. 同様に, A₂社に対する変数 (a₂, x₂) は, 他の A₁, A₃社の場合の変数を全て 0 とし, A₃社に対する変数 (a₃, x₃) は, 他の A₁, A₂社の変数を全て 0 とし, これらをまとめて (30 行×6 列) のデザイン行列 X を設定している. このデザイン行列 X には, すべてのデータに共通の切片がない. (1, 1) 標示型ダミー変数を設定したデザイン行列に対して回帰分析を行なうためには, 「定数に 0 を使用」をオンとする必要がある. 表 1.4 右下段から, 次に示すように 3 本の回帰直線のパラメータが別々に同時推定されている.

$$\left. \begin{aligned} A_1 \text{社: } \hat{y}^{(A_1)} &= \hat{\beta}_0^{(A_1)} a_1 + \hat{\beta}_1^{(A_1)} x_1 = 188.1334 + 16.4189 x_1 \\ A_2 \text{社: } \hat{y}^{(A_2)} &= \hat{\beta}_0^{(A_2)} a_2 + \hat{\beta}_1^{(A_2)} x_2 = 93.3714 + 16.9048 x_2 \\ A_3 \text{社: } \hat{y}^{(A_3)} &= \hat{\beta}_0^{(A_3)} a_3 + \hat{\beta}_1^{(A_3)} x_3 = 233.1202 + 15.3427 x_3 \end{aligned} \right\} \quad (1.5)$$

表 1.4 右上段の「分散分析表」の「回帰」の自由度が 6 となっているのは, デザイン行列 X の 6 変数を「回帰モデル」に含めたことに対応している. また, 「合計」の自由度 30 は, データ数に対応している. A₁社に対する傾き $\hat{\beta}_1^{(A_1)}$ の p 値は, $p = 0.0015$ と A₁社に対する単回帰分析の表 1.2 右下段に示した結果 $p = 0.0215$ に比べて小さくなっている. これは, 分散分析表で示されている誤差分散の推定値 (残差の分散) が $\hat{\sigma}_{(A_1 \text{単独})}^2 = 4276.17$ から $\hat{\sigma}_{(共通)}^2 = 2697.92$ と小さくなったことに起因している.

共通の傾きを持つ 3 本回帰直線

図 1.1 から, 3 本の回帰直線の傾きの差は, ほとんどなく共通の傾きを持つと推測される. そこで, 共通の傾きを持つ 3 本の回帰直線をあてはめ, 35 歳における 3 社の年収を推定し統計的に比較したい.

表 1.5 左に示すように、会社 A (A₁, A₂, A₃) ごとの切片を求めるために (1, 1) 標示型ダミー変数を (a₁, a₂, a₃) とし、年齢は共通の x とするような (30 行×4 列) のデザイン行列 X を設定する。Excel の回帰分析で、「定数に 0 を使用」をオンとすることにより、切片は異なるが、傾きが共通の回帰直線をあてはめることができる。表 1.5 右に示すように、分散分析表の「回帰」に対する自由度は 4 となり、デザイン行列 X で設定した変数の数に一致する。「合計」の自由度は、データ数 30 に一致している。

表 1.5 会社 A (A₁, A₂, A₃) に共通の傾きを持つパラメータの推定

— デザイン行列 X —						分散分析表, 「定数に 0 を使用」を on			
No	a ₁	a ₂	a ₃	x	y	自由度	変動	分散	分散比
1	1	0	0	34	684	回帰	4	17182279.29	4295569.82
2	1	0	0	33	788	残差	26	64975.71	2499.07
:						合計	30	17247255.00	
10	1	0	0	31	708				
11	0	1	0	34	692				
12	0	1	0	34	712	係数	標準誤差	t	P-値
:						切片	0	#N/A	#N/A
20	0	1	0	35	689	$\hat{\beta}_0^{A1}$	a ₁	195.2832	76.7504
21	0	0	1	34	723	$\hat{\beta}_0^{A2}$	a ₂	119.6903	86.5444
22	0	0	1	39	762	$\hat{\beta}_0^{A3}$	a ₃	201.2918	85.4293
:						$\hat{\beta}_1^{共通}$	x	16.2029	2.2690
30	0	0	1	39	793			7.1409	0.0000

推定されたパラメータから、次に示すように傾きが $\hat{\beta}_1^{共通} = 16.2029$ と共通で、

$$\left. \begin{aligned} A_1 \text{社} : \hat{y}^{(A_1)} &= \hat{\beta}_0^{(A_1)} a_1 + \hat{\beta}_1^{共通} x = 195.2832 + 16.2029x \\ A_2 \text{社} : \hat{y}^{(A_2)} &= \hat{\beta}_0^{(A_2)} a_2 + \hat{\beta}_1^{共通} x = 119.6903 + 16.2029x \\ A_3 \text{社} : \hat{y}^{(A_3)} &= \hat{\beta}_0^{(A_3)} a_3 + \hat{\beta}_1^{共通} x = 201.2918 + 16.2029x \end{aligned} \right\} \quad (1.6)$$

のように、切片が異なる 3 本の回帰直線が推定されている。

表 1.5 右下段で得られたパラメータ (係数) から、回帰直線の (始点, 終点) を会社ごとの年齢の (最小値, 最大値) に対する年収の推定値 \hat{y} を式 (1.6) を用いて推定し、図 1.3 に、年齢と年収の散布図上に回帰直線を重ね書きした結果を示す。この図からも明らかのように、同一年齢で比較すれば、A₁ 社と A₂ 社の年収は同程度であり、A₂ 社が他の 2 社よりも低いことが一目で判断される。

交互作用の検討

共通の傾きを持つ 3 本の直線のあてはめの統計的妥当性は、得られた 2 つの分散分析表の回帰の平方和 $S_{回帰}$ の差分により評価することができる。3 本の異なる直線をあてはめた場合、

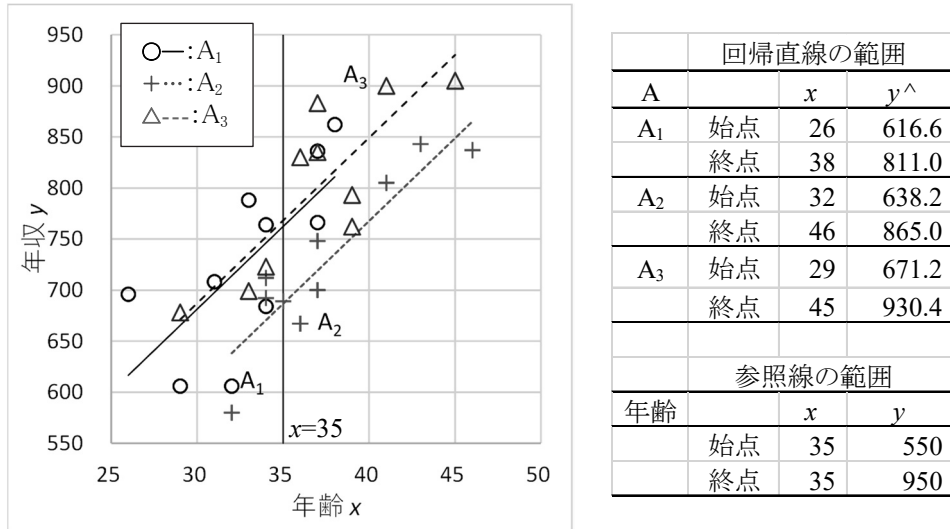


図 1.3 Excel の散布図による共通の傾きを持つ 3 本の直線

表 1.4 右上段の「分散分析表」の回帰の平方和（変動）は、 $S_{\text{回帰},df=6} = 17,182,504.95$ 、共通の傾きに対する表 1.5 右上段の「分散分析表」の回帰の平方和は、 $S_{\text{回帰},df=4} = 17,182,279.29$ であり、それらの差を計算する。この回帰の平方和の差が、交互作用の平方和 $S_{A \times x} = 225.66$ となり、自由度も差を取り $df = 6 - 4 = 2$ となる。

交互作用の平方和（変動） : $S_{A \times x} = 17,182,504.95 - 17,182,279.29 = 225.66$
 交互作用の自由度 : $df_{A \times x} = 6 - 4 = 2$
 交互作用の平均平方（分散） : $MS_{A \times x} = 225.66 / 2 = 112.83$

これらを、表 1.6 に示す分散分析表にまとめる。交互作用 $A \times x$ の平均平方（分散） $MS_{A \times x} = 112.83$ を自由度 24 の誤差分散の推定値（残差の平均平方） $\hat{\sigma}^2 = 2697.92$ で割り、検定統計量としての分散比（F 値）を求める。F 分布の右側確率のための $F.\text{dist.RT}()$ 関数で求めた p 値

$$\left. \begin{aligned} F_{A \times x} &= \frac{112.83}{2697.92} = 0.0418 \\ p_{A \times x} &= F.\text{dist.RT}(0.0418, 2, 24) = 0.9591 \end{aligned} \right\} \quad (1.7)$$

表 1.6 回帰の変動（平方和）の差分による交互作用の評価

	自由度	平方和(変動)	平均平方(分散)	分散比 F 値	p 値
回帰(別々の直線)	6	17182504.95	2863750.82	1061.47	
回帰(傾き共通)	4	17182279.29	4295569.82	1592.18	
交互作用 A×x	2	225.66	112.83	0.0418	0.9591
残差	24	64750.05	2697.92		
合計	30	17247255.00			

から、交互作用 $A \times x$ は全くないと判定される。ただし、このような統計的方法によらなくとも、図 1.1 から視覚的には交互作用がないことは明らかである。

表 1.6 に示した交互作用の評価は、残差の平方和に着目しても同じ結果が得られる。表 1.5 の自由度 26 の残差の平方和は、 $S_{残差,df=26} = 64,975.71$ であり、表 1.4 の自由度 24 の回帰の平方和は、 $S_{残差,df=24} = 64,750.05$ でありその差から、交互作用 $S_{A \times x,df=2} = 225.66$ として求めることもできる。

回帰直線上の推定値に対する 95%信頼区間

傾きが共通な平行な直線のあてはめが、“統計的”にも支持され、3 社の年齢の総平均は、 $\bar{x} = 35.87$ 歳であり、切のよい 35 歳とした場合の、3 社の年収の推定値と 95%信頼区間を推定し比較検討を行いたい。それぞれの会社の年収の推定値は、

$$\left. \begin{aligned} A_1 \text{社} : \hat{y}_{x=35}^{(A_1)} &= 195.2832 + 16.2029 \times 35 = 762.4 \\ A_2 \text{社} : \hat{y}_{x=35}^{(A_2)} &= 119.6903 + 16.2029 \times 35 = 686.8 \\ A_3 \text{社} : \hat{y}_{x=35}^{(A_3)} &= 201.2918 + 16.2029 \times 35 = 768.4 \end{aligned} \right\} \quad (1.8)$$

のように推定されたパラメータを用い、年齢 $x=35$ を式 (1.6) に代入することにより簡単に求まる。さて、これらの 95%信頼区間を求めるために、推定値の分散が必要となる。それぞれの推定されたパラメータの標準誤差 SE は、表 1.5 右下段に示されているのであるが、年齢が $x=35$ における各社の年収 y の推定値 $\hat{y}_{x=35}^{(A_i)}$ の分散は、合成分散の一般式

$$\left. \begin{aligned} Var(\hat{y}_{x=35}^{(A_i)}) &= Var(\hat{\beta}_0^{(A_i)} + \hat{\beta}_1^{共通} x) \\ &= Var(\hat{\beta}_0^{(A_i)}) + 2Cov(\hat{\beta}_0^{(A_i)}, \hat{\beta}_1^{共通})x + Var(\hat{\beta}_1^{共通})x^2 \end{aligned} \right\} \quad (1.9)$$

によって求める必要がある。

合成分散の一般式で用いられている分散 $Var(\hat{y}_{x=35}^{(A_i)})$ および $Var(\hat{\beta}_1^{共通})$ は、それぞれのパラメータの SE を 2 乗して求められるが、共分散 $Cov(\hat{\beta}_0^{(A_i)}, \hat{\beta}_1^{共通})$ は Excel の回帰分析からは得られない。そこで、線形モデルで定式化されている方法を用い、新たに計算する必要がある。パラメータの共分散行列 $\Sigma(\hat{\beta})$ は、表 1.5 左のデザイン行列を X 、表 1.5 上段の分散分析表の「残差の行の分散」を $\hat{\sigma}^2 = 2499.07$ (誤差分散の推定値) としたときに、

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2 \quad (1.10)$$

によって求めることができる。

実際に、平行な 3 本の回帰直線のパラメータの共分散行列 $\Sigma(\hat{\beta})$ を求めてみよう。表 1.8 左中段に示すように Excel の行列関数を用いて計算する。計算式中の X^T は、(30 行×4 列)

のデザイン行列 X を転置したサイズが (4 行×30 列) の行列である。行列の転置は、表 1.7 に示すように Excel のシート上でデザイン行列 X の範囲を選択・コピーし、プルダウン・メニューから「形式を選択して貼り付け」を選択し、「行/列の入れ替え」でペーストして得られる。また、転置のための行列関数 $\text{Transpose}(X \text{ の範囲})$ によっても得られる。なお、行列計算に不慣れな場合は、第 5 章に Excel による行列計算の手順について丁寧に解説してあるので先に読んでもらいたい。

表 1.7 転置行列 X^T および積和行列 ($X^T X$)

		転置行列 X^T					デザイン行列 X				積和行列 ($X^T X$)					
No		1	2	3	...	29	30	a_1	a_2	a_3	x					
a_1		1	1	1	...	0	0	1	0	0	34	=	10	0	0	331
a_2		0	0	0		0	0	1	0	0	33		0	10	0	375
a_3		0	0	0		1	1	1	0	0	34		0	0	10	370
x		34	33	34		45	39	:					331	375	370	39194
								0	0	1	45		=Mmult (Transpose (X の範囲), X の範囲)			
								0	0	1	39					
		(4×30)					(30×4)				(4×4)					

表 1.8 右に示すのは、表 1.5 右下段での Excel の回帰分析による回帰パラメータの推定に代え、デザイン行列 X を用いた Excel の行列計算による回帰パラメータの推定結果である。積和行列 ($X^T X$) は、 X^T の行方向の 30 個のデータに X の列方向の 30 個のデータを順番に掛けて加えたもので、(4×30) の転置行列 X^T 、(30×4) の X の積の計算で ($X^T X$) は、(4×4) の行列となり、Excel の行列関数 $\text{Mmult}(X^T \text{ の範囲}, X \text{ の範囲})$ により求められる。

積和行列 ($X^T X$) の逆行列 ($X^T X$)⁻¹ は、スカラーの演算での逆数に対応し、Excel の行列関数 $\text{Minverse}(X^T X \text{ の範囲})$ で求める。パラメータの共分散行列 $\Sigma(\hat{\beta})$ は、逆行列 ($X^T X$)⁻¹ のそれぞれの要素に表 1.5 右上段の残差の行の分散 (誤差分散の推定値) $\hat{\sigma}^2 = 2499.07$ を掛けて求める。

表 1.8 右下段の ($X^T y$) は、デザイン行列 X の転置行列に年収 y の列ベクトルを掛けた結果が示されている。この ($X^T y$) に逆行列 ($X^T X$)⁻¹ を前から掛けることにより、表 1.5 右下段に示した Excel の回帰分析の結果と同様に、回帰パラメータの推定値 $\hat{\beta}$ が求められている。また、推定されたパラメータの SE は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ の対角要素の平方根で求められている。このように、Excel の行列関数を使うことにより、Excel の回帰分析と同様の結果が得られる。詳しくは、第 5 章を参照のこと。

表 1.9 示すように、パラメータの共分散行列 $\Sigma(\hat{\beta})$ の対角要素は、各パラメータの分散となっており、これらの平方根は表 1.5 右下段に示した Excel の回帰分析で計算したパラメータ

表 1.8 平行な3本の回帰直線のパラメータの共分散行列 $\Sigma(\hat{\beta})$ の推定

— デザイン行列 X —						分散分析表「定数に0を使用」を on				
No	a_1	a_2	a_3	x	y		自由度	変動	分散	分散比
1	1	0	0	34	684	回帰	4	17182279	4295570	1718.87
2	1	0	0	33	788	残差	26	64976	2499.07	$=\sigma^2$
3	1	0	0	34	764	合計	30	17247255		
4	1	0	0	37	836					
5	1	0	0	29	606					
						積和行列 ($X^T X$)				
6	1	0	0	26	696	a_1	10	0	0	331
7	1	0	0	37	766	a_2	0	10	0	375
8	1	0	0	38	862	a_3	0	0	10	370
9	1	0	0	32	606	x	331	375	370	39194
10	1	0	0	31	708		=Mmult (Transpose (X の範囲), X の範囲)			
11	0	1	0	34	692					
12	0	1	0	34	712		逆行列 ($X^T X$) ⁻¹			
13	0	1	0	37	700	a_1	2.3571	2.5572	2.5231	-0.0682
14	0	1	0	43	843	a_2	2.5572	2.9971	2.8585	-0.0773
15	0	1	0	37	748	a_3	2.5231	2.8585	2.9204	-0.0762
16	0	1	0	32	580	x	-0.0682	-0.0773	-0.0762	0.0021
17	0	1	0	46	837		=Minverse ($X^T X$ の範囲)			
18	0	1	0	36	667					
19	0	1	0	41	805		パラメータの共分散行列 $\Sigma(\hat{\beta}) = (X^T X)^{-1} \sigma^2$			
20	0	1	0	35	689	β_0^{A1}	5890.62	6390.53	6305.33	-170.41
21	0	0	1	34	723	β_0^{A2}	6390.53	7489.94	7143.50	-193.07
22	0	0	1	39	762	β_0^{A3}	6305.33	7143.50	7298.16	-190.49
23	0	0	1	37	883	$\beta_1^{共通}$	-170.41	-193.07	-190.49	5.1485
24	0	0	1	29	678		$(X^T X)^{-1}$ の範囲 * σ^2			
25	0	0	1	33	699		回帰パラメータ			
26	0	0	1	36	830		$(X^T v)$	$\hat{\beta} = (X^T X)^{-1} X^T v$	SE	
27	0	0	1	41	900		7316	$\beta_0^{A1} =$	195.2832	76.7504
28	0	0	1	37	835		7273	$\beta_0^{A2} =$	119.6903	84.5192
29	0	0	1	45	905		8008	$\beta_0^{A3} =$	201.2918	85.4293
30	0	0	1	39	793		819058	$\beta_1^{共通} =$	16.2029	2.2690

の標準誤差 SE に一致する。ちなみに、共通の傾き $SE(\hat{\beta}_1^{共通}) = 2.2690$ は、

$$SE(\hat{\beta}_1) = \sqrt{Var(\hat{\beta}_1)} = \sqrt{5.1485} = 2.2690$$

となり、表 1.5 右下段に示した Excel の回帰分析で求めた $\hat{\beta}_1^{共通}$ の標準誤差に一致する。

表 1.9 パラメータの共分散行列 $\Sigma(\hat{\beta})$

$Var(\hat{\beta}_0^{A1})$				=	5890.62	6390.53	6305.33	-170.41
$Cov(\hat{\beta}_0^{A1}, \hat{\beta}_0^{A2})$	$Var(\hat{\beta}_0^{A2})$				6390.53	7489.94	7143.50	-193.07
$Cov(\hat{\beta}_0^{A1}, \hat{\beta}_0^{A3})$	$Cov(\hat{\beta}_0^{A2}, \hat{\beta}_0^{A3})$	$Var(\hat{\beta}_0^{A3})$			6305.33	7143.50	7298.16	-190.49
$Cov(\hat{\beta}_0^{A1}, \hat{\beta}_1^{共通})$	$Cov(\hat{\beta}_0^{A2}, \hat{\beta}_1^{共通})$	$Cov(\hat{\beta}_0^{A3}, \hat{\beta}_1^{共通})$	$Var(\hat{\beta}_1^{共通})$		-170.41	-193.07	-190.49	5.1485
					$\Sigma(\hat{\beta}) = (X^T X)^{-1} \sigma^2$			

表 1.10 に年齢が 35 歳に場合の回帰の推定値 ($\hat{y}_{x=35}^{(A1)}$, $\hat{y}_{x=35}^{(A2)}$, $\hat{y}_{x=35}^{(A3)}$), それらの分散 [$Var(\hat{y}_{x=35}^{(A1)})$, $Var(\hat{y}_{x=35}^{(A2)})$, $Var(\hat{y}_{x=35}^{(A3)})$], および, 95%信頼区間の計算結果を示す。

表 1.10 回帰直線の 35 歳における推定値, 分散, SE, および 95%信頼区間

会社	推定値	分散	標準誤差	幅	95%信頼区間	
	$\hat{y}_{x=35}^{A_i}$	$Var(\hat{y}_{x=35}^{A_i})$	SE	$t_{0.05} \times SE$	L95%	U95%
A ₁	762.39	268.49	16.3857	33.6814	728.70	796.07
A ₂	686.79	282.08	16.7954	34.5234	652.27	721.32
A ₃	768.39	270.50	16.4469	33.8071	734.59	802.20
		T.Inv.2T (0.05,30-4)=		2.0555		

A₁ 社の 35 歳における分散は, 推定値 $\hat{y}_{x=35}^{(A_1)} = 762.29$ に対する合成分散の一般式 (1.9) を用い,

$$\left. \begin{aligned} Var(\hat{y}_{x=35}^{(A_1)}) &= Var(\hat{\beta}_0^{(A_1)} + \hat{\beta}_1^{共通} x) \\ &= Var(\hat{\beta}_0^{(A_1)}) + 2Cov(\hat{\beta}_0^{(A_1)}, \hat{\beta}_1^{共通})x + Var(\hat{\beta}_1^{共通})x^2 \\ &= 5890.62 + 2 \times (-170.41) \times 35 + 5.1485 \times 35^2 \\ &= 268.49 \end{aligned} \right\} \quad (1.11)$$

として求められ, 95%信頼区間は, 一般的な方法により

$$\left. \begin{aligned} 95\%CL_{x=35}^{(A_1)} &= \hat{y}_{x=35}^{(A_1)} \pm t_{0.05}(30-4)\sqrt{Var(\hat{y}_{x=35}^{(A_1)})} \\ &= 762.39 \pm 2.0555 \times 16.3857 \\ &= (728.70, 796.07) \end{aligned} \right\} \quad (1.12)$$

として推定される.

会社間の推定値の差の 95%信頼区間

図 1.3 から, 年齢を揃えた場合には, A₂ 社の給料が他社に比べて低いことは, 散布図上の回帰直線が A₁ 社および A₃ 社よりも明らかに下方にあることから明らかである. 表 1.10 から, 35 歳での A₂ 社の給料の推定値は, 686.79 万円であり, 95%信頼区間の上限が 721.32 万円と A₁ 社の下限 728.70 万円および A₃ 社の下限 734.59 万円よりも低いので統計的にも明らかな差があると判断される.

統計的に 35 歳での A₂ 社の給与の推定値と A₁ 社の推定値を比較する方法を示す. A₁ 社と A₂ 社の推定値は, 次式で

$$\left. \begin{aligned} A_1 \text{社} : \hat{y}_{x=35}^{(A_1)} &= \hat{\beta}_0^{(A_1)} + \hat{\beta}_1^{共通} x \\ A_2 \text{社} : \hat{y}_{x=35}^{(A_2)} &= \hat{\beta}_0^{(A_2)} + \hat{\beta}_1^{共通} x \end{aligned} \right\} \quad (1.13)$$

で与えられている. この式で, A₁ 社から A₂ 社を引くと $\hat{\beta}_1 x_1$ 項は相殺され,

$$(A_1 \text{社}) - (A_2 \text{社}) : \left\{ \begin{aligned} \hat{y}_{x=35}^{(A_1)} - \hat{y}_{x=35}^{(A_2)} &= \hat{\beta}_0^{(A_1)} - \hat{\beta}_0^{(A_2)} \\ &= 762.39 - 686.79 = 75.59 \end{aligned} \right\} \quad (1.14)$$

それぞれの切片の差となる. その分散 $Var(\hat{y}_{x=35}^{(A_1)} - \hat{y}_{x=35}^{(A_2)})$ は,

$$\left. \begin{aligned}
 \text{Var}(\hat{y}_{x=35}^{(A_1)} - \hat{y}_{x=35}^{(A_2)}) &= \text{Var}(\hat{\beta}_0^{(A_1)} - \hat{\beta}_0^{(A_2)}) \\
 &= \text{Var}(\hat{\beta}_0^{(A_1)}) - 2\text{Cov}(\hat{\beta}_0^{(A_1)}, \hat{\beta}_0^{(A_2)}) + \text{Var}(\hat{\beta}_0^{(A_2)}) \\
 &= 5890.62 - 2 \times 6390.53 + 7489.94 = 599.49
 \end{aligned} \right\} \quad (1.15)$$

となり、表 1.11 に示すように、95%信頼区間を計算することができる。

表 1.11 回帰直線の差の推定値、分散、SE、および 95%信頼区間

会社	差の推定値	分散	標準誤差	幅	95%信頼区間	
	$\hat{y}_{x=35}^{\text{差}}$	$\text{Var}(\hat{y}_{x=35}^{\text{差}})$	SE		$t_{0.05} \times SE$	L95%
A ₁ -A ₂	75.59	599.49	24.4844	50.3285	25.26	125.92
A ₂ -A ₂	0.00	0.00	0.0000	0.0000	0.00	0.00
A ₃ -A ₂	81.60	501.10	22.3853	46.0136	35.59	127.62
		T.Inv.2T (0.05,30-4)=		2.0555		

表 1.10 および表 1.11 で計算された推定値と 95%信頼区間を Excel の折れ線グラフで作図した結果を図 1.4 に示す。A₁ 社は、A₂ 社と同程度、A₃ 社に比べて平均年収が低かったが、同一年齢で比べれば、A₂ 社よりも年収は高く、A₃ 社と同水準の年収であることが明らかになった。

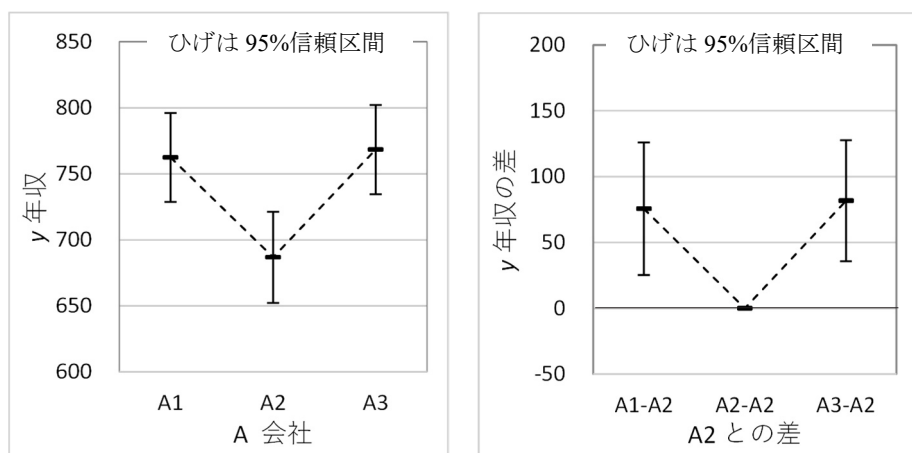


図 1.4 Excel の折れ線グラフによる年齢が 35 歳における年収の比較

デザイン行列 X とは何か

デザイン行列 (design matrix) は、計画行列とも称されるが、本書では、デザイン行列で統一している。デザイン行列は、変幻自在で説明に窮するが、最小 2 乗法によるパラメータの推定ための用いられる行列で、通常 X と表記する。単回帰分析であれば、表 1.2 左に即して言えば、切片 β_0 を推定するための変数 x_0 、傾き β_1 を推定するための変数 x_1 の (10 行×2 列) の矩形データが、デザイン行列 X となる。また、表 1.4 左では、(30 行×6 列)、表 1.5 左では、(30 行×4 列) のデザイン行列 X となり、推定の目的に合わせて行列のサイズが変わる。なお、第 5 章の 5.2 節の末尾に、design matrix について英語版の Wikipedia を引用した解説を示してあるので参考にされたい。

1.2. 季節による洗浄水の温度と回収液の濃度の関連

奥野・久米・芳賀・吉澤(1981),「多変量解析法《改訂版》」は,多くの人達に読み継がれてきた名著であり,その中の「第II部 重回帰分析 第7章 層別因子を含む回帰分析」の[例7.1]は,4水準の層別因子を含む回帰分析の事例である.原著では,第6章で示すように伝統的な層ごとの単回帰分析を積み上げる解析方法が示されている.この方法は,PCによる回帰分析が普及する以前の手計算時代の標準的な解法であり,技巧的で見通しが悪い.そこで,ダミー変数を用いた回帰分析による解析方法を用いる.なお,この事例については,高橋(2021),「最尤法によるポアソン回帰分析入門,第13.2節 交互作用を考慮した共分散分析」でも取り上げている.

季節による層別

表1.12に示すように,亜硫酸ガス回収塔の洗浄水の温度 x と,回収液の濃度 y との関係を調べたい.長期間にわたって取られたデータなので,季節の変化だけでなく,操業条件も変わっていると思われる.そこで,表1.12に示すようにデータを4つ季節に分けている.

表1.12 季節ごとの洗浄水の温度 x と回収液の濃度 y [奥野ら(1981),表7.2]

No.	A ₁ (7月・8月)		A ₂ (9月・10月)		A ₃ (11月・12月)		A ₄ (1月・2月)	
	x	y	x	y	x	y	x	y
1	30	8.0	34	9.1	27	15.0	16	23.4
2	28	10.5	19	19.4	20	21.0	16	28.2
3	30	8.2	22	20.5	14	24.2	20	29.5
4	29	13.0	25	14.2	18	15.3	16	22.2
5	28	10.1	22	11.0	12	17.3	6	40.2
6	17	16.2	25	19.1	18	24.7	16	36.6
7	24	16.1	22	16.0	13	23.9	12	35.5
8	22	13.4	23	11.1	21	20.2	11	42.1
9	30	13.0	28	12.0	24	13.6	9	42.5
10	30	7.3	24	17.1			22	25.8
11	25	14.0	30	13.3				
12			24	8.9				
平均	26.64	11.80	24.83	14.31	18.56	19.47	14.40	32.60
		総平均	21.48	19.11	$N=42$			

JMPの「グラフ・ビルダー」による層別散布図

図1.5に示すように,洗浄水の温度 x は,層別因子として取り上げた季節Aの影響を受ける.表1.12に示したように,A₁(7月・8月)で $\bar{x}_1=26.64$ 度であったのが,A₂(9月・10月)に $\bar{x}_2=24.83$ 度と減少し,さらに,A₄(1月・2月)に $\bar{x}_4=14.40$ 度と減少している.回

回収液の濃度 y の平均値は、逆に (11.80, 14.31, 19.47, 32.60) と上昇している。回収液の濃度 y の上昇は、季節 A 以外の他の因子の影響があるかも知れない。

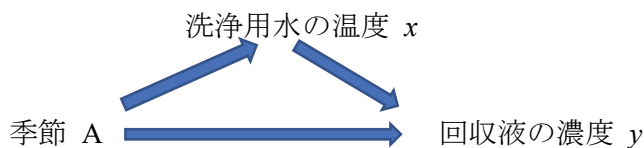


図 1.5 回収液の濃度に影響をおよぼす要因

得られたデータの関連を概観するために、JMP のグラフ・ビルダーを用いて洗淨水の温度 x と回収液の濃度 y の散布図上に季節別の回帰直線を上書きした結果を図 1.6 に示す。季節が A₄ (1月・2月) の場合に多重比較などをするまでもなく明らかに他の季節に比べ洗淨水の温度 x が低い方に分布し、他の季節と共通の洗淨水の温度 x が 20 度で比べた場合に、回収液の濃度 y が高めている。そのために、季節に関連する操業条件などの他の因子の影響があるのではないかと推測される。

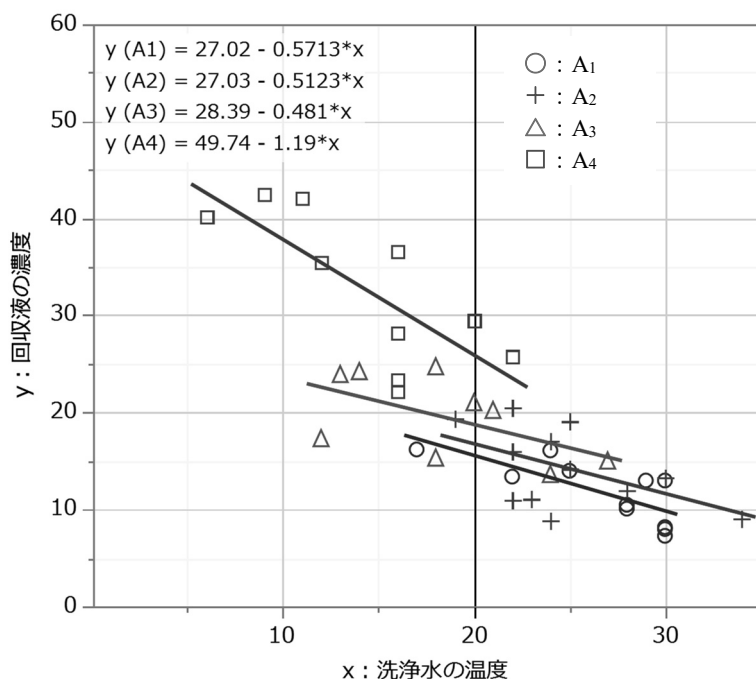


図 1.6 JMP のグラフ・ビルダーによる 4 本の回帰直線のあてはめ

JMP の「モデルのあてはめ」による解析

JMP の「モデルのあてはめ」を用いた回帰分析を用い、層別因子 A と洗淨水の温度 x との交互作用 $A \times x$ を含む回帰分析を適用し、予測プロファイルを用いて比較検討する。解析モデルは、4水準の季節 A、洗淨水の温度 x 、それらの交互作用 $A \times x$ を含め、表 1.13 に示すような分散分析表および「効果の検定」に要因ごとの平方和などが得られる。

表 1.13 季節 A と洗浄水の温度 x の交互作用を含む分散分析表

分散分析					
要因	自由度	平方和	平均平方	F値	
モデル	7	3145.7754	449.3965	30.0158	
誤差	34	509.0487	14.9720	p値(Prob>F)	
全体(修正済み)	41	3654.8240		<.0001*	
効果の検定					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
A	3	246.2924	82.0975	5.4834	0.0035*
x	1	369.0901	369.0901	24.6520	<.0001*
A*x	3	71.2462	23.7487	1.5862	0.2107

「▼ モデルの設定」の「多項式の中心化」をオフとした結果である。

交互作用 $A \times x$ が $p = 0.2107$ と有意ではないので原著では、解析モデルから交互作用を除いて季節 A に共通の傾きを持つ回帰直線をあてはめる根拠としている。だが、季節 A₄ (1月, 2月) の洗浄水の温度の範囲は、6~22度と他の季節に比べて著しく低く、低くなるほど回収液の濃度は高くなり、他の季節とは異なる別の要因が影響していると推測される。そこで、図 1.6 に示した結果から各季節に共通している洗浄水の温度が 20 度の場合に、各季節の回帰直線の推定値とその 95%信頼区間を求めて比較検討を行う。

JMP の「予測プロファイル」

JMP の「予測プロファイル」は、解析結果のグラフ表示に優れている。図 1.7 に示すように、質的変数に対しては、他の量的変数の設定値に連動し、水準ごとの推定値および 95%信頼区間が表示され、量的変数の場合には、他の質的変数で選択された水準に連動した回帰直線および 95%信頼区間が動的に表示される。さらに、それらを組み合わせた場合の推定値および 95%信頼区間の推定結果も示される。質的変数の水準の選択、量的変数の設定も GUI (Graphical User Interface) により自在に変更することができる。JMP の「予測プロファイル」は、他の統計ソフトにはない動的なグラフ表示が特徴である。

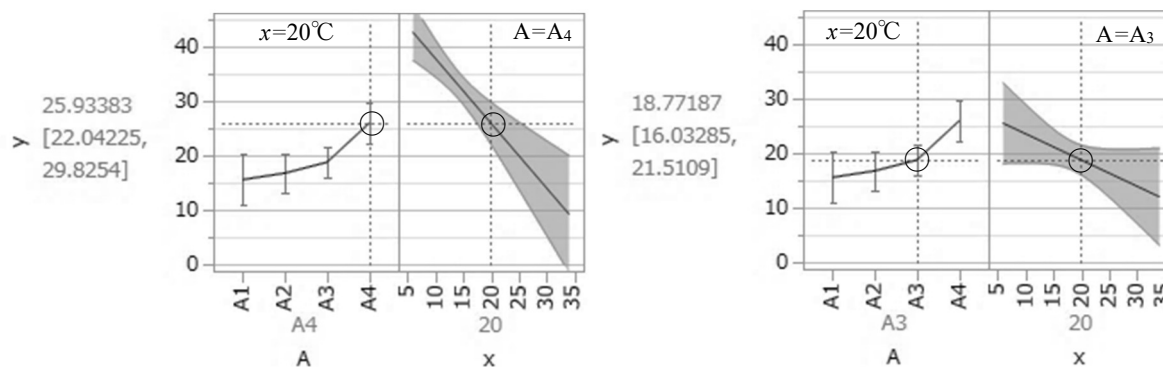
図 1.7 A₄および A₃を選択した場合の予測プロファイル

図 1.7 左は、予測プロファイル上で因子 A の A₄ 水準を選択し、洗浄水の温度 x を 20 度と設定した結果であり、それらを組み合わせた推定値として 25.93、および、その 95%信頼区間 [22.04, 29.83] が図 1.7 左端に表示され、他の (A₁, A₂, A₃) 水準に対しても、洗浄水の温度を $x=20$ とした場合の各水準の推定値とその 95%信頼区間が図示されている。この予測プロファイルから A₄ 水準と A₃ 水準の 95%信頼区間のひげが重なりあっていないので、統計的に有意な差があると判断される。

図 1.7 右は、GUI により A₃ 水準を選択し、洗浄水の温度 x を 20 度はそのままにした結果であり、A₃ 水準の推定値 18.77 と 95%信頼区間 [16.03, 21.51] が表示され、A₃ 水準に対応する洗浄水の温度 x に関する回帰直線と 95%信頼区間が示されている。回帰直線の傾きは、図 1.6 から示唆されるように A₄ 水準に比べて平坦化し、95%信頼区間の上限が洗浄水の温度 x の上昇に伴い上方に発散気味であり傾きが有意な差とは言えないことが読み取れる。

洗浄水の温度 x を 20 度と設定しているが、デフォルトでは、洗浄水の温度 x の総平均の $\bar{x} = 21.48$ 度が設定される。ただし、その温度は自由に変更することができ、図 1.8 に示すように温度を 10 度と 30 度に設定すると、交互作用を含めているために季節 A ごとの水準平均と 95%信頼区間が、がらりと変わることが示されている。洗浄水の温度が 10 度の場合には、A₄ 水準以外は外挿となり 95%信頼区間のひげの差が広がってはいるが、有意な差となっている。図 1.8 右に示すように洗浄水の温度が 30 度の場合には、A₄ 水準は外挿となることが図 1.6 から明らかであり、回収液の濃度 y の推定値に関し季節 A 間に有意な差は認められない。

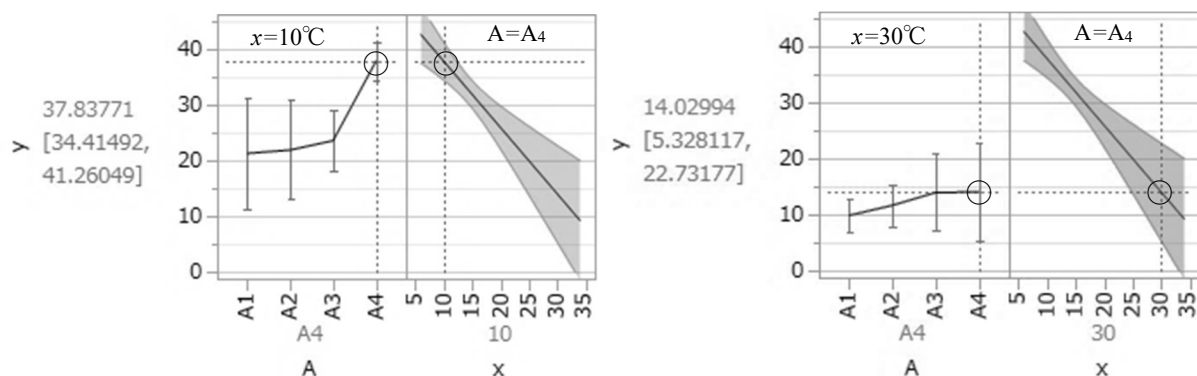


図 1.8 洗浄水の温度を変えた場合の季節 A の各水準の推定値と 95%信頼区間

対比による水準間の差の推定

JMP には、対比の機能を使って任意の水準間の差について統計量を算出する機能が備わっている。表 1.14 に示すように、洗浄水の温度を 20 度に固定し、季節 A₃ を +1、季節 A₄ を -1 とする対比の指定により、図 1.7 に示した A₃ の推定値の 18.7719 と A₄ の推定値 25.9338 との

表 1.14 洗浄水の温度 20 度の場合の季節 A₄ と季節 A₃ の差の統計量

対比		検定の詳細		パラメータ関数		
対比の指定 A A1 0 × 温度を 20°C に設定 A2 0 A3 1 A4 -1 +または-をクリックして対比值を作成。		A1	0	パラメータ		
		A2	0	切片	0	
		A3	1	A[A1]	1	
		A4	-1	A[A2]	1	
		推定値	-7.16195	A[A3]	2	
		標準誤差	2.34167	x	0	
		t値	-3.05848	x*A[A1]	20	
		p値(Prob> t)	0.004317	x*A[A2]	20	
		平方和	140.0529	x*A[A3]	40	
		下側95%	-11.9208			
		上側95%	-2.40311			
		平方和	分子自由度	分母自由度	F値	p値(Prob>F)
		140.1	1	34	9.3543	0.0043*
		x	20			

差が -7.1620, 標準誤差が 2.3417 と推定されている. 95%信頼区間の出力が, (-11.9208, -2.4031) と 0 を含まないので有意な差であることが分かる.

JMP での対比の設定は, 直観的にわかりやすい (+A₃-A₄) となっているが, JMP の内部での設定は, 表 1.14 右の「パラメータ関数」で示すように (+A₁+A₂+2A₃) が使われている. これは, JMP が (1, -1) 対比型のダミー変数をデフォルトにしていることに関係している. 表 1.15 に示す対比型ダミー変数から明らかのように, (+A₃-A₄) は, パラメータ関数で示されているように, 内部では (+a₁+a₂+2a₃) と置き換えられている.

表 1.15 (1, -1) 対比型ダミー変数における A₃ と A₄ の差

A	x ₀	a ₁	a ₂	a ₃	x	a ₁ x	a ₂ x	a ₃ x
A ₁	1	1	0	0	20	20	0	0
A ₂	1	0	1	0	20	0	20	0
A ₃	1	0	0	1	20	0	0	20
A ₄	1	-1	-1	-1	20	-20	-20	-20
+A ₃ -A ₄	0	1	1	2	0	20	20	40

パラメータ関数			
パラメータ		パラメータ	
切片	0	x	0
A[A1]	1	A[A1]*x	20
A[A2]	1	A[A2]*x	20
A[A3]	2	A[A3]*x	40

Excel による(1, -1)対比型ダミー変数を用いた解析

JMP の「モデルのあてはめ」により, 質的変数としての季節 A, 量的変数としての洗浄水の温度 x, それらの交互作用 A×x を含めた解析モデル

$$\text{解析モデル: } y = A + x + (A \times x) \tag{1.16}$$

を適用した. JMP の内部では, 質的変数について, (1, -1) 対比型ダミー変数に変換し, 線

形モデルとした解析が行われている。さらに、「予測プロファイル」による推定値の出力のグラフ表示（予測プロファイル）は、結果の解釈に有用であり、Excel で再現することにより理解を深めたい。

質的変数を線形モデルで扱うためには、量的変数のダミー変数に置き換える必要がある。表 1.3 に示したようにダミー変数は、幾つかの型に分類でき、第 1.1 節では (1, 1) 標示型ダミー変数を用い、3 本の回帰直線の切片と傾きを直接推定した。JMP の内部では、表 1.16 に示すように (1, -1) 対比型ダミー変数が使われている。季節 A は 4 水準なので、ダミー変数名として (a_1, a_2, a_3) を与え、洗浄水の温度 x との交互作用は、 (a_1x, a_2x, a_3x) のように積で与える。

表 1.16 (1, -1) 対比型ダミー変数における交互作用

A	a_1	a_2	a_3	A	a_1x	a_2x	a_3x
A ₁	1	0	0	A ₁	x	0	0
A ₂	0	1	0	A ₂	0	x	0
A ₃	0	0	1	A ₃	0	0	x
A ₄	-1	-1	-1	A ₄	$-x$	$-x$	$-x$

式 (1.16) に示した交互作用を含む解析モデルに対しダミー変数を用いた線形モデルは、

$$y_i = \theta_0 x_{0,i} + \theta_1 a_{1,i} + \theta_2 a_{2,i} + \theta_3 a_{3,i} + \theta_4 x_i + \theta_5 a_{1,i} x_i + \theta_6 a_{2,i} x_i + \theta_7 a_{3,i} x_i + \varepsilon_i \quad (1.17)$$

となる。データ数は、 $N = 11 + 12 + 9 + 10 = 42$ で、切片 x_0 を含め変数の数は 8 であり、Excel の回帰分析のためのデザイン行列 X は、表 1.17 に示すように (42 行 × 8 列) と大きくなるが、回帰分析の結果は瞬時に得られる。

表 1.17 交互作用モデルに対する回帰分析

----- デザイン行列 X -----										分散分析表「定数に 0 を使用」 on とする						
No.	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y		自由度	変動	分散	分散比	
1	A ₁	1	1	0	0	30	30	0	0	8.0		8	18486.90	2310.86	154.35	
2		1	1	0	0	28	28	0	0	10.5		残差	34	509.05	14.97 = σ^2	
:												合計	42	18995.95		
11		1	1	0	0	25	25	0	0	14.0						
12	A ₂	1	0	1	0	34	0	34	0	9.1						
13		1	0	1	0	19	0	19	0	19.4						
:												係数	SE	t	P-値	
23		1	0	1	0	24	0	24	0	8.9	θ^0	切片	0	#N/A	#N/A	#N/A
24	A ₃	1	0	0	1	27	0	0	27	15.0	θ^1	x_0	33.0450	3.1137	10.6127	0.0000
25		1	0	0	1	20	0	0	20	21.0	θ^2	a_1	-6.0281	6.3343	-0.9517	0.3480
:											θ^3	a_2	-6.0157	5.9544	-1.0103	0.3195
32		1	0	0	1	24	0	0	24	13.6	θ^4	a_3	-4.6528	4.8112	-0.9671	0.3403
33	A ₄	1	-1	-1	-1	16	-16	-16	-16	23.4	θ^5	x	-0.6887	0.1387	-4.9651	0.0000
34		1	-1	-1	-1	16	-16	-16	-16	28.2	θ^6	a_1x	0.1175	0.2473	0.4749	0.6379
:											θ^7	a_2x	0.1765	0.2450	0.7205	0.4762
42		1	-1	-1	-1	22	-22	-22	-22	25.8		a_3x	0.2077	0.2364	0.8786	0.3858
総平均						21.4762				19.1119						

Excel の回帰分析により、分散分析表およびパラメータの推定結果が得られる。ただし、第 1.1 節でも示したように、パラメータ間の共分散が得られないために、Excel の行列関数を用いてパラメータの共分散行列 $\Sigma(\hat{\theta})$ を式 (1.10) と同様に

$$\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2 \quad (1.18)$$

により計算する必要がある。行列のサイズは異なるが、第 1.1 節の表 1.8 で示したと全く同じ計算手順である。

パラメータの共分散行列 $\Sigma(\hat{\theta})$ は、式 (1.18) に示したようにデザイン行列と分散分析表の残差の分散（誤差分散の推定値） $\hat{\sigma}^2 = 14.9720$ を用いて

$$\begin{aligned} \Sigma(\hat{\theta}) &= (X^T X)^{-1} \hat{\sigma}^2 \\ &= \text{Minverse (Mmult (Transpose (Xの範囲), Xの範囲)) * 14.9720} \end{aligned}$$

	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x
θ^{\wedge}_0	9.6953	5.5188	3.1845	-2.9689	-0.4134	-0.1450	-0.0927	0.0733
θ^{\wedge}_1	5.5188	40.1234	-18.3985	-12.2452	-0.1450	-1.5302	0.6511	0.4851
θ^{\wedge}_2	3.1845	-18.3985	35.4548	-9.9109	-0.0927	0.6511	-1.4256	0.4328
θ^{\wedge}_3	-2.9689	-12.2452	-9.9109	23.1481	0.0733	0.4851	0.4328	-1.0936
θ^{\wedge}_4	-0.4134	-0.1450	-0.0927	0.0733	0.0192	0.0017	0.0011	-0.0009
θ^{\wedge}_5	-0.1450	-1.5302	0.6511	0.4851	0.0017	0.0612	-0.0221	-0.0200
θ^{\wedge}_6	-0.0927	0.6511	-1.4256	0.4328	0.0011	-0.0221	0.0600	-0.0195
θ^{\wedge}_7	0.0733	0.4851	0.4328	-1.0936	-0.0009	-0.0200	-0.0195	0.0559

のように計算することができる。表 1.17 は、表示スペースの制約で少数点以下 4 桁で表示しているが、Excel のセル内では、倍精度実数（有効数字で 15 桁）となっている。なお、これら行列計算の詳細は、第 5 章を参照のこと。

(1, -1)対比型ダミー変数における回帰直線の推定式

前節では、質的変数について (1, 1) 標示型ダミー変数を用いたので、推定されたパラメータが水準ごとの切片および共通の傾きであったが、本節では (1, -1) 対比型ダミー変数を用いているので、季節 A_i の回帰直線の切片と傾きの推定には、次に示すような計算が必要となる。

$$\left. \begin{aligned} y^{(A_1)} &= (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_4 + \hat{\theta}_5)x \\ &= 33.04 - 6.03 + (-0.69 + 0.12)x \\ &= 27.02 - 0.57x \end{aligned} \right\} \quad (1.19)$$

$$\left. \begin{aligned} y^{(A_2)} &= (\hat{\theta}_0 + \hat{\theta}_2) + (\hat{\theta}_4 + \hat{\theta}_6)x \\ &= 27.03 - 0.51x \end{aligned} \right\} \quad (1.20)$$

$$\left. \begin{aligned} y^{(A_3)} &= (\hat{\theta}_0 + \hat{\theta}_3) + (\hat{\theta}_4 + \hat{\theta}_7)x \\ &= 28.39 - 0.48x \end{aligned} \right\} \quad (1.21)$$

$$\begin{aligned}
 y^{(A_4)} &= [\hat{\theta}_0 - (\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3)] + [\hat{\theta}_4 - (\hat{\theta}_5 + \hat{\theta}_6 + \hat{\theta}_7)]x \\
 &= 33.04 - (-16.70) + (-0.69 - 0.50)x \\
 &= 49.74 - 1.19x
 \end{aligned}
 \tag{1.22}$$

これらの推定値は、JMP のグラフ・ビルダーで作成された図 1.6 に示されている回帰式に一致し、Excel による推定方法が適切であることが確認できる。

Excel による予測プロファイル

図 1.7 に示した JMP の予測プロファイルで推定した洗浄水の温度 $x=20$ とした場合の A_3 の推定を Excel の行列計算で行う。洗浄水の温度 x を季節に共通の 20 度に固定し回収液の濃度の推定値 $\hat{y}_{x=20}^{(A_3)}$ と分散 $Var(\hat{y}_{x=20}^{(A_3)})$ を推定したい。式 (1.21) で $x=20$ とすると、

$$\hat{y}_{x=20}^{(A_3)} = (\hat{\theta}_0 + \hat{\theta}_3) \times 1 + (\hat{\theta}_4 + \hat{\theta}_7)x = 28.39 - 0.48 \times 20 = 18.77
 \tag{1.23}$$

が得られ、図 1.7 右に示されている推定値に一致する。

式 (1.23) を拡張できるように一般化すると、表 1.18 に示すように、 A_3 の推定値 $\hat{y}_{x=20}^{(A_3)}$ を算出するための行ベクトルを \mathbf{l} としたときに、推定された回帰パラメータ $\hat{\boldsymbol{\theta}}$ との積についての線形和 $L_{x=20}^{(A_3)} = \mathbf{l}\hat{\boldsymbol{\theta}}$ として求めることができる。

表 1.18 季節 A_3 の推定値に対する線形和 L による計算

		$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\theta}_7$	
	$\boldsymbol{\theta}^T$	33.0450	-6.0281	-6.0157	-4.6528	-0.6887	0.1175	0.1765	0.2077	
		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	
A_3	\mathbf{l}	1	0	0	1	20	0	0	20	線形和
										L
積和	$L = \mathbf{l}\hat{\boldsymbol{\theta}}$	33.0450	0	0	-4.6528	-13.7747	0	0	4.1544	= 18.7719

線形和 $L_{x=20}^{(A_3)}$ の分散は、式 (1.9) で示した 2 変数の合成分散の一般式を 8 変数に拡張した 2 次形式、

$$Var(L_{x=20}^{(A_3)}) = \mathbf{l}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\mathbf{l}^T
 \tag{1.24}$$

	\mathbf{l}	パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$								\mathbf{l}^T	
=	1 0 0 1 20 0 0 20	9.70	5.52	3.18	-2.97	-0.41	-0.14	-0.09	0.07	1	= 1.8165
		5.52	40.12	-18.40	-12.25	-0.14	-1.53	0.65	0.49	0	
		3.18	-18.40	35.45	-9.91	-0.09	0.65	-1.43	0.43	0	
		-2.97	-12.25	-9.91	23.15	0.07	0.49	0.43	-1.09	1	
		-0.41	-0.14	-0.09	0.07	0.02	0.00	0.00	0.00	20	
		-0.14	-1.53	0.65	0.49	0.00	0.06	-0.02	-0.02	0	
		-0.09	0.65	-1.43	0.43	0.00	-0.02	0.06	-0.02	0	
		0.07	0.49	0.43	-1.09	0.00	-0.02	-0.02	0.06	20	

よって $Var(\hat{y}_{x=20}^{(A_3)})=1.8165$ として推定することができる. その 95%信頼区間は,

$$\left. \begin{aligned} (L95\%, 95\%) &= \hat{y}_{x=20}^{(A_3)} \pm t_{0.05}(42-8)\sqrt{Var(\hat{y}_{x=20}^{(A_3)})} \\ &= 18.7719 \pm 2.0322\sqrt{1.8165} \\ &= (16.0328, 21.5109) \end{aligned} \right\} \quad (1.25)$$

が得られ, 図 1.7 右に示されている推定値に一致する. 同様に他の季節 A の水準に対して計算した結果を表 1.19 に示す. 季節 A₄ の推定値は, $L_{x=20}^{(A_4)} = 25.9338$, 95%信頼区間は, (22.0423, 29.8254) と図 1.7 左の推定値に一致する.

表 1.19 交互作用を考慮した季節別の推定値と 95%信頼区間

		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$\hat{\theta}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
L ₁	A ₁	1	1	0	0	20	20	0	0	15.5912	5.0542	4.5688	11.0224	20.1600
L ₂	A ₂	1	0	1	0	20	0	20	0	16.7842	3.1520	3.6080	13.1762	20.3922
L ₃	A ₃	1	0	0	1	20	0	0	20	18.7719	1.8165	2.7390	16.0328	21.5109
L ₄	A ₄	1	-1	-1	-1	20	-20	-20	-20	25.9338	3.6669	3.8916	22.0423	29.8254
										$t_{0.05}(0.05,42-8)=$	2.0322			

この結果を Excel の「折れ線」グラフで作成した結果を図 1.9 左に示す. Excel での信頼区間の幅を付ける際には, 推定値からの長さでの設定になっているので $t_{0.05}(df) \times SE$ により, ひげの長さを計算する必要がある.

表 1.20 に示すのは, 洗浄水の温度 x を変化させた場合の回帰直線推定値および 95%信頼区間の計算結果である. 図 1.7 では別々に示されている回帰直線の推定値と 95%信頼区間の図であるが, 図 1.9 右に示すように, Excel の散布図により A₃ と A₄ についてこれらをまとめて重ね書きした結果を示す.

表 1.20 季節 A₃ と A₄ に対する回帰直線の 95%信頼区間の計算

		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$\hat{\theta}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
L ₅	A ₃	1	0	0	1	10	0	0	10	23.5820	7.0298	5.3883	18.1937	28.9703
L ₆	A ₃	1	0	0	1	15	0	0	15	21.1769	2.5904	3.2708	17.9061	24.4478
L ₇	A ₃	1	0	0	1	20	0	0	20	18.7719	1.8165	2.7390	16.0328	21.5109
L ₈	A ₃	1	0	0	1	25	0	0	25	16.3668	4.7083	4.4097	11.9571	20.7765
L ₉	A ₃	1	0	0	1	30	0	0	30	13.9618	11.2657	6.8211	7.1407	20.7829
L ₁₀	A ₄	1	-1	-1	-1	5	-5	-5	-5	43.7896	7.6105	5.6064	38.1833	49.3960
L ₁₁	A ₄	1	-1	-1	-1	10	-10	-10	-10	37.8377	2.8367	3.4228	34.4149	41.2605
L ₁₂	A ₄	1	-1	-1	-1	15	-15	-15	-15	31.8858	1.5221	2.5073	29.3785	34.3930
L ₁₃	A ₄	1	-1	-1	-1	20	-20	-20	-20	25.9338	3.6669	3.8916	22.0423	29.8254
L ₁₄	A ₄	1	-1	-1	-1	25	-25	-25	-25	19.9819	9.2710	6.1879	13.7940	26.1697
										$t_{0.05}(42-8)=$	2.0322			

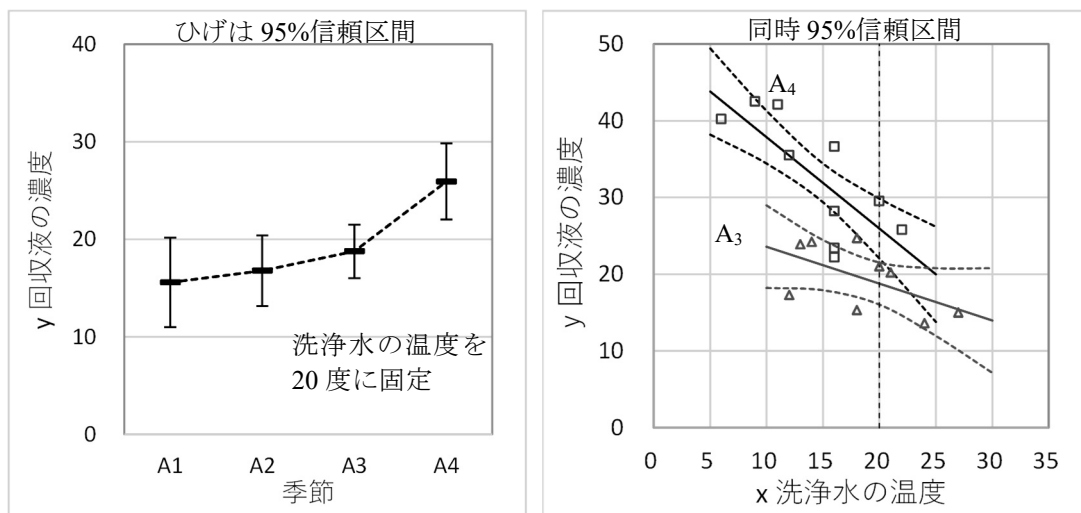


図 1.9 Excel の「折れ線」による予測プロファイルおよび散布図上の回帰直線

Excel での「折れ線グラフ」のひげの長さの設定は、かなり深いところにあり、

グラフの要素 → 誤差範囲 → その他のオプション → ユーザの設定 →
 値の設定 → 正の誤差の値 (範囲) → 負の誤差の値 (範囲)

のように慎重に設定する。気を抜くと尤もらしいまがい物の誤差範囲が表示されるので細心の注意が必要である。

水準間の差の予測プロファイル

回収液の濃度が最も高い季節 A₄ を基準とし、温度が 20 度の場合の他の季節との差の推定値と 95%信頼区間を求めたい。季節 A₃ と A₄ の差の推定値は、表 1.21 に示すように、線形和のための行ベクトル \mathbf{l} を用いて差を計算する。推定値は、

$$\begin{aligned} L_{x=20}^{(A_3-A_4)} &= L_{x=20}^{(A_3)} - L_{x=20}^{(A_4)} \\ &= 18.7719 - 25.9338 = -7.1620 \end{aligned}$$

として求められる。この差に対する分散の推定のためのベクトル $\mathbf{l}^{(A_3-A_4)}$ は、季節 A₃ のベクトル $\mathbf{l}^{(A_3)}$ と季節 A₄ のベクトル $\mathbf{l}^{(A_4)}$ の差

$$\begin{array}{r} \mathbf{l}^{(A_3)} = [1 \quad 0 \quad 0 \quad 1 \quad 20 \quad 0 \quad 0 \quad 20] \\ -) \quad \mathbf{l}^{(A_4)} = [1 \quad -1 \quad -1 \quad -1 \quad 20 \quad -20 \quad -20 \quad -20] \\ \hline \mathbf{l}^{(A_3-A_4)} = \mathbf{l}^{(A_3)} - \mathbf{l}^{(A_4)} = [0 \quad 1 \quad 1 \quad 2 \quad 0 \quad 20 \quad 20 \quad 40] \end{array}$$

であり、 $L_{x=20}^{(A_3-A_4)}$ の分散 $Var[L_{x=20}^{(A_3-A_4)}]$ は、

$$\begin{aligned} Var[L_{x=20}^{(A_3-A_4)}] &= \mathbf{l}^{(A_3-A_4)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) (\mathbf{l}^{(A_3-A_4)})^T \\ &= 5.4834 \end{aligned}$$

として求められ. この結果は, 表 1.14 に示した JMP の「パラメータ関数」に示された結果に一致する. 差 $\bar{A}_3 - \bar{A}_4 = -7.1620$ の 95%信頼区間が $(-11.9208, -2.4031)$ と推定されゼロを含まないので, 有意な差であることが示されている.

表 1.21 2 水準間の差の線形和 L と推定値

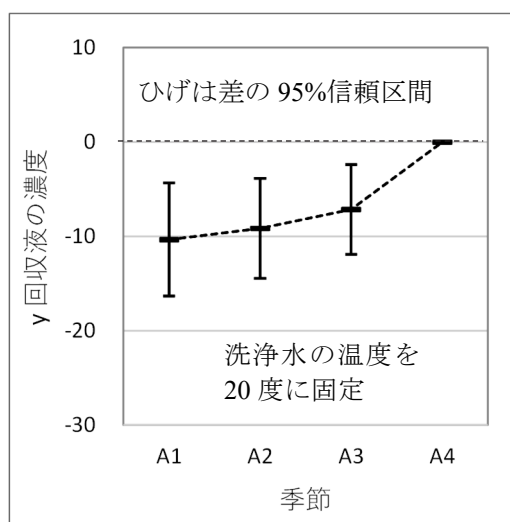
		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$l\hat{\theta}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
L_3	A_3	1	0	0	1	20	0	0	20	18.7719				
L_4	A_4	1	-1	-1	-1	20	-20	-20	-20	25.9338				
差	$A_3 - A_4$	0	1	1	2	0	20	20	40	-7.1620	5.4834	4.7588	-11.9208	-2.4031

同様に季節 A_4 を基準とし, 他の季節の差を表 1.22 に示す. 季節間の差の計算のための行ベクトル l については, 季節 A_3 と季節 A_4 の場合と同様にする. 差の分散についてもそれぞれの差の行ベクトル l の場合と同様に推定し, 95%信頼区間を計算する. JMP でも表 1.14 で示したように, ある温度設定に対し, 因子 A の水準間の差の推定ができる.

表 1.22 季節 A_4 を基準とした 2 水準間の差の推定

		l_0	l_1	l_2	l_3	l_4	l_5	l_6	l_7	$l\hat{\theta}$	分散	幅	95%信頼区間	
L	A	x_0	a_1	a_2	a_3	x	a_1x	a_2x	a_3x	y^\wedge	$Var(y^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
L_{15}	$A_1 - A_4$	0	2	1	1	0	40	20	20	-10.3426	8.7211	6.0015	-16.3441	-4.3411
L_{16}	$A_2 - A_4$	0	1	2	1	0	20	40	20	-9.1496	6.8189	5.3068	-14.4564	-3.8428
L_{17}	$A_3 - A_4$	0	1	1	2	0	20	20	40	-7.1620	5.4834	4.7588	-11.9208	-2.4031
L_{18}	$A_4 - A_4$	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	0.0000	0.0000

図 1.10 に季節 A_4 を基準とした差の推定値および 95%信頼区間の予測プロファイルを示す. この図は, 洗浄水の温度を 20 度に固定した場合であり, 図 1.8 に示したように温度が低くなれば季節 A_4 との差は広がり, 温度が高くなれば差は縮まる.

図 1.10 Excel の「折れ線」による季節 A_4 を基準とした差の予測プロファイル

1.3. 前値を共変量とする交互作用がある抗うつ剤の評価

フライス著, KR(関西臨床データ解析)研究会訳 (2004), 「臨床試験のデザインと解析, 第7章 共分散分析と変化量の検討」にランダム化ブロック試験の30組の患者のうち, 表1.23に示すように, 初めの10組について治療前のうつ症状のHamiltonの評価尺度によるスコア X , 治療後のスコア Y が示されている. この試験データは, 薬剤Aと前値 X に交互作用があるような場合について“共分散分析”による解析の考え方が丁寧に示されている. なお, Hamiltonの評価尺度スコアは, 大きいほどうつ症状は重くなる. ここでは, 簡単のためペアを考慮した解析ではなく, 前値 X を共変量, 薬剤をA, 前後差(変化量) D を反応変数とする解析を行う.

表 1.23 抗うつ薬投与前後の Hamilton 評価尺度データ [フライス (2004), 表 7.6]

No	A ₁ : プラセボ			A ₂ : イミプラミン		
	前値 x	後値 y	前後差 d	前値 x	後値 y	前後差 d
1	18	4	-14	27	6	-21
2	21	7	-14	18	4	-14
3	24	10	-14	25	6	-19
4	24	10	-14	23	7	-16
5	18	2	-16	22	5	-17
6	27	11	-16	24	6	-18
7	21	9	-12	25	8	-17
8	21	5	-16	22	7	-15
9	26	11	-15	26	8	-18
10	20	8	-12	19	3	-16
平均	22.00	7.70	-14.30	23.10	6.00	-17.10
SE	0.99	0.99	0.47	0.92	0.52	0.64
L95%	19.76	5.46	-15.37	21.01	4.83	-18.55
U95%	24.24	9.94	-13.23	25.19	7.17	-15.65

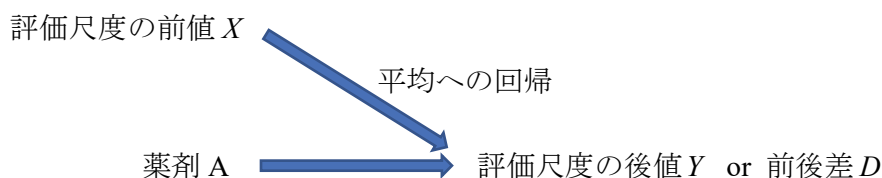


図 1.11 前値が後値に与える影響

前後差 D に対して薬剤間の比較をする際に, 前値 X を共変量とするのは, 「平均への回帰」と言われている現象の影響を取り除いて薬剤間の比較を行うためである. 後値 Y を反応とする場合でも, 同様に前値 X を共変量として含める必要がある. 前後差 D による対応のない t 検定を適用したくなるのであるが, 慎まなければならない.

「平均への回帰」について詳しくは、第9章を参照のこと。なお、「平均への回帰」とは、測定される反応に上限と下限があるような場合に、前値 X が上限に近ければ、後値 Y は、さらに増加するよりも平均値方向に減少しやすく、逆に前値 X が下限に近ければ、後値 Y は、平均方向に増加しやすくなる現象のことである。

非平行な2本の回帰直線

図1.12に示すのは、前値 x と前後差 d の散布図に Excel の「近似曲線の追加」による線形近似（回帰直線）を重ね書きした結果である。前値が小さい場合に薬剤間の前後差 d は小さく、前値が大きい場合に薬剤間の前後差 d が広がっている。このような現象は、評価スコアが0以下にならないために、スコアの減少量に下限ができてしまうためである。

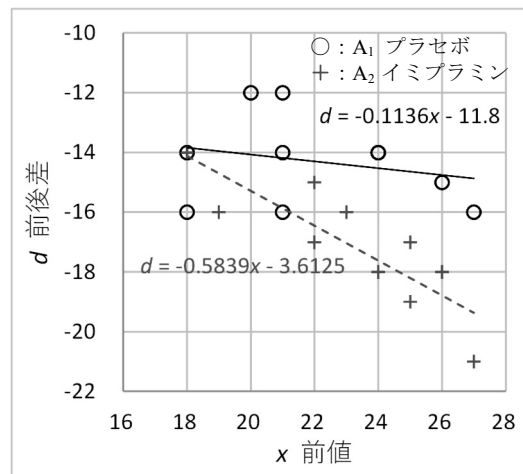


図 1.12 薬剤ごとの前後差 d に対する Excel の線形近似（回帰直線）

この図から、 A_1 プラセボ群は、回帰直線の傾きが小さいので、前値 x の大きさに関わらず前後差 d にはっきりとした違いが見いだせない。他方、 A_2 イミプラミン群は、回帰直線の傾きが大きく、前値 x が大きければ前後差 d も大きく、前値 x が小さくなると前後差 d は、 A_1 プラセボ群に近い前後差 d となることが、読み取れる。

前値 x を共変量とする共分散分析を適用する際に、薬剤 A と前値 x に交互作用 $A \times x$ がないことが必須であるとの思い込みが蔓延している。交互作用 $A \times x$ を統計的に評価するため、薬剤 A についての (1, -1) 対比型ダミー変数を a_i として、交互作用モデルの回帰式、主効果モデルの回帰式

$$\left. \begin{array}{l} \text{交互作用モデル: } d_i = \theta_0 + \theta_1 a_{1,i} + \theta_2 x_i + \theta_3 (a_{1,i} x_i) + \varepsilon_i \\ \text{主効果モデル: } d_i = \theta_0 + \theta_1 a_{1,i} + \theta_2 x_i + \varepsilon_i \end{array} \right\} \quad (1.26)$$

を用いる。薬剤 A が2水準なので、交互作用の検討は、分散分析表によらず $a_{1,i} x_i$ のパラメータ θ_3 に関する p 値で行なうことができる。

表 1.24 左に示すのは、交互作用モデルに対応する (20 行×4 列) のデザイン行列 X である。デザイン行列 X は、(20 行×4 列) であるが、「定数に 0 を使用」をオフ (デフォルト) とした場合に、選択する変数は、切片 x_0 を除いた (a_1, x, a_1x) である。表 1.24 右上段に Excel の回帰分析を適用した交互作用モデルの結果を示す。表 1.24 右下段に示す主効果モデルの場合は、交互作用のための変数 a_1x を除いた結果である。

表 1.24 前後差 d に対する交互作用モデルおよび主効果モデル

		デザイン行列 X											
		交互作用モデル											
		主効果モデル				交互作用モデル							
No	A	x_0	a_1	x	a_1x	y	d	分散分析表 「定数に 0 を使用」 off					
1	A ₁	1	1	18	18	4	-14						
2		1	1	21	21	7	-14	自由度	変動	分散	分散比		
3		1	1	24	24	10	-14	回帰	3	66.5524	22.1841	11.9721	
4		1	1	24	24	10	-14	残差	16	29.6476	1.8530	= σ^2	
5		1	1	18	18	2	-16	合計	19	96.2000			
6		1	1	27	27	11	-16						
7		1	1	21	21	9	-12	θ^{\wedge}_0	切片 x_0	-7.7062	2.4197	-3.1848	0.0058
8		1	1	21	21	5	-16	θ^{\wedge}_1	a_1	-4.0938	2.4197	-1.6918	0.1101
9		1	1	26	26	11	-15	θ^{\wedge}_2	x	-0.3488	0.1062	-3.2825	0.0047
10		1	1	20	20	8	-12	θ^{\wedge}_3	a_1x	0.2351	0.1062	2.2130	0.0418
11	A ₂	1	-1	27	-27	6	-21						
12		1	-1	18	-18	4	-14						
13		1	-1	25	-25	6	-19						
14		1	-1	23	-23	7	-16						
15		1	-1	22	-22	5	-17						
16		1	-1	24	-24	6	-18						
17		1	-1	25	-25	8	-17						
18		1	-1	22	-22	7	-15						
19		1	-1	26	-26	8	-18	$\theta^{\wedge}'_0$	切片 x_0	-8.1924	2.6717	-3.0664	0.0070
20		1	-1	19	-19	3	-16	$\theta^{\wedge}'_1$	a_1	1.2169	0.3436	3.5415	0.0025
平均				22.550	6.850		-15.70	$\theta^{\wedge}'_2$	x	-0.3329	0.1175	-2.8327	0.0115

推定されたパラメータによる交互作用の検討

表 1.24 右上段のパラメータの推定値から交互作用を含む回帰式は、

$$\left. \begin{aligned} \hat{d}^{(\ast)} &= \hat{\theta}_0 + \hat{\theta}_1 a_1 + \hat{\theta}_2 x + \hat{\theta}_3 a_1 x \\ &= -7.7062 - 4.0938 a_1 - 0.3488 x + 0.2351 a_1 x \end{aligned} \right\} \quad (1.27)$$

となる。(A₁, A₂) それぞれの回帰式は、

$$\left. \begin{aligned} \hat{d}^{(\ast, A_1)} &= -7.7062 - 4.0938 \times 1 - 0.3488 x + 0.2351 \times 1 \times x \\ &= -11.8000 - 0.1136 x \\ \hat{d}^{(\ast, A_2)} &= -7.7062 - 4.0938 \times (-1) - 0.3488 x + 0.2351 \times (-1) \times x \\ &= -3.6125 - 0.5839 x \end{aligned} \right\} \quad (1.28)$$

として求められ、これらは、図 1.12 で示されている回帰式に一致する。

交互作用の検討は、変数 a_{1x} のパラメータ $\hat{\theta}_3 = 0.2351$ 、標準誤差 $SE(\hat{\theta}_3) = 0.1062$ を用い、 t 値と p 値は、

$$\left. \begin{aligned} t &= \frac{\hat{\theta}_3}{SE(\hat{\theta}_3)} = \frac{0.2351}{0.1062} = 2.2130 \\ p &= \text{T.dist.2T}(2.2130, 16) = 0.0418 \end{aligned} \right\} \quad (1.29)$$

となり、 $p < 0.05$ なので、有意な交互作用があると判断される。

分散分析表による交互作用の検討

交互作用の検討は、交互作用モデルの分散分析表と主効果モデルの分散分析表を合成することで行うことができる。表 1.24 右下段に示されているパラメータから主効果モデルの回帰式は、

$$\left. \begin{aligned} \hat{d}^{(\pm)} &= \hat{\theta}'_0 + \hat{\theta}'_1 a_1 + \hat{\theta}'_2 x \\ &= -8.1924 + 1.2169 a_1 - 0.3329 x \end{aligned} \right\} \quad (1.30)$$

となる。共通の傾きは、 $\hat{\theta}'_2 = -0.3329$ であり、それぞれの切片は、 $\hat{\theta}'_0 \pm \hat{\theta}'_1 = (-6.9756, -9.4093)$ となる。

表 1.24 右上段の交互作用モデルの分散分析表から、交互作用を含む残差平方和 $S_{\text{残差}}^{(\text{交})}$ 、表 1.24 右下段の主効果モデルの分散分析表から、交互作用を含まない平方和 $S_{\text{残差}}^{(\text{主})}$ が、

$$\left. \begin{aligned} S_{\text{残差}}^{(\text{交})} &= \sum_i (d_i - \hat{d}_i^{(\text{交})})^2 = 29.6476 \\ S_{\text{残差}}^{(\text{主})} &= \sum_i (d_i - \hat{d}_i^{(\text{主})})^2 = 38.7222 \end{aligned} \right\} \quad (1.31)$$

求められている。表 1.25 に示すように、交互作用モデルの偏差平方和 $S_{\text{残差}}^{(\text{交})}$ と主効果モデルの偏差平方和 $S_{\text{残差}}^{(\text{主})}$ のとの差が交互作用の平方和 $S_{A \times x}$

$$\left. \begin{aligned} S_{A \times x} &= S_{\text{残差}}^{(\text{主})} - S_{\text{残差}}^{(\text{交})} \\ &= 38.7222 - 29.6476 = 9.0746 \end{aligned} \right\} \quad (1.32)$$

となる。これは、交互作用モデルの残差の平方和 $S_{\text{残差}}^{(\text{交})}$ には、交互作用の平方和 $S_{A \times x}$ が含まれていないために、交互作用が含まれている主効果モデルの残差平方和 $S_{\text{残差}}^{(\text{主})}$ との差を取ることで交互作用の平方和 $S_{A \times x}$ を取り出すことができる。

表 1.25 分散分析表 (交互作用モデル, 平方和の分離)

	自由度 df	平方和(変動)	平均平方(分散)	F (分散比)	p 値
回帰($a_1 + x + a_{1x}$)	3	66.5524	22.1841	11.9721	0.0002
($a_1 + x$) 主効果	2	57.4778	28.7389	15.5096	0.0002
(a_{1x}) 交互作用	1	9.0746	9.0746	4.8973	0.0418
残差	16	29.6476	1.8530	1.0000	
合計	19	96.2000			
				$p = \text{F.dist.RT}$ (F 値, 分子の df , 分母の df)	

残差の平方和ではなく回帰の平方和の差からも交互作用の平方和を取り出すことができる。表 1.24 から交互作用モデルの回帰の平方和は、 $S_{回帰}^{(交)} = 66.5524$ であり、主効果モデルの回帰の平方和は、 $S_{回帰}^{(主)} = 57.4778$ である。この差からも交互作用の平方和 $S_{A_{xx}}$ が

$$\left. \begin{aligned} S_{A_{xx}} &= S_{回帰}^{(交)} - S_{回帰}^{(主)} \\ &= 66.5524 - 57.4778 = 9.0746 \end{aligned} \right\} \quad (1.33)$$

となり、残差の平方和の差に一致する。表 1.25 に示すように、交互作用の F 値から p 値を求めると 0.0418 となり、式 (1.29) で示した回帰パラメータに対する p 値に一致する。

非平行な 2 本の回帰直線の差

交互作用が有意な差ではなく、平行な 2 本の直線があてはめられれば、 A_2 イミプラミン グ群の A_1 プラセボ群に対する薬効は、式 (1.14) で示したように 2 本の直線間の差が、切片の差と同じになる。さて、交互作用が有意となった場合には、前値 x の大きさにより 2 本の直線間の差は異なり、どのような要約統計量によって薬効を論ずるかが課題となる。恣意性が入り込まないように、交互作用が有意となった場合は、表 1.24 左下段に示した前値 x の総平均 $\bar{x}_{..} = 22.550$ における回帰直線の推定値の差で評価するなど、前もって定めておく必要がある。

交互作用がない場合の平行な 2 本の直線間の差は、前値 x が変化しても同じであり、そのために、前値 $x=0$ の切片のパラメータの差が群間差となり、その 95%信頼区間から群間差について論ずることができる。ただし、交互作用がある場合は、前値 x の変化によって差が変化するので、 A_2 イミプラミンと A_1 プラセボの 2 本の回帰直線の差 ($\hat{d}^{(交, A_2)} - \hat{d}^{(交, A_1)}$) について、前値 x を変化させたときの分散が必要となる。2 本の回帰直線の差は、

$$\left. \begin{aligned} A_2 \text{の回帰式: } \hat{d}^{(交, A_2)} &= (\hat{\theta}_0 - \hat{\theta}_1) + (\hat{\theta}_2 - \hat{\theta}_3)x = -3.6125 - 0.5839x \\ A_1 \text{の回帰式: } \hat{d}^{(交, A_1)} &= (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_2 + \hat{\theta}_3)x = -11.8000 - 0.1136x \end{aligned} \right\} \quad (1.34)$$

$$\text{差} \quad \left\{ \begin{aligned} (\hat{d}^{(交, A_2)} - \hat{d}^{(交, A_1)}) &= -2\hat{\theta}_1 - 2\hat{\theta}_3x \\ &= -2 \times (-4.0938) - 2 \times 0.2351x \\ &= 8.1875 - 0.4702x \end{aligned} \right\} \quad (1.35)$$

となる。この回帰式の差の推定値の分散は、合成分散の一般式から

$$\left. \begin{aligned} Var(\hat{d}^{(交, A_2)} - \hat{d}^{(交, A_1)}) &= Var(-2\hat{\theta}_1 - 2\hat{\theta}_3x) \\ &= 4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x + 4Var(\hat{\theta}_3)x^2 \end{aligned} \right\} \quad (1.36)$$

が得られる。これらのパラメータの分散および共分散は、表 1.24 に示したデザイン行列 X と交互作用モデルの分散分析表の誤差分散の推定値 $\hat{\sigma}_{(交互)}^2 = 1.8530$ を用いて、式 (1.10) および式 (1.18) と同様にパラメータの共分散行列 $\Sigma(\hat{\theta})$ が、

$$\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}_{(交互)}^2 \quad (1.37)$$

$$\Sigma(\theta^\wedge) = (X^T X)^{-1} \sigma^2_{交互}$$

$\Sigma(\theta^\wedge) =$	5.8550	-0.6666	-0.2550	0.0233	θ^\wedge_0
	-0.6666	5.8550	0.0233	-0.2550	θ^\wedge_1
	-0.2550	0.0233	0.0113	-0.0008	θ^\wedge_2
	0.0233	-0.2550	-0.0008	0.0113	θ^\wedge_3
	x_0	a_1	x	$a_1 x$	

として得られ、 $Var(\hat{\theta}_1) = 5.8550$, $Cov(\hat{\theta}_1, \hat{\theta}_3) = -0.2550$, $Var(\hat{\theta}_3) = 0.0113$ となるので、式 (1.35) の x を変化させ、分散を計算することができる。

2本の直線間の差に対する95%信頼区間およびt検定

表 1.26 に示すのは、図 1.12 に示したように、前値 x を 16 から 2 刻みで 30 まで増加させた場合の各種の推定値である。ちなみに、 $x = 22$ における $(\hat{d}_{A_2, x=22}^{(交)} - \hat{d}_{A_1, x=22}^{(交)})$ の p 値が、 $p = 0.0036$ と有意な差となっている。

表 1.26 非平行な2本の回帰直線の差の推定値に対するt検定

x	$\hat{d}_{A_1}^{(交)}$	$\hat{d}_{A_2}^{(交)}$	差	$Var(\text{差})$	$SE(\text{差})$	L95%	U95%	t値	p値
16	-13.62	-12.95	0.66	2.3433	1.5308	-2.5814	3.9088	0.4336	0.6704
18	-13.85	-14.12	-0.28	1.3342	1.1551	-2.7255	2.1719	-0.2396	0.8137
20	-14.07	-15.29	-1.22	0.6864	0.8285	-2.9736	0.5390	-1.4693	0.1611
22	-14.30	-16.46	-2.16	0.3998	0.6323	-3.4981	-0.8174	-3.4127	0.0036 *
22.550	-14.36	-16.78	-2.42	0.3843	0.6199	-3.7305	-1.1023	-3.8981	0.0013 *
24	-14.53	-17.63	-3.10	0.4743	0.6887	-4.5582	-1.6382	-4.4985	0.0004 *
26	-14.75	-18.79	-4.04	0.9101	0.9540	-6.0611	-2.0163	-4.2334	0.0006 *
28	-14.98	-19.96	-4.98	1.7072	1.3066	-7.7490	-2.2093	-3.8108	0.0015 *
30	-15.21	-21.13	-5.92	2.8654	1.6928	-9.5081	-2.3312	-3.4970	0.0030 *
				$t_{0.05(20-4)} =$	2.1199				

図 1.13 に示すのは、表 1.26 で求められた2本の回帰直線の差の推定値および差の95%信頼区間である。

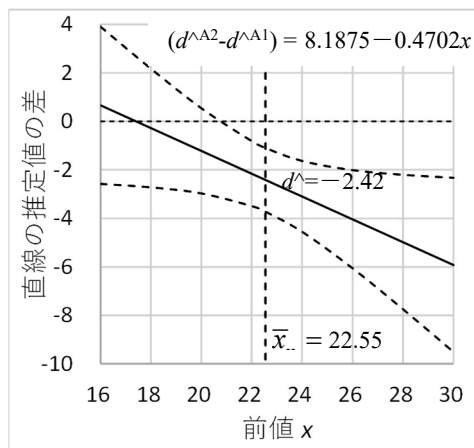


図 1.13 2本の回帰直線の差の推定値に対する95%信頼区間

前値が総平均 $\bar{x}_{..} = 22.550$ の場合についての予測値は、式 (1.34) および式 (1.35) より

$$\left. \begin{aligned} \hat{d}_{x=22.55}^{(交, A_1)} &= -11.8000 - 0.1136 \times 22.550 = -14.36 \\ \hat{d}_{x=22.55}^{(交, A_2)} &= -3.6125 - 0.5839 \times 22.550 = -16.78 \\ \hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)} &= -2.42 \end{aligned} \right\} \quad (1.38)$$

となる。分散は、式 (1.36) より、

$$\begin{aligned} Var(\hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)}) &= 4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x + 4Var(\hat{\theta}_3)x^2 \\ &= 4 \times 5.8550 + 8 \times (-0.2550) \times 22.550 + 4 \times 0.0113 \times 22.550^2 \\ &= 0.3843 \end{aligned} \quad (1.39)$$

となり、 t 値および p 値は、

$$\left. \begin{aligned} t_{x=20} &= \frac{\hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)}}{\sqrt{Var(\hat{d}_{x=22.55}^{(交, A_2)} - \hat{d}_{x=22.55}^{(交, A_1)})}} \\ &= \frac{-2.42}{\sqrt{0.3843}} = -3.8981 \end{aligned} \right\} \quad (1.40)$$

$$\left. \begin{aligned} p &= T.\text{dist}.2T(|t|, 20 - 4) \\ &= T.\text{dist}.2T(|-3.8981|, 16) = 0.0013 \end{aligned} \right\} \quad (1.41)$$

として計算されている。したがって、有意な差があると判定される。前値が $x = 22$ の場合も同様に $p = 0.0036$ が得られ、有意な差と判定される。 $x = 22$ 以上の場合も同様に有意な差となっている。これらの結果から、 A_2 イミプラミンは、 A_1 プラセボに対し、前値が $x = 22$ 以上の場合に有意な差があると判定される。

2本の回帰直線の差が有意となる前値 x_0

前値の総平均 $\bar{x}_{..} = 22.550$ における p 値による判定だけでなく、2本の回帰直線の差の95%信頼区間の上限が0となる前値 x_0 が推定できれば、「 x_0 以上であれば、統計的に有意な差があると言える」ので、歯切れがよい。95%信頼区間の上限の式は、

$$\hat{y}_0 = (2\hat{\theta}_1 + 2\hat{\theta}_3 x_0) + t_{0.05}(20 - 4) \sqrt{4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x_0 + 4Var(\hat{\theta}_3)x_0^2} \quad (1.42)$$

で与えられるので、 $\hat{y}_0 = 0$ となる x_0 を求めるためには、

$$(-2\hat{\theta}_1 - 2\hat{\theta}_3 x_0)^2 = t_{0.05}^2(16) \cdot [4Var(\hat{\theta}_1) + 8Cov(\hat{\theta}_1, \hat{\theta}_3)x_0 + 4Var(\hat{\theta}_3)x_0^2] \quad (1.43)$$

を整理すると x_0 に関して2次式となるので、2次式の解の公式を用いて解を求めればよいが、煩雑な式となる。救いは、Excelのソルバーであり、式 (1.42) で $\hat{y}_0 = 0$ となるような x_0 を探索的に求めてくれる。表 1.27 に示すのは、 $x = 21.0000$ の場合の上限が -0.1916 であり、Excelのソルバーで上限が 0.0000 となるように x を探索的に変化させた結果とし、 $x = 20.7260$ が得られた結果である。この結果として、 $p = 0.0500$ と有意な差となる x の推定値が得られたことになる。

表 1.27 Excel ソルバーによる 95%信頼区間上限が 0 となる x_0 の推定

	x	d^{\wedge}_{A1}	d^{\wedge}_{A2}	$d^{\wedge}_{A2}-d^{\wedge}_{A1}$	Var(差)	SE(差)	L95%	U95%	t 値	p 値
	20	-14.07	-15.29	-1.22	0.6864	0.8285	-2.9736	0.5390	-1.4693	0.1611
	21.0000	-14.19	-15.87	-1.69	0.4979	0.7056	-1.6875	-0.1916	-2.3915	0.0294
	22	-14.30	-16.46	-2.16	0.3998	0.6323	-2.1577	-0.8174	-3.4127	0.0036 *
	20	-14.07	-15.29	-1.22	0.6864	0.8285	-1.2173	0.5390	-1.4693	0.1611
#	20.7260	-14.16	-15.71	-1.56	0.5406	0.7352	-1.5586	0.0000	-2.1199	0.0500 *
	22	-14.30	-16.46	-2.16	0.3998	0.6323	-2.1577	-0.8174	-3.4127	0.0036 *
#	ソルバー解				$t_{0.05}(20-4)=$	2.1199			*: $p < 0.05$	

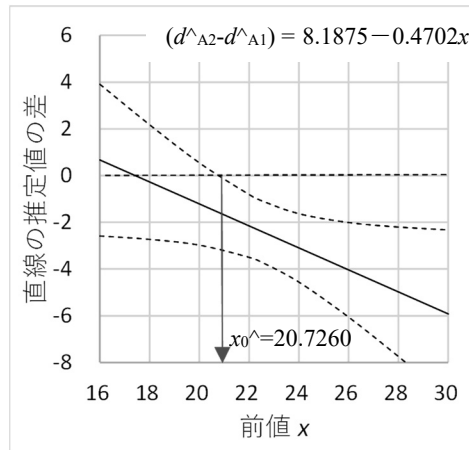


図 1.14 有意な差となる前値 x_0 の推定

JMP による検証

JMP による交互作用を含む分散分析表を表 1.28 に示す。分散分析表は、逐次（タイプ 1）検定（平方和）を選択する。薬剤 A の逐次平方和 $S_A = 39.2000$ に前値 x の逐次平方和 $S_x = 18.2778$ を加えた $S_{A+x} = 57.4778$ は、表 1.25 の自由度 2 の主効果の平方和に一致する。さらに交互作用 $S_{A*x} = 9.0746$ は、表 1.25 の交互作用の平方和に一致する。

表 1.28 JMP による交互作用を含む分散分析表

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	3	66.5524	22.1841	11.9721
誤差	16	29.6476	1.8530	p値(Prob>F)
全体(修正済み)	19	96.2000		0.0002*

逐次(タイプ1)検定				
要因	自由度	逐次平方和	F値	p値
A	1	39.2000	21.1551	0.0003*
x	1	18.2778	9.8640	0.0063*
A*x	1	9.0746	4.8973	0.0418*

1.4. 発電プラント周辺でのカキの成長

Littell ら (2002), 「SAS for Linear Models 4th ed., Chapter 7 Analysis of Covariance」に興味深い事例がある。第 7.2 節の「1 元配置型の共分散分析」は、カキの成長に関する実験データの解析事例である。ただし、2×2 元配置型の 4 群と対照群の 1 群を加えた実験であり、全体で 5 群の 1 元配置型としている。

前値を共変量とする対照群を含む 2 元配置

実験は、発電プラントの（取水路、放水路）、近くの湾の浅い部分、3 箇所でのカキの成長の比較である。発電プラントの取水路と放水路では、それぞれの水路の（上部、下部）に 4 籠を設置し、湾内では中ほどの水深に 4 籠が設置された。表 1.29 に示すように、それぞれの籠の中には、10 個のカキが入れられ、実験の最初に洗浄されカキの重量が測定され、1 か月後に再び洗浄されたカキの重量である。

実験の目的は、A₁ 取水路に比較して、発電プラントで人工的に加熱された A₂ 放水路でのカキの成長に与える影響を知りたい。さらに、B：水深（上部、下部）での成長与える影響も合わせて検討したい。それらを C：湾内での成長とも比較したいなどである。図 1.15 に示すように、initial 重量を共変量とするが、重量が減少することはなく増加のみで明確な上限もないので、「平均への回帰」現象は起きない。

表 1.29 カキの成長実験データ [Littell ら (2002), Output 7.1]

B: 上下	A: 水路						C: 湾内		
	籠	A ₁ : 取水路		A ₂ : 放水路		籠	B: 位置: 中ほど		
		x: initial	y: final	x: initial	y: final		x: initial	y: final	
B ₁ : 上部	1	27.2	32.6	28.6	35.2	1	20.4	24.6	
	2	32.0	36.6	22.4	29.1	2	19.6	23.4	
	3	33.0	37.7	23.2	28.9	3	25.1	30.3	
	4	26.8	31.0	24.4	30.2	4	18.1	21.8	
B ₂ : 下部	1	28.6	33.8	29.3	35.0				
	2	26.8	31.7	21.8	27.0				
	3	26.5	30.7	30.3	36.4				
	4	26.8	30.4	24.3	30.5				

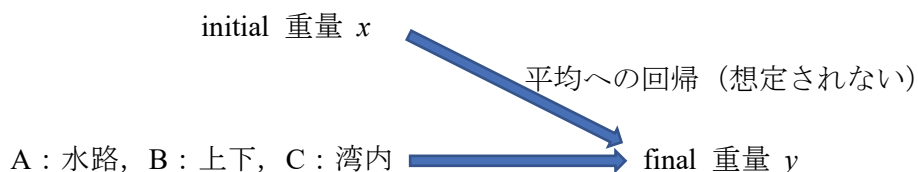


図 1.15 設置場所がカキの成長に与える影響

図 1.16 に示すのは、データの全体像を把握するために、Excel の散布図で 5 群の前値 x 、後値 y を重ね書きし、それぞれの群ごとに回帰直線を引いた結果である。C：湾内設置の前値 x が他に比べて明らかに小さい。カキの洗浄後の設置場所による 4 籠の前値の平均値を 5 群で揃えるような“群分け”が行われなかったようであり、設置が同時ではなく何日かけて逐次的に行なわれているとすれば、その間にカキがいくばくか成長した後に設置したように推測されるが、詳細は不明である。

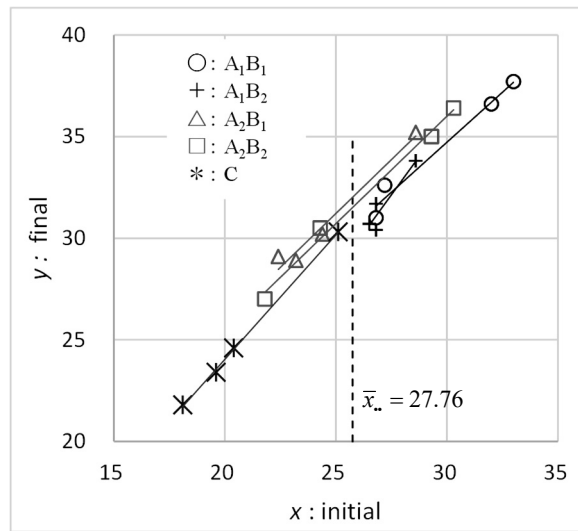


図 1.16 設置場所によるカキの成長

共変量を含む 1 元配置としてモデル化

1 か月後のカキの成長は、設置場所により 4~6 単位程度であり、C：湾内と A₁：取水路は、前値 x が全く異なるものの回帰直線を引き延ばせば互いに重なり合い、カキの成長は同程度と思われる。それらに対し、A₂：放水路は、2 単位ほど成長が早いようである。

込み入った群構成になっているので、表 1.30 に示すように全体で 5 群 (G₁~G₅) としてダミー変数の設定を行う。SAS の解析結果と合わせるため、最後の水準を 0 とする (1, 0) 型のダミー変数 (g₁, g₂, g₃, g₄) を使うことにする。推定される回帰パラメータは、G₅ 群 (C：湾内) との差となる。切片の推定値は、G₅ 群の切片となるので、平行線をあてはめた場合には、前値 x の大きさに関わらず G₅ 群 (C：湾内) と各群との差の推定値となる。

表 1.30 (1, 0) 型ダミー変数による群構成

A:水路	B:上下	群G	(1, 0)型ダミー変数			
			g ₁	g ₂	g ₃	g ₄
A ₁ :取水路	B ₁ :上部	G ₁	1	0	0	0
"	B ₂ :下部	G ₂	0	1	0	0
A ₂ :放水路	B ₁ :上部	G ₃	0	0	1	0
"	B ₂ :下部	G ₄	0	0	0	1
C:湾内	中ほど	G ₅	0	0	0	0

表 1.31 に示すように、切片を x_0 、各群に対し (g_1, g_2, g_3, g_4) 、共変量を x とした (20 行×6 列) のデザイン行列 X に対し、Excel の回帰分析を適用する。回帰のモデル式は、図 1.16 から交互作用はないと判断されるので、主効果モデル

$$y_i = \theta_0 x_{0,i} + \theta_1 g_{1,i} + \theta_2 g_{2,i} + \theta_3 g_{3,i} + \theta_4 g_{4,i} + \theta_5 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (1.44)$$

を適用する。

1 元配置の水準の組み合わせ平均

この実験で知りたいのは、水路 A (A_1, A_2) による違いはあるのか、水深 B (B_1, B_2) による違いはあるのか、さらには、A と B の交互作用はあるのか、などである。このためには、湾内 C を除いて 2 元配置型での共分散分析とすれば可能と思われるが、比較の基準となる湾内 C を外すことはできない。式 (1.44) で推定されるのは湾内 C との差のみであるが、パラメータの共分散行列を活用すれば、事後的にどのような推定でも可能となることを示す。

表 1.31 右下段に示すのは、パラメータの共分散行列 $\Sigma(\hat{\theta})$ であり、Excel の行列関数を使いデザイン行列 X と分散分析表の誤差分散の推定値 $\hat{\sigma}^2 = 0.3016$ を用いて

$$\begin{aligned} \Sigma(\hat{\theta}) &= (X^T X)^{-1} \hat{\sigma}^2 \\ &= \text{Minverse}(\text{Mmult}(\text{Transpose}(X \text{ の範囲}), X \text{ の範囲})) * \hat{\sigma}^2 \end{aligned}$$

のように計算した結果である。

表 1.31 (1, 0) 型のダミー変数を用いた回帰分析の結果

				デザイン行列 X						分散分析表「定数に 0 を使用」off							
A	B	G	籠	x_0	g_1	g_2	g_3	g_4	x	y	自由度	変動	分散	分散比			
A ₁	B ₁	G ₁	1	1	1	0	0	0	27.2	32.6							
			2	1	1	0	0	0	32.0	36.6	回帰	5	354.45	70.8894	235.05		
			3	1	1	0	0	0	33.0	37.7	残差	14	4.2223	0.3016	= σ^2		
			4	1	1	0	0	0	26.8	31.0	合計	19	358.67				
	B ₂	G ₂	1	1	0	1	0	0	28.6	33.8							
			2	1	0	1	0	0	26.8	31.7							
			3	1	0	1	0	0	26.5	30.7	θ^{\wedge}_0 切片 x_0	2.4949	1.0279	2.4272	0.0293		
			4	1	0	1	0	0	26.8	30.4	θ^{\wedge}_1 g_1	-0.2445	0.5766	-0.4240	0.6780		
A ₂	B ₁	G ₃	1	1	0	0	1	0	28.6	35.2	θ^{\wedge}_2 g_2	-0.2803	0.4929	-0.5686	0.5786		
			2	1	0	0	1	0	22.4	29.1	θ^{\wedge}_3 g_3	1.6548	0.4294	3.8534	0.0018		
			3	1	0	0	1	0	23.2	28.9	θ^{\wedge}_4 g_4	1.1071	0.4718	2.3468	0.0342		
			4	1	0	0	1	0	24.4	30.2	θ^{\wedge}_5 x	1.0832	0.0476	22.7461	0.0000		
	B ₂	G ₄	1	1	0	0	0	1	29.3	35.0							
			2	1	0	0	0	1	21.8	27.0							
			3	1	0	0	0	1	30.3	36.4	θ^{\wedge}_0	1.0565	0.3468	0.2253	0.1062	0.1899	-0.0472
			4	1	0	0	0	1	24.3	30.5	θ^{\wedge}_1	0.3468	0.3324	0.2048	0.1535	0.1896	-0.0203
C		G ₅	1	1	0	0	0	0	20.4	24.6	θ^{\wedge}_2	0.2253	0.2048	0.2430	0.1311	0.1567	-0.0145
			2	1	0	0	0	0	19.6	23.4	θ^{\wedge}_3	0.1062	0.1535	0.1311	0.1844	0.1245	-0.0087
			3	1	0	0	0	0	25.1	30.3	θ^{\wedge}_4	0.1899	0.1896	0.1567	0.1245	0.2225	-0.0128
			4	1	0	0	0	0	18.1	21.8	θ^{\wedge}_5	-0.0472	-0.0203	-0.0145	-0.0087	-0.0128	0.0023
総平均				$\bar{x}_.$	= 25.76				30.85		x_0	g_1	g_2	g_3	g_4	x	

表 1.31 右中段は, Excel の回帰分析で求めた係数(パラメータ)で, 切片の推定値 $\hat{\theta}_0 = 2.4949$ は, G₅群(C:湾内)の切片の推定値となる. G₁群(A₁:取水路, B₁:上部)推定値は, $\hat{\theta}_1 = -0.2445$ であるが, これは G₅群(C:湾内)の切片との差であり, G₁群の切片の推定値は,

$$G_1 \text{ 群の切片} = \hat{\theta}_0 + \hat{\theta}_1 = 2.4949 + (-0.2445) = 2.2504 \quad (1.45)$$

となる. 交互作用を含まない主効果モデルなので, G₅群(C:湾内)は, 対照群として設定され, 対照群と G₁群(A₁:取水路, B₁:上部)の切片の差が, 前値 x に関係なく 2 本の回帰直線の差の推定値になっている. したがって, $\hat{\theta}_1 = -0.2445$ に対する t 検定の p 値は $p_1 = 0.6780$ であり, 有意な差ではないことが分かる. G₃群および G₄群は, (A₂:放水路)に設置した場合であり, (B₁:上部, B₂:下部)ともに G₅群(C:湾内)に比べて有意な差 ($p_3 = 0.0018$, $p_4 = 0.0342$) であることが読み取れる.

パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根は, それぞれの係数(パラメータ)の標準誤差に相当する. 第 3 群 g_3 の推定値 $\hat{\theta}_3 = 1.6548$ の $SE_3 = 0.4294$ は, $\Sigma(\hat{\theta})$ の 4 行 4 列目 $Var(\hat{\theta}_3) = 0.1844$ の平方根 $\sqrt{0.1844} = 0.4294$ に一致する. パラメータの共分散行列 $\Sigma(\hat{\theta})$ は, 回帰分析で出力されていない各種の推定に対する 95%信頼区間を求めるために使われる. カキ籠の設置位置(B₁:上部, B₂:下部)間の推定値の差の 95%信頼区間を求めたい, 推定されたそれぞれの回帰直線の 95%信頼区間を図で示したい, などである.

最も基本的な推定は, 各群の共変量である前値 x の影響を除くために前値 x の総平均 $\bar{x} = 25.76$ におけるそれぞれの回帰直線の推定値とその 95%信頼区間であり, それらを図示することも結果の解釈に役立つ. そのためには, 表 1.32 に示すようにパラメータの推定値の列ベクトル $\hat{\theta}$ に対し, 推定目的に合致した行ベクトル $\mathbf{l}^{(i)}$ を設定し, それらの線形和 $\hat{L}^{(i)}$

$$\hat{L}^{(i)} = \mathbf{l}^{(i)} \hat{\theta} = \sum_{k=0}^5 l_k^{(i)} \hat{\theta}_k \quad (1.46)$$

で推定する. 線形和 $\hat{L}^{(i)}$ の分散 $Var(\hat{L}^{(i)})$ は,

$$Var(\hat{L}^{(i)}) = \mathbf{l}^{(i)} \Sigma(\hat{\theta}) (\mathbf{l}^{(i)})^T \quad (1.47)$$

表 1.32 前値の総平均 $\bar{x} = 25.76$ に対する回帰直線上の推定値 (最小 2 乗平均)

A, B,		線形和	l_0	l_1	l_2	l_3	l_4	l_5		
C	G	L	x_0	g_1	g_2	g_3	g_4	x	$\hat{\theta}$	$\mathbf{l}^{(i)} \hat{\theta}$
A ₁ B ₁	G ₁	$L^{(1)}$	1	1	0	0	0	25.76	2.4949	= 30.1531
A ₁ B ₂	G ₂	$L^{(2)}$	1	0	1	0	0	25.76	-0.2445	30.1173
A ₂ B ₁	G ₃	$L^{(3)}$	1	0	0	1	0	25.76	-0.2803	32.0523
A ₂ B ₂	G ₄	$L^{(4)}$	1	0	0	0	1	25.76	1.6548	31.5047
C	G ₅	$L^{(5)}$	1	0	0	0	0	25.76	1.1071	30.3976
									1.0832	

$L^{(i)} = \text{Mmult}(\mathbf{l}^{(i)} \text{ の範囲}, \hat{\theta} \text{ の範囲})$

によって推定することができる。各群の $\bar{x}_j = 25.76$ における推定に必要な線形和 $L^{(i)}$ に対する行ベクトル $\mathbf{l}^{(i)}$ を与え、 $\hat{L}^{(i)} = \mathbf{l}^{(i)} \hat{\boldsymbol{\theta}}$ によって求めている。 $L^{(1)}$ の場合であれば、

$$L^{(1)} = 1 \times 2.49491 + 1 \times (-0.2445) + 0 + 0 + 0 + 25.76 \times 1.0832 = 30.1531 \quad (1.48)$$

が得られる。推定値を得るためには、四則演算で対応できるのであるが、その推定値の分散を手計算レベルで計算するのは難儀であり、表 1.33 に示すように Excel の行列関数によるパラメータの共分散行列に関する 2 次形式により簡単に計算できる。

表 1.33 各群の回帰直線上の推定値および第 5 群との差の 95%信頼区間

	線形和			l_0	l_1	l_2	l_3	l_4	l_5	推定値	分散	幅	95%信頼区間	
	L	AB	G	x_0	g_1	g_2	g_3	g_4	x				L^\wedge	$Var(L^\wedge)$
前値 x	$L^{(1)}$	A ₁ B ₁	G ₁	1	1	0	0	0	25.76	30.1531	0.1115	0.72	29.44	30.87
総平均	$L^{(2)}$	A ₁ B ₂	G ₂	1	0	1	0	0	25.76	30.1173	0.0799	0.61	29.51	30.72
25.76	$L^{(3)}$	A ₂ B ₁	G ₃	1	0	0	1	0	25.76	32.0523	0.0782	0.60	31.45	32.65
	$L^{(4)}$	A ₂ B ₂	G ₄	1	0	0	0	1	25.76	31.5047	0.0764	0.59	30.91	32.10
	$L^{(5)}$	C	G ₅	1	0	0	0	0	25.76	30.3976	0.1312	0.78	29.62	31.17
G ₅ との差 (C群)	$L^{(6)}$	A ₁ B ₁	G ₁	0	1	0	0	0	0.00	-0.2445	0.3324	1.24	-1.48	0.99
	$L^{(7)}$	A ₁ B ₂	G ₂	0	0	1	0	0	0.00	-0.2803	0.2430	1.06	-1.34	0.78
	$L^{(8)}$	A ₂ B ₁	G ₃	0	0	0	1	0	0.00	1.6548	0.1844	0.92	0.73	2.58
	$L^{(9)}$	A ₂ B ₂	G ₄	0	0	0	0	1	0.00	1.1071	0.2225	1.01	0.10	2.12
	$L^{(10)}$	C	G ₅	0	0	0	0	0	0.00	0.0000	0.0000	0.00	0.00	0.00
										$t_{0.05}(20-4) =$	2.1448			

$\mathbf{l}^{(1)T} =$	l_0	l_1	l_2	l_3	l_4	l_5	パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2$						$\mathbf{l}^{(1)T}$
$Var(L^\wedge) =$	1	1	0	0	0	25.76	1.0565	0.3468	0.2253	0.1062	0.1899	-0.0472	1
	x_0	g_1	g_2	g_3	g_4	x	0.3468	0.3324	0.2048	0.1535	0.1896	-0.0203	1
							0.2253	0.2048	0.2430	0.1311	0.1567	-0.0145	0
							0.1062	0.1535	0.1311	0.1844	0.1245	-0.0087	0
							0.1899	0.1896	0.1567	0.1245	0.2225	-0.0128	0
							-0.0472	-0.0203	-0.0145	-0.0087	-0.0128	0.0023	25.76
$=$	0.1115						$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	

分散が計算できれば、推定値 \hat{y} の 95%信頼区間は、標準的な計算方式により

$$\begin{aligned} (L95\%, U95\%) &= \hat{L}^{(1)} \pm t(0.05, 20-6) \sqrt{Var(\hat{L}^{(1)})} \\ &= 30.1531 \pm 2.1448 \sqrt{0.1115} \\ &= (29.44, 30.87) \end{aligned}$$

として求められる。

さらに、第 5 群との差については、表 1.31 に示した Excel による回帰分析の推定値が使えるのであるが、表 1.33 の下段に示すように、差のベクトル $\mathbf{l}^{(6)} \sim \mathbf{l}^{(10)}$ を与えて計算式のフィルハンドルを用いて簡単に計算することもできる。これらの推定結果を、図 1.17 に示すように Excel の折れ線グラフで表す。

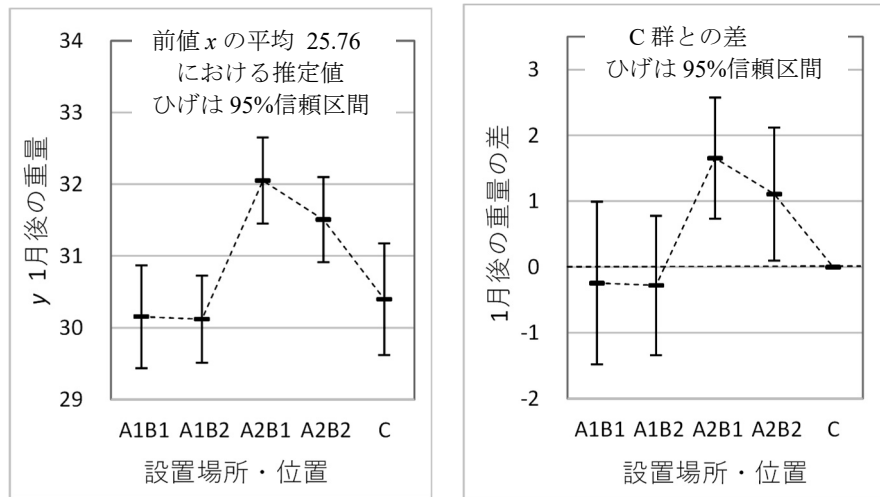


図 1.17 設置場所および位置の組み合わせによる推定平均（最小 2 乗平均）

この線グラフから A₁: 取水路 では, C: 湾内 に比べ有意な差ではなく, A₂: 排水路 では, 位置 (B₁: 上部, B₂: 下部) 共に有意な差があることが読み取れる. 各水路内の位置 (B₁: 上部, B₂: 下部) の違いは, 95%信頼区間がそれぞれの推定値を含んでいるので, 統計的な差はないと結論付けられる.

1 元配置を 2 元配置として再構築

線グラフから有意な差がないといっても報告書には, きちっとした解析結果が求められる. 設置場所 A₁ における位置の差 (B₁ - B₂) の推定値および 95%信頼区間を求めたい. そのためには, 単純に A₁B₁ の推定値 $\hat{y}_{A_1B_1}$ から A₁B₂ の推定値 $\hat{y}_{A_1B_2}$ を引けばいいのだが, 差の分散を求めるためには, パラメータに関する線形式,

$$\begin{array}{rcl}
 A_1B_1 : & \hat{y}_{A_1B_1} = \hat{\theta}_0 + \hat{\theta}_1 & + \hat{\theta}_5 x_i \\
 -) A_1B_2 : & \hat{y}_{A_1B_2} = \hat{\theta}_0 & + \hat{\theta}_2 + \hat{\theta}_5 x_i \\
 \hline
 A_1B_1 - A_1B_2 : & \hat{y}_{A_1B_1} - \hat{y}_{A_1B_2} = \hat{\theta}_1 - \hat{\theta}_2 &
 \end{array} \tag{1.49}$$

が必要となる. この分散は, パラメータが 2 つなので, 行列計算の代わりに, 合成分散の一般式でも推定できる.

$$\left. \begin{aligned}
 \text{Var}(\hat{\theta}_1 - \hat{\theta}_2) &= \text{Var}(\hat{\theta}_1) - 2\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) + \text{Var}(\hat{\theta}_2) \\
 &= 0.3324 - 2 \times 0.2048 + 0.2430 \\
 &= 0.1658
 \end{aligned} \right\} \tag{1.50}$$

設置場所 A₁ について位置を (B₁, B₂) をプールした場合の平均値を推定し, 95%信頼区間を求めたい. そのためには, A₁B₁ の推定値 $\hat{y}_{A_1B_1}$ と A₁B₂ の推定値 $\hat{y}_{A_1B_2}$ を足して 2 で割ればよい.

表 1.34 A 内 B の差, A および B の回帰直線上の推定値と 95%信頼区間

	L	A B C	l					推定値 L^\wedge	分散 $Var(L^\wedge)$	標準誤差 SE	95%信頼区間		
			l_0 x_0	l_1 g_1	l_2 g_2	l_3 g_3	l_4 g_4				l_5 x	L 95%	U 95%
A ₁ 内B ₁	$L^{(11)}$	A ₁ :(B ₁ -B ₂)	0	1	-1	0	0	0.00	0.036	0.1658	0.8734	-0.838	0.909
B ₂ との差	$L^{(12)}$	A ₁ :(B ₂ -B ₂)	0	0	0	0	0	0.00	0.000	0.0000	0.0000	0.000	0.000
A ₂ 内B ₁	$L^{(13)}$	A ₂ :(B ₁ -B ₂)	0	0	0	1	-1	0.00	0.548	0.1579	0.8524	-0.305	1.400
B ₂ との差	$L^{(14)}$	A ₂ :(B ₂ -B ₂)	0	0	0	0	0	0.00	0.000	0.0000	0.0000	0.000	0.000
A ₁ 全体	$L^{(15)}$	A ₁	1	0.5	0.5	0	0	25.76	30.135	0.0543	0.4996	29.636	30.635
A ₂ 全体	$L^{(16)}$	A ₂	1	0	0	0.5	0.5	25.76	31.779	0.0378	0.4171	31.361	32.196
B ₁ 全体	$L^{(17)}$	B ₁	1	0.5	0	0.5	0	25.76	31.103	0.0424	0.4416	30.661	31.544
B ₂ 全体	$L^{(18)}$	B ₂	1	0	0.5	0	0.5	25.76	30.811	0.0402	0.4298	30.381	31.241
C	$L^{(19)}$	C	1	0	0	0	0	25.76	30.398	0.1312	0.7768	29.621	31.174
									$t_{0.05}(20-6)=$	2.1448			

$$\begin{aligned}
 A_1B_1 : \quad & \hat{y}_{A_1B_1} = \hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_5 x_i \\
 +) \quad A_1B_2 : \quad & \hat{y}_{A_1B_2} = \hat{\theta}_0 + \hat{\theta}_2 + \hat{\theta}_5 x_i \\
 \hline
 \frac{A_1B_1 + A_1B_2}{2} : \quad & \frac{\hat{y}_{A_1B_1} + \hat{y}_{A_1B_2}}{2} = \hat{\theta}_0 + \frac{\hat{\theta}_1 + \hat{\theta}_2}{2} + \hat{\theta}_5 x_i
 \end{aligned} \tag{1.51}$$

これを、推定に用いる線形和のベクトル l として

$$l^{(15)} = [1 \quad 0.5 \quad 0.5 \quad 0 \quad 0 \quad 25.76]$$

を設定し、これまでと同様に推定値を $L^{(15)} = l^{(15)}\hat{\theta}$ で計算し、分散は

$$Var(L^{(15)}) = l^{(15)}\Sigma(\hat{\theta})(l^{(15)})^T \tag{1.52}$$

で計算する。

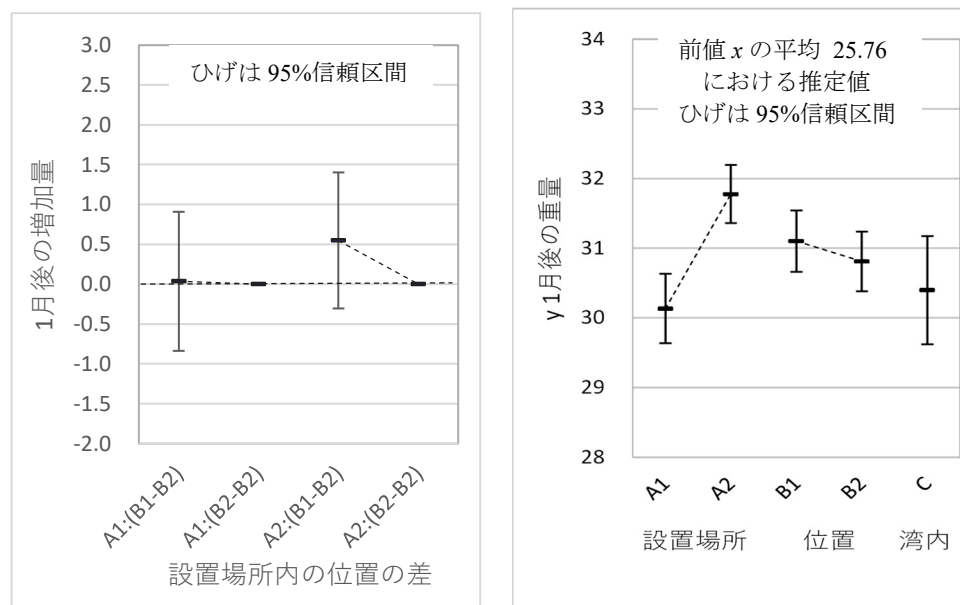


図 1.18 A 内の B 間の差, および, A の水準平均と B の水準平均

これらを総合すると、A₂の放水路においてカキの成長が他に比べて早いことが明らかとなり、その位置 (B₁上部, B₂下部) 間には有意な差が認められないと結論される。

SAS による検証

無償で継続的に提供されている OnDemand SAS の一般線形モデル GLM プロシジャにより、Excel での結果を検証する。GLM プロシジャには、最小 2 乗平均を算出するための lsmeans (least squares means) ステートメントがあり、表 1.33 に示したと同じ結果が得られる。また、estimate ステートメントで、表 1.34 に示したと同様に、各種の推定値を求めることができる。ただし、群 G の最後の水準に対してダミー変数として“0”を追加する必要がある。このことが、交互作用がある場合にダミー変数のインフレーションを引き起こし、estimate ステートメントの設定が煩雑となる原因である。

```
SAS プログラム 1.1 /* SAS データファイルの作成, GLM プロシジャ */
TITLE1 ' Oysters.SAS 2020/08/23 Y.Takahashi ' ;
data d01 ;
  input G$ @ ;
  do rep=1 to 4 ;
    input x y @ ;
    output ;
  end ;
datalines ;
G1_A1B1 27.2 32.6 32.0 36.6 33.0 37.7 26.8 31.0
G2_A1B2 28.6 33.8 26.8 31.7 26.5 30.7 26.8 30.4
G3_A2B1 28.6 35.2 22.4 29.1 23.2 28.9 24.4 30.2
G4_A2B2 29.3 35.0 21.8 27.0 30.3 36.4 24.3 30.5
G5_C 20.4 24.6 19.6 23.4 25.1 30.3 18.1 21.8
;
proc glm data=d01 ;
  class G ;
  model Y = G x / solution ;
  lsmeans G / stderr CL ;
  estimate 'A1:(B1-B2)' G 1 -1 0 0 0 ;
  estimate 'A2:(B1-B2)' G 0 0 1 -1 0 ;
  estimate 'A1' intercept 1 G 0.5 0.5 0 0 0 x 25.76;
  estimate 'A2' intercept 1 G 0 0 0.5 0.5 0 x 25.76;
  estimate 'B1' intercept 1 G 0.5 0 0.5 0 0 x 25.76;
  estimate 'B2' intercept 1 G 0 0.5 0 0.5 0 x 25.76;
run ;
```

SAS の結果:HTML 出力を Excel に取り込み、有効数字などを揃えている。

Model ステートメントによる分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	5	354.4472	70.8894	235.05	<.0001
Error	14	4.2223	0.3016		
Corrected Total	19	358.6695			

Excel の表 1.31 右上段の結果に一致する。

群 G および前値 x に対する F 検定

要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
G	4	12.0894	3.0223	10.02	0.0005
x	1	156.0402	156.0402	517.38	<.0001

Excel による各因子の Type III の平方和は、回帰の平方和の差分から別途計算する必要がある

Model ステートメントによる回帰パラメータの推定 (solution オプション)

パラメータ	推定値		標準誤差	t 値	Pr > t		係数	標準誤差
Intercept	2.4949	B	1.0279	2.43	0.0293	切片	2.4949	1.0279
G G1_A1B1	-0.2445	B	0.5766	-0.42	0.6780	g ₁	-0.2445	0.5766
G G2_A1B2	-0.2803	B	0.4929	-0.57	0.5786	g ₂	-0.2803	0.4929
G G3_A2B1	1.6548	B	0.4294	3.85	0.0018	g ₃	1.6548	0.4294
G G4_A2B2	1.1071	B	0.4718	2.35	0.0342	g ₄	1.1071	0.4718
G G5_C	0	B	.	.	.	x	1.0832	0.0476
x	1.0832		0.0476	22.75	<.0001			

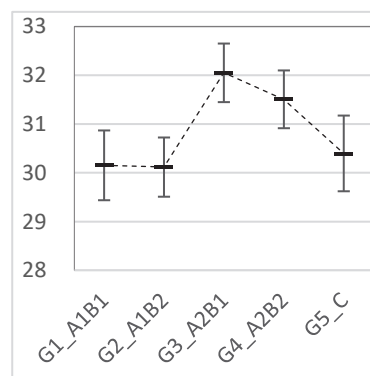
Note: X'X は特異行列です。w 正規方程式には一般化逆行列が使用されています。
w 文字 'B' が付けられた推定値は一意的な推定値ではありません。

Excel の表 1.31 右中段の結果に一致する。

lsmeans ステートメントによる最小 2 乗平均 (表 1.33 上段に対応)

G	y の最小 2 乗平均	標準誤差	Pr > t	95% 信頼限界		信頼幅
G1_A1B1	30.1531	0.3339	<.0001	29.44	30.87	0.72
G2_A1B2	30.1173	0.2827	<.0001	29.51	30.72	0.61
G3_A2B1	32.0523	0.2796	<.0001	31.45	32.65	0.60
G4_A2B2	31.5047	0.2764	<.0001	30.91	32.10	0.59
G5_C	30.3976	0.3622	<.0001	29.62	31.17	0.78

表 1.35 の上段に一致
信頼幅は、Excel で計算
Excel の線グラフによる作図。図 1.17 左と同じ⇒



estimate ステートメントによる推定値 (表 1.34 に対応)

パラメータ	推定値	標準誤差	t 値	Pr > t	95% 信頼区間	
					L95%	U95%
A1:(B1-B2)	0.0358	0.4072	0.09	0.9312	-0.84	0.91
A2:(B1-B2)	0.5476	0.3974	1.38	0.1898	-0.30	1.40
A1	30.1352	0.2329	129.37	<.0001	29.64	30.63
A2	31.7785	0.1945	163.43	<.0001	31.36	32.20
B1	31.1027	0.2059	151.05	<.0001	30.66	31.54
B2	30.8110	0.2004	153.76	<.0001	30.38	31.24
$t_{0.05}(20-6) =$					2.1448	

estimate ステートメントでの各種の推定は、SAS/GLM プロシジャでの内部で生成される「デザイン行列」について理解した上で設定する必要がある。

1.5. 除草剤の散布による雑草の成長抑制

奥野ら (1978), 「応用統計ハンドブック, 3.7.3 節 乱塊法における共分散分析」に除草剤の効果と比較するための実験データが示されている. 表 1.35 に示すように, 無散布群と 4 種類の除草剤に対し, 2 または 3 濃度を設定した実験であり, 延べ 11 群の処理について 3 ブロックで除草剤の散布実験を行ない, 所定期日の雑草の重量 y を測定した. 雑草重量 y は, 初期の雑草の生え方 x の影響を受けるので, 被覆度を目測で 10~50%の間で採点して前値 x とした.

表 1.35 除草剤の散布試験データ [奥野ら (1978), 表 7.4]

除草剤	散布量 g/3m ²	群	ブロック I		ブロック II		ブロック III	
			x	y	x	y	x	y
無散布		1	10	41	30	157	30	168
M.C.P.	100	2	15	56	10	28	15	68
	150	3	15	31	35	129	30	92
C.M.u.	25	4	10	17	20	54	40	123
	50	5	10	12	10	13	50	141
	75	6	10	10	15	20	40	61
P.C.P.	300	7	20	49	15	45	40	115
	600	8	20	49	15	33	20	25
	900	9	10	9	20	43	30	60
24-D	100	10	15	27	20	90	40	123
	150	11	10	17	15	51	20	98
平均			13.2		18.6		32.3	

除草剤の種類は 4 剤で, それぞれ 2~3 用量となっており, 対照群として「無散布」が設定されている. ブロック別の被覆度 x の平均は, (13.2%, 18.6%, 32.3%) となっており, 雑草の生育状況でブロック化されて, 各ブロック内で全 11 群が割り付けられている. ただし, 仔細にみると, ブロック内での前値 x の大きさの逆転が若干あるので, ブロックを無視して除草剤別に前値 x の小さい順位に並べ変えた結果を表 1.36 に示す.

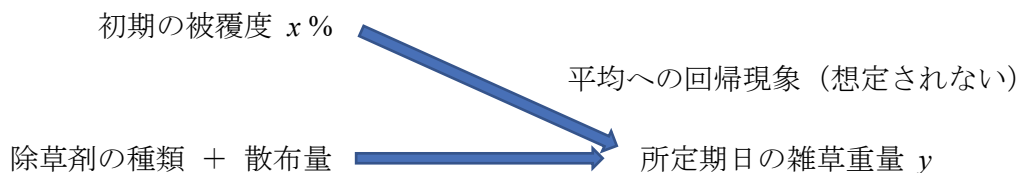


図 1.19 除草剤の雑草の成長抑制に与える影響

表 1.36 右には, 除草剤の種類別に散布量を (+ : 低, Δ : 中, \square : 高) と区分けした散布図に, 線形近似による回帰直線を重ね書きしている. ほとんどの回帰直線は, 原点方向に向

かっており、平行でないことは明らかである。また、第1.3節で取り上げた交互作用を含む共分散分析ではなく、切片を共通とする回帰直線のあてはめ、あるいは、原点を共通の「切片」とする回帰直線のあてはめが妥当である。

所定期日の雑草重量 y (後値とする) は、初期の雑草の被覆度 $x\%$ (前値とする) の影響が大きいので、前値 x の被覆度 (15%, 30%) とした場合に推定される回帰直線の推定値とその95%信頼区間により、各除草剤の用量反応関係の形状を概観することにする。

表 1.36 除草剤別の前値 x に対する後値 y の関連

G	処置	ブロック	x	y
G ₀	無散布	I	10	41
		II	30	157
		III	30	168
A ₂	MCP 100	II	10	28
		I	15	56
		III	15	68
A ₃	MCP 150	I	15	31
		III	30	92
		II	35	129
B ₂	CMu 25	I	10	17
		II	20	54
		III	40	123
B ₃	CMu 50	I	10	12
		II	10	13
		III	50	141
B ₄	CMu 75	I	10	10
		II	15	20
		III	40	61
C ₂	PCP 300	II	15	45
		I	20	49
		III	40	115
C ₃	PCP 600	II	15	33
		I	20	49
		III	20	25
C ₄	PCP 900	I	10	9
		II	20	43
		III	30	60
D ₂	24D 100	I	15	27
		II	20	90
		III	40	123
D ₃	24D 150	I	10	17
		II	15	51
		III	20	98

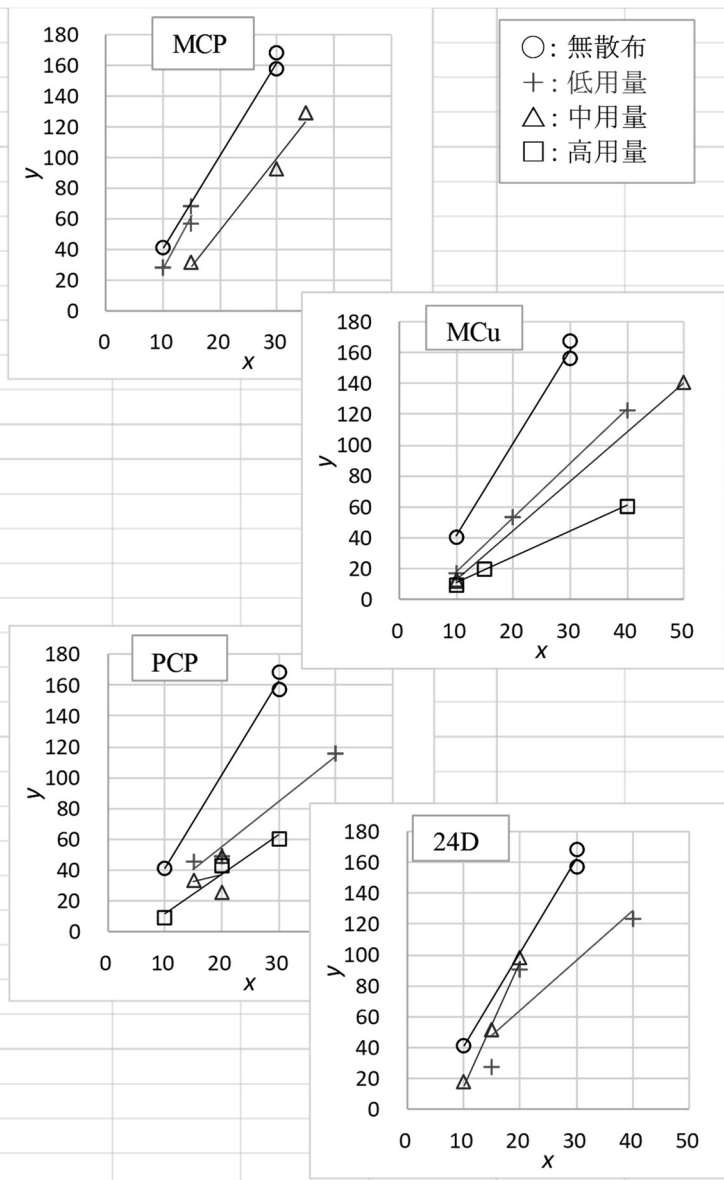


図 1.20 除草剤別の反応関係

MCP 剤に対する共通の切片を持つ回帰直線

無処置と MCP 剤の 2 用量を質的変数, 前値を共変量とし, 切片 β_0 を共通とする 3 本の回帰式を

$$\left. \begin{aligned} A_1 \text{ 無散布: } & y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \\ A_2 \text{ MCP100: } & y_i = \beta_0 + \beta_2 x_i + \varepsilon_i \\ A_3 \text{ MCP150: } & y_i = \beta_0 + \beta_3 x_i + \varepsilon_i \end{aligned} \right\} \quad (1.53)$$

とする. これらの式を表 1.3 で例示した (1, 1) 標示型ダミー変数を用いて統合すると,

$$y_i = \beta_0 + \beta_1 a_1 x_i + \beta_2 a_2 x_i + \beta_3 a_3 x_i + \varepsilon_i \quad (1.54)$$

のように一つの式にまとめることができる. 表 1.37 に示すように, (1, 1) 標示型ダミー変数を (a_1, a_2, a_3) とし, 解析に用いるデザイン行列 X は, ダミー変数 (a_1, a_2, a_3) と前値 x との積とする.

表 1.37 に示すのは, Excel の回帰分析を用いて共通の切片を持つ回帰直線をあてはめた結果である. 共通の切片 $\hat{\beta}_0 = -29.7171$ は, t 検定の結果 $p = 0.0213$ であり, 原点よりも統計的に下回っている. 切片がマイナスであることは現実には解釈できない推定値であり, 共通の切片の推定値に対し, 意味付けすることには無理がある. 推定された回帰パラメータ (係数) から, 図 1.21 に示すように初期の被覆度 $x = 15\%$ における各群の推定値は,

$$\begin{aligned} A_1 : \hat{y}_{x=15}^{(A_1)} &= -29.7271 + 6.4426 \times 15 = 66.9116 \\ A_2 : \hat{y}_{x=15}^{(A_2)} &= -29.7271 + 6.0529 \times 15 = 61.0661 \\ A_3 : \hat{y}_{x=15}^{(A_3)} &= -29.7271 + 4.3056 \times 15 = 34.8570 \end{aligned}$$

表 1.37 共通の切片を持つ回帰直線のあてはめ

A	処置					デザイン行列 X					分散分析表(共通切片)「定数に 0 を使用」off				
		a_1	a_2	a_3	x	x_0	$a_1 x$	$a_2 x$	$a_3 x$	y	自由度	変動	分散	分散比	
A ₁	無散布 (G ₀)	1	0	0	10	1	10	0	0	41	回帰	3	22967.47	7655.82	120.09
		1	0	0	30	1	30	0	0	157	残差	5	318.75	63.75	= σ^2
		1	0	0	30	1	30	0	0	168	合計	8	23286.22		
A ₂	MCP 100	0	1	0	10	1	0	10	0	28			$t_{0.05(9-4)} =$	2.5706	
		0	1	0	15	1	0	15	0	56	係数	標準誤差	t	P -値	
		0	1	0	15	1	0	15	0	68	β^{\wedge}_0 切片 x_0	-29.7271	8.9913	-3.3062	0.0213
A ₃	MCP 150	0	0	1	15	1	0	0	15	31	β^{\wedge}_1 $a_1 x$	6.4426	0.3785	17.0200	0.0000
		0	0	1	30	1	0	0	30	92	β^{\wedge}_2 $a_2 x$	6.0529	0.7372	8.2103	0.0004
		0	0	1	35	1	0	0	35	129	β^{\wedge}_3 $a_3 x$	4.3056	0.3476	12.3871	0.0001
											パラメータの共分散行列 $\Sigma(\beta) = (X^T X)^{-1} \sigma^2$				
											β^{\wedge}_0	80.8432	-2.9784	-5.8795	-2.7521
											β^{\wedge}_1	-2.9784	0.1433	0.2166	0.1014
											β^{\wedge}_2	-5.8795	0.2166	0.5435	0.2002
											β^{\wedge}_3	-2.7521	0.1014	0.2002	0.1208
											x_0	$a_1 x$	$a_2 x$	$a_3 x$	

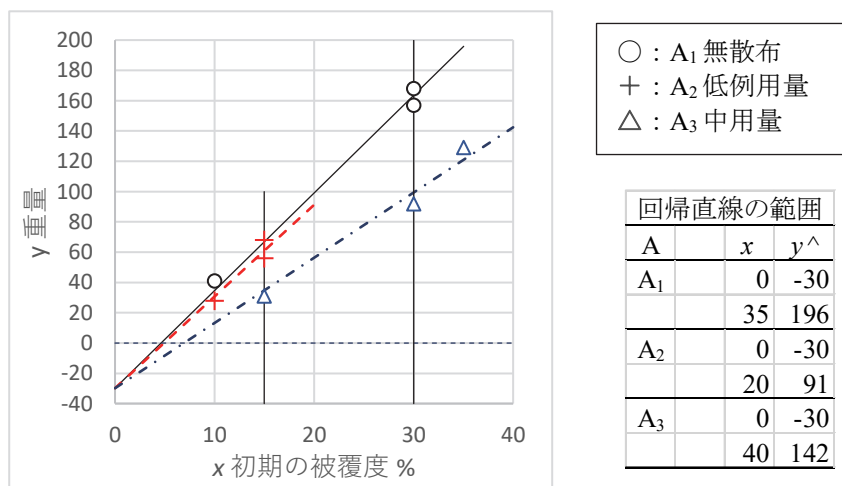


図 1.21 MCP 剤に対する切片を共通とする回帰直線のあてはめ

と推定される. それぞれの 95%信頼区間を求めるために, パラメータの共分散行列 $\Sigma(\hat{\beta})$ が, 表 1.37 右下段に計算されている. 無散布 A₁ に対する推定値の分散 $Var(\hat{y}_{x=15}^{(A_1)})$ は, 合成分散の一般式により

$$\begin{aligned} Var(\hat{y}_{x=15}^{(A_1)}) &= Var(\hat{\beta}_0 + \hat{\beta}_1 \times 15) \\ &= Var(\hat{\beta}_0) + 2 \times Cov(\hat{\beta}_0, \hat{\beta}_1) \times 15 + Var(\hat{\beta}_1) \times 15^2 \\ &= 80.8432 + 2 \times (-2.9784) \times 15 + 0.1433 \times 15^2 = 23.7291 \end{aligned}$$

として求めることができる.

表 1.38 に示すのは, $x=15$ の場合も含め, 前値 $x=30$ における各種の推定値, 無散布 A₁ からの減少量に対する推定値, 分散, 95%信頼区間を計算した結果を示す. 推定値および分散の計算には, これまでと同様に Excel の行列関数を用いた 2 次形式で求めている.

表 1.38 共通の切片を持つ回帰直線における各種の推定

x	線形和		用量	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
	A	L		x_0	a_1x	a_2x	a_3x				$L^$	$Var(L^)$
15	A ₁	$L^{(1)}$	0	1	15	0	0	66.91	23.73	12.52	54.39	79.43
	A ₂	$L^{(2)}$	100	1	0	15	0	61.07	26.75	13.29	47.77	74.36
	A ₃	$L^{(3)}$	200	1	0	0	15	34.86	25.46	12.97	21.89	47.83
30	A ₁	$L^{(4)}$	0	1	30	0	0	163.55	31.09	14.33	149.22	177.88
	A ₂	$L^{(5)}$	100	1	0	30	0	151.86	217.23	37.89	113.97	189.75
	A ₃	$L^{(6)}$	200	1	0	0	30	99.44	24.45	12.71	86.73	112.15
15	A ₁ -A ₁	$L^{(7)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	A ₂ -A ₁	$L^{(8)}$	100	0	-15	15	0	-5.85	57.05	19.42	-25.26	13.57
	A ₃ -A ₁	$L^{(9)}$	200	0	-15	0	15	-32.05	13.80	9.55	-41.60	-22.51
30	A ₁ -A ₁	$L^{(10)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	A ₂ -A ₁	$L^{(11)}$	100	0	-30	30	0	-11.69	228.21	38.83	-50.52	27.14
	A ₃ -A ₁	$L^{(12)}$	200	0	-30	0	30	-64.11	55.18	19.10	-83.20	-45.01
								T.inv.2T(0.05,9-4)=		2.5706		

図 1.22 に表 1.38 で推定された切片を共通とする回帰直線の前値 $x=15$ および $x=30$ における回帰直線の推定値と 95%信頼区間を Excel の折れ線グラフで作図した結果を示す. 無散布 A_1 群に対し, A_2 MCP 100 g/3m² 群は, 有意な差ではなく, A_2 : MCP 200 g/3m² 群では, 無散布群に対し 40%~50%程度の除草効果があると推定される.

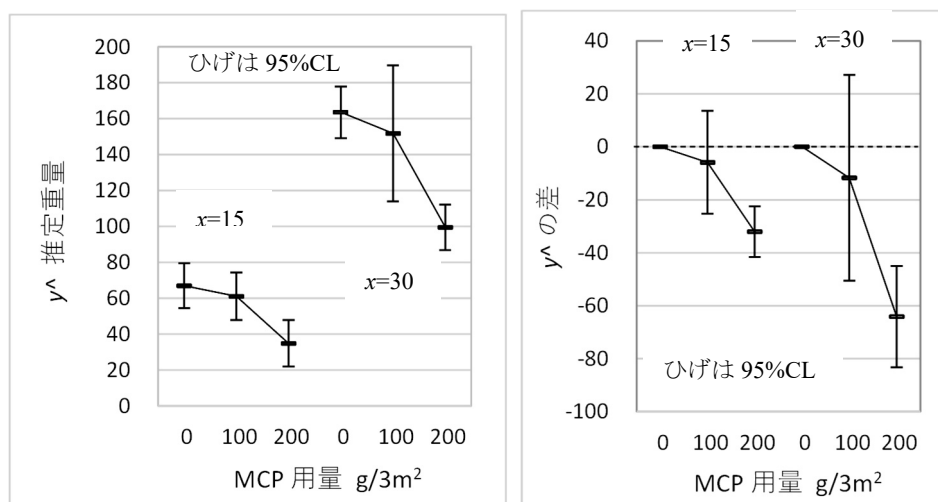


図 1.22 共通の切片を持つ回帰直線における各種の推定値と 95%信頼区間

PCP 剤に対する原点を共通とする回帰直線

PCP 剤については, 原点を共通とする 4 本の回帰直線をあてはめる. 原点を通る回帰式の場合は, 切片がゼロなので, パラメータとして設定せず, それぞれ別々の傾きをもつ 4 本の回帰直線の同時あてはめとなる.

$$\left. \begin{array}{l} C_1 \text{ 無処置: } y_i = \beta_1 x_i + \varepsilon_i \\ C_2 \text{ PCP 300: } y_i = \beta_2 x_i + \varepsilon_i \\ C_3 \text{ PCP 600: } y_i = \beta_3 x_i + \varepsilon_i \\ C_4 \text{ PCP 900: } y_i = \beta_4 x_i + \varepsilon_i \end{array} \right\} \quad (1.55)$$

(1, 1) 標示型ダミー変数 (c_1, c_2, c_3, c_4) を用いると式 (1.55) は,

$$y_i = \beta_1 c_1 x_i + \beta_2 c_2 x_i + \beta_3 c_3 x_i + \beta_4 c_4 x_i + \varepsilon_i \quad (1.56)$$

となり, 一つの式でまとめて表すことができる. 表 1.39 に示すように, 共通の切片がない傾きだけの ($c_1 x_i, c_2 x_i, c_3 x_i, c_4 x_i$) をデザイン行列 \mathbf{X} としている. Excel の回帰分析でも切片がないモデルの設定は, 「定数に 0 を使用」設定を オンとすることに可能となる.

原点を通る回帰直線のデザイン行列 \mathbf{X} は, 切片がないので積和行列 ($\mathbf{X}^T \mathbf{X}$) は, 対角行列となり, その逆行列もまた対角行列となるので, 誤差分散の推定値 $\hat{\sigma}^2 = 91.15$ を掛けたパラメータの共分散行列も表 1.39 右下に示すように対角行列となり, パラメータ間の共分散は, 全て 0 である.

表 1.39 原点を共通とする回帰直線のあてはめ

C	処置	x	デザイン行列 X								y	分散分析表「定数に0を使用」を on					
			c ₁	c ₂	c ₃	c ₄	c ₁ x	c ₂ x	c ₃ x	c ₄ x		自由度	変動	分散	分散比		
C ₁	無散布 (G ₀)	10	1	0	0	0	10	0	0	0	41	回帰	4	81120.8	20280.2	222.50	
		30	1	0	0	0	30	0	0	0	157	残差	8	729.2	91.15	= σ ²	
		30	1	0	0	0	30	0	0	0	168	合計	12	81850.0			
C ₂	PCP 300	15	0	1	0	0	0	15	0	0	45	<i>t</i> _{0.05(12-4)} = 2.3060					
		20	0	1	0	0	0	20	0	0	49	係数	標準誤差	<i>t</i>	<i>P</i> -値		
		40	0	1	0	0	0	40	0	0	115	切片	#N/A	#N/A	#N/A		
C ₃	PCP 600	15	0	0	1	0	0	0	15	0	33	β [^] ₁	c ₁ x	5.3474	0.2190	24.4143	0.0000
		20	0	0	1	0	0	0	20	0	49	β [^] ₂	c ₂ x	2.8112	0.2024	13.8896	0.0000
		20	0	0	1	0	0	0	20	0	25	β [^] ₃	c ₃ x	1.9268	0.2982	6.4615	0.0002
C ₄	PCP 900	10	0	0	0	1	0	0	0	10	9	β [^] ₄	c ₄ x	1.9643	0.2552	7.6983	0.0001
		20	0	0	0	1	0	0	0	20	43						
		30	0	0	0	1	0	0	0	30	60						
パラメータの共分散行列 Σ(β [^]) = (X ^T X) ⁻¹ σ ²											β [^] ₁	0.0480	0.0000	0.0000	0.0000		
											β [^] ₂	0.0000	0.0410	0.0000	0.0000		
											β [^] ₃	0.0000	0.0000	0.0889	0.0000		
											β [^] ₄	0.0000	0.0000	0.0000	0.0651		
											c ₁ x						
											c ₂ x						
											c ₃ x						
											c ₄ x						

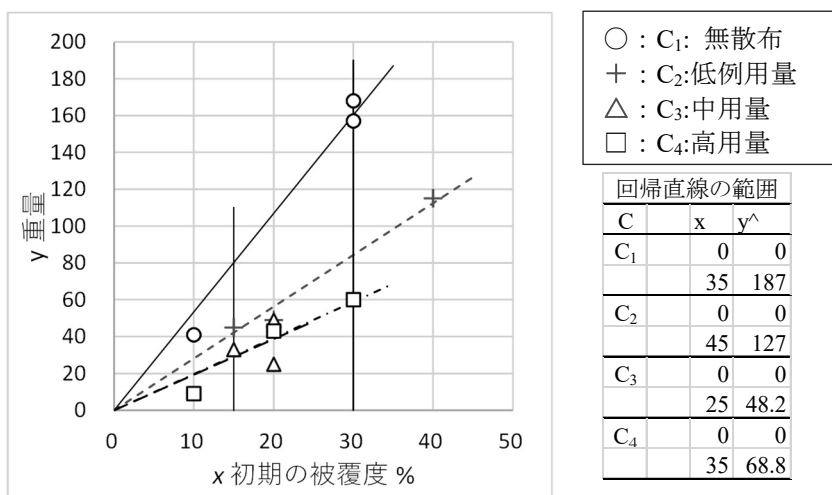


図 1.23 原点を共通とする回帰直線のあてはめ

表 1.40 に示すように、前値が $x=15$ および $x=30$ 単位の場合の各群の推定値とその 95% 信頼区間、無散布群との差についての推定値とその 95% 信頼区間の計算結果を示す。これまでと同様に行列計算を使っているが、共分散が全て 0 なので、 $\mathbf{l}^{(1)} = [15 \ 0 \ 0 \ 0]$ の推定値と分散は、

$$\begin{aligned}
 \hat{y}_{x=15}^{(C_1)} &= \mathbf{l}^{(1)} \hat{\boldsymbol{\beta}} \\
 &= 15 \hat{\beta}_1 \\
 &= 15 \times 5.3474 = 80.21
 \end{aligned} \tag{1.57}$$

$$\begin{aligned}
 \text{Var}(\hat{y}_{x=15}^{(C_1)}) &= \text{Var}(15 \hat{\beta}_1) \\
 &= 15^2 \times 0.0480 = 10.79
 \end{aligned} \tag{1.58}$$

として計算できる。

無散布からの差の推定値のためのベクトル $L^{(10)} = [-15 \ 15 \ 0 \ 0]$ は、すでに計算されている

$$L^{(1)} = \hat{y}_{x=15}^{(C_1)} = 80.21, \quad L^{(2)} = \hat{y}_{x=15}^{(C_2)} = 42.17 \quad (1.59)$$

から

$$L^{(2)} - L^{(1)} = \hat{y}_{x=15}^{(C_2-C_1)} = 42.17 - 80.21 = -38.04 \quad (1.60)$$

として計算され、その分散は、分散の加法性により

$$\begin{aligned} \text{Var}(\hat{y}_{x=15}^{(C_2-C_1)}) &= \text{Var}(\hat{y}_{x=15}^{(C_2)}) + \text{Var}(\hat{y}_{x=15}^{(C_1)}) \\ &= 9.22 + 10.79 = 20.01 \end{aligned}$$

によっても計算できる。図 1.24 に表 1.40 で推定された結果を用いて、Excel の線グラフで作図した結果を示す。

表 1.40 原点を共通とする回帰直線における各種の推定

x	C	L	用量	l_i				推定値	分散	幅	95%信頼区間	
				c_{1x}	c_{2x}	c_{3x}	c_{4x}				y^{\wedge}	$\text{Var}(y^{\wedge})$
15	C ₁	$L^{(1)}$	0	15	0	0	0	80.21	10.79	7.58	72.63	87.79
	C ₂	$L^{(2)}$	300	0	15	0	0	42.17	9.22	7.00	35.17	49.17
	C ₃	$L^{(3)}$	600	0	0	15	0	28.90	20.01	10.31	18.59	39.22
	C ₄	$L^{(4)}$	900	0	0	0	15	29.46	14.65	8.83	20.64	38.29
30	C ₁	$L^{(5)}$	0	30	0	0	0	160.42	43.18	15.15	145.27	175.57
	C ₂	$L^{(6)}$	300	0	30	0	0	84.34	36.87	14.00	70.34	98.34
	C ₃	$L^{(7)}$	600	0	0	30	0	57.80	80.03	20.63	37.18	78.43
	C ₄	$L^{(8)}$	900	0	0	0	30	58.93	58.60	17.65	41.28	76.58
15	C ₁ -C ₁	$L^{(9)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	C ₂ -C ₁	$L^{(10)}$	300	-15	15	0	0	-38.04	20.01	10.32	-48.36	-27.73
	C ₃ -C ₁	$L^{(11)}$	600	-15	0	15	0	-51.31	30.80	12.80	-64.11	-38.51
	C ₄ -C ₁	$L^{(12)}$	900	-15	0	0	15	-50.75	25.44	11.63	-62.38	-39.11
30	C ₁ -C ₁	$L^{(13)}$	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
	C ₂ -C ₁	$L^{(14)}$	300	-30	30	0	0	-76.08	80.04	20.63	-96.72	-55.45
	C ₃ -C ₁	$L^{(15)}$	600	-30	0	30	0	-102.62	123.21	25.60	-128.21	-77.02
	C ₄ -C ₁	$L^{(16)}$	900	-30	0	0	30	-101.49	101.77	23.26	-124.76	-78.23
T.inv.2T (0.05,12-4)=									2.3060			

Excel の回帰分析で、原点を通る（「定数に 0 を使用」をオン）回帰式をあてはめた場合の分散分析表の結果が怪しいとの指摘が、岩崎（2006）、「単回帰分析、第 4-3 節 原点を通る回帰式（Excel）」にあるので、同じデータを用いて Excel 2016 で確認したところ、SPSS および JMP の出力に一致しているので、このバグは解消されている。また、表 1.39 の Excel の回帰分析での分散分析表は、後述の R での検証で、一致することが確認されている。この指摘に限らず、Excel での計算自体に操作ミス、計算式の設定ミスなどがあっても気が付きにくいなどの脆弱性があることに注意する必要がある。

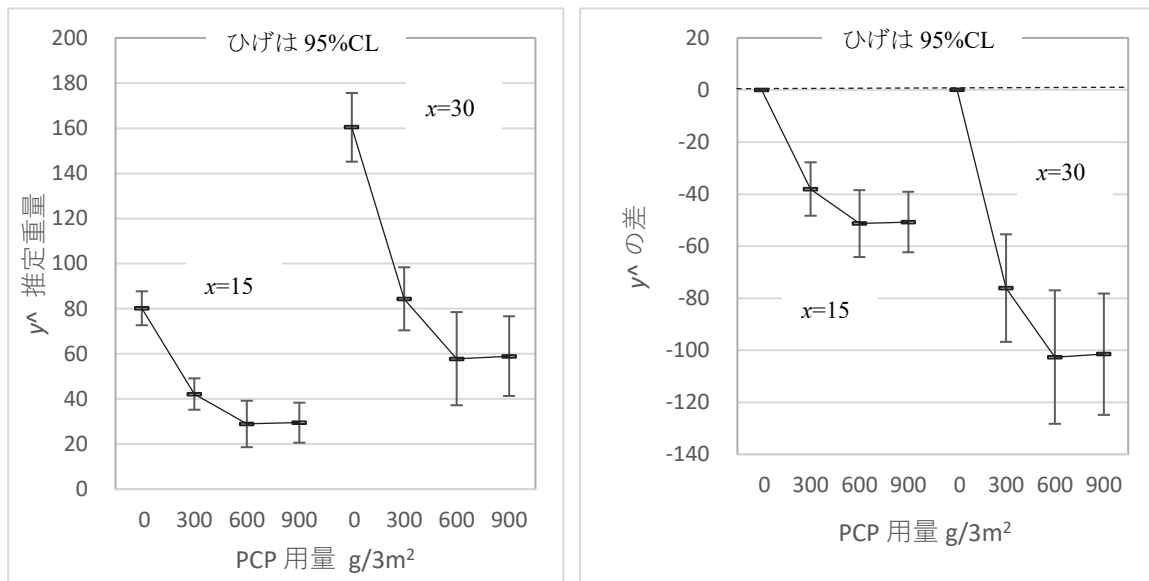


図 1.24 原点を共通とする回帰直線における各種の推定と 95%信頼区間

原点を通る回帰直線の推定は、Excel で「定数に 0 を使用」をオンとする対応であり、SAS の GLM プロシジャで Model ステートメントの「nointercept」オプションを使い、JMP で「切片なし」をオンとし、R の lm() 関数で lm の回帰式の末尾に「+0」を加えることにより実現できる。

R による検証(原点共通の場合)

R の lm() 関数で原点を通る回帰直線は、モデル式の最後に「+0」を加えることにより切片無のモデルになる。実際に表 1.39 で示したデザイン行列を用いて計算した結果を次に示す。表 1.39 の Excel でのパラメータの推定値は、完全に一致し、「Residual standard error: 9.547」の平方は、 $9.547^2 = 91.15$ と残差の平均平方(分散)に一致する。

```
Call:
lm(formula = y ~ c1x + c2x + c3x + c4x + 0, data = Dataset)

Residuals:
    Min       1Q   Median       3Q      Max
-13.537  -8.079   1.811   3.810  10.463

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
c1x   5.3474    0.2190  24.414 0.0000000846 ***
c2x   2.8112    0.2024  13.890 0.00000069849 ***
c3x   1.9268    0.2982   6.461  0.000196 ***
c4x   1.9643    0.2552   7.698 0.00005752696 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.547 on 8 degrees of freedom
Multiple R-squared:  0.9911,    Adjusted R-squared:  0.9866
F-statistic: 222.5 on 4 and 8 DF,  p-value: 0.00000003127
```

2. デザイン行列を活用した 1 因子実験データの解析

平方和の分解、および、分散の加法性に基づく各種の実験データの解析法は、実験計画法を発展させてきた。伝統的な実験計画法は、多くの要因を含む多彩な実験データに対して手計算でも対処できるように、長年の創意工夫の集大成とも言うべき優れた方法である。ただし、量的変数に対しては、幾つかの水準を設定し質的変数として扱うことを前提にしなければならない問題を内在している。先進的な統計ソフトでは、質的変数を量的変数（ダミー変数）に置き換え、量的変数に対する線形モデルによる解析が行われている。だが、その効果的な使い方について十分に知られていない。第 1 章では、伝統的な平方和の分解によって対応できない各種の事例について身近にある Excel の回帰分析に加え、デザイン行列 X を用いた Excel の行列計算を用いることにより、きめ細かな解析ができることを示した。本章では、1 因子実験データについて、平方和の分解による解析法と対比しつつ、ダミー変数を用いた線形モデルによる解析方法を詳しく解説する。

2.1. 繰り返しが等しい 1 因子実験データ

芳賀 (2014), 「医薬品開発のための統計解析 第 2 部 実験計画法 改訂版」, の「第 1 章 質的因子の 1 因子実験」に示されている繰り返しが 5 の 1 因子実験データを表 2.1 に示す。実験は、4 種類の薬剤を取り上げ、薬効の違いを調べるため、各薬剤を 5 匹の動物に投与し、薬効を評価した結果であり、各薬剤について平均値と標準偏差が示されている。

表 2.1 データと平均・標準偏差 [芳賀 (2014), 表示 1.1.1]

水準	繰り返し					平均	標準偏差
	1	2	3	4	5		
A ₁	10.8	9.9	9.7	10.4	10.7	10.30	0.48
A ₂	10.7	10.6	11.0	10.8	10.9	10.80	0.16
A ₃	11.4	10.7	10.9	11.3	11.7	11.20	0.40
A ₄	11.9	11.2	11.0	11.1	11.3	11.30	0.35

Excelの関数, 平均:Avarage (データの範囲), 標準偏差:StDev.S (データの範囲)

データが得られたら何らかのグラフを作成し概観するのが、解析の最初の第一歩である。表 2.2 左に示すのは、表 2.1 の (4 行×5 列) の Excel シート上の矩形データを行方向に (20

行×1列) に並べ替え, 因子 A の水準番号を i として含めている. これは, Excel の散布図の X 軸の変数として使うためである. 図 2.1 右上段の散布図は, 水準番号 i とデータ y を選択して基本の「散布図」を作成し, 散布図上のデフォルトの「●印」を「データ系列の書式設定」で「○印」に置き換え, 軸の目盛りを整えた結果である. X 軸の元の目盛りは (1 2 3 4) なので, (A₁ A₂ A₃ A₄) を図形の「テキストボックス」を用いた重ね書きした結果である. 各水準の平均を, 「データの選択」により散布図上に重ね書きし, ●印を「マーカのオプション」で「種類 -」に変更し, 実線を点線に変更し書式を整えている. 図 2.1 右下段「箱ひげ図」は, 因子 A とデータ y を選択して作成し, 目盛りなどを整えた結果である.

表 2.2 行方向へのデータの並べ替え

因子	番号	反応	水準	番号	平均
A	i	y	A	i	\bar{y}_i
A ₁	1	10.8	A ₁	1	10.3
A ₁	1	9.9	A ₂	2	10.8
A ₁	1	9.7	A ₃	3	11.2
A ₁	1	10.4	A ₄	4	11.3
A ₁	1	10.7			
A ₂	2	10.7			
A ₂	2	10.6			
A ₂	2	11.0			
A ₂	2	10.8			
A ₂	2	10.9			
A ₃	3	11.4			
A ₃	3	10.7			
A ₃	3	10.9			
A ₃	3	11.3			
A ₃	3	11.7			
A ₄	4	11.9			
A ₄	4	11.2			
A ₄	4	11.0			
A ₄	4	11.1			
A ₄	4	11.3			

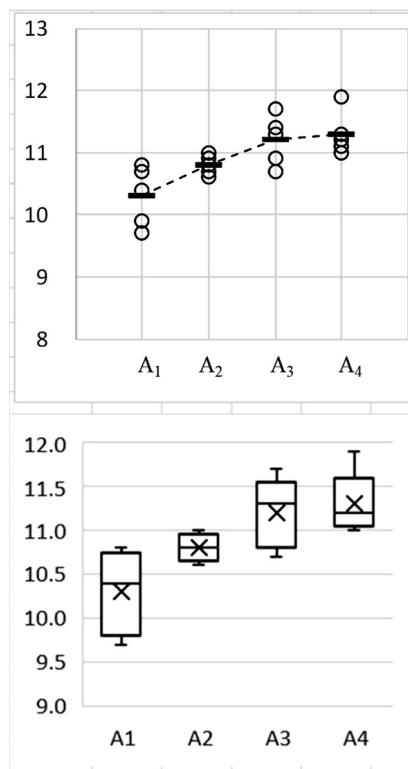


図 2.1 Excel による平均値の折れ線入り散布図および箱ひげ図

各種の平方和の分解による分散分析表

表 2.3 に示すのは, 各種の平方和の計算のための Excel シートである. デザイン行列 X を用いた回帰分析を適用する際には, 全てのデータが, 行方向に並んでいることを前提にしている. それに合わせ, 行方向のデータの並びに対して平方和の分解による解析を行う. データの構造式は,

$$\left. \begin{aligned}
 y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\
 &= \bar{y}_{..} + (\bar{y}_i - \bar{y}_{..}) + (y_{ij} - \bar{y}_i) \\
 \text{ただし, } \sum_i \alpha_i &= 0, \quad \sum_i \sum_j \varepsilon_{ij} = 0
 \end{aligned} \right\} \quad (2.1)$$

($\mu: \bar{y}_i$ の平均, α_i : 因子Aの効果, ε_{ij} : 誤差) であり, それらに対応する各種の平方和

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 \quad (2.2)$$

$$S_T = S_A + S_e$$

を計算する.

表 2.3 Excel による各種の平方和の計算

			反応	1: μ^{\wedge}		2: α^{\wedge}_i	3: $\varepsilon^{\wedge}_{ij}$	1+2+3	分散分析表			
A	i	j	y_{ij}	$\bar{y}_{..}$	$y_{ij} - \bar{y}_{..}$	$\bar{y}_{i.}$	$\bar{y}_{i.} - \bar{y}_{..}$	$y_{ij} - \bar{y}_{i.}$	y_{ij}	要因	自由度	平方和
A ₁	1	1	10.80	10.90	-0.10	10.30	-0.60	0.50	10.80	A	3	3.10
	1	2	9.90	10.90	-1.00	10.30	-0.60	-0.40	9.90	e	16	2.18
	1	3	9.70	10.90	-1.20	10.30	-0.60	-0.60	9.70	T	19	5.28
	1	4	10.40	10.90	-0.50	10.30	-0.60	0.10	10.40			
	1	5	10.70	10.90	-0.20	10.30	-0.60	0.40	10.70			
A ₂	2	1	10.70	10.90	-0.20	10.80	-0.10	-0.10	10.70			
	2	2	10.60	10.90	-0.30	10.80	-0.10	-0.20	10.60			
	2	3	11.00	10.90	0.10	10.80	-0.10	0.20	11.00			
	2	4	10.80	10.90	-0.10	10.80	-0.10	0.00	10.80			
	2	5	10.90	10.90	0.00	10.80	-0.10	0.10	10.90			
A ₃	3	1	11.40	10.90	0.50	11.20	0.30	0.20	11.40			
	3	2	10.70	10.90	-0.20	11.20	0.30	-0.50	10.70			
	3	3	10.90	10.90	0.00	11.20	0.30	-0.30	10.90			
	3	4	11.30	10.90	0.40	11.20	0.30	0.10	11.30			
	3	5	11.70	10.90	0.80	11.20	0.30	0.50	11.70			
A ₄	4	1	11.90	10.90	1.00	11.30	0.40	0.60	11.90			
	4	2	11.20	10.90	0.30	11.30	0.40	-0.10	11.20			
	4	3	11.00	10.90	0.10	11.30	0.40	-0.30	11.00			
	4	4	11.10	10.90	0.20	11.30	0.40	-0.20	11.10			
	4	5	11.30	10.90	0.40	11.30	0.40	0.00	11.30			
平均			10.90	平方和	5.28		3.10	2.18				
				S_T		S_A	S_e					

表 2.3 に示した各平方和の計算は, 以下の手順で行っている. まず, 総平均 $\bar{y}_{..}$ の計算は, Excel の Average () 関数を用いて

$$\left. \begin{aligned} \bar{y}_{..} &= \sum_i \sum_j y_{ij} / N, \quad N: \text{データ数} \\ &= \text{Average}(y_{ij} \text{の範囲}) = 10.90 \end{aligned} \right\} \quad (2.3)$$

として計算し「 μ 」の推定値としている. 総平方和 S_T は, 反応 y_{ij} と総平均 $\bar{y}_{..}$ の差 ($y_{ij} - \bar{y}_{..}$) の平方和であり, SumSq () 関数により,

$$\left. \begin{aligned} S_T &= \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 \\ &= \text{SumSq}((y_{ij} - \bar{y}_{..}) \text{の範囲}) = 5.28 \end{aligned} \right\} \quad (2.4)$$

として求めている. なお, S_T は, DevSq (y_{ij} の範囲) 関数によって偏差平方和を直接求められるが, 元のシグマを用いた式に沿った計算法としている. 因子 A の各水準の平均値は, 水準ごとに Average () 関数を適用し, 水準平均 $\bar{y}_{i.}$ を求めている.

$$\left. \begin{aligned} \bar{y}_{i.} &= \sum_j (y_{ij}) / n_i, & n_i: \text{各水準のデータ数} \\ &= \text{Average}(y_{i.} \text{の範囲}) \end{aligned} \right\} \quad (2.5)$$

因子 A の平方和 S_A は、水準平均 $\bar{y}_{i.}$ と総平均 $\bar{y}_{..}$ との差 $(\bar{y}_{i.} - \bar{y}_{..})$ を計算し、SumSq () 関数で

$$\left. \begin{aligned} S_A &= \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= \text{SumSq}((\bar{y}_{i.} - \bar{y}_{..}) \text{の範囲}) = 3.10 \end{aligned} \right\} \quad (2.6)$$

を求めている。残差平方和 S_e は、反応 y_{ij} と因子 A の水準平均 $\bar{y}_{i.}$ との差 $(y_{ij} - \bar{y}_{i.})$ を計算し、SumSq () 関数で

$$\left. \begin{aligned} S_e &= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 \\ &= \text{SumSq}((y_{ij} - \bar{y}_{i.}) \text{の範囲}) = 2.18 \end{aligned} \right\} \quad (2.7)$$

を求めている。総平方和 S_T が、

$$\left. \begin{aligned} S_T &= S_A + S_e \\ 5.28 &= 3.10 + 2.18 \end{aligned} \right\} \quad (2.8)$$

のように S_A と S_e の和に分解できるも表 2.3 右上段の分散分析表の形式で確認できる。また、式 (2.1) のデータの構造式 $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ に対し、

$$\left. \begin{aligned} y_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\varepsilon}_{ij} \\ y_{1,1} & 10.80 = 10.90 + (-0.60) + 0.50 \\ y_{1,2} & 9.90 = 10.90 + (-0.60) + (-0.40) \\ & \vdots \\ y_{4,5} & 11.30 = 10.90 + 0.40 + 0.00 \end{aligned} \right\} \quad (2.9)$$

のように成り立っていることも表 2.3 右端の (1+2+3) の列で確認できる。

ダミー変数を用いた線形モデル

平方和の分解による分散分析表の作成は、取り上げる因子 (変数) が全て質的変数であることを前提にしている。線形モデルによる解析では、質的変数を量的変数、いわゆるダミー変数に変換する必要がある。2 水準の (なし, あり) を (0, 1) 型ダミー変数とすると、線形モデルで推定されたパラメータが、「なし」に対する「あり」の差の推定値となり、結果の解釈がしやすい。そのために、ダミー変数といえば「(0, 1) 型ダミー変数に変換する方法である」と断定する人が多数いる。

2 水準の (1: 男, 2: 女) or (1: Female, 2: Male) とコード化されている場合に、(0, 1) 型ダミー変数か、(1, 0) 型ダミー変数か、どちらの型を選ぶかは、悩ましい問題でもある。そのまま (1, 2) 型ダミー変数としたらどうなのであろうか。実際に試してみると、パラメータの推定値は (0, 1) 型ダミー変数と同じであり、切片の推定値が異なるだけである。(1, 0) 型ダミー変数とすると、パラメータの推定値の符号が変化する。

多水準の因子、複数の因子を同時に取り扱い、交互作用なども考慮するような場合には、(0, 1) 型ダミー変数は、各種の推定値とその 95%信頼区間の計算に際し、不都合な側面が出てくる。これは、式 (2.1) で示したデータの構造式では、効果 α_i について

$$\sum_i \alpha_i = 0 \quad (2.10)$$

という制約条件が付いており、2水準であれば、

$$\left. \begin{array}{l} \alpha_1 + \alpha_2 = 0 \\ \alpha_2 = -\alpha_1 \end{array} \right\} \quad (2.11)$$

を満たす必要がある。このような制約条件を実現するためのダミー変数は、足して 0 となる (1, -1) 対比型ダミー変数が必要となる。因子 A が 4 水準の場合には、表 2.4 左に示すように第 4 水準が -1 となるような (1, -1) 対比型ダミー変数とすることにより、データの構造式の制約条件を満たすことができる。もちろん、表 2.4 右に示すように (0, 1) 型ダミー変数とすることも可能である。ダミー変数名を (a'_2 , a'_3 , a'_4) としたのは、ダミー変数が「1」となる因子 A の水準に対比させるための配慮である。

表 2.4 (1, -1) 対比型ダミー変数 vs. (0, 1) 型ダミー変数

	(1, -1) 対比型				(0, 1) 型		
	a_1	a_2	a_3		a'_2	a'_3	a'_4
A ₁	1	0	0	A ₁	0	0	0
A ₂	0	1	0	A ₂	1	0	0
A ₃	0	0	1	A ₃	0	1	0
A ₄	-1	-1	-1	A ₄	0	0	1
和	0	0	0	和	1	1	1
平均	0	0	0	平均	1/4	1/4	1/4

表 2.5 左に示すのは、切片の変数として $x_0 = 1$ を明示的に加え、因子 A を (1, -1) 対比型ダミー変数とし (20 行×4 列) のデザイン行列 \mathbf{X} を設定している。線形モデルは、切片の推定のための変数 x_0 を含めた式を

$$\left. \begin{array}{l} y_{ij} = \beta_0 x_{0,ij} + \beta_1 a_{1,ij} + \beta_2 a_{2,ij} + \beta_3 a_{3,ij} + \varepsilon_{ij} \\ \text{or} \\ y_{ij} = \theta_0 x_{0,ij} + \theta_1 a_{1,ij} + \theta_2 a_{2,ij} + \theta_3 a_{3,ij} + \varepsilon_{ij} \end{array} \right\} \quad (2.12)$$

とする。通常の回帰分析では、切片のための変数は内部で補ってくれるが、Excel により各種の推定値を求め、その 95%信頼区間を求めるためには、パラメータの分散と共分散が必要となる。それらを求めるためには、切片 $x_0 = 1$ を含めた (20 行×4 列) のデザイン行列 \mathbf{X} が必要不可欠である。

表 2.5 右上段に Excel の「分析ツール」の「回帰分析」を適用した結果を示す。デザイン行列には、「切片 x_0 」が含まれているが、「入力 X 範囲」の設定で「定数に 0 を使用」をオフとし、ダミー変数 (a_1 , a_2 , a_3) を選択した結果が示されている。ここでは、Excel の「回帰分析」の結果の出力の左側 5 列分に限定している。

表 2.5 対比型ダミー変数によるデザイン行列 X を用いた Excel の回帰分析

				デザイン行列 X								
A	i	j	y	x_0	a_1	a_2	a_3	分散分析表, 「定数に 0 を使用」 off				
A ₁	1	1	10.8	1	1	0	0	自由度	変動	分散	分散比	
	1	2	9.9	1	1	0	0	回帰	3	3.1000	1.0333	7.5841
	1	3	9.7	1	1	0	0	残差	16	2.1800	0.1363	$=\sigma^2$
	1	4	10.4	1	1	0	0	合計	19	5.2800		
	1	5	10.7	1	1	0	0					
A ₂	2	1	10.7	1	0	1	0	係数	標準誤差	t	P -値	
	2	2	10.6	1	0	1	0	$\hat{\beta}_0$ 切片 x_0	10.9000	0.0825	132.0606	0.0000
	2	3	11.0	1	0	1	0	$\hat{\beta}_1$ a_1	-0.6000	0.1430	-4.1970	0.0007
	2	4	10.8	1	0	1	0	$\hat{\beta}_2$ a_2	-0.1000	0.1430	-0.6995	0.4943
	2	5	10.9	1	0	1	0	$\hat{\beta}_3$ a_3	0.3000	0.1430	2.0985	0.0521
A ₃	3	1	11.4	1	0	0	1	パラメータの共分散行列 $\Sigma(\hat{\beta})=(X^T X)^{-1}\sigma^2$				
	3	2	10.7	1	0	0	1	$\hat{\beta}_0$	0.0068	0.0000	0.0000	0.0000
	3	3	10.9	1	0	0	1	$\hat{\beta}_1$	0.0000	0.0204	-0.0068	-0.0068
	3	4	11.3	1	0	0	1	$\hat{\beta}_2$	0.0000	-0.0068	0.0204	-0.0068
	3	5	11.7	1	0	0	1	$\hat{\beta}_3$	0.0000	-0.0068	-0.0068	0.0204
A ₄	4	1	11.9	1	-1	-1	-1	x_0	a_1	a_2	a_3	
	4	2	11.2	1	-1	-1	-1	=Minverse (Mmult (Transpose (X の範囲), X の範囲))* σ^2				
	4	3	11.0	1	-1	-1	-1					
	4	4	11.1	1	-1	-1	-1					
	4	5	11.3	1	-1	-1	-1					

Excel の「回帰分析」は、データを変更した場合などに対し、手動による再計算を行う必要があるため、自動計算の機能がある線形モデル（回帰分析）のための LinEst () 関数を使用することも可能である。ただし、LinEst () 関数の結果を活用するためには、あらかじめ表頭・表側に結果の解釈ができるように書式を整えておく必要があること、さらにパラメータの表示形式が逆順であることなどの理由で、Excel の「回帰分析」を使っている。ここに示した回帰分析の結果は、別シートに出力された結果の一部を選択・コピーし、デザイン行列の横にペーストして書式を整えた結果である。再計算が必要な場合には、新たなシート上に出力された結果の数値のみをコピーし「値のみ」をペーストすることにより、あらかじめ整えた書式が維持される。

表 2.5 右上段に示した分散分析表の「回帰」の行の変動が、表 2.3 に示した平方和の計算による $S_A = 3.10$ となり、残差の変動が $S_e = 2.18$ 、合計が $S_T = 5.28$ に対応する。切片の推定値 $\hat{\beta}_0 = 10.90$ は、一般平均 μ の推定値であり $\bar{y}_.$ に等しい。ダミー変数 a_1 のパラメータの推定値 $\hat{\beta}_1 = -0.60$ が、データの構造式における効果 α_1 に対応し $p = 0.0007$ と有意な差であり、他は有意な差でないことが読み取れる。表 2.5 右下段が (4 行 × 4 列) のパラメータの共分散行列 $\Sigma(\hat{\beta}) = (X^T X)^{-1}\sigma^2$ であり、その対角要素が、回帰パラメータ（係数）の分散の推定値であり、平方根を取ると標準誤差 $SE(\hat{\beta}_1)$ となる。2 行 2 列目の $Var(\hat{\beta}_1) = 0.0204$ の平方根が、 $\sqrt{Var(\hat{\beta}_1)} = 0.1430$ となり標準誤差 $SE(\hat{\beta}_1)$ に等しいことが確かめられる。なお、行列計算に不慣れな場合は、第 5 章を参照のこと。

デザイン行列 X を用いた計算の実際

行列による回帰分析の計算式は、統計の理論を重視する成書で多用されているが、実際のデータの解析方法を例示する成書では、極まれに飾り程度に示されている。Excel の行列関数による計算は、従来のスクリプト型の行列計算言語にない特徴を持っていて、四則演算のごとく誰にでも簡単に Excel シート上に表示された行列計算の結果を直接見ることができる。表 2.5 右下段に示したパラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算は、計算結果が出力される (4 行 × 4 列) のセルを最初に選択し、

$$\left. \begin{aligned} \Sigma(\hat{\beta}) &= (X^T X)^{-1} \hat{\sigma}^2 \\ &= \text{Minverse}(\text{Mmult}(\text{Transpose}(X\text{の範囲}), X\text{の範囲})) * \hat{\sigma}^2 \end{aligned} \right\} \quad (2.13)$$

のように行列関数による計算式を入力した結果が、(4 行 × 4 列) のセルに直接表示されている。ただし、周辺の変数名などは、Excel シート上に別途書き込んだものである。このように、元のデザイン行列 X に並列して共分散行列 $\Sigma(\hat{\beta})$ の計算結果を表示できることが Excel の特徴であり、良さでもある。従来のスクリプト型の行列計算言語を使った場合には、結果がテキスト形式で出力されるのが常であり、表 2.5 に示したようなコンパクトな結果を示すためには、表示形式を整える努力を必要とする。

実験データに限らず観察データに対する解析結果の解釈には、各種の推定値とその 95% 信頼区間のグラフ表示が欠かせない。と、言うのはやさしいが、それを実践するのは難儀である。Excel では、パラメータの共分散行列 $\Sigma(\hat{\beta})$ が同一シート上に得られ、それを活用することにより、どのような推定値に対しても分散を手軽に求めることができる。推定値の分散が計算できれば、95% 信頼区間は標準的な方法で求めることができ、Excel シート上にある推定値に 95% 信頼区間を重ね書きした予測プロファイルの作成も容易にできる。理論的には周知のことであり、新規性はまったくないのであるが、行列計算に引き続き、ストレスなく予測プロファイルの作成が手軽できるようになったことは、うれしいかぎりである。

伝統的な実験計画法の解析方法は、分散分析表の作成のための計算方法を示すことを目的にしており、各種の推定値の分散を求める際には、「分散の加法性」を注意深く適用した手計算を前提としてきた。パラメータの共分散行列 $\Sigma(\hat{\beta})$ を使った分散の計算方法は、線形モデルの理論で示されているものの、行列計算を前提にしているために難儀なことであった。そこで、「手計算」の方法と対比しつつ、パラメータの共分散行列 $\Sigma(\hat{\beta})$ を使った計算方法を対比して示す。

データの構造式における効果の推定

式(2.1)に示したデータの構造式に基づき、表 2.5 左に示したように、デザイン行列 X を設定し、変数 (x_0, a_1, a_2, a_3) を用いた回帰分析により得た推定値 $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$

を表 2.5 右に示した. 式 (2.1) のデータの構造式に対応づけると, $\hat{\mu} = \hat{\beta}_0 x_0 = \hat{\beta}_0$, $\hat{\alpha}_1 = \hat{\beta}_1 a_1 = \hat{\beta}_1$, $\hat{\alpha}_2 = \hat{\beta}_2 a_2 = \hat{\beta}_2$, $\hat{\alpha}_3 = \hat{\beta}_3 a_3 = \hat{\beta}_3$ が得られる. A_4 に対応する効果 $\hat{\alpha}_4$ は, 直接推定されていないので, 表 2.4 の (1. -1) 対比型ダミー変数の設定により,

$$\left. \begin{aligned} \hat{\alpha}_4 &= \hat{\beta}_1(-a_1) + \hat{\beta}_2(-a_2) + \hat{\beta}_3(-a_3) \\ &= -\hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3 \end{aligned} \right\} \quad (2.14)$$

で求められる.

表 2.6 に示すのは, パラメータの推定値 $\hat{\beta}$ を用いた各種の線形和 L の推定と 95%信頼区間の計算結果である. 線形和 $L^{(0)}$ に対し行ベクトル $\mathbf{l}^{(0)} = [1 \ 0 \ 0 \ 0]$ が設定され, $L^{(0)} = \hat{\mu} = \mathbf{l}^{(0)} \hat{\beta}$ によって一般平均 $\hat{\mu} = 10.900$ が推定されている. 同様に行ベクトル $\mathbf{l}^{(1)} = [0 \ 1 \ 0 \ 0]$ を用いて $L^{(1)} = \mathbf{l}^{(1)} \hat{\beta} = \hat{\beta}_1 a_1 = \hat{\alpha}_1 = -0.60$ が推定されて, 行ベクトル $\mathbf{l}^{(4)} = [0 \ -1 \ -1 \ -1]$ に対しては, $\hat{\alpha}_4 = 0.40$ が推定されている. これらの線形和 $L^{(i)} = \mathbf{l}^{(i)} \hat{\beta}$ の分散は, $Var(\mathbf{l}^{(i)} \hat{\beta}) = \mathbf{l}^{(i)} \Sigma(\hat{\beta})(\mathbf{l}^{(i)})^T$ によって求められ, この平方根が表 2.5 右中段のパラメータの推定値の標準誤差 SE に一致することが確かめられる.

表 2.6 データの構造式の効果 α_i の推定と 95%信頼区間

	L	l_0	l_1	l_2	l_3	$\hat{\beta}$	推定値	分散	標準誤差	幅	95%信頼区間	
		x_0	a_1	a_2	a_3		$\mathbf{l} \hat{\beta}$	$Var(\mathbf{l} \hat{\beta})$	SE	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
μ	$L^{(0)}$	1	0	0	0	10.900	10.9000	0.0068	0.0825	0.1750	10.7250	11.0750
α_1	$L^{(1)}$	0	1	0	0	-0.600	-0.6000	0.0204	0.1430	0.3031	-0.9031	-0.2969
α_2	$L^{(2)}$	0	0	1	0	-0.100	-0.1000	0.0204	0.1430	0.3031	-0.4031	0.2031
α_3	$L^{(3)}$	0	0	0	1	0.300	0.3000	0.0204	0.1430	0.3031	-0.0031	0.6031
α_4	$L^{(4)}$	0	-1	-1	-1		0.4000	0.0204	0.1430	0.3031	0.0969	0.7031
推定値 $\mathbf{l} \hat{\beta} = \text{Mmult}(\mathbf{l} \text{ 範囲}, \hat{\beta} \text{ の範囲})$							$t_{0.05}(20-4) =$		2.1199			
分散 $Var(\mathbf{l} \hat{\beta}) = \text{Mmult}(\text{Mmult}(\mathbf{l} \text{ の範囲}, \Sigma(\hat{\beta}) \text{ の範囲}), \text{Transpose}(\mathbf{l} \text{ の範囲}))$												

表 2.6 でデータの構造式に対応する効果 $\hat{\alpha}_i$ の推定値と 95%信頼区間が得られたので, Excel の折れ線グラフを用いて結果の表示を行なう. 図 2.2 に示すように, 1 因子実験での効果 $\hat{\alpha}_i$ と 95%信頼区間が示されている. 水準効果 $\hat{\alpha}_i$ は, 定義により全て加えると 0.0 となるので, 効果 $\hat{\alpha}_i$ は, 0.0 を挟む形になる.

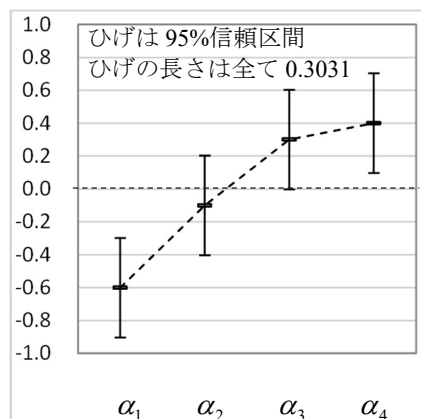


図 2.2 折れ線グラフによる効果 $\hat{\alpha}_i$ に対する予測プロファイル

水準平均の95%信頼区間

表 2.5 右に示したパラメータの推定値 $\hat{\beta}$ を用い、因子 A の A_i 水準平均 $\bar{y}_{i\cdot}$ 、および、その95%信頼区間を求める。表 2.7 に示すように、行ベクトル $L^{(5)}$ を用いて $L^{(5)}\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 = 10.30$ が推定され、 A_1 水準の表 2.1 右下段に示した算術平均と一致する。その分散は、 $Var(L^{(5)}\hat{\beta}) = L^{(5)}\Sigma(\hat{\beta})(L^{(5)})^T$ であり、 $\hat{\beta}_0$ と $\hat{\beta}_1$ の共分散は表 2.5 右下段から 0.0 なので

$$\left. \begin{aligned} Var(\hat{\beta}_0 + \hat{\beta}_1) &= Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1) + Var(\hat{\beta}_1) \\ &= 0.0068 + 0.0 + 0.0204 = 0.0273 \end{aligned} \right\} \quad (2.15)$$

と計算される。他の A_i 水準についても同様に求められている。

表 2.7 水準平均の95%信頼区間

	$\bar{y}_{i\cdot}$	L	$l_0 \quad l_1 \quad l_2 \quad l_3$				$\hat{\beta}$	推定値	分散	標準誤差	幅	95%信頼区間	
			x_0	a_1	a_2	a_3		$L\hat{\beta}$	$Var(L\hat{\beta})$	SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
A_1	$\bar{y}_{1\cdot}$	$L^{(5)}$	1	1	0	0	10.900	10.3000	0.0273	0.1651	0.3499	9.9501	10.6499
A_2	$\bar{y}_{2\cdot}$	$L^{(6)}$	1	0	1	0	-0.600	10.8000	0.0273	0.1651	0.3499	10.4501	11.1499
A_3	$\bar{y}_{3\cdot}$	$L^{(7)}$	1	0	0	1	-0.100	11.2000	0.0273	0.1651	0.3499	10.8501	11.5499
A_4	$\bar{y}_{4\cdot}$	$L^{(8)}$	1	-1	-1	-1	0.300	11.3000	0.0273	0.1651	0.3499	10.9501	11.6499

この結果を表 2.6 に示したデータの構造式の効果 $\hat{\alpha}_i$ の $SE(\hat{\alpha}_i) = 0.1430$ と比べると、 $SE(\bar{y}_{i\cdot}) = 0.1651$ と大きくなっている。これは、水準平均に $\bar{y}_{i\cdot} = \hat{\mu} + \hat{\alpha}_i$ のように一般平均の推定値 $\hat{\mu}$ が含まれているためである。

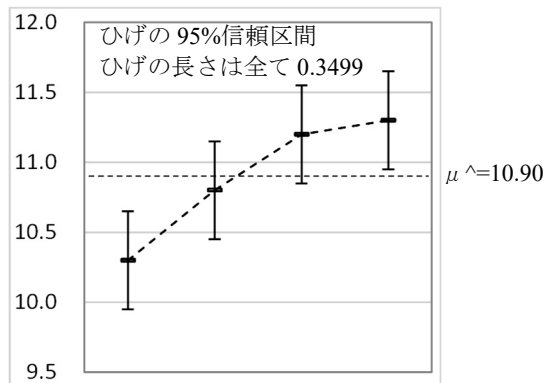


図 2.3 因子 A の水準平均と 95%信頼区間

水準間の差と95%信頼区間

A_1 水準を基準として他の水準の差および 95%信頼区間を求めてみよう。表 2.8 に示すように「 $(A_2 - A_1) L^{(9)}$ 」は、 A_2 水準の推定のための行ベクトル $L^{(6)}$ から A_1 水準の推定のために行ベクトル $L^{(5)}$ を引いたものである。表には示していないが、 $[(A_3 + A_4) / 2 - A_1]$ などのように、有意ではない A_3 水準と A_4 水準の平均値と A_1 水準との差なども同様に設定することができる。

表 2.8 A_2 と A_1 の差の推定のためのベクトル

			l_0	l_1	l_2	l_3	
			x_0	a_1	a_2	a_3	
	A_2	$L^{(6)}$	1	0	1	0	
-)	A_1	$L^{(5)}$	1	1	0	0	
A_2	-	A_1	$L^{(9)}$	0	-1	1	0

表 2.9 に各種の水準間の差について計算した結果を示す. (推定値, 分散, SE , t 値, $t_{0.05} \times SE$, $L95\%$, $U95\%$) の計算式は, 表 2.6 に示した効果 $\hat{\alpha}_i$ に対す場合と全く同じで, 線形和 $L^{(i)} = \mathbf{l}^{(i)} \hat{\boldsymbol{\beta}}$ の行ベクトル $\mathbf{l}^{(i)}$ を推定目的に合わせて設定している. また, これらの計算方法は, 繰り返しが不揃いの場合でも, 1 因子実験の場合に限らず, あらゆる実験モデルの解析で共通な汎用的な方法である.

表 2.9 因子 A の水準間の比較と 95%信頼区間

			l_0	l_1	l_2	l_3			推定値	分散	標準誤差	幅	95%信頼区間	
			x_0	a_1	a_2	a_3	$\hat{\boldsymbol{\beta}}$		$\mathbf{l} \hat{\boldsymbol{\beta}}$	$Var(\mathbf{l} \hat{\boldsymbol{\beta}})$	SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
A_2	-	A_1	$L^{(9)}$	0	-1	1	0	10.900	0.5000	0.0545	0.2335	0.4949	0.0051	0.9949
A_3	-	A_1	$L^{(10)}$	0	-1	0	1	-0.600	0.9000	0.0545	0.2335	0.4949	0.4051	1.3949
A_4	-	A_1	$L^{(11)}$	0	-2	-1	-1	-0.100	1.0000	0.0545	0.2335	0.4949	0.5051	1.4949
A_3	-	A_2	$L^{(12)}$	0	0	-1	1	0.300	0.4000	0.0545	0.2335	0.4949	-0.0949	0.8949
A_4	-	A_2	$L^{(13)}$	0	-1	-2	-1		0.5000	0.0545	0.2335	0.4949	0.0051	0.9949
A_4	-	A_3	$L^{(14)}$	0	-1	-1	-2		0.1000	0.0545	0.2335	0.4949	-0.3949	0.5949

表 2.9 で推定された因子 A の水準間の差と 95%信頼区間について, 図 2.4 に Excel の折れ線グラフを用いて描いた結果を示す.

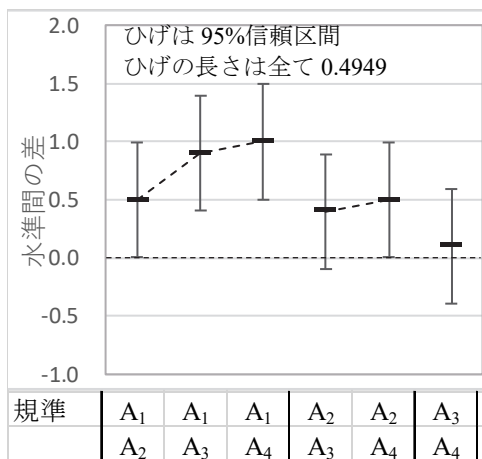


図 2.4 因子 A の水準間の差と 95%信頼区間

これまでも各種の推定値と 95%信頼区間の折れ線グラフを示してきたのであるが, 作図する際の幾つかのヒントを示す. 1) 推定値のマーカは, 組み込みオプションで \square を選択しサイズを拡大する. 2) 水準間のつなぎ線は, 適当な太さで適切な線種を選択する. 3) 上下の

ひげの幅を $t_{0.05}(df) \times SE$ で前もって計算し、ユーザ設定の誤差範囲とする。4) 水準間のつながり線の消去は、右側の点のみを選択し「線なし」にする。5) 込み入った軸ラベルは、Excelシート上に設定し、全て込みでコピーし、「図（拡張メタファイル）」として貼り付ける。

水準平均に対する伝統的な分散の推定

伝統的な実験データの解析では、各種の偏差平方和を計算し、それらを分散分析表としてまとめ、結果を吟味することを主体にしており、各種の推定値と95%信頼区間の計算については、統計の基礎である分散の加法性による対応を必要とする。

Excelの回帰分析などで作成された分散分析表には、「残差の分散」、「誤差の平均平方」などの表記で推定された誤差分散の推定値 $\hat{\sigma}^2$ が必ず示されている。表2.5右上段に示した分散分析表から $\hat{\sigma}^2 = 0.1363$ が得られている。この $\hat{\sigma}^2$ を使い、 A_1 水準の平均10.30は、5個のデータ y_{ij} の平均であり、それぞれの誤差 ε_{ij} は、互いに独立なので分散の加法性を用い、 A_1 水準平均の分散 $Var(\bar{A}_1)$ は、

$$\left. \begin{aligned} Var(\bar{A}_1) &= Var\left(\frac{y_{1,1} + y_{1,2} + \dots + y_{1,5}}{5}\right) \\ &= \frac{Var(y_{1,1})}{5^2} + \frac{Var(y_{1,2})}{5^2} + \dots + \frac{Var(y_{1,5})}{5^2} \\ &= \frac{5\hat{\sigma}^2}{5^2} = \frac{0.1363}{5} = 0.0273 \end{aligned} \right\} \quad (2.16)$$

との計算結果を得る。他の A_i 水準も同様に計算することができ、表2.7に示した分散に一致する。

水準平均の差とその分散の推定

A_2 水準と A_1 水準の差の分散は、 A_2 の平均と A_1 の平均が互いに独立なので、

$$\left. \begin{aligned} Var(\bar{A}_2 - \bar{A}_1) &= Var(\bar{A}_2) + (-1)^2 Var(\bar{A}_1) \\ &= 2 \frac{\hat{\sigma}^2}{5} = 2 \times 0.0273 = 0.0545 \end{aligned} \right\} \quad (2.17)$$

との計算結果を得る。このように、どのような推定なのかを見極めて分散の計算を別々行う必要がある。繰り返しが全て同じなので、因子Aの全ての水準の組合せの差についても同じ分散となり、表2.9に示した結果に一致する。

このように、「水準平均に対する分散の計算」と「水準間の差の分散の計算」が分散分析表の誤差分散の推定値 $\hat{\sigma}^2$ を用いて簡単に求めることができる事は素晴らしいのであるが、第1章で示した量的変数を含む実験データの解析に応用することができない。

水準効果の分散の推定

厄介なのは、因子 A の効果としての $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3, \hat{\alpha}_4)$ の推定値の分散の導出である。効果の推定値は、各水準の平均値 $\bar{y}_{i\cdot}$ から総平均 $\bar{y}_{\cdot\cdot}$ を差し引くことで、

$$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot} \quad (2.18)$$

簡単に求められる。因子 A の各水準の分散は、5 個のデータの平均なので、分散分析表の誤差分散の推定値 $\hat{\sigma}^2 = 0.1363$ を使って、 $Var(\bar{A}_i) = \hat{\sigma}^2 / 5 = 0.0273$ として求められた。ただし、因子 A の効果 $\hat{\alpha}_i$ は、

$$Var(\hat{\alpha}_i) = Var(\bar{A}_i - \hat{\mu}) = Var(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \quad (2.19)$$

で定義される。ここでは、 $\bar{y}_{\cdot\cdot}$ の中に $\bar{y}_{i\cdot}$ も含まれているので、互いに独立ではなく、分散の加法性が成り立たない。

デザイン行列 X を用いた場合に $Var(\hat{\alpha}_i) = Var(\hat{\beta}_i)$ は、表 2.5 に示したようパラメータの推定値の標準誤差 SE が直接推定されているので、 SE の 2 乗が分散として得られる。しかし、伝統的な分散分析による誤差分散 $\hat{\sigma}^2$ を用いた導出に際しては、個々のデータ y_{ij} に戻って分散を計算する必要がある。実際に計算すると

$$\left. \begin{aligned} Var(\hat{\alpha}_1) &= Var(\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot}) \\ &= Var\left(\frac{y_{1,1} + y_{1,1} + \dots + y_{1,5}}{5} - \frac{y_{1,1} + y_{1,1} + \dots + y_{2,1} + y_{2,1} + \dots + y_{4,5}}{20}\right) \\ &= Var\left(\frac{3y_{1,1} + 3y_{1,1} + \dots + 3y_{1,5}}{20} - \frac{y_{2,1} + y_{2,1} + \dots + y_{4,5}}{20}\right) \\ &= \frac{(3^2 \times 5)\hat{\sigma}^2}{20^2} + \frac{15\hat{\sigma}^2}{20^2} \\ &= \frac{60\hat{\sigma}^2}{400} = \frac{0.1363}{6.6667} = 0.0204 \end{aligned} \right\} \quad (2.20)$$

のように、 $\hat{\sigma}^2 = 0.1363$ を有効反復数 $n_e = 6.6667$ で割って分散 $Var(\hat{\alpha}_1) = 0.0204$ が得られる。この分散の平方根が、表 2.5 の回帰パラメータの $SE(\hat{\beta}_1) = 0.1430$ に一致する。

このような場合に伝統的な解析法では、水準効果の分散の推定に「伊奈の式」を使った有効反復数 n_e を算出して分散を計算する方法が知られている。他にも、「田口の式」による有効反復数 n_e の計算法などがあり、秘伝的な香りに満ちあふれている。しかも、繰り返し数が等しい場合のみ適用でき、繰り返し数が不揃いの場合には適用できないこので、発展性に欠ける。手計算時代には、素晴らしい解決方法であったことは間違いないが、新たな方法の発展・普及を阻害しているがごとくである。なお、「伊奈の式」については、芳賀 (2014) の「第 1.6 節 補遺 (1) 伊奈の法則」に詳細に示されている。

見方を変えれば、 \bar{A}_1 の推定値は $\bar{A}_1 = \hat{\mu} + \hat{\beta}_1$ であり、表 2.5 右下段から共分散が $Cov(\hat{\mu}, \hat{\beta}_1) = 0$ なので、分散の加法性が成り立ち

$$Var(\bar{A}_1) = Var(\hat{\mu}) + Var(\hat{\beta}_1) \quad (2.21)$$

のように成り立つ。したがって、効果の分散は、分散の加法性ではなく、分散の引き算

$$\left. \begin{aligned} Var(\hat{\beta}_1) &= Var(\bar{A}_1) - Var(\hat{\mu}) \\ &= Var(\bar{y}_{i.}) - Var(\bar{y}_{..}) \\ &= \frac{\hat{\sigma}^2}{5} - \frac{\hat{\sigma}^2}{20} \\ &= 0.0273 - 0.0068 = 0.0204 \end{aligned} \right\} \quad (2.22)$$

で求められる。このような各種の推定値の分散を誤差分散 $\hat{\sigma}^2$ から推定することは、統計の基礎的な素養を高めるために必要ではあるが、繰り返しが等しい場合に限定され、汎用性に欠ける。パラメータの共分散行列 $\Sigma(\hat{\beta})$ を使う方法は、線形和 $L^{(i)}$ を計算するためにベクトル $I^{(i)}$ を設定する必要があるが、同じ式で計算できる汎用的な方法である。

伝統的な「平方和の分解」と「分散の加法性」による1因子実験による解析では、各種の推定値の分散を求める際に、手作業的な計算を必要としている。このこと自体は、統計の理論の実践には欠かせないことも理解できるが、手計算の方法を強いる方法だけでいいのだろうか。現代の算盤である Excel を活用した線形モデルによる1元配置の解析方法も合わせて示すことが、実践のための応用力を付けるため不可欠のように思えてならない。線形モデルに対してデザイン行列 X を用いた Excel による一連の解析方法は、多くの統計ソフトと同様の解析方法であり、統計ソフトを使用する前の入門として適している。逆説的には、ブラックボックス的な統計ソフトによる解析方法を理解し、更なる応用力を付けるために、Excel による一連の解析方法の習得は、更なる学習の助けになる。

分散の加法性が成り立たない効果の差

分散の加法性が成り立つことを前提に実験計画法は、平方和の分解を使った分散分析表の作成が、基本的な解析方法として定着し、ほとんど全ての「実験計画法」に関する書物で踏襲されて続けている。嘆かわしいとしか言いようのないことであり、量的な変数が含まれる場合、欠測値が生じた場合、繰り返しが不揃いの場合などへの適用などの応用に際し、融通の利かない方法として自らの発展を阻害している。

デザイン行列 X を用いた回帰分析の結果を表 2.5 に示したのであるが、パラメータの共分散行列 $\Sigma(\hat{\beta})$ から $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ の間には、共通の共分散

$$Cov(\beta_i, \beta_j) = -0.0068$$

が存在していることが示されている。そのために、 A_2 水準の平均と A_1 水準の平均の差の推定値は

$$\begin{aligned}\bar{A}_2 - \bar{A}_1 &= (\hat{\mu} + \hat{\beta}_2) - (\hat{\mu} + \hat{\beta}_1) \\ &= \hat{\beta}_2 - \hat{\beta}_1 \\ &= -0.60 - (-0.10) = 0.50\end{aligned}$$

として求まるが、効果の差の分散を求める際に「分散の加法性」が成り立たないために、合成分散の一般式

$$\begin{aligned}Var(\hat{\beta}_2 - \hat{\beta}_1) &= Var(\hat{\beta}_1) - 2Cov(\hat{\beta}_2, \hat{\beta}_1) + Var(\hat{\beta}_2) \\ &= 0.0204 - 2 \times (-0.0068) + 0.0204 = 0.0545\end{aligned}$$

のようにパラメータ間の共分散を考慮しなければならない。

見方を変えて、 \bar{A}_2 の分散と \bar{A}_1 の分散を使った場合には、互いに独立なので、分散の加法性が成り立ち、

$$\begin{aligned}Var(\bar{A}_2 - \bar{A}_1) &= Var(\bar{A}_2) + Var(\bar{A}_1) \\ &= 0.0273 + 0.0273 = 0.0545\end{aligned}$$

のように、共分散を含めずに分散を求めることができる。

セル平均モデル

1因子実験のデータの構造式を式(2.1)では、因子Aの効果をもとに、(1) $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ としたのであるが、 $\mu + \alpha_i$ を $\mu_i = \mu + \alpha_i$ とし、

$$(2) \quad y_{ij} = \mu_i + \varepsilon_{ij} \tag{2.23}$$

のようなデータの構造式として表すことも可能である。式(2.23)は、一般平均 μ を含まないセル平均モデル (cell means model) と言われ、このモデルに対応するためには、表2.10に示すように (1, 1) 標示型ダミー変数を用いてデザイン行列 X を生成する必要がある。Excelの回帰分析では、「定数に0を設定」オプションをオンとすることにより、因子Aの各水準の平均値、SEなどを直接推定することができる。

表 2.10 (1, 1) 標示型ダミー変数

	(1, 1)型			
	a_1	a_2	a_3	a_4
A_1	1	0	0	0
A_2	0	1	0	0
A_3	0	0	1	0
A_4	0	0	0	1

2.2. 繰り返しが不揃いな1因子実験データ

繰り返しが等しい1因子実験の場合には、平方和の分解による従来の解析方法とデザイン行列 \mathbf{X} を用いた線形モデルの解析結果が一致することを前節で示した。繰り返しが不揃いの1因子実験の場合は、伝統的な平方和の分解による方法とデザイン行列 \mathbf{X} を用いた方法には、微妙な食い違いが発生する。芳賀（2014）の第1.2節で示されているデータを表2.11に示す。

実験は、 A_0 を対照群とし、処置群 $A_1 \sim A_4$ と比較したい。その際、基準となる対照群の繰り返し数を増やすことにより、 A_0 と $A_1 \sim A_4$ 群の差の標準誤差が小さくなることが期待される。表2.11に示すように全データ18個から求めた総平均は $\bar{y}_{..} = 50.50$ であり、水準平均の平均は、 $\hat{\mu} = 51.40$ と異なる。

表2.11 繰り返しが不揃いの1因子実験データ [芳賀（2014），表示1.2.1]

A	n	平均	繰り返し					
			1	2	3	4	5	6
A_0	6	46.00	43	45	42	47	49	50
A_1	3	49.00	47	51	49			
A_2	3	53.00	54	48	57			
A_3	3	58.00	55	58	61			
A_4	3	51.00	52	48	53			
総平均	18	50.50	$= \bar{y}_{..}$					
水準平均の平均		51.40	$= \hat{\mu}$					

図2.5に示すのは、前節と同様に表2.11のデータを行方向に並べ直し、データの散布図に平均値を上書きした結果である。さらに、全データに対する総平均 $\bar{y}_{..}$ 、水準平均の平均 $\hat{\mu}$ についても重ね書きしてある。

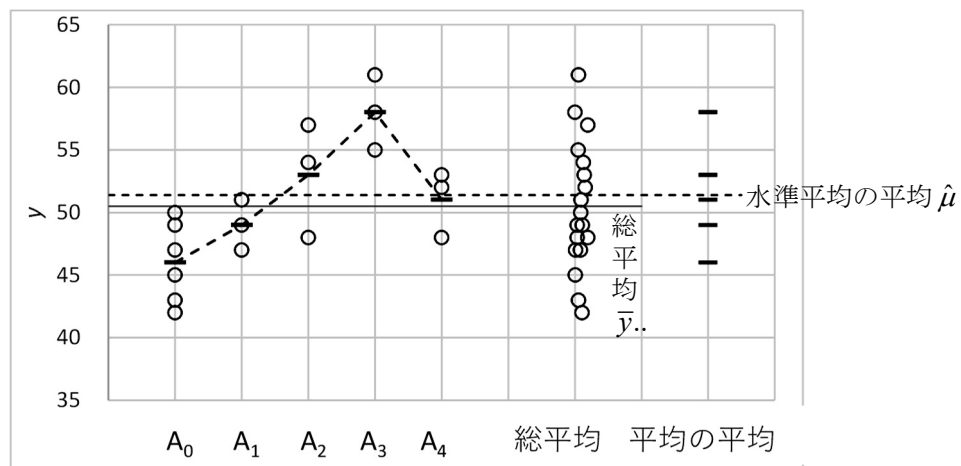


図2.5 平均の折れ線入りの散布図

平方和の分解

表 2.12 に示すように、平方和の計算を前節の表 2.3 に準じて行う。平方和 (S_T, S_A, S_e) の計算結果が示され、

$$\left. \begin{aligned} S_T &= S_A + S_e \\ &= 316.50 + 134.00 = 450.50 \end{aligned} \right\} \quad (2.24)$$

のように $S_T = S_A + S_e$ となることから、平方和の分解が成り立っていることが確認される。総平方和 S_T は、データの総平均 $\bar{y}_{..} = 50.50$ からの偏差平方和で計算している。計算量を減らすために、 S_A を計算しなくても $S_A = S_T - S_e$ として求めることもできるが、平方和の定義に基づいた計算が、各平方和の分解の意味を理解するために適している。

表 2.12 各種の平方和の計算

				1: μ^{\wedge}'			2: $\alpha^{\wedge}'_i$	3: $\varepsilon^{\wedge}'_{ij}$	1+2+3	分散分析表		
A	i	j	y_{ij}	$\bar{y}_{..}$	$y_{ij} - \bar{y}_{..}$	$\bar{y}_{i.}$	$\bar{y}_{i.} - \bar{y}_{..}$	$y_{ij} - \bar{y}_{i.}$	y_{ij}	要因	自由度	平方和
A ₀	0	1	43	50.50	-7.50	46.00	-4.50	-3.00	43.00	A	4	316.50
		2	45	50.50	-5.50	46.00	-4.50	-1.00	45.00	e	13	134.00
		3	42	50.50	-8.50	46.00	-4.50	-4.00	42.00	T	17	450.50
		4	47	50.50	-3.50	46.00	-4.50	1.00	47.00			
		5	49	50.50	-1.50	46.00	-4.50	3.00	49.00			
		6	50	50.50	-0.50	46.00	-4.50	4.00	50.00			
A ₁	1	1	47	50.50	-3.50	49.00	-1.50	-2.00	47.00			
		2	51	50.50	0.50	49.00	-1.50	2.00	51.00			
		3	49	50.50	-1.50	49.00	-1.50	0.00	49.00			
A ₂	2	1	54	50.50	3.50	53.00	2.50	1.00	54.00			
		2	48	50.50	-2.50	53.00	2.50	-5.00	48.00			
		3	57	50.50	6.50	53.00	2.50	4.00	57.00			
A ₃	3	1	55	50.50	4.50	58.00	7.50	-3.00	55.00			
		2	58	50.50	7.50	58.00	7.50	0.00	58.00			
		3	61	50.50	10.50	58.00	7.50	3.00	61.00			
A ₄	4	1	52	50.50	1.50	51.00	0.50	1.00	52.00			
		2	48	50.50	-2.50	51.00	0.50	-3.00	48.00			
		3	53	50.50	2.50	51.00	0.50	2.00	53.00			
平均		50.50	平方和	450.50		316.50	134.00					
				S_T		S_A	S_e					

因子 A の平方和 S_A は、各群の平均 $\bar{y}_{i.}$ と総平均 $\bar{y}_{..}$ の差の平方和を加算している。その差を効果 α'_i

$$\alpha'_i = \bar{y}_{i.} - \bar{y}_{..} \quad (2.25)$$

として表わす。ただし、効果 α'_i の合計は、表 2.13 に示すように

$$\sum_{i=0}^4 \alpha'_i = 4.50, \quad \sum_{i=0}^4 n_i \alpha'_i = 0 \quad (2.26)$$

となり、0 とはならないが、それぞれの群のデータ数 n_i での重み付きで合計とした場合には、0 となる。

表 2.13 総平均 $\bar{y}_{..}$ からの差の効果 $\hat{\alpha}'_i$ および水準平均の平均 $\hat{\mu}$ からの効果 $\hat{\alpha}_i$

	繰返し数	水準平均	総平均	$\bar{y}_i - \bar{y}_{..}$	重み付き	\bar{y}_i の平均	$\bar{y}_i - \hat{\mu}$
A	n_i	$\bar{A}_i : \bar{y}_i$	$\bar{y}_{..}$	$\hat{\alpha}'_i$	$n_i \hat{\alpha}'_i$	$\hat{\mu}$	$\hat{\alpha}_i$
A ₀	6	46.00	50.50	-4.50	-27.00	51.40	-5.40
A ₁	3	49.00	50.50	-1.50	-4.50	51.40	-2.40
A ₂	3	53.00	50.50	2.50	7.50	51.40	1.60
A ₃	3	58.00	50.50	7.50	22.50	51.40	6.60
A ₄	3	51.00	50.50	0.50	1.50	51.40	-0.40
	\bar{y}_i の平均	51.40	計	4.50	0.00	計	0.00

因子の効果 α_i を因子 A の水準平均の平均 $\hat{\mu} = 51.40$ (一般平均) からの差とすれば,

$$\hat{\alpha}_i = \bar{y}_i - (\text{水準平均の平均}) = \bar{y}_i - \hat{\mu} \quad (2.27)$$

$\hat{\alpha}_i$ の合計は, 表 2.13 に示すように 0.0 となる.

デザイン行列 X を用いた回帰分析

表 2.14 に示すのは, 前節と同様に因子 A に対し (1, -1) 対比型ダミー変数で設定された (18 行×5 列) のデザイン行列 X を用いて Excel の回帰分析を適用した結果である. 分散分析表は, 平方和の分解で求めた結果に一致する. ただし, 切片の推定値 $\hat{\beta}_{00} = 51.40$ は, 表 2.13 に示した水準平均の平均 $\hat{\mu}$ となり, 総平均 $\bar{y}_{..}$ とは異なる. 因子 A の各水準の推定値 $\hat{\beta}_i$ は, 各水準の平均値 \bar{y}_i と一般平均 $\hat{\mu}$ との差となる効果 $\hat{\alpha}_i$ が推定されている. デザイン行列 X を用いた場合の分散分析表は, どのようにして求められているのであろうか.

表 2.14 (1, -1) 対比型のダミー変数を用いた Excel による回帰分析

番号	Y		—デザイン行列 X—					分散分析表, 「定数に 0 を使用」 off						
群	i	j	y	x ₀	a ₀	a ₁	a ₂	a ₃						
A ₀	0	1	43	1	1	0	0	0		自由度	変動	分散	分散比	
	0	2	45	1	1	0	0	0	回帰	4	316.50	79.1250	7.6763	
	0	3	42	1	1	0	0	0	残差	13	134.00	10.3077	= σ^2	
	0	4	47	1	1	0	0	0	合計	17	450.50			
	0	5	49	1	1	0	0	0		係数	標準誤差	t	P-値	
	0	6	50	1	1	0	0	0						
A ₁	1	1	47	1	0	1	0	0	$\hat{\beta}^{\wedge}_{00}$ 切片 x ₀	51.400	0.7864	65.3592	0.0000	
	1	2	51	1	0	1	0	0	$\hat{\beta}^{\wedge}_{0}$ a ₀	-5.400	1.2842	-4.2049	0.0010	
	1	3	49	1	0	1	0	0	$\hat{\beta}^{\wedge}_{1}$ a ₁	-2.400	1.6371	-1.4660	0.1664	
A ₂	2	1	54	1	0	0	1	0	$\hat{\beta}^{\wedge}_{2}$ a ₂	1.600	1.6371	0.9774	0.3462	
	2	2	48	1	0	0	1	0	$\hat{\beta}^{\wedge}_{3}$ a ₃	6.600	1.6371	4.0316	0.0014	
	2	3	57	1	0	0	1	0						
A ₃	3	1	55	1	0	0	0	1	パラメータの共分散行列 $\Sigma(\hat{\beta}^{\wedge}) = (X^T X)^{-1} \sigma^2$					
	3	2	58	1	0	0	0	1	$\hat{\beta}^{\wedge}_{00}$	0.6185	-0.2749	0.0687	0.0687	0.0687
	3	3	61	1	0	0	0	1	$\hat{\beta}^{\wedge}_{0}$	-0.2749	1.6492	-0.4123	-0.4123	-0.4123
A ₄	4	1	52	1	-1	-1	-1	-1	$\hat{\beta}^{\wedge}_{1}$	0.0687	-0.4123	2.6800	-0.7559	-0.7559
	4	2	48	1	-1	-1	-1	-1	$\hat{\beta}^{\wedge}_{2}$	0.0687	-0.4123	-0.7559	2.6800	-0.7559
	4	3	53	1	-1	-1	-1	-1	$\hat{\beta}^{\wedge}_{3}$	0.0687	-0.4123	-0.7559	-0.7559	2.6800
									x ₀	a ₀	a ₁	a ₂	a ₃	
=Minverse (Mmult (Transpose (Xの範囲), Xの範囲))* σ^2														

デザイン行列 \mathbf{X} を用いた場合の分散分析表と平方和の分解による分散分析表が完全に一致するのは、なぜなのだろうか。分散分析表は、デザイン行列 \mathbf{X} を用いて得られたパラメータから推定される各水準の推定値をベースに組み立てられている。対照群 A_0 の場合では、

$$\left. \begin{aligned} \text{平方和の分解: } \bar{A}'_0 &= \bar{y}_{..} + \alpha'_0 = 50.50 - 4.50 = 46.00 \\ \text{デザイン行列: } \bar{A}_0 &= \hat{\mu} + \hat{\alpha}_0 = 51.40 - 5.40 = 46.00 \end{aligned} \right\} \quad (2.28)$$

と水準平均が一致するので、因子 A の平方和 S_A も一致する。また、回帰分析の場合でも総平方和 S_T は、平方和の分解と同様に y_{ij} と $\bar{y}_{..}$ の偏差平方和で計算されており、回帰パラメータとして推定された一般平均 $\hat{\mu} = 51.4$ からの平方和でない。

Excel の回帰分析の出力から、因子 A の全体（回帰）に対し $F = 7.6763$ と有意な差であること、ダミー変数 a_0 のパラメータの推定値が $\hat{\beta}_0 = -5.40$ 、 $p = 0.0010$ と有意な差、ダミー変数 a_3 のパラメータの推定値が $\hat{\beta}_3 = 6.60$ 、 $p = 0.0014$ と有意な差であり、他は有意な差でないことが読み取れる。各種の推定値の標準誤差 SE の算出には、表 2.14 右下段のパラメータの共分散行列 $\Sigma(\hat{\beta})$ が必要で、ちなみに $\Sigma(\hat{\beta})$ の（1行・1列）目は、切片 x_0 の分散で、標準誤差の2乗 $Var(\hat{\beta}_1) = 0.7864^2 = 0.6185$ に一致する。

パラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算は、（5行×18列）の転置行列 \mathbf{X}^T と（18行×5列）の \mathbf{X} の積和行列 $(\mathbf{X}^T \mathbf{X})$ は（5行×5列）の正方行列となり、 $(\mathbf{X}^T \mathbf{X})$ の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ も同じ（5行×5列）の正方行列となる。Excel シート上では、（5行×5列）のセルを最初に選択し、Excel の行列関数による計算結果に誤差分散の推定値 $\hat{\sigma}^2$ を掛けて

$$\left. \begin{aligned} \Sigma(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2 \\ &= \text{Minverse}(\text{Mmult}(\text{Transpose}(\mathbf{X}\text{の範囲}), \mathbf{X}\text{の範囲})) * \hat{\sigma}^2 \end{aligned} \right\} \quad (2.29)$$

として求めることができる。もちろん、繰り返しが等しい場合でも、計算式は全く同じである。

データの構造式における効果

表 2.13 に示したように繰り返し不揃いの1因子実験のデータの構造式は、繰り返しが等しい場合の式 (2.1) とまったく同じ、

$$\left. \begin{aligned} y_{ij} &= \mu + \alpha_i + \varepsilon_{ij} \\ \text{ただし, } \sum_i \alpha_i &= 0, \quad \sum_i \sum_j \varepsilon_{ij} = 0 \end{aligned} \right\} \quad (2.30)$$

であり、表 2.14 で示したパラメータ推定値に示すように、切片 x_0 のパラメータの推定値として、一般平均 $\hat{\mu} = 51.40$ が推定されている。効果 $\hat{\alpha}_0$ は、 $\hat{\alpha}_0 = \hat{\beta}_0 = -5.40$ であり、同様に $\hat{\alpha}_1 = -2.40$ 、 $\hat{\alpha}_2 = 1.60$ 、 $\hat{\alpha}_3 = 6.60$ が求められている。効果 $\hat{\alpha}_4$ に対応する変数がないので、

$\hat{\alpha}_4 = -\hat{\beta}_0 - \hat{\beta}_1 - \hat{\beta}_2 - \hat{\beta}_3$ で計算する必要がある. 表 2.14 には, $\hat{\alpha}_3$ までの効果についての SE が計算されているが, 効果 $\hat{\alpha}_4$ の SE は, どのように求めたらよいのであろうか.

表 2.15 に示すように効果 $\hat{\alpha}_4$ は, 線形和

$$\hat{\alpha}_4 : L^{(6)} = I^{(6)}\hat{\beta} = [0 \quad -1 \quad -1 \quad -1 \quad -1] \begin{bmatrix} 51.40 \\ -5.40 \\ -2.40 \\ 1.60 \\ 6.60 \end{bmatrix} = -0.400 \quad (2.31)$$

として計算している. SE は, 表 2.14 右下段で計算されているパラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いて Excel シートのセルに埋め込まれた計算式によって

$$SE(\hat{\alpha}_4) = \sqrt{Var(\hat{\alpha}_4)} = \sqrt{I^{(6)}\Sigma(\hat{\beta})(I^{(6)})^T} = 1.6371 \quad (2.32)$$

として計算できる. 結果として, 効果 $\hat{\alpha}_4$ の SE は, 繰り返しが 3 の他の $\hat{\alpha}_i$ の SE と等しいことが分かる. 表 2.13 で省略した 95%信頼区間についても表 2.15 で改めて計算した結果を示す.

表 2.15 効果 α_i の推定と 95%信頼区間

		l_{00}	l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間	
		x_0	a_0	a_1	a_2	a_3	$\hat{\beta}$	$I\hat{\beta}$	$Var(I\hat{\beta})$	SE	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
	μ	$L^{(1)}$	1	0	0	0	51.400	51.40	0.6185	0.7864	1.6990	49.70	53.10
効果	α_0	$L^{(2)}$	0	1	0	0	-5.400	-5.40	1.6492	1.2842	2.7744	-8.17	-2.63
	α_1	$L^{(3)}$	0	0	1	0	-2.400	-2.40	2.6800	1.6371	3.5367	-5.94	1.14
	α_2	$L^{(4)}$	0	0	0	1	1.600	1.60	2.6800	1.6371	3.5367	-1.94	5.14
	α_3	$L^{(5)}$	0	0	0	1	6.600	6.60	2.6800	1.6371	3.5367	3.06	10.14
	α_4	$L^{(6)}$	0	-1	-1	-1	-0.400	-0.40	2.6800	1.6371	3.5367	-3.94	3.14
										$t_{0.05}(18-5) =$	2.1604		

図 2.6 に, 表 2.15 で計算された, 効果 $\hat{\alpha}_i$ の推定値と 95%信頼区間を Excel の折れ線グラフで作図した結果を示す. 効果 $\hat{\alpha}_0$ の 95%信頼区間が他よりも短くなっているのは, 繰り返し数が 6 で, 他の群よりも大きいためである.

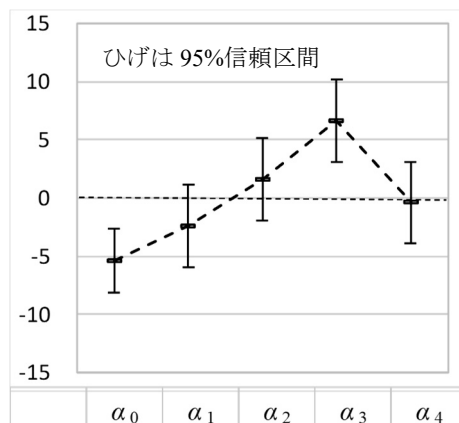


図 2.6 折れ線グラフによる効果に対する予測プロファイル

水準平均に対する95%信頼区間

水準平均は、単純な算術平均であり、表2.11に示したごとくであり、その分散も表2.14左上段に示した分散分析表の誤差分散 $\hat{\sigma}^2 = 10.3077$ を用い、式(2.16)と同様に個々の y_{ij} に戻ると分散の加法性が成り立つので、反復数が6の A_0 の $\bar{y}_{0.}$ の場合であれば、

$$Var(\bar{y}_{0.}) = \frac{\hat{\sigma}^2}{6} = \frac{10.3077}{6} = 1.7179$$

であり、反復数が3の A_1 場合は

$$Var(\bar{y}_{1.}) = \frac{\hat{\sigma}^2}{3} = \frac{10.3077}{3} = 3.4359$$

と計算することができる。表2.16に示すように行列関数を用いて大げさに計算する必要はないのであるが、線形和による推定値、行ベクトル $\mathbf{l}^{(i)}$ の2次形式による分散の計算で結果が一致することが確認できる。

表 2.16 水準平均の推定と95%信頼区間

		l_{00}	l_0	l_1	l_2	l_3		推定値	分散	標準誤差	幅	95%信頼区間		
		L	x_0	a_0	a_1	a_2	a_3	$\hat{\beta}$	$\mathbf{l}\hat{\beta}$	$Var(\mathbf{l}\hat{\beta})$	SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
平均	A_0	$L^{(7)}$	1	1	0	0	0	51.400	46.00	1.7179	1.3107	2.8316	43.17	48.83
	A_1	$L^{(8)}$	1	0	1	0	0	-5.400	49.00	3.4359	1.8536	4.0045	45.00	53.00
	A_2	$L^{(9)}$	1	0	0	1	0	-2.400	53.00	3.4359	1.8536	4.0045	49.00	57.00
	A_3	$L^{(10)}$	1	0	0	0	1	1.600	58.00	3.4359	1.8536	4.0045	54.00	62.00
	A_4	$L^{(11)}$	1	-1	-1	-1	-1	6.600	51.00	3.4359	1.8536	4.0045	47.00	55.00
											$t_{0.05}(18-5) =$	2.1604		

水準平均の差に対する95%信頼区間

対照群と設定された A_0 群と他の群との差およびその95%信頼区間を求めたい。表2.16に示した A_0 水準の平均の推定値を求めるために $\mathbf{l}^{(7)} = [1 \ 1 \ 0 \ 0 \ 0]$ が設定され、同様に A_1 水準では、 $\mathbf{l}^{(8)} = [1 \ 0 \ 1 \ 0 \ 0]$ なので、その差は、表2.17に示すように $\mathbf{l}^{(12)} = \mathbf{l}^{(8)} - \mathbf{l}^{(7)}$ となる。

表 2.17 水準間の差の推定のためのベクトル \mathbf{x}'_{12} の生成

				l_{00}	l_0	l_1	l_2	l_3	
				L	x_0	a_0	a_1	a_2	a_3
		A_1	$L^{(8)}$	1	0	1	0	0	
-)		A_0	$L^{(7)}$	1	1	0	0	0	
A_1	-	A_0	$L^{(12)}$	0	-1	1	0	0	

他の群に対しても同様に差のベクトルのように行ベクトル $\mathbf{l}^{(i)}$ を求め、表2.18に示すように各種の差の推定値と95%信頼区間が計算されている。

表 2.18 パラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた各種の推定

		l_{00}	l_0	l_1	l_2	l_3			推定値	分散	標準誤差	幅	95%信頼区間	
		L	x_0	a_0	a_1	a_2	a_3	$\hat{\beta}$	$l\hat{\beta}$	$Var(l\hat{\beta})$	SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
A_1	- A_0	$L^{(12)}$	0	-1	1	0	0	51.400	3.00	5.1538	2.2702	4.9045	-1.90	7.90
A_2	- A_0	$L^{(13)}$	0	-1	0	1	0	-5.400	7.00	5.1538	2.2702	4.9045	2.10	11.90
A_3	- A_0	$L^{(14)}$	0	-1	0	0	1	-2.400	12.00	5.1538	2.2702	4.9045	7.10	16.90
A_4	- A_0	$L^{(15)}$	0	-2	-1	-1	-1	1.600	5.00	5.1538	2.2702	4.9045	0.10	9.90
A_2	- A_1	$L^{(16)}$	0	0	-1	1	0	6.600	4.00	6.8718	2.6214	5.6632	-1.66	9.66
:	:													
A_4	- A_3	$L^{(17)}$	0	-1	-1	-1	-2		-7.00	6.8718	-2.6703	5.6632	-12.66	-1.34
											$t_{0.05}(18-5)=$	2.1604		

これらの推定値と 95%信頼区間は、全て何らかのグラフで表示することが、解析結果を理解し伝えるために必須である。図 2.6 に示した「効果」の予測プロファイルは、有効性の量的な判断のためには有益であるが、図 2.7 左に示すように伝統的な水準平均の予測プロファイルで表わすことも理解を深めるために有益である。また、図 2.7 右に示すように水準間の差の予測プロファイルは、統計的な判定を直接確認でき結果の解釈に役立つ。

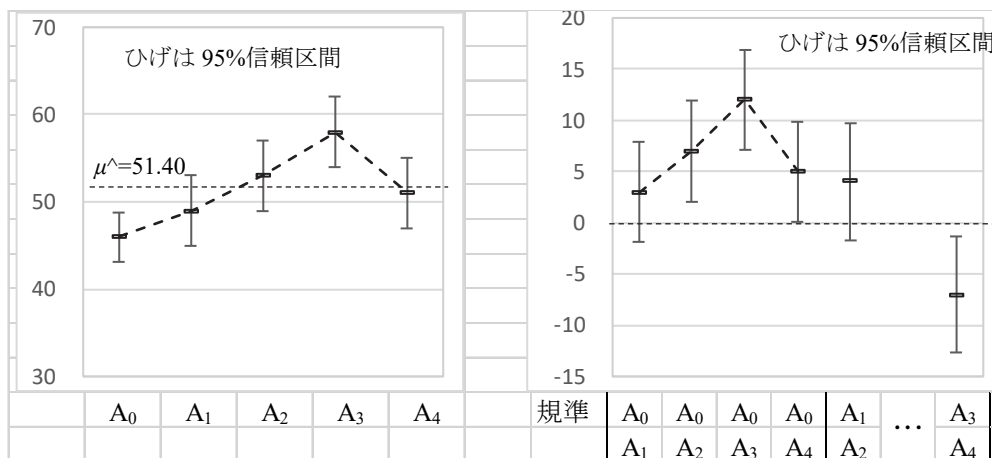


図 2.7 折れ線グラフによる水準平均と水準間の差の予測プロファイル

分散の加法性での対応と限界

パラメータの共分散行列 $\Sigma(\hat{\beta})$ を使わない場合の 95%信頼区間の算出には、分散分析表の誤差分散の推定値 $\hat{\sigma}^2 = 10.3077$ を使って、どのような推定なのかを見極めて計算式を設定する必要がある、画一的な計算式で表わすことができない。 A_0 群の平均 46.00 は、6 個のデータ y_{ij} の平均なので、分散の加法性を用い

$$\begin{aligned}
 Var(\bar{A}_0) &= Var\left(\frac{y_{1,1} + y_{1,2} + \dots + y_{1,6}}{6}\right) \\
 &= \frac{Var(y_{1,1})}{6^2} + \frac{Var(y_{1,2})}{6^2} + \dots + \frac{Var(y_{1,6})}{6^2} \\
 &= \frac{6\hat{\sigma}^2}{6^2} = \frac{10.3077}{6} = 1.7179
 \end{aligned}
 \quad (2.33)$$

との計算結果を得る. もちろん, 表 2.16 に示す $L^{(7)}$ の分散に一致する. 表 2.18 に示す $L^{(12)}$ の分散は, A_1 群の平均と A_0 群の平均が互いに独立なので,

$$\left. \begin{aligned} \text{Var}(\bar{A}_1 - \bar{A}_0) &= \text{Var}(\bar{A}_1) + (-1)^2 \text{Var}(\bar{A}_0) \\ &= \frac{\hat{\sigma}^2}{3} + \frac{\hat{\sigma}^2}{6} \\ &= \frac{\hat{\sigma}^2}{2} = \frac{10.3077}{2} = 5.1538 \end{aligned} \right\} \quad (2.34)$$

との計算結果に一致する. このように, どのような推定なのかを見極めて分散の計算を別々行う必要がある.

因子 A の各水準の分散は, 6 個または 3 個のデータの平均なので, 分散分析表の誤差分散 $\hat{\sigma}^2 = 10.3077$ を使って,

$$\left. \begin{aligned} \text{Var}(\bar{A}_0) &= \frac{\hat{\sigma}^2}{6} = \frac{10.3077}{6} = 1.7179 \\ \text{Var}(\bar{A}_1) &= \frac{\hat{\sigma}^2}{3} = \frac{10.3077}{3} = 3.4359 \end{aligned} \right\} \quad (2.35)$$

として求められる. ただし, 厄介なのは, 因子 A の効果としての α_i (β_i に対応) の分散の導出であり, $\text{Var}(\hat{\beta}_i)$ は,

$$\text{Var}(\hat{\beta}_i) = \text{Var}(\bar{A}_i - \hat{\mu}) = \text{Var}\left(\bar{y}_i - \frac{\bar{y}_0. + \bar{y}_1. + \bar{y}_2. + \bar{y}_3. + \bar{y}_4.}{5}\right) \quad (2.36)$$

となり, $\hat{\mu}$ の計算式の中に \bar{y}_i が含まれているので, 互いに独立ではなく, 分散の加法性が成り立たない. 繰り返しが等しい場合は, $\hat{\mu}$ からの差ではなく $\text{Var}(\bar{A}_i - \bar{y}_{..})$ のように $\bar{y}_{..}$ からの差であったので, 式 (2.20) に示したように元の y_{ij} の分散に戻り, 再構成することにより有効反復数 n_e を求めることができたが, 繰り返し不揃いの場合は, 計算することは容易ではない. (1, -1) 対比型ダミー変数を用いた回帰分析を適用すれば, パラメータ $\hat{\beta}_i$ の SE が求められているので, 有効反復数 n_e を計算する必要はない.

式 (2.30) に示した 1 因子実験に関するデータの構造式は, 一般平均 μ , 水準の効果 α_i , 誤差 ε_{ij} の和として定義されている. 繰り返しが不揃いな 1 因子実験データの解析において (1, -1) 対比型ダミー変数によるデザイン行列 \mathbf{X} を設定することにより, 回帰分析のパラメータの推定値として, μ および α_i が得られ, それらの標準誤差 SE も同時に得られることを示した. 伝統的な平方和の分解による方法は, 分散分析表を作成, 表 2.13 に示したように μ および α_i の推定はできるのであるが, 統計的な考察に必要な標準誤差 SE を求めるすべがない.

2.3. (0, 1) 型ダミー変を用いた 1 因子実験

繰り返しが等しい 1 因子実験データに対し第 2.1 節では、(1, -1) 対比型のダミー変数を用いたデザイン行列 \mathbf{X} による線形モデルを取り上げた。その際に (0, 1) 型ダミー変数については、線形モデルを念頭にした場合に不向きであることを示唆しただけで、理由を示さなかった。そこで、第 2.2 節で取り上げた繰り返しが不揃いの事例を用い、実際に適用した結果により明らかにする。

表 2.19 右に示すように、(0, 1) 型ダミー変数は、質的変数の最初の水準 A_0 に対して (0, 0, 0, 0) を与え、2 番目の水準 A_1 に (1, 0, 0, 0) を与え、最後の水準 A_4 に (0, 0, 0, 1) を与える。ダミー変数名としては、1 と設定されている水準に対応付けることにすると、因子 A が (A_0, A_1, \dots, A_4) となっているので、ダミー変数名とし (a_1, a_2, a_3, a_4) を付与している。

表 2.19 (1, -1) 対比型 vs. (0, 1) 型ダミー変数

	切片	(1, -1) 対比型					切片	(0, 1) 型			
	x_0	a_0	a_1	a_2	a_3		x_0	a_1	a_2	a_3	a_4
A_0	1	1	0	0	0	A_0	1	0	0	0	0
A_1	1	0	1	0	0	A_1	1	1	0	0	0
A_2	1	0	0	1	0	A_2	1	0	1	0	0
A_3	1	0	0	0	1	A_3	1	0	0	1	0
A_4	1	-1	-1	-1	-1	A_4	1	0	0	0	1
和	5	0	0	0	0	和	5	1	1	1	1
平均	1	0	0	0	0	平均	1	1/5	1/5	1/5	1/5

(0, 1) 型ダミー変数によるデザイン行列 \mathbf{X} を用いた場合

表 2.20 に示すのは、(0, 1) 型ダミー変数のデザイン行列 \mathbf{X} を用いた回帰分析の結果である。表 2.14 に示した (1, -1) 対比型ダミー変数と比べて大きく異なるのは、最後の水準 A_4 が (-1, -1, -1, -1) から (0, 0, 0, 1) となり、すっきりとしている。このことが、(0, 1) 型ダミー変数が好まれる理由でもある。

表 2.20 右上段に示す分散分析表は、表 2.14 右上段の分散分析表と完全に一致しているが、パラメータの推定値は、まったく異なる。どのように理解したら良いのであろうか。切片 x_0 の推定値が、 $\hat{\beta}_0 = 46.00$ となっている。 A_0 のダミー変数は、 $a_1 = 0, \dots, a_4 = 0$ であり、切片が $x_0 = 1$ となっているので、切片のパラメータ $\hat{\beta}_0 = 46.00$ は、最初の水準 A_0 の推定値になる。

表 2.20 (0, 1) 型のデザイン行列 X に対する回帰分析

群	番号		Y	X					分散分析表, 「定数に 0 を使用」 off					
	i	j	y	x ₀	a ₁	a ₂	a ₃	a ₄	自由度	変動	分散	分散比		
A ₀	0	1	43	1	0	0	0	0	回帰	4	316.50	79.1250	7.6763	
	0	2	45	1	0	0	0	0						
	0	3	42	1	0	0	0	0						
	0	4	47	1	0	0	0	0						
	0	5	49	1	0	0	0	0						
	0	6	50	1	0	0	0	0						
A ₁	1	1	47	1	1	0	0	0	β ⁰ _{切片 x₀}	46.0000	1.3107	35.0956	0.0000	
	1	2	51	1	1	0	0	0	β ¹ _{a₁}	3.0000	2.2702	1.3215	0.2091	
	1	3	49	1	1	0	0	0	β ² _{a₂}	7.0000	2.2702	3.0834	0.0087	
A ₂	2	1	54	1	0	1	0	0	β ³ _{a₃}	12.0000	2.2702	5.2859	0.0001	
	2	2	48	1	0	1	0	0	β ⁴ _{a₄}	5.0000	2.2702	2.2024	0.0463	
	2	3	57	1	0	1	0	0						
A ₃	3	1	55	1	0	0	1	0	パラメータの共分散行列 Σ(β ⁰) = (X ^T X) ⁻¹ σ ²					
	3	2	58	1	0	0	1	0	β ⁰	1.7179	-1.7179	-1.7179	-1.7179	-1.7179
	3	3	61	1	0	0	1	0	β ¹	-1.7179	5.1538	1.7179	1.7179	1.7179
A ₄	4	1	52	1	0	0	0	1	β ²	-1.7179	1.7179	5.1538	1.7179	1.7179
	4	2	48	1	0	0	0	1	β ³	-1.7179	1.7179	1.7179	5.1538	1.7179
	4	3	53	1	0	0	0	1	β ⁴	-1.7179	1.7179	1.7179	1.7179	5.1538
									x ₀	a ₁	a ₂	a ₃	a ₄	
=Minverse (Mmult (Transpose (Xの範囲), Xの範囲))*σ ²														

第2の水準 A₁ の推定値は、表 2.19 の (0, 1) 型ダミー変数から

$$\left. \begin{aligned}
 \bar{A}_1 &= \hat{\beta}_0 x_0 + \hat{\beta}_1 a_1 \\
 &= \hat{\beta}_0 + \hat{\beta}_1 \\
 &= 46.0 + 3.0 = 49.0
 \end{aligned} \right\} \quad (2.37)$$

として求められる。この関係から $\hat{\beta}_1 = 3.00$ は、 \bar{A}_1 は、 \bar{A}_0 からの増分 (差分) であることが分かる。

表 2.21 に示すように、(0, 1) 型ダミー変数を用いた場合の推定値は、最初の水準の推定値からの差分となると理解される。対照群 A₀ を基準とし A₁~A₄ との差分だけに興味がある場合に、(0, 1) 型ダミー変数を用いた線形モデルを適用すれば、対照群 A₀ との差分に関する t 値および p 値が得られる。表 2.21 には含めていないが、95%信頼区間も得られるので、対照群 A₀ との差の検定を目的とする場合に最も適している。

表 2.21 (0, 1) 型ダミー変数のデザイン行列 X を用いたパラメータ推定

A	n	推定値		和 β ⁰ + β ⁱ	水準平均 ȳ _i
		β ⁱ			
A ₀	6	β ⁰	46.00	—	46.00
A ₁	3	β ¹	3.00	49.00	49.00
A ₂	3	β ²	7.00	53.00	53.00
A ₃	3	β ³	12.00	58.00	58.00
A ₄	3	β ⁴	5.00	51.00	51.00

(0, 1)型ダミー変数の場合の各水準の 95%信頼区間

表 2.22 に示すようにダミー変数の型が異なったとしても、因子 A の各水準の推定値は、得られたパラメータから、

$$\bar{A}_0 = 46.0, \bar{A}_1 = 46.0 + 3.0 = 49.0, \dots, \bar{A}_4 = 46.0 + 5.0 = 51.0$$

となり、(1, -1) 対比型ダミー変数の場合の推定値は、

$$\bar{A}_0 = 51.4 - 5.4 = 46.0, \bar{A}_1 = 51.4 - 2.4 = 49.0, \dots, \bar{A}_4 = 51.4 + 5.4 + 2.4 - 1.6 - 6.6 = 51.0$$

と、全く同じになる。

表 2.22 パラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた平均値の 95%信頼区間

		l_0	l_1	l_2	l_3	l_4		推定値	分散		幅	95%信頼区間		
L		x_0	a_1	a_2	a_3	a_4	$\hat{\beta}$	$l\hat{\beta}$	$Var(l\hat{\beta})$	SE	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$	
平均	$A_0 L^{(1)}$	1	0	0	0	0	46.00	46.00	1.7179	1.3107	2.8316	43.17	48.83	
	$A_1 L^{(2)}$	1	1	0	0	0	3.00	49.00	3.4359	1.8536	4.0045	45.00	53.00	
	$A_2 L^{(3)}$	1	0	1	0	0	7.00	53.00	3.4359	1.8536	4.0045	49.00	57.00	
	$A_3 L^{(4)}$	1	0	0	1	0	12.00	58.00	3.4359	1.8536	4.0045	54.00	62.00	
	$A_4 L^{(5)}$	1	0	0	0	1	5.00	51.00	3.4359	1.8536	4.0045	47.00	55.00	
											$t_{0.05}(18-5)=$	2.1604		

因子 A の各水準の 95%信頼区間も報告書に掲載したい。そのためには、表 2.20 右下段に示したパラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた各種の推定が必要となる。 A_1 水準の平均値は、

$$\bar{A}_1 = \hat{\beta}_0 + \hat{\beta}_1 = 46.0 + 3.0 = 49.0$$

であり、その分散は、

$$\begin{aligned} Var(\hat{\beta}_0 + \hat{\beta}_1) &= Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1) + Var(\hat{\beta}_1) \\ &= 1.7179 + 2 \times (-1.7179) + 5.1538 = 3.4359 \end{aligned}$$

として求められる。

さて、式 (2.30) に示したデータの構造式に示した一般平均 μ と各種準の効果 α_i を推定したい。そのためには、表 2.19 右の下段の「平均」の係数 $l^{(\mu)} = [1 \ 1/5 \ 1/5 \ 1/5 \ 1/5]$ を用いて

$$\begin{aligned} \hat{\mu} &= 1 \times \hat{\beta}_0 + \frac{1}{5} \times (\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 + \hat{\beta}_4) \\ &= 46.00 + \frac{1}{5} \times (3.00 + 7.00 + 12.00 + 5.00) = 51.40 \end{aligned} \tag{2.38}$$

のような計算をする必要がある。効果 α_i は、水準平均 \bar{A}_i と一般平均 μ との差なので、表 2.22 の線形和のための係数ベクトル $l^{(i)}$ と一般平均 μ の係数ベクトル $l^{(\mu)}$ との差を必要となり、かなり面倒な設定を必要とする。単に分散分析表と \bar{A}_0 と \bar{A}_i の水準平均の差と 95%信頼区間の推定のためだけならば (0, 1) 型ダミー変数が適しているが、効果 α_i の推定に際しては、

(1, -1) 対比型ダミー変数が最善である。したがって、解析の目的に合わせたダミー変数の選択ができるようになることが肝要である。

2.4. 1 因子実験の量的変数に対する多項式回帰

量的変数に対して幾つかの水準を設定して質的変数として解析する方法が広く普及している。手計算時代には、量的変数のまま回帰分析を適用すること自体が難儀であった。そのために、量的変数を質的変数として扱うことにより平方和の分解による分散分析表を作成し、 F 検定による統計的な考察が行えることは、画期的な方法であり、実験データの解析法として合理的な方法であったともいえる。さて、 F 検定で何らかの水準間に差が統計的にあった場合に、さらに、直線的な関係なのか、あるいは曲線的な関係なのか明らかにしたい。

量的変数を質的変数に置き換えた分散分析では、直線関係か否かの判定に LOF (Lack Of Fit, あてはまりの悪さ) 解析が知られている。芳賀 (2014) の第2章「量的因子の1因子実験」に LOF 解析が展開されているが、伝統的な平方和の分解による分散分析と回帰分析を併合した解析法となっていて煩雑で見通しが悪い。そこで、デザイン行列 X を用いた多項式回帰の適用により、見通しの良い解析法を提示する。表 2.23 に示すのは、対照群 0 mg/kg に対し、10, 20, 30 mg/kg を 5 匹の動物に対し投与し、ある薬理反応を得た結果である。

表 2.23 ある薬理作用に対する用量反応実験データ [芳賀 (2014), 表示 2.2.1]

	用量 <i>dose</i>	繰り返し						
A	mg/kg	1	2	3	4	5	平均	標準偏差
A ₁	0	10.5	9.6	10.4	10.2	9.4	10.020	0.4919
A ₂	10	10.8	10.7	11.1	10.9	11.0	10.900	0.1581
A ₃	20	11.4	10.7	10.9	11.3	11.7	11.200	0.4000
A ₄	30	11.9	11.2	11.0	11.1	11.3	11.300	0.3536
	全体						10.855	

この実験の目的は、どの程度の用量から反応が対照群に対して統計的な差 (有意な差) があるのかを確認するのが主たる目的である。A₂ の 10 mg/kg 群の反応は、A₁ の 0 mg/kg 対照群の最大値である 10.5 よりも (10.7~11.1) と全て大きく、統計的な方法によらずとも明らかな反応の差があると判断される。さらに、用量反応関係が直線的なのか、何らかの曲線関係が示唆されるのかを統計的に調べたい。

デザイン行列 X を用いた多項式回帰

これまで示してきた質的因子 A の場合のデザイン行列 X と同様に、1 次式, 2 次式, ... をあてはめる回帰分析を効率よく行うためにデザイン行列 X を前もって設定する。表 2.24 右に示すように用量 *dose* は 2 桁なので、冪乗すると桁数のインフレーションが起きるので、10 で割って 1 桁とし、2 乗および 3 乗とした“多項式型ダミー変数”とする。

表 2.24 多項式回帰のためのダミー変数

	(0, 1) 型				dose	x	多項式型			
	a_2	a_3	a_4				x^0	x^1	x^2	x^3
A ₁	0	0	0	A ₁	0	0	1	0	0	0
A ₂	1	0	0	A ₂	10	1	1	1	1	1
A ₃	0	1	0	A ₃	20	2	1	2	4	8
A ₄	0	0	1	A ₄	30	3	1	3	9	27

3 次式のあてはめ

表 2.25 右上段に示すのは、3 次の多項式型ダミー変数を (20 行×4 列) のデザイン行列 X とした回帰分析の結果である。分散分析表の分散比 F 値は、12.2742 と十分に大きく 3 次式がよくあてはまっているとの結果であるが、パラメータ (係数) の変数 x^3 の p 値からは、統計的に有意ではないとの結果である。これは、低次の (x^0 , x^1 , x^2) の変数が回帰モデルにすでに含まれている条件下で、さらに、 x^3 を加えた時の x^3 自身の寄与分を表して、2 次式のあてはめに加えて 3 次式にする必要がないと解釈される。同様に 1 次の係数 $\hat{\beta}_1^{(3)} = 1.2967$ の場合は、(x^0 , x^2 , x^3) の変数が回帰モデルにすでに含まれている条件下でさらに x^1 を加えて時の x^1 自身の寄与分を表して、単純に 1 次式をあてはめた結果ではないことに注意が必要である。

表 2.25 デザイン行列を用いた 3 次式に対する回帰分析

A	dose		— デザイン行列 X —				$X\hat{\beta}^{\wedge}$						
	mg/kg	y	x^0	x^1	x^2	x^3		y [^]	分散分析表, 「定数に 0 使用」 off				
A ₁	0	10.5	1	0	0	0	10.02						
	0	9.6	1	0	0	0	10.02		自由度	変動	分散	分散比	
	0	10.4	1	0	0	0	10.02		回帰	3	5.0815	1.6938	12.2742
	0	10.2	1	0	0	0	10.02		残差	16	2.2080	0.1380	
	0	9.4	1	0	0	0	10.02		合計	19	7.2895		
A ₂	10	10.8	1	1	1	1	10.90						
	10	10.7	1	1	1	1	10.90	$\hat{\beta}_0^{(3)}$	切片 x^0	10.0200	0.1661	60.3133	0.0000
	10	11.1	1	1	1	1	10.90	$\hat{\beta}_1^{(3)}$	x^1	1.2967	0.6374	2.0342	0.0589
	10	10.9	1	1	1	1	10.90	$\hat{\beta}_2^{(3)}$	x^2	-0.4800	0.5634	-0.8520	0.4068
	10	11.0	1	1	1	1	10.90	$\hat{\beta}_3^{(3)}$	x^3	0.0633	0.1238	0.5115	0.6160
A ₃	20	11.4	1	2	4	8	11.20						
	20	10.7	1	2	4	8	11.20						
	20	10.9	1	2	4	8	11.20						
	20	11.3	1	2	4	8	11.20	$\hat{\beta}_0^{(3)}$	パラメータの共分散行列 $\Sigma(\hat{\beta}^{\wedge(3)}) = (X^T X)^{-1} \sigma^{\wedge 2}$				
	20	11.7	1	2	4	8	11.20	$\hat{\beta}_1^{(3)}$	0.0276	-0.0506	0.0276	-0.0046	
								$\hat{\beta}_2^{(3)}$	-0.0506	0.4063	-0.3450	0.0721	
								$\hat{\beta}_3^{(3)}$	0.0276	-0.3450	0.3174	-0.0690	
A ₄	30	11.9	1	3	9	27	11.30		$\hat{\beta}_0^{(3)}$	-0.0046	0.0721	-0.0690	0.0153
	30	11.2	1	3	9	27	11.30		x_0	x^1	x^2	x^3	
	30	11.0	1	3	9	27	11.30		=Minverse (Mmult (Transpose (X の範囲), X の範囲)) * $\sigma^{\wedge 2}$				
	30	11.1	1	3	9	27	11.30						
	30	11.3	1	3	9	27	11.30						

パラメータの推定値からは、3次式のあてはめに関して否定的な結果であったが、表 2.26 に示すように3次曲線のあてはめを例示する。3次式は、

$$\hat{y}^{(3)} = \hat{\beta}_0^{(3)}x^0 + \hat{\beta}_1^{(3)}x^1 + \hat{\beta}_2^{(3)}x^2 + \hat{\beta}_3^{(3)}x^3$$

$$= 10.0200x^0 + 1.2967x^1 - 0.4800x^2 + 0.0633x^3 \quad (2.39)$$

であり、その分散 $Var(\hat{y}^{(3)})$ は、これまで用いてきたの線形和の係数ベクトル \mathbf{l} を、 $[\mathbf{l}_0^{(3)} = x^0 = 1, \mathbf{l}_1^{(3)} = x^1, \mathbf{l}_2^{(3)} = x^2, \mathbf{l}_3^{(3)} = x^3]$ としたときに、これまでと同様にパラメータの共分散行列に間するベクトル \mathbf{l} の2次形式により、

$$Var(\hat{y}^{(3)}) = \mathbf{l}^{(3)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(3)}) (\mathbf{l}^{(3)})^T \quad (2.40)$$

として分散の推定ができる。

表 2.26 に示すように3次曲線を滑らかに描くために、用量 $dose$ を -10 から 5 刻みに 40 まで与え、式 (2.39) に代えて

$$\hat{y}^{(3)} = \mathbf{l}_0^{(3)} \hat{\beta}_0^{(3)} + \mathbf{l}_1^{(3)} \hat{\beta}_1^{(3)} + \mathbf{l}_2^{(3)} \hat{\beta}_2^{(3)} + \mathbf{l}_3^{(3)} \hat{\beta}_3^{(3)} \quad (2.41)$$

により推定し、分散 $Var(\hat{y}^{(3)})$ は、式 (2.40) によって求めている。推定値結果は、当然のことながら、それぞれの用量 $dose$ の平均値に完全にフィットしている。図 2.8 に示すように95%信頼区間は、6次式となり複雑な曲線となっている。統計的には、より少ないパラメータでの推定が、より安定した回帰式となることが知られている。この95%信頼区間から、得られたデータの範囲を超えた場合に上下に跳ね上がっていることから、オーバーフィットによる予測式の不安定さが現われている。

表 2.26 3次曲線をあてはめた予測プロファイル

dose mg/kg	l_0 x^0	l_1 x^1	l_2 x^2	l_3 x^3	推定値 \hat{y}	分散 $Var(\hat{y})$	幅 $t_{0.05} \times SE$	95%信頼区間 $L95\%$ $U95\%$	
-10	1	-1.0	1.0	-1.0	8.18	1.904	2.93	5.25	11.11
-5	1	-0.5	0.3	-0.1	9.24	0.314	1.19	8.06	10.43
0	1	0.0	0.0	0.0	10.02	0.028	0.35	9.67	10.37
5	1	0.5	0.3	0.1	10.56	0.030	0.37	10.19	10.92
10	1	1.0	1.0	1.0	10.90	0.028	0.35	10.55	11.25
15	1	1.5	2.3	3.4	11.10	0.018	0.28	10.82	11.38
20	1	2.0	4.0	8.0	11.20	0.028	0.35	10.85	11.55
25	1	2.5	6.3	15.6	11.25	0.030	0.37	10.89	11.62
30	1	3.0	9.0	27.0	11.30	0.028	0.35	10.95	11.65
35	1	3.5	12.3	42.9	11.39	0.314	1.19	10.21	12.58
40	1	4.0	16.0	64.0	11.58	1.904	2.93	8.65	14.51
					$t_{0.05}(20-4) = 2.1199$				

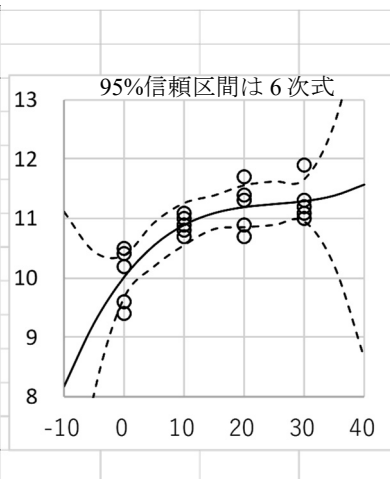


図 2.8 3次曲線のあてはめ

2次式のあてはめ

2次式のあてはめは、デザイン行列 \mathbf{X} の (x^0, x^1, x^2) について回帰分析を行なう。表 2.27 に示すように、2次式をあてはめた場合に、 x^2 変数の係数は、3次式のあてはめで p

表 2.27 デザイン行列を用いた 2 次式に対する回帰分析

dose		— デザイン行列 X —					$X\beta^{\wedge}$					
A	mg/kg	y	x^0	x^1	x^2	x^3	y^{\wedge}	分散分析表, 「定数に 0 使用」 off				
A ₁	0	10.5	1	0	0	0	10.04	自由度	変動	分散	分散比	
	0	9.6	1	0	0	0	10.04	回帰	2	5.0454	2.5227	19.1105
	0	10.4	1	0	0	0	10.04	残差	17	2.2441	0.1320	
	0	10.2	1	0	0	0	10.04	合計	19	7.2895		
	0	9.4	1	0	0	0	10.04					
A ₂	10	10.8	1	1	1	1	10.84	係数	標準誤差	t	P-値	
	10	10.7	1	1	1	1	10.84	$\beta^{\wedge}_0^{(2)}$ 切片 x^0	10.0390	0.1584	63.3895	0.0000
	10	11.1	1	1	1	1	10.84	$\beta^{\wedge}_1^{(2)}$ x^1	0.9990	0.2543	3.9280	0.0011
	10	10.9	1	1	1	1	10.84	$\beta^{\wedge}_2^{(2)}$ x^2	-0.1950	0.0812	-2.4002	0.0281
	10	11.0	1	1	1	1	10.84					
A ₃	20	11.4	1	2	4	8	11.26	パラメータの共分散行列 $\Sigma(\beta^{\wedge(2)}) = (X^T X)^{-1} \sigma^{\wedge 2}$				
	20	10.7	1	2	4	8	11.26	$\beta^{\wedge}_0^{(2)}$	0.0251	-0.0277	0.0066	
	20	10.9	1	2	4	8	11.26	$\beta^{\wedge}_1^{(2)}$	-0.0277	0.0647	-0.0198	
	20	11.3	1	2	4	8	11.26	$\beta^{\wedge}_2^{(2)}$	0.0066	-0.0198	0.0066	
	20	11.7	1	2	4	8	11.26	x_0	x^1	x^2		
A ₄	30	11.9	1	3	9	27	11.28	=Minverse (Mmult (Transpose (X の範囲), X の範囲))* $\sigma^{\wedge 2}$				
	30	11.2	1	3	9	27	11.28					
	30	11.0	1	3	9	27	11.28					
	30	11.1	1	3	9	27	11.28					
	30	11.3	1	3	9	27	11.28					

=0.4068 と有意でなかったが, $\hat{\beta}_2^{(2)} = -0.1950$, $p = 0.0281$ と有意な差に変化する. この結果から, 1 次式ではなく 2 次式のあてはめが支持されることになる.

表 2.28 に推定された 2 次式の回帰パラメータ $\hat{\beta}^{(2)}$ およびパラメータの共分散行列 $\Sigma(\hat{\beta}^{(2)})$ を用いて計算した推定値 \hat{y} および 95%信頼区間を散布図上に重ね書きした予測プロファイル

表 2.28 2 次曲線をあてはめた予測プロファイル

dose	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間		
mg/kg	x^0	x^1	x^2	x^3	y^{\wedge}	$Var(y^{\wedge})$	$t_{0.05} \times SE$	L95%	U95%	
-10	1	-1.0	1.0	-1.0	8.85	0.205	0.95	7.89	9.80	
-5	1	-0.5	0.3	-0.1	9.49	0.078	0.59	8.90	10.08	
0	1	0.0	0.0	0.0	10.04	0.025	0.33	9.70	10.37	
5	1	0.5	0.3	0.1	10.49	0.012	0.23	10.26	10.72	
10	1	1.0	1.0	1.0	10.84	0.015	0.25	10.59	11.10	
15	1	1.5	2.3	3.4	11.10	0.017	0.27	10.82	11.37	
20	1	2.0	4.0	8.0	11.26	0.015	0.25	11.00	11.51	
25	1	2.5	6.3	15.6	11.32	0.012	0.23	11.08	11.55	
30	1	3.0	9.0	27.0	11.28	0.025	0.33	10.95	11.62	
35	1	3.5	12.3	42.9	11.15	0.078	0.59	10.56	11.73	
40	1	4.0	16.0	64.0	10.92	0.205	0.95	9.96	11.87	
							$t_{0.05}(20-3) =$	2.1098		

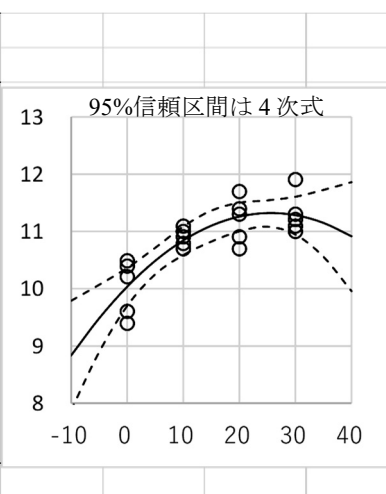


図 2.9 2 次曲線のあてはめ

を示す. 統計ソフトでは, 2次式に限らず多項式回帰の95%信頼区間を図示できるものもあるが, その計算方法はブラックボックス的である.

多くの統計入門書で, 1次式の回帰直線の場合の95%信頼区間の計算式は, 平方和を用いた式で示されているが, 2次式以上の回帰曲線の95%信頼区間の計算式を見いだすことは不可能に近い. この理由は, 1次式の回帰直線のパラメータ推定に際し, 平方和に基づく計算式が示されているが, 2次式への拡張性が全くない式となっているためである. デザイン行列 \mathbf{X} に基づいた計算法では, これまでも示してきたように, 線形和 $L = \mathbf{l}\hat{\boldsymbol{\beta}}$ の係数ベクトル \mathbf{l} を用いた2次形式

$$\text{Var}(\mathbf{l}\hat{\boldsymbol{\beta}}) = \text{Var}(\hat{y}) = \mathbf{l} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{l}^T \quad (2.42)$$

によって, 推定値 \hat{y} の分散 $\text{Var}(\hat{y})$ を計算し

$$95\%CL(\hat{y}) = \hat{y} \pm t_{0.05}(df_e) \sqrt{\text{Var}(\hat{y})} \quad (2.43)$$

によって95%信頼区間の計算をしてきた. ここでは, 2次式の推定値の分散 $\text{Var}(\hat{y})$ の計算過程を詳しく示す.

外挿となるが, $dose = 40 \text{ mg/kg}$ の場合を示す. 表 2.28 の係数ベクトル $\mathbf{l}_{dose=40}^{(2)}$ は,

$$\mathbf{l}_{dose=40}^{(2)} = [1 \quad 4 \quad 16]$$

である. 推定値 $\hat{y}_{dose=40}^{(2)}$ は,

$$\begin{aligned} \hat{y}_{dose=40}^{(2)} &= \mathbf{l}_{dose=40}^{(2)} \hat{\boldsymbol{\beta}}^{(2)} \\ &= [1 \quad 4 \quad 16] \begin{bmatrix} 10.0390 \\ 0.9990 \\ -0.1950 \end{bmatrix} \\ &= 1 \times 10.0390 + 4 \times 0.9990 + 16 \times (-0.1950) = 10.92 \end{aligned}$$

であり, その分散は,

$$\text{Var}(\hat{y}_{dose=40}^{(2)}) = \mathbf{l}_{dose=40}^{(2)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(2)}) (\mathbf{l}_{dose=40}^{(2)})^T$$

=	1	4	16	0.0251	-0.0277	0.0066	1
				-0.0277	0.0647	-0.0198	4
				0.0066	-0.0198	0.0066	16
			=	0.0198	-0.0858	0.0330	1
							4
							16
			=	0.2046			

として計算されている.

シグマでの計算式では, パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{(2)})$ の要素を $c_{ii'}$ とし, $\mathbf{l}_{dose=40}^{(2)}$ の係数を $l_i^{(2)}$ とすると

$$\text{Var}(\hat{y}^{(2)}) = \sum_{i=0}^2 \sum_{i'=0}^2 [l_i^{(2)} l_{i'}^{(2)} c_{ii'}] \quad (2.44)$$

となる。この2次形式の式からも明らかのように、95%信頼区間の計算式は、4次式となり、これを四則演算式で示すこと自体が難儀なことであるし、実際に計算することも煩雑である。

表 2.28 で示した Excel の計算シートでの計算式は、1 行目で

$$\begin{aligned} \text{Var}(I^{(2)} \hat{\beta}^{(2)}) &= \text{Mmult}(\text{Mmult}(I_{\text{dose}=40}^{(2)} \text{ の範囲}, \Sigma(\hat{\beta}^{(2)}) \text{ の範囲}), \text{Transpose}(I_{\text{dose}=40}^{(2)} \text{ の範囲})) = 0.205 \\ t_{0.05} \times SE &= \text{T.inv.2T}(0.05, 20-3) * \text{Sqrt}(\text{Var}(I^{(2)} \hat{\beta}^{(2)})) = 0.95 \\ L95\% &= \text{推定値} - (t_{0.05} \times SE) = 7.89, \quad U95\% = \text{推定値} + (t_{0.05} \times SE) = 9.80 \end{aligned}$$

として計算し、2 行目以後はフィルハンドルで計算式をコピーして求めている。これらの計算式は、汎用的であり線形モデルに限らずロジスティック回帰でも、ポアソン回帰でも共通である。

デザイン行列 X を用いた単回帰分析

一般的に、回帰分析というと直線をあてはめる単変数の場合が想定され、2 変数以上の場合に重回帰分析と区別されるが、本書では全て回帰分析として区別しないで使っている。ダミー変数を用いた場合に線形モデルとし、回帰モデルを区別しているが、どちらもデザイン行列 X を用いた解析方法としては、同一である。

表 2.29 にデザイン行列 X を用いた単回帰分析の結果を示す。分散分析表および回帰パラメータの推定は、Excel による回帰分析の結果であるが、その下の 2×2 のパラメータの共分散行列 $\Sigma(\hat{\beta}^{(1)})$ は、切片を含めた (20 行 \times 2 列) のデザイン行列 X から計算されている。

表 2.29 デザイン行列を用いた単回帰分析

dose		— デザイン行列 X —					$X\beta^{\wedge}$					
A	mg/kg	y	x^0	x^1	x^2	x^3	y^{\wedge}	分散分析表, 「定数に 0 使用」 off				
A ₁	0	10.5	1	0	0	0	10.04	自由度	変動	分散	分散比	
:								回帰	1	4.2849	4.2849	25.6700
	0	9.4	1	0	0	0	10.04	残差	18	3.0046	0.1669	
A ₂	10	10.8	1	1	1	1	10.84	合計	19	7.2895		
:												
	10	11.0	1	1	1	1	10.84	係数	標準誤差	t	P-値	
A ₃	20	11.4	1	2	4	8	11.26	$\beta^{\wedge}_0^{(1)}$ 切片 x^0	10.2340	0.1529	66.9460	0.0000
:								$\beta^{\wedge}_1^{(1)}$ x^1	0.4140	0.0817	5.0666	0.0001
	20	11.7	1	2	4	8	11.26					
A ₄	30	11.9	1	3	9	27	11.28	パラメータの共分散行列 $\Sigma(\hat{\beta}^{(1)}) = (X^T X)^{-1} \sigma^{\wedge 2}$				
:								$\beta^{\wedge}_0^{(1)}$	0.0234	-0.0100		
	30	11.3	1	3	9	27	11.28	$\beta^{\wedge}_1^{(1)}$	-0.0100	0.0067		
								x_0	x^1			
								=Minverse (Mmult (Transpose (X の範囲), X の範囲)) * $\sigma^{\wedge 2}$				

パラメータの共分散行列 $\Sigma(\hat{\beta}^{(1)})$ の活用は、平方和を用いる 95%信頼区間の計算方法からの脱却を図ることができる。Excel の単回帰分析を使っている人たちは、どのようにしたら 95%信頼区間を計算し、図示したいとの思いがあり、その際に頼りにするのは、平方和 S_{xx} などを用いた計算公式であり、元データを用いてシグマを用いた計算が別途必要となる。第5章で詳しく示すが、伝統的な計算公式は、過度に標準化されており、2次式に対する計算式への拡張のヒントすら示されないのが最大の問題点である。まさに、応用力を発揮できないようにするための「ガラスの天井」がはめ込まれているがごとくである。

表 2.30 に示すように、回帰直線の 95%信頼区間を計算するために必要な推定値 \hat{y} の分散 $Var(\hat{y}^{(1)})$ は、

$$\begin{aligned} Var(\hat{y}^{(1)}) &= l^{(1)} \Sigma(\hat{\beta}^{(1)}) l^{(1)T} \\ &= [x^0 \ x^1] \begin{bmatrix} Ver(\beta_0^{(1)}) & Cov(\beta_0^{(1)}, \beta_1^{(1)}) \\ Cov(\beta_0^{(1)}, \beta_1^{(1)}) & Ver(\beta_1^{(1)}) \end{bmatrix} \begin{bmatrix} x^0 \\ x^1 \end{bmatrix} \\ &= (x^0)^2 Ver(\beta_0^{(1)}) + 2x^0 x^1 Cov(\beta_0^{(1)}, \beta_1^{(1)}) + (x^1)^2 Ver(\beta_1^{(1)}) \\ &= Ver(\beta_0^{(1)}) + 2x Cov(\beta_0^{(1)}, \beta_1^{(1)}) + x^2 Ver(\beta_1^{(1)}) \end{aligned} \quad (2.45)$$

のように合成分散の一般式で表すこともでき、 $l=[1 \ -1.0]$ の場合であれば、

$$\begin{aligned} Var(\hat{y}^{(1)}) &= l^{(1)} \Sigma(\hat{\beta}^{(1)}) l^{(1)T} \\ &= 0.0234 + 2 \times (-1) \times (-0.0100) + (-1)^2 \times 0.0067 = 0.0501 \end{aligned} \quad (2.46)$$

のように計算されている。この計算式が示されている成書を見いだすことができるが、偏差平方和 S_{xx} を用いた式へ誘導されるのが常である。平方和の分解と分散の加法性による解析法に対する先人たちの“こだわり”にほとんどの人達がかからめとられている。

表 2.30 に示すように、パラメータの共分散 $Cov(\hat{\beta}_0, \hat{\beta}_1)$ を素直に使った計算により、回帰直線の分散を計算し、95%信頼区間を重ね書きした予測プロファイルを図 2.10 に示す。2次

表 2.30 単回帰式の回帰直線と 95%信頼区間の予測プロファイル

dose mg/kg	l_0 x^0	l_1 x^1	l_2 x^2	l_3 x^3	推定値 \hat{y}	分散 $Var(\hat{y})$	幅 $t_{0.05} \times SE$	95%信頼区間 L95% U95%	
-10	1	-1.0	1.0	-1.0	9.82	0.0501	0.47	9.35	10.29
-5	1	-0.5	0.3	-0.1	10.03	0.035	0.39	9.63	10.42
0	1	0.0	0.0	0.0	10.23	0.023	0.32	9.91	10.56
5	1	0.5	0.3	0.1	10.44	0.015	0.26	10.18	10.70
10	1	1.0	1.0	1.0	10.65	0.010	0.21	10.44	10.86
15	1	1.5	2.3	3.4	10.86	0.008	0.19	10.66	11.05
20	1	2.0	4.0	8.0	11.06	0.010	0.21	10.85	11.27
25	1	2.5	6.3	15.6	11.27	0.015	0.26	11.01	11.53
30	1	3.0	9.0	27.0	11.48	0.023	0.32	11.15	11.80
35	1	3.5	12.3	42.9	11.68	0.035	0.39	11.29	12.08
40	1	4.0	16.0	64.0	11.89	0.050	0.47	11.42	12.36
					$t_{0.05}(20-2) = 1.9727$				

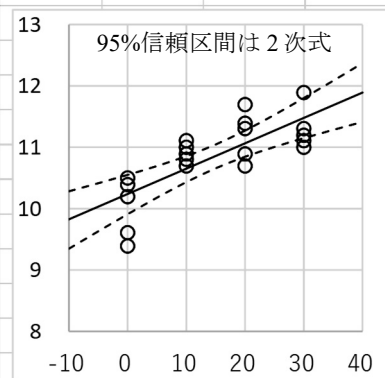


図 2.10 1次式のあてはめ

式のあてはまりの検討をしなければ、少々あてはまりは悪いようにも見えるが、直線があてはまっていると結論したくなるような予測プロファイルとなっている。

あてはまりの悪さ LOF 解析

量的変数に対する 1 因子実験データに多項式をあてはめ、次数を順次減らすことにより、2 次式のあてはめが適していることを示した。平方和の分解による分散分析表の作成に引き続き、回帰直線をあてはめて分散分析表を作成し、有意な差が得られたのであるが、散布図に直線を引いてみると 2 次曲線らしき形状があり気になる。統計的に曲線の「あてはまり」は統計的に良くないと言えれば、直線なあてはめが正当化される。このような方法は、LOF (Lack OF Fit) 解析として知られている。

表 2.25 に示した 3 次式のあてはめの分散分析表は、質的変数 A としての分散分析表に一致する。表 2.29 に示した 1 次式のあてはめの分散分析表は、回帰分析そのものである。表 2.31 に示すように質的変数 A としての自由度 3 の平方和は、 $S_A = 5.0815$ であり、回帰の平方和は、 $S_{\text{回帰}}^{(1)} = 4.2849$ である。あてはまりの悪さ LOF の平方和は、

$$\begin{aligned} S_{\text{LOF}} &= S_A - S_{\text{回帰}}^{(1)} \\ &= 5.0815 - 4.2849 = 0.7966 \end{aligned} \quad (2.47)$$

として計算される。

表 2.31 質的変数 A としての分散分析表および回帰分析の分散分析表

要因	df	平方和	平均平方	F 値	p 値	要因	df	平方和	平均平方	F 値	p 値
A	3	5.0815	1.6938	12.27	0.0001	Reg	1	4.2849	4.2849	25.67	0.0001
e	16	2.2080	0.1380			e	18	3.0046	0.1669		
T	19	7.2895				T	19	7.2895			

表 2.32 に示すのは、これらの分散分析表を統合して LOF 解析のための分散分析表である。 $S_{\text{LOF}} = 0.7966$ なので、その自由度 2 で割った平均平方を、さらに誤差分散の推定値 $\hat{\sigma}^2 = 0.1380$ で割って求めた $F = 2.8862$ から $p = 0.0851$ が算出される。微妙な大きさの p 値であるが、これを無視すれば、統計的に LOF が有意でないので、回帰直線なあてはまりに問題

表 2.32 LOF 解析のための分散分析表

要因	df	平方和	平均平方	F 値	p 値
A	3	5.0815	1.6938	12.2742	0.0001
Reg	1	4.2849	4.2849	31.0500	0.0000
LOF	2	0.7966	0.3983	2.8862	0.0851
e	16	2.2080	0.1380		
T	19	7.2895			

がないと言うことも可能である。別の見方をすれば、2次式をあてはめ、更なるLOF解析を行ってみる必要性もある。

LOF（あてはまりの悪さ）解析は、直線をあてはめていいのか、それで十分なのか、などの疑問に対する解析法であるが、込みいった解析法であるので、これまでの結果を整理し理解を深めたい。表 2.29 に示すように、1次式をあてはめた場合に回帰係数 $\hat{\beta}_1^{(1)} = 0.4140$ ($p < 0.001$) である。2次式をあてはめた場合に2次の項 $\hat{\beta}_2^{(2)} = -0.1950$ ($p < 0.0281$) と有意となっている。3次式をあてはめた場合に3次の項 $\hat{\beta}_3^{(3)} = 0.0633$ (NS) となるので、2次式をあてはめが支持される。

LOF 解析に代わる逐次平方和（タイプ I の平方和）

LOF 解析は、量的変数に対する1因子の分散分析表に回帰分析から得られた回帰の平方和を組み込む方法として知られているが、表 2.24 に示したように多項式型ダミー変数をデザイン行列 \mathbf{X} とした用いた場合には、表 2.33 に示すように、1次、2次、3次、... のように回帰分析を行ない、得られた回帰パラメータの p 値の変化を評価することにより平方和を主体にした LOF 解析の代わりとなる。

表 2.33 回帰パラメータに関する有意差検定

	1次式			2次式			3次式		
	係数	p		係数	p		係数	p	
x^0	10.2340	0.0000		10.0390	0.0000		10.0200	0.0000	
x^1	0.4140	0.0001	***	0.9990	0.0011	**	1.2967	0.0589	
x^2				-0.1950	0.0281	*	-0.4800	0.4068	
x^3							0.0633	0.6160	NS

Excel の回帰分析によるパラメータの p 値は、Type III の平方和による p 値に一致する。

JMP の「モデルのあてはめ」で表 2.34 に示すように3次式のあてはめを行い、「逐次検定」を選択することにより、多項式回帰の次数を一目で判定できる。

表 2.34 JMP の「モデルのあてはめ」による3次式の逐次（タイプ 1）の平方和

逐次(タイプ1)検定					
要因	自由度	逐次平方和	F値	p値(Prob>F)	Type III p値
x	1	4.2849	31.0500	<.0001*	0.0589
x*x	1	0.7605	5.5109	0.0321*	0.4068
x*x*x	1	0.0361	0.2616	0.6160	0.6160

$x*x$ の $p=0.0321$ は、自由度 16 の $\hat{\sigma}^2 = 0.1380$ を用いているが、

表 2.33 の「2次式」の $p=0.0281$ は、自由度 17 の $\hat{\sigma}^2 = 0.1320$ を用いている。

3. 繰り返しが不揃いの 2 因子実験データの解析

繰り返しが不揃いの 2 因子実験データの解析に対し、平方和の分解による伝統的な方法では、分散分析表の作成ができない。ただし、質的変数をダミー変数に展開し、線形モデルを適用することにより分散分析が可能となり実用化できることは、40 年以上前から線形モデルを活用した SAS の GLM プロシジャによって明らかにされていた。だが、知る人ぞ知るがごとく、その存在は、日の目を見ることなく静かに時が流れてきた。観察データの解析において、2 つの層別因子に着目した場合に、それらの組み合わせたセルのデータ数が等しくなることはまったく期待できない。そこで、繰り返しが等しい 2 因子実験データに対する平方和の分解による解析法と対比できるように、線形モデルによる解析方法を導入する。次に、繰り返しが不揃いの 2 因子実験データでは、平方和の分解による分散分析表ができないことを明らかにし、線形モデルによる解析方法を Excel の回帰分析および行列関数を用いて平易に説明する。

3.1. 繰り返しが等しい 2 因子実験

高橋・大橋・芳賀 (1989), 「SAS による実験データの解析」の「第 15 章 4 種の平方和と LSMEANS」の図表 15.1 で例示されているデータを用い、伝統的な平方和の分解による分散分析を例示する。さらに、同じデータを用いて (1, -1) 対比型ダミー変数を用い、Excel の回帰分析を適用した結果を示し、平方和の分解による方法と対比する。

平方和の分解による分散分析表の作成

表 3.1 に示すのは、完全ランダム化により得られた繰り返しが 2 の 2 因子実験データである。これらのデータおよびセル平均を含めた散布図を、図 3.1 に示す。因子 A の A₂ 水準と

表 3.1 繰り返しが等しい 2 因子実験データ [高橋ら (1989), 図表 15.1]

	B ₁			B ₂			B ₃			y の
	y		平均	y		平均	y		平均	平均
A ₁	10	13	11.5	14	12	13.0	22	19	20.5	15.0
A ₂	15	14	14.5	16	18	17.0	21	18	19.5	17.0
平均	13.0			15.0			20.0			16.0

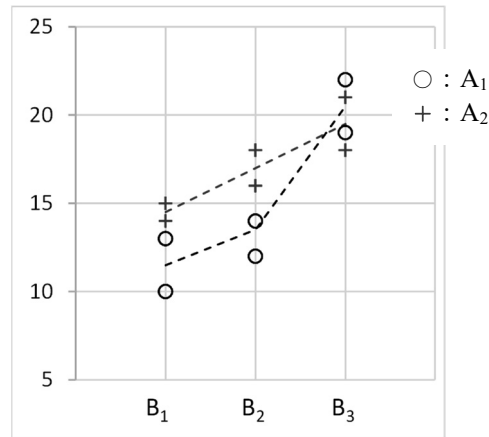


図 3.1 散布図に 2 因子実験の平均値の重ね書き

A₁ 水準の差に注目すると、B₁ 水準と B₂ 水準で 3~4 単位程度の差があるが、B₃ 水準で A₁ 水準と A₂ 水準がわずかに逆転している。

データの構造式は、

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (3.1)$$

であり、主効果については、

$$\sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0 \quad (3.2)$$

交互作用については、

$$\left. \begin{aligned} \sum_i (\alpha\beta)_{ij} &= 0, \quad i=1, 2 \\ \sum_j (\alpha\beta)_{ij} &= 0, \quad j=1, 2, 3 \end{aligned} \right\} \quad (3.3)$$

のような制約条件が付く。

総平方和 S_T は、表 3.2 に示すように、各 y_{ijk} から総平均 $\bar{y}_{...}$ を引いた平方和であり、

$$\bar{y}_{...} = \left(\sum_i \sum_j \sum_k y_{ijk} \right) / N = 16.0, \quad N = 12 : \text{データ総数} \quad (3.4)$$

$$S_T = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 = 148.0 \quad (3.5)$$

として求められる。

因子 A の平方和 S_A を計算する場合、A₁ 水準の平均を $\bar{y}_{1..}$ 、A₂ 水準の平均を $\bar{y}_{2..}$ としたときに、 α_i の推定値を $\hat{\alpha}_i = (\bar{y}_{i..} - \bar{y}_{...})$ とすると因子 A の平方和 S_A は、

$$\bar{y}_{A_i} : \bar{y}_{i..} = \left(\sum_j \sum_k y_{ijk} \right) / (n_B \cdot n_K), \quad n_B = 3, n_K = 2 : \text{繰り返し数} \quad (3.6)$$

$$S_A = \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}_{...})^2 = 12.0 \quad (3.7)$$

によって求められる。計算の冗長性をなくすために

$$S_A = n_B n_K \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2, \quad n_B = 3, n_K = 2 \quad (3.8)$$

表 3.2 Excel の SumSq () 関数を活用した平方和の計算

			総平均		主効果 A		主効果 B		セル平均	期待平均	交互作用	残差
			μ^{\wedge}		α^{\wedge}		β^{\wedge}				$(\alpha\beta)^{\wedge}$	ε^{\wedge}
A	B	y_{ijk}	$\bar{y}_{...}$	$y_{ijk} - \bar{y}_{...}$	\bar{y}_A	$\bar{y}_A - \bar{y}_{...}$	\bar{y}_B	$\bar{y}_B - \bar{y}_{...}$	\bar{y}_{AB}	\hat{y}'_{AB}	$\bar{y}_{AB} - \hat{y}'_{AB}$	$y_{ijk} - \bar{y}_{AB}$
A ₁	B ₁	10	16.0	-6.0	15.0	-1.0	13.0	-3.0	11.5	12.0	-0.5	-1.5
		13	16.0	-3.0	15.0	-1.0	13.0	-3.0	11.5	12.0	-0.5	1.5
	B ₂	14	16.0	-2.0	15.0	-1.0	15.0	-1.0	13.0	14.0	-1.0	1.0
		12	16.0	-4.0	15.0	-1.0	15.0	-1.0	13.0	14.0	-1.0	-1.0
	B ₃	22	16.0	6.0	15.0	-1.0	20.0	4.0	20.5	19.0	1.5	1.5
		19	16.0	3.0	15.0	-1.0	20.0	4.0	20.5	19.0	1.5	-1.5
A ₂	B ₁	15	16.0	-1.0	17.0	1.0	13.0	-3.0	14.5	14.0	0.5	0.5
		14	16.0	-2.0	17.0	1.0	13.0	-3.0	14.5	14.0	0.5	-0.5
	B ₂	16	16.0	0.0	17.0	1.0	15.0	-1.0	17.0	16.0	1.0	-1.0
		18	16.0	2.0	17.0	1.0	15.0	-1.0	17.0	16.0	1.0	1.0
	B ₃	21	16.0	5.0	17.0	1.0	20.0	4.0	19.5	21.0	-1.5	1.5
		18	16.0	2.0	17.0	1.0	20.0	4.0	19.5	21.0	-1.5	-1.5
平方和				148.0		12.0		104.0			14.0	18.0
			S_T		S_A		S_B				$S_{A \times B}$	S_e
$S_T = S_A + S_B + S_{A \times B} + S_e =$											148.0	

としたいくなるが、Excel の SumSq () 関数を用いる場合は、式 (3.7) による計算の方が簡便である。因子 B の平方和 S_B も同様に、 β_j の推定値を $\hat{\beta}_j = (\bar{y}_{.j} - \bar{y}_{...})$ として

$$\bar{y}_{B_j} : \bar{y}_{.j} = (\sum_i \sum_k y_{ijk}) / (n_A \cdot n_k), \quad n_A = 2, n_k = 2 : \text{繰り返し数} \quad (3.9)$$

$$S_B = \sum_i \sum_j \sum_k (\bar{y}_{.j} - \bar{y}_{...})^2 = 104.0 \quad (3.10)$$

で求める。

交互作用 A×B の平方和 $S_{A \times B}$ は、因子 A と因子 B の組み合わせ (セル) 平均 $\bar{y}_{ij.}$ を、

$$\bar{y}_{A_i B_j} : \bar{y}_{ij.} = (\sum_k y_{ijk}) / n_k \quad (3.11)$$

で求め、交互作用がないとした場合の期待平均 $\hat{y}'_{ij.}$ の差から求める。なお、期待平均 $\hat{y}'_{ij.}$ は、交互作用がないと仮定した場合の効果の和 $(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)$ に対応し、

$$\left. \begin{aligned} \hat{y}'_{ij.} &= \bar{y}_{...} + (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) \\ &= \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...} \end{aligned} \right\} \quad (3.12)$$

によって推定することができる。交互作用 $S_{A \times B}$ は、セル平均 $\bar{y}_{ij.}$ と対応する期待平均 $\hat{y}'_{ij.}$ の差の平方和から

$$S_{A \times B} = \sum_i \sum_j \sum_k (\bar{y}_{ij.} - \hat{y}'_{ij.})^2 = 14.0 \quad (3.13)$$

として求められる。残差平方和 S_e は、各 y_{ijk} とセル平均 $\bar{y}_{ij.}$ との残差 $\hat{\varepsilon}_{ijk} = (y_{ijk} - \bar{y}_{ij.})$ の平方和から

$$S_e = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 = 18.0 \quad (3.14)$$

によって求められる。

表 3.2 で示したような繰り返しが等しい場合には、総平方和 S_T が、別々に計算された平方和の和

$$\left. \begin{aligned} S_T &= S_A + S_B + S_{A \times B} + S_e \\ &= 12.0 + 104.0 + 14.0 + 18.0 = 148.0 \end{aligned} \right\} \quad (3.15)$$

と等しくなる。総平方和 S_T がそれぞれ別々に求めた平方和の和となることから、「平方和の分解」と称されている。

式 (3.1) で示したデータの構造式が、それぞれの効果の和となっているか確認する。表 3.2 の最初の行に対するデータの構造式は、

$$\left. \begin{aligned} y_{111} &= \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 + (\widehat{\alpha\beta})_{11} + \hat{\varepsilon}_{111} \\ &= 16.0 + (-1.0) + (-3.0) + (-0.5) + (-1.5) = 10.0 \end{aligned} \right\} \quad (3.16)$$

と計算され、 $y_{111} = 10.0$ に一致する。

修正項 CT を用いた平方和の計算

Excel を用いて、平方和の定義に沿った計算方法を示したのであるが、表 3.3 に示すように多くの実験計画法の成書では、手計算を前提とした修正項 CT (*Correction Term*) を用いた計算式

$$\left. \begin{aligned} S_T &= \left(\sum_i \sum_j \sum_k y_{ijk}^2 \right) - CT \\ &= 3220 - \frac{192^2}{12} \\ &= 3220 - 3072 = 148.0 \end{aligned} \right\} \quad (3.17)$$

ただし、 $CT = T_{...}^2 / N$ 、 $T_{...}$: データの総合計

が多用されている。

因子 A、因子 B の平方和は、

$$\left. \begin{aligned} S_A &= \frac{\sum_i \left(\sum_j \sum_k y_{ijk} \right)^2}{n_B n_K} - CT \\ &= \frac{8100 + 10404}{3 \times 2} - 3072 = 12.0 \end{aligned} \right\} \quad (3.18)$$

$$\left. \begin{aligned} S_B &= \frac{\sum_j \left(\sum_i \sum_k y_{ijk} \right)^2}{n_A n_K} - CT \\ &= \frac{2704 + 3600 + 6400}{2 \times 2} - 3072 = 104.0 \end{aligned} \right\} \quad (3.19)$$

で求める。交互作用 $S_{A \times B}$ の計算のために S_{AB}

$$\left. \begin{aligned} S_{AB} &= \frac{\sum_j \sum_i \left(\sum_k y_{ijk} \right)^2}{n_K} - CT \\ &= \frac{529 + 676 + 1681 + 841 + 1156 + 1521}{2} - 3072 = 130.0 \end{aligned} \right\} \quad (3.20)$$

を求め、 $S_T = S_{AB} + S_e$ の関係から S_e は、

$$S_e = S_T - S_{AB} = 148.0 - 130.0 = 18.0 \tag{3.21}$$

として求められる。さらに、 $S_{AB} = S_A + S_B + S_{A \times B}$ の関係から、交互作用 $S_{A \times B}$ が

$$S_{A \times B} = S_{AB} - S_A - S_B = 130.0 - 12.0 - 104.0 = 14.0 \tag{3.22}$$

として求められる。これらの平方和の計算から表 3.4 に示す分散分析表を作成する。

表 3.3 修正項 CT を用いた平方和の計算表

		y_{ijk}	y_{ijk}^2	T_A	T_A^2	T_B	T_B^2	T_{AB}	T_{AB}^2				
A ₁	B ₁	10	100	90	8100	52	2704	23	529				
		13	169										
	B ₂	14	196			60	3600			26	676		
		12	144										
	B ₃	22	484			80	6400			41	1681		
		19	361										
A ₂	B ₁	15	225	102	10404			29	841				
		14	196										
	B ₂	16	256							34	1156		
		18	324										
	B ₃	21	441							39	1521		
		18	324										
合計 $T_{...} =$		192	3220		18504		12704		6404				
				$n_{\cdot ik} = 6$		$n_{i \cdot k} = 4$		$n_{\dots k} = 2$					
	$N =$	12	148.0		12.0		104.0		130.0	14.0	18.0		
	$CT =$	3072.00	S_T		S_A		S_B		S_{AB}	$S_{A \times B}$	S_e		

表 3.4 繰り返しが等しい場合の分散分析表

要因	自由度	平方和	平均平方	F値	p値
A	1	12.00	12.00	4.0000	0.0924
B	2	104.00	52.00	17.3333	0.0032
A×B	2	14.00	7.00	2.3333	0.1780
残差	6	18.00	3.00		
全体	11	148.00			

さて、「交互作用とはどのようなものなのか？」と問われたときにこの計算手順の説明をしても、謎が深まるだけで本質的な理解は得られない。さらに、折れ線グラフ上で交互作用 $S_{A \times B}$ の説明を尽くしても、元々が式(3.22)に示したように平方和の差で求めているので、謎は深まるばかりである。表 3.2 に即して言えば、交互作用 $S_{A \times B}$ は、組み合わせたセル平均 $\bar{y}_{A_i B_j}$ と期待平均 $\hat{y}'_{A_i B_j}$ の差(ずれ)の平方和と直接的に説明することができる。さらに、抽象的な交互作用の説明ではなく、図 3.2 に示すように可視化する。実線で示したのが、式(3.12)で示した期待平均 $\hat{y}'_{A_i B_j}$ であり、点線で示したのがセル平均 $\bar{y}_{A_i B_j}$ であり期待平均 $\hat{y}_{A_i B_j}$ とセル平均 $\bar{y}_{A_i B_j}$ の差の平方和が交互作用の平方和 $S_{A \times B}$ となる。この方法の詳細は、表 3.2 の交互作用 $S_{A \times B}$ の計算例で示されている。

表 3.5 セル平均 \bar{y}_{A,B_j} と期待平均 \hat{y}'_{A,B_j}

		セル平均	期待平均	差	
		\bar{y}_{AB}	\hat{y}'_{AB}	$A \times B$	
		j			
A ₁	B ₁	1	11.5	12.0	-0.5
	B ₂	2	13.0	14.0	-1.0
	B ₃	3	20.5	19.0	1.5
A ₂	B ₁	1	14.5	14.0	0.5
	B ₂	2	17.0	16.0	1.0
	B ₃	3	19.5	21.0	-1.5
				平方和=	7.0
				$S_{A \times B} = 2 \times$ 平方和=	14.0

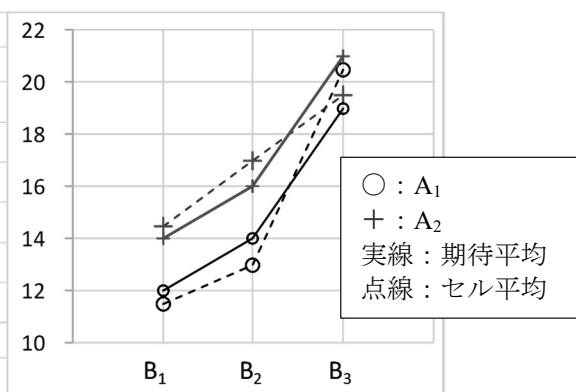


図 3.2 交互作用の図的表現

推定値に対する 95%信頼区間のグラフ表示

分散分析表の作成に引き続き、残差の平均平方（誤差分散の推定値 $\hat{\sigma}^2$ ）を用い、因子 A、因子 B、交互作用 $A \times B$ についての水準平均の 95%信頼区間、水準間の差の 95%信頼区間などを Excel シート上で計算した結果を表 3.6 に示す。

それぞれの要因の各水準の平均の分散は、それらの反復数を n_e としたときに $(\hat{\sigma}^2 / n_e)$ として計算される。A₁ 水準の平均は、6 個の y_{ijk}

$$\bar{y}_{1..} = \frac{y_{111} + y_{112} + y_{121} + y_{122} + y_{131} + y_{132}}{6} = 15.0 \tag{3.23}$$

の平均であり、それぞれの y_{ijk} の分散は、 $\hat{\sigma}^2 = 3.00$ なので、6 個の平均の分散は、

$$\left. \begin{aligned} Var(\bar{y}_{1..}) &= Var\left(\frac{y_{111}}{6} + \frac{y_{112}}{6} + \frac{y_{121}}{6} + \frac{y_{122}}{6} + \frac{y_{131}}{6} + \frac{y_{132}}{6}\right) \\ &= \frac{6\hat{\sigma}^2}{6^2} = \frac{3.00}{6} = 0.50 \end{aligned} \right\} \tag{3.24}$$

表 3.6 水準平均と 95%信頼区間

要因	平均	反復数 n_e	分散	SE	$t_{0.05} \times SE$	L95%	U96%
A ₁	15.0	6	0.500	0.707	1.730	13.270	16.730
A ₂	17.0	6	0.500	0.707	1.730	15.270	18.730
B ₁	13.0	4	0.750	0.866	2.119	10.881	15.119
B ₂	15.0	4	0.750	0.866	2.119	12.881	17.119
B ₃	20.0	4	0.750	0.866	2.119	17.881	22.119
A ₁ B ₁	11.5	2	1.500	1.225	2.997	8.503	14.497
A ₁ B ₂	13.0	2	1.500	1.225	2.997	10.003	15.997
A ₁ B ₃	20.5	2	1.500	1.225	2.997	17.503	23.497
A ₂ B ₁	14.5	2	1.500	1.225	2.997	11.503	17.497
A ₂ B ₂	17.0	2	1.500	1.225	2.997	14.003	19.997
A ₂ B ₃	19.5	2	1.500	1.225	2.997	16.503	22.497
			$\hat{\sigma}^2 =$	3.00	$t_{0.05}(6) =$	2.447	

として求められる。他の水準も同様に求めることができ、それらの水準平均の分散を計算し、誤差の自由度 6 の t 分布の 5% 点から、95% 信頼区間の幅 $t_{0.05} \times SE$ を計算し、下側 L95% および上側 U95% が求められている。

図 3.3 に示すのは、Excel の折れ線グラフを用いて、因子 A、因子 B、交互作用 A×B の水準平均に対して 95% 信頼区間を付けた結果である。作図は、表 3.6 の「平均」の列の 11 個を選び、基本の折れ線を描き、「ユーザ設定」の「信頼区間」で平均に対して「正の誤差」および「負の誤差」で 11 行分の $t_{0.05} \times SE$ の範囲を選択し、書式を整えた結果である。

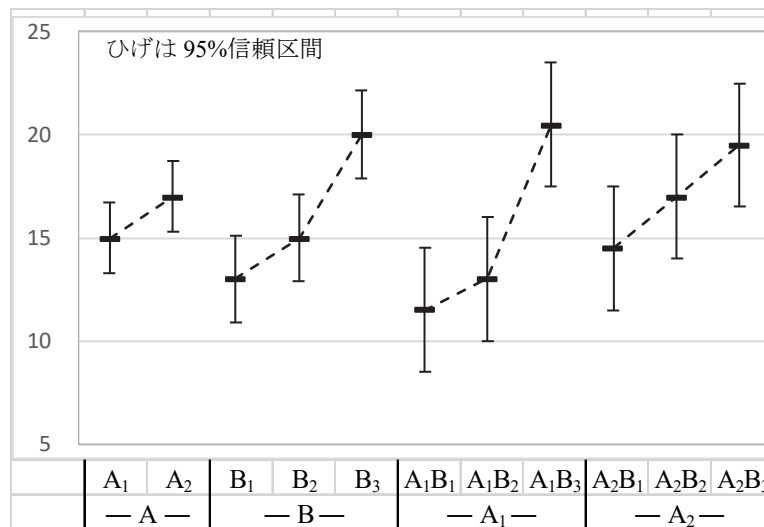


図 3.3 因子 A、因子 B、交互作用 A×B の水準平均と 95% 信頼区間

表 3.4 の分散分析表だけでは、結果の解釈が十分に行なえないので、図 3.3 に示すように水準平均と 95% 信頼区間を示すことにより解釈が容易になる。分散分析表で得られる最も有用な統計量は、残差の平均平方（誤差分散の推定値） $\hat{\sigma}^2 = 3.00$ であり、これを用いて、各水準の反復数 n_e を考慮することにより、表 3.6 に示したように各水準に対する 95% 信頼区間の計算が、フィルハンドルによる計算式のコピー機能により一括してできる。その結果を図 3.3 に示すことは、簡単とは言えないが若干の手間と工夫により作成できることは、うれしいかぎりである。

推定値の差に関する 95% 信頼区間のグラフ表示

図 3.3 に示した各種準に対する線グラフでは、2 水準間のおおよその統計的な判断が可能であるが、きちっとした統計量として表したい。そこで、表 3.7 に示すように、第 1 水準を基準とした差の推定値に対する 95% 信頼区間を求める。ここでも、水準間の差の分散を求めるための“反復数 n_e ”が活躍する。

表 3.7 第1水準との差と 95%信頼区間

要因	平均	n_e	分散	$t_{0.05} \times SE$	L95%	U96%
A_1-A_1	0.0	0	0.000	0.000	0.000	0.000
A_2-A_1	2.0	3	1.000	2.447	-0.447	4.447
B_1-B_1	0.0	0	0.000	0.000	0.000	0.000
B_2-B_1	2.0	2	1.500	2.997	-0.997	4.997
B_3-B_1	7.0	2	1.500	2.997	4.003	9.997
$A_1B_1-A_1B_1$	0.0	0	0.000	0.000	0.000	0.000
$A_1B_2-A_1B_1$	1.5	1	3.000	4.238	-2.738	5.738
$A_1B_3-A_1B_1$	9.0	1	3.000	4.238	4.762	13.238
$A_2B_1-A_2B_1$	0.0	0	0.000	0.000	0.000	0.000
$A_2B_2-A_2B_1$	2.5	1	3.000	4.238	-1.738	6.738
$A_2B_3-A_2B_1$	5.0	1	3.000	4.238	0.762	9.238
$\sigma^2 =$	3.00		$t_{0.05}(6) =$	2.447		

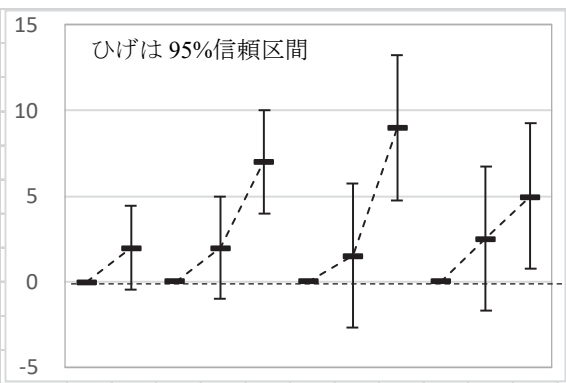


図 3.4 第1水準との差の 95%信頼区間

因子 A の第 2 水準 A_2 と第 1 水準 A_1 は、互いに独立なので、平均値の差に対する分散は、分散の加法性により

$$\left. \begin{aligned}
 Var(\bar{y}_{2..} - \bar{y}_{1..}) &= Var(\bar{y}_{2..}) + Var(\bar{y}_{1..}) \\
 &= \frac{\hat{\sigma}^2}{6} + \frac{\hat{\sigma}^2}{6} = \frac{3.00}{3} = 1.00
 \end{aligned} \right\} \quad (3.25)$$

が求められ、 $n_e = 3$ となる。他の水準間も同様に計算した結果が表 3.7 に示されている。表 3.7 の右の図 3.4 は、図 3.3 に示した水準平均の差の線グラフに 95%信頼区間の幅を付け、見栄えが良くなるように整えた結果である。

(1, -1)対比型ダミー変数

繰り返しが等しい 2 因子実験データの解析については、平方和の分解による解析手順が確立しており、ダミー変数を用いた線形モデルによる解析を行う必然性はない。しかし、繰り返しが不揃いとなった場合に、平方和の分解による解析が行えない。そのために、質的変数（因子）に対しダミー変数を用い、すべて量的変数に置き換えた線形モデルを適用する必要がある。そのために、繰り返しが等しい場合について、線形モデルの適用法を示すことにより、段階的な学習の助けになることを期待している。

性別（男性，女性）などを（0，1）に置き換えるダミー変数が広く使われている。ただし、この（0，1）型ダミー変数とした場合に、式 (3.1) に示したデータの構造式に対する制約条件として設定されている式 (3.2)，(3.3) を満足させることができない。そのためには、性別（男性，女性）などを（1，-1）に置き換える対比型ダミー変数とする必要がある。表 3.8 に示すのは、（1，-1）対比型ダミー変数である。交互作用は、主効果のダミー変数の積として設定する。

表 3.8 (1, -1) 対比型のダミー変数

	a_1		b_1	b_2			a_1b_1	a_1b_2
A ₁	1	B ₁	1	0	A ₁	B ₁	1	0
A ₂	-1	B ₂	0	1		B ₂	0	1
和	0	B ₃	-1	-1		B ₃	-1	-1
		和	0	0	A ₂	B ₁	-1	0
						B ₂	0	-1
						B ₃	1	1
						和	0	0

デザイン行列 X を用いた回帰分析の適用

表 3.9 左に示すのは、(1, -1) 対比型ダミー変数を用いた (12 行×6 列) のデザイン行列 X である。表 3.9 右に示すのは、Excel の回帰分析を適用し、モデル (回帰) 全体に対する分散分析表、その下に各変数 (ダミー変数) についてのパラメータ (係数) が推定されている。回帰分析では、一般的に変数を (x_1, x_2, \dots) とし、それらのパラメータ $(\beta_1, \beta_2, \dots)$ の積和をモデル式としている。ただし、2 因子実験の場合には、ダミー変数として (a, b) などを用いているので、 β を使うのは紛らわしいため切片を含めてパラメータとして $(\theta_0, \theta_1, \theta_2, \dots)$ を使い

$$y_{ijk} = \theta_0 x_0 + \theta_1 a_{1,i} + \theta_2 b_{1,j} + \theta_3 b_{2,j} + \theta_4 (a_1 b_1)_{ij} + \theta_5 (a_1 b_2)_{ij} + \varepsilon_{ijk} \quad (3.26)$$

のように、線形モデル式とする。

表 3.9 対比型ダミー変数を用いた交互作用を含む 2 因子実験データの解析

				— デザイン行列 X —				分散分析表, 「定数に 0 を使用」 off						
A	B	y	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2	自由度	変動	分散	分散比		
A ₁	B ₁	10	1	1	1	0	1	0	回帰	5	130.00	26.0000	8.6667	
		13	1	1	1	0	1	0	残差	6	18.00	3.0000		
	B ₂	14	1	1	0	1	0	1	合計	11	148.00			
		12	1	1	0	1	0	1						
B ₃	22	1	1	-1	-1	-1	-1							
	19	1	1	-1	-1	-1	-1							
									係数	標準誤差	t	P-値		
A ₂	B ₁	15	1	-1	1	0	-1	0	θ^{\wedge}_0	切片 x_0	16.00	0.5000	32.0000	0.0000
		14	1	-1	1	0	-1	0	θ^{\wedge}_1	a_1	-1.00	0.5000	-2.0000	0.0924
	B ₂	16	1	-1	0	1	0	-1	θ^{\wedge}_2	b_1	-3.00	0.7071	-4.2426	0.0054
		18	1	-1	0	1	0	-1	θ^{\wedge}_3	b_2	-1.00	0.7071	-1.4142	0.2070
B ₃	21	1	-1	-1	-1	1	1		θ^{\wedge}_4	a_1b_1	-0.50	0.7071	-0.7071	0.5060
	18	1	-1	-1	-1	1	1		θ^{\wedge}_5	a_1b_2	-1.00	0.7071	-1.4142	0.2070

表 3.9 右に示すのは、Excel の回帰分析の適用結果であるが、切片の変数 x_0 を除いた一般的な回帰分析の適用結果である。デザイン行列 X としては、切片に対する変数を明示しなければならない。そのために、全て 1 となる変数 x_0 を明示的に与えている。通常回帰分析では、切片の変数 x_0 を内部で補うことが一般的であるが、各種の推定のためには、変数 x_0 を常に意識する必要がある。変数 x_0 をわざわざ含めてあるのは、各種の推定のために必要と

なるパラメータの共分散行列 $\Sigma(\hat{\theta})$

$$\Sigma(\hat{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2, \quad \hat{\sigma}^2 : \text{誤差分散の推定値} \quad (3.27)$$

を求めるために必要不可欠なためである。

要因効果に対する推定と95%信頼区間

表 3.9 右上段に示されている分散分析表は、モデル全体 (A+B+A×B) に対する平方和なので、それぞれの要因についての平方和は示されていない。表 3.9 右下段に式 (3.26) に対応したパラメータ θ_i の推定値が“係数”の欄に示されていて、それぞれについての標準誤差 SE, t 値, および, p 値が示されている。ダミー変数 a_1 に対して $p=0.0924$ と表 3.4 の分散分析表の因子 A についての p 値に等しくなる。したがって、因子が2水準の場合は、パラメータに関する p 値が分散分析表の p 値に等しくなる。

因子 B の (1, -1) 対比型ダミー変数 b_1 と b_2 については、一般平均 (ここでは総平均) の推定値 $\hat{\mu}=16.0$ からの差の p 値が示されており, b_1 のパラメータ $\hat{\theta}_2=-3.0$ に対し $p=0.0054$ と有意な差であるが, b_2 については, $\hat{\theta}_3=-1.0$, $p=0.2070$ と有意ではない。表 3.11 に示すように B₃ 水準については,

$$\text{推定値 } \hat{\beta}_3 = -\hat{\theta}_2 - \hat{\theta}_3 = 3.0 + 1.0 = 4.0 \quad (3.28)$$

として推定される。平方和の分解による分散分析, それに引き続く水準平均の 95%信頼区間のグラフ表示について示したと同様に, 式 (3.1) のデータの構造式の効果 ($\hat{\alpha}_i, \hat{\beta}_j, (\hat{\alpha\beta})_{ij}$) を用いて推定を行なう。図 3.3 に示した水準平均の 95%信頼区間には, 総平均 \bar{y}_{\dots} の分散も含まれており, 効果に対する 95%信頼区間よりも大きめになっている。

デザイン行列 \mathbf{X} のために設定された (1, -1) 対比型ダミー変数に対し, 通し番号の変数 l_i を別名として設定し, 線形和 L を

$$L = \sum_i l_i \hat{\theta}_i = \mathbf{l} \hat{\theta} \quad (3.29)$$

とする。行ベクトル \mathbf{l} は, 推定目的に即した数値を設定する。推定値 $\hat{\beta}_3$ の場合であれば, 表 3.8 の B₃ に対するダミー変数を用いて

$$\mathbf{l}^{(5)} = [0 \quad 0 \quad -1 \quad -1 \quad 0 \quad 0] \quad (3.30)$$

と設定することにより,

$$\left. \begin{aligned} \hat{\beta}_3 &= \mathbf{l}^{(5)} \hat{\theta} \\ &= 0 \times 16.0 + 0 \times (-1.0) - 1 \times (-3.0) - 1 \times (-1.0) + 0 \times (-0.5) + 0 \times (-1.0) \\ &= 3.0 + 1.0 = 4.0 \end{aligned} \right\} \quad (3.31)$$

が推定される。表 3.9 に示したパラメータ (係数) の標準誤差から, $\hat{\beta}_1$ と $\hat{\beta}_2$ の分散は, $Var(\hat{\beta}_1) = Var(\hat{\beta}_2) = 0.7071^2 = 0.50$ であることが示されている。分散の加法性を仮定して $Var(\hat{\beta}_3) = Var(-\hat{\beta}_1 - \hat{\beta}_2)$ を $Var(\hat{\beta}_3) = 2 \times 0.50 = 1.00$ として良いのだろうか。

正しい分散の計算は、パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$ を用い

$$Var(\mathbf{l}\hat{\theta}) = \mathbf{l}\Sigma(\hat{\theta})\mathbf{l}^T \tag{3.32}$$

として求めることができる。表 3.10 に示すパラメータの共分散行列 $\Sigma(\hat{\theta})$ の $\hat{\beta}_1 = \hat{\theta}_2$, $\hat{\beta}_2 = \hat{\theta}_3$ の共分散は $Cov(\hat{\theta}_2, \hat{\theta}_3) = -0.25$ であり、合成分散の一般式を用いて

$$\begin{aligned} Var(\hat{\beta}_3) &= Var(-\hat{\beta}_2 - \hat{\beta}_3) \\ &= Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_2, \hat{\beta}_3) + Var(\hat{\beta}_3) \\ &= 0.50 + 2 \times (-0.25) + 0.50 = 0.50 \end{aligned} \tag{3.33}$$

のように $\hat{\beta}_1$ と $\hat{\beta}_2$ の分散に等しくなる。

表 3.10 パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$

θ^{\wedge}_0	0.25	0.00	0.00	0.00	0.00	0.00
θ^{\wedge}_1	0.00	0.25	0.00	0.00	0.00	0.00
θ^{\wedge}_2	0.00	0.00	0.50	-0.25	0.00	0.00
θ^{\wedge}_3	0.00	0.00	-0.25	0.50	0.00	0.00
θ^{\wedge}_4	0.00	0.00	0.00	0.00	0.50	-0.25
θ^{\wedge}_5	0.00	0.00	0.00	0.00	-0.25	0.50
	x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$

表 3.11 に示すのは、交互作用 A×B の効果も含めて、効果についての合成分散の一般式を用いて計算した結果である。Excel の折れ線グラフに 95%信頼区間を付けるためには、 $t_{0.05}(6) \times SE$ の半幅が必要となる。図 3.5 に示すのは、表 3.11 で計算された線形和 $L^{(i)} = \mathbf{l}^{(i)} \hat{\theta}$ に対し、Excel の折れ線グラフを用いて描き、 $t_{0.05}(12-6) \times SE$ を上下に付加し、95%信頼区間としている。このようにデータの構造式で定義された要因の効果に 95%信頼区間を重ね書きした線グラフによる予測プロファイルは、統計的な判断に役に立つ。

表 3.11 線形和による因子 A, 因子 B, 交互作用 A×B の効果の推定

	効果	L	l_0	l_1	l_2	l_3	l_4	l_5	$\hat{\theta}$	推定値	分散	幅
			x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$				
	α^{\wedge}_1	$L^{(1)}$	0	1	0	0	0	0	16.00	-1.0	0.25	1.223
	α^{\wedge}_2	$L^{(2)}$	0	-1	0	0	0	0	-1.00	1.0	0.25	1.223
	β^{\wedge}_1	$L^{(3)}$	0	0	1	0	0	0	-3.00	-3.0	0.50	1.730
	β^{\wedge}_2	$L^{(4)}$	0	0	0	1	0	0	-1.00	-1.0	0.50	1.730
	β^{\wedge}_3	$L^{(5)}$	0	0	-1	-1	0	0	-0.50	4.0	0.50	1.730
$\alpha^{\wedge}_1 + \beta^{\wedge}_1$	$(\alpha\beta)^{\wedge}_{11}$	$L^{(6)}$	0	1	1	0	1	0	-1.00	-4.5	1.25	2.736
$\alpha^{\wedge}_1 + \beta^{\wedge}_2$	$(\alpha\beta)^{\wedge}_{12}$	$L^{(7)}$	0	1	0	1	0	1		-3.0	1.25	2.736
$\alpha^{\wedge}_1 + \beta^{\wedge}_3$	$(\alpha\beta)^{\wedge}_{13}$	$L^{(8)}$	0	1	-1	-1	-1	-1		4.5	1.25	2.736
$\alpha^{\wedge}_2 + \beta^{\wedge}_1$	$(\alpha\beta)^{\wedge}_{21}$	$L^{(9)}$	0	-1	1	0	-1	0		-1.5	1.25	2.736
$\alpha^{\wedge}_2 + \beta^{\wedge}_2$	$(\alpha\beta)^{\wedge}_{22}$	$L^{(10)}$	0	-1	0	1	0	-1		1.0	1.25	2.736
$\alpha^{\wedge}_2 + \beta^{\wedge}_3$	$(\alpha\beta)^{\wedge}_{23}$	$L^{(11)}$	0	-1	-1	-1	1	1		3.5	1.25	2.736
											$t_{0.05}(12-6) =$	2.447

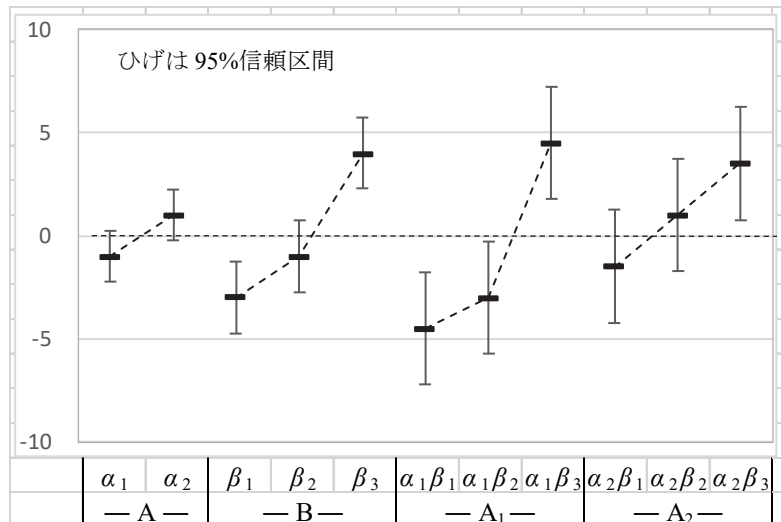


図 3.5 要因効果についての線グラフによる予測プロファイル

効果 α などに対する有効反復数 n_e を用いた分散の推定

表 3.6 で示した水準平均の分散は、元のデータに戻って水準平均の反復数 n_e を求めて計算したのであるが、効果 α などの分散の推定には、少々ややこしい問題が発生する。因子 A の第 1 水準 A_1 の水準平均は、

$$\bar{y}_{1..} = \bar{y}_{...} + \alpha_1 \quad (3.34)$$

なので、 α_1 について解くと

$$\alpha_1 = \bar{y}_{1..} - \bar{y}_{...} \quad (3.35)$$

が得られる。効果 α_1 の分散 $Var(\alpha_1)$ は、

$$Var(\alpha_1) = Var(\bar{y}_{1..} - \bar{y}_{...}) \quad (3.36)$$

となるが、 $\bar{y}_{1..}$ と $\bar{y}_{...}$ には、共通のデータが含まれており、互いに独立ではなく、分散の加法性が成り立たないので、簡単に計算できない。元のデータに戻って

$$\begin{aligned} Var(\bar{y}_{1..} - \bar{y}_{...}) &= Var\left(\frac{y_{111} + y_{112} + \dots + y_{132}}{6} - \frac{y_{111} + y_{112} + \dots + y_{211} + y_{212} + \dots + y_{232}}{12}\right) \\ &= Var\left(\frac{y_{111} + y_{112} + \dots + y_{132}}{12} - \frac{y_{211} + y_{212} + \dots + y_{232}}{12}\right) \\ &= \frac{6\hat{\sigma}^2}{12^2} + \frac{6\hat{\sigma}^2}{12^2} \\ &= \frac{\hat{\sigma}^2}{12} = \frac{3.00}{12} = 0.250 \end{aligned} \quad (3.37)$$

となり、有効反復数は $n_e = 12$ となる。表 3.6 に示されている A_1 水準の平均値に対する反復数は $n_e = 6$ なので $Var(\bar{A}_1) = 3.00/6 = 0.50$ であり、推定される分散 $Var(\alpha_1) = 3.00/12 = 0.25$ と明らかに小さくなり、図 3.5 に示したように 95%信頼区間も小さくなる。

Excel の回帰分析を用いた分散分析表の作成

因子 A のように 2 水準の場合であれば、分散分析の F 検定の p 値と回帰分析の t 検定の p 値は同じになるので、あえて分散分析表による検討は必要ない。ただし、因子 B のように 3 水準の場合に分散分析表で示したような因子全体としての p 値も欲しい場合には、デザイン行列 X の一部分を使った回帰分析の分散分析表から平方和が得られる。

表 3.12 に示すのは、デザイン行列 X の (a_1) , (b_1, b_2) , (a_1b_1, a_1b_2) 列を別々に用いた回帰分析の結果で、回帰の変動（平方和）から、表 3.4 に示す因子 A の 12.0、因子 B の 104.00、交互作用 A×B の 14.00 が得られている。なお、このようなデザイン行列 X の列の一部を使った平方和が適切なのは、繰り返しが等しい場合に限られる。次節で示すように、繰り返しが不揃い場合には使えない。

表 3.12 因子 A, B, および交互作用 A×B についての回帰分析

分散分析表, 因子 A, 2水準			分散分析表, 因子 B, 3水準			分散分析表, 交互作用 A×B		
	自由度	変動		自由度	変動		自由度	変動
回帰 A	1	12.00	回帰 B	2	104.00	回帰A×B	2	14.00
残差	10	136.00	残差	9	44.00	残差	9	134.00
合計	11	148.00	合計	11	148.00	合計	11	148.00

このような方法で平方和が計算できるのは、デザイン行列 X に関して“因子が直交”している場合に限られる。“因子が直交”していることを確認するためには、正規方程式で使われている $(X^T X)$ を計算した際に、各要因（因子）間の積和の部分がすべて 0 となっている場合である。表 3.13 に示すように、 (a_1) 列 (b_1, b_2) 列の積和が全て 0、 (b_1, b_2) 列と (a_1b_1, a_1b_2) 列の積和が全て 0 であることから因子 A, 因子 B および交互作用 A×B が全体として直交していることが確認される。なお、 b_1 と b_2 の間は 0 でないので直交していない。

表 3.13 デザイン行列 X の直交性の確認

	$X^T X$					
	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2
x_0	12	0	0	0	0	0
a_1	0	12	0	0	0	0
b_1	0	0	8	4	0	0
b_2	0	0	4	8	0	0
a_1b_1	0	0	0	0	8	4
a_1b_2	0	0	0	0	4	8
	=Mmult(Transpose(X の範囲), X の範囲)					

3.2. 繰り返しが不揃いの2因子実験

きちっと計画され実施された実験であっても、何らかの手違いなどでデータが得られないことも経験する。観察データを整理して2元表にまとめ、分散分析による解析をしようとする場合には、必然的に繰り返しが不揃いになる。繰り返しが不揃いの1因子実験の場合は、平方和の分解による解析手順が確立している。しかし、繰り返しが不揃いの2因子実験の場合には、平方和の分解による解析が行なえない。線形モデルの解析手順による解析を行なえば、良いのではないかと思うが、日本語での解析事例を見いだすことは、困難である。ただし、SAS Institute Inc. (2020), 「The GLM Procedure, , SAS/STAT 15.2 User's Guide」を紐解くと、最初の Getting Started 節で、「PROC GLM for Unbalanced ANOVA」が取り上げられているようにポピュラーな課題である。なお、本節のテーマを主体にした JMP の Discovery Summit Japan 2022 での講演スライドを元に、高橋 (2023), 「JMP で繰り返しが不揃いの2元配置データの解析ができるの? - 平方和の分解ではなくデザイン行列と最小2乗平均の活用 -」を「続・高橋セミナー 第11回」として先行して公表したので参考にしてもらいたい。

繰り返しが不揃いの2因子実験データ

表 3.14 に示すのは、高橋ら (1989) の第 15-3 節、「因子が直交しない場合」に示されているデータである。繰り返しが不揃いで因子が直交しないこのデータは、表 3.1 の2因子実験データに対し A_1B_2 のセル平均が変わらないように2つのデータ (15, 11) を加えたものである。この繰り返しが不揃いのデータは SAS の GLM プロシジャのタイプ I, II, III, IV の各平方和の特徴を浮き彫りにするために例示したものである。

表 3.14 繰り返しが不揃いで因子が直交しない場合 [高橋ら (1989), 図表 15.11]

	B ₁		B ₂				B ₃		平均 \bar{y}
A ₁	10	13	14	12	15	11	22	19	14.5
A ₂	15	14	16	18			21	18	17.0
平均 \bar{y}	13.0		14.33				20.0		15.57

新たなデータに対し、何らかのグラフを作成して全体を概観することが、統計解析の最初の一步である。因子 A は2水準、因子 B は3水準なので、因子 B の水準番号 j を X 軸とし、反応 y を Y 軸とする Excel による層別散布図を作成する。表 3.15 に示すようにグラフの作成のために行方向にデータを並べ替え、 A_iB_j のセルごとの平均値を計算し、層別散布図に重ね書きしたグラフも示す。なお、 A_1B_2 セルのデータ数は4と異なるが、セル平均は、表 3.1 と同じ 13.0 である。

表 3.15 組み合わせ平均

A	B	j	y_{ijk}	A	B	j	$\bar{y}_{ij\cdot}$	
A ₁	B ₁	1	10	A ₂	B ₁	1	11.5	
		1	13		B ₂	2	13.0	
		2	14		B ₃	3	20.5	
	B ₂	2	12		B ₁	1	14.5	
		2	15		B ₂	2	17.0	
		2	11		B ₃	3	19.5	
	B ₃	3	22		一般平均 $\hat{\mu} =$			16.0
		3	19					
A ₂	B ₁	1	15					
		1	14					
	B ₂	2	16					
		2	18					
	B ₃	3	21					
		3	18					
総平均 $\bar{y}_{\dots} =$			15.57					

図 3.6 層別散布図に組合せ平均の重ね書き

きれいな層別散布図の作成手順

- 1) A₁水準の (j, y_{1jk}) の範囲を選択し, Excel のデフォルトの散布図を作成し, マーカを「○」に変更する.
- 2) 「データの選択」, 新たなデータの「追加」で, 因子 A₂水準の (j, y_{2j}) の範囲を選択して, マーカを重ね書きし, マーカを「+」に整える.
- 3) 「データの選択」, 新たなデータの「追加」で, 因子 A₁の平均 ($j, \bar{y}_{1j\cdot}$) の範囲を選択し, 平均値のマーカを重ね書きし, 点線「---」で結び, マーカを「なし」とする.
- 4) 因子 A₂について同様に重ね書きし, ドット線「…」とする.
- 5) 因子 B の水準番号の軸ラベルを消去し, 「図形→キストボックス」の挿入で「B₁ B₂ B₃」を書き込む.

適切なグラフを見れば, どのように結果を解釈するのかは自明となる. ただし, グラフによる結果の判断には過度な思い込みが入り込みがちである. そのためには, 統計解析の結果を謙虚に受け止め解釈に反映する必要がある.

データの構造式

データの構造式は, 式 (3.1) で示した場合と同様で

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (3.38)$$

あるが, 繰り返し数の違いを考慮する必要がある. 主効果については, n_i, n_j を周辺のデータ数として主効果は,

$$\sum_i n_i \alpha_i = 0, \quad \sum_j n_j \beta_j = 0 \quad (3.39)$$

重み付き平均を0とし、交互作用については、 n_{ij} を組合せセルのデータ数として

$$\left. \begin{aligned} \sum_j n_{ij}(\alpha\beta)_{ij} &= 0, \quad i=1, 2 \\ \sum_i n_{ij}(\alpha\beta)_{ij} &= 0, \quad j=1, 2, 3 \end{aligned} \right\} \quad (3.40)$$

重み付き平均を0とするような関係が成り立つとする。

偏差平方和の計算

繰り返しが不揃いの場合には、「分散の加法性が成り立たない」ので平方和の分解による分散分析表が作成できないと言われているが、本当なのだろうか。表3.3に示した修正項CTを用いた計算は、 (n_A, n_B, n_K) が一定でなく使えないので、表3.2に示したように全データを行方向に並べて、ExcelのSumSq()関数を用いた平方和の計算を行なう。

表3.16に示すように、Average()関数により総平均 $\bar{y}_{...}=15.57$ を全ての行に書き込み、偏差 $(y_{ijk}-\bar{y}_{...})$ を求めSumSq()関数で、総平方和 $S_T=171.43$ を得る。因子Aのそれぞれの平均 \bar{y}_A を書き込み $(\bar{y}_A-\bar{y}_{...})$ についての平方和から $S_A=21.43$ を得る。因子Bのそれぞれの平均 \bar{y}_B を書き込み $(\bar{y}_B-\bar{y}_{...})$ についての平方和から $S_B=114.10$ を得る。因子Aと因子Bの組合せ平均を \bar{y}_{AB} 全ての行に、さらに式(3.12)に示した期待平均 $\hat{y}'_{ij.}=\bar{y}_{i..}+\bar{y}_{.j.}-\bar{y}_{...}$ を計算し、 $(\bar{y}_{AB}-\hat{y}'_{AB})$ についての平方和から $S_{A \times B}=16.10$ を得る。最後に反応 y_{ijk} と組合せ平均 \bar{y}_{AB} の差 $(y_{ijk}-\bar{y}_{AB})$ から残差平方和 $S_e=26.00$ を得る。

表 3.16 Excel の Sumsq() 関数による偏差平方和の計算

		総平均			主効果		主効果		セル	期待	交互作用	残差
		μ^{\wedge}			A	α^{\wedge}	B	β^{\wedge}	平均	平均	$(\alpha\beta)^{\wedge}$	ε^{\wedge}
A	B	y_{ijk}	$\bar{y}_{...}$	$y-\bar{y}_{...}$	\bar{y}_A	$\bar{y}_A-\bar{y}_{...}$	\bar{y}_B	$\bar{y}_B-\bar{y}_{...}$	\bar{y}_{AB}	\hat{y}'_{AB}	$\bar{y}_{AB}-\hat{y}'_{AB}$	$y-\bar{y}_{AB}$
A ₁	B ₁	10	15.57	-5.57	14.50	1.07	13.00	2.57	11.50	11.93	-0.43	-1.50
		13	15.57	-2.57	14.50	1.07	13.00	2.57	11.50	11.93	-0.43	1.50
	B ₂	14	15.57	-1.57	14.50	1.07	14.33	1.24	13.00	13.26	-0.26	1.00
		12	15.57	-3.57	14.50	1.07	14.33	1.24	13.00	13.26	-0.26	-1.00
		15	15.57	-0.57	14.50	1.07	14.33	1.24	13.00	13.26	-0.26	2.00
		11	15.57	-4.57	14.50	1.07	14.33	1.24	13.00	13.26	-0.26	-2.00
B ₃	22	15.57	6.43	14.50	1.07	20.00	-4.43	20.50	18.93	1.57	1.50	
	19	15.57	3.43	14.50	1.07	20.00	-4.43	20.50	18.93	1.57	-1.50	
A ₂	B ₁	15	15.57	-0.57	17.00	-1.43	13.00	2.57	14.50	14.43	0.07	0.50
		14	15.57	-1.57	17.00	-1.43	13.00	2.57	14.50	14.43	0.07	-0.50
	B ₂	16	15.57	0.43	17.00	-1.43	14.33	1.24	17.00	15.76	1.24	-1.00
		18	15.57	2.43	17.00	-1.43	14.33	1.24	17.00	15.76	1.24	1.00
	B ₃	21	15.57	5.43	17.00	-1.43	20.00	-4.43	19.50	21.43	-1.93	1.50
		18	15.57	2.43	17.00	-1.43	20.00	-4.43	19.50	21.43	-1.93	-1.50
			171.43		21.43		114.10				16.10	26.00
			S_T		S_A		S_B				$S_{A \times B}$	S_e
$S'_T = S_A + S_B + S_{A \times B} + S_e =$										177.62		

表 3.16 で計算された総平方和は、 $S_T = 171.43$ であり、各要因の平方和の合計は、

$$S'_T = S_A + S_B + S_{A \times B} + S_e = 21.43 + 114.10 + 16.10 + 26.00 = 177.62 \quad (3.41)$$

となり、一致しないことは明らかである。したがって、分散分析表としてまとめることができない。

デザイン行列 X を用いた回帰分析

表 3.8 に示した (1, -1) 対比型ダミー変数を用いて因子 A, 因子 B, および交互作用 A×B に関する (14 行×6 列) のデザイン行列 X を作成し、Excel の回帰分析を適用した結果を表 3.17 に示す。分散分析表から、モデル全体のあてはまりの程度は、「回帰」に対する F 値で示されている。 F 値に対する p 値を表示スペースの都合でカットしたので、Excel の F.dist.RT () 関数を用いて p 値を求めると

$$p = \text{F.dist.RT}(8.9495, 5, 8) = 0.0039 \quad (3.42)$$

と高度に有意であることが確認される。

(1, -1) 対比型ダミー変数を用いた線形モデル (回帰モデル) に対するパラメータ (係数) は、次のように推定され、

$$\left. \begin{aligned} \hat{y}_{ijk} &= \hat{\theta}_0 x_0 + \hat{\theta}_1 a_1 + \hat{\theta}_2 b_1 + \hat{\theta}_3 b_2 + \hat{\theta}_4 a_1 b_1 + \hat{\theta}_5 a_1 b_2 \\ &= 16.0x_0 + (-1.0)a_1 + (-3.0)b_1 + (-1.0)b_2 + (-0.5)a_1 b_1 + (-1.0)a_1 b_2 \end{aligned} \right\} \quad (3.43)$$

推定されたパラメータは、繰り返しが等しい場合の表 3.9 に示した結果と一致している。これは、データを追加してもセル平均が変わらないようにしたためである。このことから、(1, -1) 対比型ダミー変数を用いた線形モデルは、セル平均を主体にした解析となっていると理解される。

表 3.17 繰り返しが不揃いの2因子実験データに対する回帰分析

		— デザイン行列 X —							分散分析表, 「定数に 0 を使用」 off																		
		y	x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$		自由度	変動	分散	分散比														
A ₁	B ₁	10	1	1	1	0	1	0	回帰	5	145.4286	29.0857	8.9495														
		13	1	1	1	0	1	0																			
	B ₂	14	1	1	0	1	0	1						残差	8	26.0000	3.2500										
		12	1	1	0	1	0	1																			
		15	1	1	0	1	0	1																			
B ₃	11	1	1	0	1	0	1	合計	13	171.4286																	
	22	1	1	-1	-1	-1	-1					θ^{\wedge}_0 切片 x_0	16.00					0.4983	32.1117	0.0000							
	19	1	1	-1	-1	-1	-1							θ^{\wedge}_1 a_1	-1.00	0.4983	-2.0070				0.0797						
	A ₂	B ₁	15	1	-1	1	0															-1	0	θ^{\wedge}_2 b_1	-3.00	0.7205	-4.1639
14			1	-1	1	0	-1															0	θ^{\wedge}_3 b_2				
B ₂	B ₂	16	1	-1	0	1	0	-1	θ^{\wedge}_4 $a_1 b_1$	-0.50	0.7205	-0.6940	0.5073														
		18	1	-1	0	1	0	-1						θ^{\wedge}_5 $a_1 b_2$	-1.00	0.6719	-1.4884	0.1750									
	B ₃	B ₃	21	1	-1	-1	-1	1	1																		
18			1	-1	-1	-1	1	1																			

推定されたパラメータ（係数）の p 値で $p < 0.05$ となっているのは、ダミー変数 b_1 の推定 $\hat{\theta}_2 = -3.00$ のみである。この結果をどのように解釈したらよいのであろうか。一般平均（セル平均の平均）の推定値は、 $\hat{\mu} = \hat{\theta}_0 = 16.00$ で、 B_1 水準と総平均 $\hat{\mu}$ の差が $\hat{\theta}_2 = -3.00$ となり、統計的に有意であるとの解釈となる。 B_3 水準の推定値は、対比型ダミー変数なので $\hat{\beta}_3 = -\hat{\theta}_2 - \hat{\theta}_3 = 4.00$ となるが、 p 値はどうなるのであろうか。交互作用に対する a_1b_1 、 a_1b_2 ダミー変数の p 値から、全体としては有意ではないように思われるが、全体としてはどうなのだろうか。

分散分析表の作成

表 3.16 での偏差平方和による計算手順では、分散分析表の作成ができないことを明らかにした。また、表 3.17 に示した Excel の回帰分析では、モデル全体の平方和を計算することができたのであるが、因子 A、因子 B、交互作用 A×B の平方和を計算し、分散分析表を作成し統計的に評価する必要もある。線形モデルによる解析を基本としている先進的な SAS の GLM プロシジャでは、(タイプ I, II, III, IV の平方和) として平方和の計算を行っている。最も基本的な平方和は、逐次平方和であり、タイプ I の平方和とも言われている。

逐次平方和は、表 3.18 に示すようにデザイン行列のダミー変数の一部を使い Excel の回帰分析を行なう。表 3.19 に示すように回帰の平方和を $S_{(A)} = S_A^{(1)}$ 、 $S_{(A+B)}$ 、 $S_{(A+B+A \times B)}$ を求め、それらの差分

$$S_B^{(1)} = S_{(A+B)} - S_{(A)} \tag{3.44}$$

$$S_{A \times B}^{(1)} = S_{(A+B+A \times B)} - S_{(A+B)} \tag{3.45}$$

をそれぞれの要因の平方和とする。このようにして求めた平方和を加えた場合に回帰モデルの平方和 S_T が

$$S_T = S_A^{(1)} + S_B^{(1)} + S_{A \times B}^{(1)} \tag{3.46}$$

のように平方和に分解が成り立つ。

表 3.18 逐次（タイプ I）平方和の算出のためのダミー変数の選択

番号		デザイン行列 X						回帰の平方和
		切片	A	—B—	—A×B—			
		x_0	a_1	b_1	b_2	a_1b_1	a_1b_2	
1	因子Aのみ	△	○					$S_{(A)}$
2	因子A+因子B	△	○	○	○			$S_{(A+B)}$
3	因子A+因子B+交互作用A×B	△	○	○	○	○	○	$S_{(A+B+A \times B)}$

表 3.19 に示すように、Excel の回帰分析で、デザイン行列 X の変数を逐次増加させることにより、因子 A のみの平方和 $S_{(A)} = 21.4286$ 、さらに因子 B を加えた $S_{(A+B)} = 130.2286$ 、フルモデル $S_{(A+B+A \times B)} = 145.4286$ が得られている。

表 3.19 Excel の回帰分析による逐次平方和の算出

		— デザイン行列 X —							1. 分散分析表 因子A 「定数に0を使用」off				
		y	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2	自由度	変動	分散	分散比	
A ₁	B ₁	10	1	1	1	0	1	0	回帰	1	21.4286	21.4286	1.7143
		13	1	1	1	0	1	0	残差	12	150.0000	12.5000	
	B ₂	14	1	1	0	1	0	1	合計	13	171.4286		
		12	1	1	0	1	0	1	2. 分散分析表 因子A+因子B				
		15	1	1	0	1	0	1	自由度	変動	分散	分散比	
	B ₃	11	1	1	0	1	0	1	回帰	3	130.2286	43.4095	10.5363
		22	1	1	-1	-1	-1	-1	残差	10	41.2000	4.1200	
		19	1	1	-1	-1	-1	-1	合計	13	171.4286		
A ₂	B ₁	15	1	-1	1	0	-1	0	3. 分散分析表 因子A+因子B+交互作用A×B				
		14	1	-1	1	0	-1	0	自由度	変動	分散	分散比	
	B ₂	16	1	-1	0	1	0	-1	回帰	5	145.4286	29.0857	8.9495
		18	1	-1	0	1	0	-1	残差	8	26.0000	3.2500	
	B ₃	21	1	-1	-1	-1	1	1	合計	13	171.4286		
		18	1	-1	-1	-1	1	1					
			A		B		A×B						

表 3.20 に示すのは、ダミー変数を逐次増加して得られた回帰の平方和（変動）の差分を計算し、分散分析表にまとめた結果である。因子 A と因子 B は、 $p < 0.05$ と有意な差であるが、交互作用 A×B は、有意な差ではない。このようにして、繰り返しが不揃いの2因子実験データについて逐次平方和による分散分析表が得られる。

表 3.20 逐次（タイプ I）平方和による分散分析表

要因	回帰の平方和	差し引く平方和		df	タイプ I	平均平方	F 値	p 値	
A	$S_{(A)}$	21.43	=	$S_{(A)}^{(1)}$	1	21.43	21.43	6.59	0.0332
B	$S_{(A+B)}$	130.23	- $S_{(A)}$	$S_{(B)}^{(1)}$	2	108.80	54.40	16.74	0.0014
A×B	$S_{(A+B+A×B)}$	145.43	- $S_{(A+B)}$	$S_{(A×B)}^{(1)}$	2	15.20	7.60	2.34	0.1586
e					8	26.00	3.25		
T					13	171.43			

最適水準と現行水準の差の推定と95%信頼区間

繰り返しが揃っている場合でも、不揃いの場合でもデザイン行列 X を用いた解析方法は、デザイン行列 X のサイズが異なるだけで変わることがない。第 3.1 節では、表 3.12 に示したように各要因の平方和を直接求める方法を示したが、表 3.20 に示したと同様の方法により分散分析表を作成することもできる。第 3.1 節で示した線形和 L による各種の水準平均の 95% 信頼区間の算出方法と同様に、繰り返しが不揃いの場合についても同じ手順により推定することができる。そこで、第 3.1 節で扱わなかった最適水準と現行水準の差の推定と 95% 信頼区間のグラフ表示に焦点をあてる。

表 3.20 に示した逐次平方和を用いた分散分析表で、交互作用 A×B の p 値は、 $p=0.1586$ であるが、交互作用 A×B があることを前提にする。表 3.17 の回帰分析で推定されたパラメータを用い、線形和を用いて交互作用を考慮した効果の推定を行なう。表 3.21 に示すのは、交互作用 A×B を含む切片 $x_0=0$ とした (1, -1) 対比型ダミー変数を用いた線形和による効果の推定値である。なお、切片を $x_0=1$ とすれば、表 3.15 に示した組み合わせ平均に一致する。

表 3.21 交互作用を考慮した効果の推定

A	B	j	L	l_0	l_1	l_2	l_3	l_4	l_5	θ^\wedge	$l\theta^\wedge$
				x_0	a_1	b_1	b_2	a_1b_1	a_1b_2		
A ₁	B ₁	1	$L^{(1)}$	0	1	1	0	1	0	16.00	-4.50
	B ₂	2	$L^{(2)}$	0	1	0	1	0	1	-1.00	-3.00
	B ₃	3	$L^{(3)}$	0	1	-1	-1	-1	-1	-3.00	4.50
A ₂	B ₁	1	$L^{(4)}$	0	-1	1	0	-1	0	-1.00	-1.50
	B ₂	2	$L^{(5)}$	0	-1	0	1	0	-1	-0.50	1.00
	B ₃	3	$L^{(6)}$	0	-1	-1	-1	1	1	-1.00	3.50

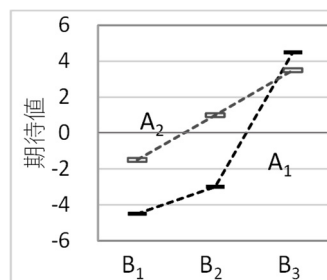


図 3.7 交互作用を考慮した効果の推定

実験は、A₁B₁の現行水準に対し、反応 y が大きくなる組合せ水準の探索を目的としている。A₁B₁の効果 $(\hat{\alpha\beta})_{11} = -4.50$ に対し、最も大きな反応（最適水準）A₁B₃の効果 $(\hat{\alpha\beta})_{13} = 4.50$ であり、それらの差に対する 95%信頼区間を求めて統計的な判断をしたい。また、A₁B₃と同程度の A₂B₃についても検討したい。これは、最適でなくとも現行の A₁水準から A₂水準に変更することにより製造コストなどの低減が図れれば、魅力的でもあるためである。さらに、A₂B₂との差が統計的に有意なのか 95%信頼区間を示し判断したい。これも、B₃水準に変更するよりも B₂水準への変更が技術的に容易などの理由があれば、最適水準でもなくとも、統計的に有意な差があれば、採用したいとの目的にそったものである。

表 3.22 に示すように現行水準 $L^{(1)}$ と最適水準 $L^{(3)}$ の効果の差の推定するための係数ベクトル l は、

$$l^{(7)} = l^{(3)} - l^{(1)} = [0 \quad 0 \quad -2 \quad -1 \quad -2 \quad -1] \tag{3.47}$$

として求められる。

表 3.22 現行水準と最適水準の期待値の差

				l_0	l_1	l_2	l_3	l_4	l_5	
				L	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2
最適水準	A ₁	B ₃	$L^{(3)}$	0	1	-1	-1	-1	-1	
現行水準	A ₁	B ₁	$L^{(1)}$	0	1	1	0	1	0	
		差	$L^{(7)}$	0	0	-2	-1	-2	-1	

同様に、最適水準と同程度の A_2B_3 の $L^{(6)}$ 、気になる A_2B_2 の $L^{(5)}$ も同様に係数ベクトル \boldsymbol{l} を求めることができる。各種の線形和 $L^{(k)}$ の分散 $Var(L^{(k)})$ は、パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ に関する $\boldsymbol{l}^{(k)}$ の2次形式で求められることを式(3.32)で示した。繰り返しが不揃いであっても $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の計算は、 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\sigma}^2$ と常に一定である。もちろんデザイン行列 \boldsymbol{X} および誤差分散の期待値 $\hat{\sigma}^2$ の内容は異なるが、計算式は同じである。

表 3.23 に示すパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ は、Excel の行列計算を用いれば容易に計算できることをこれまで繰り返し示してきた。表 3.23 右の SE は、表 3.17 の「標準誤差」に一致する。これは、 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の対角要素が、回帰パラメータの分散 $Var(\hat{\boldsymbol{\theta}})$ となるので、その平方根を取ったものである。

表 3.23 パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$

		パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\sigma}^2$					$Var(\hat{\boldsymbol{\theta}})$	SE	標準誤差	
$\hat{\theta}_0$	x_0	0.2483	-0.0226	0.0226	-0.0451	0.0226	-0.0451	0.2483	0.4983	0.4983
$\hat{\theta}_1$	a_1	-0.0226	0.2483	0.0226	-0.0451	0.0226	-0.0451	0.2483	0.4983	0.4983
$\hat{\theta}_2$	b_1	0.0226	0.0226	0.5191	-0.2257	-0.0226	0.0451	0.5191	0.7205	0.7205
$\hat{\theta}_3$	b_2	-0.0451	-0.0451	-0.2257	0.4514	0.0451	-0.0903	0.4514	0.6719	0.6719
$\hat{\theta}_4$	$a_1 b_1$	0.0226	0.0226	-0.0226	0.0451	0.5191	-0.2257	0.5191	0.7205	0.7205
$\hat{\theta}_5$	$a_1 b_2$	-0.0451	-0.0451	0.0451	-0.0903	-0.2257	0.4514	0.4514	0.6719	0.6719
=Minverse (Mmult (Transpose (Xの範囲), Xの範囲)) * $\hat{\sigma}^2$								対角要素	平方根	回帰分析

表 3.24 に示すのは、現行水準 A_1B_1 と最適水準 A_1B_3 の差など含め、線形和による推定を行ない、パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ を用いた分散の推定を行ない、95%信頼区間を計算した結果であり、どれも95%信頼区間の下限 $L_{95\%}$ がゼロ0を含まないので、有意な差となることが示される。

表 3.24 現行水準と最適水準などの期待値の差の95%信頼区間

		l_0	l_1	l_2	l_3	l_4	l_5						
L		x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$	$\hat{\boldsymbol{\theta}}$	$\boldsymbol{l} \hat{\boldsymbol{\theta}}$	$Var(\boldsymbol{l} \hat{\boldsymbol{\theta}})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
$A_1B_1 - A_1B_1$	$L^{(0)}$	0	0	0	0	0	0	16.00	= 0.00	0.0000	0.0000	0.00	0.00
$A_1B_3 - A_1B_1$	$L^{(7)}$	0	0	-2	-1	-2	-1	-1.00	9.00	3.2500	4.1572	4.84	13.16
$A_2B_3 - A_1B_1$	$L^{(8)}$	0	-2	-2	-1	0	1	-3.00	8.00	3.2500	4.1572	3.84	12.16
$A_2B_2 - A_1B_1$	$L^{(9)}$	0	-2	-1	1	-1	-1	-1.00	5.50	3.2500	4.1572	1.34	9.66
									-0.50	$t_{0.05}(14-6) =$	2.3060		
									-1.00				

図 3.8 に示すのは、推定された線形和 $L^{(k)} = \boldsymbol{l}^{(k)} \boldsymbol{\theta}$ について、Excel の折れ線グラフに $t_{0.05} \times SE$ の幅を付けた結果であり、効果の差の95%信頼区間が示されている。期待される効果の差に対する95%信頼区間の相対的な関係が示されている。

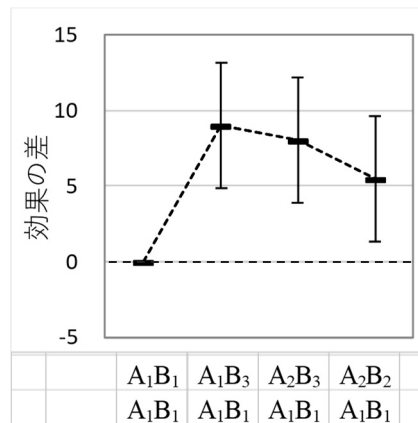


図 3.8 現行水準 A₁B₁と最適水準 A₁B₃などとの差の 95%信頼区間

啓蒙活動

2元配置分散分析と言えば、「繰り返しが無い」、あるいは、「繰り返しが有る」について論じられ、「繰り返しが不揃い」について論じられることはまれである。そこで、本章で取り上げた事例を用い 2022 年の SAS ユーザ総会では、「平方和の分解では解けない繰り返しが不揃いの 2 元配置データの解析入門」、2022 年の Discovery Summit Japan では、「JMP で繰り返しが不揃いの 2 元配置データの解析ができるの?」、2022 年度統計関連学会連合大会では、「Excel による繰り返しが不揃いの 2 元配置データの解析—Lsmeans とタイプ III の平方和の活用—」と題して講演を行った。さらに、これらを取りまとめ、高橋 (2023)、第 11 回 続・高橋セミナーとして公表したので、参考にしてもらいたい。

BioStat
高橋セミナー | 続・高橋セミナー | 芳賀敏郎先生 | ユックムス

メニュー	<p style="text-align: center;">続・高橋セミナー</p> <p>続・高橋セミナー 第11回 JMPで繰り返しが不揃いの2元配置データの解析ができるの? - 平方和の分解ではなくデザイン行列と最小2乗平均の活用 -</p> <p>2023年03月</p> <p>要約</p> <p>実データで、2つの質的変数、量的な変数を応答として解析しようとする必率的に「繰り返しが不揃いの2元配置」に帰着するが、この問題を扱っている日本語の書籍を見出すことができない。そのためか、Web上では、「繰り返しが等しくない」と解析できない」などが蔓延している。その原因は、平方和の分解に頼り切りで、タミー変数を用いたデザイン行列Xを用いた解析法が、まったく普及していないためである。「モデルのあてはめ」を使っているJMPユーザは、繰り返しが等しくても、不揃いでも、全く気にせず解析しているはずである。「効果の詳細」での質的変数の水準平均を求めようすると、「最小2乗平均」と「平均」が並列して出力されるので、それらに食い違いが出ることをうすうす知っているに違いない。ただし、その違いをきちっと説明することができるのだろうか。そこで、2×3の繰り返しが不揃いの2元配置データについて、Excelの行列関数を用いた解析方法を示し、JMPによる解析結果と対比することにより、「モデルのあてはめ」の計算原理の理解を深め、更なる応用ができるように、平方和の分解に代わる計算方法を示す。</p>
トップ	
セミナー記録	

3.3. タイプ I, タイプ II, タイプ III の平方和

繰り返しが等しい 2 因子実験データ場合には、取り上げた要因ごとの偏差平方和の合計が総平方和 S_T に一致するように、いわゆる平方和の分解が成り立つことにより分散分析表が作成できることを第 3.1 節で示した。繰り返しが不揃いの 2 因子実験データ場合には、要因ごとの偏差平方和の計算が可能ではあるが、それぞれの平方和を加えた合計が総平方和 S_T に等しくならないために、分散分析表にまとめられない。

繰り返しが不揃いの場合でも分散分析表を作成したい、との要望に沿う形で、ダミー変数を用いた線形モデルに対し、タイプ I, タイプ II, タイプ III の平方和が考案された。Excel の回帰分析を段階的に適用し、モデル（回帰）の平方和の増分に着目することで平方和の分解ができ、逐次増加型の平方和（タイプ I の平方和）として分散分析表が作成できることを示した。

逐次増加型の平方和は、第 2.4 節の多項式回帰による Lack of fit の解析でも取り上げた。これは、1 次式をあてはめた回帰分析のモデル（回帰）の平方和、2 次式をあてはめた場合、3 次式をあてはめた場合の平方和が逐次増加することに着目し、その差分の平方和を分散分析表にまとめられることを示した。この逐次増加型の平方和が、タイプ I の平方和と称されていることも示した。

主効果を主体にしたタイプ II の平方和

繰り返しが不揃いの 2 因子実験データに対する逐次増加型の平方和は、モデルに含めるダミー変数を解析目的に合わせて逐次的に増加する必要がある。表 3.20 に示した逐次（タイプ I）平方和による分散分析表は、変数の増加を

$$(A) \rightarrow (A+B) \rightarrow (A+B+A \times B)$$

とした。ただし、因子 B を最初のモデルとし、A を加え、さらに $A \times B$ を加えることも可能である。

$$(B) \rightarrow (A+B) \rightarrow (A+B+A \times B)$$

モデルに含めるダミー変数の順番を変えた場合に各要因の平方和が、一致すれば幸いなのであるが、不幸なことに一致しない。そのために逐次平方和による分散分析表の結果には恣意性が入り込んでしまう。恣意性を排除するため、モデルに含める順番に関係なく平方和が一致するような“平方和”が望まれる。そこで、考案されたのが、主効果を主体にしたタイプ II の平方和である。

表 3.25 に示すようにタイプ II の平方和は、(主効果のみを含むモデル)、(主効果と交互作用を含むモデル) のように 2 段階のモデルを想定する。すべての主効果を含むモデルの平方和を $S_{(A+B)}$ とし、この主効果モデルから主効果の因子を 1 つずつ除いたモデル、因子 A を除いたモデルの平方和を $S_{(B)}$ とし、主効果モデルの平方和 $S_{(A+B)}$ から因子 B を除いたモデルの平方和の減少分

$$S_A^{(2)} = S_{(A+B)} - S_{(B)}$$

が、因子 A のタイプ II の平方和 $S_A^{(2)}$ となる。言い換えると因子 A が含まれていないモデルに対し、因子 A を加えた場合の平方和の増加分とも言える。

同様に、因子 A を除いたモデルの平方和が $S_{(B)}$ となり、主効果モデルの平方和から因子 A を除いたモデルの平方和の減少分

$$S_B^{(2)} = S_{(A+B)} - S_{(A)}$$

が、因子 B に関するタイプ II の平方和 $S_B^{(2)}$ となる。

交互作用の場合も同様に、主効果と交互作用を含むフルモデルの平方和を $S_{(A+B+A \times B)}$ とし、フルモデルから交互作用 $A \times B$ を除いたモデルの平方和の減少分

$$S_{A \times B}^{(2)} = S_{(A+B+A \times B)} - S_{(A+B)}$$

表 3.25 主効果を主体にしたタイプ II の平方和算出のための分散分析表

		— デザイン行列 X —							1. 分散分析表 (A+B)				
		y	x ₀	a ₁	b ₁	b ₂	a ₁ b ₁	a ₁ b ₂	自由度	変動	分散	分散比	
A ₁	B ₁	10	1	1	1	0	1	0	回帰	3	130.2286	43.4095	10.5363
		13	1	1	1	0	1	0	残差	10	41.2000	4.1200	
	B ₂	14	1	1	0	1	0	1	合計	13	171.4286		
		12	1	1	0	1	0	1					
		15	1	1	0	1	0	1					
		11	1	1	0	1	0	1					
A ₂	B ₃	22	1	1	-1	-1	-1	-1	回帰	2	114.0952	57.0476	10.9452
		19	1	1	-1	-1	-1	-1	残差	11	57.3333	5.2121	
	B ₁	15	1	-1	1	0	-1	0	合計	13	171.4286		
		14	1	-1	1	0	-1	0					
		16	1	-1	0	1	0	-1					
		18	1	-1	0	1	0	-1					
A ₂	B ₃	21	1	-1	-1	-1	1	1	回帰	1	21.4286	21.4286	1.7143
		18	1	-1	-1	-1	1	1	残差	12	150.0000	12.5000	
	B ₁	15	1	-1	1	0	-1	0	合計	13	171.4286		
		14	1	-1	1	0	-1	0					
		16	1	-1	0	1	0	-1					
		18	1	-1	0	1	0	-1					
				A	B	A×B		3. 分散分析表 (A)					
									自由度	変動	分散	分散比	
									回帰	1	21.4286	21.4286	1.7143
									残差	12	150.0000	12.5000	
									合計	13	171.4286		
									4. 分散分析表 (A+B+A×B)				
									自由度	変動	分散	分散比	
									回帰	5	145.4286	29.0857	8.9495
									残差	8	26.0000	3.2500	
									合計	13	171.4286		

が、交互作用 $A \times B$ に関するタイプ II の平方和 $S_{A \times B}^{(2)}$ となる。見方を変えれば、 $S_{(A+B)}$ に交互作用 $A \times B$ を加えた $S_{(A+B+A \times B)}$ の平方和の増分で、逐次（タイプ I）の平方和 $S_{A \times B}^{(1)}$ に一致する。

表 3.25 で得られた分散分析表のモデル（回帰）についての平方和（変動）に注目し、それらを組み合わせてタイプ II の平方和を計算した結果を表 3.26 に示す。

表 3.26 主効果を主体にしたタイプ II の平方和による分散分析表

要因	回帰の平方和	差し引く平方和		df	タイプ II	平均平方	F 値	p 値		
A	$S_{(A+B)}$	130.23	- $S_{(B)}$	114.10	= $S_{A}^{(2)}$	1	16.13	16.13	4.96	0.0565
B	$S_{(A+B)}$	130.23	- $S_{(A)}$	21.43	= $S_{B}^{(2)}$	2	108.80	54.40	16.74	0.0014
$A \times B$	$S_{(A+B+A \times B)}$	145.43	- $S_{(A+B)}$	130.23	= $S_{A \times B}^{(2)}$	2	15.20	7.60	2.34	0.1586
e						8	26.00	3.25		
T	S_T	171.43			合計	13	166.13	#		
										# タイプ II の平方和の合計は S_T に一致しない

タイプ II の平方和の合計は、総平方和 S_T に一致しないので、表 3.26 のような分散分析表で示すと計算ミスがあるかのように錯誤が起きてしまうので、モデル全体の分散分析表の下に、タイプ II の平方和の結果のみを別途示す配慮が必要である。

主効果と交互作用を平等に扱うタイプ III の平方和

タイプ II の平方和は、主効果のみのモデルと交互作用を含むモデルを段階的に適用したのであるが、タイプ III の平方和は、主効果と交互作用を平等に扱う。表 3.27 に示すようにタイプ III の平方和は、主効果と交互作用すべてを含むフルモデルの平方和を $S_{(A+B+A \times B)}$ とし、このフルモデルから、因子 A を除いたモデルの平方和を $S_{(B+A \times B)}$ とし、フルモデルからの平方和の減少分

$$S_A^{(3)} = S_{(A+B+A \times B)} - S_{(B+A \times B)}$$

が、因子 A のタイプ III の平方和 $S_A^{(3)}$ である。同様に、因子 B を除いたモデルの平方和を $S_{(A+A \times B)}$ とし、フルモデルの平方和からの平方和の減少分

$$S_B^{(3)} = S_{(A+B+A \times B)} - S_{(A+A \times B)}$$

が、因子 B のタイプ III の平方和である。

交互作用の場合も同様に、フルモデルの平方和を $S_{(A+B+A \times B)}$ から主効果 (A+B) を除いた平方和の減少分

$$S_{A \times B}^{(3)} = S_{(A+B+A \times B)} - S_{(A+B)}$$

が、交互作用 $A \times B$ のタイプ III の平方和である。見方を変えれば、 $S_{(A+B)}$ に交互作用 $A \times B$ を加えた $S_{(A+B+A \times B)}$ の増分となり、逐次（タイプ I）の平方和 $S_{A \times B}^{(1)}$ に一致する。

表 3.27 右に示した 4 通りの分散分析表は、Excel の回帰分析のパラメータの推定に関する部分を省略している。Excel の回帰分析における「入力 X 範囲」の設定に際し、複数のダミー変数が隣あっている必要があり、(A+A×B) の場合には、因子 B のダミー変数 (b_1, b_2) が中にあるので、このままでは設定できない。そのために、因子 A と交互作用 A×B のダミー変数を入れ替えたデザイン行列 X を別枠に設定して対応している。

表 3.27 主効果と交互作用を平等に扱うタイプ III の平方和算出のための分散分析表

		— デザイン行列 X —							1. 分散分析表 (A+B+A×B)				
		y	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2	自由度	変動	分散	分散比	
A ₁	B ₁	10	1	1	1	0	1	0	回帰	5	145.4286	29.0857	8.9495
		13	1	1	1	0	1	0	残差	8	26.0000	3.2500	
	B ₂	14	1	1	0	1	0	1	合計	13	171.4286		
		12	1	1	0	1	0	1	2. 分散分析表 (B+A×B)				
		15	1	1	0	1	0	1	自由度	変動	分散	分散比	
B ₃	11	1	1	0	1	0	1	回帰	4	132.3377	33.0844	7.6171	
	19	1	1	-1	-1	-1	-1	残差	9	39.0909	4.3434		
A ₂	B ₁	15	1	-1	1	0	-1	0	合計	13	171.4286		
		14	1	-1	1	0	-1	0	3. 分散分析表 (A+A×B)				
	B ₂	16	1	-1	0	1	0	-1	自由度	変動	分散	分散比	
		18	1	-1	0	1	0	-1	回帰	3	40.2286	13.4095	1.0221
		21	1	-1	-1	-1	1	1	残差	10	131.2000	13.1200	
B ₃	18	1	-1	-1	-1	1	1	合計	13	171.4286			
				A	B	A×B		4. 分散分析表 (A+B)					
		「3. 分散分析表 (A+A×B)」の場合には、							自由度	変動	分散	分散比	
		デザイン行列 X の並びを変えている。							回帰	3	130.2286	43.4095	10.5363
		x_0	a_1	a_1b_1	a_1b_2	b_1	b_2	残差	10	41.2000	4.1200		
		これは、Excel での変数の設定に際し、							合計	13	171.4286		
		連続していることが必須のためである。											

表 3.28 に示すのは、表 3.27 で計算されたそれぞれの分散分析表のモデル (回帰) の平方和 (変動) を取り出し、それらの差から、主効果と交互作用を平等に扱うタイプ III の平方和による分散分析表が作成されている。

表 3.28 主効果と交互作用を平等に扱うタイプ III の平方和による分散分析表

要因	回帰の平方和	差し引く平方和	df	タイプIII	平均平方	F値	p値
A	$S_{(A+B+A×B)}$ 145.43	- $S_{(B+A×B)}$ 132.34	$S_A^{(3)}$ 1	13.09	13.09	4.03	0.0797
B	$S_{(A+B+A×B)}$ 145.43	- $S_{(A+A×B)}$ 40.23	$S_B^{(3)}$ 2	105.20	52.60	16.18	0.0015
A×B	$S_{(A+B+A×B)}$ 145.43	- $S_{(A+B)}$ 130.23	$S_{A×B}^{(3)}$ 2	15.20	7.60	2.34	0.1586
e			8	26.00	3.25		
T	S_T 171.43		合計 13	159.49 #			
# タイプIII の平方和の合計は S_T に一致しない							

表 3.20 に示したタイプ I (逐次) の平方和, 表 3.26 に示したタイプ II の平方和と同様にタイプ III の平方和は, SAS の GLM プロシジャで定式化された呼び方である. どの平方和を使うことが妥当なのか, そもそも交互作用の存在を認めた場合に, それらの主効果の平方和を云々すること自体に意味がないとも言え, それらの使い分けを示すこと自体が難儀なことである. 大切なのは, 図 3.8 で示したように, 実験の目的に沿った結果の解釈に役立つ各種の推定であり, 分散分析表に示された結果の解釈は, 隔靴搔痒的な解釈しかできないとの現状を認識してもらいたい.

SAS の GLM プロシジャによるタイプ I, タイプ II, タイプ III の平方和

Excel の回帰分析を用い タイプ I, タイプ II, タイプ III の平方和の生成方法を示した. この結果が正しいことの検証に, タイプ I, タイプ II, タイプ III の平方和の本家であり, 無償で継続的に提供されている OnDemand SAS の GLM プロシジャを用いた結果と照合する. これらの 3 種の平方和は, GLM プロシジャの Model ステートメントのオプションで「SS1」, 「SS2」, 「SS3」を設定することで出力される. これまで, Excel の回帰分析を段階的に適用して求めた平方和と一致することを確認する.

SAS のプログラム 3.1

```
Title1 'TwoWay_不揃 ANOVA. sas 2022-5-22 Y.Takahashi ' ;

data d01 ;
  input A$ B$ @ ;
  do k = 1 to 4 ;
    input y @; output ;
  end ;
datalines ;
A1 B1 10 13 . .
A1 B2 14 12 15 11
A1 B3 22 19 . .
A2 B1 15 14 . .
A2 B2 16 18 . .
A2 B3 21 18 . .
;
proc glm data=d01 ;
  class A B ;
  model y = A B A*B / SS1 SS2 SS3;
run ;
```

表 3.29 に GLM プロシジャによる分散分析表を示す. 表 3.25 および表 3.27 に示した要因 (A+B+A×B) を含んだ Excel の回帰分析で得られた分散分析表に一致することが確認される.

表 3.29 GLM プロシジャによる分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	5	145.4286	29.0857	8.95	0.0039
Error	8	26.0000	3.2500		
Corrected Total	13	171.4286			

SAS の HTML 出力を Excel に貼り付け、書式を整えている。

表 3.30 に示すのは、タイプ I の平方和であり、表 3.20 に示した Excel の回帰分析の分散分析表から算出した逐次（タイプ I）平方和に一致する。

表 3.30 GLM プロシジャによるタイプ I の平方和

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
A	1	21.4286	21.4286	6.59	0.0332
B	2	108.8000	54.4000	16.74	0.0014
A*B	2	15.2000	7.6000	2.34	0.1586

表 3.31 に示すのは、タイプ II の平方和であり、表 3.26 に示した Excel の回帰分析の分散分析表の主効果を主体にしたタイプ II の平方和に一致する。

表 3.31 GLM プロシジャによるタイプ II の平方和

要因	自由度	Type II 平方和	平均平方	F 値	Pr > F
A	1	16.1333	16.1333	4.96	0.0565
B	2	108.8000	54.4000	16.74	0.0014
A*B	2	15.2000	7.6000	2.34	0.1586

表 3.33 に示すのは、タイプ III の平方和であり、表 3.28 に示した Excel の回帰分析の分散分析表の主効果と交互作用を平等に扱うタイプ III の平方和に一致する。

表 3.32 GLM プロシジャによるタイプ III の平方和

要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
A	1	13.0909	13.0909	4.03	0.0797
B	2	105.2000	52.6000	16.18	0.0015
A*B	2	15.2000	7.6000	2.34	0.1586

繰り返しが不揃いにもかかわらず各要因の平方和の計算結果を表 3.16 に示したのであるが、GLM プロシジャの 3 種の平方和と対比した結果を表 3.33 に示す。因子 A についてのみ、タイプ I の平方和に一致するが、他に一致する平方和を見いだすことができない。

表 3.33 平方和の計算結果と GLM プロシジャによる各種の平方和との対比

要因	自由度	Type I		Type II		Type III	平方和の計算	表3.16
A	1	21.4286	≠	16.1333	≠	13.0909	21.4286	
B	2	108.8000	=	108.8000	≠	105.2000	114.0952	不一致
A*B	2	15.2000	=	15.2000	=	15.2000	16.0952	不一致
	計	145.4286					151.6190	

(1, -1) 対比型ダミー変数を用いた Excel の回帰分析では、GLM プロシジャのように手早く各種の平方和を求めることはできないが、若干の手間により GLM プロシジャと同等の結果を得られことを示した。GLM プロシジャの世界だけでは、理解が上滑りがちになりやすいが、Excel での回帰分析を適用することにより、理解が深められることを期待する。

対比によるタイプ III の計算

表 3.28 に示したようにタイプ III の平方和は、全変数を含んだモデル（回帰）の平方和から、求めたい因子あるいは交互作用について除いた場合の平方和の差分から求められることを示した。ただし、SAS の GLM プロシジャでは、平方和の差分ではなく、求めたい因子あるいは交互作用に関する対比を用いて平方和を直接推定している。この方法は、一線形モデルにおける平方和の計算方法としても知られているので、Excel の行列関数を用いて計算方法を示す。

1) 対比による因子 A に関するタイプ III の平方和

A₁ と A₂ の差のベクトル $l_{A_2-A_1}$ を表 3.34 に示す。平方和の計算のために $L_A = l_{A_2-A_1}$ と置き直すと

表 3.34 因子 A の水準間の差のベクトル

		l_0	l_1	l_2	l_3	l_4	l_5
		x_0	a_1	b_1	b_2	a_1b_1	a_1b_2
	l_{A1}	1	1	0	0	0	0
	l_{A2}	1	-1	0	0	0	0
L_A	差 l_{A1-A2}	0	2	0	0	0	0

表 3.17 から、フルモデルの推定値は、 $\hat{\theta} = [16.0 \ -1.0 \ -3.0 \ -1.0 \ -0.5 \ -1.0]^T$ なので、 $L_A \hat{\theta} = 2 \times (-1.0) = -2$ であり、 $L_A (X^T X)^{-1} L_A^T = (-2) \times 0.07638889 \times (-2)$ と共にスカラーになるので、式 (3.48) に示すように $S_A^{(3)} = 13.0909$ となり、タイプ III の平方和に一致する。

$$\left. \begin{aligned}
 SS(L_A \theta) &= (L_A \theta)^T [L_A (X^T X)^{-1} L_A^T]^{-1} (L_A \theta) \\
 &= [2 \times (-1.00)] \times [2 \times 0.07638889 \times 2]^{-1} [2 \times (-1.00)] \\
 &= (-2)^2 / (4 \times 0.07638889) \\
 &= 13.0909
 \end{aligned} \right\} \quad (3.48)$$

表 3.35 対比による平方和の計算に必要な行列とパラメータの推定値

	$(X^T X)$						$(X^T X)^{-1}$						θ^{\wedge}
θ^{\wedge}_0	14	2	0	2	0	2	0.0764	-0.0069	0.0069	-0.0139	0.0069	-0.0139	16.00
θ^{\wedge}_1	2	14	0	2	0	2	-0.0069	0.0764	0.0069	-0.0139	0.0069	-0.0139	-1.00
θ^{\wedge}_2	0	0	8	4	0	0	0.0069	0.0069	0.1597	-0.0694	-0.0069	0.0139	-3.00
θ^{\wedge}_3	2	2	4	10	0	2	-0.0139	-0.0139	-0.0694	0.1389	0.0139	-0.0278	-1.00
θ^{\wedge}_4	0	0	0	0	8	4	0.0069	0.0069	-0.0069	0.0139	0.1597	-0.0694	-0.50
θ^{\wedge}_5	2	2	0	2	4	10	-0.0139	-0.0139	0.0139	-0.0278	-0.0694	0.1389	-1.00
	x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$	x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$	

2) 対比による因子 B に関するタイプ III の平方和

表 3.36 に示すように因子 B については、 B_1 水準と B_3 水準との差、 B_2 水準と B_3 水準との差 ($l_{B_1-B_2}, l_{B_1-B_3}$) を合わせて L_B ととする。

表 3.36 因子 B の水準間の差のベクトル

		l_0	l_1	l_2	l_3	l_4	l_5
		x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$
	l_{B_1}	1	0	1	0	0	0
	l_{B_2}	1	0	0	1	0	0
	l_{B_3}	1	0	-1	-1	0	0
L_B	差 $l_{B_1-B_3}$	0	0	2	1	0	0
	差 $l_{B_2-B_3}$	0	0	-1	2	0	0

計算結果は、以下に示すように 2×2 の行列計算に帰着し、平方和として $S_B^{(3)} = 105.2000$ を得る。

$$L_B \theta = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -3.00 \\ -1.00 \end{bmatrix} = \begin{bmatrix} -7.00 \\ 1.00 \end{bmatrix}$$

$$\begin{aligned} (L_B (X^T X)^{-1} L_B^T)^{-1} &= \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0.1597 & -0.0694 \\ -0.0694 & 0.1389 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}^T^{-1} \\ &= \begin{bmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{bmatrix}^{-1} \end{aligned}$$

$$\begin{aligned} (L_B \theta)^T (L_B (X^T X)^{-1} L_B^T)^{-1} (L_B \theta) &= \begin{bmatrix} -7.00 \\ 1.00 \end{bmatrix}^T \begin{bmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{bmatrix}^{-1} \begin{bmatrix} -7.00 \\ 1.00 \end{bmatrix} \\ &= 105.2000 \end{aligned}$$

3) 対比による交互作用 A×B に関するタイプ III の平方和

表 3.37 に交互作用 A×B について $A_1 B_1$ および $A_1 B_2$ について $A_1 B_3$ との差のベクトルを示

す. 因子 B と同様に, 計算結果は, 以下に示すように 2×2 の行列計算に帰着し平方和として, $S_{A \times B}^{(3)} = 15.2000$ を得る.

表 3.37 交互作用 A×B の水準間の差のベクトル

		l_0	l_1	l_2	l_3	l_4	l_5
		x_0	a_1	b_1	b_2	$a_1 b_1$	$a_1 b_2$
	$l_{A1 \times B1}$	1	0	0	0	1	0
	$l_{A1 \times B2}$	1	0	0	0	0	1
	$l_{A1 \times B3}$	1	0	0	0	-1	-1
$L_{A \times B}$	差 $l_{A \times B1-B3}$	0	0	0	0	2	1
	差 $l_{A \times B2-B3}$	0	0	0	0	-1	2

$$L_{A \times B} \theta = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} -0.50 \\ -1.00 \end{pmatrix} = \begin{pmatrix} -2.00 \\ -1.50 \end{pmatrix}$$

$$(L_{A \times B} (X^T X)^{-1} L_{A \times B}^T)^{-1} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 0.1597 & -0.0694 \\ -0.0694 & 0.1389 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{pmatrix}^{-1}$$

$$(L_{A \times B} \theta)^T (L_{A \times B} (X^T X)^{-1} L_{A \times B}^T)^{-1} (L_{A \times B} \theta) = \begin{pmatrix} -2.00 \\ -1.50 \end{pmatrix}^T \begin{pmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{pmatrix}^{-1} \begin{pmatrix} -2.00 \\ -1.50 \end{pmatrix} = 15.2000$$

以上の結果を表 3.38 に, タイプ III の平方和としてまとめて示す. この結果は, 表 3.32 に示した SAS の GLM プロシジャによるタイプ III の平方和に一致する.

表 3.38 対比の計算によるタイプ III の平方和

要因	df					df	タイプ III		
A	1	$S_A = (L_A \theta^{\wedge})^T (L_A (X^T X) L_A^T)^{-1} (L_A \theta^{\wedge})$			$S_{(A)}^{(3)}$	1	13.0909		
B	2	$S_B = (L_B \theta^{\wedge})^T (L_B (X^T X) L_B^T)^{-1} (L_B \theta^{\wedge})$			$S_{(B)}^{(3)}$	2	105.2000		
A×B	2	$S_{A \times B} = (L_{A \times B} \theta^{\wedge})^T (L_{A \times B} (X^T X) L_{A \times B}^T)^{-1} (L_{A \times B} \theta^{\wedge})$			$S_{(A \times B)}^{(3)}$	2	15.2000		
e	13	$S_{(T)}$	171.4286	-	$S_{(A+B+A \times B)}$	145.4286	= $S_{(e)}^{(3)}$	8	26.0000
T		各平方和の合計は, $S_T = 171.4286$ に一致しない					13	159.4909	

タイプ III の平方和の使い方

統計ソフトでタイプ III の平方和の計算は, 以上のように Excel の行列計算によって求められる. ただし, Excel の行列関数を使ってこのような平方和の計算を求めても, 他に応用するすべは見当たらない. 特に, 交互作用の存在に否定的な場合に, 該当する交互作用をモデ

ルから外した場合に、該当する主効果の平方和が変化するのみならず、他の要因の平方和も少なからず変化する。

そもそも、分散分析表は、要因（因子）全体に対して統計的な判断には役立つが、結果の考察に際しては、水準間の差の95%信頼区間などが重要であり、分散分析表だけの結果の考察は不十分である。どうしてもタイプ III の平方和が必要とする場合は、無償で継続的に提供されている SAS の GLM プロシジャを使うことを推奨する。R については、タイプ III の平方和を出すための方法が Web 上で散見するが、結果の妥当性は未確認である。

大切なのは、Excel の回帰分析に限らず回帰パラメータとパラメータの共分散行列 $\Sigma(\hat{\theta})$ または $\Sigma(\hat{\beta})$ から計算する各種の水準間の差の推定と 95%信頼区間の推定である。伝統的な実験計画法とそれに引き続く平方和の分解による分散分析表の作成手順は、因子が互いに直交している場合に限ってしか適用できない限定的な方法であり、発展性に欠けることを自覚しなければならない。

現行水準と最適水準の差の 95%信頼区間

GLM プロシジャでは、プログラム 3.2 に示すように model ステートメントに引き続き lsmeans ステートメントによる各種の最小 2 乗平均の推定のみならず、estimate ステートメントにより現行水準と最適水準の差の推定も行える。

SAS のプログラム 3.2

```
proc glm data=d01 ;
  class A B ;
  model y = A B A*B ;
  estimate "A1 B1"      intercept 1  A 1 0  B  1 0 0  A*B  1 0 0 0 0 0 ;
  estimate "A1 B3"      intercept 1  A 1 0  B  0 0 1  A*B  0 0 1 0 0 0 ;
  estimate "A1B3 - A1B1" intercept 0  A 0 0  B -1 0 1  A*B -1 0 1 0 0 0 ;
run ;
```

表 3.39 に示すように、現行水準と最適水準の推定値と標準誤差が出力され、それらの差に対する標準誤差も出力される。95%信頼区間の出力はないので、Excel シート上で追加している。この結果は、表 3.24 の線形和 $L^{(7)} = 9.00$ 、(4.84, 13.16) に一致している。

表 3.39 現行水準と最適水準の推定値および差の 95%信頼区間

パラメータ	推定値	標準誤差	t 値	Pr > t	95%信頼区間	
					L95%	U95%
A1 B1	11.50	1.2748	9.02	<.0001	8.56	14.44
A1 B3	20.50	1.2748	16.08	<.0001	17.56	23.44
A1B3 - A1B1	9.00	1.8028	4.99	0.0011	4.84	13.16
					$t_{0.05}(14-6)=$	2.3060

4. 欠測値がある直交表の線形モデルによる解析

「線形推定論」は、実験データの解析法としての分散分析に対する理論的な枠組みを与えてきた。ただし、朝香・石川・山口 監修 (1988), 「新版 品質管理便覧 第2版」には、「線形推定論」を見いだすことができない。幸いなことに、楠・辻谷・松本・和田 (1995), 「応用 実験計画法」の「第6章 線形モデル」の第2節に「線形推定論」が含まれている。伝統的な線形推定論では、各種の実験デザインで取り上げる因子をダミー変数するプロセスがきちっと導入され、デザイン行列、正規方程式、逆行列などをベースにした線形モデル入門となっている。各種の実験モデルの解析は、線形モデルを拡張することによりすべて対応できることを高橋・大橋・芳賀 (1989), 「SASによる実験データの解析」で示してきた。あらためて、ダミー変数の数理およびそれらを活用したデザイン行列を活用した解析を Excel の行列計算を用いて実用的に活用できるようにすると共に、最先端の統計ソフト JMP を用いて検証し、先駆的な SAS の GLM プロシジャで導入された各種の平方和についても示す。

4.1. 繰り返しのない2因子実験データ

構造モデル・回帰モデル・線形モデル

表 4.1 に示すのは、繰り返しのない2因子実験のデータである。データの構造式は、一般平均 μ と要因効果を示す母数 α_i, β_j および誤差 ε_{ij} の1次式

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{Normal}(0, \sigma^2) \quad (4.1)$$

で表すことができる。このモデルは、構造モデルまたは DE (Design of Experiment) モデルとも言われている。

表 4.1 2因子実験のデータ [楠ら (1995), 表 6.5, p191]

	B ₁	B ₂
A ₁	10.9	11.5
A ₂	12.4	12.9
A ₃	12.2	12.3

データ y_i が実験条件を表す連続変数 x_i と直線関係にある場合には、 y_i を x_i の一次式

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{Normal}(0, \sigma^2) \quad (4.2)$$

となる。これは、回帰モデルと呼ばれている。式 (4.1), 式 (4.2) の共通点は、

- 1) データの構造式が、式 (4.1) の母数 μ, α_i, β_j あるいは式 (4.2) のパラメータ β_0, β_1 と、確率変数である誤差を加えて1次式で表わされる。
- 2) 誤差に、独立性、等分散性、不偏性および正規性の仮定を設けている。

ことなどである。

式 (4.1), 式 (4.2) は、一般的に

$$y_i = \theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i} + \varepsilon_i, \quad \varepsilon_i \sim \text{Normal}(0, \sigma^2) \quad (4.3)$$

と置き換えられ、これを線形モデルと呼ぶことにする。ただし、式 (4.3) で $x_{0,i}$ をすべて1とした場合に、データの構造式の場合 θ_0 は、一般平均 μ となり、 $(\theta_0, \theta_1, \dots, \theta_p)$ は、式 (4.1) における各要因の主効果あるいは交互作用であり、式 (4.2) の回帰分析では、パラメータ (係数) と呼ばれる $(\beta_0, \beta_1, \dots, \beta_p)$ に対応し、 β_0 は切片となる。

式 (4.1) において、 3×2 の繰返しのない2因子実験のデータの構造式は、

$$\left. \begin{aligned} y_{11} &= \mu + \alpha_1 && + \beta_1 && + \varepsilon_{11} \\ y_{12} &= \mu + \alpha_1 && + \beta_2 && + \varepsilon_{12} \\ y_{21} &= \mu &+ \alpha_2 &+ \beta_1 && + \varepsilon_{21} \\ y_{22} &= \mu &+ \alpha_2 &+ \beta_2 && + \varepsilon_{21} \\ y_{31} &= \mu && + \alpha_3 + \beta_1 && + \varepsilon_{31} \\ y_{32} &= \mu && + \alpha_3 &+ \beta_2 &+ \varepsilon_{32} \end{aligned} \right\} \quad (4.4)$$

として展開される。これは、式 (4.3) において $p=5$ とし、母数 $(\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ をパラメータ $(\theta_0, \theta_1, \theta_2, \dots, \theta_5)$ に対応付け、それらの変数を $(x_{0,i}, x_{1,i}, x_{2,i}, \dots, x_{5,i})$ とした線形モデル

$$\left. \begin{aligned} y_1 &= \theta_0 x_{0,1} + \theta_1 x_{1,1} + \theta_2 x_{2,1} + \cdots + \theta_5 x_{5,1} + \varepsilon_1 \\ y_2 &= \theta_0 x_{0,2} + \theta_1 x_{1,2} + \theta_2 x_{2,2} + \cdots + \theta_5 x_{5,2} + \varepsilon_2 \\ &\vdots \\ y_6 &= \theta_0 x_{0,6} + \theta_1 x_{1,6} + \theta_2 x_{2,6} + \cdots + \theta_5 x_{5,6} + \varepsilon_6 \end{aligned} \right\} \quad (4.5)$$

で表わすことができる。さて、式 (4.4) と式 (4.5) を比べると、式 (4.5) の変数 $(x_{1,i}, x_{2,i}, \dots, x_{5,i})$ の間には、

$$\left. \begin{aligned} x_{1,i} + x_{2,i} + x_{3,i} &= 1, && i=1, 2, \dots, 6 \\ x_{4,i} + x_{5,i} &= 1, && i=1, 2, \dots, 6 \end{aligned} \right\} \quad (4.6)$$

のように1次従属の関係がある。構造 (DE) モデルにおける母数に関する制約

$$\alpha_1 + \alpha_2 + \alpha_3 = 0, \quad \beta_1 + \beta_2 = 0 \quad (4.7)$$

を用い、最後の水準について整理すると

$$\alpha_3 = -\alpha_1 - \alpha_2, \quad \beta_1 = -\beta_2 \quad (4.8)$$

となり、これらを式 (4.4) に代入すると

$$\left. \begin{aligned} y_{11} &= \mu + \alpha_1 && + \beta_1 + \varepsilon_{11} \\ y_{12} &= \mu + \alpha_1 && - \beta_1 + \varepsilon_{12} \\ y_{21} &= \mu && + \alpha_2 + \beta_1 + \varepsilon_{21} \\ y_{22} &= \mu && + \alpha_2 - \beta_1 + \varepsilon_{22} \\ y_{31} &= \mu - \alpha_1 - \alpha_2 + \beta_1 + \varepsilon_{31} \\ y_{32} &= \mu - \alpha_1 - \alpha_2 - \beta_1 + \varepsilon_{32} \end{aligned} \right\} \quad (4.9)$$

のように書き換えられる。

線形モデルに対する正規方程式

線形モデルの一般式である式 (4.3) を、データ数が n の線形モデル式に展開する。この線形モデルに基づくパラメータ θ_i の推定方法について示す。

$$\left. \begin{aligned} y_1 &= \theta_0 x_{0,1} + \theta_1 x_{1,1} + \cdots + \theta_p x_{p,1} + \varepsilon_1 \\ &\vdots \\ y_i &= \theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i} + \varepsilon_i \\ &\vdots \\ y_n &= \theta_0 x_{0,n} + \theta_1 x_{1,n} + \cdots + \theta_p x_{p,n} + \varepsilon_n \end{aligned} \right\} \quad (4.10)$$

式 (4.10) を誤差 ε_i について

$$\varepsilon_i = y_i - (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i}) \quad (4.11)$$

として、解いたとき、 $i=1, 2, \dots, n$ に関する残差平方和 Q

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [y_i - (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})]^2 \quad (4.12)$$

を最小にする $(\theta_0, \theta_1, \dots, \theta_p)$ を求めたい。このような方法を最小2乗法と呼び、そのとき得られる解を $(\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p)$ とする。

解を求めるためには、式 (4.12) の残差平方和 Q をすべての θ_k で偏微分して、0とおいた式

$$\frac{\partial Q}{\partial \theta_k} = 0, \quad k = 0, 1, 2, \dots, p \quad (4.13)$$

すなわち、

$$\left. \begin{aligned} \sum_{i=1}^n [x_{0,i} y_i - x_{0,i} (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})] &= 0 \\ \cdots \cdots \cdots \\ \sum_{i=1}^n [x_{k,i} y_i - x_{k,i} (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})] &= 0 \\ \cdots \cdots \cdots \\ \sum_{i=1}^n [x_{p,i} y_i - x_{p,i} (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})] &= 0 \end{aligned} \right\} \quad (4.14)$$

となる. さらに, 式 (4.14) を θ_k に関して解き, その解を $(\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p)$ とする. なお, 式の導出の詳細は, 第 5.3 節を参照のこと.

ここで, 簡単のため,

$$\left. \begin{aligned} a_{k,k'} &= \sum_{i=1}^n x_{k,i} x_{k',i}, & k=0, 1, 2, \dots, p, & \quad k'=0, 1, 2, \dots, p \\ B_k &= \sum_{i=1}^n x_{k,i} y_i, & k=0, 1, 2, \dots, p \end{aligned} \right\} \quad (4.15)$$

とおくと, 式 (4.14) は,

$$\left. \begin{aligned} \hat{\theta}_0 a_{0,0} + \hat{\theta}_1 a_{0,1} + \dots + \hat{\theta}_p a_{0,p} &= B_0 \\ \dots & \\ \hat{\theta}_0 a_{k,0} + \hat{\theta}_1 a_{k,1} + \dots + \hat{\theta}_p a_{k,p} &= B_k \\ \dots & \\ \hat{\theta}_0 a_{p,0} + \hat{\theta}_1 a_{p,1} + \dots + \hat{\theta}_p a_{p,p} &= B_p \end{aligned} \right\} \quad (4.16)$$

となり, これを正規方程式と呼ぶ. この解を, 最小 2 乗推定量といい, 線形不偏推定量の中で分散最小の最良線形不偏推定量 (BLUE: best linear unbiased estimator) になる.

デザイン行列 X を用いた正規方程式

母数に無駄のない構造 (DE) モデルの式 (4.9) を, 線形モデル式で表すと

$$\left. \begin{aligned} y_1 &= \theta_0 x_{0,1} + \theta_1 x_{1,1} + \theta_2 x_{2,1} + \theta_3 x_{3,1} + \varepsilon_1 \\ y_2 &= \theta_0 x_{0,2} + \theta_1 x_{1,2} + \theta_2 x_{2,2} + \theta_3 x_{3,2} + \varepsilon_2 \\ &\vdots \\ y_6 &= \theta_0 x_{0,6} + \theta_1 x_{1,6} + \theta_2 x_{2,6} + \theta_3 x_{3,6} + \varepsilon_6 \end{aligned} \right\} \quad (4.17)$$

となる. 行列を用いると, 表 4.2 のように表わすことができる.

表 4.2 は, デザイン行列 X の各行に列ベクトル θ を掛けて加える. すなわち, 第 i 行について行と列の積和は,

$$\theta_0 x_{0,i} + \theta_1 x_{1,i} + \theta_2 x_{2,i} + \theta_3 x_{3,i} + \varepsilon_i = y_i \quad (4.18)$$

と表すことができる.

表 4.2 デザイン行列 X を用いた線形モデル

			— デザイン行列 X —				パラメータ		誤差	反応	
A	B	i	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	θ		ε	=	y
A ₁	B ₁	1	1	1	0	1	θ_0	+	ε_1	=	y_1
	B ₂	2	1	1	0	-1	θ_1		ε_2		y_2
A ₂	B ₁	3	1	0	1	1	θ_2		ε_3		y_3
	B ₂	4	1	0	1	-1	θ_3		ε_4		y_4
A ₃	B ₁	5	1	-1	-1	1			ε_5		y_5
	B ₂	6	1	-1	-1	-1			ε_6		y_6

式 (4.16) に示した正規方程式は、表 4.3 に示すように行列の表記を用いて表わすことができる。転置行列 X^T は、Excel の行列関数 $X^T = \text{Transpose}(X \text{ の範囲})$ で求めことができ、デザイン行列 X の行方向を列方向に、列方向を行方向に並び変えた行列である。行列の積 $(X^T X)$ は、 X^T の行方向と X の列方向の積和であり、その要素は、式 (4.15) の $a_{k,k'}$ となっている。式 (4.15) の B_k は、転置行列 X^T と列ベクトル $X^T Y$ の積 $X^T Y$ の各セルに対応する。

表 4.3 デザイン行列 X を用いた正規方程式

	— 転置行列 X^T —						デザイン行列 X 推定値					— 転置行列 X^T —						Y	
	1	2	3	4	5	6	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	θ^{\wedge}	1	2	3	4	5	6	y	y_i
$x_{0,i}$	1	1	1	1	1	1	1	1	0	1	$\theta^{\wedge}_0 =$	1	1	1	1	1	1	10.9	y_1
$x_{1,i}$	1	1	0	0	-1	-1	1	1	0	-1	θ^{\wedge}_1	1	1	0	0	-1	-1	11.5	y_2
$x_{2,i}$	0	0	1	1	-1	-1	1	0	1	1	θ^{\wedge}_2	0	0	1	1	-1	-1	12.4	y_3
$x_{3,i}$	1	-1	1	-1	1	-1	1	0	1	-1	θ^{\wedge}_3	1	-1	1	-1	1	-1	12.9	y_4
							1	-1	-1	1								12.2	y_5
							1	-1	-1	-1								12.3	y_6

— $X^T X (a_{k,k'})$ —				θ^{\wedge}	$X^T Y$	B_k
6.0	0	0	0	$\theta^{\wedge}_0 =$	72.20	B_0
0	4.0	2.0	0	θ^{\wedge}_1	-2.10	B_1
0	2.0	4.0	0	θ^{\wedge}_2	0.80	B_2
0	0	0	6.0	θ^{\wedge}_3	-1.20	B_3

正規方程式の解

表 4.4 に示すように、正規方程式の解は、表 4.3 の下段の正規方程式の両辺に $(X^T X)$ の逆行列 $(X^T X)^{-1}$ を掛けることにより得られる。逆行列の計算は、伝統的に煩雑な「掃き出し法」が必要とされていたのであるが、Excel の行列関数 $\text{Minverse}(X^T X \text{ の範囲})$ によって直

表 4.4 デザイン行列 X を用いた正規方程式の解

$(X^T X)^{-1}$	$(c_{k,k'})$	$X^T X$	$(a_{k,k'})$	θ^{\wedge}	$(X^T X)^{-1}$	$(c_{k,k'})$	$X^T Y$
0.167	0	6.0	0 0 0	$\theta^{\wedge}_0 =$	0.167	0	72.20
0	0.333	0	4.0 2.0 0	θ^{\wedge}_1	0	0.333	-2.10
0	-0.167	0	2.0 4.0 0	θ^{\wedge}_2	0	-0.167	0.80
0	0	0	0 6.0	θ^{\wedge}_3	0	0	-1.20
$I = (X^T X)^{-1} (X^T X)$				θ^{\wedge}	$(X^T X)^{-1}$	$(c_{k,k'})$	$X^T Y$
1.0	0	6.0	0 0 0	$\theta^{\wedge}_0 =$	0.167	0	72.20
0	1.0	0	4.0 2.0 0	θ^{\wedge}_1	0	0.333	-2.10
0	0	1.0	2.0 4.0 0	θ^{\wedge}_2	0	-0.167	0.80
0	0	0	0 6.0	θ^{\wedge}_3	0	0	-1.20
				θ^{\wedge}	$(X^T X)^{-1} X^T Y$		
				μ^{\wedge}	$\theta^{\wedge}_0 =$	12.033	
				α^{\wedge}_1	θ^{\wedge}_1	-0.833	
				α^{\wedge}_2	θ^{\wedge}_2	0.617	
				β^{\wedge}_1	θ^{\wedge}_3	-0.200	

接計算できるようになった。デザイン行列 \mathbf{X} を使うことによりシグマ表記の煩雑な計算法を行列表記で簡潔に示せるだけでなく、実際の計算にもそのまま使えることができるようになったことは、うれしいかぎりである。

逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ の各要素を $c_{k,k'}$ 、 $(\mathbf{X}^T \mathbf{Y})$ の各要素を B_k とすれば、 $\hat{\theta}_k$ は、

$$\hat{\theta}_k = c_{k,0} B_0 + c_{k,1} B_1 + c_{k,2} B_2 + c_{k,3} B_3 \quad (4.19)$$

と線形推定論で定式化されている式となる。逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ と $(\mathbf{X}^T \mathbf{X})$ の積は、対角要素が 1 の単位行列 \mathbf{I} となり、ベクトル $\hat{\theta}$ との積は、元の $\hat{\theta}$ がそのまま残る。右辺の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ と $(\mathbf{X}^T \mathbf{Y})$ の積が、 θ の推定値 $\hat{\theta}$ となる。回帰モデルでは、 $\hat{\theta}$ に代わりに $\hat{\beta}$ を用いて式 (4.19) を次式

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (4.20)$$

で表わすのが一般的である。式 (4.16) および式 (4.19) は、手計算時代における正規方程式の解を求めるために工夫された表記法である。

手計算時代であっても、繰り返しが揃っている場合の正規方程式の解は、パズル的な代数計算によって $(\mathbf{X}^T \mathbf{X})$ の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ を求めることができる。ただし、欠測値などがあり繰り返しが不揃いになった場合は、コンピュータを用いた逆行列の計算が必要不可欠であり、楠ら(1995)の「第7.3節 一般の重回帰」に、「Gauss-Doolittle法による逆行列の計算」の詳細が表として示されている。また、新村(1983)、「重回帰分析における掃き出し演算子」も計算過程が詳細に示されている。なお、Excelで逆行列の計算をする際には、計算精度を保つために15桁の有効数字(倍精度実数)となっていることを確認すること。

パラメータの共分散行列の活用

このようにして得られた最小2乗推定量 $\hat{\theta}_k$ の期待値、分散および共分散は、

$$\left. \begin{aligned} E(\hat{\theta}_k) &= \hat{\theta}_k \\ \text{Var}(\hat{\theta}_k) &= c_{k,k} \hat{\sigma}^2 \\ \text{Cov}(\hat{\theta}_k, \hat{\theta}_{k'}) &= c_{k,k'} \hat{\sigma}^2 \end{aligned} \right\} \quad (4.21)$$

となる。式 (4.12) の θ_k に $\hat{\theta}_k$ を代入して得られる残差平方和 S_e

$$\left. \begin{aligned} S_e &= \sum_{i=1}^n [y_i - (\hat{\theta}_0 x_{0,i} + \hat{\theta}_1 x_{1,i} + \cdots + \hat{\theta}_p x_{p,i})]^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned} \right\} \quad (4.22)$$

について、(自由度 = データ数 - 独立な母数の個数) により、

$$E[S_e] = (n - p - 1) \sigma^2 \quad (4.23)$$

となる。ゆえに、誤差分散 σ^2 の不偏推定量は、

$$\hat{\sigma}^2 = \frac{S_e}{n - p - 1} \quad (4.24)$$

で与えられる. なお, 残差平方和を S_e としているが, 本来は S_ε とすべきところを, 表記の簡単化のための $\hat{\varepsilon}$ を e と置き換えている. このような置き換えは, β の推定値 $\hat{\beta}$ とすべきところを, 簡単化のために b と置き換えるような表記法と同様である.

誤差分散の推定値 $\hat{\sigma}^2$ の推定のために表 4.5 に示すように, $\hat{y} = X\hat{\theta}$ により予測値 \hat{y} , 残差 $\hat{\varepsilon} = y - \hat{y}$ の平方和から $S_e = 0.0700$ を求め, 残差の自由度 $df = n - (p + 1) = 6 - (3 + 1) = 2$ で割って, 誤差分散の推定値 $\hat{\sigma}^2 = 0.0350$ が得られる.

表 4.5 予測値 \hat{y} の推定および誤差分散の推定値 $\hat{\sigma}^2$

A	B	i	デザイン行列 X				予測値 y^\wedge の推定		誤差分散 σ^2 の推定					
			$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	θ^\wedge	y^\wedge	y	y^\wedge	$\varepsilon^\wedge = y - y^\wedge$			
A ₁	B ₁	1	1	1	0	1	12.033	=	11.000	10.9	-	11.000	=	-0.100
	B ₂	2	1	1	0	-1	-0.833		11.400	11.5		11.400		0.100
A ₂	B ₁	3	1	0	1	1	0.617		12.450	12.4		12.450		-0.050
	B ₂	4	1	0	1	-1	-0.200		12.850	12.9		12.850		0.050
A ₃	B ₁	5	1	-1	-1	1			12.050	12.2		12.050		0.150
	B ₂	6	1	-1	-1	-1			12.450	12.3		12.450		-0.150
									$n =$	6		$S_e =$	0.0700	
									$p =$	3		$\hat{\sigma}^2 =$	0.0350	
									$df = n - p - 1 =$	2		$\hat{\sigma}^\wedge =$	0.1871	

デザイン行列 X とパラメータの推定値 $\hat{\theta}$ の積 $X\hat{\theta}$ は, 表 4.6 に示すように式 (4.9) に示した構造 (DE) モデルに対応する.

表 4.6 予測値の構造 (DE) モデルによる推定

A	B	i	θ^\wedge_0 μ^\wedge	θ^\wedge_1 α^\wedge_1	θ^\wedge_2 α^\wedge_2	θ^\wedge_3 β^\wedge_1	y^\wedge
A ₁	B ₁	1	12.033	+ -0.833	+ 0	+ -0.200	= 11.000
	B ₂	2	12.033	-0.833	0	0.200	11.400
A ₂	B ₁	3	12.033	0	0.617	-0.200	12.450
	B ₂	4	12.033	0	0.617	0.200	12.850
A ₃	B ₁	5	12.033	0.833	-0.617	-0.200	12.050
	B ₂	6	12.033	0.833	-0.617	0.200	12.450

推定されたパラメータ $\hat{\theta}$ の分散および共分散は, 式 (4.21) で示されているのであるが, 行列として扱いたい. 誤差分散は, 表 4.5 で $\hat{\sigma}^2 = 0.0350$ と推定されているので, $c_{k,k'}$ の行列

表 4.7 パラメータの共分散行列 $\Sigma(\hat{\theta})$

	$(X^T X)^{-1}$	$(c_{k,k'})$	σ^2	パラメータの共分散行列 $\Sigma(\hat{\theta})$					
θ^\wedge_0	0.1667	0	0	0	0.0058	0	0	0	θ^\wedge_0
θ^\wedge_1	0	0.3333	-0.1667	0	0	0.0117	-0.0058	0	θ^\wedge_1
θ^\wedge_2	0	-0.1667	0.3333	0	0	-0.0058	0.0117	0	θ^\wedge_2
θ^\wedge_3	0	0	0	0.1667	0	0	0	0.0058	θ^\wedge_3
	μ^\wedge	α^\wedge_1	α^\wedge_2	β^\wedge_1		μ^\wedge	α^\wedge_1	α^\wedge_2	β^\wedge_1

$(X^T X)^{-1}$ に $\hat{\sigma}^2$ を掛けてパラメータの共分散行列 $\Sigma(\hat{\theta})$ を求めることができる。共分散行列は、一般的にデータに関する共分散行列 $\Sigma(\mathbf{x})$ が連想されるので、パラメータの共分散行列は、 $\Sigma(\hat{\theta})$ または $\Sigma(\hat{\beta})$ と表記し、明確に区別する必要がある。

計算されたパラメータの共分散行列 $\Sigma(\hat{\theta})$ から、 $\hat{\theta}_1$ の分散は

$$\text{Var}(\hat{\theta}_1) = c_{1,1} \hat{\sigma}^2 = 0.3333 \times 0.0350 = 0.0117$$

となり、 $\hat{\theta}_1$ と $\hat{\theta}_2$ の共分散は、

$$\text{Cov}(\hat{\theta}_1, \hat{\theta}_2) = c_{1,2} \hat{\sigma}^2 = -0.1667 \times 0.0350 = -0.0058$$

となる。

線形和の推定と95%信頼区間

実験計画法に基づいた実験データの解析上の最大のメリットは、取り上げた多くの因子の組合せデータから、ある注目する因子に対し、他の因子の影響を統計的に取り除いて、あたかも単独で繰り返した実験データのごとく水準の平均と標準誤差を推定することができ、さらに95%信頼区間も精度よく推定することができることにある。

線形推定論によって求められる線形和とその95%信頼区間の推定方法は、定式化はされてはいるが、行列を用いた計算は複雑であり、手計算で対応できる範囲を超えている。ここでは、線形推定論の原点に戻り、Excelの行列関数を用いた計算手順を示す。

線形推定論による定式化は、次のようである。母数(パラメータ) $\theta_0, \theta_1, \dots, \theta_q$ ($q \leq p$) に対する任意の1次式

$$L = \sum_{k=0}^q l_k \theta_k \tag{4.25}$$

の推定量は、

$$\hat{L} = \sum_{k=0}^q l_k \hat{\theta}_k \tag{4.26}$$

となる。その分散は、

$$\text{Var}(\hat{L}) = \sum_k \sum_{k'} l_k l_{k'} c_{k,k'} \hat{\sigma}^2 \tag{4.27}$$

で与えられる。ゆえに、線形和 L の $100(1-\alpha)\%$ 信頼限界は、

$$\sum_k l_k \hat{\theta}_k \pm t_\alpha(n-p-1) \sqrt{\sum_k \sum_{k'} l_k l_{k'} c_{k,k'} \hat{\sigma}^2} \tag{4.28}$$

となる。

線形和 L に対する分散 $\text{Var}(\hat{L})$ が、式(4.27)として2次形式で示されているのであるが、手計算の範囲を超えた式であり、何らかのコンピュータ言語を用いて計算するしかなく、実用性に欠けていた。救いの神が、Excelの行列関数であり、 k のすべての組み合わせに対する

$(l_k l_k' c_{k,k} \hat{\sigma}^2)$ の和の計算が、Excel の行列関数によって、何のストレスもなく四則演算のごとく計算できるようになった。実際に、Excel シート上に示されている矩形データに対し、Excel の行列関数を用いた計算方法について例示する。

組合せ水準の 95%信頼区間

表 4.8 に示す $A_1 B_1$ の推定値を求めるための係数ベクトルを $l^{(1)}$

$$l^{(1)} = [l_0^{(1)} \quad l_1^{(1)} \quad l_2^{(1)} \quad l_3^{(1)}] = [1 \quad 1 \quad 0 \quad 1] \tag{4.29}$$

としたときに、式 (4.26) は、

$$\hat{L}^{(1)} = \sum_{k=0}^3 l_k^{(1)} \hat{\theta}_k = l^{(1)} \hat{\theta} = 11.0000 \tag{4.30}$$

	$l^{(1)}_0$	$l^{(1)}_1$	$l^{(1)}_2$	$l^{(1)}_3$			$y^{(1)}$
	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$		$\hat{\theta}$	$L^{(1)}$
$l^{(1)}$	1	1	0	1	12.0333	=	11.0000
					-0.8333		
					0.6167		
					-0.2000		

のように、Excel の Mmult ($l^{(1)}$ の範囲, $\hat{\theta}$ 範囲) 関数で計算できる。式 (4.27) の分散の推定量は、

$$\hat{Var}(\hat{L}^{(1)}) = \sum_k \sum_{k'} l_k^{(1)} l_{k'}^{(1)} c_{k,k'} \hat{\sigma}^2 = l^{(1)} \Sigma(\hat{\theta}) (l^{(1)})^T = 0.0233 \tag{4.31}$$

	$l^{(1)}_0$	$l^{(1)}_1$	$l^{(1)}_2$	$l^{(1)}_3$	パラメータの共分散行列 $\Sigma(\hat{\theta})$				$l^{(1)T}$		
$l^{(1)}$	1	1	0	1	0.0058	0	0	0	1		
					0	0.0117	-0.0058	0	1		
					0	-0.0058	0.0117	0	0		
					0	0	0	0.0058	1		
					=				$Var(L^{(1)})$		
					0.0058	0.0117	-0.0058	0.0058	1	=	0.0233
									1		
									0		
									1		

によって計算することができる。

表 4.8 組合せ水準に対するパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いた分散の計算

$L^{(i)}$	A	B	$l^{(i)}_0$	$l^{(i)}_1$	$l^{(i)}_2$	$l^{(i)}_3$	$\hat{\theta}$	$y^{(i)}$	分散	幅	95%信頼区間	
			$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$		$L^{(i)}$	$Var(L^{(i)})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
$L^{(1)}$	A ₁	B ₁	1	1	0	1	12.0333	= 11.000	0.0233	0.6572	10.3428	11.6572
$L^{(2)}$		B ₂	1	1	0	-1	-0.8333	11.400	0.0233	0.6572	10.7428	12.0572
$L^{(3)}$	A ₂	B ₁	1	0	1	1	0.6167	12.450	0.0233	0.6572	11.7928	13.1072
$L^{(4)}$		B ₂	1	0	1	-1	-0.2000	12.850	0.0233	0.6572	12.1928	13.5072
$L^{(5)}$	A ₃	B ₁	1	-1	-1	1		12.050	0.0233	0.6572	11.3928	12.7072
$L^{(6)}$		B ₂	1	-1	-1	-1		12.450	0.0233	0.6572	11.7928	13.1072
										$t_{0.05}(6-3-1) = T.inv.2T(0.05, 6-1-3) =$	4.3027	

$\hat{L}^{(1)}$ に対する 95%信頼区間は、式 (4.28) によって、

$$\left. \begin{aligned} & \sum_k l_k^{(1)} \hat{\theta}_k \pm t_{0.05}(n-p-1) \sqrt{\sum_k \sum_{k'} l_k^{(1)} l_{k'}^{(1)} c_{k,k'} \hat{\sigma}^2} \\ & = \hat{L}^{(1)} \pm t_{0.05}(6-3-1) \sqrt{Var(\hat{L}^{(1)})} \\ & = 11.000 \pm 4.3027 \sqrt{0.0233} \\ & = (10.3428, 11.6572) \end{aligned} \right\} \quad (4.32)$$

として求めることができる。行列計算は、ややこしいように思われるかも知れないが、行列のサイズは異なっても常に一定形式で計算できることに汎用性がある。また、表 4.8 に示すように Excel のフィルハンドルによる計算式のコピー機能により複数の線形和 L とその分散などの計算が一括してできる。このように複数の線形和（推定値）の分散および 95%信頼区間が一括して計算ができる。

表 4.8 で推定した各因子の組み合わせの推定値に 95%信頼区間を付けた線グラフを図 4.1 に示す。因子 A と因子 B の各組み合わせによる実験データは 1 回しかないのに、その 95%信頼区間を算出できるのことは、2 因子に対する実験デザインの良さの賜物である。推定結果をきれいな図で示すことは、実験データの結果を解釈する際に必須であり、Excel 行列計算の結果を用いて簡便に作図できる。

図 4.1 左に示すのは、Excel の折れ線グラフで、6 個のデータ y に対して折れ線をあてはめ、マークを○印、折れ線を「線なし」とし、「データの追加」で推定値 \hat{y} を重ね書きし、折れ線を --- 点線に変更し、 A_1B_2 と A_2B_1 、 A_2B_2 と A_3B_1 の間の点線を消した結果である。図 4.1 右は、推定値の 95%信頼区間の設定に際し「ユーザ設定」の機能を用いてひげを付け加えた結

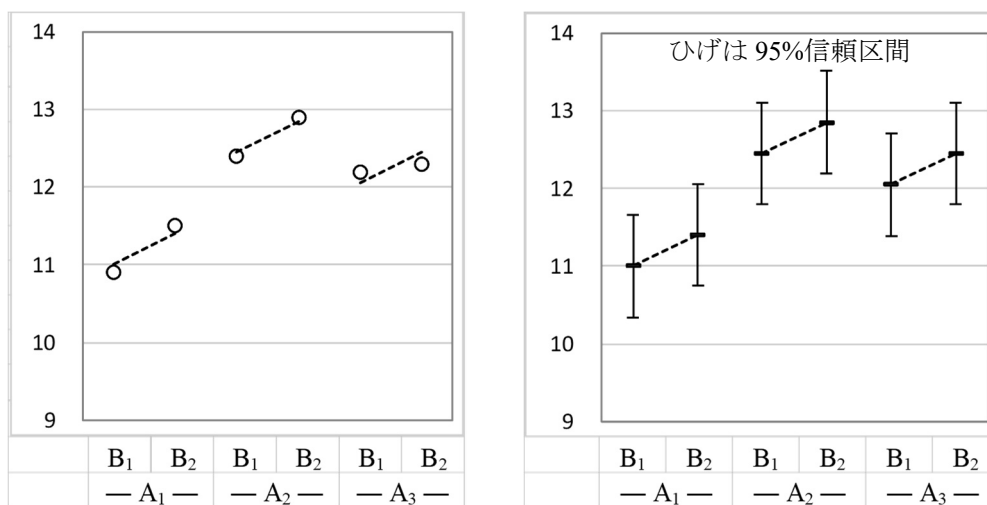


図 4.1 因子 A および B の組み合わせによる推定値と 95%信頼区間

果である。なお、ひげの長さは「 $t_{0.05}(df) \times SE$ 」を用いている。X軸の設定は、Excelのセル内で記述し、折れ線グラフのサイズに合わせて幅を調整し、折れ線グラフごとコピーして、ワードに「(図) 拡張メタファイル」形式で貼り付けたものである。

A_1B_1 の推定値は、表4.6から

$$\left. \begin{aligned} \hat{y}_{1,1} &= \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 \\ &= \hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3 \\ &= 12.0333 - 0.8333 - 0.2000 \\ &= 11.0000 \end{aligned} \right\} \quad (4.33)$$

であることは、自明である。この推定値の分散は、合成分散の一般式を拡張し、表4.7の共分散行列を用いて

$$\left. \begin{aligned} \text{Var}(\hat{y}_{1,1}) &= \text{Var}(\hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3) \\ &= \text{Var}(\hat{\theta}_0) + \text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_3) \\ &\quad + 2\text{Cov}(\hat{\theta}_0, \hat{\theta}_1) + 2\text{Cov}(\hat{\theta}_0, \hat{\theta}_3) + 2\text{Cov}(\hat{\theta}_1, \hat{\theta}_3) \\ &= 0.0058 + 0.0117 + 0.0058 + 0.0 + 0.0 + 0.0 \\ &= 0.0233 \end{aligned} \right\} \quad (4.34)$$

のように計算できるが、式(4.31)と本質的に同じ計算方法となっている。

伝統的な実験計画法では、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いない便宜的な方法、すなわち、式(4.34)の推定に用いているパラメータの自由度の合計の3を用いて、いわゆる田口の式を使い、次式で計算している。

$$\left. \begin{aligned} \frac{1}{n_e} &= \frac{1 + \text{推定に無視しない要因の自由度の和}}{\text{実験数 } N} = \frac{1+3}{6} \\ \text{Var}(y_{11}) &= \hat{\sigma}^2 \times \frac{1}{n_e} = 0.0350 \times \frac{4}{6} = 0.0233 \end{aligned} \right\} \quad (4.35)$$

ただし、田口の式が使えるのは、完備型あるいは直交表の実験データの場合のみで、欠測値がある場合、繰り返し数が不揃いの場合には適用できない。

因子の水準間の差の推定および95%信頼区間

表4.1に示した繰り返しのない2元配置のデータに戻って各種の推定にチャレンジする。表4.9に各水準の平均と基準となる水準との差の平均値を計算した結果を示す。

因子Aの各水準の平均は、因子Bの2つの水準の影響が含まれており、平均値そのもの自体はピュアな推定値とは言い難い。ただし、 A_1 水準と A_2 水準との差は、因子Bの影響が除去され、信頼できる推定値になっている。そこで、その差1.450の95%信頼区間を求めたい。

表 4.9 水準平均と基準水準からの差

	B ₁	B ₂	平均	差
A ₁	10.9	11.5	11.2000	規準
A ₂	12.4	12.9	12.6500	1.4500
A ₃	12.2	12.3	12.2500	1.0500
平均	11.8333	12.2333	12.0333	
差	規準	0.4000		

ここで、パラメータ $\theta_0, \theta_1, \dots, \theta_q$ ($q \leq p$) の任意の1次式 $L^{(j)} = \sum_{k=0}^q l_k^{(j)} \theta_k$ に着目する。因子 A₁ 水準と A₂ 水準の差の推定値と分散を求めて 95%信頼区間を求める。さらに、A₁ 水準と A₃ 水準の差についても同様に求めたい。

表 4.10 に示すように因子 B の水準をプールした因子 A₁ の係数ベクトル $\mathbf{l}^{(7)}$

$$A_1 : \mathbf{l}^{(7)} = [1 \quad 1 \quad 0 \quad 0]$$

は、因子 B に対応するデザイン変数 $x_{3,i}$ について、因子 B の水準のダミー変数の平均

$$l_3^{(7)} = \frac{[B_1 : x_{3,i} = 1] + [B_2 : x_{3,i} = -1]}{2} = \frac{1 + (-1)}{2} = 0 \tag{4.36}$$

とすることにより水準平均および分散が求められる。他の A₂ および A₃ についても表 4.10 に示すように計算することができる。

表 4.10 因子ごとの水準および差に対する推定値と 95%信頼区間

$L^{(i)}$	A	B	$l_0^{(i)}$	$l_1^{(i)}$	$l_2^{(i)}$	$l_3^{(i)}$	$\hat{\theta}$	線形和 L^\wedge	分散 $Var(L^\wedge)$	幅 $t_{0.05} \times SE$	95%信頼区間	
			$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$					$L_{95\%}$	$U_{95\%}$
$L^{(7)}$	A ₁		1	1	0	0	12.0333	11.200	0.0175	0.5692	10.6308	11.7692
$L^{(8)}$	A ₂		1	0	1	0	-0.8333	12.650	0.0175	0.5692	12.0808	13.2192
$L^{(9)}$	A ₃		1	-1	-1	0	0.6167	12.250	0.0175	0.5692	11.6808	12.8192
$L^{(10)}$	A ₁ -A ₁		0	0	0	0	-0.2000	0.000	0.0000	0.0000	0.0000	0.0000
$L^{(11)}$	A ₂ -A ₁		0	-1	1	0		1.450	0.0350	0.8050	0.6450	2.2550
$L^{(12)}$	A ₃ -A ₁		0	-2	-1	0		1.050	0.0350	0.8050	0.2450	1.8550
$L^{(13)}$		B ₁	1	0	0	1		11.833	0.0117	0.4647	11.3686	12.2981
$L^{(14)}$		B ₂	1	0	0	-1		12.233	0.0117	0.4647	11.7686	12.6981
$L^{(15)}$	B ₁ -B ₁		0	0	0	0		0.000	0.0000	0.0000	0.0000	0.0000
$L^{(16)}$	B ₂ -B ₁		0	0	0	-2		0.400	0.0233	0.6572	-0.2572	1.0572
										$t_{0.05}(6-3-1) = T.inv.2T(0.05, 2) =$	4.3027	

水準 A₁ を基準とした (A₂ - A₁) の係数ベクトル $\mathbf{l}^{(11)}$ は、

$$(A_2 - A_1) : \begin{cases} \mathbf{l}^{(11)} = \mathbf{l}^{(8)} - \mathbf{l}^{(7)} \\ = [1 \quad 0 \quad 1 \quad 0] - [1 \quad 1 \quad 0 \quad 0] \\ = [0 \quad -1 \quad 1 \quad 0] \end{cases} \tag{4.37}$$

となり、同様に (A₃ - A₁) は、

$$(A_3 - A_1) : \begin{cases} I^{(12)} = I^{(9)} - I^{(7)} \\ = [1 -1 -1 \ 0] - [1 \ 1 \ 0 \ 0] \\ = [0 -2 -1 \ 0] \end{cases} \quad (4.38)$$

となり，推定値，分散，95%信頼区間が計算されている。

因子 B については，因子 A についてのダミー変数 $x_{1,i}$ および $x_{2,i}$ について，因子 A の水準のダミー変数の平均は，それぞれ 0 になることから，

$$\left. \begin{aligned} B_1 : I^{(13)} &= [1 \ 0 \ 0 \ 1] \\ B_2 : I^{(14)} &= [1 \ 0 \ 0 -1] \end{aligned} \right\} \quad (4.39)$$

であり，水準 B_1 を基準とした $(B_2 - B_1)$ のダミー変数のベクトルは，

$$(B_2 - B_1) : \begin{cases} I^{(16)} = I^{(14)} - I^{(13)} \\ = [1 \ 0 \ 0 -1] - [1 \ 0 \ 0 \ 1] \\ = [0 \ 0 \ 0 -2] \end{cases} \quad (4.40)$$

となる．これらの計算は，表 4.8 の $A_1 B_1$ に対する $I^{(i)}$ の計算式をコピーしペーストしたものである．このように，求めたい推定値に対し，対応する線形和のためのベクトル I を設定するだけで半自動的に計算することができる．

表 4.10 で推定した各因子の水準平均と 95%信頼区間，基準となる水準との差の推定値と 95%信頼区間を図 4.2 に示す．図 4.1 右の因子を組み合わせた場合に比べ 95%信頼区間の幅が小さくなっていることが確認される．

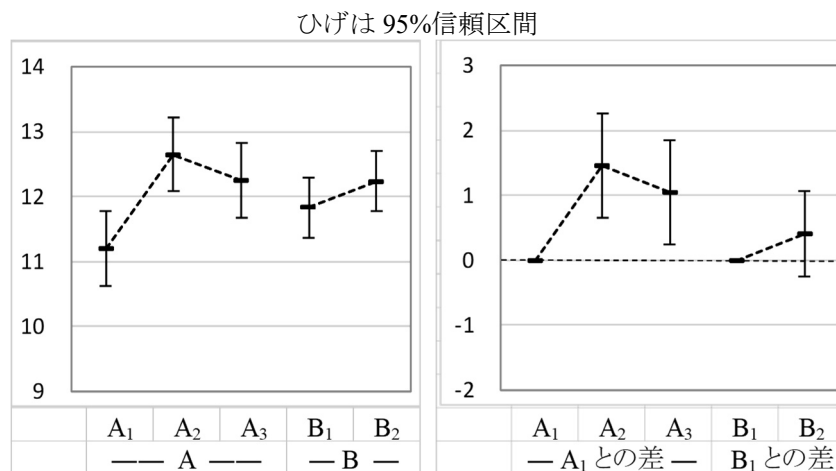


図 4.2 因子 A および B の水準平均および第 1 水準と差の予測プロファイル

現行水準と最適水準の差および 95%信頼区間

実験の目的が，現行の操業条件から何らかの改善をしたいとする．現行の水準を $A_1 B_1$ としたときに，最適水準が $A_2 B_2$ であった場合に，その差の 95%信頼区間を算出し，統計的な判断

をしたい。2つの組み合わせ水準の推定値および分散は、表4.8で既に計算している。表4.11に示すように、それぞれの係数ベクトルの差は、

$$\begin{aligned} \mathbf{l}^{(17)} &= \mathbf{l}^{(4)} - \mathbf{l}^{(1)} \\ &= [1 \ 0 \ 1 -1] - [1 \ 1 \ 0 \ 1] \\ &= [0 -1 \ 1 -2] \end{aligned} \quad (4.41)$$

となり、

$$\hat{Var}[\hat{L}^{(17)}] = \sum_k \sum_{k'} l_k^{(17)} l_{k'}^{(17)} c_{k,k'} \hat{\sigma}^2 = \mathbf{l}^{(17)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{l}^{(17)T} = 0.0583 \quad (4.42)$$

が計算され、差の95%信頼区間は、(0.8108, 2.8892)と0を含んでいないので、有意な差であると判断される。

表4.11 現行水準と最適水準の差の95%信頼区間

$L^{(i)}$	A	B	$l_{0i}^{(i)}$	$l_{1i}^{(i)}$	$l_{2i}^{(i)}$	$l_{3i}^{(i)}$	$\hat{\boldsymbol{\theta}}$	線形和 L^\wedge	分散 $Var(L^\wedge)$	幅 $t_{0.05} \times SE$	95%信頼区間 $L_{95\%}$ $U_{95\%}$	
$L^{(1)}$	A ₁	B ₁	1	1	0	1	12.0333	11.0000	0.0233	0.6572	10.3428	11.6572
$L^{(4)}$	A ₂	B ₂	1	0	1	-1	-0.8333	12.8500	0.0233	0.6572	12.1928	13.5072
	差0		0	0	0	0	0.6167	0.0000	0.0000	0.0000	0.0000	0.0000
$L^{(17)}$	差		0	-1	1	-2	-0.2000	1.8500	0.0583	1.0392	0.8108	2.8892
$t_{0.05}(6-3-1) =$										4.3027		

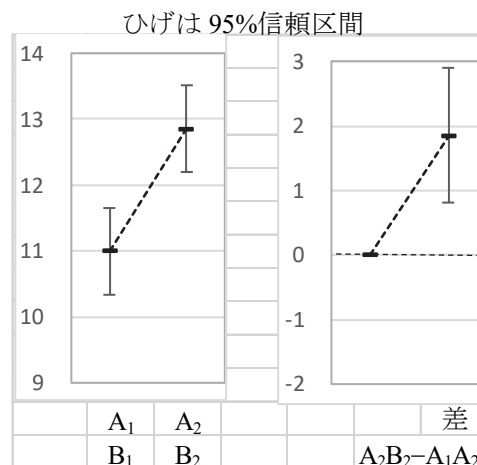


図4.3 現行水準と最適水準との差の予測プロファイル

従来の実験計画法においては、現行水準および最適水準のそれぞれの推定値の分散を求めるために、式(4.35)に示した田口の式によって有効反復数 n_e を求め、それぞれの分散を求めることができた。現行水準および最適水準の差は、簡単に求められるが、差の分散をどのようにして求めたら良いのだろうか。A₁B₁の推定値とA₂B₂の推定値は、表4.7に示すように、A₁とA₂の効果 $\hat{\alpha}_1$ と $\hat{\alpha}_2$ は互いに独立でではなく、共分散 $Cov(\hat{\alpha}_1, \hat{\alpha}_2) = -0.0058$ が存在し、互いに独立ではないので、それぞれの分散を加えて求めることができない。

4.2. 直交表に対する線形モデル

直交表を用いた実験データの解析

直交表を用いた実験データの解析は、手計算を前提とした平方和の分解に基づく方法が確立している。平方和の分解による計算の結果は、分散分析表にまとめられ、取り上げた多くの要因の効果を一括して統計的に評価できる優れたものである。ただし、2水準あるいは3水準の質的な変数に限定されており、量的な変数の場合は、ある特定の2または3水準を設定して質的な変数として扱うことになる。2水準の直交表に対し、一部の因子に対し3水準したい場合に、「組合せ法」として、直交表の2列を用いて(1, 1 → 1), (1, 2 → 2), (2, 1 → 3), (2, 2 → 3)のように4水準を3水準に縮小するような割り付けテクニックを用いて平方和を計算するなど、様々な割り付けに対応した平方和の計算なども整備され、実験計画法と言えば、直交表と言われるほど普及し、多くの実験計画法に関する書籍により解説されている。

直交表に限らず伝統的な実験計画による解析法は、偏差平方和に基づく平方和の分解を主体にしている。直交表を用いた実験データに線形モデルを適用し解析する方法は、要因を全て量的な変数として取り扱うことになる。標準的に用いられている(1, 2)型の2水準系の直交表も元々は、(1, -1)対比型ダミー変数を手作業での計算が容易になるように「-1」を「2」に置き換えたものである。

直交表における構造(DE)モデルと線形モデル

2水準系の L_8 直交表のルーツは、全て2水準の(2×2×2)の3元配置である。要因をA, B, Cとしたときに、データの構造式は、次式で与えられる。

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijk} \quad (4.43)$$

y_{111}	$= \mu$	$+ \alpha_1$	$+ \beta_1$	$+ (\alpha\beta)_{11}$	$+ \gamma_1$	$+ (\alpha\gamma)_{11}$	$+ (\beta\gamma)_{11}$	$+ (\alpha\beta\gamma)_{111}$	$+ \varepsilon_{111}$	}	(4.44)
y_{112}	$= \mu$	$+ \alpha_1$	$+ \beta_1$	$+ (\alpha\beta)_{11}$	$+ \gamma_2$	$+ (\alpha\gamma)_{12}$	$+ (\beta\gamma)_{12}$	$+ (\alpha\beta\gamma)_{112}$	$+ \varepsilon_{112}$		
y_{121}	$= \mu$	$+ \alpha_1$	$+ \beta_2$	$+ (\alpha\beta)_{12}$	$+ \gamma_1$	$+ (\alpha\gamma)_{11}$	$+ (\beta\gamma)_{21}$	$+ (\alpha\beta\gamma)_{121}$	$+ \varepsilon_{121}$		
y_{122}	$= \mu$	$+ \alpha_1$	$+ \beta_2$	$+ (\alpha\beta)_{12}$	$+ \gamma_2$	$+ (\alpha\gamma)_{12}$	$+ (\beta\gamma)_{22}$	$+ (\alpha\beta\gamma)_{122}$	$+ \varepsilon_{122}$		
y_{211}	$= \mu$	$+ \alpha_2$	$+ \beta_1$	$+ (\alpha\beta)_{21}$	$+ \gamma_1$	$+ (\alpha\gamma)_{21}$	$+ (\beta\gamma)_{11}$	$+ (\alpha\beta\gamma)_{211}$	$+ \varepsilon_{211}$		
y_{212}	$= \mu$	$+ \alpha_2$	$+ \beta_1$	$+ (\alpha\beta)_{21}$	$+ \gamma_2$	$+ (\alpha\gamma)_{22}$	$+ (\beta\gamma)_{12}$	$+ (\alpha\beta\gamma)_{212}$	$+ \varepsilon_{212}$		
y_{221}	$= \mu$	$+ \alpha_2$	$+ \beta_2$	$+ (\alpha\beta)_{22}$	$+ \gamma_1$	$+ (\alpha\gamma)_{21}$	$+ (\beta\gamma)_{21}$	$+ (\alpha\beta\gamma)_{221}$	$+ \varepsilon_{221}$		
y_{222}	$= \mu$	$+ \alpha_2$	$+ \beta_2$	$+ (\alpha\beta)_{22}$	$+ \gamma_2$	$+ (\alpha\gamma)_{22}$	$+ (\beta\gamma)_{22}$	$+ (\alpha\beta\gamma)_{222}$	$+ \varepsilon_{222}$		

式(4.43)を線形モデルとして表すと

$$y_i = \theta_0 x_{0,i} + \theta_1 x_{1,i} + \theta_2 x_{2,i} + \cdots + \theta_p x_{p,i} + \varepsilon_i \quad (4.45)$$

となり、母数 $(\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \dots)$ を $(\theta_0, \theta_1, \theta_2, \dots)$ に対応付け、その変数を $(x_{0,i},$

$x_{1,i}, x_{2,i}, \dots$) とすることにより線形モデルとなる. 変数の数 p は, $p = 3 \times 2 + 3 \times 4 + 8 = 26$ と大きくなるが, 変数の間に

$$x_{1,i} + x_{2,i} = 1, \quad x_{3,i} + x_{4,i} = 1, \dots \quad (4.46)$$

などの関係 (1次従属) がある. そのために, 母数に関する制約条件

$$\left. \begin{aligned} \alpha_1 + \alpha_2 = 0, \quad \beta_1 + \beta_2 = 0, \quad \gamma_1 + \gamma_2 = 0, \\ (\alpha\beta)_{11} + (\alpha\beta)_{12} + (\alpha\beta)_{21} + (\alpha\beta)_{22} = 0, \\ (\alpha\gamma)_{11} + \dots \end{aligned} \right\} \quad (4.47)$$

を用い

$$\left. \begin{aligned} \alpha_1 = \alpha, \quad \alpha_2 = -\alpha \\ \beta_1 = \beta, \quad \beta_2 = -\beta \\ \gamma_1 = \gamma, \quad \gamma_2 = -\gamma \\ (\alpha\beta)_{11} = (\alpha\beta)_{22} = \alpha\beta, \quad (\alpha\beta)_{12} = (\alpha\beta)_{21} = -\alpha\beta \\ (\alpha\gamma)_{11} = (\alpha\gamma)_{22} = \alpha\gamma, \quad \dots \\ \dots \end{aligned} \right\} \quad (4.48)$$

とする. 式 (4.48) を式 (4.44) に代入すると

y_{111}	$= \mu$	$+ \alpha$	$+ \beta$	$+ \alpha\beta$	$+ \gamma$	$+ \alpha\gamma$	$+ \beta\gamma$	$+ \alpha\beta\gamma$	$+ \varepsilon_{111}$	}	(4.49)
y_{112}	$= \mu$	$+ \alpha$	$+ \beta$	$+ \alpha\beta$	$- \gamma$	$- \alpha\gamma$	$- \beta\gamma$	$- \alpha\beta\gamma$	$+ \varepsilon_{112}$		
y_{121}	$= \mu$	$+ \alpha$	$- \beta$	$- \alpha\beta$	$+ \gamma$	$+ \alpha\gamma$	$- \beta\gamma$	$- \alpha\beta\gamma$	$+ \varepsilon_{121}$		
y_{122}	$= \mu$	$+ \alpha$	$- \beta$	$- \alpha\beta$	$- \gamma$	$- \alpha\gamma$	$+ \beta\gamma$	$+ \alpha\beta\gamma$	$+ \varepsilon_{122}$		
y_{211}	$= \mu$	$- \alpha$	$+ \beta$	$- \alpha\beta$	$+ \gamma$	$- \alpha\gamma$	$+ \beta\gamma$	$- \alpha\beta\gamma$	$+ \varepsilon_{211}$		
y_{212}	$= \mu$	$- \alpha$	$+ \beta$	$- \alpha\beta$	$- \gamma$	$+ \alpha\gamma$	$- \beta\gamma$	$+ \alpha\beta\gamma$	$+ \varepsilon_{212}$		
y_{221}	$= \mu$	$- \alpha$	$- \beta$	$+ \alpha\beta$	$+ \gamma$	$- \alpha\gamma$	$- \beta\gamma$	$+ \alpha\beta\gamma$	$+ \varepsilon_{221}$		
y_{222}	$= \mu$	$- \alpha$	$- \beta$	$+ \alpha\beta$	$- \gamma$	$+ \alpha\gamma$	$+ \beta\gamma$	$- \alpha\beta\gamma$	$+ \varepsilon_{222}$		

と書き換えることができる. 式 (4.49) において $p=7$ となり, 母数 ($\mu, \alpha, \beta, \alpha\beta, \gamma, \alpha\gamma, \beta\gamma, \alpha\beta\gamma$) を ($\theta_0, \theta_1, \theta_2, \dots, \theta_7$) に対応付け, それらの変数を ($x_{0,i}, x_{1,i}, x_{2,i}, \dots, x_{7,i}$) とする.

これらの母数に関する制約により, 変数の間に式 (4.43) のような従属関係は存在せず, ($x_{0,i}, x_{1,i}, x_{2,i}, \dots, x_{7,i}$) は1次独立となる.

$$\left. \begin{aligned} y_1 &= \theta_0 x_{0,1} + \theta_1 x_{1,1} + \theta_2 x_{2,1} + \dots + \theta_7 x_{7,1} + \varepsilon_1 \\ y_2 &= \theta_0 x_{0,2} + \theta_1 x_{1,2} + \theta_2 x_{2,2} + \dots + \theta_7 x_{7,2} + \varepsilon_2 \\ &\vdots \\ y_8 &= \theta_0 x_{0,6} + \theta_1 x_{1,6} + \theta_2 x_{2,6} + \dots + \theta_7 x_{7,6} + \varepsilon_8 \end{aligned} \right\} \quad (4.50)$$

式 (4.50) をデザイン行列として表せば, 表 4.12 に示すようになる. 一般的に普及している直交表は, 手計算の利便性のためにデザイン行列 \mathbf{X} の「-1」を「2」に置き換えている. そのために, 列と列との積和が0となる直交性が失われている. (1, 2)型のままで線形モデルとすることも可能であるが, 結果の解釈が煩雑になるので, (1, -1)対比型ダミー変数とするこ

表 4.12 L_8 直交表に対するデザイン行列 X および (1, 2) 型 L_8 直交表

—— デザイン行列 X ——									Y			列 番							
No.	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	θ	ε	y	No.	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	1	1	1	1	1	1	1	1	θ_0	ε_1	y_1	1	1	1	1	1	1	1	1
2	1	1	1	1	-1	-1	-1	-1	θ_1	ε_2	y_2	2	1	1	1	2	2	2	2
3	1	1	-1	-1	1	1	-1	-1	θ_2	ε_3	y_3	3	1	2	2	1	1	2	2
4	1	1	-1	-1	-1	-1	1	1	θ_3	ε_4	y_4	4	1	2	2	2	2	1	1
5	1	-1	1	-1	1	-1	1	-1	θ_4	ε_5	y_5	5	2	1	2	1	2	1	2
6	1	-1	1	-1	-1	1	-1	1	θ_5	ε_6	y_6	6	2	1	2	2	1	2	1
7	1	-1	-1	1	1	-1	-1	1	θ_6	ε_7	y_7	7	2	2	1	1	2	2	1
8	1	-1	-1	1	-1	1	1	-1	θ_7	ε_8	y_8	8	2	2	1	2	1	1	2
	μ	a	b	ab	c	ac	bc	abc				成分	a	b	ab	c	ac	bc	abc
												群	-1-	-2-	-3-	-3-	-3-	-3-	-3-

とが望ましい。ダミー変数といえば (0, 1) 型ではないかとの疑問も生ずるかもしれない。もちろん (0, 1) 型とすることは可能であるが、直交性が失われてしまうので避けた方がよい。

線形モデルによる解析は、直交の条件が崩れても共通の解析の手順により各種の推定ができる汎用的な方法であり、多くの統計ソフトの内部の計算ロジックに使われている。ただし、行列計算を伴うので、多くの実験計画法の入門書では、避けられている。また、実験計画の専門書では、理論が重視され実際の解析では、統計ソフトに丸投げするのが現状であり、線形モデルを実際に使った事例を扱っている成書は、ごく限られている。

これまで、Excel の行列関数を用いて、線形モデルによる解析が手軽にできることを示してきたのであるが、Excel の分析ツールとして提供されている「回帰分析」は、線形モデルの解析のかなりの部分をカバーするので、解析手順の簡素化に役立つ。

直交表に対する手計算による解析手順を、Basic などの計算機言語により実現することは、田中・垂水 (1986)、「パソコン統計解析ハンドブック III 実験計画法編」に示されているように、容易ではあったと思われる。ただし、線形モデルによる各種の推定および 95% 信頼区間の計算は、行列計算になしには実現不可能に近いので、多くの実験計画法の入門書では、伝統的な手計算による分散分析表の作成に引き続き、各種の推定とその分散を求めるために、田口の公式および伊奈の公式などによる有効反復数 n_e の計算が示されているのが現状である。

4.3. 欠測値がある直交表に対する線形モデルの活用

直交表による実験で欠測値が生じた場合

最も基本的な L_8 直交表を対象に、何らかのトラブルでデータが得られない状況を設定し、線形モデルによる解析法を示す。取り上げる事例としては、手計算による解析結果が丁寧に示されている楠ら（1995）「応用実験計画法」の第6章の「例題 6.6」を用いる。

表 4.13 に示すのは、 L_8 直交表による実験で、4つの要因（A, B, C, D）を取り上げ、互いに交互作用がないと想定されるが、もしも交互作用があった場合に備えて、要因の割り付けは A を (1)列, B を (2)列, C を (4)列, D を (7)列とした。割り付けがない (3)列には、成分表示から $A \times B$ と $C \times D$, (5)列には、 $A \times C$ と $B \times D$, (6)列には、 $A \times D$ と $B \times C$ が出現することがわかる。実際の実験データで、残差の平均平方が大きい場合には、これらの列を組み込んだ解析を行うことにより、いくばくかの情報が得られる。

因子 D を第7列に割り付けたのは、成分表示が abc であり、成分 a の列, 成分 b の列, 成分 c の列の3因子交互作用が現れる列なので、それらの列に割り付けた3つの要因（A, B, C）間の2因子交互作用があったとしても、その影響を受けないようにするための配慮である。

表 4.13 直交表への要因の割り付け [楠ら（1995），例題 6.6]

	A	B		C			D	データ
No.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	y
1	1	1	1	1	1	1	1	32
2	1	1	1	2	2	2	2	23
3	1	2	2	1	1	2	2	19
4	1	2	2	2	2	1	1	17
5	2	1	2	1	2	1	2	19
6	2	1	2	2	1	2	1	20
7	2	2	1	1	2	2	1	欠測
8	2	2	1	2	1	1	2	8
成分	a	b	ab	c	ac	bc	abc	
交互作用			$A \times B$		$A \times C$	$A \times D$		
			$C \times D$		$B \times D$	$B \times C$		

表 4.13 に示すように、実験 No = 7 において欠測が生じた。そのために、伝統的な平方和の計算による分散分析表が作成できないので、デザイン行列 X を用いた線形モデルを用いた解析が必要となる。表 4.14 に、表 4.3 と同様にデザイン行列 X を用いた正規方程式を示す。積和行列 ($X^T X$) の対角要素がデータ数の 7 となり、非対角セルに 1 および -1 が出現しており、各列が互いに直交していないことが明らかである。欠測セルがなければ、対角要素がす

べて 8 となり、他のセルはすべて 0 となる。なお、Excel による行列計算に不慣れな場合には、第 5 章を参照のこと。

表 4.14 デザイン行列 X を用いた正規方程式

		転置行列 X^T						デザイン行列 X					θ^\wedge	転置行列 X^T						Y			
		1	2	3	4	5	6	8	x_0	x_1	x_2	x_3	x_4		1	2	3	4	5	6	8	y	
μ	x_0	1	1	1	1	1	1	1	1	1	1	1	1	$\theta^\wedge_0 =$	1	1	1	1	1	1	1	32	y_1
A	x_1	1	1	1	1	-1	-1	-1	1	1	1	-1	-1	$\theta^\wedge_1 =$	1	1	1	1	-1	-1	-1	23	y_2
B	x_2	1	1	-1	-1	1	1	-1	1	1	-1	1	-1	$\theta^\wedge_2 =$	1	1	-1	-1	1	1	-1	19	y_3
C	x_3	1	-1	1	-1	1	-1	-1	1	1	-1	-1	1	$\theta^\wedge_3 =$	1	-1	1	-1	1	-1	-1	17	y_4
D	x_4	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	$\theta^\wedge_4 =$	1	-1	-1	1	-1	1	-1	19	y_5
									1	-1	1	-1	1									20	y_6
									1	-1	-1	-1	-1									8	y_8
									μ	A	B	C	D										

$X^T X (a_{kk'})$				θ^\wedge	$X^T Y$		
7	1	1	-1	-1	$\theta^\wedge_0 =$	138.0	B_0
1	7	-1	1	1	$\theta^\wedge_1 =$	44.0	B_1
1	-1	7	1	1	$\theta^\wedge_2 =$	50.0	B_2
-1	1	1	7	-1	$\theta^\wedge_3 =$	2.0	B_3
-1	1	1	-1	7	$\theta^\wedge_4 =$	0.0	B_4

表 4.14 で示されている積和行列 $(X^T X)$ の逆行列 $(X^T X)^{-1}$ を正規方程式の両辺に掛けると表 4.15 の結果を得る。得られた線形式は、

$$\left. \begin{aligned}
 \hat{y}_i &= \hat{\theta}_0 x_{0,i} + \hat{\theta}_1 x_{1,i} + \dots + \hat{\theta}_4 x_{4,i} \\
 &= 19.17x_{0,i} + 3.583x_{1,i} + 4.333x_{2,i} + 2.167x_{3,i} + 1.917x_{4,i}
 \end{aligned} \right\} \quad (4.51)$$

μ A B C D

となる。

表 4.15 欠測値がある直交表に対する正規方程式の解

$(X^T X)^{-1} (c_{kk'})$					$X^T X (a_{kk'})$					θ^\wedge	$(X^T X)^{-1} (c_{kk'})$					$X^T Y$
0.167	-0.042	-0.042	0.042	0.042	7	1	1	-1	-1	$\theta^\wedge_0 =$	0.167	-0.042	-0.042	0.042	0.042	138.0
-0.042	0.167	0.042	-0.042	-0.042	1	7	-1	1	1	$\theta^\wedge_1 =$	-0.042	0.167	0.042	-0.042	-0.042	44.0
-0.042	0.042	0.167	-0.042	-0.042	1	-1	7	1	1	$\theta^\wedge_2 =$	-0.042	0.042	0.167	-0.042	-0.042	50.0
0.042	-0.042	-0.042	0.167	0.042	-1	1	1	7	-1	$\theta^\wedge_3 =$	0.042	-0.042	-0.042	0.167	0.042	2.0
0.042	-0.042	-0.042	0.042	0.167	-1	1	1	-1	7	$\theta^\wedge_4 =$	0.042	-0.042	-0.042	0.042	0.167	0.0
										θ^\wedge	$(X^T X)^{-1} X^T Y$					
										$\theta^\wedge_0 =$	19.17			x_0	μ	
										$\theta^\wedge_1 =$	3.583			x_1	A	
										$\theta^\wedge_2 =$	4.333			x_2	B	
										$\theta^\wedge_3 =$	2.167			x_3	C	
										$\theta^\wedge_4 =$	1.917			x_4	D	

表 4.16 に示すのは、予測値 \hat{y}_i を $X\hat{\theta}$ で計算し、元のデータ y_i との残差 $\hat{\varepsilon}_i = y_i - \hat{y}_i$ の平方和 $S_e = 4.3333$ が計算されている。さらに、自由度 $df = 7 - (4 + 1) = 2$ で割った誤差分散の推定値 $\hat{\sigma}^2 = S_e / df = 2.1667$ が求められている。

表 4.16 予測値 \hat{y} の推定および誤差分散 $\hat{\sigma}^2$ の推定

No.	デザイン行列 X					予測値 \hat{y} の計算		誤差分散 $\hat{\sigma}^2$ の推定		
	x_0	x_1	x_2	x_3	x_4	$\hat{\theta}$	\hat{y}	y	y^\wedge	$\varepsilon^\wedge = y - y^\wedge$
1	1	1	1	1	1	19.167	= 31.167	32	- 31.167	= 0.8333
2	1	1	1	-1	-1	3.583	23.000	23	23.000	0.0000
3	1	1	-1	1	-1	4.333	18.667	19	18.667	0.3333
4	1	1	-1	-1	1	2.167	18.167	17	18.167	-1.1667
5	1	-1	1	1	-1	1.917	20.167	19	20.167	-1.1667
6	1	-1	1	-1	1		19.667	20	19.667	0.3333
8	1	-1	-1	-1	-1		7.167	8	7.167	0.8333
		A	B	C	D			$n = 7$	$S_e =$	4.3333
								$p = 4$	$\hat{\sigma}^2 =$	2.1667
								$df = n - p - 1 = 2$	$\hat{\sigma} =$	1.4720

ここまでの行列計算に代えて、表 4.18 に示すようにデザイン行列を用いた Excel の回帰分析によっても全く同じ結果が得られる。各種の推定のために必要なパラメータの共分散行列 $\Sigma(\hat{\theta})$ は、表 4.17 に示すように正規方程式の解を求めるために用いた積和行列 $(X^T X)$ の逆行列 $(X^T X)^{-1}$ に誤差分散の推定値 $\hat{\sigma}^2 = 2.1667$ を掛けて求めることができる。パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根は、表 4.18 に示す回帰パラメータ（係数）の標準誤差 SE となる。

表 4.17 パラメータの共分散行列 $\Sigma(\hat{\theta})$

	$(X^T X)^{-1} (c_{kk'})$					$\hat{\sigma}^2$	パラメータの共分散行列 $\Sigma(\hat{\theta})$					SE
$\hat{\theta}_0$	0.167	-0.042	-0.042	0.042	0.042	2.1667	0.361	-0.090	-0.090	0.090	0.090	0.6009
$\hat{\theta}_1$	-0.042	0.167	0.042	-0.042	-0.042		-0.090	0.361	0.090	-0.090	-0.090	0.6009
$\hat{\theta}_2$	-0.042	0.042	0.167	-0.042	-0.042		-0.090	0.090	0.361	-0.090	-0.090	0.6009
$\hat{\theta}_3$	0.042	-0.042	-0.042	0.167	0.042		0.090	-0.090	-0.090	0.361	0.090	0.6009
$\hat{\theta}_4$	0.042	-0.042	-0.042	0.042	0.167		0.090	-0.090	-0.090	0.090	0.361	0.6009
	x_0	x_1	x_2	x_3	x_4		x_0	x_1	x_2	x_3	x_4	
	A	B	C	D			A	B	C	D		

Excel の回帰分析を用いたパラメータの推定

表 4.18 に示す Excel の回帰分析では、通常「切片 x_0 」列を含めない“デザイン行列 X ”を説明変数として設定する。下段に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ は、Excel の行列関数と分散分析表に示されている残差の行の分散 $\hat{\sigma}^2 = 2.1667$ を用いて、

$$\Sigma(\hat{\theta}) = \text{Minverse}(\text{Mmult}(\text{Transpose}(X \text{ の範囲}), X \text{ の範囲})) * \hat{\sigma}^2 \quad (4.52)$$

として一括計算した結果である。この計算に用いるデザイン行列 X は、もちろん「切片 x_0 」を含んだデザイン行列 X である。

表 4.18 に示す Excel の回帰分析の結果は、表 4.15 に示した線形モデルによる結果と一致する。したがって、欠測値がある直交表の解析には、ごく一般的な回帰分析を適用すればよい

表 4.18 Excel の回帰分析によるパラメータの推定

No.	デザイン行列 X					θ	ε	Y	分散分析表, 「定数に 0 を使用」を off					
	x_0	x_1	x_2	x_3	x_4					自由度	変動	分散	分散比	
1	1	1	1	1	1	θ_0	ε_1	32	回帰	4	303.0952	75.7738	34.9725	
2	1	1	1	-1	-1	θ_1	ε_2	23	残差	2	4.3333	2.1667		
3	1	1	-1	1	-1	θ_2	ε_3	19	合計	6	307.4286			
4	1	1	-1	-1	1	θ_3	ε_4	17						
5	1	-1	1	1	-1	θ_4	ε_5	19						
6	1	-1	1	-1	1			20	θ^{\wedge}_0 切片 x_0	19.1667	0.6009	31.8953	0.0010	
8	1	-1	-1	-1	-1			8	θ^{\wedge}_1 x_1	3.5833	0.6009	5.9630	0.0270	
		A	B	C	D				θ^{\wedge}_2 x_2	4.3333	0.6009	7.2111	0.0187	
	μ	a	b	c	abc				θ^{\wedge}_3 x_3	2.1667	0.6009	3.6056	0.0691	
									θ^{\wedge}_4 x_4	1.9167	0.6009	3.1895	0.0858	
									パラメータの共分散行列 $\Sigma(\hat{\theta})=(X^T X)^{-1}\sigma^{\wedge 2}$					
									θ^{\wedge}_0	0.3611	-0.0903	-0.0903	0.0903	0.0903
									θ^{\wedge}_1	-0.0903	0.3611	0.0903	-0.0903	-0.0903
									θ^{\wedge}_2	-0.0903	0.0903	0.3611	-0.0903	-0.0903
									θ^{\wedge}_3	0.0903	-0.0903	-0.0903	0.3611	0.0903
									θ^{\wedge}_4	0.0903	-0.0903	-0.0903	0.0903	0.3611

=Minverse (Mmult (Transpose (X の範囲), X の範囲))* $\sigma^{\wedge 2}$

ことが分かる。分散分析表がほしい場合には、パラメータの SE と誤差分散を用いて、各因子の平方和を計算することができる。分子が自由度 1 の分散分析表の F 値は、パラメータの t 値との間で、 $F=t^2$ との関係が知られおり、また、 F 値は、各因子の平均平方を $\hat{\sigma}^2$ で割って得られていることから、次式が導かれる。

$$t_A^2 = F_A = \frac{S_A / 1}{\hat{\sigma}^2} \tag{4.53}$$

式 (4.53) を S_A について解くと、平方和 S_A を求めることができる。

$$\begin{aligned} S_A &= t_A^2 \hat{\sigma}^2 \\ &= 5.9630^2 \times 2.1667 \\ &= 77.0417 \end{aligned} \tag{4.54}$$

他の因子についても同様に平方和を計算することができる。これらの結果を表 4.19 の分散分析表にまとめる。回帰の平方和は、表 4.18 のモデル全体の分散分析表から $S_{\text{回帰}} = 303.0952$

表 4.19 欠測値がある直交表に対する分散分析表

要因	自由度	変動	分散	分散比	p 値	t 値	$\hat{\sigma}^2$	$t^2 \hat{\sigma}^2$	平方和 S
	df	平方和 S	平均平方	F					
回帰	4	303.0952	75.7738	34.9725					
A	1	77.0417	77.0417	35.5577	0.0270	5.9630	2.1667	77.0417	
B	1	112.6667	112.6667	52.0000	0.0187	7.2111	2.1667	112.6667	
C	1	28.1667	28.1667	13.0000	0.0691	3.6056	2.1667	28.1667	
D	1	22.0417	22.0417	10.1731	0.0858	3.1895	2.1667	22.0417	
	小計	239.9167							
残差	2	4.3333	2.1667						
合計	6	307.4286							

となるが、それぞれの因子の平方和の合計には一致しない。そこで、表 4.19 に示すようにモデル全体の分散分析表に、それぞれの因子の平方和を別の列に挿入する形式により、整合性を保つ必要がある。

欠測値がある場合の各種の推定

欠測値がない場合には、各因子の水準平均と、推定された回帰パラメータを用いた推定値が一致するのであるが、欠測値がある場合には一致しない。因子 A の場合であれば、

$$\left. \begin{aligned} A_1 \text{の平均} &= (32 + 23 + 19 + 17) / 4 = 22.750 \\ A_2 \text{の平均} &= (19 + 20 + 8) / 3 = 15.667 \end{aligned} \right\} \quad (4.55)$$

であり、回帰パラメータを用いた場合は、表 4.20 に示すように A₁ は一致するが、A₂ は、

$$\left. \begin{aligned} A_1 \text{の推定値} &= \hat{\theta}_0 + \hat{\theta}_1 = \hat{\mu} + \hat{\alpha} = 19.167 + 3.583 = 22.750 \\ A_2 \text{の推定値} &= \hat{\theta}_0 - \hat{\theta}_1 = \hat{\mu} - \hat{\alpha} = 19.167 - 3.583 = 15.583 \end{aligned} \right\} \quad (4.56)$$

のように異なることに注意が必要である。これらの推定値は、行ベクトル \mathbf{l} とパラメータの推定値 $\hat{\boldsymbol{\theta}}$ の積和 $L^{(1)} = \mathbf{l}^{(1)} \hat{\boldsymbol{\theta}}$

A ₁ の推定値	$\mathbf{l}^{(1)}$	1	1	0	0	0	19.1667	=	22.7500
							3.5833		最小2乗平均
							4.3333		Lsmeans
							2.1667		
							1.9167		

で求められている。統計ソフト SAS および JMP では、このようにパラメータ $\hat{\boldsymbol{\theta}}$ を用いた線形和で求められた推定値を最小 2 乗平均 (lsmeans) と称している。

これらの線形和 L の分散 $Var(L)$ は、A₁ および A₂ の推定値に用いた行ベクトル \mathbf{l}

$$\left. \begin{aligned} \mathbf{l}^{(1)} &= [1 \quad 1 \quad 0 \quad 0 \quad 0] \\ \mathbf{l}^{(2)} &= [1 \quad -1 \quad 0 \quad 0 \quad 0] \end{aligned} \right\} \quad (4.57)$$

を用いて、

$$\left. \begin{aligned} Var(L^{(1)}) &= \mathbf{l}^{(1)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{l}^{(1)T} \\ &= Var(\hat{\theta}_0) \times 1^2 + 2 \times Cov(\hat{\theta}_0, \hat{\theta}_1) \times 1 \times 1 + Var(\hat{\theta}_1) \times 1^2 \\ &= 0.3611 + 2 \times (-0.0903) + 0.3611 = 0.5417 \\ Var(L^{(2)}) &= \mathbf{l}^{(2)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{l}^{(2)T} \\ &= Var(\hat{\theta}_0) \times 1^2 + 2 \times Cov(\hat{\theta}_0, \hat{\theta}_1) \times 1 \times (-1) + Var(\hat{\theta}_1) \times (-1)^2 \\ &= 0.3611 - 2 \times (-0.0903) + 0.3611 = 0.5417 = 0.9028 \end{aligned} \right\} \quad (4.58)$$

として計算されている。 $\mathbf{l}^{(1)}$ に比べ $\mathbf{l}^{(2)}$ の分散が大きくなっているのは、欠測値の影響で、デー

タ数が3となっていることの反映でもある．表 4.20 に示すように，因子 B, ..., D の水準平均と分散が計算されている．

表 4.20 因子ごとの水準に対する 95%信頼区間

$L^{(i)}$	A	B	C	D	$l^{(i)}_0$	$l^{(i)}_1$	$l^{(i)}_2$	$l^{(i)}_3$	$l^{(i)}_4$	θ^\wedge	線形和	分散	幅	95%信頼区間	
					x_0	x_1	x_2	x_3	x_4		L^\wedge	$Var(L^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
$L^{(1)}$	A ₁				1	1	0	0	0	19.167	22.750	0.5417	3.167	19.583	25.917
$L^{(2)}$	A ₂				1	-1	0	0	0	3.583	15.583	0.9028	4.088	11.495	19.671
$L^{(3)}$		B ₁			1	0	1	0	0	4.333	23.500	0.5417	3.167	20.333	26.667
$L^{(4)}$		B ₂			1	0	-1	0	0	2.167	14.833	0.9028	4.088	10.745	18.921
$L^{(5)}$			C ₁		1	0	0	1	0	1.917	21.333	0.9028	4.088	17.245	25.421
$L^{(6)}$			C ₂		1	0	0	-1	0		17.000	0.5417	3.167	13.833	20.167
$L^{(7)}$				D ₁	1	0	0	0	1		21.083	0.9028	4.088	16.995	25.171
$L^{(8)}$				D ₂	1	0	0	0	-1		17.250	0.5417	3.167	14.083	20.417
					A	B	C	D					$t_{0.05}(7-5)=$	4.303	

直交表実験では，現行の操業状態 (A₁, B₁, C₁, D₁) での推定値に対し，水準を変更したときに良い結果（小さい方がよいとする）が得られる水準に変更したときの推定値を求め，その差の 95%信頼区間により，実験結果の総合評価をするのが常である．表 4.21 に示すように線形和 $L^{(9)}$ が，現行水準の組み合わせで， $L^{(10)}$ が最適水準である．その差は $L^{(11)}$ で計算されていて，差の推定値は $L^{(12)} = -24.000$ ，分散が 4.333，95%信頼区間は (-32.957, -15.043) と計算されている．表 4.20 に示した各因子に対する推定値と 95%信頼区間，表 4.21 の現行水準と最適水準の差の推定値と 95%信頼区間の結果を図 4.4 に示す．

表 4.21 現行水準と最適水準の差および 95%信頼区間

$L^{(i)}$	A	B	C	D	$l^{(i)}_0$	$l^{(i)}_1$	$l^{(i)}_2$	$l^{(i)}_3$	$l^{(i)}_4$	θ^\wedge	線形和	分散	幅	95%信頼区間	
					x_0	x_1	x_2	x_3	x_4		L^\wedge	$Var(L^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
$L^{(9)}$	A ₁	B ₁	C ₁	D ₁	1	1	1	1	1	19.167	31.167	1.4444	5.171	25.996	36.338
$L^{(10)}$	A ₂	B ₂	C ₂	D ₂	1	-1	-1	-1	-1	3.583	7.167	1.4444	5.171	1.996	12.338
$L^{(11)}$	現行-最適				0	-2	-2	-2	-2	4.333	-24.000	4.3333	8.957	-32.957	-15.043
										2.167			$t_{0.05}(7-5)=$	4.303	
										1.917					

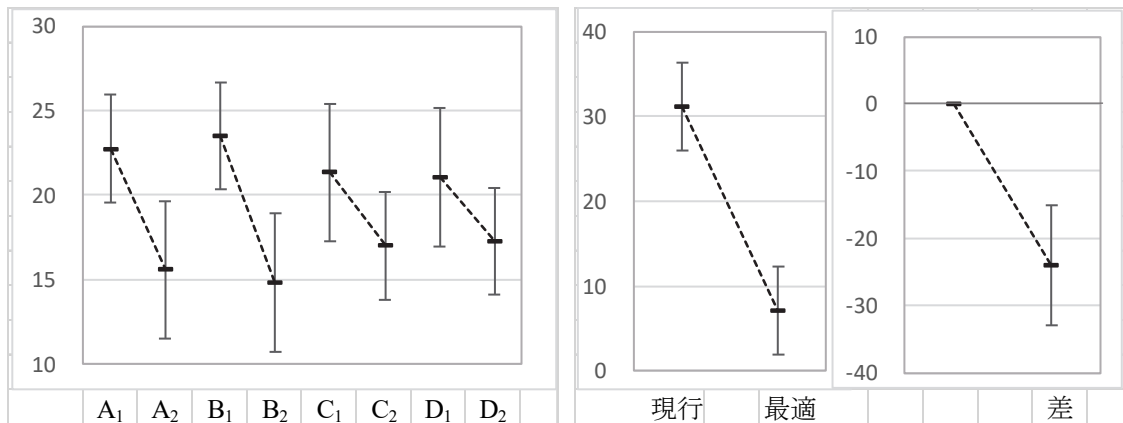


図 4.4 要因ごとの水準平均および現行水準と最適水準についての推定

4.4. JMP の「モデルのあてはめ」による解析

本章の原稿を推敲するために、2021年12月にオンラインで開催された JMP の Discovery Summit Japan 2021 にて、「線形モデルによる欠測値がある直交表の解析，一現行水準と最適水準の差の 95%信頼区間の推定を例にして」と題して講演した。この講演で用いたスライドを主体に、説明を付け加えて高橋（2022）を同じタイトルで、「続・高橋セミナー第 10 回」として公表した。ここでは、そのエッセンスを示す。

(1, -1)対比型ダミー変数

JMP の「モデルのあてはめ」を用いた「欠測値がある直交表の解析」を例示する。JMP ユーザは、直交表に限らず欠測値があるか否かなどに頓着なく「モデルのあてはめ」による分散分析を行なっているに違いない。これは、JMP の「モデルのあてはめ」は、表 4.22 左に示すような質的変数によるモデルの設定に対し、表 4.22 右の「コーディングのテーブル」に示すように内部で (1, -1) 対比型ダミー変数に変換してデザイン行列を内部で生成し、線形モデルによる解析をしている。ただし、このことがユーザには悟られないよう、徹底した配慮がなされている。

表 4.22 JMP の解析用ファイルとコーディングのテーブル

解析用ファイル							コーディングのテーブル						
No.	A	B	C	D	y		切片	A[1]	B[1]	C[1]	D[1]	y	
1	1	1	1	1	1	32	1	1	1	1	1	32	
2	2	1	1	2	2	23	2	1	1	-1	-1	23	
3	3	1	2	1	2	19	3	1	1	-1	1	19	
4	4	1	2	2	1	17	4	1	1	-1	-1	17	
5	5	2	1	1	2	19	5	1	-1	1	1	19	
6	6	2	1	2	1	20	6	1	-1	1	-1	20	
7	7	2	2	1	1	.	7	
8	8	2	2	2	2	8	8	1	-1	-1	-1	8	

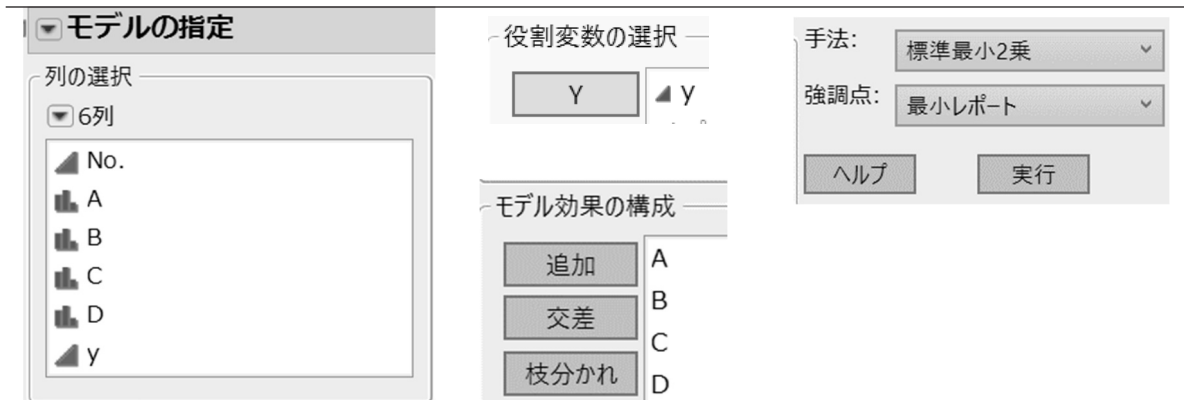
データは数値だが、文字変数として設定。
欠測値は、ピリオド「.」

内部で使用しているデザイン行列をファイルに出力。
(1, -1) 対比型ダミー変数が内部で設定されている。

JMP の「モデルのあてはめ」では、表 4.23 に示すように「役割変数の選択」で反応変数 y を、「モデル効果の構成」で説明変数 (A, B, C, D) を設定し、手法は「標準最小 2 乗」とし、強調点としては「最小レポート」を選択する。

表 4.24 に JMP の「モデルのあてはめ」の実行結果を示す。因子が互いに直交することを大前提にする分散分析表では、各因子の平方和が先に表示され、その後に「誤差」と「全体 or 合計」が示すのが一般的である。しかし、因子が互いに直交しない場合に、「各因子の平方和

表 4.23 JMP の「モデルのあてはめ」の設定画面



の合計」+「誤差平方和」=「合計」とはならない。そのために、直交を前提にしない JMP の分散分析表は、「モデル」に自由度 4 として全ての因子を込みにした平方和 303.0952 が先に示され、引き続き「効果の検定」で、各要因の平方和が示される。

表 4.24 JMP の「モデルのあてはめ」による分散分析表

分散分析					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
モデル	4	303.0952	75.7738	34.9725	
誤差	2	4.3333	2.1667		
全体(修正済み)	6	307.4286			0.0280*
効果の検定					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
A	1	77.0417	77.0417	35.5577	0.0270*
B	1	112.6667	112.6667	52.0000	0.0187*
C	1	28.1667	28.1667	13.0000	0.0691
D	1	22.0417	22.0417	10.1731	0.0858

「効果の検定」に出力されている「平方和」は、タイプ III の平方和である。
表 4.19 に示した Excel による結果に一致する。

効果の検定における (A, B, C, D) 4 因子の平方和の合計は、

$$S_A + S_B + S_C + S_D = 77.0417 + 112.6667 + 28.1667 + 22.0417 = 239.9168$$

となり、モデルの平方和 303.0952 とは異なる。このことは、従来の分散分析の解析方法である「平方和の分解」が成り立っていないことの証拠である。

どうしても直交性を保ちたいとの願望が、欠測値に対し「何かしらの値で補間する」などといった便宜的な方法が、まことしやかに語り続けられる元凶となっている。なお、JMP が 1990 年代に発売された当時から、「モデルのあてはめ」では、欠測値を含む解析に対応していた。

予測プロファイルを用いた推定

JMP の解析結果が、表 4.24 に示した分散分析表のみであれば、本節で取り上げたりしない。直交表を用いた実験で、各因子の「現行水準」の組合せによる推定値に対し、各因子の「最適水準」を組合せた推定値を統計的に比較することが、従来の解析法の中で定式化されてきた。そのために、「現行水準」の組合せ、および、「最適水準」の組合せそれぞれについての推定値と 95%信頼区間、さらに、それらの差の推定値と 95%信頼区間の推定も標準的な解析方法として定着してきた。統計的には、かなり込み入った解析方法であるが、現場の技術者が実践できることを目的にした解析方法が整備されてきたことは、先人たちの努力の賜物である。

欠測値が存在しない直交表の解析は、電卓レベルの計算で対応できるのであるが、不幸にも欠測値が出てしまったら最後、完全にお手上げの状況となる。なお、JMP に先立つこと 15 年前から SAS の GLM プロシジャでは、欠測値ある直交表の解析が行なえるような線形モデルを前提にした解析方法であった。図 4.5 に示すように「予測プロファイル」での GUI 操作により、「現行水準」および「最適水準」についての推定値および 95%信頼区間が得られている。さて、それらの差の推定と 95%信頼区間はどのようにして求めるのであろうか。

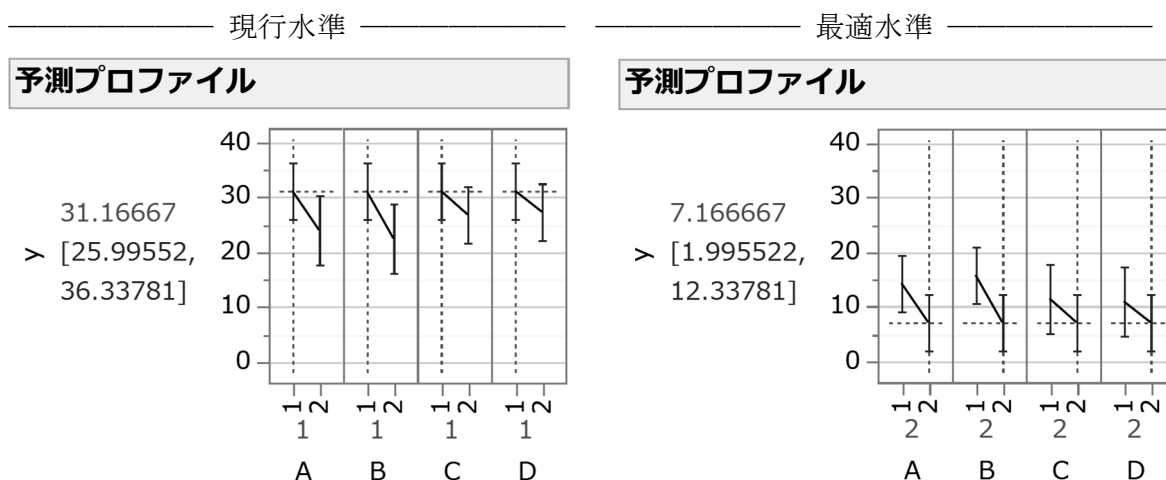


図 4.5 予測プロファイルを用いた現行水準と最適水準の推定
表 4.21 に示した Excel による結果に一致する。

カスタム検定による最適水準と現行水準の差の推定

表 4.25 に示すのは、(最適水準) と (現行水準) の差を求めるために「推定値」のプルダウン・メニューにある「カスタム検定」の設定画面で、「最適 - 現行」と入力し、パラメータに謎めいた (-2, -2, -2, -2) を設定している。これは、表 4.21 で示した Excel による設定に一

致している. この「カスタム検定」では, JMP の内部のデザイン行列 X の生成方式を理解し, 線形モデルによる線形和についての推定方法に精通していること必要である.

表 4.25 「カスタム検定」による謎めいたパラメータ設定

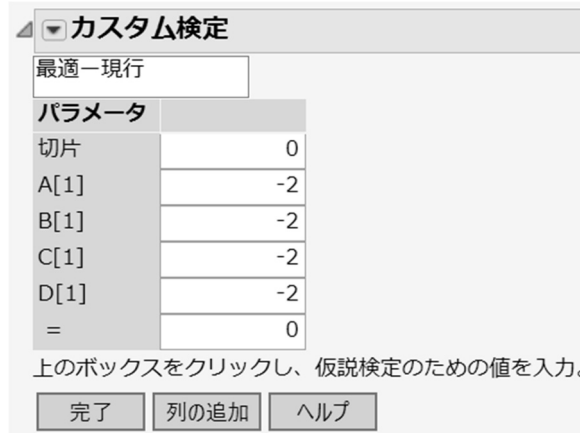
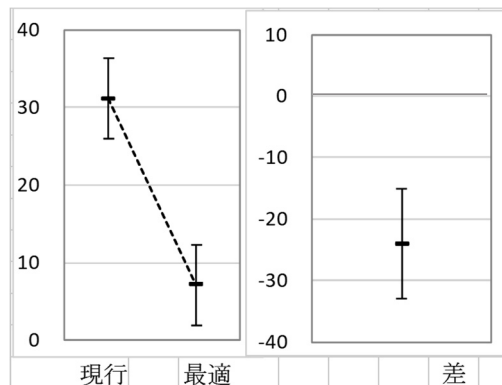


表 4.21 の「現行-最適」で示されている $I^{(1)} = (0 \ -2 \ -2 \ -2 \ -2)$ を用いる.

表 4.26 に示すのは, 「カスタム検定」による結果であり, 差の推定値 = 24.0000 と標準誤差 2.0817 が出力されている. ただし, 95%信頼区間は出力されないので, 表 4.21 に示すように別途 Excel で計算し, 図 4.4 に示した折れ線グラフの結果を示している.

表 4.26 最適水準と現行水準の差の推定結果

最適 - 現行	
パラメータ	
切片	0
A[1]	-2
B[1]	-2
C[1]	-2
D[1]	-2
=	0
値	-24.0000
標準誤差	2.0817
t値	-11.5292
p値(Prob> t)	0.0074
平方和	288.0000



Excel の折れ線グラフ
ひげは 95%信頼区間

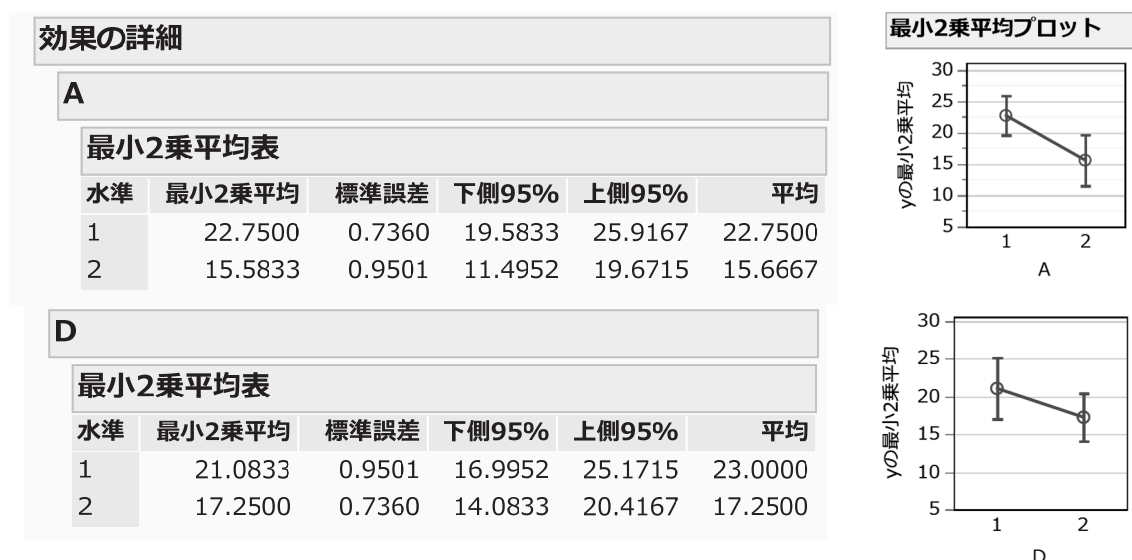
最小 2 乗平均

辛口の論評ではあるが, JMP の「モデルのあてはめ」は, 汎用的であるがために, 出力結果が冗長でまとまりに欠けるきらいがある. 表 4.27 で示す「最小 2 乗平均」とは, 何なのか? どのように計算されたものなのか? 「最小 2 乗平均 (Lsmeans)」は, JMP および SAS で使

用されている方言のようなもので、統計の専門用語として認知されていない。ただし、線形モデルで推定されたパラメータを用いて各種の推定値を求めた場合に、その推定値のことを「最小2乗平均, Lsmeans」とした定義は、絶妙であり、実務家には受け入れられているが、その定義には、様々な講釈が付きまとい、謎めいた存在となっている。

欠測値あるは欠測セルがある場合の水準平均をどのように定義するのかに関連しているが、きちっと解説している日本語の成書を見いだすのは困難である。表 4.27 に示すのは、JMP の「効果の詳細」で出力される各因子の水準に対する「最小2乗平均」である。表 4.20 に示したように Excel の基本機能だけを用いて JMP および SAS から出力される「最小2乗平均」を再現することにより、謎めいた「最小2乗平均」について理解が深められる。

表 4.27 「モデルのあてはめ」による最小2乗平均



各種のダミー変数の歴史的背景

平方和の分解を前提にした場合に、直交表として使いやすいのは (1, 2) 型である。これは、各因子の第 1 水準、第 2 水準に対応していて「第 1 水準の合計、第 2 水準の合計を求め」など手順の説明として分かりやすいからであろう。表 4.28 に示すように、朝香ら (1988), 「新版 品質管理便覧 第 2 版, p446-7.」, 「7.1.4.2 節 直交表の性質とその利用」に直交表の歴史的背景が詳しく述べられている。

- (a) (1, -1) 基本型
- (b) (1, 2) 型, +1 を 1, -1 を 2 におきかえた。実験計画に使いやすいように、田口¹⁴⁾・小西氏¹⁵⁾によって提唱された。
- (c) (0, 1) 型, +1 を 0, -1 を 1 におきかえたもの。嶋田氏¹⁶⁾によって提唱された。

従って、JMP の内部で使われているデザイン行列でのダミー変数は、(0, 1) 型ではなく直交表の基本型である (1, -1) 型であることが確認される。

表 4.28 各種の直交用「朝香ら(1988), 新版 品質管理便覧 第2版, p446-7.」

(a) 基本型					(b) $L_4(2^3)$				(c) $H_{2,4}$			
No.	(0)	列 番			No.	列 番			No.	列 番		
		(1)	(2)	(3)		(1)	(2)	(3)		(1)	(2)	(3)
1	+1	+1	+1	+1	1	1	1	1	1	0	0	0
2	+1	+1	-1	-1	2	1	2	2	2	0	1	1
3	+1	-1	+1	-1	3	2	1	2	3	1	0	1
4	+1	-1	-1	+1	4	2	2	1	4	1	1	0

(b) : 14) 田口玄一(1977,1962) : 改訂新版実験計画法(上), 実験計画法(下), 丸善

(b) 15) 小西省三(1965) : 例題演習実験計画法, 日刊工業新聞社

(c) 16) 島田正三(1958) : やさしい直交配列の話, 日本規格協会

JMP の使い勝手

さて、JMP の使い勝手であるが、因子を (1, 2) 型の名義尺度として解析することにより、欠測値のある直交表の解析が容易にできる。最適水準の組合せ、現行水準の組合せの推定と 95%信頼区間の計算も予測プロファイルで GUI 的にできることは、現場の技術者にとって扱い易い。ただし、カスタム検定を用いて (最適水準と現行水準) の差の推定は、JMP の内部計算の仕組みを理解なしには不可能である。因子ごとの水準平均と 95%信頼区間のグラフは、個別の表示となり、まとめて表示できないもどかしさが残るが、JMP の出力画面を、範囲を自由に選択しコピーし、Excel のセルにテキストとしてペーストできるので、再活用することは容易である。

JMP の「モデルのあてはめ」を用いて欠測値がある直交表の解析方法を示してきたのであるが、JMP はライセンス契約が必要な有償ソフトであり、誰でも手軽に使える統計ソフトではない。無償の統計ソフトとして R が脚光を浴びているが、欠測値がある直交表の解析は、一筋縄では行かない。無償で継続的に提供されている OnDemand SAS の GLM プロシジャを用いれば、R よりも豊富な結果が得られるので、次節でプログラムと結果の要約を示す。

4.5. SAS/GLM による欠測値がある直交表の解析

無償で継続的に提供されている On Demand SAS の GLM プロシジャを用い、欠測値がある直交表の解析方法を示す。GLM プロシジャを用いた解析方法の詳細は、高橋ら (1989)、「SAS による実験データの解析」に記されているので参照されたい。ここでは、JMP による解析結果を SAS で得るためのプログラムと結果の見方について示す。

GLM プロシジャによる解析プログラム

GLM プロシジャでは、変数を名義尺度として扱う場合は、class ステートメントで定義し、各因子の最小 2 乗平均の算出は、lsmeans ステートメントを用いる。現行水準と最適水準およびその差は、estimate ステートメントで設定するが、その際に GLM プロシジャの内部でデザイン行列への展開方法に留意しなければならない。GLM プロシジャでは、JMP の場合と異なり、式 (4.44) で示した母数に無駄の有る 1 次従属型のデザイン行列に展開されている。ただし、解析に際しては、最後の水準を基準とした (1, 0) 型のダミー変数に準じた解析が行なわれている。

SAS のプログラム 4.1

```
Title1 'L8_missing.sas 2022-3-15 Y.Takahashi ' ;

data d01 ;
  input No A B C D y ;
/*No. A B C D y */
datalines ;
  1 1 1 1 1 32
  2 1 1 2 2 23
  3 1 2 1 2 19
  4 1 2 2 1 17
  5 2 1 1 2 19
  6 2 1 2 1 20
  7 2 2 1 1 .
  8 2 2 2 2 8
;

proc print data=d01 ; run ;

proc glm data=d01 ;
  class A B C D ;
  model y = A B C D / solution ;
  lsmeans A B C D / stderr cl ;
  estimate '現行' intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
  estimate '最適' intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
  estimate '最適-現行' intercept 0 A -1 1 B -1 1 C -1 1 D -1 1 ;
  estimate '最小 2 乗平均 A1f' intercept 1 A 1 0 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;
  estimate '最小 2 乗平均 A1s' intercept 1 A 1 0 ;
  estimate '最小 2 乗平均 A2f' intercept 1 A 0 1 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;
  estimate '最小 2 乗平均 A2s' intercept 1 A 0 1 ;
run ;
```

GLM プロシジャの出力

表 4.29 に示すのは、DATA ステップで読み込んだテキストデータを PRINT プロシジャの HTML 形式の出力を Excel に取り込み、体裁を整えてワードに「図 (拡張メタファイル)」で貼り付けた結果である。データはすべて「数値データ」となっている。Obs=7 では、y=. のように欠測値となっている。

表 4.29 DATA ステップ読み込んだデータの proc print による出力

Obs	No	A	B	C	D	y
1	1	1	1	1	1	32
2	2	1	1	2	2	23
3	3	1	2	1	2	19
4	4	1	2	2	1	17
5	5	2	1	1	2	19
6	6	2	1	2	1	20
7	7	2	2	1	1	.
8	8	2	2	2	2	8

表 4.30 に示すのは、model ステートメントで設定した変数をすべて込みにしたモデル全体に対する分散分析表で、表 4.24 の JMP の結果に対応する。

表 4.30 proc glm モデル全体に対する分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	4	303.0952	75.7738	34.97	0.028
Error	2	4.3333	2.1667		
Corrected Total	6	307.4286			
R2 乗	変動係数	Root MSE	y の平均		
0.9859	7.4665	1.47196	19.71429		

表 4.31 に示すのは、デフォルトで出力されるタイプ I の平方和とタイプ III の平方和であり、タイプ I の平方和は、第 3.2 節で示した逐次平方和であり、タイプ III の平方和は、表 4.24 に示した JMP のデフォルトの平方和に一致する。

表 4.31 タイプ I の平方和 および タイプ III の平方和

要因	自由度	Type I 平方和	平均平方	F 値	Pr > F
A	1	86.0119	86.0119	39.70	0.0243
B	1	176.8167	176.8167	81.61	0.0120
C	1	18.2250	18.2250	8.41	0.1012
D	1	22.0417	22.0417	10.17	0.0858
要因	自由度	Type III 平方和	平均平方	F 値	Pr > F
A	1	77.0417	77.0417	35.56	0.0270
B	1	112.6667	112.6667	52.00	0.0187
C	1	28.1667	28.1667	13.00	0.0691
D	1	22.0417	22.0417	10.17	0.0858

表 4.32 に示すのは、model ステートメントの solution オプションで出力された結果である。推定値は、第2水準を基準とした第1水準との差となっている。さて、切片 Intercept の推定値 7.1667 は、何を推定しているのであろうか？

表 4.32 最後の水準を基準とする (1, 0) 型ダミー変数でのパラメータ推定値

パラメータ	推定値		標準誤差	t 値	Pr > t
Intercept	7.1667	B	1.2019	5.96	0.027
A 1	7.1667	B	1.2019	5.96	0.027
A 2	0	B	.	.	.
B 1	8.6667	B	1.2019	7.21	0.0187
B 2	0	B	.	.	.
C 1	4.3333	B	1.2019	3.61	0.0691
C 2	0	B	.	.	.
D 1	3.8333	B	1.2019	3.19	0.0858
D 2	0.0000	B	.	.	.

Note: 'X' は特異行列です。w 正規方程式には一般化逆行列が使用されています。
w 文字 'B' が付けられた推定値は一意的な推定値ではありません。

切片は、解析モデルに含まれている変数が、すべて 0 の場合の推定値である。したがって、(1, 0) 型ダミー変数とみなす GLM プロシジャでは、(A₂, B₂, C₂, D₂) 水準の場合に「すべて 0」となることから、切片の推定値 7.1667 は、すべての因子が第2水準の場合の推定値と理解される。(1, -1) 対比型ダミー変数の場合は、表 4.18 に示したように、切片の推定値は、データの構造式における一般平均 $\hat{\mu} = 19.1667$ と意味付けられるのであるが、(1, 0) 型ダミー変数の場合は、すべての因子が第2水準の場合の

$$\widehat{A_2 B_2 C_2 D_2} = 7.1667 \tag{4.59}$$

の推定値であると説明できる。では、一般平均の推定値 $\hat{\mu} = 19.1667$ は、どのように推定したら良いのだろうか、切片に他の推定値を 1/2 倍して加えたもので、

$$\hat{\mu} = 7.1667 + \frac{7.1667 + 8.6667 + 4.3333 + 3.8333}{2} = 19.1667 \tag{4.60}$$

のように推定することができる。

Lsmeans ステートメントによる最小 2 乗平均

このような厄介な推定をから理解しやすい推定値を得るために GLM プロシジャでは、表 4.33 に示すように lsmeans ステートメントで質的変数としての (A, B, C, D) を設定するだけで、各因子の全ての水準に対し最小 2 乗平均を出力できるようになっている。因子 A の A₁ 水準の最小 2 乗平均は、

$$\hat{A}_1 = 7.1667 + 7.1667 + \frac{8.6667 + 4.3333 + 3.8333}{2} = 22.7500 \tag{4.61}$$

であり、A₂ 水準の最小 2 乗平均は

$$\hat{A}_2 = 7.1667 + 0.0 + \frac{8.6667 + 4.3333 + 3.8333}{2} = 15.5833 \tag{4.62}$$

として計算されている。

表 4.33 各因子の最小 2 乗平均および 95%信頼区間

A	最小 2 乗平均	標準誤差	Pr > t	95% 信頼区間	
1	22.7500	0.7360	0.0010	19.5833	25.9167
2	15.5833	0.9501	0.0037	11.4952	19.6715
B	最小 2 乗平均	標準誤差	Pr > t	95% 信頼区間	
1	23.5000	0.7360	0.0010	20.3333	26.6667
2	14.8333	0.9501	0.0041	10.7452	18.9215
C	最小 2 乗平均	標準誤差	Pr > t	95% 信頼区間	
1	21.3333	0.9501	0.0020	17.2452	25.4215
2	17.0000	0.7360	0.0019	13.8333	20.1667
D	最小 2 乗平均	標準誤差	Pr > t	95% 信頼区間	
1	21.0833	0.9501	0.0020	16.9952	25.1715
2	17.2500	0.7360	0.0018	14.0833	20.4167

図 4.5 に示した JMP での現行水準と最適水準の推定は、GUI 操作により推定できるが、SAS の GLM プロシジャでは、GLM の内部で生成される式 (4.44) で示した母数に無駄の有るデザイン行列に即して estimate ステートメントで次のように設定する必要がある。

```
estimate '現行' intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
estimate '最適' intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
```

最適水準から現行水準を引く場合は、「最適」から「現行」を引けばよい。JMP では、表 4.26 に示したようにカスタム検定を使うのであるが、内部で生成されるデザイン行列の知識が求められる。JMP に比べて、GLM プロシジャでは首尾一貫している。ただし、「最適」と「現行」水準の設定ができる事が前提である。

```
estimate '最適' intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
-) estimate '現行' intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
estimate '最適-現行' intercept 0 A -1 1 B -1 1 C -1 1 D -1 1 ;
```

表 4.34 現行水準と最適水準の差の推定

パラメータ	推定値	標準誤差	t 値	Pr > t
現行	31.1667	1.2019	25.93	0.0015
最適	7.1667	1.2019	5.96	0.0270
最適-現行	-24.0000	2.0817	-11.53	0.0074

Estimate ステートメントによる最小 2 乗平均

各因子の水準に対する最小 2 乗平均計算は、lsmeans ステートメントで求めるのが一般的であるが、estimate ステートメントを用いてもできる。因子 A の第 1 水準の場合であれば、他の因子のそれぞれの水準の平均値 0.5 を与える必要があるが、設定しなくても 0.5 が自動的に補われる。

```
estimate '最小2乗平均A1f' intercept 1 A 1 0 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;
estimate '最小2乗平均A1s' intercept 1 A 1 0 ;
```

表 4.35 estimate ステートメントによる最小2乗平均

パラメータ	推定値	標準誤差	t 値	Pr > t
最小2乗平均A1f	22.7500	0.7360	30.9100	0.0010
最小2乗平均A1s	22.7500	0.7360	30.9100	0.0010
最小2乗平均A2f	15.5833	0.9501	16.4000	0.0037
最小2乗平均A2s	15.5833	0.9501	16.4000	0.0037

(1, 0)型および(0, 1)型ダミー変数の扱いにくさ

各種の最小2乗平均は、推定された回帰パラメータを用いて、各因子の水準平均を算出しているのであるが、(1, 0)型ダミー変数の場合でも、(0, 1)型ダミー変数の場合でも、ここに示したように推定されたパラメータから簡単には計算することができない。(1, -1)対比型ダミー変数の場合であれば、表 4.20 に示したように、切片（一般平均 μ の推定値）にデータの構造式のパラメータの推定値をプラス・マイナスするといった直観的な計算で求められるし、他者に説明することも容易である。

SAS の GLM プロシジャとの出会いは、30 年以上前であったが、lsmeans ステートメントによって計算される最小2乗平均は、種々の実験計画のモデルにおける推定の問題について、華麗に対応できることに感激した。そこで、その計算原理を高橋ら（1989）の第 15 章「4 種の平方和と LSMEAN」、第 16 章「GLM プロシジャの計算方式」としてまとめた。ただし、一般的な方法として普及活動を行なうことがためられた。

多くの実験計画のモデルを記述する際の線形モデル、欠測値のある L_8 直交表であれば、

$$y_i = \mu + \theta_1 a + \theta_2 b + \theta_3 c + \theta_4 d + \varepsilon_i \quad (4.63)$$

とするのが自然であり、ここで μ は、一般平均の推定値となると思うのが普通であろう。ところが、(a, b, c, d) に (1, 0) 型ダミー変数を与えると、式 (4.59) で示したように μ は、すべての因子が第 2 水準の場合の推定値 $\widehat{A_2 B_2 C_2 D_2} = 7.1667$ となることを示した。(1, -1)対比型ダミー変数を (a, b, c, d) に与えた場合には、表 4.18 に示すように、 μ は一般平均の推定値 19.1667 となり、ダミー変数に対するパラメータの推定値は、 μ に対する効果となることが示され、線形モデルを素直に解釈できる。

無償で継続的に提供されている On Demand SAS の GLM プロシジャにより、欠測値のある直交表の解析が行なえ、現行水準と最適水準との差の推定も行なえることを示した。ただし、GLM プロシジャが採用している (1, 0) 型ダミー変数の特質を理解した上で使うことが必要である。

5. デザイン行列を用いた回帰分析の基礎

シグマが出てきただけで読み飛ばしたくなる人たちの気持ちは、その式を見ても実際のデータで自ら計算を行なう意欲がわからないからだと思われる。ましてや、行列の計算式が出てきたときに、実際のデータで自ら計算することができないものに対し、興味を持つことができないだけでなく嫌悪感さえ抱くと推測する。そのために、多くの人達が、統計ソフトの使い方、結果の見方さえ習得すれば、十分だと思ってしまいがちになるのではないだろうか。このような人達を念頭にし、シグマによって記述されている回帰分析を題材にし、Excelの基本的な計算機能を用いることにより、何の苦痛もなくサクサクとシグマの計算ができる事を示す。さらに、Excelシート上に展開された矩形データが行列の基本であることを導入し、Excelの行列関数により、複雑なシグマで表記された計算が、瞬時にできることを示す。本章では、これらの計算手段を用いることにより、伝統的なシグマ表記の回帰分析に代わる、新しい解析手順を導入する。

5.1. 回帰分析における各種の推定

第1章では、各種の質的変数を含む回帰分析の事例について、Excelの分析ツールの一つである「回帰分析」を用いた解析方法を示し、統計ソフトによる検証も合わせて示してきた。Excelの統計解析の機能は、限定的であり正確性について MaCullough ら (2008), 「On the accuracy of statistical procedures in Microsoft Excel 2007」での Excel 2007 に対する批判があることは十分承知している。幸いなことに、Excelの回帰分析に限ってみれば、各種の推定結果のグラフ表示への連携などにより、総合的には統計ソフトに勝るとも劣らないことを示してきた。

質的変数を含む場合に、いわゆるダミー変数を使うことにより、統計ソフトの回帰分析による解析ができることは、良く知られている。これまでの章では、質的変数に対するダミー変数と量的変数の積で生成された交互作用の変数を含むデザイン行列 X を用い、Excelの回帰分析を適用し、さらに、行列関数による回帰パラメータの共分散行列 $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ の推定方法を示してきた。回帰分析に派生する各種の推定値の分散を求めるために、パラメータの共分散行列 $\Sigma(\hat{\beta})$ が中心的な役割を演じていることを強調してきた。

各種の推定値に対して 95%信頼区間を計算し、その結果のグラフ表示に散布図および折れ線グラフを活用し、統計ソフトでは対応できない複合的なグラフ表示が、同じ計算シート内で容易にできることも示してきた。特に交互作用がある場合のこれらのグラフ表示は、結果の統計的な解釈をする際に助けになる。

質的変数が含まれる場合に Excel の「回帰分析」を用いる場合は、自らダミー変数を生成する必要があり煩わしい。多くの統計ソフトでは、内部でダミー変数に展開し、交互作用を含めた場合でも内部でしかるべき変数に展開するので、大変便利である。ただし、第 1.4 節の SAS による結果の検証では、各種の推定をする際には、内部で生成されたデザイン行列を念頭にし、推定のため `estimate` ステートメントによる細かな設定が必要で、容易なことではないことも示してきた。

これらの状況を鑑みるにつれ、回帰分析から派生する各種の推定の課題に対処するためには、Excel によるデザイン行列 X の生成、回帰分析の実施、各種の推定、推定値の 95%信頼区間のグラフ化といった一連の解析マナーの習得が必要不可欠と思われる。その結果とし、統計ソフトを自在に使いこなし、活用するために必須の知識と経験が得られるのではないだろうか。

第 1 章では、全く触れなかったが、ほとんど全ての統計に関する成書での回帰分析（重回帰分析を含む）の計算手順は、手計算時代の簡便な画一的な計算法が踏襲されている。そのために、第 1.2 節に示したように共分散分析の解析法もその影響を受け、第 6 章で詳しく述べるが、分散分析と回帰分析の複合的な解析法が示されており、理解し応用することが難しい。そこで、第 1 章で導入した事例を用いて、デザイン行列 X を用いた回帰分析についての基礎を最初に示し、合わせて伝統的な手計算による解析法を対比し、その問題点を浮き彫りにする。

ドレーパ・スミス著、中村訳（1968）「応用回帰分析」の第 1 章には、シグマを使った回帰分析、第 2 章には、同じデータを用いたデザイン行列 X を用いた計算法が丁寧に例示され、重回帰分析へ橋渡しがなされている。さらに、Draper, Smith (1998), *Applied Regression Analysis* 3rd ed. も参考にし、シグマを使った計算に対比したデザイン行列 X を用いた計算法の関係を並列的に示すことにより、シグマを用いた計算になれ親しんだ人達にデザイン行列 X を用いた解析の利便性を強調したい。用いるデータは、第 1.3 節の抗うつ薬の臨床試験のデータである。

5.2. デザイン行列 X を用いた行列計算の基礎Excel シート上にデザイン行列 X を設定

反応変数 Y_i の列ベクトルを Y ，説明変数のデザイン行列を X ，推定したいパラメータのベクトルを β ，誤差の列ベクトルを ε とする。データは，表 1.23 抗うつ薬のプラセボ群のデータである。まず，Excel シートの好きな場所に，これと同じものを再現してもらいたい。ここに示したのは，Excel シート上の該当セルをコピーして，ワード上で「図(拡張メタファイル)」によりペーストした結果が示されている。

			X_0	X_1				
$Y =$	4	$X =$	1	18	$\beta =$	β_0	$\varepsilon =$	ε_1
	7		1	21		β_1		ε_2
	10		1	24				ε_3
	10		1	24				ε_4
	2		1	18				ε_5
	11		1	27				ε_6
	9		1	21				ε_7
	5		1	21				ε_8
	11		1	26				ε_9
	8		1	20				ε_{10}
	10×1		10×2			2×1		10×1

フォントのサイズは 10 ポイント，半角英数字のフォントは Times New Roman とする。ギリシャ文字 β は「ベータ」と入力すると変換され全角の「 β 」となる。半角の「 β 」にするためにフォントを Times New Roman に変更してイタリック β を選択すると「 β 」となる。さらに太字にするためにボールド β を選択するとベクトル「 β 」となる。

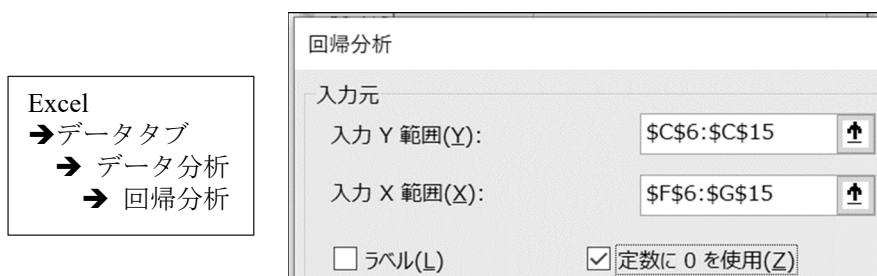
同様に「いぷしろん」で全角の「 ε 」に変換し，フォントを Times New Roman に変更してイタリックにすると「 ε 」となる。添え字の「1」を加えてイタリックを外して「 ε_1 」とし，「1」を「セル書式の設定」で「下付き」にすると「 ε_1 」となる。「 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{10}$ 」は，フィルハンドルで「 ε_1 」を引き延ばすと自動的に得られる。パラメータの β_0 と β_1 も同様の手順で作成する。

このように統計で使うギリシャ文字を Excel で自由に扱えるようになることから初めてもらいたい。デザイン行列の「 X 」は，半角の大文字「 X 」をボールド「 X 」とし，さらにイタリックにすると「 X 」となる。

デザイン行列 X の一般的な表記は，角括弧 $[\cdot \cdot \cdot]$ あるいは括弧 $(\cdot \cdot \cdot)$ で挟むのであるが，Excel シート上では表記しづらいので，「 β 太い外枠(I)」を用いて矩形データ全体を

囲む表記 $\boxed{\dots}$ を用いる。デザイン行列 \mathbf{X} は、大きさが 10 行×2 列で、10 行×1 列のベクトル \mathbf{X}_0 と \mathbf{X}_1 を並べたものである。列ベクトル \mathbf{X}_0 の値を全て 1 としているのは、回帰直線の Y 切片を推定するためである。列ベクトル $\boldsymbol{\beta}$ は、推定したい切片「 β_0 」と傾き「 β_1 」を含んでいる。

多くの統計ソフトの回帰分析では、モデル式の設定で、説明変数 \mathbf{X}_1 のみを与え、切片を推定するための変数 \mathbf{X}_0 の設定を必要としない。原点を通る回帰直線を求めたい場合には、切片を含まない 10 行×1 列の列ベクトル \mathbf{X}_1 のみのデザイン行列とする必要がある。そのために、統計ソフトでは、「切片を含まない」などのオプションを設定することにより原点を通る回帰直線を推定することができる。Excel の「データ分析ツール」の一つである「回帰分析」では、次に示すように、



「定数に 0 を使用」オプションをオンと設定する。通常回帰分析では、「定数に 0 を使用」オプションを選択（オフのまま）せず、列ベクトル \mathbf{X}_1 の範囲のみを設定する。そして、列ベクトル \mathbf{Y} の範囲を設定して、回帰分析を行なう。このように、Excel でも他の統計ソフトの回帰分析では、切片を含まない設定を暗黙の前提としている。そのため、原点を通る回帰直線のあてはめるために「切片を含まない」とのオプションが必要となる。

列ベクトル \mathbf{Y} 、デザイン行列 \mathbf{X} 、列ベクトル $\boldsymbol{\beta}$ 、列ベクトル $\boldsymbol{\varepsilon}$ を用いた回帰モデル式

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{5.1}$$

は、何を意味するのだろうか。デザイン行列 \mathbf{X} と列ベクトル $\boldsymbol{\beta}$ の積 $\mathbf{X}\boldsymbol{\beta}$ は、デザイン行列 \mathbf{X} のある行のセルと、列ベクトル $\boldsymbol{\beta}$ のセル同士を順番に掛けて足した「積和」の計算を意味する。これをデザイン行列 \mathbf{X} の 1 行目から 10 行目まで繰り返し、大きさが 10×1 の新たな列ベクトルが生成される。右辺の $(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$ は、大きさが 10×1 の列ベクトル同士の足し算で、同じ行のセル同士の足し算となり、大きさが 10×1 の列ベクトルとなる。

列ベクトル \mathbf{Y} と $(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$ を等号で結んだ式

$$\mathbf{Y} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

は、 \mathbf{Y} の行方向に展開して

	$(X\beta)$	ε	$(X\beta + \varepsilon)$
$Y = X\beta + \varepsilon =$	$1\beta_0 + 18\beta_1$	ε_1	$1\beta_0 + 18\beta_1 + \varepsilon_1$
	$1\beta_0 + 21\beta_1$	ε_2	$1\beta_0 + 21\beta_1 + \varepsilon_2$
	$1\beta_0 + 24\beta_1$	ε_3	$1\beta_0 + 24\beta_1 + \varepsilon_3$
	$1\beta_0 + 24\beta_1$	ε_4	$1\beta_0 + 24\beta_1 + \varepsilon_4$
	$1\beta_0 + 18\beta_1$	ε_5	$1\beta_0 + 18\beta_1 + \varepsilon_5$
	$1\beta_0 + 27\beta_1$	ε_6	$1\beta_0 + 27\beta_1 + \varepsilon_6$
	$1\beta_0 + 21\beta_1$	ε_7	$1\beta_0 + 21\beta_1 + \varepsilon_7$
	$1\beta_0 + 21\beta_1$	ε_8	$1\beta_0 + 21\beta_1 + \varepsilon_8$
	$1\beta_0 + 26\beta_1$	ε_9	$1\beta_0 + 26\beta_1 + \varepsilon_9$
	$1\beta_0 + 20\beta_1$	ε_{10}	$1\beta_0 + 20\beta_1 + \varepsilon_{10}$

$$\left. \begin{aligned} Y_1 &= 1\beta_0 + 18\beta_1 + \varepsilon_1 \\ Y_2 &= 1\beta_0 + 21\beta_1 + \varepsilon_2 \\ &\vdots \\ Y_{10} &= 1\beta_0 + 20\beta_1 + \varepsilon_{10} \end{aligned} \right\} \quad (5.2)$$

として展開できる。また、添え字を使った式

$$Y_i = \beta_0 X_{0,i} + \beta_1 X_{1,i} + \varepsilon_i \quad (i=1, 2, \dots, 10) \quad (5.3)$$

としても、同じ回帰モデルである。デザイン行列 X を使った式 (5.1) は、最も簡潔な表記となっている。その意味することを Excel で表記した矩形データと対比し、連想することにより行列計算が身近なものとなることを期待している。

行列計算の実際

Excel シート上の列ベクトル β に適当な数値 (-11.0, 0.8) を入れて、実際に $X\beta$ の計算にチャレンジしてみよう。手順は、以下に示すように、

	A	B	C	D	E	F	G	H	I
1		X							
2		X ₀	X ₁		β		X β		
3	X =	1	18		-11.0	=	=Mmult(B3:C12,E3:E4)		
4		1	21		0.8				
5		1	24						
6		1	24						
7		1	18						
8		1	27						
9		1	21						
10		1	21						
11		1	26						
12		1	20						
13		10×2			2×1		10×1		

	A	B	C	D	E	F	G
1		X					
2		X ₀	X ₁		β		X β
3	X =	1	18		-11.0	=	3.40
4		1	21		0.8		5.80
5		1	24				8.20
6		1	24				8.20
7		1	18				3.40
8		1	27				10.60
9		1	21				5.80
10		1	21				5.80
11		1	26				9.80
12		1	20				5.00
13		10×2			2×1		10×1

=Mmult (B3:C12, E3:E4)
 =Mmult (X の範囲, β の範囲)

- 1) $X\beta$ の計算結果となる (10 行×1 列) の矩形を選択し太い外枠で囲み, 上部に「 $X\beta$ 」を入力する.
- 2) $X\beta$ の (10 行×1 列) の矩形を選択し, 数式バーに, 行列の積の関数「=Mmult (X の範囲を選択, β の範囲を選択)」のように入力する. なお, 「 X の範囲を選択」は, マウスで (10 行×2 列) の行列 X の範囲の選択することを意味する.
- 3) $X\beta$ の最初の行で計算式の入力が終わったら「コントロールキー」と「シフトキー」を同時に押しながら「エンター」すると行列の積和の計算が行なわれる.

デザイン行列 X の転置と積和

行列 A と行列 B の掛け算は, 左側の行列 A の「行」のセルと右側の行列 B の「列」のセルを順番に掛けて加えた和 (積和) の行列として定義されている. 行列の積和の計算は, “行” 方向と “列” 方向であって, “列” 方向と “行” 方向の “列・行” ではなく, あくまで “行・列” の順番である.

A			B			AB			
→	→	→	↓			=	$A_{11}B_{11}+A_{12}B_{21}+A_{13}B_{31}$		
→	→	→					$A_{21}B_{11}+A_{22}B_{21}+A_{23}B_{31}$		
→	→	→					$A_{31}B_{11}+A_{32}B_{21}+A_{33}B_{31}$		

(10 行×2 列) のデザイン行列 X と (10 行×2 列) X の積 XX の内側のサイズは, (2 列 vs. 10 行) と列と行の数が異なり積和の計算ができない. そこで, 最初の X について行と列を入れ替え, (2 行×10 列) の X^T (転置行列) とする. 行列の転置は, X の範囲をコピーし, 転置したい先頭位置に対し, 「形式を選択し貼り付け」, 「行/列の入れ替え」で作成する.

積 $X^T X$ は, (2 行×10 列 vs. 10 行×2 列) と内側の数が一致し, 積和の計算ができる. このように, 行列の積が成り立つのは, (2 行×10 列) の X^T と (10 行×2 列) の X のように隣り合う行列の内側が, 10 列と 10 行のように大きさが完全に一致する必要がある. 行列の積 $X^T X$ の結果は, 外側の (2 行×2 列) の大きさとなる.

行と列の入れ替えは, 転置 (Transpose) といい, X^T のように表記する. なお, X' あるいは X' と表記する場合もあり, 様々である. 行列の積は, Excel の Mmult () 関数を使い

- X_0^T 行と X_0 列の 積和=10 を ($X^T X$) の 1 行 1 列目へ
- X_1^T 行と X_0 列の 積和=220 を ($X^T X$) の 2 行 1 列目へ
- X_0^T 行と X_1 列の 積和=220 を ($X^T X$) の 1 行 2 列目へ
- X_1^T 行と X_1 列の 積和=4928 を ($X^T X$) の 2 行 2 列目へ

5.3. 偏差平方和による回帰パラメータの推定

回帰パラメータの推定

求めたい回帰式を,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i=1,2,\dots,n \quad (5.6)$$

としたときに, 誤差 ε_i の平方和 Q は,

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (5.7)$$

となる. 通常回帰分析において誤差 ε_i は, 平均が 0, 分散が σ^2 の正規分布に従うと仮定するのだが, 実際にはどんな分布であっても最小 2 乗法での計算が可能である. そのために, 最小 2 乗法による回帰分析が, ゆうずうむげに無批判的に使われ続けている最大の理由である. しかし, 基本中の基本であるので, 丁寧に説明する. 誤差 ε_i の平方和 Q をパラメータ β_0 および β_1 で偏微分すると, 式 (5.8) が得られる.

$$\left. \begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum_{i=1}^n [(Y_i - \beta_0 - \beta_1 X_i) X_i] \end{aligned} \right\} \quad (5.8)$$

偏微分に不慣れな場合には, 段階的な学習が必要である. それぞれの誤差の平方 ε_i^2 は,

$$\left. \begin{aligned} \varepsilon_i^2 &= (Y_i - \beta_0 - \beta_1 X_i)^2 \\ &= Y_i^2 + \beta_0^2 + \beta_1^2 X_i^2 - 2\beta_0 Y_i - 2\beta_1 X_i Y_i + 2\beta_0 \beta_1 X_i \end{aligned} \right\} \quad (5.9)$$

と式を展開できるので, β_0 での偏微分は, 他の変数を定数と見なした微分なので,

$$\left. \begin{aligned} \frac{\partial \varepsilon_i^2}{\partial \beta_0} &= \frac{\partial (Y_i^2 + \beta_0^2 + \beta_1^2 X_i^2 - 2\beta_0 Y_i - 2\beta_1 X_i Y_i + 2\beta_0 \beta_1 X_i)}{\partial \beta_0} \\ &= 0 + 2\beta_0 + 0 \quad -2Y_i \quad - \quad 0 \quad + 2\beta_1 X_i \\ &= 2(\beta_0 - Y_i + \beta_1 X_i) \\ &= -2(Y_i - \beta_0 - \beta_1 X_i) \end{aligned} \right\} \quad (5.10)$$

のように要素に分解し, シグマで再統合すれば式 (5.8) の 1 行目となる. β_1 での偏微分も同様である.

式 (5.8) を 0 と置き, β_0 と β_1 の推定値としての $\hat{\beta}_0$ と $\hat{\beta}_1$ を求める.

$$\left. \begin{aligned} -2\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) &= 0 \\ -2\sum_{i=1}^n [(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i] &= 0 \end{aligned} \right\} \quad (5.11)$$

正規方程式

式 (5.11) を $\hat{\beta}_0$ と $\hat{\beta}_1$ について解くために, 両辺を -2 で割り式を整理すると,

$$\left. \begin{aligned} \sum_{i=1}^n Y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i &= 0 \\ \sum_{i=1}^n X_i Y_i - \hat{\beta}_0 \sum_{i=1}^n X_i - \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= 0 \end{aligned} \right\} \quad (5.12)$$

を得る. 推定値 $\hat{\beta}_0$ と $\hat{\beta}_1$ が, 含まれない項を右辺に移して整理すると,

$$\left. \begin{aligned} (1) \quad n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ (2) \quad \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i Y_i \end{aligned} \right\} \quad (5.13)$$

が得られる. この方程式は, 正規方程式と呼ばれている. 式 (5.13) の左辺は, $\mathbf{X}^T \mathbf{X}$ と $\hat{\boldsymbol{\beta}}$ の積に等しく, 右辺は $\mathbf{X}^T \mathbf{Y}$ と等しいので, 行列で表わすと

$\mathbf{X}^T \mathbf{X}$		$\hat{\boldsymbol{\beta}}$	=	$\mathbf{X}^T \mathbf{Y}$
n	ΣX_i	$\hat{\beta}_0$		ΣY_i
ΣX_i	ΣX_i^2	$\hat{\beta}_1$		$\Sigma(X_i Y_i)$
2×2		2×1		2×1

となり, 行列表記での正規方程式

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y} \quad (5.14)$$

が得られる.

正規方程式の解によるパラメータの推定

正規方程式 (5.13) を $\hat{\beta}_0$ と $\hat{\beta}_1$ について解くために, (1) の両辺に $1/n$ を掛け, さらに ΣX_i を掛けると次式を得る. (2) は Σ の範囲の表示を外してある. 以後, Σ 記号の添え字 i は省略する.

$$\left. \begin{aligned} (1) \quad \hat{\beta}_0 \Sigma_i X_i + \hat{\beta}_1 \frac{(\Sigma_i X_i)^2}{n} &= \frac{(\Sigma_i X_i)(\Sigma_i Y_i)}{n} \\ (2) \quad \hat{\beta}_0 \Sigma X_i + \hat{\beta}_1 \Sigma X_i^2 &= \Sigma(X_i Y_i) \end{aligned} \right\} \quad (5.15)$$

式 (5.15) の (2) 式から (1) 式を引いて、 $\hat{\beta}_1$ について解くと

$$\left. \begin{aligned} \hat{\beta}_1 &= \frac{\Sigma(X_i Y_i) - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}} \\ &= \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \\ &= \frac{S_{XY}}{S_{XX}} \end{aligned} \right\} \quad (5.16)$$

のように、多くの統計の教科書で示されている結果が得られる。ここで、 S_{XX} は、 \mathbf{X} の平均 \bar{X} からの偏差の平方和であり、 S_{XY} は、 \mathbf{X} の平均 \bar{X} からの偏差と \mathbf{Y} の平均 \bar{Y} からの偏差の積和である。この式をさらに、 $\bar{X} = (\Sigma X_i) / n$ 、 $\bar{Y} = (\Sigma Y_i) / n$ などの関係を用い、式 (5.16) の 2 行目の分子 $\Sigma(X_i - \bar{X})(Y_i - \bar{Y})$ について

$$\left. \begin{aligned} \Sigma(X_i - \bar{X})(Y_i - \bar{Y}) &= \Sigma(X_i Y_i) - \bar{X}\Sigma Y_i - \bar{Y}\Sigma X_i + n\bar{X}\bar{Y} \\ &= \Sigma(X_i Y_i) - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} \\ &= \Sigma(X_i Y_i) - n\bar{X}\bar{Y} \\ &= \Sigma(X_i Y_i) - \frac{(\Sigma X_i)(\Sigma Y_i)}{n} \end{aligned} \right\} \quad (5.17)$$

と、展開し、整理すると式 (5.16) の 1 行目の分子となることを利用している。式 (5.16) の 2 行目の分母 $\Sigma(X_i - \bar{X})^2$ については、式 (5.17) の Y を X に置き換えることにより式 (5.16) の 1 行目の分母となる。

$$\left. \begin{aligned} \Sigma(X_i - \bar{X})^2 &= \Sigma X_i^2 - 2\bar{X}\Sigma X_i + n\bar{X}^2 \\ &= \Sigma X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \Sigma X_i^2 - n\left(\frac{\Sigma X_i}{n}\right)^2 \\ &= \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n} \end{aligned} \right\} \quad (5.18)$$

このように、平均値を差し引いた平方和の計算は、手計算あるいは電卓の時代では、端数が出るので計算がめんどろであり、もとの数字 X_i の 2 乗和の計算で済ませられるとの理由で、標準的な計算法として普及し、多くの統計の教科書に引き継がれている。ただし、Excel を用いた場合には、平均値を差し引いた平方和の計算の方が簡潔で扱いやすい。

正規方程式 (5.13) の 1 行目は、

$$n\hat{\beta}_0 + \hat{\beta}_1 \Sigma X_i = \Sigma Y_i \quad (5.19)$$

なので、 $\hat{\beta}_0$ について解くと、 $\hat{\beta}_1$ を含む次式が得られる。

$$\left. \begin{aligned} \hat{\beta}_0 &= \frac{\sum Y_i - \hat{\beta}_1 \sum X_i}{n} \\ &= \bar{Y} - \hat{\beta}_1 \bar{X} \end{aligned} \right\} \quad (5.20)$$

まとめると、回帰式の切片 $\hat{\beta}_0$ と傾き $\hat{\beta}_1$ の推定値は、

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (5.21)$$

と簡潔な式となる。パラメータの導出だけであれば、この簡便な式 (5.21) によってパラメータ $\hat{\beta}_1$ および $\hat{\beta}_0$ を計算することができ、表 5.1 に示すように Excel を用いて実に簡便にパラメータ $\hat{\beta}_1$ および $\hat{\beta}_0$ を推定することができる。

偏差平方和を用いたパラメータの推定の実際

表 5.1 に示す Excel による計算シートの結果を用いて、回帰直線のパラメータは、

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{78.00}{88.00} = 0.8864$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 7.70 - 0.8864 \times 22.00 = -11.8000$$

として求められる。

表 5.1 偏差平方和を用いた回帰パラメータの Excel シート上での推定

i	Y	X	Y 偏差	X 偏差	X 偏差 ²	XY 偏差		
1	4	18	-3.70	-4.00	16.00	14.80	$\hat{\beta}_1 =$	0.8864
2	7	21	-0.70	-1.00	1.00	0.70	$\hat{\beta}_0 =$	-11.8000
3	10	24	2.30	2.00	4.00	4.60		
4	10	24	2.30	2.00	4.00	4.60		
5	2	18	-5.70	-4.00	16.00	22.80		
6	11	27	3.30	5.00	25.00	16.50		
7	9	21	1.30	-1.00	1.00	-1.30		
8	5	21	-2.70	-1.00	1.00	2.70		
9	11	26	3.30	4.00	16.00	13.20		
10	8	20	0.30	-2.00	4.00	-0.60		
	7.70	22.00	0.00	0.00	88.00	78.00		
	平均	平均	合計	合計	平方和	平方和		
	\bar{Y}	\bar{X}	$Y_i - \bar{Y}$	$X_i - \bar{X}$	S_{XX}	S_{XY}		

Excel シート上で数値を入力すると表示形式が「標準」となり、右側の罫線に貼り付いてしまい、見栄えが悪い。「セルの書式設定」で表示形式を「数値」に変更し、小数点以下の桁数を調整する見栄えが向上する。

計算シートを用いなくとも、Excel の関数を用いて平方和の計算を

$$\bar{X} = \text{Average}(X\text{の範囲}) = 7.70$$

$$\bar{Y} = \text{Average}(Y\text{の範囲}) = 22.00$$

$$S_{XX} = \text{DevSq}(X\text{の範囲}) = 88.00$$

$$S_{XY} = \text{SumProduct}(X\text{の範囲}-\bar{X}, Y\text{の範囲}-\bar{Y}) = 78.00$$

のように直接推定することもできる。さらに、Excel には、 $\hat{\beta}_1$ および $\hat{\beta}_0$ を推定する便利な関数

$$\hat{\beta}_1 = \text{Slope}(Y\text{の範囲}, X\text{の範囲}) = 0.8864$$

$$\hat{\beta}_0 = \text{Intercept}(Y\text{の範囲}, X\text{の範囲}) = -11.8000$$

もある。

回帰分析は、このように定式化され、ほとんどの統計の教科書で取り上げられているが、いわゆる有意差検定よりもかなり高級であり、きちっと理解し、さらなる応用のために学習したいと思っても難解であり、教科書に示されている計算公式の範囲内に多くの人達が留まざるをえなくなってしまう。単回帰分析に引き続き、2次式あるいは3次式をあてはめたいと思っても、パラメータを推定する式を見つけることができるのだろうか。見つけることができなければ、自ら導出することができるのだろうか。伝統的な回帰分析の解法は、更に回帰分析を拡張し活用したいと思う人達の学習意欲をへし折るような、まさにガラスの天井のごとくである。第2.4節で示した多項式回帰に対しても、デザイン行列 X を用いた回帰分析を適用すれば、同一手順ですべて解決することができる。

ガラスの天井を超えるためには、回帰分析を Excel の行列計算で実行できるようになることが最初の一步である。とは言え、いきなりパラメータが3以上の場合に取り組むと、敷居が高すぎて挫折しかねない。段階的な学習としては、式(5.14)の正規方程式を次に示す行列表記に対応付け

$$(X^T X)\hat{\beta} = X^T Y \quad \longleftrightarrow \quad \begin{array}{|c|c|} \hline X^T X & \beta^{\wedge} \\ \hline \end{array} = \begin{array}{|c|} \hline X^T Y \\ \hline \end{array}$$

n	ΣX_i	β_0^{\wedge}	=	ΣY_i
ΣX_i	ΣX_i^2	β_1^{\wedge}		$\Sigma X_i Y_i$
2×2		2×1		2×1

学習することが望ましい。ただし、推奨できる日本語の成書は、残念ながら絶版となっているドレーパら(1968)、「応用回帰分析」しか見当たらない。なお、Net書店では中古本が手に入る場合もあるが、Draperら(1989)もネット書店で手軽に入手できる。

そこで、「偏差平方和を用いた回帰分析」と「デザイン行列を用いた回帰分析」について、相互の関連を丁寧に示すことにする。そして、読者が「偏差平方和による回帰分析」から、「デザイン行列 X を用いた回帰分析」を主体にした解析法に親しみを感じてもらいたいと願っている。

5.4. デザイン行列 vs. 偏差平方和を用いた回帰分析

行列計算によるパラメータ推定

前節では、偏差平方和を用いた回帰分析のパラメータの推定方法を示し、式 (5.13) のシグマ表記の正規方程式は、式 (5.14) に示したように行列を用いて表すと、

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}$$

となる。推定値 $\hat{\boldsymbol{\beta}}$ を得るために、行列計算では両辺を $(\mathbf{X}^T \mathbf{X})$ で割ることができない。そこで、逆行列の定義により $(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{X}) = \mathbf{I}$ が単位行列となることを活用し、逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ を式 (5.14) 両辺に掛けると

$$\left. \begin{aligned} (\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\ \mathbf{I} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \end{aligned} \right\} \quad (5.22)$$

のように $\hat{\boldsymbol{\beta}}$ を推定することができる。2x2 の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ ならば、 $(\mathbf{X}^T \mathbf{X})$ の行列式 D を計算し、次のようにして計算することができる。実際に手軽に手計算できるのは、2x2 の場合までで、(3x3) 以上の場合には、行列式の Mdetarm() 関数を用いても煩雑であり勧められない。

$\mathbf{X}^T \mathbf{X}$		-1	$(\mathbf{X}^T \mathbf{X})^{-1}$		
n	ΣX_i		$\Sigma X_i^2 / D$	$-\Sigma X_i / D$	
ΣX_i	ΣX_i^2		$-\Sigma X_i / D$	n / D	
			$D = (n \Sigma X_i^2) - (\Sigma X_i)^2$		
$\mathbf{X}^T \mathbf{X}$		-1	$(\mathbf{X}^T \mathbf{X})^{-1}$		$(\mathbf{X}^T \mathbf{X})^{-1}$
10	220		$4928 / D$	$-220 / D$	5.6000
220	4928		$-220 / D$	$10 / D$	-0.2500
			$D = 49280 - 48400 = 880$		
			$= \text{Mdeterm}(\mathbf{X}^T \mathbf{X} \text{の範囲})$		

実際に逆行列を前掛けすると

$(\mathbf{X}^T \mathbf{X})^{-1}$		$\mathbf{X}^T \mathbf{X}$		=	$(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{X})$	
5.6000	-0.2500	10	220		1.0000	0.0000
-0.2500	0.0114	220	4928		0.0000	1.0000

のように単位行列 \mathbf{I} になることが確認され、単位行列と $\hat{\boldsymbol{\beta}}$ の積は、

$(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{X}^T \mathbf{X})$		$\hat{\boldsymbol{\beta}}$	=	$\hat{\boldsymbol{\beta}}$
1.0000	0.0000	β_0^{\wedge}		β_0^{\wedge}
0.0000	1.0000	β_1^{\wedge}		β_1^{\wedge}

と元の $\hat{\boldsymbol{\beta}}$ と同じになる。

逆行列は、Excel の Minverse () 関数によって簡単に求めることができる。推定値 $\hat{\beta}$ を求めるために必要な $X^T Y$ は、第5.2節の「デザイン行列 X と反応ベクトル Y との積」の項の結果を用い、これらの計算結果を組み合わせ、次のような手順で推定値 $\hat{\beta}$ を簡単に求めることができる。

$$\hat{\beta} = (X^T X)^{-1} X^T Y =$$

$(X^T X)^{-1}$		$X^T Y$	$\hat{\beta}$
5.6000	-0.2500	77	-11.8000
-0.2500	0.0114	1772	0.8864
=Minverse ($X^T X$ の範囲)			
		=Mmult (Transpose (X の範囲), Y の範囲)	
		=Mmult ($(X^T X)^{-1}$ の範囲, $X^T Y$ の範囲)	

デザイン行列 X での推定式と偏差平方和での推定式の相違

行列計算によって推定値 $\hat{\beta}$ が得られるのであるが、正規方程式 (5.22) から、導出された $\hat{\beta}_0$ と $\hat{\beta}_1$ の計算式は、式 (5.16) から式 (5.21) に示した式とは異なるので、 $(\sum X_i) = n\bar{X}$, $(\sum Y_i) = n\bar{Y}$ などの関係を用いて行列計算の式を整理すると、一致することが確認できる。

$$\left[\begin{array}{c} \hat{\beta}_0 \\ \hat{\beta}_1 \end{array} \right] = \left[\begin{array}{cc} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{array} \right]^{-1} \left[\begin{array}{c} \sum Y_i \\ \sum X_i Y_i \end{array} \right]$$

$$= \left[\begin{array}{cc} \frac{\sum X_i^2}{D} & \frac{-\sum X_i}{D} \\ \frac{-\sum X_i}{D} & \frac{n}{D} \end{array} \right] \left[\begin{array}{c} \sum Y_i \\ \sum X_i Y_i \end{array} \right]$$

$$= \left[\begin{array}{c} \frac{(\sum X_i^2)(\sum Y_i) - (\sum X_i)(\sum X_i Y_i)}{n\sum X_i^2 - (\sum X_i)^2} \\ \frac{-\sum X_i \sum Y_i + n\sum X_i Y_i}{n\sum X_i^2 - (\sum X_i)^2} \end{array} \right] \quad (5.23)$$

$$= \left[\begin{array}{c} \frac{\bar{Y}(\sum X_i^2) - \bar{X}(\sum X_i Y_i)}{\sum X_i^2 - n\bar{X}^2} \\ \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} \end{array} \right]$$

ここで、式 (5.23) の最後の行列の2行目は、式 (5.21) で導出された $\hat{\beta}_1$ の推定値に一致する。

$$\hat{\beta}_1 = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{S_{XY}}{S_{XX}} \quad (5.24)$$

さて、式 (5.21) で導出された $\hat{\beta}_0$ の推定式とは、式 (5.23) の最後の行列の1行目は明らかに異なる。そこで、式 (5.21) からスタートし、式の変形を行なうと

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
&= \frac{\bar{Y}S_{XX} - \bar{X}S_{XY}}{S_{XX}} \\
&= \frac{\left(\bar{Y}(\Sigma X_i^2) - \bar{Y} \frac{(\Sigma X_i)^2}{n} \right) - \left(\bar{X}(\Sigma X_i Y_i) - \bar{X} \frac{(\Sigma X_i)(\Sigma Y_i)}{n} \right)}{S_{XX}} \\
&= \frac{(\bar{Y}(\Sigma X_i^2) - n\bar{Y}\bar{X}^2) - (\bar{X}(\Sigma X_i Y_i) - n\bar{Y}\bar{X}^2)}{S_{XX}} \\
&= \frac{\bar{Y}(\Sigma X_i^2) - \bar{X}(\Sigma X_i Y_i)}{\Sigma X_i^2 - n\bar{X}^2}
\end{aligned} \tag{5.25}$$

となり、式 (5.23) の最後の行列の 1 行目が導出される。私にとっても見るのも嫌になる数式の変形であり、シグマによる計算と行列による回帰係数の計算結果が一致することを数式で示すことは、難儀である。実用上は、事例により数値計算の結果が一致することは、これまでの結果で明らかである。

したがって、偏差平方和ベースの回帰分析およびデザイン行列ベースの計算方法を両建てで説明することは冗長であり、気持ちとしては避けなかったのであるが、偏差平方和を用いた回帰分析を「ガラスの天井」と言い切るため、あえて両者の関係について丁寧に示した。

偏差平方和を用いたパラメータの分散の推定

パラメータ $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$ は、式 (5.16) の正規方程式の解を用いて、

$$\begin{aligned}
Var(\hat{\beta}_1) &= Var \left[\frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \right] \\
&= Var \left[\frac{\Sigma(X_i - \bar{X})Y_i}{\Sigma(X_i - \bar{X})^2} \right] \\
&= \frac{Var(Y_i)}{\Sigma(X_i - \bar{X})^2} \\
&= \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2}
\end{aligned} \tag{5.26}$$

となる。分子の変形は、

$$\begin{aligned}
\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) &= \Sigma(X_i Y_i - X_i \bar{Y} - \bar{X} Y_i + \bar{X} \bar{Y}) \\
&= \Sigma(X_i Y_i - \bar{X} Y_i) - \Sigma X_i \bar{Y} + \Sigma \bar{X} \bar{Y} \\
&= \Sigma(X_i Y_i - \bar{X} Y_i) - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} \\
&= \Sigma(X_i - \bar{X})Y_i
\end{aligned} \tag{5.27}$$

を用いて, $\Sigma(X_i - \bar{X})$ を Y_i に関してコンスタント化するためである. $\hat{\beta}_0$ の分散 $Var(\hat{\beta}_0)$ も, 正規方程式の解を用いて, \bar{Y} と $\hat{\beta}_1$ が無相関であり, \bar{X} はコンスタントなので,

$$\left. \begin{aligned} Var(\hat{\beta}_0) &= Var(\bar{Y} - \hat{\beta}_1 \bar{X}) \\ &= Var(\bar{Y}) + \bar{X}^2 Var(\hat{\beta}_1) \\ &= \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{\Sigma(X_i - \bar{X})^2} \\ &= \frac{\sigma^2 \Sigma X_i^2}{n \Sigma(X_i - \bar{X})^2} \end{aligned} \right\} \quad (5.28)$$

となる. 共分散 $Cov(\hat{\beta}_0, \hat{\beta}_1)$ は,

$$\left. \begin{aligned} Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\bar{Y} - \hat{\beta}_1 \bar{X}, \hat{\beta}_1) \\ &= -\bar{X} Var(\hat{\beta}_1) \\ &= \frac{-\bar{X} \sigma^2}{\Sigma(X_i - \bar{X})^2} \end{aligned} \right\} \quad (5.29)$$

となる.

パラメータ $\hat{\beta}$ の共分散行列 $\Sigma(\hat{\beta})$ は, これらの計算式から

$$\left. \begin{aligned} \Sigma(\hat{\beta}) &= \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\Sigma X_i^2}{n \Sigma(X_i - \bar{X})^2} & \frac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} & \frac{1}{\Sigma(X_i - \bar{X})^2} \end{bmatrix} \sigma^2 \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2 \end{aligned} \right\} \quad (5.30)$$

となる. したがって, パラメータの分散共分散行列は, デザイン行列 \mathbf{X}^T と \mathbf{X} の積の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ に誤差分散 σ^2 を掛けた結果に一致する. ただし, 誤差分散 σ^2 は, 未知なので, 残差の推定値 $\hat{\varepsilon}_i$ の平方和を自由度 $(n-2)$ で割った誤差分散の推定値 $\hat{\sigma}^2$ を用いて計算する.

$$\left. \begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} \\ &= \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{n-2} \end{aligned} \right\} \quad (5.31)$$

これまで, $(\mathbf{X}^T \mathbf{X})^{-1}$ については, 次の式を示してきたのであるが, 式 (5.30) と異なるので, 式の変形を行なう.

$\mathbf{X}^T \mathbf{X}$		^{-1}	$(\mathbf{X}^T \mathbf{X})^{-1}$	
n	ΣX_i		$\Sigma X_i^2 / D$	$-\Sigma X_i / D$
ΣX_i	ΣX_i^2	$-\Sigma X_i / D$	n / D	
		$D = (n \Sigma X_i^2) - (\Sigma X_i)^2$		

$(\mathbf{X}^T \mathbf{X})$ の行列式 D は,

$$\left. \begin{aligned} D &= n \Sigma X_i^2 - (\Sigma X_i)^2 \\ &= n(\Sigma X_i^2 - n \bar{X}^2) \\ &= n \Sigma (X_i - \bar{X})^2 \end{aligned} \right\} \quad (5.32)$$

と変形できるので,

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} \frac{\Sigma X_i^2}{n \Sigma (X_i - \bar{X})^2} & \frac{-\Sigma X_i}{n \Sigma (X_i - \bar{X})^2} \\ \frac{-\Sigma X_i}{n \Sigma (X_i - \bar{X})^2} & \frac{n}{n \Sigma (X_i - \bar{X})^2} \end{bmatrix} = \begin{bmatrix} \frac{\Sigma X_i^2}{n \Sigma (X_i - \bar{X})^2} & \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} & \frac{1}{\Sigma (X_i - \bar{X})^2} \end{bmatrix} \quad (5.33)$$

と一致することが確かめられる. このように, 従来のシグマを用いたパラメータの分散の計算式とデザイン行列 \mathbf{X} を用いた計算結果が同じであることが示された. したがって, 従来の計算式ではなく, デザイン行列 \mathbf{X} を用いた計算手順で簡単に得れるパラメータの共分散行列 $\Sigma(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$ を基本とすることは, 1変数の回帰分析にしか適用できない偏差平方和を用いた計算方法によって作り出されているガラスの天井を超えるための必須の知識である.

ここに示したパラメータの共分散行列 $\Sigma(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$ は, 1変数のデザイン行列 \mathbf{X} のみならず, 2次式の回帰 $y = \beta_0 + \beta_1 x + \beta_2 x^2$, 2変数の重回帰 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, ダミー変数 a を用いた共分散分析 $y = \beta_0 + \beta_1 a + \beta_2 x$, 交互作用を含む2因子実験など, 一般線形モデルのパラメータの共分散行列においても, デザイン行列 \mathbf{X} のサイズが異なるだけで, 計算手順はすべて同じである.

シグマを用いた回帰分析の功罪

得られたデータに対し1次式(回帰直線)をあてはめてみたが, 直線とはいいがたいので, 2次式(回帰曲線)をあてはめたい. パラメータの推定は, 回帰分析の説明変数 x に加えて x^2 を加えることにより得られ, Excelの散布図上に2次曲線の推定値を重ね書きすることができた. さらに, 2次曲線の95%信頼区間を重ね書きしたと思い, 計算式がどこにあるか手を尽くして探したが見つからない. 1次式(回帰直線)の計算式を参考にし, 拡張しようと思ってもまったく分からない. 統計ソフトの多項式回帰回帰を使えば, 95%信頼区間を図示できることは分かったが, 計算式が分からない. このように, シグマを用いた回帰分析の方法は, 拡張性が全くない. デザイン行列 \mathbf{X} を活用した場合については, 第2.4節, 「1因子実験の量的変数に対する多項式回帰」に詳細が示されている.

5.5. Excel の行列計算による回帰分析の実際

Excel の行列計算による回帰パラメータの推定

表 5.2 に示すのは、これまで示してきた計算方法を総合し、デザイン行列 X を用いた回帰分析の計算事例である。全ての計算過程を 1 枚の Excel シートで示したために、やや見づらくなっているのですが、丁寧な解説を付け加える。なお、Excel の分析ツールの「回帰分析」を活用した表 5.6 と比較することにより理解を深めてもらいたい。

表 5.2 Excel の行列計算を用いた回帰分析

i	X		Y	Y^{\wedge}	$\varepsilon^{\wedge}=Y-Y^{\wedge}$						
1	1	18	4	4.1545	-0.1545						
2	1	21	7	6.8136	0.1864						
3	1	24	10	9.4727	0.5273						
4	1	24	10	9.4727	0.5273						
5	1	18	2	4.1545	-2.1545						
6	1	27	11	12.1318	-1.1318						
7	1	21	9	6.8136	2.1864						
8	1	21	5	6.8136	-1.8136	項	推定値	分散	SE	t 値	p 値
9	1	26	11	11.2455	-0.2455	$\beta_0^{\wedge} =$	-11.8000	13.2745	3.6434	-3.2387	0.0119
10	1	20	8	5.9273	2.0727	$\beta_1^{\wedge} =$	0.8864	0.0269	0.1641	5.4005	0.0006
						$(X^T X)^{-1} X^T Y$		sqrt(分散)	T.dist.2T (t, 10-2)		
						$\varepsilon^{\wedge T} \varepsilon^{\wedge} =$	18.9636	共分散	13.2745	-0.5926	
						$\sigma^{\wedge 2} =$	2.3705	行列	-0.5926	0.0269	
						$(X^T X)$	$(X^T X)^{-1}$	$(X^T Y)$	$\Sigma(\beta^{\wedge}) = (X^T X)^{-1} \sigma^{\wedge 2}$		
						10	220				
						220	4928				
						5.6000	-0.2500				
						-0.2500	0.0114				
						77					
						1772					

デザイン行列 X に対し、 $X^T X$ の結果が表の下段に (2×2) の矩形内に Excel の Mmult () 関数および Transpose () 関数を用いて

$$X^T X = \text{Mmult}(\text{Transpose}(X\text{の範囲}), X\text{の範囲}) = \begin{bmatrix} 10 & 220 \\ 220 & 4928 \end{bmatrix}$$

と計算されている。その横に $(X^T X)$ の逆行列 $(X^T X)^{-1}$ が、Minverse () 関数を用いて

$$(X^T X)^{-1} = \text{Minverse}(X^T X\text{の範囲}) = \begin{bmatrix} 5.6000 & -0.2500 \\ -0.2500 & 0.0114 \end{bmatrix}$$

となり、デザイン行列 X の転置行列 X^T と列ベクトル Y との積 $X^T Y$ が

$$X^T Y = \text{Mmult}(\text{Transpose}(X\text{の範囲}), Y\text{の範囲}) = \begin{bmatrix} 77 \\ 1772 \end{bmatrix}$$

として計算されている。回帰パラメータの推定値 $\hat{\beta}$ は、表の中段の「推定値」の欄に

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \text{Mmult}((X^T X)^{-1}\text{の範囲}, X^T Y\text{の範囲}) = \begin{bmatrix} -11.8000 \\ 0.8864 \end{bmatrix}$$

$\hat{\beta}_0 = -11.8000$, $\hat{\beta}_1 = 0.8864$ として計算されている. 誤差分散は, 回帰の推定値 \hat{Y} を

$$\hat{Y} = X\hat{\beta} = \text{Mmult}(X\text{の範囲}, \hat{\beta}\text{の範囲})$$

で求め, 残差ベクトル $\hat{\epsilon}$ が列ベクトル Y と推定ベクトル \hat{Y} の差

$$\hat{\epsilon} = Y - \hat{Y} = (Y\text{の範囲} - \hat{Y}\text{の範囲})$$

として求めている. 残差平方和

$$S_e = \hat{\epsilon}^T \hat{\epsilon} = \text{Mmult}(\text{Transpose}(\hat{\epsilon}\text{の範囲}), \hat{\epsilon}\text{の範囲})$$

$$\text{または, } S_e = \text{SumSq}(\hat{\epsilon}\text{の範囲})$$

を計算し, データ数 $n = 10$ からパラメータの数 2 を引いた自由度で割った平均平方が誤差分散の推定値 $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n-2} = \frac{18.9636}{10-2} = 2.3705$$

として計算されている.

回帰パラメータの分散および共分散の推定

デザイン行列 X を用いた回帰分析の最大の利点は, 回帰パラメータ $\hat{\beta}_0$ と $\hat{\beta}_1$ の分散および共分散が (2×2) の行列として簡単に得られることである. 式 (5.30) からパラメータの共分散行列 $\Sigma(\hat{\beta})$ は,

$$\begin{aligned} \Sigma(\hat{\beta}) &= \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \end{bmatrix} \\ &= (X^T X)^{-1} \hat{\sigma}^2 \end{aligned} \tag{5.34}$$

であることを示した. デザイン行列の積和の逆行列 $(X^T X)^{-1}$ および誤差分散の推定値 $\hat{\sigma}^2$ は,

5.6000	-0.2500	$\hat{\epsilon}^T \hat{\epsilon} =$	18.9636
-0.2500	0.0114	$\hat{\sigma}^2 =$	2.3705
$(X^T X)^{-1}$			

として計算されているので, パラメータの共分散行列 $\Sigma(\hat{\beta})$ は,

$$\Sigma(\hat{\beta}) = \begin{bmatrix} 5.6000 & -0.2500 \\ -0.2500 & 0.0114 \end{bmatrix} \begin{bmatrix} 2.3705 \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} 13.2745 & -0.5926 \\ -0.5926 & 0.0269 \end{bmatrix} \Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$$

と計算される. 分散 $\text{Var}(\hat{\beta}_0) = 13.2745$, $\text{Var}(\hat{\beta}_1) = 0.0269$ は, $(X^T X)^{-1} \hat{\sigma}^2$ の対角要素となっているので, 次のように推定値の分散が得られる.

項	推定値	分散
$\hat{\beta}_0 =$	-11.8000	13.2745
$\hat{\beta}_1 =$	0.8864	0.0269

更に標準誤差 SE を分散の平方根 `sqrt()` 関数で求め, 推定値/ SE で t 値を計算し, t 分布の両側確率の p 値を `T.dist.2T()` 関数

$$p \text{ 値} = T.\text{dist}.2T(t\text{値}, (10-2))$$

で計算し、回帰パラメータについての推定および t 検定が次のように行える。

項	推定値	分散	SE	t 値	p 値
$\beta_0^{\wedge} =$	-11.8000	13.2745	3.6434	-3.2387	0.0119
$\beta_1^{\wedge} =$	0.8864	0.0269	0.1641	5.4005	0.0006
$(X^T X)^{-1} X^T Y$		sqrt(分散)		T.dist.2T(t , 10-2)	

表 5.1 で示した単に回帰パラメータを計算した Excel シートより計算は複雑ではあるが、同程度の Excel シート 1 枚の中に、基本的な関数のみで、回帰分析の基本的な結果が網羅されている。なお、回帰分析に付きものの、分散分析表は、行列計算で求めるよりも推定値ベースで求めることが簡潔であり、本質的である。

Excel の関数を用いた分散分析表の作成

一般的に最小 2 乗法による回帰分析には、分散分析表が付きものである。表 5.2 では、反応 Y_i に対して、推定された回帰パラメータを用いた推定値 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ との残差 $\hat{\epsilon}_i = Y_i - \hat{Y}_i$ を計算し、その平方和を求めた。この残差平方和 S_e は、推定された回帰直線からのズレの大きさを表している。

$$S_e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \tag{5.35}$$

表 5.3 に示すように、平均 \bar{Y} からの偏差を $(Y_i - \bar{Y})$ としたときの偏差平方和を「全体の偏差平方和」の意味で S_T とする。

$$S_T = \sum_{i=1}^n (Y_i - \bar{Y})^2 \tag{5.36}$$

回帰直線の推定値は、 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ であり、平均 \bar{Y} からの偏差 $R_i = \hat{Y}_i - \bar{Y}$ の平方和が、回帰の平方和 S_R となる。

$$S_R = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \tag{5.37}$$

表 5.3 分散分析表作成のための平方和の計算

i	X		Y_i	\bar{Y}	$Y_i - \bar{Y}$	Y_i^{\wedge}	$Y_i^{\wedge} - \bar{Y}$	$Y_i - Y_i^{\wedge}$		
1	1	18	4	7.70	-3.70	4.1545	-3.5455	-0.1545		
2	1	21	7	7.70	-0.70	6.8136	-0.8864	0.1864	$\beta_0^{\wedge} =$	-11.8000
3	1	24	10	7.70	2.30	9.4727	1.7727	0.5273	$\beta_1^{\wedge} =$	0.8864
4	1	24	10	7.70	2.30	9.4727	1.7727	0.5273		
5	1	18	2	7.70	-5.70	4.1545	-3.5455	-2.1545		
6	1	27	11	7.70	3.30	12.1318	4.4318	-1.1318		
7	1	21	9	7.70	1.30	6.8136	-0.8864	2.1864		
8	1	21	5	7.70	-2.70	6.8136	-0.8864	-1.8136		
9	1	26	11	7.70	3.30	11.2455	3.5455	-0.2455		
10	1	20	8	7.70	0.30	5.9273	-1.7727	2.0727		
				7.70	88.10		69.1364	18.9636		
				\bar{Y}	S_T		S_R	S_e		
自由度			10	1	10-1=9	2	2-1=1	10-2=8		

このように、元の反応 Y_i についての平方和には、

$$\left. \begin{aligned} S_T &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \\ &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \\ &= S_e + S_R \end{aligned} \right\} \quad (5.38)$$

との関係があり、平方和の分解とも言われている（本節の末尾に式の証明を示す）。実際に、これらの平方和を計算した結果を表 5.4 に示す。

回帰パラメータの推定には、デザイン行列 \mathbf{X} を使うことを勧めるが、分散分析表の作成には、各種の平方和の計算に基づいた方法が、理解しやすい。平方和の計算は、`SumSq()` 関数の使用が効率的である。

$$\begin{aligned} S_T &= \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = \text{SumSq}(\mathbf{Y} \text{の範囲} - \bar{Y}) = 88.10, & df &= 10 - 1 = 9 \\ \hat{\mathbf{Y}} &= \text{Mmult}(\mathbf{X} \text{の範囲}, \hat{\boldsymbol{\beta}} \text{の範囲}), & df &= 2 \\ S_R &= \sum_{i=1}^{10} (\hat{Y}_i - \bar{Y})^2 = \text{SumSq}(\hat{\mathbf{Y}} \text{の範囲} - \bar{Y}) = 69.1364, & df &= 2 - 1 = 1 \\ S_e &= \sum_{i=1}^{10} (Y_i - \hat{Y}_i)^2 = \text{SumSq}(\mathbf{Y} \text{の範囲} - \hat{\mathbf{Y}} \text{の範囲}) = 18.9636, & df &= 10 - 2 = 8 \end{aligned}$$

計算結果を、表 5.4 の分散分析表にまとめる。計算原理を Excel の行列計算で習得した後は、表 5.6 に示すように Excel の分析ツールの回帰分析によっても同じ結果が得られるので、自ら計算することにこだわる必要はない。

表 5.4 回帰に対する分散分析表

要因		平方和	自由度	平均平方	F 値	p 値
回帰	S_R	69.1364	2-1=1	69.1364	29.1659	0.0006
残差	S_e	18.9636	10-2=8	2.3705		
全体	S_T	88.1000	10-1=9			

反応 \mathbf{Y} の自由度は、データ数 n であるが、全体の平方和 S_T の自由度は、計算のために反応 \mathbf{Y} から求めた算術平均 \bar{Y} を用いているので、自由度が 1 つ分減少して $df_T = n - 1$ となる。誤差平方和 S_e は、回帰の推定値 \hat{Y}_i の計算のために $\hat{\beta}_0$ と $\hat{\beta}_1$ を用いているので自由度が 2 つ減り $df_e = n - 2$ となる。回帰の平方和 S_R は、自由度が 2 の \hat{Y}_i に対し、自由度 1 の \bar{Y} の差の平方和なので 1 つ減り $df_R = 1$ となる。平方和と同様に自由度の推定にも

$$\left. \begin{aligned} df_T &= df_R + df_e \\ &= 1 + (n - 2) = n - 1 \end{aligned} \right\} \quad (5.39)$$

が成り立つ。分散分析表における自由度については、各種の便宜的な説明が行なわれているが、偏差平方和の定義式に立ち返ることにより、自由度の本質的な理解となる。

F 値は、回帰の平均平方 69.1364 を残差の平均平方 2.3705 で割って 29.1659 が計算されている。この F 値は、分子の自由度 1、分母の自由度 8 の F 分布の右側 (Right Tail) 確率で、 $F.dist.RT()$ 関数を使い、 $F.dist.RT(29.1659, 1, 8)$ により計算されている。これらの Excel の確率分布の計算になれることも大切である。Excel の 2007 年以前にも多くの分布関数が提供されていたのであるが、分布関数の引数の使い方などが不統一であり、計算ミスを起こしやすく、Excel に対する不信感の原因でもあった。それ以後に改善がはかられ、Excel の分布関数を使うときに「2007 年以前のバージョンと互換性があります。」とのメッセージが出た場合には、使用を避けることが望ましい。

回帰パラメータの共分散行列の活用

Excel のデータ分析ツールの「回帰分析」により、パラメータの推定などが手軽にできることから Excel の行列関数を使った回帰分析をする意義はないと思われるかもしれない。多くの Excel ユーザの悩みは、回帰直線の 95%信頼区間および 95%予測区間 (個別データの 95%信頼区間) を散布図上に描きたいと思っても手軽に解決する機能が見いだせないことにある。

デザイン行列を用いた行列計算による回帰分析の計算過程では、表 5.2 に示したようにパラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算結果が含まれており、この行列 $\Sigma(\hat{\beta})$ を使って、パラメータの標準誤差 SE を計算した。パラメータの共分散行列 $\Sigma(\hat{\beta})$ があれば、95%信頼区間を散布図上に描くのは容易である。Excel のデータ分析ツールの「回帰分析」で出力される分散分析表の誤差分散の推定値 $\hat{\sigma}^2$ が得られるので、切片を含めた説明変数をデザイン行列 X とし、行列計算で $(X^T X)^{-1}$ を求め

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$$

により、パラメータの共分散行列 $\Sigma(\hat{\beta})$ が容易に得られる。

回帰直線の 95%信頼区間

回帰直線の 95%信頼区間を求めるためには、回帰の推定値 \hat{Y}_i の分散が必要となる。合成分散の一般式により、

$$\left. \begin{aligned} \text{Var}(\hat{Y}_i) &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 X_i) \\ &= \text{Var}(\hat{\beta}_0) + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) X_i + \text{Var}(\hat{\beta}_1) X_i^2 \end{aligned} \right\} \quad (5.40)$$

により、計算することができる。任意の行ベクトルを

$$\mathbf{x} = [1 \quad x]$$

とすれば、次の 2 次形式により、

$$\begin{aligned}
 \text{Var}(\hat{y}) &= \mathbf{x}\Sigma(\hat{\boldsymbol{\beta}})\mathbf{x}^T \\
 &= \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Var}(\hat{\beta}_1) \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} \\
 &= \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} \text{Var}(\hat{\beta}_0) + \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)x \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \text{Var}(\hat{\beta}_1)x \end{bmatrix} \\
 &= \text{Var}(\hat{\beta}_0) + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)x + \text{Var}(\hat{\beta}_1)x^2
 \end{aligned} \tag{5.41}$$

合成分散の一般式に一致する。

Excel できれいな 95%信頼区間の滑らかな曲線を描くためには、表 5.5 に示すように、描きたい X 軸の範囲内で、適当な間隔の x を設定し、推定値を $\hat{y} = \mathbf{x}\hat{\boldsymbol{\beta}}$ 、分散 $\text{Var}(\hat{y})$ を式 (5.40)，または、式 (5.41) で計算する。

表 5.5 回帰直線の 95%信頼区間

切片	x	\hat{y}	$\text{Var}(\hat{y})$	L95%	U95%	個別 L95%	個別 U95%
1	14	0.61	1.96	-2.62	3.84	-4.19	5.41
1	16	2.38	1.21	-0.15	4.92	-1.98	6.74
1	18	4.15	0.67	2.27	6.04	0.13	8.17
1	20	5.93	0.34	4.57	7.28	2.13	9.73
1	22	7.70	0.24	6.58	8.82	3.98	11.42
1	24	9.47	0.34	8.12	10.83	5.67	13.27
1	26	11.25	0.67	9.36	13.13	7.23	15.27
1	28	13.02	1.21	10.48	15.55	8.66	17.38
1	30	14.79	1.96	11.56	18.02	9.99	19.59
1	32	16.56	2.93	12.62	20.51	11.25	21.87
T.inv.2T(0.05, 10-2)=				2.3060			

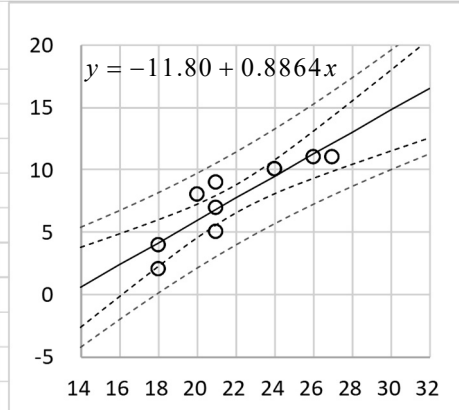


図 5.1 回帰直線の 95%信頼区間

表 5.5 の 1 行目の $\mathbf{x} = [1 \ 14]$ に対する推定値 \hat{y}_1 は、

$$\begin{aligned}
 \hat{y}_1 &= \mathbf{x}\hat{\boldsymbol{\beta}} \\
 &= [1 \ 14] \begin{bmatrix} -11.8000 \\ 0.8864 \end{bmatrix} = 0.6091
 \end{aligned}$$

となり、2 次形式により回帰直線の推定値 \hat{y}_1 の分散 $\text{Var}(\hat{y})$ は、

$$\begin{aligned}
 \text{Var}(\hat{y}_1) &= \mathbf{x}\Sigma(\hat{\boldsymbol{\beta}})\mathbf{x}^T \\
 &= [1 \ 14] \begin{bmatrix} 13.2745 & -0.5926 \\ -0.5926 & 0.0269 \end{bmatrix} \begin{bmatrix} 1 \\ 14 \end{bmatrix} = 1.9610
 \end{aligned}$$

として求められ、

95%信頼区間は、一般的な方法で

$$\begin{aligned} L95\% &= \hat{y}_1 - t_{0.05}(10-2)\sqrt{\text{Var}(\hat{y}_1)} \\ &= 0.6091 - 2.3060 \times \sqrt{1.9610} = -2.6201 \\ U95\% &= 0.6091 + 2.3060 \times \sqrt{1.9610} = 3.8383 \end{aligned}$$

として計算されている。

個別データの95%信頼区間は、回帰直線の分散 $\text{Var}(\hat{Y}_i)$ に1個分のデータの分散 $\hat{\sigma}^2$ を加えた分散を使う。

$$\begin{aligned} \text{個別}L95\% &= \hat{y}_1 - t_{0.05}(10-2)\sqrt{\text{Var}(\hat{y}_1) + \hat{\sigma}^2} \\ &= 0.6091 - 2.3060 \times \sqrt{1.9610 + 2.3705} = -4.1902 \\ \text{個別}U95\% &= 0.6091 + 2.3060 \times \sqrt{1.9610 + 2.3705} = 5.4084 \end{aligned}$$

第1行目で作成した計算式は、行方向にフィルハンドルでコピーすることによりすべて計算される。この結果をExcel散布図にまとめた結果が示されている。

伝統的な方法

多くの成書で回帰直線の95%信頼区間についての記述は、ほとんどが、以下の形式で示されている（過度な標準化となっている）。ある x_0 についての回帰式を

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{5.42}$$

とする。ところで、式(5.20)から $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ なので、式(5.42)に $\hat{\beta}_0$ を代入して

$$\left. \begin{aligned} \hat{y}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0 \\ &= \bar{y} + \hat{\beta}_1 (x_0 - \bar{x}) \end{aligned} \right\} \tag{5.43}$$

が得られ、 \hat{y}_0 の分散 $\text{Var}(\hat{y}_0)$ は、 \bar{y} と $\hat{\beta}_1$ の共分散が0なので、式(5.26)の $\text{Var}(\hat{\beta}_1)$ を用いて、次式で与えられる。 $\text{Var}(\bar{y})$ は、 n 個データの平均なので、 $\text{Var}(\bar{y}) = \hat{\sigma}^2 / n$ となり、

$$\left. \begin{aligned} \text{Var}(\hat{y}_0) &= \text{Var}[\bar{y} + \hat{\beta}_1 (x_0 - \bar{x})] \\ &= \text{Var}(\bar{y}) + (x_0 - \bar{x})^2 \text{Var}(\hat{\beta}_1) \\ &= \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right] \hat{\sigma}^2 \end{aligned} \right\} \tag{5.44}$$

分散が求められれば、標準的な方法で95%信頼区間を求めることができる。また、個別データに対する分散 $\text{Ver}(\hat{y}_{0, \text{個別}})$ は、

$$\left. \begin{aligned} \text{Ver}(\hat{y}_{0, \text{個別}}) &= \text{Ver}(\hat{y}_0) + \hat{\sigma}^2 \\ &= \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right] \hat{\sigma}^2 \end{aligned} \right\} \tag{5.45}$$

で与えられる。

このように、多くの成書で定式化されている単回帰についてシグマを用いた計算公式は、単回帰分析のみに対するものであり、2変数以上の問題に対して応用することができない。このような状況は、再度繰り返すが、多くの読者に対して応用力を封じ込めるような「ガラスの天井」そのものである。例えば2次式 $y = \beta_0 + \beta_1 x + \beta_2 x^2$ をあてはめ、その95%信頼区間を描きたいとしても、従来の計算方法では、解決の糸口はつかめない。

行列計算の場合ならば、変数の数が増えてもパラメータの共分散行列は、常に $\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$ であり、第2.4節で例示したように、変数の数が増えても95%信頼区間の計算方法は2次形式の適用方法同じである。ただし、行列計算の式を使わない場合には、95%信頼区間の計算式は、4次式になり、計算式を示すこと自体が難儀でもあり、それを理解し応用しようとする読者にとっても優しくない。そのために、式を示さずに図による表示に留められるのが常である。

現実的な対応

Excel の「回帰分析」は手軽に使える優れものであるので、パラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算を付け加えることにより、95%信頼区間の計算を自在にできるようになる。表5.6に示すように、Excel の「回帰分析」を使い、分散分析表から残差の分散（誤差分散の推定値） $\hat{\sigma}^2 = 2.3705$ を、係数の表から $\hat{\beta}_0 = -11.8000$ 、 $\hat{\beta}_1 = 0.8864$ を得る。表5.2で示したと同様にパラメータの共分散行列 $\Sigma(\hat{\beta})$ は、デザイン行列 X を用い

表 5.6 分析ツールの回帰分析およびパラメータの共分散行列の計算

i	X		Y	分散分析表 (定数に 0 を使用) off						
				自由度	変動	分散	分散比	有意 F		
1	1	18	4							
2	1	21	7	回帰	1	69.1364	69.1364	29.1659	0.0006	
3	1	24	10	残差	8	18.9636	2.3705			
4	1	24	10	合計	9	88.1000				
5	1	18	2							
6	1	27	11		係数	標準誤差	t	P-値	下限 95%	上限 95%
7	1	21	9	切片	-11.8000	3.6434	-3.2387	0.0119	-20.2018	-3.3982
8	1	21	5	x	0.8864	0.1641	5.4005	0.0006	0.5079	1.2648
9	1	26	11							
10	1	20	8							
	10.0	220.0		5.6000	-0.2500		13.2745	-0.5926		2.3060
	220.0	4928.0		-0.2500	0.0114		-0.5926	0.0269		$t_{0.05}(10-2)$
	$X^T X$			$(X^T X)^{-1}$			$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$			

$$X^T X = \text{Mmult}(\text{Transpose}(X \text{ の範囲}), X \text{ の範囲})$$

$$= \begin{matrix} \begin{matrix} 10.0 & 220.0 \\ 220.0 & 4928.0 \end{matrix} \\ X^T X \end{matrix}$$

$$\begin{aligned}
 \Sigma(\hat{\beta}) &= (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2 \\
 &= \text{Minverse}(\mathbf{X}^T \mathbf{X} \text{の範囲}) \times \hat{\sigma}^2 \\
 &= \begin{array}{|c|c|} \hline 5.6000 & -0.2500 \\ \hline -0.2500 & 0.0114 \\ \hline \end{array} \times \boxed{2.3060} \\
 &\quad \quad \quad (\mathbf{X}^T \mathbf{X})^{-1} \\
 &= \begin{array}{|c|c|} \hline 13.2745 & -0.5926 \\ \hline -0.5926 & 0.0269 \\ \hline \end{array} \\
 &\quad \quad \quad \Sigma(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2
 \end{aligned}$$

によって計算できる。これらを用いて、表 5.5 で示した回帰直線の 95%信頼区間の計算が可能となる。

平方和の分解に対する補足

式 (5.38) で、全体の平方和 S_T が、回帰の平方和 S_R と誤差平方和 S_e の和に分解できるとしたが、式の展開を省略して結論だけを示したので補足をする。式 (5.38) では、第3項があり、これがゼロとなることを示さなかった。

$$\left. \begin{aligned}
 S_T &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \\
 &= \sum_{i=1}^n [(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})]^2 \\
 &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + 2 \sum_{i=1}^n [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})] \\
 &= S_e + S_R + 2 \sum_{i=1}^n [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})] \\
 &= S_e + S_R
 \end{aligned} \right\} \quad (5.46)$$

第3項は、次のように展開して0となる。

$$\left. \begin{aligned}
 \sum_{i=1}^n [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})] &= \sum_{i=1}^n \hat{\varepsilon}_i \hat{Y}_i - \bar{Y} \sum_{i=1}^n \hat{\varepsilon}_i \\
 &= \sum_{i=1}^n \hat{\varepsilon}_i (\hat{\beta}_0 + \hat{\beta}_1 X_i) \\
 &= \hat{\beta}_0 \sum_{i=1}^n \hat{\varepsilon}_i + \hat{\beta}_1 \sum_{i=1}^n \hat{\varepsilon}_i X_i \\
 &= 0
 \end{aligned} \right\} \quad (5.47)$$

これは、式 (5.11) の正規方程式を ε_i で置き換えた次の式が、0となることを用いている。

$$\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_{i=1}^n (Y_i - \hat{Y}_i) = \sum_{i=1}^n \hat{\varepsilon}_i = 0 \quad (5.48)$$

$$\sum_{i=1}^n [(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i] = \sum_{i=1}^n [(Y_i - \hat{Y}_i) X_i] = \sum_{i=1}^n \hat{\varepsilon}_i X_i = 0 \quad (5.49)$$

5.6. 逆推定値に対する各種の95%信頼区間の推定

逆推定とは何か

前節の図 5.1 を図 5.2 に再掲し，得られた回帰直線に対して反応が $y_0 = 6$ となる x_0 を逆推定するための矢印を重ね書きしている．逆推定値は，回帰直線の式

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_0$$

を \hat{x}_0 について解き，表 5.2 から， $\hat{\beta}_0 = -11.8000$ ， $\hat{\beta}_1 = 0.8864$ なので，

$$\begin{aligned} \hat{x}_0 &= \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \\ &= \frac{6.0 - (-11.800)}{0.8864} = 20.0821 \end{aligned} \quad (5.50)$$

となり，容易に逆推定値 $\hat{x}_0 = 20.0821$ を得ることができる．難しいのは， \hat{x}_0 の95%信頼区間の算出である．これは，推定された逆推定値 \hat{x}_0 が回帰パラメータの比で表されており，線形式に対する合成分散の一般式が使えないからである．

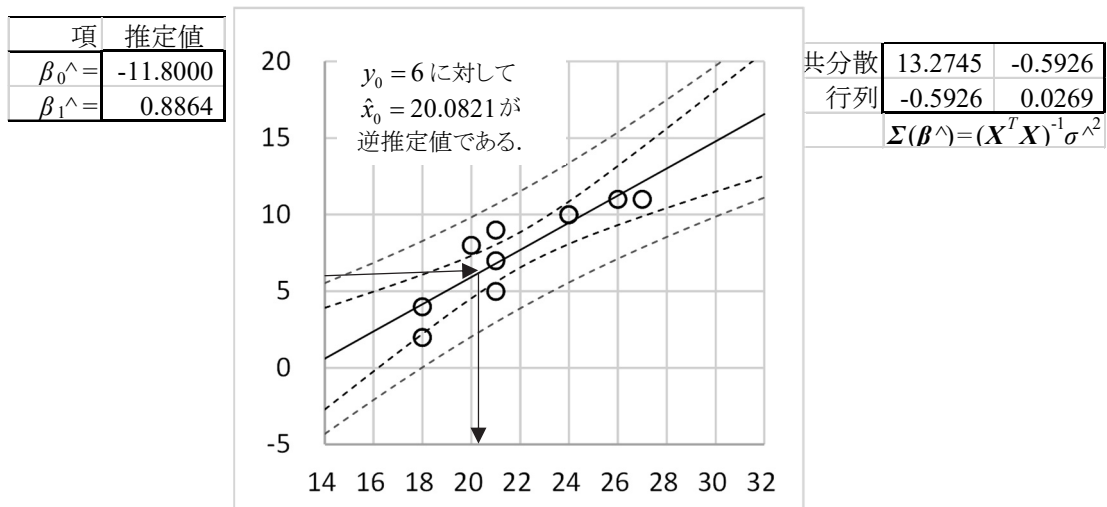


図 5.2 逆推定の例示

逆推定値の95%信頼区間の求め方について，参考になるのは，竹内(1963)，「数理統計学，第29章 IV 回帰直線自体についての推論」である．JMP を用いた逆推定については，芳賀(2016)，「医薬品開発のための統計学，第3部 非線形モデル 改訂版，第1章 (4) JMP による逆推定の解析」が詳しい．パラメータの共分散分散行列を活用については，Collett(2003)，「Modeling Binary Data 2nd. ed., 4.2.1 Approximate standard error of an estimated effective dose」が参考になる．逆推定の95%信頼区間に関して，高橋(2013a)，「応用回帰分析 I - 各種の重み付き回帰における逆推定-」が詳しい．また，高橋(2013b)，「回帰分析・再入門 - 統計ソフ

トが対応していない生物統計の各種の課題を Excel でサクサク解こう」は、「基礎セミナー じっくり勉強すれば身につく統計入門」シリーズの第7回目の公表資料で、本章でのデザイン行列を用いた回帰分析についてスライドを用いて説明をしている。逆推定値の分散の推定について示されていないが、アーミテイジ著、椿・椿訳（2001）、「医学研究のための統計的方法、第3.6節 分散に関するその他の公式」に「比の分散」、「積の分散」、「一般の関数の分散」についても参考になる。

デルタ法による逆推定値に対する近似 95%信頼区間

推定されたパラメータの比についての合成分散は、いわゆるデルタ法によって求めることができる。そのために、求めたい逆推定値 \hat{x}_0 の式 (5.50) に対して、パラメータ $\hat{\beta}_0$ と $\hat{\beta}_1$ で偏微分し、

$$d_0 = \frac{\partial \hat{x}_0}{\partial \hat{\beta}_0} = \frac{\partial (y_0 - \hat{\beta}_0) / \hat{\beta}_1}{\partial \hat{\beta}_0} = -\frac{1}{\hat{\beta}_1} \tag{5.51}$$

$$d_1 = \frac{\partial \hat{x}_0}{\partial \hat{\beta}_1} = \frac{\partial (y_0 - \hat{\beta}_0) / \hat{\beta}_1}{\partial \hat{\beta}_1} = \frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2} \tag{5.52}$$

を得る。これらの偏微分式を行ベクトル \mathbf{d}

$$\mathbf{d} = [d_0 \ d_1] \tag{5.53}$$

としたとき、 \hat{x}_0 の分散 $Var(\hat{x}_0)$ は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する \mathbf{d} の2次形式

$$\begin{aligned} Var(\hat{x}_0) &= \mathbf{d} \Sigma(\hat{\beta}) \mathbf{d}^T \\ &= [d_0 \ d_1] \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \end{bmatrix} \end{aligned} \tag{5.54}$$

で求めることができる。

表 5.7 逆推定の近似 95%信頼区間

y_0	\hat{x}_0	d_0	d_1	$Var(\hat{x}_0)$	L95%	U95%	個別L95	個別U95
2	15.5692	-1.1282	-17.5653	1.7196	12.5453	18.5932	?	?
4	17.8256	-1.1282	-20.1110	0.8992	15.6390	20.0123		
6	20.0821	-1.1282	-22.6567	0.4278	18.5737	21.5904		
8	22.3385	-1.1282	-25.2024	0.3057	21.0636	23.6134		
10	24.5949	-1.1282	-27.7481	0.5326	22.9120	26.2778		
12	26.8513	-1.1282	-30.2938	1.1087	24.4232	29.2793		
T.inv.2T (0.05, 10-2)=					2.3060			

表 5.7 の1行目の $y_0 = 2$ に対する逆推定値 \hat{x}_0 は、表 5.6 または図 5.2 に示されている $\hat{\beta}_0$ と $\hat{\beta}_1$ の推定値を用いて

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{2 - (-11.80)}{0.8864} = 15.5692$$

であり、 \hat{x}_0 に対する $\hat{\beta}_0$ と $\hat{\beta}_1$ の偏微分式は、

$$d_0 = -\frac{1}{\hat{\beta}_1} = -\frac{1}{0.8864} = -1.1282$$

$$d_1 = \frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2} = \frac{-(2 - (-11.80))}{0.8864^2} = -17.5653$$

$$\mathbf{d} = [-1.1282 \quad -17.5653]$$

なので、分散 $Var(\hat{x}_0)$ は、

$$\begin{aligned} Var(\hat{x}_0) &= \mathbf{d} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{d}^T \\ &= [-1.1282 \quad -17.5653] \begin{bmatrix} 13.2745 & -0.5926 \\ -0.5926 & 0.0269 \end{bmatrix} \begin{bmatrix} -1.1282 \\ -17.5653 \end{bmatrix} \\ &= 1.7196 \end{aligned}$$

と推定されている。下側95%点と上側95%点は、

$$\begin{aligned} L95\% &= \hat{x}_0 - t_{0.05}(10-2)\sqrt{Var(\hat{x}_0)} \\ &= 15.5692 - 2.3060 \times \sqrt{2.3705} = 12.5453 \\ U95\% &= \hat{x}_0 + t_{0.05}(10-2)\sqrt{Var(\hat{x}_0)} \\ &= 15.5692 + 2.3060 \times \sqrt{2.3705} = 18.5932 \end{aligned}$$

となる。第1行目で作成した計算式は、行方向にフィルハンドルでコピーすることにより計算される。なお、この方法では、個別データの95%信頼区間の計算ができない。

逆推定値に対する正確な95%信頼区間

逆推定値に対する正確な95%信頼区間の算出方法には、いくつかの方法があるので、図5.2で示した回帰直線の推定値 \hat{y} の95%信頼区間を活用する方法を示す。ある $y_0 = 6$ に対する逆推定値は、

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{6.0 - (-11.800)}{0.8864} = 20.0821$$

と推定される。図5.3に示すように、ある $y_0 = 6$ の水平線2つの95%信頼区間の交点に着目する。水平線と95%信頼上側曲線の交点のX軸 \hat{x}_{L95} が、X軸方向の下側95%点となるが、このままでは推定できない。何らかの探索的な方法が必要となる。

交点の正確な95%信頼区間の下限の推定値 \hat{x}_{L95} に対する回帰の推定値 \hat{y}_{L95} は、未知の \hat{x}_{L95} を用いて

$$\hat{y}_{L95} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} \quad (5.55)$$

である。この \hat{y}_{L95} に対するY軸方向の上側95%点は、 $y_0 = 6$ と等しいので、次の等式が成り立つ。

$$y_0 = \hat{y}_{L95} + t_{0.05} \sqrt{Var(\hat{y}_{L95})} \quad (5.56)$$

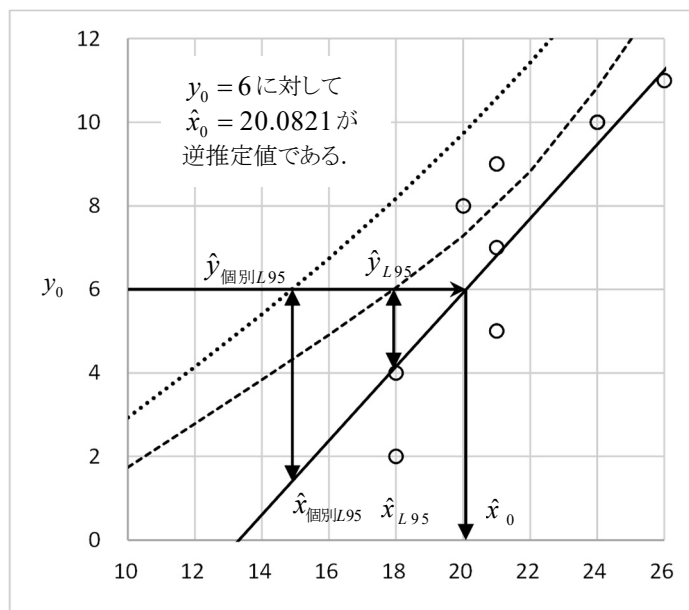


図 5.3 逆推定の 95%信頼区間の算出の例示

推定したいのは、正確な 95%信頼区間の下限の推定値 \hat{x}_{L95} なので、 \hat{y}_{L95} を $\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}$

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} + t_{0.05} \sqrt{\text{Var}(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} \quad (5.57)$$

に置き換える. この式を \hat{x}_{L95} について解くことにより, 逆推定値 \hat{x}_{L95} の下側 95%点 が推定できる. 上の式を右辺の $\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}$ を左辺に移し, 右辺の $t_{0.05}$ を左辺の分母とし, 両辺を平方すると, 次式が得られる.

$$\left(\frac{y_0 - \hat{\beta}_0 - \hat{\beta}_1 \hat{x}_{L95}}{t_{0.05}} \right)^2 = \text{Var}(\hat{\beta}_0) + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \hat{x}_{L95} + \text{Var}(\hat{\beta}_1) \hat{x}_{L95}^2 \quad (5.58)$$

左辺を右辺に移して \hat{x}_{L95} について整理すると,

$$\left[\text{Var}(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_{\alpha}^2} \right] + \left[2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_{\alpha}^2} \right] \hat{x}_{L95} + \left[\text{Var}(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_{\alpha}^2} \right] \hat{x}_{L95}^2 = 0 \quad (5.59)$$

----- a ----- ----- b ----- ----- c -----

が得られる. この複雑な式は, 幸い正確な 95%信頼区間の下限の推定値 \hat{x}_{L95} に関する 2 次式

$$a + b\hat{x}_{L95} + c\hat{x}_{L95}^2 = 0 \quad (5.60)$$

になるので, 2 次式の解の公式

$$\hat{x}_{L95} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \quad (5.61)$$

により, \hat{x}_{L95} を求めることができる. 解は 2 つあるが, 小さい方が下側 95%点 \hat{x}_{L95} となり, 大きい方が上側 95%点 \hat{x}_{U95} となる.

個別データに対する逆推定値の正確な95%信頼区間

個別の上側95%曲線と $y_0 = 6$ を通る水平線との交点は、個別データの下側95%点の推定値 $\hat{x}_{\text{個別}L95}$ に対する回帰の推定値

$$\hat{y}_{\text{個別}L95} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{個別}L95} \tag{5.62}$$

である。個別のデータの上側95%点 $\hat{y}_{\text{個別}L95}$ は、 $y_0 = 6$ に等しいので、次の等式が成り立つ。

$$y_0 = \hat{y}_{L95} + t_{0.05} \sqrt{\text{Var}(\hat{y}_{\text{個別}L95}) + \hat{\sigma}^2} \tag{5.63}$$

推定したいのは、 $\hat{x}_{\text{個別}L95}$ なので、

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{個別}L95} + t_{0.05} \sqrt{\text{Var}(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{個別}L95}) + \hat{\sigma}^2} \tag{5.64}$$

と置き換える。式(5.63)を $\hat{x}_{\text{個別}L95}$ について解くと式(5.59)と同様な結果となるが、2次式のパラメータ a について

$$\text{個別 } a' = a + \hat{\sigma}^2$$

となるだけで、他は同様である。

これらから、表5.8に示すように、与えられた y_0 に対する逆推定値の95%信頼区間、および、個別データの逆推定値の95%信頼区間を推定することができる。

表 5.8 逆推定の正確な2種類の95%信頼区間の推定値

y_0	2次式のパラメータ			逆推定	95%信頼区間			個別	個別95% CL	
	a	b	c	\hat{x}_0	\hat{x}_{L95}	\hat{x}_{U95}	a'	個別L95	個別U95	
2	-22.5382	3.4152	-0.1208	15.5692	10.4967	17.7739	-20.1678	8.4028	19.8678	
4	-33.6710	4.0820	-0.1208	17.8256	14.3037	19.4860	-31.3005	11.7630	22.0267	
6	-46.3081	4.7487	-0.1208	20.0821	17.9324	21.3764	-43.9377	14.9018	24.4070	
8	-60.4497	5.4154	-0.1208	22.3385	21.0020	23.8258	-58.0792	17.7647	27.0632	
10	-76.0957	6.0822	-0.1208	24.5949	23.2245	27.1224	-73.7252	20.3340	30.0130	
12	-93.2461	6.7489	-0.1208	26.8513	25.0382	30.8279	-90.8756	22.6413	33.2247	

表5.8の1行目の $y_0 = 2$ の逆推定値 \hat{x}_0 は、表5.2の $\hat{\beta}_0$ と $\hat{\beta}_1$ の推定値を用いて

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{2 - (-11.80)}{0.8864} = 15.5692$$

である。2次式のパラメータ a , b , c は、

$$a = \text{Var}(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} = 13.2745 - \frac{(2 - (-11.80))^2}{2.3060^2} = -22.5382$$

$$b = 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2 \times (-0.5926) + \frac{2 \times (2 - (-11.80)) \times 0.886}{2.3060^2} = 3.4152$$

$$c = \text{Var}(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_\alpha^2} = 0.0269 - \frac{0.8864^2}{2.3060^2} = -0.1208$$

表 5.2 :

項	推定値	$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} =$	18.9636	共分散	13.2745	-0.5926
$\hat{\beta}_0 =$	-11.8000	$\hat{\sigma}^2 =$	2.3705	行列	-0.5926	0.0269
$\hat{\beta}_1 =$	0.8864	$t_{0.05} =$	2.3060		$\boldsymbol{\Sigma} \wedge = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$	
$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$						

となる。2次式の解は、2つあり下側95%点 \hat{x}_{L95} と上側95%点 \hat{x}_{U95} が同時に求められ、

$$\begin{aligned} \hat{x}_{L95} &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \\ &= \frac{-3.4152 \pm \sqrt{3.4152^2 - 4 \times (-22.5382) \times (-0.1208)}}{2 \times (-0.1208)} = (10.4967, 17.7739) \end{aligned}$$

となる。個別の下側95%点は、

$$\text{個別}a' = a + \hat{\sigma}^2 = -22.5382 + 2.3705 = -20.1678$$

なので、

$$\begin{aligned} \hat{x}_{\text{個別}L95} &= \frac{-b \pm \sqrt{b^2 - 4a'c}}{2c} \\ &= \frac{-3.4152 \pm \sqrt{3.4152^2 - 4 \times (-20.1678) \times (-0.1208)}}{2 \times (-0.1208)} = (8.4028, 19.8678) \end{aligned}$$

と計算される。図 5.3 の逆推定の 95%信頼区間の算出の例示で用いた \hat{x}_{L95} に関する逆推定値は、表 5.8 から該当部分を抜粋した結果を表 5.9 に示すように、 $\hat{x}_0 = -0.296$ 、 $\hat{x}_{L95} = -0.675$ 、 $\hat{x}_{\text{個別}L95} = -1.243$ となる。

表 5.9 近似および正確 95%信頼区間の逆推定値の比較

		逆推定	95%信頼区間		個別95% CL	
	y_0	\hat{x}_0	\hat{x}_{L95}	\hat{x}_{U95}	個別L95	個別U95
正確	6	20.0821	17.9324	21.3764	14.9018	24.4070
近似	"	"	18.5737	21.5904	?	?

Excel ソルバーを用いた逆推定の正確な 95%信頼区間

Excel ソルバーには、目標値を最大化または最小化するパラメータの推定だけでなく、任意に設定した推定値となるようなパラメータを推定することができる。この方法を使うことにより、表 5.5 に示した回帰直線の 95%信頼区間の計算シートを用い、95%信頼区間の曲線が y_0 となるような逆推定値 \hat{x}_0 を Excel ソルバーで探索する。

表 5.10 に示すのは、表 5.5 をコピーした表に対し、1行目の「U95%」が「6」となるように1行目の「x」をソルバーで探索した結果であり、表頭の「x」を「逆推定値 \hat{x}_0 」に代えた結

果である。さらに2行目の「L95%」が「6」となるように2行目の「x」をソルバーで探索した結果である。さらに、3行目の「ŷ」が「6」となるように3行目の「x」をソルバーで探索した結果は、 $y_0 = 6$ となる逆推定値 $\hat{x}_{(y_0=6)} = 20.0821$ が推定されている。この逆推定値 $\hat{x}_{(y_0=6)}$ に対する95%信頼区間は、1行目目と2行目が95%信頼区間 (17.9324, 21.3764) となっている。

表 5.10 Excel ソルバーによる逆推定の実際

逆推定値	切片	逆推定値 x_0^{\wedge}	y^{\wedge}	$Var(y^{\wedge})$	L95%	U95%	個別 L95%	個別 U95%
L95%	1	17.9324	4.0946	0.6827	2.1892	6.0000	:目標値(6)にセット	
U95%	1	21.3764	7.1473	0.2475	6.0000	8.2945		
$x^{\wedge}_{(y_0=6)}$	1	20.0821	6.0000	0.3361	4.6630	7.3370		
個別L95%	1	14.9017	1.4084	1.5943			-3.1833	6.0000
個別U95%	1	24.4070	9.8335	0.3931			6.0000	13.6670
$x^{\wedge}_{(y_0=6)}$	1	20.0821	6.0000	0.3361			2.2062	9.7938
T.inv.2T(0.05, 10-2)=					2.3060			

Excel のソルバーで「目標セルの設定」で当該のセルを選択し、「指定値」が「6」となるように元の表の「x」を変化させた結果で、表頭の「x」は、「逆推定値 \hat{x}_0 」に変更してある。

個別のU95%およびL95%についても同様な手順により、逆推定値 $\hat{x}_{(y_0=6)}$ に対する95%信頼区間は、4行目目と5行目が個別の95%信頼区間 (14.9017, 24.4070) となっている。これらは、表 5.9 に示した推定結果に一致する。

面積の標準偏差を推定したい (<https://mathwords.net/gosadenpa>)

長方形の横の長さを測定し $x=10\text{ cm}$ を得た。標準偏差は 0.3 cm だとする。同様に、縦の長さは $y=5\text{ cm}$ で少しだけ正確に測定できて標準偏差は 0.2 cm だとする。面積は、 $z = xy = 10 \times 5 = 50\text{ cm}^2$ である。面積 z の標準偏差を誤差の伝播式を用いて推定したい。

横の長さ x の分散は、 $Var(x) = 0.3^2$ 、縦の長さの分散は、 $Var(y) = 0.2^2$ であり、 x と y の共分散は、互いに独立して計測しているので $Cov(x, y) = 0$ である。面積 $z = xy$ の分散を推定したい。面積 z を x と y で偏微分すると $d_1 = \partial z / \partial x = y = 5$ 、 $d_2 = \partial z / \partial y = x = 10$ となるので、分散共分散行列を Σ 、偏微分ベクトルを $d = [d_1 \ d_2]$ としたときに、面積 z の分散は、 Σ に関する d の2次形式 $Var(z) = d^T \Sigma d$ によって推定できる (誤差の伝播式の一般化)。面積 z の標準偏差の推定値は、分散の平方根なので、 $SD(z) = \sqrt{Var(z)} = \sqrt{6.25} = 2.5\text{ cm}^2$ となる。

表 5.11 面積の標準偏差の推定 (誤差の伝播式での計算に一致)

		分散共分散行列 Σ		d		Σ		d^T	$Var(z)$	$SD(z)$
$x =$	10	0.3^2	0	5	10	0.09	0	5	6.25	2.50
$y =$	5	0	0.2^2	d_1	d_2	0	0.04	10		
$z = xy =$	50 cm^2	x	y							

5.7. JMP の回帰分析による逆推定

これまで、Excel を用いて回帰分析のパラメータの共分散行列を活用し、各種の 95%信頼区間について示してきた。統計ソフト JMP の「二変量のあてはめ」によって回帰直線の 95%信頼区間のきれいなグラフを手軽に作成できるのであるが、内部でどのような計算式が使われているか出力させることが可能であり、大変興味深い。また、JMP の「モデルのあてはめ」を使うことにより、逆推定も手軽に行なうことができる。

JMP の「二変量の関係」による回帰分析

表 5.12 に示すのは、JMP の「二変量の関係」による回帰分析の結果である。「分散分析」、および、「パラメータの推定値」は、表 5.6 に示した Excel の回帰分析の結果に相当する。図 5.4 に示すのは、デフォルトで出力される回帰直線の散布図に 95%信頼区間、および、個別データの 95%信頼区間を重ね書きした結果であり、図 5.1 に相当する。

表 5.12 JMP の「二変量の関係」による回帰分析

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	1	69.1364	69.1364	29.1659
誤差	8	18.9636	2.3705	p値(Prob>F)
全体(修正済み)	9	88.1000		0.0006*

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	-11.8000	3.6434	-3.2387	0.0119*
x	0.8864	0.1641	5.4005	0.0006*

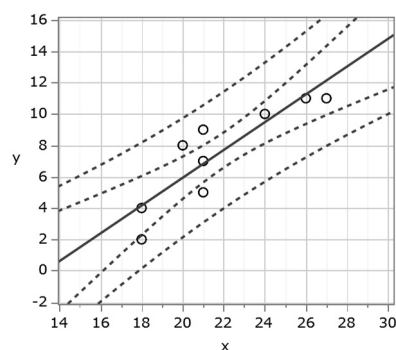


図 5.4 95%信頼区間の出力

回帰直線の推定値の 95%信頼区間の計算式

図 5.4 に示した 2 種類の 95%信頼区間は、どのような計算式が使われているのであろうか。詳細は省くが、「平均の信頼限界の計算式」を選択すると表 5.13 に示すように、「平均 y の下側 95%」および「平均 y の上側 95%」が、元の JMP データファイルに新たな変数として追加される。さらに、「個別の信頼限界の計算式」を選択すると「個別 y の下側 95%」および「個別 y の上側 95%」が新たな変数として追加される。表 5.14 に示すように、これらの追加された変数には、計算式が含まれており、内部での計算式を確認できるようになっている。これらの JMP 内部での計算式は、過度に標準化されている平方和を用いた式 (5.44) および式 (5.45) と異なることが確認できる。

表 5.13 JMP の「二変量の関係」による 95%信頼区間の出力

	x	y	予測値 y	平均 y の 下側95%	平均 y の 上側95%	個別 y の 下側95%	個別 y の 上側95%	
○	1	18	4	4.15	2.27	6.04	0.13	8.17
○	2	21	7	6.81	5.63	8.00	3.07	10.56
○	3	24	10	9.47	8.12	10.83	5.67	13.27
○	4	24	10	9.47	8.12	10.83	5.67	13.27
○	5	18	2	4.15	2.27	6.04	0.13	8.17
○	6	27	11	12.13	9.93	14.33	7.95	16.31
○	7	21	9	6.81	5.63	8.00	3.07	10.56
○	8	21	5	6.81	5.63	8.00	3.07	10.56
○	9	26	11	11.25	9.36	13.13	7.23	15.27
○	10	20	8	5.93	4.57	7.28	2.13	9.73

表 5.14 に示すのは、JMP のデータファイル上で「平均 y の上側 95%」の計算式を表示させた結果である。計算式の `Vec Quadratic ()` 関数は、表 5.2 に示したデザイン行列 \mathbf{X} の積和に対する逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ に対し、デザイン行列 \mathbf{X} の行ベクトルとの式 (5.41) に示した 2 次形式の計算を行なう関数であり、誤差分散 $\hat{\sigma}^2 = 2.3705$ を掛け、 \hat{y}_i の分散 $Var(\hat{y}_i)$ の平方根が計算されている。 $(\mathbf{X}^T \mathbf{X})^{-1}$ に $\hat{\sigma}^2$ を掛けた結果は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ であり、推定値 \hat{y}_i の分散の計算にも $\Sigma(\hat{\beta})$ を挟んだ 2 次形式になっていて、伝統的な偏差平方和を用いた計算ではないことが確認できる。

表 5.14 JMP の「二変量の関係」による上側 95%の計算式

$$\begin{aligned}
 & -11.8 \\
 & + 0.8863636364 \cdot x \\
 & + 2.3060041352 \cdot \sqrt{\text{Vec Quadratic} \left(\begin{bmatrix} 5.6 & -0.25 \\ -0.25 & 0.011 \end{bmatrix}, [1] \parallel x \right) \cdot 2.3704545455}
 \end{aligned}$$

表 5.6 で示されている Excel による回帰分析と対比すると `VecQuadratic ()` 関数は、

$$Var(\hat{y}_i) = \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} 5.6000 & -0.2500 \\ -0.2500 & 0.0114 \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} * 2.3705$$

$(\mathbf{X}^T \mathbf{X})^{-1}$
 σ^2

と解釈される。回帰の 95%信頼区間の上側の計算式は、

$$\begin{aligned}
 L95\% &= \hat{\beta}_0 + \hat{\beta}_1 x_i + t_{0.05}(10-2) \sqrt{Var(\hat{y}_i)} \\
 &= -11.80 + 0.8864x + 2.3060 \sqrt{Var(\hat{y}_i)}
 \end{aligned}$$

であることが、読み取れる。

表 5.15 に示すのは、「個別 y の上側 95%」の計算式である。この式は、

$$\begin{aligned} \text{個別}U95\% &= \hat{\beta}_0 + \hat{\beta}_1 x_i + t_{0.05}(10-2)\sqrt{\text{Var}(\hat{y}_i) + \hat{\sigma}^2} \\ \sqrt{\text{Var}(\hat{Y}_i) + \hat{\sigma}^2} &= \sqrt{\text{Vec Quadratic}((X^T X)^{-1}, [1 \ x])\hat{\sigma}^2 + \hat{\sigma}^2} \end{aligned}$$

のように 2 次形式の関数が使われていることがわかる。これらのことから、JMP での回帰分析は、デザイン行列 X を用いた計算となっていることが確認できる。

表 5.15 JMP の「二変量の関係」による個別データの上側 95%の計算式

The screenshot shows a JMP calculation script with the following content:

```

-11.8
+ 0.8863636364 * x
+ 2.3060041352
+ sqrt(Vec Quadratic([5.6, -0.25; -0.25, 0.011], [1] || x) * 2.3704545455 + 2.3704545455)
    
```

JMP の「モデルのあてはめ」による逆推定値の 95%信頼区間

逆推定を行い、Excel での計算結果との相互検証を行なう。JMP の「二変量の関係」での回帰分析は、信頼区間の標示などで豊富な機能があるが、逆推定には対応していない。表 5.16 に示す「モデルのあてはめ」による回帰分析を用いて逆推定を行ない、表 5.17 に結果を示す。

表 5.16 JMP の「逆推定」の設定

The screenshot shows the JMP 'Inverse Prediction' dialog box with the following settings:

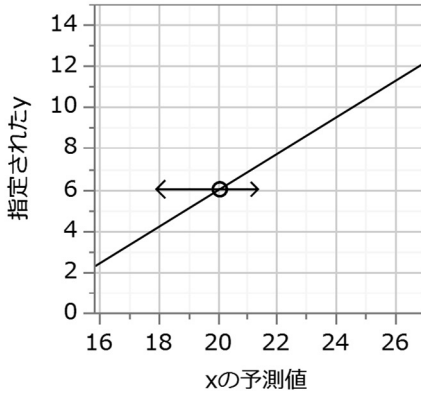
- Model: 二変量の関係
- Response: y
- Confidence Level: 0.95
- Direction: 両側
- Checkbox: 応答変数の期待値ではなく、個々の値に対する信頼区間

A tooltip at the bottom right states: "Yの値(およびその他の説明変数の値)から、Xの値を予測したい場合。信頼区間も求められる。"

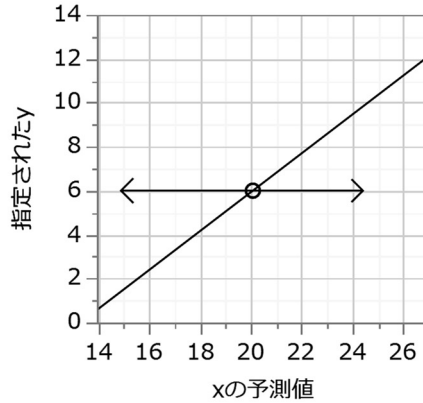
表 5.17 JMP の「モデルのあてはめ」による逆推定の結果とグラフ表示

逆推定				逆推定			
指定されたy	xの予測値	下側95%	上側95%	指定されたy	xの予測値	下側95%	上側95%
6.0000	20.0821	17.9324	21.3764	6.0000	20.08205	14.9018	24.4070

応答の期待値に対する信頼区間



個別の応答に対する信頼区間



この JMP の逆推定による結果は，表 5.9 の正確な 95%信頼区間の結果と完全に一致する．

表 5.9 抜粋

	逆推定	95%信頼区間		個別95% CL	
		x^{\wedge}_0	x^{\wedge}_{L95}	個別L95	個別U95
正確	6	20.0821	17.9324	14.9018	24.4070

逆推定値の 95%信頼区間の直接推定

推定された回帰直線のパラメータを用い，ある y_0 に対する逆推定値を $\hat{x}_0 = (y_0 - \hat{\beta}_0) / \hat{\beta}_1$ によって求め，その分散の推定のために \hat{x}_0 をパラメータで偏微分し，パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式の計算を必要とし，煩雑な手順を示した．そこで，逆推定値 \hat{x}_0 およびその分散を直接推定する方法を示す．

一般的に回帰直線は，切片 β_0 と傾き β_1 をパラメータとして推定しているのだが，ある y_0 とその逆推定値 \hat{x}_0 を原点とするようなオフセット式を想定しよう．Y 軸となる反応 y に対し $y' = y - y_0$ ，X 軸の変数 x に対し $x' = x - \hat{x}_0$ とする通る回帰式は，

$$\left. \begin{aligned} y_i - y_0 &= \beta_1(x_i - \hat{x}_0) + \varepsilon_i \\ y_i &= y_0 + \beta_1(x_i - \hat{x}_0) + \varepsilon_i \end{aligned} \right\} \quad (5.65)$$

となる．ただし， y_0 は，与えられた定数であるが，逆推定値 \hat{x}_0 は，データから推定される推定値であるので， $\beta_2 = \hat{x}_0$ と置き換え，傾き β_1 と逆推定値 β_2 を推定し，それらの分散から，逆推定値 \hat{x}_0 の 95%信頼区間を求めることができる．

式 (5.65) は、線形モデル式ではないので、残差 $\varepsilon_i = y_i - \hat{y}_i$ に対し線形モデルを適用する方法でパラメータの推定ができる。この方法は、第10章から第13章で示す非線形回帰なのであるが、手短かに計算方法を示す。回帰式 (5.65) をパラメータ β_1 および β_2 で偏微分

$$z_{1,i} = \frac{\partial y_i}{\partial \beta_1} = x_i - \beta_2, \quad z_{2,i} = \frac{\partial y_i}{\partial \beta_2} = -\beta_1 \quad (5.66)$$

し、(10行×2列)の偏微分行列 \mathbf{Z} とする。表 5.18 に示すように、 $y_0 = 6.0$ を設定し、初期値 $\hat{\beta}_1 = 1, \hat{\beta}_2 = 20$ に対し、式 (5.65) により推定値 \hat{y}_i を計算する。残差 $\varepsilon_i = y_i - \hat{y}_i$ について平方和 S_e を計算する。Excel のソルバーで、 S_e を最小にするように $(\hat{\beta}_1, \hat{\beta}_2)$ を変化させた結果が示され、逆推定値 $\hat{\beta}_2 = 20.0820$ が推定されている。偏微分行列 \mathbf{Z} を“デザイン行列 \mathbf{X} ”と見なすことにより、 $(\hat{\beta}_1, \hat{\beta}_2)$ に対する共分散行列 $\Sigma(\hat{\beta})$ が得られるので、その対角要素から $Var(\hat{\beta}_2) = 0.4278$ が得られ、一般的な方法により近似 95%信頼区間が (18.5737, 21.5904) が推定されている。この結果は、表 5.9 に示した近似計算の結果に一致する。

表 5.18 逆推定値の 95%信頼区間の直接推定

$y_0 = 6.0$				共分散行列		近似95%信頼区間		
初期値		推定値		$\Sigma(\hat{\beta}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \sigma^2$		$SE(\beta_i)$	$L95\%$	$U95\%$
$\hat{\beta}_1 =$	1.0	$\hat{\beta}_1 =$	0.8864	0.0269	-0.0583	0.1641	0.5079	1.2648
$\hat{\beta}_2 =$	20.0	$\hat{\beta}_2 =$	20.0820	-0.0583	0.4278	0.6541	18.5737	21.5904
		$S_e =$	18.9636	偏微分行列 \mathbf{Z}		$t_{0.05(10-2)} =$	2.3060	
i	x	y	\hat{y}	$\varepsilon = y - \hat{y}$	$z_1 = \partial_{\beta_1}$	$z_2 = \partial_{\beta_2}$		
1	18	4	4.1546	-0.1546	-2.0820	0.8864		
2	21	7	6.8137	0.1863	0.9180	0.8864		
3	24	10	9.4727	0.5273	3.9180	0.8864		
4	24	10	9.4727	0.5273	3.9180	0.8864		
5	18	2	4.1546	-2.1546	-2.0820	0.8864		
6	27	11	12.1318	-1.1318	6.9180	0.8864		
7	21	9	6.8137	2.1863	0.9180	0.8864		
8	21	5	6.8137	-1.8137	0.9180	0.8864		
9	26	11	11.2455	-0.2455	5.9180	0.8864		
10	20	8	5.9273	2.0727	-0.0820	0.8864		

式 (5.65) は、一つの変数に対し、一つのパラメータの積の和で定義される線形モデルではないことから、非線形最小2乗法によりパラメータの推定を行なう必要がある。式 (5.65) に対して JMP の非線形回帰を使えば、表 5.17 右に示した回帰の逆推定 \hat{x}_0 の正確な 95%信頼区間を直接推定することができる。詳しくは、芳賀 (2016) の非線形モデルについての解説資料、大和田 (2010)、線形モデルと非線形モデルの基本的な考え方—逆推定の解析、標準誤差と信頼限界—、中西 (2016)、じっくり勉強すれば身につく統計入門 第12回—非線形回帰を用いた逆推定の基礎—を参照のこと。

6. 伝統的な共分散分析からの脱却

共分散分析は、歴史的に難解な解析法として認識されてきた。質的変数に対する伝統的な平方和の分解による解析方法は、量的変数である共変量を含めた共分散分析に適用できない。そのために、偏差平方和を用いた単回帰分析を重層的に適用しつつ、交互作用を含む分散分析表を作成し、交互作用のないことを確認して交互作用項を除き、平行な直線のあてはめに帰着し、質的因子の水準間の比較を行なう方法が定式化されてきた。他方、質的変数をダミー変数に置き換えて線形モデルとした共分散分析の解析法も知られてはいるのだが、統計ソフトを使うことを前提とするために、解析方法がブラック・ボックス的となり、理解の妨げになっている。そこで、先人たちの創意工夫に満ちた伝統的な解析方法を Excel の力を借りつつ忠実に再現する。さらに、Excel の回帰分析により、質的変数をダミー変数に置き換え、線形モデルによる共分散分析の解析法を導入すると共に、無償で継続的に提供されている Ondemand SAS の GLM プロシジャを用いた共分散分析の新たな活用方法についても示す。

6.1. 伝統的な共分散分析の解析手順

第 1.2 節の表 1.12 で取り上げた「季節ごとの洗浄水の温度 x と回収液の濃度 y 」では、JMP の「モデルのあてはめ」を用いて交互作用を考慮した共分散分析の解析法を示した。さらに、Excel による交互作用を含むデザイン行列 \mathbf{X} を活用した解析法を合わせて示した。取り上げるデータは、奥野ら(1981)「多変量解析法《改訂版》、第 II 部 重回帰分析、第 7 章 層別因子を含む回帰分析」の事例 7.1 である。原著に示されている解析法は、伝統的な共分散分析の解析手順であり、共分散分析を取り上げている成書で共通の解析方法として定着している。ここでは、原著で示されている解析方法を Excel の力を借りつつ忠実に再現することから始める。

季節ごとの回帰直線

表 6.1 に表 1.12 のデータを再掲する。このデータは、亜硫酸ガス回収塔の洗浄水の温度 x と、回収液の濃度 y との関係を調べたが、測定回数が少ないので、長期間にわたって収集されたデータを使わなければならない。その間に季節の変化ばかりでなく、操業条件も変わっているため、データを 4 つ季節に分けた結果が示されている。

表 6.1 季節ごとの洗浄水の温度 x と回収液の濃度 y [表 1.12 再掲]

No.	A ₁ (7月・8月)		A ₂ (9月・10月)		A ₃ (11月・12月)		A ₄ (1月・2月)	
	x	y	x	y	x	y	x	y
1	30	8.0	34	9.1	27	15.0	16	23.4
2	28	10.5	19	19.4	20	21.0	16	28.2
3	30	8.2	22	20.5	14	24.2	20	29.5
4	29	13.0	25	14.2	18	15.3	16	22.2
5	28	10.1	22	11.0	12	17.3	6	40.2
6	17	16.2	25	19.1	18	24.7	16	36.6
7	24	16.1	22	16.0	13	23.9	12	35.5
8	22	13.4	23	11.1	21	20.2	11	42.1
9	30	13.0	28	12.0	24	13.6	9	42.5
10	30	7.3	24	17.1			22	25.8
11	25	14.0	30	13.3				
12			24	8.9				
平均	26.64	11.80	24.83	14.31	18.56	19.47	14.40	32.60
		総平均	21.48	19.11	$N=42$			

図 6.1 に示すのは、得られたデータの関連を概観するために、洗浄水の温度 x と回収液の濃度 y について Excel で作成した層別散布図上に季節別の回帰直線を上書きした結果である。季節が A₄ (1月・2月) の場合に多重比較などをするまでもなく明らかに他の季節に比べ洗浄水の温度 x が低い。伝統的な共分散分析の観点からは、季節 A のそれぞれの共変量 x の分布 (平均値) が明らかに異なり、共分散分析の適用の前提を満たしていない。全ての季節で共通となる洗浄水の温度 x が 20 度で比べた場合でも、A₄ (1月・2月) の回収液の濃度 y が高めになっている。そのために、季節に関連する操業条件などの他の因子の影響があると示唆される。

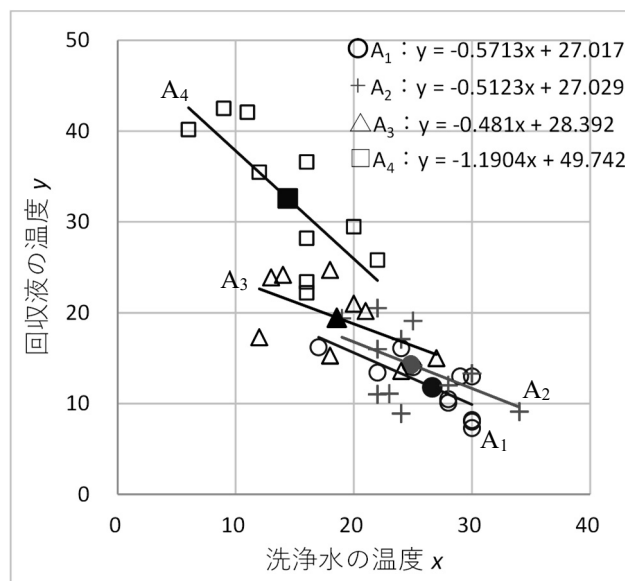


図 6.1 Excel の散布図による季節ごとの 4 本の回帰直線のあてはめ

Excel の散布図で最初に A₁ の (x, y) をプロット・○印に変更する。さらに、A₂, A₃, A₄ を順次「データの選択」で重ね書きし、書式を整え、軸目盛の調整などを行っている。回帰直線は、「近似曲線の追加」の「線形近似」で重ね書きしている。

伝統的な共分散分析は、1 因子実験モデルでの解析に際し、反応 y に影響を与えることがはっきりしている量的変数があるが、実験に際し一定に保つことがどうしてもできない。この量的変数を「共変量」とし、統計的に共変量の影響を除去し、季節 A の水準間の比較をしたい。そのために、共変量の分布が季節 A に対し均一でなければ共変量として扱えないと認識されている。季節 A により洗浄水の温度 x の分布が異なるので共変量としての前提が満たされないことを無視したとしても、更なる共分散分析の適用条件とし、季節 A と洗浄水の温度 x の間に交互作用がないことも確認しなければならない。

重層的な回帰分析の適用

伝統的な共分散分析は、洗浄水の温度 x と季節 A との交互作用の平方和 S_{xxA} を含む分散分析表を作成するために、単回帰分析を重層的に適用する解析方法となっている。

(手順 1) 全データに対する回帰分析

表 6.1 に示した季節 A の水準ごとのデータを全て行方向に並び替え、偏差平方和を用いた回帰分析を行なう。まず、 x と y に関する偏差平方和 ($S_{xx}^{(全体)}$, $S_{xy}^{(全体)}$, $S_{yy}^{(全体)}$) を計算する。

$$\left. \begin{aligned} S_{xx}^{(全体)} &= \sum_{i=1}^4 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})^2 = 1788.48 \\ S_{xy}^{(全体)} &= \sum_{i=1}^4 \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) = -2124.24 \\ S_{yy}^{(全体)} &= S_T^{(全体)} = \sum_{i=1}^4 \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 = 3654.82 \end{aligned} \right\} \quad (6.1)$$

$S_{xy}^{(全体)}$ の計算は Excel で $\bar{x}_{..} = \text{Average}(x \text{ の範囲})$, $\bar{y}_{..} = \text{Average}(y \text{ の範囲})$ を前もって計算し、SumProduct (x の範囲 $-\bar{x}_{..}$, y の範囲 $-\bar{y}_{..}$) 関数により配列として計算する。 $S_{xx}^{(全体)}$, $S_{yy}^{(全体)}$ は、SumSq (x の範囲 $-\bar{x}_{..}$) 関数で同様に配列として計算する。これらの結果からパラメータ ($\hat{\beta}_1^{(全体)}$, $\hat{\beta}_0^{(全体)}$) を

$$\left. \begin{aligned} \text{傾き: } \hat{\beta}_1^{(全体)} &= \frac{S_{xy}^{(全体)}}{S_{xx}^{(全体)}} = \frac{-2124.24}{1788.48} = -1.1877 \\ \text{切片: } \hat{\beta}_0^{(全体)} &= \bar{y}_{..} - \hat{\beta}_1^{(全体)} \bar{x}_{..} = 19.1119 - (-1.1877) \times 21.4762 = 44.6200 \end{aligned} \right\} \quad (6.2)$$

として推定する。次いで分散分析に必要な回帰の平方和 $S_R^{(全体)}$ および残差平方和 $S_e^{(全体)}$ を推定する。なお、 $S_e^{(全体)}$ には、季節 A の平方和 S_A および交互作用の平方和 $S_{R \times A}$ が含まれている。

$$\left. \begin{aligned} S_R^{(全体)} &= \frac{(S_{xy}^{(全体)})^2}{S_{xx}^{(全体)}} = \frac{(-2124.24)^2}{1788.48} = 2523.03 \\ S_e^{(全体)} &= S_T^{(全体)} - S_R^{(全体)} = 3654.82 - 2523.03 = 1131.79 \end{aligned} \right\} \quad (6.3)$$

(手順 2) 季節 A の水準ごとの回帰分析の併合

表 6.2 に示す計算シート上で、季節 A の水準ごとの偏差平方和 ($S_{xy}^{(i)}$, $S_{xx}^{(i)}$, $S_{yy}^{(i)}$) を計算し、 i に関する合計を

表 6.2 季節ごとの回帰分析の併合

A ₁		A ₂		A ₃		A ₄							
x	y	x	y	x	y	x	y		A ₁	A ₂	A ₃	A ₄	和
30	8.0	34	9.1	27	15.0	16	23.4	n_i	11	12	9	10	42
28	10.5	19	19.4	20	21.0	16	28.2	\bar{x}_i	26.64	24.83	18.56	14.40	—
30	8.2	22	20.5	14	24.2	20	29.5	\bar{y}_i	11.80	14.31	19.47	32.60	—
29	13.0	25	14.2	18	15.3	16	22.2	$S_{xx}^{(i)}$	178.55	183.67	204.22	216.40	782.83
28	10.1	22	11.0	12	17.3	6	40.2	$S_{xy}^{(i)}$	-102.00	-94.08	-98.23	-257.60	-551.92
17	16.2	25	19.1	18	24.7	16	36.6	$\hat{\beta}_1^{(i)}$	-0.57	-0.51	-0.48	-1.19	—
24	16.1	22	16.0	13	23.9	12	35.5	$\hat{\beta}_0^{(i)}$	27.02	27.03	28.39	49.74	—
22	13.4	23	11.1	21	20.2	11	42.1	$S_{yy}^{(i)}$	100.36	181.85	148.76	538.44	969.41
30	13.0	28	12.0	24	13.6	9	42.5	$S_R^{(i)}$	58.27	48.19	47.25	306.64	460.36
30	7.3	24	17.1			22	25.8	$S_e^{(i)}$	42.09	133.65	101.51	231.80	509.05
25	14.0	30	13.3										
		24	8.9										

$$\left. \begin{aligned}
 S_{xx}^{(\text{和})} &= \sum_{i=1}^4 S_{xx}^{(i)} = 178.55 + 183.67 + 204.22 + 216.40 = 782.83 \\
 S_{xy}^{(\text{和})} &= \sum_{i=1}^4 S_{xy}^{(i)} = (-102.00) + (-94.08) + (-98.23) + (-257.60) = -551.92 \\
 S_{yy}^{(\text{和})} &= \sum_{i=1}^4 S_{yy}^{(i)} = 100.36 + 181.85 + 148.76 + 538.44 = 969.41
 \end{aligned} \right\} \quad (6.4)$$

として求め ($S_{xy}^{(\text{和})}$, $S_{xx}^{(\text{和})}$, $S_{yy}^{(\text{和})}$) とする。また、季節 A の水準ごとの回帰係数を

$$\hat{\beta}_1^{(i)} = \frac{S_{xy}^{(i)}}{S_{xx}^{(i)}}, \quad \hat{\beta}_0^{(i)} = \bar{y}_i - \hat{\beta}_1^{(i)} \bar{x}_i. \quad (6.5)$$

で求め、図 6.1 に示された回帰係数に一致することを確認する。水準ごとの回帰の平方和 $S_R^{(i)}$ を求め、それらの和を計算する。 $S_R^{(i)}$ の和を $S_R^{(\text{和})}$, $S_{yy}^{(i)}$ の和を $S_{yy}^{(\text{和})} = S_T^{(\text{和})}$ とし、それらの差から残差平方和 $S_e^{(\text{和})}$ を求める。

$$\left. \begin{aligned}
 S_R^{(\text{和})} &= \sum_{i=1}^4 \frac{(S_{xy}^{(i)})^2}{S_{xx}^{(i)}} = 58.27 + 48.19 + 47.25 + 306.64 = 460.36 \\
 S_e^{(\text{和})} &= S_T^{(\text{和})} - S_R^{(\text{和})} = 969.41 - 460.36 = 509.05
 \end{aligned} \right\} \quad (6.6)$$

(手順 3) 季節 A について共通の傾き $\hat{\beta}_1^{(\text{共通})}$ の推定

手順 2 で求めた $S_{xx}^{(\text{和})} = 782.83$, $S_{xy}^{(\text{和})} = -551.92$ から共通の傾き $\hat{\beta}_1^{(\text{共通})} = -0.7050$ を算出する。共通の傾きでの回帰の平方和 $S_R^{(\text{共通})} = 389.11$ を推定し、 $S_T^{(\text{和})} = S_{yy}^{(\text{和})} = 969.41$ との差から交互作用 $S_{xxA}^{(\text{共通})}$ を含む $S_e^{(\text{共通})} = 580.29$ を算出する。

$$\hat{\beta}_1^{(\text{共通})} = \frac{S_{xy}^{(\text{和})}}{S_{xx}^{(\text{和})}} = \frac{-551.92}{782.83} = -0.7050 \quad (6.7)$$

$$\left. \begin{aligned}
 S_R^{(\text{共通})} &= \frac{(S_{xy}^{(\text{和})})^2}{S_{xx}^{(\text{和})}} = \frac{(-551.92)^2}{782.83} = 389.11 \\
 S_e^{(\text{共通})} &= S_T^{(\text{和})} - S_R^{(\text{共通})} = 969.41 - 389.11 = 580.29
 \end{aligned} \right\} \quad (6.8)$$

原点を通る回帰直線

(手順 2 および 3 の別解) 原点を通る回帰分析の適用

手順 2 および手順 3 は、季節 A の水準ごとに共通の傾きを推定するために単回帰分析を重層的に適用する方法であるが、見通しが悪い。そこで、別の方法として、表 6.3 に示すように季節 A の水準ごとの偏差 $x'_{ij} = x_{ij} - \bar{x}_i$, $y'_{ij} = y_{ij} - \bar{y}_i$ を計算し、全データ $n=42$ に対する原点を通る Excel の回帰分析を適用することにより、共通の傾きを求めることができる。表 6.4 に分散分析表およびパラメータ (係数) の推定結果を示す。

表 6.3 季節 A の水準ごとの偏差の計算

No.	A	x	y	水準ごとの偏差		
				x'	y'	
1	A ₁	30	8.0	3.4	-3.8	
2	A ₁	28	10.5	1.4	-1.3	
:						
11	A ₁	25	14.0	-1.6	2.2	
12	A ₂	34	9.1	9.2	-5.2	
:						
23	A ₂	24	8.9	-0.8	-5.4	
24	A ₃	27	15.0	8.4	-4.5	
:						
32	A ₃	24	13.6	5.4	-5.9	
33	A ₄	16	23.4	1.6	-9.2	
:						
41	A ₄	9	42.5	-5.4	9.9	
42	A ₄	22	25.8	7.6	-6.8	
				782.83	969.41	-551.92
				$S_{x'x'}^{(和)}$	$S_{y'y'}^{(和)}$	$S_{x'y'}^{(和)}$

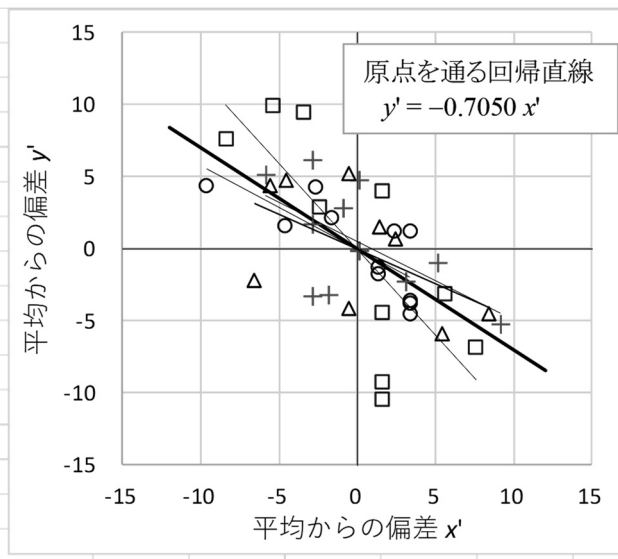


図 6.2 原点を通る回帰直線

表 6.4 Excel による原点を通る回帰分析 (定数に 0 を使用) on

	自由度	変動	分散	分散比	係数	標準誤差	t	P-値
回帰	1	389.11	389.11	27.49	切片	0.0000	#N/A	#N/A
残差	41	580.29	14.15		x'	-0.7050	0.1345	-5.2433
合計	42	969.41						0.0000

傾きが $\hat{\beta}_1^{(原点)} = -0.7050$ と推定され、式 (6.7) の $\hat{\beta}_1^{(共通)}$ に一致する。また、分散分析表の平方和も

$$\left. \begin{aligned}
 \hat{\beta}_1^{(原点)} &= \hat{\beta}_1^{(共通)} = -0.7050 \\
 S_T^{(原点)} &= S_{yy}^{(和)} = 969.41 \\
 S_R^{(原点)} &= S_R^{(共通)} = 389.11 \\
 S_e^{(原点)} &= S_e^{(共通)} = 580.29
 \end{aligned} \right\} \quad (6.9)$$

のように一致する。

交互作用を含む分散分析表の作成

(手順4) 手順1~3で求めた平方和から分散分析表の作成

手順1の $S_{yy}^{(全体)}$ が総平方和 $S_T = S_{yy}^{(全体)} = 3654.82$, $S_R^{(全体)}$ が共変量 x の平方和 $S_x = S_R^{(全体)} = 2523.03$ となる. 手順1の $S_e^{(全体)} = 1131.79$ から手順2で示した $S_e^{(共通)} = 580.29$ との差が, 季節Aの平方和 $S_A = 551.49$ となる. 手順3の $S_e^{(共通)} = 580.29$ から手順2で示した $S_e^{(和)} = 509.05$ との差が, 交互作用 $S_{x \times A} = 71.25$ となる.

$$\left. \begin{aligned} S_T &= S_{yy}^{(全体)} = 3654.82 \\ S_x &= S_R^{(全体)} = 2523.03 \\ S_A &= S_e^{(全体)} - S_e^{(共通)} = 1131.79 - 580.29 = 551.49 \\ S_{x \times A} &= S_e^{(共通)} - S_e^{(和)} = 580.29 - 509.05 = 71.25 \\ S_e &= S_e^{(和)} = 509.05 \end{aligned} \right\} \quad (6.10)$$

これらの平方和を用いて表6.5に示す分散分析表を作成し, 交互作用の有無について統計的判断を行なう. 表6.5に示すように交互作用の p 値は, $p = 0.2107$ と有意ではないので, 平行な直線があてはめられるとみなす. なお, 第1.2節では, 交互作用を無視しないとした.

表 6.5 交互作用を含む分散分析表

要因	df	平方和	平均平方	F	p
x	1	2523.03	2523.03	168.52	0.0000
A	3	551.49	183.83	12.28	0.0000
$x \times A$	3	71.25	23.75	1.59	0.2107
e	34	509.05	14.97	1.00	
全体	41	3654.82			

ここに示した平方和は, 逐次平方和 (タイプ I) であり, 表 1.13 に示した平方和は, タイプ III の平方和であり一致しない.

平行な回帰直線のあてはめ

(手順5) 交互作用の平方和 $S_{x \times A}$ を残差平方和へプーリング

交互作用 ($x \times A$) が有意ではないので, 交互作用の平方和 $S_{x \times A} = 71.25$ を残差平方和 $S_e = 509.05$ にプーリングして表 6.6 示すように新たな残差平方和 S'_e

$$S'_e = S_e^{(共通)} + S_{x \times A} = 509.05 + 71.25 = 580.29$$

とし, 分散分析表を作り直す. なお, S'_e は, 手順3別解で求めた $S_e^{(原点)} = 580.29$ に一致する.

表 6.6 交互作用をプーリングした分散分析表

要因	df	平方和	平均平方	F	p
x	1	2523.03	2523.03	160.87	0.0000
A	3	551.49	183.83	11.72	0.0000
e	37	580.29	15.68	1.00	
全体	41	3654.82			

これまでの手順は, 先人たちの努力の賜物であるが, 化石的な手順なのでまったく推奨できない. 次節のダミー変数を用いた Excel の回帰分析を参照のこと.

(手順 6) 平行線をあてはめた場合の季節 A の水準ごとの切片の推定

手順 3 で求めた季節 A についての共通の傾き $\hat{\beta}_1^{(共通)} = -0.7050$ とする切片 $\hat{\beta}_0^{(i)}$ を

$$\hat{\beta}_0^{(i)} = \bar{y}_{i\cdot} - \hat{\beta}_1^{(共通)} \bar{x}_{i\cdot} \quad (6.11)$$

で推定すると表 6.9 に示すように $\hat{\beta}_0^{(1)} = 30.58$, $\hat{\beta}_0^{(2)} = 31.82$, $\hat{\beta}_0^{(3)} = 32.55$, $\hat{\beta}_0^{(4)} = 42.75$ が得られる.

表 6.7 傾きを共通とする場合の切片の推定

A	$\bar{x}_{i\cdot}$	$\bar{y}_{i\cdot}$	切片 $\hat{\beta}_0^{(i)}$	$\hat{\beta}_1^{(共通)}$
A ₁	26.64	11.80	30.58	-0.705
A ₂	24.83	14.31	31.82	-0.705
A ₃	18.56	19.47	32.55	-0.705
A ₄	14.40	32.60	42.75	-0.705

これらを用いて図 6.3 に示すように共通の傾きを持つ平行な回帰直線を引く.

表 6.8 回帰直線の範囲の設定

A		x	共通 y [^]
A ₁	min	17	18.59
	平均	26.64	11.80
	max	30	9.43
A ₂	min	19	18.42
	平均	24.83	14.31
	max	34	7.85
A ₃	min	12	24.09
	平均	18.56	19.47
	max	27	13.51
A ₄	min	6	38.52
	平均	14.40	32.60
	max	22	27.24

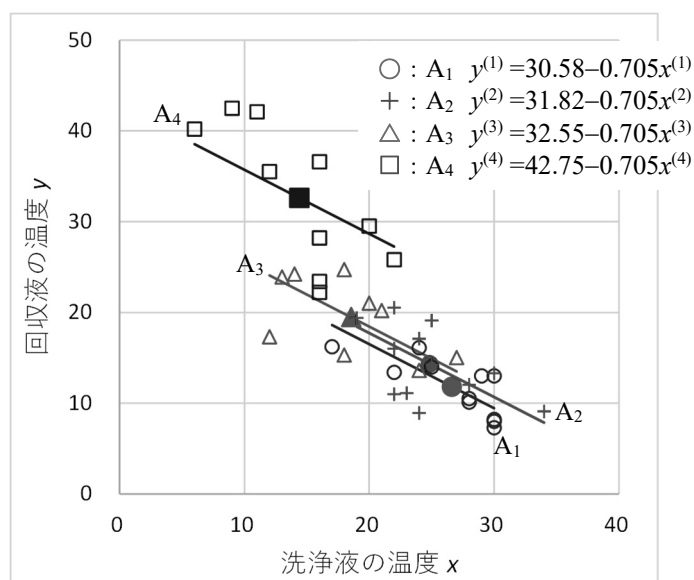


図 6.3 平行な 4 本の回帰直線のあてはめ

平行な直線のあてはめは, 図の左の直線の始点と終点をあらかじめ設定し, 「データの選択」で順次「追加」して描く. 各水準の重心については, マーカ ● などを設定している.

平行な直線間の差の 95%信頼区間.

(手順 7) 季節 A の第 4 水準を基準とする水準間の差

図 6.3 に示したように, 季節 A ごとの平行線に着目したときに, 季節 A の A₄ 水準を基準とし, それぞれの A_i 水準との平行線間の差と 95%信頼区間を推定し, 図 6.4 に示すようにひげ付き線グラフを作成したい.

平行な回帰直線の任意の x に対し、基準となる水準を i' とすると、 Y 軸方向の差 $d^{(i-i')}$ は、

$$\left. \begin{aligned} d^{(i-i')} &= (\hat{\beta}_0^{(i)} + \hat{\beta}_1^{(共通)}x) - (\hat{\beta}_0^{(i')} + \hat{\beta}_1^{(共通)}x) \\ &= \hat{\beta}_0^{(i)} - \hat{\beta}_0^{(i')} \end{aligned} \right\} \quad (6.12)$$

のように、切片の差に帰着する。切片の差の分散を求めるために、 $\bar{y}^{(i)} = \hat{\beta}_0^{(i)} + \hat{\beta}_1^{(共通)}\bar{x}^{(i)}$ を $\hat{\beta}_0^{(i)}$ について解いた式、および、 $\bar{y}^{(i')} = \hat{\beta}_0^{(i')} + \hat{\beta}_1^{(共通)}\bar{x}^{(i')}$ を $\hat{\beta}_0^{(i')}$ について解いた式を代入し、

$$\left. \begin{aligned} \hat{\beta}_0^{(i)} - \hat{\beta}_0^{(i')} &= (\bar{y}^{(i)} - \hat{\beta}_1^{(共通)}\bar{x}^{(i)}) - (\bar{y}^{(i')} - \hat{\beta}_1^{(共通)}\bar{x}^{(i')}) \\ &= \bar{y}^{(i)} - \bar{y}^{(i')} - (\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(共通)} \end{aligned} \right\} \quad (6.13)$$

を得る。

(手順 8) 水準間の差の 95%信頼区間

回帰直線間の差 $d^{(i-i')}$ の分散 $Var(d^{(i-i')})$ は、平均値 $\bar{y}^{(i)}$, $\bar{y}^{(i')}$, $(\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(共通)}$ で表すことができ、これらは互いに独立で、 $(\bar{x}^{(i)} - \bar{x}^{(i')})$ は誤差を持たない定数であり、 $\hat{\beta}_1^{(共通)}$ の分散は、式 (5.26) に示した $Var(\beta_1) = \sigma^2 / \sum_i (X_i - \bar{X})^2$ を用いて

$$\left. \begin{aligned} Var(d^{(i-i')}) &= Var\left[\bar{y}^{(i)} - \bar{y}^{(i')} - (\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(共通)}\right] \\ &= Var(\bar{y}^{(i)}) + Var(\bar{y}^{(i')}) + Var\left[(\bar{x}^{(i)} - \bar{x}^{(i')})\hat{\beta}_1^{(共通)}\right] \\ &= \frac{\hat{\sigma}^2}{n^{(i)}} + \frac{\hat{\sigma}^2}{n^{(i')}} + (\bar{x}^{(i)} - \bar{x}^{(i')})^2 \frac{\hat{\sigma}^2}{S_{xx}^{(共通)}} \\ &= \left(\frac{1}{n^{(i)}} + \frac{1}{n^{(i')}} + \frac{(\bar{x}^{(i)} - \bar{x}^{(i')})^2}{S_{xx}^{(共通)}}\right) \hat{\sigma}^2 \end{aligned} \right\} \quad (6.14)$$

となる。ここで、誤差分散の推定値 $\hat{\sigma}^2$ は、表 6.6 に示した交互作用をプーリングした残差の平均平方の 15.6836 を用いる。表 6.9 に示すように、季節 A_4 の切片 $\hat{\beta}_0^{(4)}$ 基準とした季節 A_i の切片 $\hat{\beta}_0^{(i)}$ の差について式 (6.14) を用いて差の分散を計算し、次いで、95%信頼区間を求める。

表 6.9 季節 A_4 を基準とする水準間の差の 95%信頼区間

季節	切片	平均	差 ($\beta_0^{(i)}$)	$Var(\beta_0^{(i)})$	幅	差の95%CL			
A_i	n_i	\bar{x}_i	$-\beta_0^{(i')}$	$-\beta_0^{(i')}$	$t_{0.05} \times SE$	L95% U95%			
規準	A_4	10	42.75	14.40	—	—			
	A_1	11	30.58	26.64	-12.17	5.9939	4.9606	-17.13	-7.21
	A_2	12	31.82	24.83	-10.94	5.0562	4.5561	-15.49	-6.38
	A_3	9	32.55	18.56	-10.20	3.6570	3.8747	-14.08	-6.33
	A_4	10	42.75	14.40	0.00				
	$\sum_i S_{xx}^{(i)} = S_{xx}^{(共通)} =$		782.83				$\hat{\sigma}^2 =$	15.6836	
		$df =$	37				$t(0.05, 37) =$	2.0262	

次節で示すパラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた表 6.19 の結果に一致する。

Excel の線グラフを用い図 6.4 季節 A_4 を基準とする水準間の差の 95%信頼区間を重ね書きする。図から明らかなように、他の季節との間に、有意な差があることが示されている。

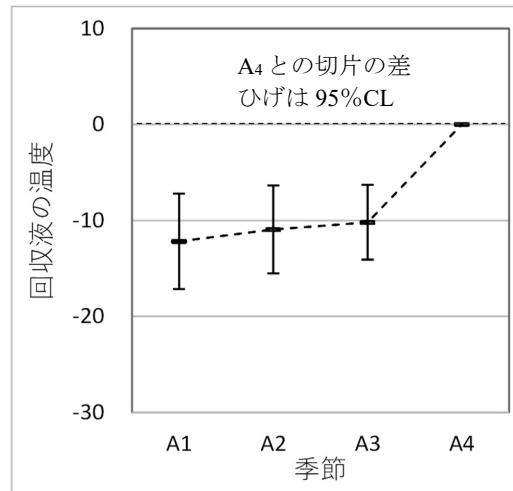


図 6.4 季節 A₄ を基準とする水準間の差の 95%信頼区間

伝統的な共分散分析で定式化されているのは、ここまでである。更なる結果の吟味をするために、推定された回帰直線の 95%信頼区間を示し、季節 A ごとの洗浄水の温度 x が共通となる 20 度における回収液の推定温度 \hat{y}_i での 95%信頼区間付きの折れ線グラフも欲しいが、どうしたら良いのであろうか。これについては、次節の図 6.5 を参照のこと。

共分散分析は難解なのか？

奥野ら（1981）に示されている共分散分析の手順に忠実に Excel の力も借りて再現したのであるが、なんとも技巧的であり気が遠くなりそうである。私も実務で共分散分析を使ってきたのであるが、ダミー変数を用いた線形モデルによる共分散分析を主体にしてきた。伝統的な解析手順については、面倒なことをしていると流し読みしていた。あらためて、共分散分析について記載されている成書を紐解いてみると、伝統的な解析法のオンパレードである。他方、Web 上での解説記事では、ダミー変数を用いた回帰分析を用いた解析方法がメジャーであるが、表層的な説明で終わっている。

この事例は、「層別因子を含む回帰分析」であって、いわゆる共変量を含む「共分散分析」の事例ではない。だが、質的因子と量変数を含む解析は、短絡的に「共分散分析」の事例とみなされがちで、「交互作用のないことを確認し、平行な直線をあてはめる」ような紋きり型の解析に絡めとられるが常である。交互作用があると、式 (6.14) に示したような「水準間の差の 95%信頼区間」が適用できなくなってしまうために、交互作用がないことが「共分散分析」の適用の絶対条件であるかの風評が蔓延している。第 7 章では、本物の“共変量”と質的変数との間で交互作用が存在する事例を示す。

6.2. Excel の回帰分析を活用した共分散分析

第 6.1 節で示した回帰分析を重層的に用いる共分散分析の解析法は、手計算時代の先人たちの創意工夫の賜物である。Excel を用いて奥野ら（1981）に示されている計算方法を忠実に再現したのは、デザイン行列 X を用いた共分散分析と対比することにより、統計解析の多面性を浮き彫りにするためである。

(0, 1)型ダミー変数を用いたデザイン行列 X

季節 A は 4 水準なので、表 6.10 に示すように (0, 1) 型ダミー変数 (a_2, a_3, a_4) を与える。洗浄水の温度 x との交互作用のための変数は、(xa_2, xa_3, xa_4) のように積で与える。

表 6.10 (0, 1) 型ダミー変数

A	a_2	a_3	a_4	A	xa_2	xa_3	xa_4
A ₁	0	0	0	A ₁	0	0	0
A ₂	1	0	0	A ₂	x	0	0
A ₃	0	1	0	A ₃	0	x	0
A ₄	0	0	1	A ₄	0	0	x

表 6.11 左に (0, 1) 型ダミー変数を用いた交互作用を含むデザイン行列 X を示す。表 6.11 右には、Excel の回帰分析で得られた（回帰モデル、主効果モデル、交互作用モデル）での分散分析表を示す。

表 6.11 交互作用モデルに対する段階的な Excel の回帰分析の適用

3) 交互作用モデル										1) 回帰モデル (定数に 0 を使用) off					
2) 主効果モデル										自由度	変動	分散	分散比		
1) 回帰モデル										回帰	残差	合計			
No.	A	x_0	x	a_2	a_3	a_4	xa_2	xa_3	xa_4	y					
1	A ₁	1	30	0	0	0	0	0	0	8.0	1	2523.03	2523.03	89.17	
2	A ₁	1	28	0	0	0	0	0	0	10.5	40	1131.79	28.29		
:	:	:	:	:	:	:	:	:	:	:	41	3654.82			
11	A ₁	1	25	0	0	0	0	0	0	14.0	2) 主効果モデル (定数に 0 を使用) off				
12	A ₂	1	34	1	0	0	34	0	0	9.1	自由度	変動	分散	分散比	
:	:	:	:	:	:	:	:	:	:	:	回帰	4	3074.53	768.63	49.01
23	A ₂	1	24	1	0	0	24	0	0	8.9	残差	37	580.29	15.68	
24	A ₃	1	27	0	1	0	0	27	0	15.0	合計	41	3654.82		
:	:	:	:	:	:	:	:	:	:	:	3) 交互作用モデル (定数に 0 を使用) off				
32	A ₃	1	24	0	1	0	0	24	0	13.6	自由度	変動	分散	分散比	
33	A ₄	1	16	0	0	1	0	0	16	23.4	回帰	7	3145.78	449.40	30.02
:	:	:	:	:	:	:	:	:	:	:	残差	34	509.05	14.97	
41	A ₄	1	9	0	0	1	0	0	9	42.5	合計	41	3654.82		
42	A ₄	1	22	0	0	1	0	0	22	25.8					

逐次平方和による分散分析表

表 6.11 右に示した 3 通りの分散分析表は、説明変数を逐次増加させた Excel の回帰分析の分散分析表のみを示した結果である。表 6.12 に示すのは、3 通りの分散分析表の回帰の平方和（変動）、残差の平方和（変動）に含まれている各要因の平方和（ S_x , S_A , $S_{(x \times A)}$, S_e ）の合成方法を示している。回帰分析に含める説明変数を増加すれば、「回帰」の平方和も逐次増加し、その差分から S_A および $S_{(x \times A)}$ を得ることができる。同様に残差の平方和の減少分からも S_A および $S_{(x \times A)}$ が計算できる。

表 6.12 逐次的な回帰分析による平方和の増加

1) 回帰モデル				2) 主効果モデル				3) 交互作用モデル			
要因	df	平方和		要因	df	平方和		要因	df	平方和	
x	1	2523.03	S_x	x	1	3074.53	S_{x+A}	x	1	3145.78	$S_{x+A+(x \times A)}$
A	(3)	1131.79	$S_{A+(x \times A)+e}$	A	3			x \times A	3		
x \times A	(3)			e	37	580.29	$S_{(x \times A)+e}$	e	34		
e	40			T	41	3654.82		T	41	3654.82	
T	41	3654.82									
		$S_x = 2523.03$				$S_{x+A} - S_x = 551.49$	S_A			$S_{x+A+(x \times A)} - S_{x+A} = 71.25$	$S_{x \times A}$

表 6.13 に示すのは、表 6.12 のそれぞれの「回帰」の平方和から

$$\left. \begin{aligned} S_A &= S_{x+A} - S_x = 3074.53 - 2523.03 = 551.49 \\ S_{x \times A} &= S_{x+A+(x \times A)} - S_{x+A} = 3145.78 - 3074.53 = 71.25 \end{aligned} \right\} \quad (6.15)$$

S_A および $S_{(x \times A)}$ を計算し、逐次（タイプ I）平方和による分散分析表にまとめた結果で、表 6.5 に示した分散分析表に一致する。このように変数を逐次増加した Excel の回帰分析により、先人たちの苦勞を全くせずに、共分散分析における交互作用が、あっさりと求められることは、喜ばしいことである。なお、交互作用を残差平方和 S_e にプールした結果は、すでに表 6.6 に示してあるので割愛する。

表 6.13 逐次（タイプ I）平方和による分散分析表（表 6.5 に一致）

要因	df	逐次平方	平均平方	F	p
x	1	2523.03	2523.03	168.52	0.0000
A	3	551.49	183.83	12.28	0.0000
x \times A	3	71.25	23.75	1.59	0.2107
e	34	509.05	14.97	1.00	
全体	41	3654.82			

共通の傾きを持つ場合の各種の推定

前節の手順 7 で、季節 A の A_4 水準を基準として、他の水準との平行な直線間の差については、式 (6.12) に示した切片の差に帰着し、式 (6.14) に示した分散を用いて 95%信頼区間を推定し、統計的な考察を行った。ここでは、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を活用し、各種

の95%信頼区間の算出法を示す。表6.14に主効果モデルにおけるExcelの回帰分析の結果を示し、パラメータの共分散行列 $\Sigma(\hat{\theta})$ の計算結果も合わせて示す。

これまでも繰り返し述べてきたが、パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素が、推定された回帰パラメータ $\hat{\theta}_i$ の分散 $Var(\hat{\theta}_i)$ となり、この平方根が標準誤差 $SE(\hat{\theta}_i)$ となる。傾きの推定値は、 $\hat{\theta}_1 = -0.7050$ の $SE(\hat{\theta}_1) = 0.1415$ は、 $\Sigma(\hat{\theta})$ の2行2列目の平方根 $\sqrt{0.0200} = 0.1415$ となっている。パラメータの共分散行列 $\Sigma(\hat{\theta})$ の1行2列目は、 $\hat{\theta}_0$ と $\hat{\theta}_1$ の共分散 $Cov(\hat{\theta}_0, \hat{\theta}_1) = -0.5336$ であり、各種の線形和 L の分散の推定に用いられる。

表 6.14 主効果モデルに対するパラメータの共分散行列

デザイン行列 X							2) 主効果モデル (定数に0を使用) off						
No.	A	x_0	x	a_2	a_3	a_4	y		自由度	変動	分散	分散比	
1	A ₁	1	30	0	0	0	8.0						
2	A ₁	1	28	0	0	0	10.5	回帰	4	3074.53	768.63	49.01	
:		:	:	:	:	:	:	残差	37	580.29	15.68	$=\sigma^2$	
11	A ₁	1	25	0	0	0	14.0	合計	41	3654.82			
12	A ₂	1	34	1	0	0	9.1						
:		:	:	:	:	:	:		係数	標準誤差	t	P-値	
23	A ₂	1	24	1	0	0	8.9	θ^{\wedge}_0 切片	30.5793	3.9548	7.7323	0.0000	
24	A ₃	1	27	0	1	0	15.0	θ^{\wedge}_1 x	-0.7050	0.1415	-4.9810	0.0000	
:		:	:	:	:	:	:	θ^{\wedge}_2 a ₂	1.2372	1.6727	0.7396	0.4642	
32	A ₃	1	24	0	1	0	13.6	θ^{\wedge}_3 a ₃	1.9695	2.1158	0.9309	0.3580	
33	A ₄	1	16	0	0	1	23.4	θ^{\wedge}_4 a ₄	12.1731	2.4482	4.9722	0.0000	
:		:	:	:	:	:	:						
41	A ₄	1	9	0	0	1	42.5	パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$					
42	A ₄	1	22	0	0	1	25.8	x_0	x	a_2	a_3	a_4	
								θ^{\wedge}_0	15.6401	-0.5336	-2.3880	-5.7381	-7.9557
								θ^{\wedge}_1	-0.5336	0.0200	0.0361	0.1619	0.2451
								θ^{\wedge}_2	-2.3880	0.0361	2.7979	1.7177	1.8678
								θ^{\wedge}_3	-5.7381	0.1619	1.7177	4.4767	3.4068
								θ^{\wedge}_4	-7.9557	0.2451	1.8678	3.4068	5.9939
								=Minverse (Mmult (Transpose (Xの範囲), Xの範囲)) * σ^2					

デザイン行列 X の変数と回帰分析で推定されたパラメータ（係数）を組み合わせ、各種の線形和 $L^{(i)}$ による推定値を定義する。デザイン行列の変数 (x_0, x_1, a_2, a_3) に対し別名として、行ベクトル $\mathbf{l}^{(i)} = (l_0^{(i)}, l_1^{(i)}, l_2^{(i)}, l_3^{(i)}, l_4^{(i)})$ を与える。推定されたパラメータを列ベクトル $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4)^T$ としたときに、線形和 $L^{(i)}$ を

$$L^{(i)} = \sum_{j=0}^4 l_j^{(i)} \hat{\theta}_j = \mathbf{l}^{(i)} \hat{\theta} \tag{6.16}$$

として定義すると、線形和 $L^{(i)}$ の分散 $Var(L^{(i)})$ は、パラメータの共分散行列 $\Sigma(\hat{\theta})$ の j 行 j' 列の要素を $c_{j,j'}$ として

$$Var(L^{(i)}) = \sum_{j=0}^4 \sum_{j'=0}^4 l_j^{(i)} c_{j,j'} l_{j'}^{(i)} = \mathbf{l}^{(i)} \Sigma(\hat{\theta}) \mathbf{l}^{(i)T} \tag{6.17}$$

で与えられ、パラメータの共分散行列 $\Sigma(\hat{\theta})$ に関する $\mathbf{l}^{(i)}$ の2次形式となる。

表 6.14 に示した傾きを共通とする回帰パラメータの切片は、表 6.15 に示すように、それぞれの線形和 $L^{(i)}$ に対応する行ベクトル $\mathbf{l}^{(i)}$ とパラメータの推定値の列ベクトル $\hat{\boldsymbol{\theta}}$ の積 $L^{(i)} = \mathbf{l}^{(i)} \hat{\boldsymbol{\theta}}$ で計算されており、表 6.7 に示した $\hat{\beta}_0^{(i)}$ に一致する。

表 6.15 線形和 $L^{(i)}$ を用いた傾きを共通とする回帰パラメータの切片の推定

A	$L^{(i)}$	l_0	l_1	l_2	l_3	l_4	$\hat{\boldsymbol{\theta}}$	=	切片 $L^{(i)}$	
		x_0	x	a_2	a_3	a_4				
A ₁	$L^{(1)}$	1	0	0	0	0	30.5793		30.58	$\hat{\beta}_0^{(1)}$
A ₂	$L^{(2)}$	1	0	1	0	0	-0.7050		31.82	$\hat{\beta}_0^{(2)}$
A ₃	$L^{(3)}$	1	0	0	1	0	1.2372		32.55	$\hat{\beta}_0^{(3)}$
A ₄	$L^{(4)}$	1	0	0	0	1	1.9695		42.75	$\hat{\beta}_0^{(4)}$
							12.1731			

前節の表 6.9 では、A₄ 水準と他水準の切片の差の分散を求め、95%信頼区間の推定を行ったが、推定された切片の 95%信頼区間の推定は行わなかった。これは、推定された回帰直線の外挿であり、結果の解釈に不必要なためである。

さて、季節 A の水準ごとの意味のある推定値として、洗浄水の温度 x が 20 度場合の回収液の温度 $\hat{y}_{x=20}^{(i)}$ を推定し、その 95%信頼区間を求め、報告書に記載したいとしよう。表 6.15 では、 $x=l_1=0$ としたので、 $x=l_1=20$ とすれば、20 度における線形和の推定値が計算される。表 6.18 に示すように、A₁ 水準の場合であれば、推定値は $\hat{y}_{x=20}^{(5)} = 16.4788$ であり、

表 6.16 A₁ 水準の洗浄水の温度が 20 度場合の回収液の温度の推定

$y^{(5)}$ $x=20$	l_0	l_1	l_2	l_3	l_4	$\hat{\boldsymbol{\theta}}$	=	
	x_0	x	a_2	a_3	a_4			
	1	20	0	0	0	30.5793		16.4788
						-0.7050		
						1.2372		
						1.9695		
						12.1731		

その線形和の分散は、式 (6.17) に示したパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ に関する $\mathbf{l}^{(i)}$ の 2 次形式を用い、表 6.17 に示すように $Var(\hat{y}_{x=20}^{(1)}) = 2.3081$ と推定される。

表 6.17 A₁ 水準の洗浄水の温度が 20 度場合の回収液の温度の分散の推定

l_0	l_1	l_2	l_3	l_4	$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$	\mathbf{l}^T	=	$Var(\hat{y}_{x=20}^{(5)})$	
x_0	x	a_2	a_3	a_4					
1	20	0	0	0	15.64	1		2.3081	L95%= 13.4005
					-0.53	20			U95%= 19.5571
					-2.39	0			
					-5.74	0			
					-7.96	0			$t_{0.05}(42-5) = 2.0262$

95%信頼区間は,

$$\begin{aligned}
 (L95\%, U95\%) &= \hat{y}_{x=20}^{(5)} \pm t(0.05, 37) \sqrt{\text{Var}(\hat{y}_{x=20}^{(1)})} \\
 &= 16.4788 \pm 2.0262 \times \sqrt{2.3081} \\
 &= (13.4005, 19.5571)
 \end{aligned}
 \tag{6.18}$$

として求められる. 表 6.16 に示した計算式により, 他の季節における推定結果を表 6.18 に示す. なお, 計算式の設定を A₁ 水準で行い, セルのフィルハンドルを用いて他の季節の計算式をコピーして求めている.

表 6.18 季節別の洗浄水の温度が 20 度の場合の回収液の濃度の推定

A	$L^{(i)}$	l_0 x_0	l_1 x	l_2 a_2	l_3 a_3	l_4 a_4	推定値 $L^{(i)}$	分散 $\text{Var}(L^{(i)})$	幅 $t_{0.05} \times SE$	95%信頼区間 $L95\%$ $U95\%$	
A ₁	$L^{(5)}$	1	20	0	0	0	16.48	2.3081	3.08	13.40	19.56
A ₂	$L^{(6)}$	1	20	1	0	0	17.72	1.7750	2.70	15.02	20.42
A ₃	$L^{(7)}$	1	20	0	1	0	18.45	1.7844	2.71	15.74	21.15
A ₄	$L^{(8)}$	1	20	0	0	1	28.65	2.1966	3.00	25.65	31.65
$t(0.05, 42-5) =$									2.0262		

この結果を Excel の「折れ線」グラフで作成した結果を図 6.5 に示す. Excel での信頼区間の幅を付ける際には, 推定値からの幅が必要となるので $t_{0.05}(df) \times SE$ によりひげの長さを別途計算する必要がある.

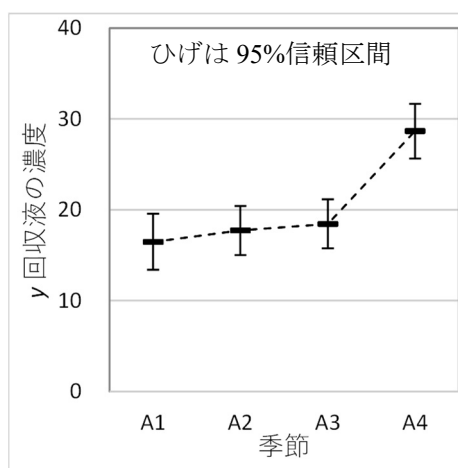


図 6.5 洗浄水の温度が 20 度の場合の予測プロフィール

Excel での「折れ線グラフ」のひげの長さの設定は, かなり深いところにあり,

グラフの要素 → 誤差範囲 → その他のオプション → ユーザの設定 →
 値の設定 → 正の誤差の値 (範囲) → 負の誤差の値 (範囲)

のように設定する. 少しでも気を抜くとまがい物のひげが付くので細心の注意が必要である.

水準間の差の予測プロファイル

回収液の濃度が最も高い季節 A_4 を基準とし、温度が 20 度の場合の他の季節との差の推定値と 95%信頼区間を求めたい。季節 A_3 と A_4 の差の推定値は、デザイン行列の変数 $(x_0, x_1, a_2, a_2, a_3)$ に対し別名として与えた行ベクトル $\mathbf{l}^{(i)} = (l_0^{(i)}, l_1^{(i)}, l_2^{(i)}, l_3^{(i)}, l_4^{(i)})$ 間の差を計算する。推定値は、

$$\left. \begin{aligned} \hat{y}_{x=20}^{(A_3-A_4)} &= \hat{y}_{x=20}^{A_3} - \hat{y}_{x=20}^{A_4} \\ &= 18.45 - 28.65 = -10.20 \end{aligned} \right\} \quad (6.19)$$

として求められる。この差に対する分散の推定のためのベクトル $\mathbf{d}^{(A_3-A_4)}$ は、季節 A_3 のベクトル $\mathbf{l}^{(A_3)}$ と季節 A_4 のベクトル $\mathbf{l}^{(A_4)}$ の差

$$\left. \begin{aligned} \mathbf{l}^{(A_3)} &= [1 \quad 20 \quad 0 \quad 1 \quad 0] \\ -) \mathbf{l}^{(A_4)} &= [1 \quad 20 \quad 0 \quad 0 \quad 1] \\ \hline \mathbf{d}^{(A_3-A_4)} &= [0 \quad 0 \quad 0 \quad 1 \quad -1] \end{aligned} \right\} \quad (6.20)$$

であり、 $\hat{y}^{(A_3-A_4)}$ の分散 $Var[\hat{y}^{(A_3-A_4)}]$ は、

$$\begin{aligned} Var[\hat{y}^{(A_3-A_4)}] &= \mathbf{d}^{(A_3-A_4)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) [\mathbf{d}^{(A_3-A_4)}]^T \\ &= 3.6570 \end{aligned} \quad (6.21)$$

として求められ、差 -10.20 の 95%信頼区間が $(-18.10, -2.31)$ と推定されゼロを含まないので、有意な差であることが示されている。

同様に季節 A_4 を基準とし、他の季節の差を表 6.19 に示す。差のデザイン行列については、季節 A_3 と季節 A_4 の場合と同様にする。差の分散についてもそれぞれの差のデザイン行列の場合と同様に推定し、95%信頼区間を計算する。この結果は、前節の表 6.9 に示した結果に一致し、また、図 6.4 に示した水準間の差の 95%信頼区間の折れ線グラフが作成することができる。

表 6.19 季節 A_4 を基準とした 2 水準間の差の 95%信頼区間

A	$L^{(i)}$	l_0 x_0	l_1 x	l_2 a_2	l_3 a_3	l_4 a_4	推定値 $L^{(i)}$	分散 $Var(L^{(i)})$	幅 $t_{005} \times SE$	95%信頼区間 $L_{95\%}$ $U_{95\%}$	
A_1-A_4	$L^{(9)}$	0	0	0	0	-1	-12.17	5.9939	4.96	-17.13	-7.21
A_2-A_4	$L^{(10)}$	0	0	1	0	-1	-10.94	5.0562	4.56	-15.49	-6.38
A_3-A_4	$L^{(11)}$	0	0	0	1	-1	-10.20	3.6570	3.87	-14.08	-6.33
A_4-A_4	$L^{(12)}$	0	0	0	0	0	0.00	0.0000	0.00	0.00	0.00
							$t(0.05, 42-5) =$		2.0262		

線形モデルにおける各種の推定は、線形和 $L^{(i)}$ に対する式 (6.16)、その分散 $Var(L^{(i)})$ は、式 (6.17) として定式化されていることが、汎用的であり伝統的な方法に対する優位性でもある。

洗浄水の温度 x に関する予測プロファイル

季節 A_3 と季節 A_4 について洗浄水の温度を変化させた場合の予測プロファイルを作成する。表 6.20 に示すように、季節 A_3 の洗浄水の温度 x を 5 度から 35 度まで変化させる。季節 A_3 も同様に洗浄水の温度 x を 1 度から 31 度まで変化させた行ベクトル $L^{(i)}$ を生成する。これらのベクトルに対し、パラメータの推定値 $\hat{\theta}$ およびパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いて予測値と分散を推定し、95%信頼区間を求める。

表 6.20 洗浄水の温度 x を変えた場合の回収液の濃度の 95%信頼区間

A	$L^{(i)}$	l_0 x_0	l_1 x	l_2 a_2	l_3 a_3	l_4 a_4	推定値 $L^{(i)}$	分散 $Var(L^{(i)})$	幅 $t_{0.05} \times SE$	95%信頼区間 $L95\%$ $U95\%$	
A_3	$L^{(13)}$	1	5	0	1	0	29.02	5.42	4.72	24.30	33.74
	$L^{(14)}$	1	10	0	1	0	25.50	3.21	3.63	21.87	29.13
	$L^{(15)}$	1	20	0	1	0	18.45	1.78	2.71	15.74	21.15
	$L^{(16)}$	1	30	0	1	0	11.40	4.37	4.23	7.16	15.63
	$L^{(17)}$	1	35	0	1	0	7.87	7.16	5.42	2.45	13.29
A_4	$L^{(18)}$	1	1	0	0	1	42.05	5.17	4.61	37.44	46.65
	$L^{(19)}$	1	10	0	0	1	35.70	1.96	2.83	32.87	38.54
	$L^{(20)}$	1	20	0	0	1	28.65	2.20	3.00	25.65	31.65
	$L^{(21)}$	1	30	0	0	1	21.60	6.44	5.14	16.46	26.75
	$L^{(22)}$	1	31	0	0	1	20.90	7.09	5.39	15.50	26.29
								$t(0.05,42-5)=$	2.0262		

図 6.6 に示す予測プロファイルは、表 6.20 で計算された季節 A_4 と季節 A_3 の予測値と 95% 信頼区間を散布図に重ね書きした結果である。季節 A_4 の洗浄水の温度 x が 10 度の場合、回収液の濃度 \hat{y} は、35.70 と季節 A_3 の 25.50 と 10.20 の差がある。95%信頼区間が互いに重なっていないので統計的な差があると判断される。このように、推定された平行な回帰直線に対し、95%信頼区間を計算し重ね書きして示せることは、一連の定式化された線形モデルの解析手順の賜物であるし、Excel の行列計算の柔軟性のおかげでもある。

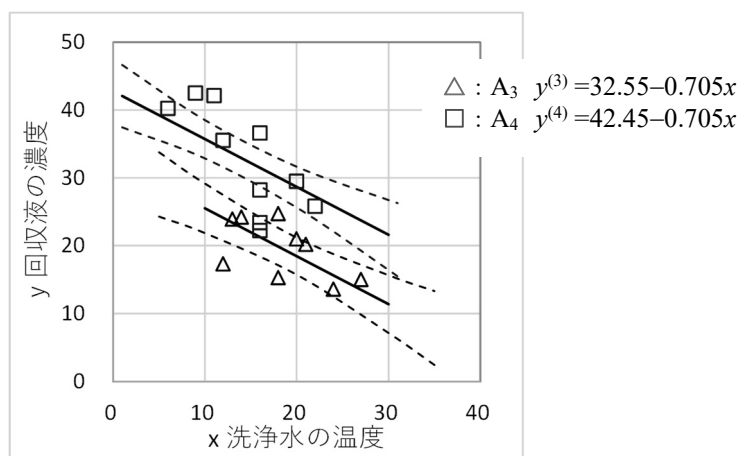


図 6.6 洗浄水の温度 x に対する季節 A_4 および季節 A_3 の予測プロファイ

6.3. SAS の GLM プロシジャによる共分散分析

SAS でのテキストデータの取り込み

無償で継続的に提供されている OnDemand SAS の GLM プロシジャを用いて、これまでに示した解析法と対応付けた解析方法を示す。GLM プロシジャは、共変量 x 、質的変数 A 、それらの交互作用を扱うことができる。まず、GLM プロシジャを使うための SAS ファイルの作成から始める。SAS は、様々な形式の外部ファイルを読み込めるが、特徴的なのは、テキストデータを読み込むための種々のプログラミング機能を持っているので少量のデータの入力に対し、柔軟な対応が可能である。

表 6.1 に示した Excel で作成した表形式のデータを、テキストエディタ（メモ帳）にコピー & ペーストし、tab 記号をスペースに変換した結果を SAS プログラムの中に取り込む。SAS の DATA ステップで、do ステートメントなどを使い、各行に 4 セット分の季節 A に対応する共変量 x 、反応 y を読み込み、行方向に 42 行を SAS データファイル「d01」に出力する。元のデータの並びのテキストデータがどのような形であれ、解析用のプログラミング機能を用いて解析に必要とされる SAS データファイルを作成することができる。

```
SAS プログラム 6.1      /* SAS データファイルの作成 */
Title1 'ANCOVA_奥野.sas 2022-6-13 Y.Takahashi ' ;
data d01 ;
  input No @ ;
  do A = "A1", "A2", "A3", "A4" ;
    input x y @;
    if (x ne .) then output ;
  end ;
/*      A1          A2          A3          A4
No     x      y      x      y      x      y      x      y */ ;
datalines ;
1  30  8.0   34  9.1   27 15.0   16 23.4
2  28 10.5   19 19.4   20 21.0   16 28.2
3  30  8.2   22 20.5   14 24.2   20 29.5
4  29 13.0   25 14.2   18 15.3   16 22.2
5  28 10.1   22 11.0   12 17.3    6 40.2
6  17 16.2   25 19.1   18 24.7   16 36.6
7  24 16.1   22 16.0   13 23.9   12 35.5
8  22 13.4   23 11.1   21 20.2   11 42.1
9  30 13.0   28 12.0   24 13.6    9 42.5
10 30  7.3   24 17.1    .   .      22 25.8
11 25 14.0   30 13.3    .   .      .   .
12 .   .      24  8.9    .   .      .   .
;
proc print data=d01 ; run ;
```

GLM プロシジャによる交互作用 $x \times A$ の評価

GLM プロシジャを用いて交互作用 $x \times A$ の評価を行なう。季節 A は、文字データであるので、GLM プロシジャでは、class ステートメントで質的変数であることを明示し、model ステートメントで `model y = x A x*A` のようにモデル式を与え、交互作用は「*」で設定する。「/」以後がオプションで、solution オプションは、パラメータの推定値の出力要求であり、SS1 SS2 SS3 は、タイプ I, II, III の平方和の出力要求である。

SAS プログラム 6.2 /* GLM プロシジャによる交互作用の解析 */

```
proc glm data=d01 ;
  class A ;
  model y = x A x*A / solution SS1 SS2 SS3 ;
run ;
```

OnDemand SAS のデフォルトの出力は、HTML 形式なので、出力された分散分析表コピーし、Excel にペーストし、若干の整形後にワードにペーストした結果を表 6.21 に示す。この結果は、表 6.11 右下段の Excel の回帰分析の分散分析表に一致する。

表 6.21 GLM プロシジャによる包括的な分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	7	3145.7754	449.3965	30.02	<.0001
Error	34	509.0487	14.9720		
Corrected Total	41	3654.8240			

表 6.22 に示すのは、SAS の GLM プロシジャで出力される 3 種の平方和を Excel 上で編集した結果である。交互作用 $x \times A$ の平方和は、3 種の平方和が共に $S_{xxA} = 71.2462$ であり、交互作用が統計的に無いとみなせる。主効果 x および季節 A のタイプ I の平方和は、Excel での回帰分析から導出した表 6.13 の逐次 (タイプ I) の平方和に一致する。なお、3 種の平方和についての詳細は、第 3.3 節を参照のこと。

表 6.22 交互作用モデルにおけるタイプ I, II, III の平方和の比較

要因	自由度	Type I 平方和		Type II 平方和		Type III 平方和	F 値	Pr > F
x	1	2523.0347	≠	389.1143	≠	369.0901	-	-
A	3	551.4944	≠	246.2924	=	246.2924	-	-
x*A	3	71.2462	=	71.2462	=	71.2462	1.59	0.2107

GLM プロシジャが内部で生成するデザイン行列 X

分散分析表の出力結果だけで満足するならば、SAS が内部で生成しているデザイン行列の知識は必要ない。表 6.23 に示すように solution オプションで出力したパラメータの推定値から 4 本の回帰直線のパラメータ推定を行いたいが、どのようにしたら良いのだろうか。

表 6.23 GLM プロシジャによるパラメータの推定値

パラメータ	推定値		標準誤差	t 値	Pr > t
Intercept	49.7416	B	3.9804	12.5	<.0001
x	-1.1904	B	0.2630	-4.53	<.0001
A A1	-22.7247	B	8.7579	-2.59	0.0139
A A2	-22.7124	B	8.2075	-2.77	0.0091
A A3	-21.3495	B	6.5383	-3.27	0.0025
A A4	0	B	.	.	.
x*A A1	0.6191	B	0.3912	1.58	0.1228
x*A A2	0.6781	B	0.3882	1.75	0.0897
x*A A3	0.7094	B	0.3775	1.88	0.0688
x*A A4	0	B	.	.	.

注) 「B」は、一意的な推定値ではないとの警告である。

(1, 0) 型ダミー変数を用いて Excel の回帰分析で再現できる。

表 6.24 に示すように GLM プロシジャの内部のデザイン行列 X は、表 6.23 に示すパラメータの出力から、(1, 1) 標示型ダミー変数としていることがうかがわれる。ダミー変数間が一次従属となり、このままでは解析できない。そのために GLM プロシジャでは、(A A4) および (x*A A4) を解析に際し無視してパラメータの推定を行っている。そのために、パラメータの推定は、最後の水準を基準とした (1, 0) 型ダミー変数と同じ結果となっている。

表 6.24 SAS が内部で生成しているデザイン行列

SAS元データ														
No.	A	x	y	切片	x	A	A	A	A	x*A	x*A	x*A	x*A	y
						A1	A2	A3	A4	A1	A2	A3	A4	
1	A1	30	8.0	1	30	1	0	0	0	30	0	0	0	8.0
2	A1	28	10.5	1	28	1	0	0	0	28	0	0	0	10.5
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
11	A1	25	14.0	1	25	1	0	0	0	25	0	0	0	14.0
12	A2	34	9.1	1	34	0	1	0	0	0	34	0	0	9.1
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
23	A2	24	8.9	1	24	0	1	0	0	0	24	0	0	8.9
24	A3	27	15.0	1	27	0	0	1	0	0	0	27	0	15.0
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
32	A3	24	13.6	1	24	0	0	1	0	0	0	24	0	13.6
33	A4	16	23.4	1	16	0	0	0	1	0	0	0	16	23.4
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
41	A4	9	42.5	1	9	0	0	0	1	0	0	0	9	42.5
42	A4	22	25.8	1	22	0	0	0	1	0	0	0	22	25.8

SAS/GLM プロシジャには、JMP のように内部で生成されたデザイン行列を出力するすべがないので、別途 Excel で作成した結果を示している。

したがって、表 6.25 に示すように切片 $\hat{\theta}_0 = 49.7416$ が A₄ 水準の切片となり、 $\hat{\theta}_1 = -1.1904$ が傾きとなる。A₁ 水準の切片 $\hat{\beta}_0^{(1)}$ は、 $\hat{\theta}_0 = 49.7416$ に $\hat{\theta}_2 = -22.7247$ を加えて

$$\hat{\beta}_0^{(1)} = \hat{\theta}_0 + \hat{\theta}_2 = 49.7416 - 22.7247 = 27.0169 \quad (6.22)$$

表 6.25 GLM プロシジャの推定値から4本の回帰直線の切片と傾きの推定

タ	推定値				切片	
Intercept	49.7416	$=\theta^{\wedge}_0$		A ₁	$\beta^{\wedge}_0^{(1)}=$	27.0169 $=\theta^{\wedge}_0+\theta^{\wedge}_2$
x	-1.1904	$=\theta^{\wedge}_1$		A ₂	$\beta^{\wedge}_0^{(2)}=$	27.0292 $=\theta^{\wedge}_0+\theta^{\wedge}_3$
A A1	-22.7247	$=\theta^{\wedge}_2$		A ₃	$\beta^{\wedge}_0^{(3)}=$	28.3921 $=\theta^{\wedge}_0+\theta^{\wedge}_4$
A A2	-22.7124	$=\theta^{\wedge}_3$		A ₄	$\beta^{\wedge}_0^{(4)}=$	49.7416 $=\theta^{\wedge}_0$
A A3	-21.3495	$=\theta^{\wedge}_4$				
A A4	0					傾き
x*A A1	0.6191	$=\theta^{\wedge}_5$		A ₁	$\beta^{\wedge}_1^{(1)}=$	-0.5713 $=\theta^{\wedge}_1+\theta^{\wedge}_5$
x*A A2	0.6781	$=\theta^{\wedge}_6$		A ₂	$\beta^{\wedge}_1^{(2)}=$	-0.5123 $=\theta^{\wedge}_1+\theta^{\wedge}_6$
x*A A3	0.7094	$=\theta^{\wedge}_7$		A ₃	$\beta^{\wedge}_1^{(3)}=$	-0.4810 $=\theta^{\wedge}_1+\theta^{\wedge}_7$
x*A A4	0			A ₄	$\beta^{\wedge}_1^{(4)}=$	-1.1904 $=\theta^{\wedge}_1$

GLM プロシジャの出力を Excel に取り込み、回帰直線のパラメータを計算した結果である。

として求められる。A₁水準の傾き $\hat{\beta}_1^{(1)}$ は、 $\hat{\theta}_1 = -1.1904$ に $\hat{\theta}_5 = 0.6191$ を加えて

$$\hat{\beta}_1^{(1)} = \hat{\theta}_1 + \hat{\theta}_5 = -1.1904 + 0.6191 = -0.5713 \quad (6.23)$$

として求められる。

なお、表 6.25 に示した回帰パラメータの推定値は、表 6.2 に示した別々に推定した4本の回帰直線の推定値に一致する。

表 6.26 季節ごとの回帰パラメータ推定 (表 6.2 抜粋)

	A ₁	A ₂	A ₃	A ₄	和
n_i	11	12	9	10	
$\bar{x}_{i\cdot}$	26.64	24.83	18.56	14.40	—
$\bar{y}_{i\cdot}$	11.80	14.31	19.47	32.60	—
$S_{xx}^{(i)}$	178.55	183.67	204.22	216.40	782.83
$S_{xy}^{(i)}$	-102.00	-94.08	-98.23	-257.60	-551.92
$\beta^{\wedge}_1^{(i)}$	-0.57	-0.51	-0.48	-1.19	—
$\beta^{\wedge}_0^{(i)}$	27.02	27.03	28.39	49.74	—
$\beta^{\wedge}_1^{(i)} = S_{xy}^{(i)} / S_{xx}^{(i)}$	$\beta^{\wedge}_0^{(i)} = \bar{y}_{i\cdot} - \beta^{\wedge}_1^{(i)} \bar{x}_{i\cdot}$				

散布図上に4本の回帰直線を重ね書きした結果が、図 6.7 に示すようにデフォルトで出力される。この結果に、何らかの方法で、手を加えてブラッシュアップしようとしたが、まったく手が加えられなかった。結果の信頼性の確保の観点からは、望ましいのであるが、残念なので、表 6.25 の回帰パラメータの推定値をワードで重ね書きしてある。

交互作用を含まない解析

交互作用を含まない解析は、model y = x A のように x*A を落とせばよい。表 6.27 に示すように、包括的な分散分析表が得られる。Excel の回帰分析による表 6.14 の分散分析表に一致する。

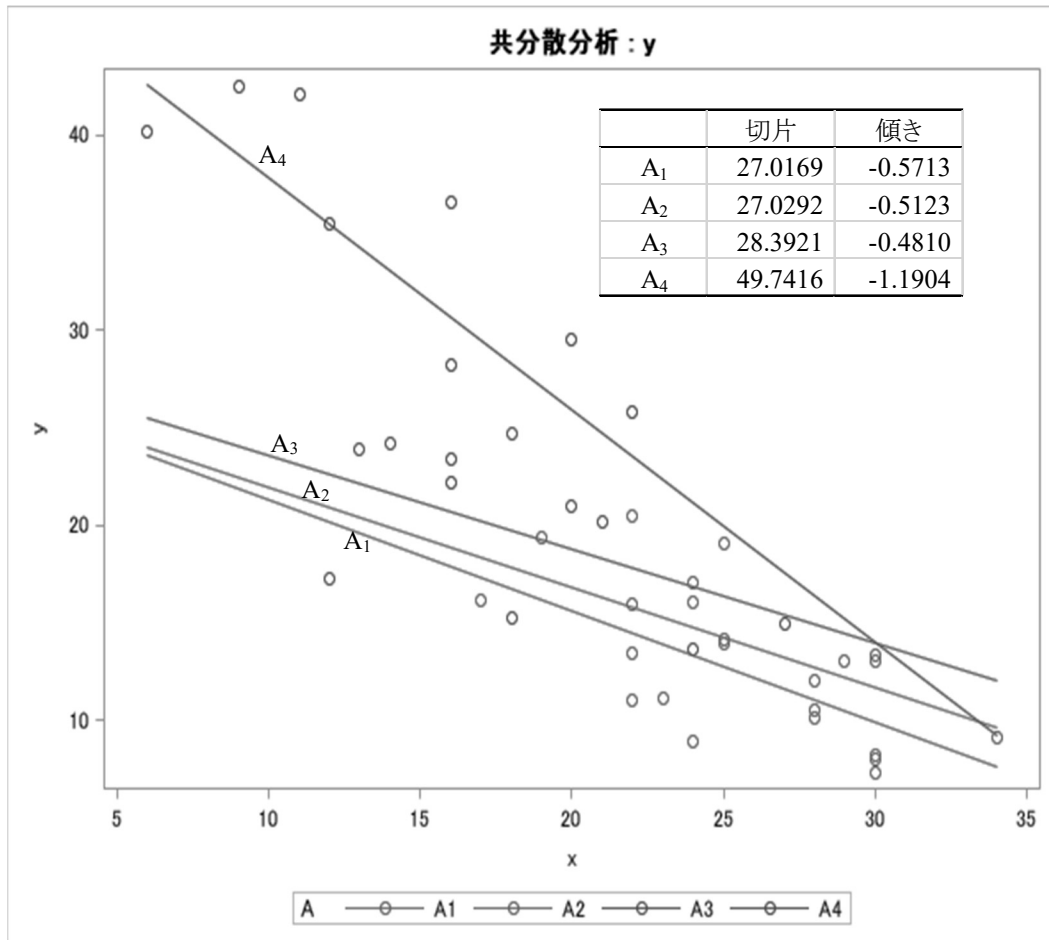


図 6.7 GLM プロシジャでデフォルト出力される散布図上の 4 本の回帰直線

SAS プログラム 6.3 /* GLM プロシジャによる主効果モデル */

```
proc glm data=d01 ;
  class A ;
  model y = x A / solution SS1 SS2 SS3 ;
run ;
```

表 6.27 GLM プロシジャによる包括的な分散分析表

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	4	3074.5291	768.63229	49.01	<.0001
Error	37	580.2949	15.6836		
Corrected Total	41	3654.8240			

表 6.28 に示すように、交互作用を含まない主効果モデルでは、タイプ II, III の平方和が一致することが確認される。もちろんタイプ I の平方和は、季節 A については、他と同じ $S_A = 551.4944$ であるが、共変量 x については $S_x^{(I)} = 2523.0347$ とタイプ I の平方和のみが異

なる。主効果 x および A のタイプ I の平方和は、Excel での回帰分析から導出した表 6.12 の逐次 (タイプ I) の平方和に一致する。季節 A は、統計的に明らかな差があるので、これまでと同様に各種の推定値による比較検討を行なう。

表 6.28 主効果モデルにおけるタイプ I, II, III の平方和の比較

要因	自由度	Type I 平方和	Type II 平方和	Type III 平方和	F 値	Pr > F
x	1	2523.0347	\neq 389.1143	$=$ 389.1143	-	-
A	3	551.4944	$=$ 551.4944	$=$ 551.4944	11.72	<.0001

表 6.29 に、solution オプションで出力したパラメータ推定値から、平行な 4 本の回帰直線のパラメータを推定する。 A_1 の切片は、

$$\hat{\beta}_0^{(1)} = \hat{\theta}_0 + \hat{\theta}_2 = 42.7523 - 12.1731 = 30.5793 \quad (6.24)$$

であり、傾きは、 $\hat{\beta}_1^{(1)} = \hat{\theta}_1 = -0.7050$ である。この結果は、Excel の回帰分析の結果から推定した表 6.15 に一致する。

表 6.29 GLM プロシジャのパラメータ推定値から 4 本の回帰直線のパラメータ推定

パラメータ	推定値								
Intercept	42.7523	$=\theta^0_0$							
x	-0.7050	$=\theta^0_1$			切片			傾き	
A A1	-12.1731	$=\theta^0_2$	A_1	$\hat{\beta}^{(1)}_0 =$	30.5793	$=\theta^0_0 + \theta^0_2$	$\hat{\beta}^{(1)}_1 =$	-0.7050	$=\theta^0_1$
A A2	-10.9359	$=\theta^0_3$	A_2	$\hat{\beta}^{(2)}_0 =$	31.8164	$=\theta^0_0 + \theta^0_3$	$\hat{\beta}^{(2)}_1 =$	-0.7050	$=\theta^0_1$
A A3	-10.2036	$=\theta^0_4$	A_3	$\hat{\beta}^{(3)}_0 =$	32.5488	$=\theta^0_0 + \theta^0_4$	$\hat{\beta}^{(3)}_1 =$	-0.7050	$=\theta^0_1$
A A4	0		A_4	$\hat{\beta}^{(4)}_0 =$	42.7523	$=\theta^0_0$	$\hat{\beta}^{(4)}_1 =$	-0.7050	$=\theta^0_1$

散布図上に 4 本の回帰直線を重ね書きした結果が、図 6.8 に示すようにデフォルトで出力される。

Lsmeans ステートメントによる最小 2 乗平均の推定

SAS の GLM プロシジャは、推定値としての最小 2 乗平均を最初に提供した統計ソフトである。最小 2 乗平均は、GLM プロシジャの lsmeans ステートメントを用いて推定することができる。最小 2 乗平均は、デフォルトでは、 x の総平均 $\bar{x}.. = 21.4762$ に対する推定値が用いられるが、 x の値を自由に設定して最小 2 乗平均を求めることができる。そこで、at オプションにより、表 6.18 で設定したと同様の $x=20$ を設定する。

SAS プログラム 6.4 /* GLM プロシジャによる最小 2 乗平均の推定 */

```
proc glm data=d01 ;
  class A ;
  model y = x A / ;
  lsmeans A / at x=20 stderr cl ;
run ;
```

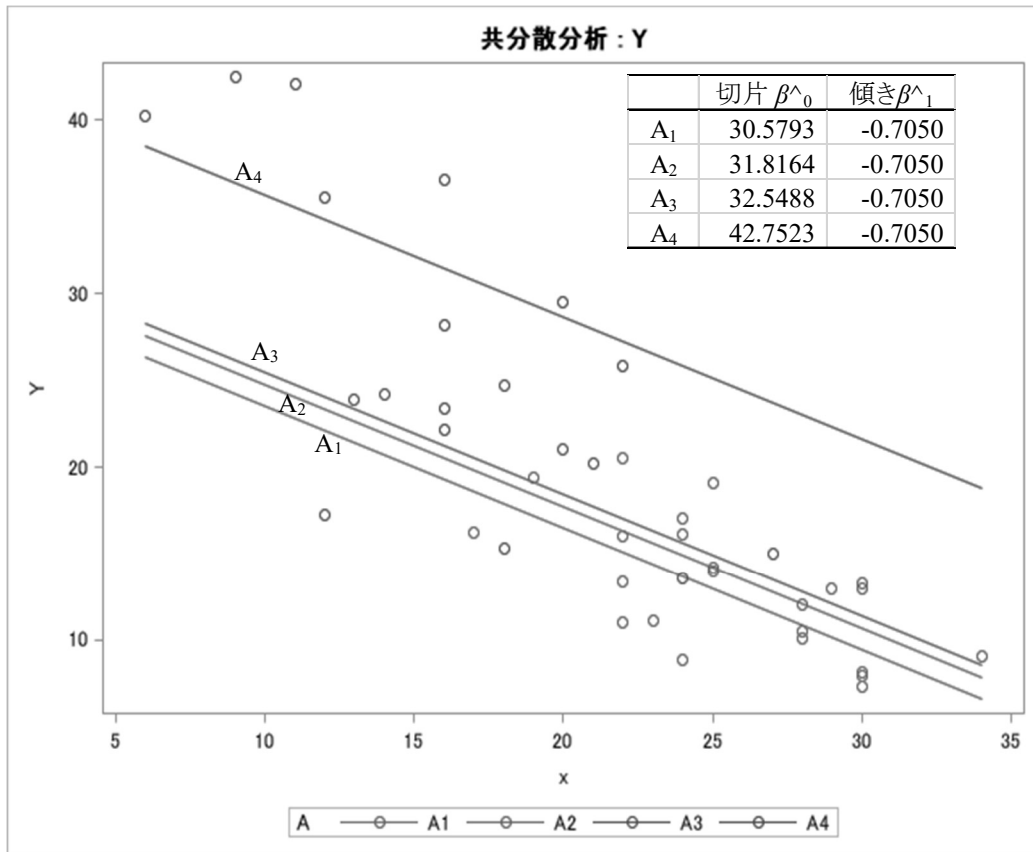



図 6.8 GLM プロシジャによるデフォルトの散布図上の平行な 4 本の回帰直線

表 6.30 に GLM プロシジャによって計算された $x=20$ における最小 2 乗平均と **stderr** および **cl** オプションによって出力された標準誤差と 95%信頼区間が示されている。この結果は、Excel で作成した表 6.18 に一致する。SAS でこの結果を線グラフで作成するのは難儀であるので、Excel で作成した図 6.5 を図 6.9 に再掲する。

表 6.30 GLM プロシジャによる最小 2 乗平均

最小 2 乗平均 : $x=20$				
A	最小 2 乗平均	標準誤差	95% 信頼限界	
A1	16.4788	1.5193	13.4005	19.5571
A2	17.7159	1.3323	15.0165	20.4154
A3	18.4483	1.3358	15.7417	21.1549
A4	28.6519	1.4821	25.6488	31.6549

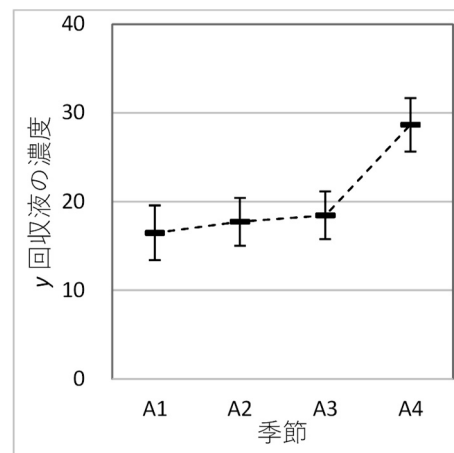


図 6.9 最小 2 乗平均プロット (図 6.5 再掲)

SAS ユーザにとって GLM プロシジャによって計算された最小 2 乗平均は、極めて有用な統計量であり、結果の提示に際し何の断りもなく LSMEANS と称して使われているのが現状である。ただし、SAS・JMP ユーザ以外にとっては、未知の統計量であることを認識し、どのような計算によって推定されたかを Excel などによって再現できることが、他者からの質疑に答える基礎知識を与える。たとえば、「線形モデルで推定されたパラメータを用いた線形和」と一般論を示し、「共分散分析では、推定された平行な直線上の推定値」などと続け、表 6.30 のような推定値を示し、表 6.18 に示した線形和について説明することで理解が得れる。

Estimate ステートメントによる最小 2 乗平均の差の推定

Lsmeans ステートメントを用いた最小 2 乗平均について示したきたのであるが、ある基準となる水準、たとえば A₄ 水準からの他の水準の差および標準誤差、95%信頼区間を推定したい。Lsmeans ステートメントでは、pdiff オプションを使って総当り式に水準間の差の推定値が得られるが、出力が冗長なので、estimate ステートメントを用いて、水準間の差の係数ベクトルを与える。

```
SAS プログラム 6.5 /* estimate ステートメントを用いた最小 2 乗平均の差の推定 */
proc glm data=d01 ;
  class A ;
  model y = x A ;
  estimate 'A1-A4 full' intercept 0 x 0.0 A 1 0 0 -1 ;
  estimate 'A1-A4' A 1 0 0 -1 ;
  estimate 'A2-A4' A 0 1 0 -1 ;
  estimate 'A3-A4' A 0 0 1 -1 ;
  estimate 'A4-A4' A 0 0 0 0 ;
run ;
```

Estimate ステートメントで、水準間の差の推定に際し、0 となる係数 intercept 0 あるいは x 0.0 などは、省略することができる。表 6.31 に示すように、Excel による追加の計算によって、表 6.19 に完全に一致することが確かめられる。

表 6.31 Estimate ステートメントによる最小 2 乗平均の計算

パラメータ	推定値	標準誤差	t 値	Pr > t	L95%	U95%
A1-A4 full	-12.1731	2.4482	-4.97	<.0001	-17.1337	-7.2125
A1-A4	-12.1731	2.4482	-4.97	<.0001	-17.1337	-7.2125
A2-A4	-10.9359	2.2486	-4.86	<.0001	-15.4920	-6.3798
A3-A4	-10.2036	1.9123	-5.34	<.0001	-14.0783	-6.3288
A4-A4	0	0	.	.	0.0000	0.0000
				$t_{0.05}(42-5)=$	2.0262	

パラメータの推定値、95%信頼区間が得られたならば、その結果をグラフ化する必要がある。統計ソフトで対応できないならば、何らかの手段で作成する努力を怠ってはならない。

7. 共変量を含む 3 因子実験データの探索的解析

これまで取り上げてきたほとんど事例は、1 つの質的変数と 1 つの量的変数の場合であった。現実には、複数の質的変数と複数の量的変数を同時に取り上げ、一気に問題解決を図りたい。多くの実験研究あるいは観察研究の場で、対象分野の特質により交互作用よりも主効果に重きを置く分野、常に交互作用を念頭に置く分野があり、実験計画の考え方が異なり、解析の考え方も異なる。宮川（2008）、「問題発見の解決の科学—SQC の基本—、第 6.3 節 計画的に採取した 3 元配置データの解析事例」で取り上げられている「冷蔵庫の塗装不良の改善」の実験データを取り上げる。原書で実践されているのは、層別散布図に回帰直線を重ね書きした図を主体にしたアプローチである。そこで、Excel による格子状の層別散布図の作成方法を示すと共に、Excel による探索的な回帰分析と結果のグラフ化による問題解決法を例示する。さらに、JMP による交互作用も考慮した変数選択機能を活用した探索的な解決法も合わせて示す。

7.1. Excel の散布図を主体にした問題解決

表 7.1 に示すのは、ある大手の家電メーカーでの現業実習で遭遇した事例であり、冷蔵庫の塗装不良の改善のために、取り上げられた因子は、(A : シンナーの種類, B : 塗装の色, C : 冷蔵庫の機種) であり、塗装の中心膜厚 Y を測定し、共変量としてガンスプレーの吐出量 X が測定された。因子は各 2 水準で全ての組み合わせに対し 10 回の塗装が反復して行われた。それぞれの因子の組合せで 10 回の反復の平均と SD を示してある。中心膜厚の管理限界は、(28 ~ 36 単位) であり、管理限界を超えた場合には不良品となる。シンナー A_1 の場合には、40 個のデータ中 3 件の管理下限を下回る不良品が発生し、シンナー A_2 の場合には、40 個のデータ中 3 件の管理上限を上回る不良品が発生していることが示されている。

共変量がなければ、各セルの 10 個中の不良品の割合に注目し、不良の割合に対する解析で済ませたくなる。シンナー A_1 ・色 B_2 の場合の機種 (国内・国外) に関わらず管理限界の下側を超える不良が発生し、シンナー A_2 ・色 B_1 の場合には逆に機種に関わらず管理限界の上側を

表 7.1 冷蔵庫の塗装データ [宮川 (2008), 表 6.1]

	機種: 国内 C ₁					機種: 国外 C ₂				
	色 B ₁		色 B ₂		注	色 B ₁		色 B ₂		注
シンナー	x	y	x	y		x	y	x	y	
A ₁	91.3	33.5	88.6	29.2		89.0	31.1	90.9	30.7	
	89.3	32.7	88.4	27.6	▼	90.1	30.7	88.0	28.5	
	88.9	31.6	89.7	28.8		90.3	33.4	90.7	29.3	
	90.4	32.6	90.4	29.7		89.6	32.1	92.9	31.2	
	88.5	30.3	91.2	28.8		90.7	32.0	92.6	30.4	
	92.6	33.0	89.3	28.0	▼	92.2	32.2	89.0	29.9	
	91.3	31.4	91.7	31.0		88.3	33.2	89.3	29.2	
	91.5	32.4	92.3	29.8		90.4	31.3	88.7	28.4	
	87.7	31.2	89.7	30.3		92.2	33.7	90.0	28.0	▼
	89.7	31.5	91.2	30.1		87.5	30.2	91.5	29.4	
平均	90.12	32.02	90.25	29.33		90.03	31.99	90.36	29.50	
SD	1.55	0.98	1.32	1.06		1.52	1.18	1.65	1.05	
A ₂	89.6	32.3	90.7	31.2		91.0	33.3	89.7	29.7	
	90.8	35.0	91.0	32.4		91.6	34.0	92.0	31.6	
	88.7	32.8	89.3	31.0		89.7	35.0	89.6	31.7	
	88.8	31.7	92.4	34.3		92.4	36.7	91.3	33.4	△
	90.6	35.6	92.7	33.7		90.4	34.2	91.6	32.2	
	89.3	33.1	89.7	29.1		91.2	36.1	91.8	30.8	△
	92.4	35.4	89.2	30.2		90.4	32.6	90.3	30.1	
	89.1	33.9	90.4	30.1		91.0	34.3	91.5	30.3	
	89.7	33.7	90.1	32.1		89.3	33.0	90.1	30.9	
	92.2	36.1	88.9	28.9	△	90.2	33.5	91.0	31.5	
平均	90.12	33.96	90.44	31.30		90.72	34.27	90.89	31.22	
SD	1.34	1.51	1.30	1.82		0.91	1.33	0.89	1.10	

注) △:流れ不良(yが36以上), ▼:ウス不良(yが28以下)

超える不良が発生している。このような不良の発生原因の探求し、不良品の発生を防止するための方法を見いだしたい。中心膜厚 Y の平均値も不良の発生に関連して上下することが確認でき、不良発生の代用特性であることが推測される。

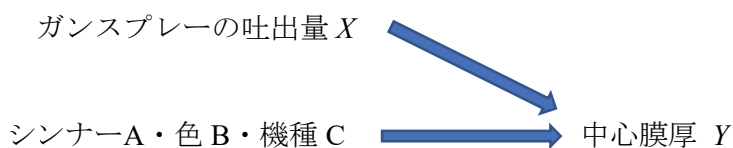


図 7.1 吐出量 X などが中心膜厚 Y に与える影響

(2×2) 格子状の層別散布図に回帰直線の重ね書き

塗装に用いるガンスプレーの吐出量 X が、中心膜厚 Y に対する共変量であることの確認は、 X と Y の散布図を描き、回帰直線を描くことにより確認できる。表 7.1 のデータの並びに対応し、Excel で作成した (2×2) の格子状の層別散布図を図 7.2 に示す。この図の作成は、まず、左上段に 機種: 国内 C₁・色 B₁・シンナー A₁ の散布図を作成、シンナー A₂ を重ね書きし、マーカの種類と色を整え、「近似曲線の追加」により回帰直線を引く。これをコピーし (2×2)

の格子状に並べ、コピーしたそれぞれの散布図に対して、散布図の点をマウスで選択すると（ x の欄, y の欄）の最初の参照先が Excel シート上で色別し表示されるので、それらに対応する位置にマウスで移動する。すると自動的に散布図の点が新たに表示され、回帰直線と回帰式も更新される。

管理限界の上限および下限を示す点線は、Excel シート上に始点と終点の (x, y) の座標を与え、「データの選択」により追加で重ね書きし点線に変更している。各々の図に対する説明は、Excel シートのセルの幅を図の大きさに調整して説明を加える。ワードに張り付ける際には、 (2×2) 格子状の図を含むセル全体をコピーすることで全体のバランスを保つことができる。

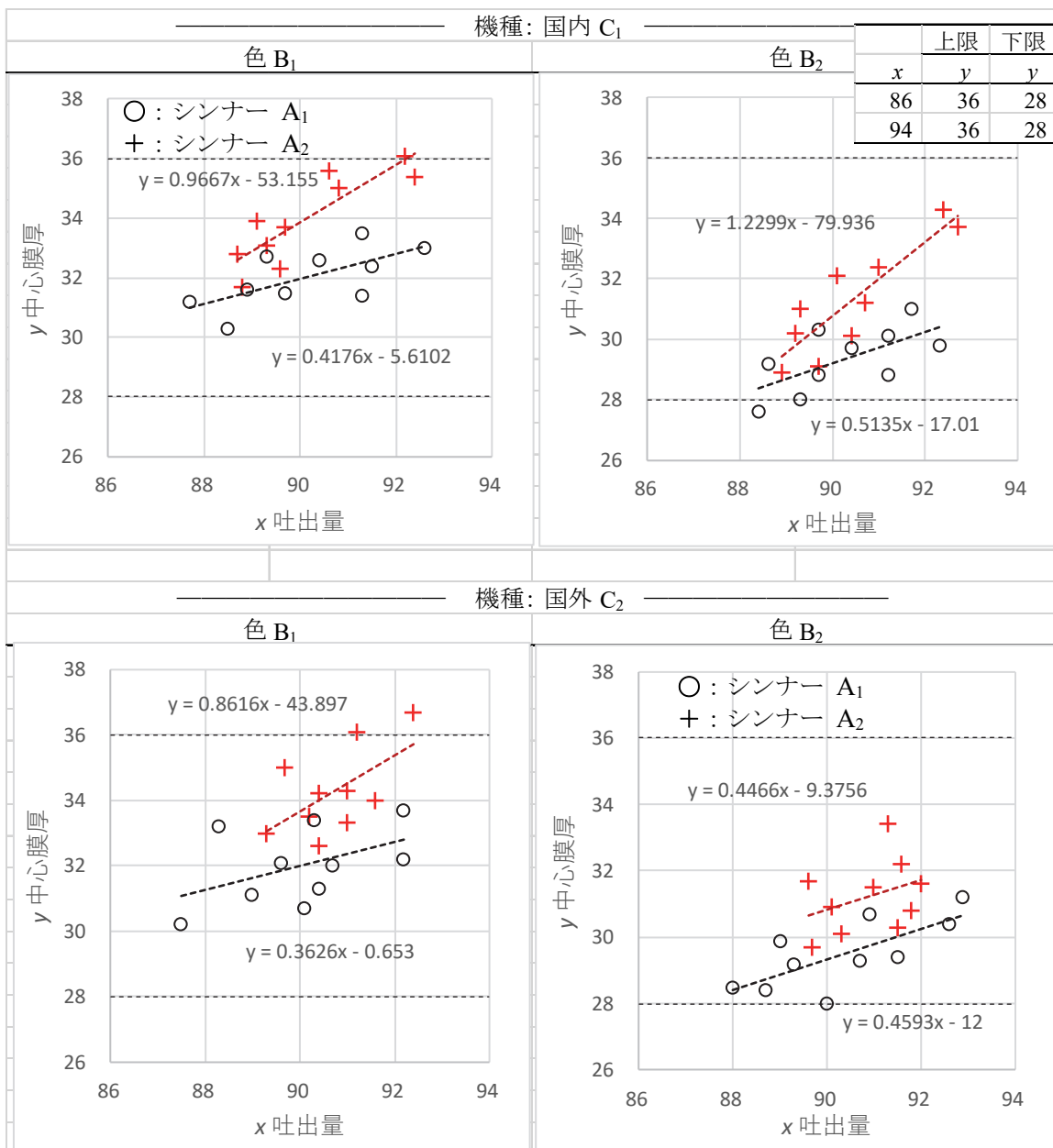


図 7.2 Excel による格子状の層別散布図行列に回帰直線の重ね書き

図 7.2 から、吐出量 X と中心膜厚 Y の関係は、直線的であることが確認できる。また、色 B_1 は、色 B_2 に対して中心膜厚 Y が厚めで、シンナー A_2 の場合に中心膜厚が 36 を超える流れ不良が発生している。色 B_2 でシンナー A_1 の場合には、中心膜厚が 28 を下回るウス不良が発生している。シンナー A_1 は、シンナー A_2 に対して散布図上で回帰直線の傾きが小さ目である。機種 C (国内, 国外) を色 B とシンナー A で対応づけて比較すると、ほぼ同様の散布図であることも確認される。

これらの視察から、色 B_2 でシンナー A_1 の場合は、吐出量を多めに、色 B_1 でシンナー A_2 の場合には、吐出量 X を少なめにすることが、不良を出にくくするために必要と思われる。これらの格子状の層別散布図から得られた結果から、3 因子間の交互作用はないと判断されるが、シンナーと吐出量 X には、交互作用が統計的にあることを確認する必要がある。

共変量 X の分布の均一性の確認

ガンスプレーの吐出量 X は、中心膜厚 Y に対して影響を与える共変量であることが、図 7.2 で確認することができた。また、各実験条件の下で吐出量 X の範囲が微妙に異なるようでもある。そこで、図 7.3 に $(2 \times 2 \times 2)$ の各実験条件の下での吐出量 X を散布図にし、平均 \pm SD を重ね書きした結果を示す。なお、この図は、表 7.1 のデータを行方向に並べ直し、 $(2 \times 2 \times 2)$ の因子の組み合わせに対して 1, 2, ..., 8 の組み合わせ番号を付与し散布図を作成している。さらに、平均と SD を計算し、平均 \pm SD を求めて散布図に重ね書きし、デフォルトのマークを「データ系列の書式設定」によって「+」などに変更し、全体を点線で結んだ後に一部を消している。

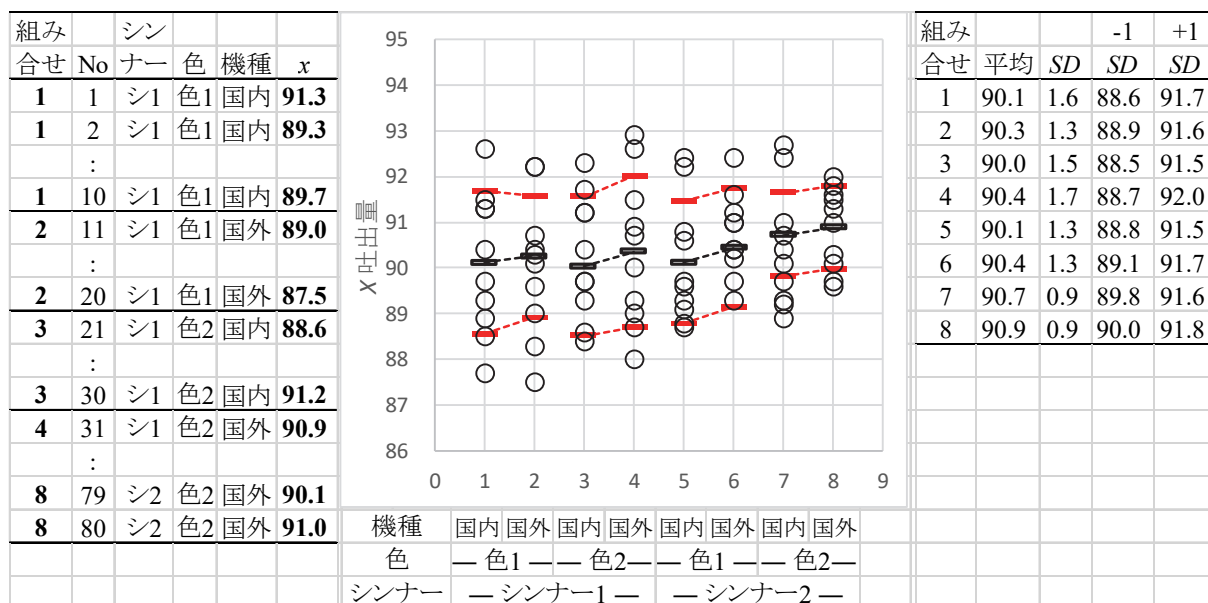


図 7.3 ガンスプレーの吐出量 X の Excel の散布図に平均 \pm SD を重ね書き

(2×2) 格子状の層別散布図の併合

闇雲に散布図を作るのではなく、どのような散布図が問題解決の糸口になるか深い洞察力が必要である。図 7.2 から、冷蔵庫の機種 C (国内, 国外) による差異が相対的に小さいので、表 7.1 を表 7.2 のように、行方向に機種 C (国内, 国外) を展開し、それらを併合した (2×1) 格子状の層別散布図を図 7.4 示す。

表 7.2 機種 C (国内, 国外) を行方向に並べ替え

	----- シンナー A ₁ -----				----- シンナー A ₂ -----			
	色 B ₁		色 B ₂		色 B ₁		色 B ₂	
機種	x	y	x	y	x	y	x	y
国内	91.3	33.5	88.6	29.2	89.6	32.3	90.7	31.2
C ₁	89.3	32.7	88.4	27.6	90.8	35.0	91.0	32.4
⋮								
	89.7	31.5	91.2	30.1	92.2	36.1	88.9	28.9
国外	89.0	31.1	90.9	30.7	91.0	33.3	89.7	29.7
C ₂	⋮				⋮			
	92.2	33.7	90.0	28.0	89.3	33.0	90.1	30.9
	87.5	30.2	91.5	29.4	90.2	33.5	91.0	31.5

図 7.4 から色 B₁ に対するシンナー A₂ の場合に中心膜厚が 36 を超えて「流れ不良」発生し、色 B₂ に対するシンナー A₁ の場合に中心膜厚が 28 以下になる「ウス不良」発生していることが鮮明になり、不良を発生しないような解決方法の糸口が見いだされる。Excel では、散布図に対する回帰直線の重ね書きは容易にできるのであるが、残念ながら回帰直線の 95%信頼区間を描く機能はない。そのためには、きちっとした統計解析のスキルが必要であり、グラフ表示のための作図のスキルも必要となる。

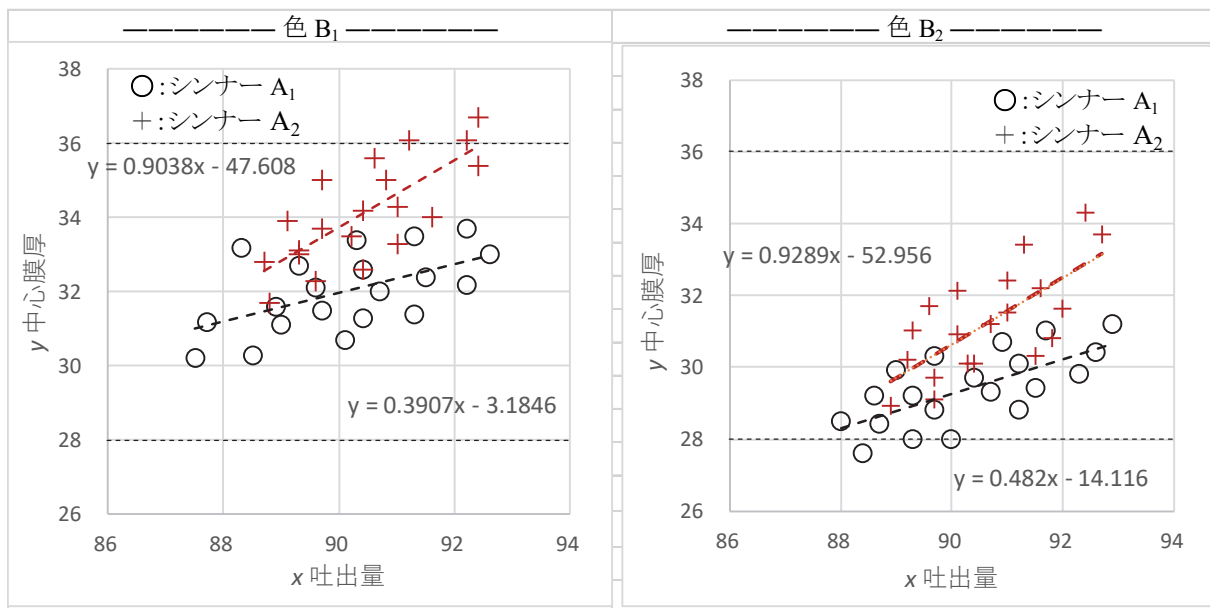


図 7.4 機種 (国内, 国外) を合わせた色別のシンナーに関する層別散布図

7.2. 回帰直線に対する個別データの95%信頼区間

管理限界がある場合の統計的な判断をするためには、それぞれの回帰直線に対する個別データの95%信頼区間の重ね書きが必要である。図7.2および図7.4でExcelの近似曲線の機能を用いて回帰直線を重ね書きしたのであるが、さらに個別データの95%信頼区間を重ね書きしたい。Excelによる回帰直線の95%信頼区間の重ね書きによる作図は、第1.2節の図1.9に2本の傾きが異なる直線に対する事例、第5.5節の図5.1による個別データの95%信頼区間の事例、第6.2節の図6.6により平行な直線に対する事例を示してきた。図7.4は、層別因子としての「色B」に2種類の「シンナーA」に対する回帰直線が重ね書きされ、全体で4本の回帰直線となっている。この4本の回帰直線についての個別データについての個別データの同時95%信頼区間を重ね書きし、統計的な考察をしたい。

交互作用を含む線形モデル

前節の格子状の層別散布図に対する回帰直線の重ね書き図による検討から、線形モデルで取り上げる変数を、シンナーA (シンナーA₁: a=0, シンナーA₂: a=1), 色B (色B₁: b=0, 色B₂: b=1) のように (0, 1) 型ダミー変数で与え、ガンスプレーの吐出量 x, シンナーの種類 a と吐出量 x の交互作用のための変数を ax とする。線形モデルを、

$$y_i = \theta_0 + \theta_1 a_i + \theta_2 b_i + \theta_3 x_i + \theta_4 (ax)_i + \varepsilon_i \quad (7.1)$$

とし、表7.3に解析用のデータを示す。

表7.3 交互作用を含む解析用のデザイン行列 X

No.	シンナー A	色 B	機種 C	膜厚 y	デザイン行列 X				
					x ₀	a	b	x	ax
1	1	1	国内	33.5	1	0	0	91.3	0.0
:									
20	1	1	国外	30.2	1	0	0	87.5	0.0
21	1	2	国内	29.2	1	0	1	88.6	0.0
:									
40	1	2	国外	29.4	1	0	1	91.5	0.0
41	2	1	国内	32.3	1	1	0	89.6	89.6
:									
60	2	1	国外	33.5	1	1	0	90.2	90.2
61	2	2	国内	31.2	1	1	1	90.7	90.7
:									
80	2	2	国外	31.5	1	1	1	91.0	91.0

デザイン行列 X の設定は、最初の1行目で切片の変数 x₀ に対し「1」とし、変数 a は、Excel の if() 関数を用い「if((シンナーA=1), 0, 1)」によって (0, 1) 型ダミー変数とし、変数

b も同様に「if(色B=1), 0, 1)」と(0, 1)型ダミー変数を生成する. 交互作用 ax は, 「 $a*x$ 」と積で与える. 2行目から80行目は, 変数 x を除いてフィルハンドルで数式のコピー機能でセットする. このような操作により, (80行×5列)のデザイン行列 X を作成する. 表7.4左にExcelの回帰分析による分散分析表とパラメータの推定値を示す.

表7.4 デザイン行列 X に対する回帰分析の結果

No	シ	色	機	種	y	デザイン行列 X					分散分析表 (切片に0を使用) off				
						x_0	a	b	x	ax	自由度	変動	分散	分散比	
1	1	1	内		33.5	1	0	0	91.3	0.0	回帰	4	282.7236	70.6809	84.3497
:											残差	75	62.8463	0.8380	
20	1	1	外		30.2	1	0	0	87.5	0.0	合計	79	345.5699		
21	1	2	内		29.2	1	0	1	88.6	0.0					
:															
40	1	2	外		29.4	1	0	1	91.5	0.0	θ^0 切片	-7.5835	9.0466	-0.8383	0.4045
41	2	1	内		32.3	1	1	0	89.6	89.6	θ^1 a	-40.3310	14.8599	-2.7141	0.0082
:											θ^2 b	-2.8842	0.2056	-14.0253	0.0000
60	2	1	外		33.5	1	1	0	90.2	90.2	θ^3 x	0.4406	0.1004	4.3904	0.0000
61	2	2	内		31.2	1	1	1	90.7	90.7	θ^4 ax	0.4656	0.1643	2.8329	0.0059
:															
80	2	2	外		31.5	1	1	1	91.0	91.0					

パラメータ(係数)の推定値から, 交互作用 ax と主効果 b が共に有意な差であることが確認される. なお, 主効果 a および x については, 交互作用 ax がモデルに含まれているので, 慎重に取り扱わなければならない. 特に主効果 a の推定値 $\hat{\theta}_1 = -40.3310$ は, 共変量が, $x=0$ における切片の差の推定値であることを認識する必要がある. この回帰分析の結果の解釈のためには, 推定されたパラメータを用いて, 回帰式を明示し図に重ね書きする必要がある. パラメータの推定値から, 因子 A と因子 B の組み合わせによる回帰式は,

$$\left. \begin{aligned}
 \text{シンナーA}_1, \text{色B}_1: \hat{y}^{(1,1)} &= \hat{\theta}_0 + \hat{\theta}_3 x = -7.5835 + 0.4656x \\
 \text{シンナーA}_1, \text{色B}_2: \hat{y}^{(1,2)} &= (\hat{\theta}_0 + \hat{\theta}_2) + \hat{\theta}_3 x = -10.4677 + 0.4656x \\
 \text{シンナーA}_2, \text{色B}_1: \hat{y}^{(2,1)} &= (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_3 + \hat{\theta}_4)x = -47.9145 + 0.9061x \\
 \text{シンナーA}_2, \text{色B}_2: \hat{y}^{(2,2)} &= (\hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_2) + (\hat{\theta}_3 + \hat{\theta}_4)x = -50.7987 + 0.9061x
 \end{aligned} \right\} \quad (7.2)$$

として得られる. なお, これらの回帰式は, 図7.4で示した色別・シンナー別に求めた個別データから求めた回帰式とは, 微妙に異なる. これは, 全データを用いた回帰分析のパラメータから推定した式となっているためである.

個別データの95%信頼区間

式(7.2)に示された4本の回帰式の個別データの95%信頼区間を描くためには, それぞれの回帰直線の推定値 \hat{y} に対する分散を推定する必要があり, 別々に回帰分析をすることも考えられるが, 交互作用の検討に用いた表7.4に示したパラメータの推定値を反映したい. その

ためには、表 7.4 に示した (80 行×5 列) のデザイン行列 X を用いたパラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$ を必要とする。

表 7.5 に示すのは、転置行列 X^T とデザイン行列 X の積 $(X^T X)$ の計算方法である。転置行列 X^T は、Transpose (X の範囲) 関数を使い、行列の積 $(X^T X)$ は、Mmult (X^T の範囲, X の範囲) 関数を用いて行う。転置行列 X^T を Excel シート上に作成するのは難儀なので、Transpose () 関数を Mmult () 関数の引数とすることにより一括計算することができる。

表 7.5 転置行列 X^T とデザイン行列 X の積 $(X^T X)$

転置行列 $X^T = \text{Transpose}(X \text{ の範囲})$						デザイン行列 X					=Mmult (X^T の範囲, X の範囲) $(X^T X)$				
1	2	3	...	79	80	x_0	a	b	x	ax	x_0	a	b	x	ax
1	1	1	...	1	1	1	0	0	91.3	0.0	80	40	40	7229	3622
0	0	0	...	1	1	1	0	0	89.3	0.0	40	40	20	3622	3622
0	0	0	...	1	1	1	0	0	88.9	0.0	40	20	40	3619	1813
91.3	89.3	88.9	...	90.1	91.0	:	:	:	:	:	7229	3622	3619	653420	327967
0.0	0.0	0.0	...	90.1	91.0	1	1	1	90.1	90.1	3622	3622	1813	327967	327967
						1	1	1	91.0	91.0	x_0	a	b	x	ax
(5×80)						(80×5)					(5×5)				

表 7.6 に示すのは、 $(X^T X)$ について Minverse () 関数により逆行列 $(X^T X)^{-1}$ を求め、回帰分析の分散分析表の残差の分散 (誤差分散の推定値) $\hat{\sigma}^2 = 0.8380$ を掛けて、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を求めた結果である。

表 7.6 パラメータの共分散行列 $\Sigma(\hat{\theta})$

逆行列 $(X^T X)^{-1} = \text{Minverse}(X^T X \text{ の範囲})$					$\hat{\sigma}^2$	パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$				
97.67	-97.27	0.10	-1.08	1.08	0.8380	81.8408	-81.5062	0.0839	-0.9077	0.9035
-97.27	263.52	0.10	1.08	-2.91		-81.5062	220.8166	0.0847	0.9030	-2.4419
0.10	0.10	0.05	0.00	0.00		0.0839	0.0847	0.0423	-0.0012	-0.0009
-1.08	1.08	0.00	0.01	-0.01		-0.9077	0.9030	-0.0012	0.0101	-0.0100
1.08	-2.91	0.00	-0.01	0.03		0.9035	-2.4419	-0.0009	-0.0100	0.0270
(5×5)						(5×5)				

表 7.7 に示すのは、パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根が表 7.4 に示した Excel の回帰分析の標準誤差 SE に一致することの確認である。

表 7.7 パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根

パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$					対角要素の平方根		表 7.4 回帰分析の結果		
					$\text{Var}(\hat{\theta})$	SE	係数 標準誤差		
81.8408	-81.5062	0.0839	-0.9077	0.9035	81.8408	9.0466	切片	-7.5835	9.0466
-81.5062	220.8166	0.0847	0.9030	-2.4419	220.8166	14.8599	a	-40.3310	14.8599
0.0839	0.0847	0.0423	-0.0012	-0.0009	0.0423	0.2056	b	-2.8842	0.2056
-0.9077	0.9030	-0.0012	0.0101	-0.0100	0.0101	0.1004	x	0.4406	0.1004
0.9035	-2.4419	-0.0009	-0.0100	0.0270	0.0270	0.1643	ax	0.4656	0.1643

色B別の個別データの95%信頼区間

表 7.8 に示すのは、吐出量 x を $[87, 88, \dots, 93]^T$ のように設定し、推定値 \hat{y} および分散 $Var(\hat{y})$ を計算し、個別データの95%信頼区間の推定値である。変数名 (x_0, a, b, x, ax) に対して別名を $l=(l_0, l_1, l_2, l_3, l_4)$ とし、 l に推定目的に沿った係数を与え、推定されたパラメータ $\hat{\theta}$ との積和（線形和） $L=l\hat{\theta}$ によって推定値 \hat{y} を計算し、分散を $Var(L)=Var(l\hat{\theta})=l\Sigma(\hat{\theta})l^T$ によって求めた結果である。

表 7.4 で推定された回帰パラメータは、

$$\hat{\theta} = [-7.5835 \quad -40.3310 \quad -2.8842 \quad 0.4406 \quad 0.4656]^T$$

であり、表 7.8 の最初の行のベクトルを $l^{(1)}$ としたときに、線形和

$$\hat{y}_{x=87}^{(1)} = l^{(1)}\hat{\theta}$$

$$= \text{Mmult}(l^{(1)} \text{ の範囲}, \hat{\theta} \text{ の範囲})$$

	x_0	a	b	x	ax	$\hat{\theta}$	\hat{y}
=	1	0	0	87	0	-7.5835	30.7466
						-40.3310	
						-2.8842	
						0.4406	
						0.4656	

は、色 B_1 ・シンナー A_1 について $x=87$ の場合の回帰直線の推定値が得られている。その分散 $Var(L^{(1)})=Var(l^{(1)}\hat{\theta})$ は、表 7.7 に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いて

表 7.8 色 B_1 についてのシンナー A_1 と A_2 についての個別データの95%信頼区間

シ	色		l_0	l_1	l_2	l_3	l_4	$l\hat{\theta}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	\hat{y}	$Var(\hat{y})$	個別L95%	個別U95%
1	1	$L^{(1)}$	1	0	0	87	0	30.7466	0.1303	28.7864	32.7068
1	1	$L^{(2)}$	1	0	0	88	0	31.1872	0.0773	29.2814	33.0930
1	1	$L^{(3)}$	1	0	0	89	0	31.6278	0.0444	29.7566	33.4990
1	1	$L^{(4)}$	1	0	0	90	0	32.0684	0.0317	30.2107	33.9261
1	1	$L^{(5)}$	1	0	0	91	0	32.5090	0.0391	30.6434	34.3745
1	1	$L^{(6)}$	1	0	0	92	0	32.9495	0.0666	31.0549	34.8442
1	1	$L^{(7)}$	1	0	0	93	0	33.3901	0.1143	31.4461	35.3341
シ	色		l_0	l_1	l_2	l_3	l_4	$l\hat{\theta}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	\hat{y}	$Var(\hat{y})$	個別L95%	個別U95%
2	1	$L^{(8)}$	1	1	0	87	87	30.9196	0.2381	28.8531	32.9861
2	1	$L^{(9)}$	1	1	0	88	88	31.8257	0.1364	29.8593	33.7921
2	1	$L^{(10)}$	1	1	0	89	89	32.7319	0.0689	30.8349	34.6289
2	1	$L^{(11)}$	1	1	0	90	90	33.6380	0.0354	31.7763	35.4997
2	1	$L^{(12)}$	1	1	0	90.56	90.56	34.1424	0.0316	32.2848	36.0000 #
2	1	$L^{(13)}$	1	1	0	92	92	35.4503	0.0708	33.5512	37.3493
2	1	$L^{(14)}$	1	1	0	93	93	36.3564	0.1397	34.3867	38.3261

Excel のソルバーで個別U95%が、上限の36になるように x を変化させ 90.56 が得られている。

$$Var(L^{(1)}) = l^{(1)} \Sigma(\hat{\theta}) l^{(1)T}$$

	x_0	a	b	x	ax	$\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$					x'^T	$Var(y^{\wedge})$
=	1	0	0	87	0	81.8408	-81.5062	0.0839	-0.9077	0.9035	1	= 0.1303
						-81.5062	220.8166	0.0847	0.9030	-2.4419	0	
						0.0839	0.0847	0.0423	-0.0012	-0.0009	0	
						-0.9077	0.9030	-0.0012	0.0101	-0.0100	87	
						0.9035	-2.4419	-0.0009	-0.0100	0.0270	0	

によって計算され、個別データの95%信頼区間は、

$$\begin{aligned} \text{個別 } 95\%CL &= L^{(1)} \pm t_{0.05}(80-5) \sqrt{\hat{\sigma}^2 + Var(L^{(1)})} \\ &= 30.7466 \pm 1.9921 \times \sqrt{0.8380 + 0.1303} \\ &= (28.7864, 32.7068) \end{aligned}$$

となる。個別データの95%信頼区間が管理限界の上限 $y=36$ を超える吐出量 x を求めるために、Excelのソルバーを用いて、表7.8の個別U95%が、管理限界の上限の $y=36$ となるように x を変化させると $\hat{x}=90.56$ を求めることができる。

表7.9に示すは、色 B_2 に対する個別データの95%信頼区間の計算結果であり、図7.5に示すのは、表7.8および表7.9で求めた、回帰直線の推定値を重ね書きし、それらの個別データの95%信頼区間をさらに重ね書きした結果である。

表7.9 色 B_2 についてのシンナー A_1 と A_2 についての個別データの95%信頼区間

シ	色		l_0	l_1	l_2	l_3	l_4	$l\theta^{\wedge}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	y^{\wedge}	$Var(y^{\wedge})$	個別L95%	個別U95%
1	2	$L^{(15)}$	1	0	1	87	0	27.8625	0.1377	25.8948	29.8302
1	2	$L^{(16)}$	1	0	1	88	0	28.3031	0.0824	26.3920	30.2141
1	2	$L^{(17)}$	1	0	1	89	0	28.7436	0.0472	26.8694	30.6178
1	2	$L^{(18)}$	1	0	1	90	0	29.1842	0.0321	27.3260	31.0424
1	2	$L^{(19)}$	1	0	1	91	0	29.6248	0.0372	27.7612	31.4884
1	2	$L^{(20)}$	1	0	1	91.57	0	29.8762	0.0491	28.0000	31.7525 #
1	2	$L^{(21)}$	1	0	1	93	0	30.5059	0.1078	28.5687	32.4432
シ	色		l_0	l_1	l_2	l_3	l_4	$l\theta^{\wedge}$	分散	個別95%信頼区間	
A	B	L	x_0	a	b	x	ax	y^{\wedge}	$Var(y^{\wedge})$	個別L95%	個別U95%
2	2	$L^{(22)}$	1	1	1	87	87	28.0354	0.2530	25.9547	30.1161
2	2	$L^{(23)}$	1	1	1	88	88	28.9416	0.1471	26.9644	30.9187
2	2	$L^{(24)}$	1	1	1	89.06	89	29.9005	0.0722	28.0000	31.8010 #
2	2	$L^{(25)}$	1	1	1	90	90	30.7538	0.0377	28.8897	32.6179
2	2	$L^{(26)}$	1	1	1	91	91	31.6600	0.0341	29.7996	33.5203
2	2	$L^{(27)}$	1	1	1	92	92	32.5661	0.0647	30.6735	34.4588
2	2	$L^{(28)}$	1	1	1	93	93	33.4723	0.1294	31.5130	35.4315

#ソルバーで個別L95%が、下限の28になるように x を変化させ 91.57 および 89.06 が得られている。

図7.5左に示す色 B_1 でシンナー A_1 (+) で上側管理限界 36 を超える吐出量は、個別データの95%信頼区間から、目の子で吐出量 x が 90.5 ぐらいと読め、Excelのソルバーによる正確な

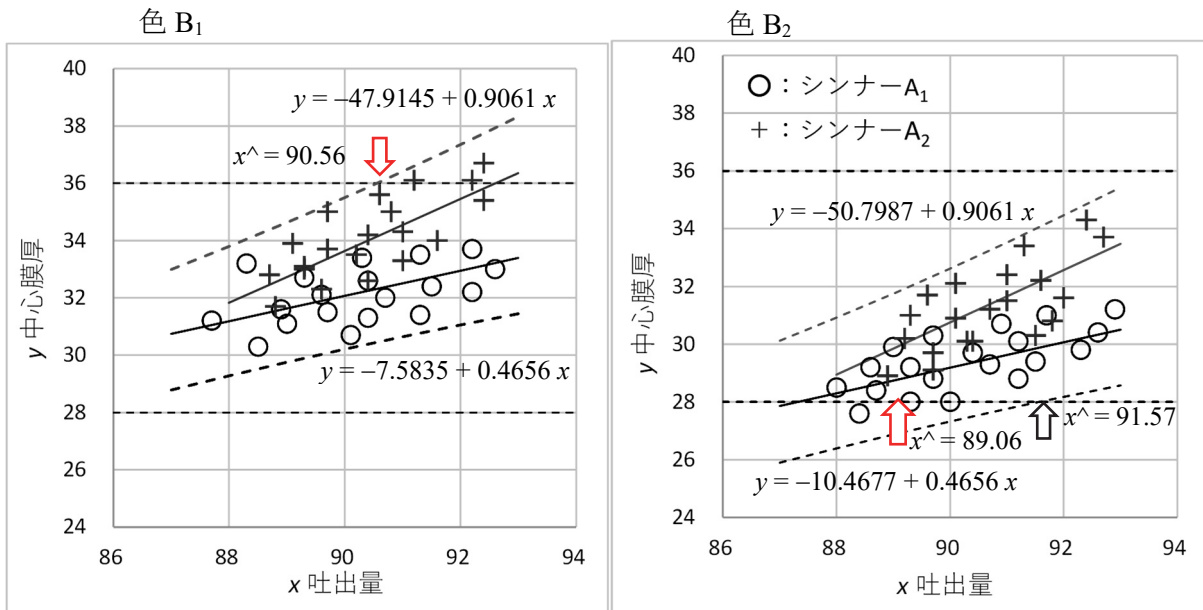


図 7.5 色 B₁ および B₂ についての回帰直線に対する個別データの 95% 信頼区間の表示

推定では、 $\hat{x}=90.56$ となる。吐出量の最大値が 90.56 を超えないようにすることで、不良率の軽減が図られる。また、交互作用 ax が統計的に有意であることから、シンナー A₂ の傾きが、シンナー A₁ に比べて急であるので不良が出やすくなった原因でもある。実験結果の報告に際し、このような統計的分析を行なうことにより、よりの確な数値に基づく対策が提言できる。

図 7.5 右に示す色 B₂ の場合は、色 B₁ の場合に比べて、全体的に下限方向にシフトしている。表 7.4 のパラメータ b の推定値から、シフト量は $\hat{\theta}_2 = -2.8842$ であり、 $p < 0.0000$ と高度な差となっている。その結果としてシンナー A₁ の場合に下側管理限界 28 を下回る不良が発生する原因となっている。Excel のソルバーを用いて、個別データの下側 95% 信頼区間が、下側管理限界 28 を超える吐出量を推定すると $\hat{x} = 91.57$ となり、吐出量 x の平均である 90.5 を超える値となっており、吐出量が少ないことが不良発生の原因である。また、シンナー A₂ の場合には、表 7.9 から、吐出量が 89.06 で、下側管理限界 28 を下回ることも読み取れる。

以上の結果から、不良品を軽減する対策として短絡的には、吐出量 X の変化に対し勾配が大きいシンナー 2 をやめてシンナー 1 に統一し、色 2 に対しては、吐出量を 91.57 以上とすることにより中心被膜 y が、管理限界の下限以上となり、不良率の軽減となることが推測される。シンナー 2 を使う別の理由があるならば、色 1 に対しては、吐出量を 90.56 以下とすることが望まれる。

7.3. Excelの回帰分析による交互作用を含む変数減少法

第7.1節で共変量を含む3因子実験データの解析に際し、適切な散布図を作成することにより、問題解決のための方法が見い出された。さらに、第7.2節でExcelの回帰分析に加え、Excelの行列関数を用いて回帰直線の個別データの95%信頼区間を重ね書きすることにより、きめ細かな統計解析を実現できることを示した。

本節では、表7.3に示したデータを用い、Excelの回帰分析による探索的な解析を試みる。共変量がなければ、繰り返しが10の3因子の分散分析による解析ができ、一般的な分散分析の手順により問題解決となる。だが、共変量が一つ含まれるだけで、従来の平方和の分解を主体にする解析法が適用できない。実験前に特定の値に設定することができない量的変数がある場合に、第6章に示した伝統的な共分散分析として解析手順が定式化されている。ただし、現実的には1因子実験の場合に限定されている。共変量を含む3因子実験データの解析は、どのように行ったらよいのだろうか。

ダミー変数

質的変数を量的変数に変換する方法として、ダミー変数が使われる。統計ソフトでは、質的変数をダミー変数に自動的に変換する機能があり、Rでは表7.10に示すように最初の水準を0とする(0, 1)型、SASでは最後の水準を0とする(1, 0)型、JMPでは表7.11に示すような水準ごとに足して0となるような(1, -1)対比型が使われている。

表 7.10 (0, 1) 型ダミー変数 (R の場合)

シナナー	<i>a</i>	色	<i>b</i>	機種	<i>c</i>	シナナー	色	<i>a</i>	<i>b</i>	<i>ab</i>
1	0	1	0	国内	0	1	1	0	0	0
2	1	2	1	国外	1		2	0	1	0
						2	1	1	0	0
							2	1	1	1

表 7.11 (1, -1) 対比型ダミー変数 (JMP の場合)

シナナー	<i>a</i>	色	<i>b</i>	機種	<i>c</i>	シナナー	色	<i>a</i>	<i>b</i>	<i>ab</i>
1	1	1	1	国内	1	1	1	1	1	1
2	-1	2	-1	国外	-1		2	1	-1	-1
						2	1	-1	1	-1
							2	-1	-1	1

Excelの回帰分析では、このような自動変換の機能がないので自ら設定する必要がある。どのような型のダミー変数とするかは自由なので、ここでは最初の水準を0とする(0, 1)型を用いる。

交互作用を含む線形モデル

経験的に交互作用を考慮する必要性がないような分野ならば、考慮すべき変数をすべて含む線形モデルを適用し、 p 値などを見ながらモデルに含める変数を適宜減らす方法が適している。ただし、Excel の回帰分析で、取り上げる変数は、すべて隣り合っていないといけないので、モデル化から外したい変数ベクトルを **Ctrl+X** で選択し、適当な列の前に同じサイズのセルを選択し、「切り取ったセルの挿入」することにより、デザイン行列 X の変数の入れ替えが、サイズを変更することなく容易にできる。

ここでは、経験的に交互作用があることを考慮にするような実験データの解析とする。量的変数 X の場合には、2 乗項 X^2 も考慮する必要もあるが、吐出量 X に関しては、直線のあてはめとし、質的変数 (A シンナー、B 色、C 機種) について (0, 1) 型ダミー変数とする。解析モデルとして、2 因子交互作用は、すべての組合せについて設定し、3 因子交互作用は、念のために $A \times B \times C$ を含めるが、吐出量 X との 3 因子交互作用は、無いものとみなす。

主効果： A シンナー、B 色、C 機種、吐出量 X ，
 交互作用： $A \times X$ ， $B \times X$ ， $C \times X$ ，
 $A \times B$ ， $A \times C$ ， $B \times C$ ，
 $A \times B \times C$

表 7.12 に最大モデルでの回帰分析を行うために、(80 行×12 列) のデザイン行列 X を示す。交互作用の設定は、主効果の変数 (a , b , c , x) に対し掛け算で与え、Excel のフィルハンドルを用いて計算式のコピー機能を用いることにより、ストレスなく作成できる。

表 7.12 (80 行×12 列) のデザイン行列 X

A	B	C	No.	x_0	a	b	c	x	ax	bx	cx	ab	ac	bc	abc	y
1	1	1	1	1	0	0	0	91.3	0.0	0.0	0.0	0	0	0	0	33.5
			:													:
		2	11	1	0	0	1	89.0	0.0	0.0	89.0	0	0	0	0	31.1
			:													:
	2	1	21	1	0	1	0	88.6	0.0	88.6	0.0	0	0	0	0	29.2
			:													:
		2	31	1	0	1	1	90.9	0.0	90.9	90.9	0	0	1	0	30.7
			:													:
2	1	1	41	1	1	0	0	89.6	89.6	0.0	0.0	0	0	0	0	32.3
			:													:
		2	51	1	1	0	1	91.0	91.0	0.0	91.0	0	1	0	0	33.3
			:													:
	2	1	61	1	1	1	0	90.7	90.7	90.7	0.0	1	0	0	0	31.2
			:													:
		2	71	1	1	1	1	89.7	89.7	89.7	89.7	1	1	1	1	29.7
			:													:
			80	1	1	1	1	91.0	91.0	91.0	91.0	1	1	1	1	31.5

直交表の解析を想定すると、すべての変数 (a, b, \dots, bc, abc) を対象に回帰分析を行い、有意でない変数の平方和をまとめて誤差項にプールすればと思われるかもしれない。しかし、(0, 1) 型ダミー変数としたこと、さらに、共変量 x が含まれることにより、変数間の直交性が失われているために、一つの変数を除いただけで、他のすべての変数の平方和が変化してしまい戸惑うことになる。

変数選択の手順

取り上げた要因を (主効果, 2 因子交互作用, 3 因子交互作用) のように区別する。そして、高次の交互作用が含まれている場合には、主効果も含め低次の交互作用の p 値は、完全に無視し、同じレベルの交互作用の要因のみの p 値を用いて、要因の取捨選択を前もって定めた ($p=0.15$) などの基準で行なう。

表 7.13 左に示すのは、すべての変数 (a, b, \dots, bc, abc) を含んだ 3 因子モデルでの回帰分析の結果で、3 因子交互作用 abc が有意ではないとの結論を得る。表 7.13 右に示すのは、交互作用 abc を除いた 2 因子モデルでの回帰分析の結果である。分散分析表を比較すると、3 因子モデルの回帰の自由度が 11 から 2 因子モデルで 10 に減少し変動 (平方和) がわずかに減り、その分が残差の平方和に移行している。

表 7.13 3 因子モデル vs. 2 因子モデル

分散分析表(3因子モデル)					分散分析表(2因子モデル)					
		自由度	変動	分散	分散比		自由度	変動	分散	分散比
	回帰	11	286.264	26.024	29.839	回帰	10	286.061	28.606	33.169
	残差	68	59.305	0.872		残差	69	59.508	0.862	
	合計	79	345.570			合計	79	345.570		
		係数	標準誤差	t	P -値		係数	標準誤差	t	P -値
θ^{\wedge}_0	切片	-12.077	14.082	-0.858	0.3941	切片	-12.552	13.969	-0.899	0.3720
θ^{\wedge}_1	a	-41.352	15.865	-2.606	0.0112 *	a	-41.303	15.776	-2.618	0.0109 *
θ^{\wedge}_2	b	-10.703	14.899	-0.718	0.4750 ?	b	-9.882	14.719	-0.671	0.5042 ?
θ^{\wedge}_3	c	16.105	15.281	1.054	0.2956	c	16.142	15.195	1.062	0.2918
θ^{\wedge}_4	x	0.489	0.156	3.132	0.0026 *	x	0.494	0.155	3.186	0.0022 *
θ^{\wedge}_5	ax	0.480	0.176	2.730	0.0081 *	ax	0.481	0.175	2.748	0.0076 *
θ^{\wedge}_6	bx	0.088	0.165	0.533	0.5955	bx	0.080	0.163	0.490	0.6256
θ^{\wedge}_7	cx	-0.179	0.170	-1.054	0.2956	cx	-0.178	0.169	-1.056	0.2948
θ^{\wedge}_8	ab	-0.233	0.593	-0.394	0.6949	ab	-0.433	0.421	-1.028	0.3074
θ^{\wedge}_9	ac	-0.163	0.604	-0.269	0.7888	ac	-0.367	0.429	-0.856	0.3949
θ^{\wedge}_{10}	bc	0.151	0.592	0.256	0.7989	bc	-0.049	0.419	-0.117	0.9069
θ^{\wedge}_{11}	abc	-0.405	0.840	-0.482	0.6311					

回帰係数を比較するとわずかではあるが、すべて異なる。伝統的な直交表の解析では、有意でない要因の平方和を残差平方和にプールし、誤差の自由度を大きくすることが手順とし

て確立して、交互作用も含めた全ての要因が直交しているので、平方和も回帰係数（パラメータ）も変化することはない。表 7.4 では、因子 B（変数 b ）の p 値が $p < 0.0001$ なのに対し、表 7.13 に示す 3 因子モデルで $p = 0.4750$ 、2 因子モデルで $p = 0.5042$ と全く異なる結果となっている。この原因は、因子 B の主効果を含む交互作用（ $B \times X$ 、 $A \times B$ 、 $B \times C$ 、 $A \times B \times C$ ）が含まれていることに起因している。

表 7.14 右に示すのは、2 因子モデルで $p = 0.15$ 以上となる交互作用をモデルから外し、表 7.14 右に示すように、交互作用 ax のみを含めたモデルとする。因子 B の p 値が、 $p = 0.504$ から、 $p < 0.0001$ へ大変化を起こしている。このことを以ってしても、交互作用が含まれているときの主効果の p 値による判断は、全く意味をなさないことが示されている。推定された回帰パラメータも $\hat{\theta}_2' = -9.882$ から $\hat{\theta}_2'' = -2.886$ へと大きく変化している。これは、量的変数 x との交互作用がある場合に、 $x = 0$ における（ $b = 0$ と $b = 1$ ）に対する切片の差となるためであり、まったく意味のない推定値である。

主効果+交互作用モデルで、因子 C が有意ではないので、モデルから外すことにする。因子 A と共変量 X の主効果は有意な差なので、モデルから外す誘惑は起きないが、もしもどちらかが有意でない場合に外したくなる誘惑にかられるかもしれないが、絶対に外してはならない。

表 7.14 2 因子モデル vs. 主効果+交互作用モデル

分散分析表(2因子モデル)					分散分析表(主効果+交互作用モデル)					
	自由度	変動	分散	分散比		自由度	変動	分散	分散比	
回帰	10	286.061	28.606	33.169	回帰	5	283.172	56.634	67.164	
残差	69	59.508	0.862		残差	74	62.398	0.843		
合計	79	345.570			合計	79	345.57			
	係数	標準誤差	t	P -値		係数	標準誤差	t	P -値	
θ^{\wedge}_0'	切片	-12.552	13.969	-0.899	0.3720	切片	-7.528	9.075	-0.830	0.4095
θ^{\wedge}_1'	a	-41.303	15.776	-2.618	0.0109 *	a	-41.778	15.038	-2.778	0.0069 *
θ^{\wedge}_2'	b	-9.882	14.719	-0.671	0.5042 ?	b	-2.886	0.206	-13.99	0.0000 *
θ^{\wedge}_3'	c	16.142	15.195	1.062	0.2918	c	-0.152	0.208	-0.729	0.4683
θ^{\wedge}_4'	x	0.494	0.155	3.186	0.0022 *	x	0.441	0.101	4.379	0.0000 *
θ^{\wedge}_5'	ax	0.481	0.175	2.748	0.0076 *	ax	0.482	0.166	2.895	0.0050 *
θ^{\wedge}_6'	bx	0.080	0.163	0.490	0.6256					
θ^{\wedge}_7'	cx	-0.178	0.169	-1.056	0.2948					
θ^{\wedge}_8'	ab	-0.433	0.421	-1.028	0.3074					
θ^{\wedge}_9'	ac	-0.367	0.429	-0.856	0.3949					
θ^{\wedge}_{10}'	bc	-0.049	0.419	-0.117	0.9069					

表 7.15 に示すように、因子 C を外して最終モデルにようやく到達し、表 7.4 に示した層別散布図を使った探索的な結果にたどり着いた。推定されたパラメータの解釈を試みよう。パ

ラメータの表記として $\hat{\theta}_k'''$ としたいのであるが、煩雑なので $\hat{\theta}_k''' \rightarrow \hat{\theta}_k$ とする。切片の $\hat{\theta}_0 = -7.5835$ は、 $a=b=x=0$ の場合であり、図 7.5 左の (シンナーA₁・色 B₁) ○印の切片である。因子 A の $\hat{\theta}_1 = -40.3310$ は、(シンナーA₁・色 B₁) の切片から (シンナーA₂・色 B₁) の切片の差となっている。したがって、表 7.15 右に示した因子 A の $p=0.0082$ は、図 7.5 右に示したデータの範囲での推論される p 値でないことに注意が必要である。このような不可解な推定値となるのは、交互作用 A×X が回帰モデルに含まれているためであり、交互作用がある場合の主効果の推定値および p 値に全く意味がないことに注意しなければならない。

表 7.15 (主効果+交互作用) モデル vs. 最終モデル

分散分析表(主効果+交互作用モデル)					分散分析表(最終モデル)				
	自由度	変動	分散	分散比		自由度	変動	分散	分散比
回帰	5	283.172	56.634	67.164	回帰	4	282.724	70.681	84.350
残差	74	62.398	0.843		残差	75	62.846	0.838	
合計	79	345.57			合計	79	345.57		
	係数	標準誤差	t	P-値		係数	標準誤差	t	P-値
θ_0'' 切片	-7.528	9.075	-0.830	0.4095	θ_0 切片	-7.5835	9.0466	-0.8383	0.4045
θ_1'' a	-41.778	15.038	-2.778	0.0069 *	θ_1 a	-40.3310	14.8599	-2.7141	0.0082 *
θ_2'' b	-2.886	0.206	-13.99	0.0000 *	θ_2 b	-2.8842	0.2056	-14.0253	0.0000 *
θ_3'' c	-0.152	0.208	-0.729	0.4683	θ_3 x	0.4406	0.1004	4.3904	0.0000 *
θ_4'' x	0.441	0.101	4.379	0.0000 *	θ_4 ax	0.4656	0.1643	2.8329	0.0059 *
θ_5'' ax	0.482	0.166	2.895	0.0050 *					

推定されたパラメータを用いた回帰直線の推定

量的変数 x と 2 水準の質的変数をダミー変数とした回帰分析に際し、交互作用がなければ、推定されたパラメータの解釈は楽なのであるが、交互作用が入っていると説明に窮することになる。そのために、表 7.16 に示すように、量的変数 x の存在範囲を設定し、推定されたパラメータを用いて反応 y の推定値 \hat{y} を計算し、結果を考察することが必須である。因子 A の水準、因子 B の水準の組み合わせごとに、共変量 X の範囲を (0, 88, 93) と変化させた場合の推定値 \hat{y} の推定値を計算した結果である。なお、この推定された結果は、図 7.5 左および図 7.5 右に示した回帰直線に一致する。これらの 4 本の層別回帰直線の内、管理限界を超え

表 7.16 推定されたパラメータを用いた中心膜厚 y の推定値

シンナー	色	切片		傾き		x		
A	B					0	88	93
A ₁	B ₁ (b=0)	θ_0	-7.5835	θ_3	0.4406	-7.58	31.19	33.39
a=0	B ₂ (b=1)	$\theta_0 + \theta_2$	-10.4677	θ_3	0.4406	-10.47	28.31	30.51
A ₂	B ₁ (b=0)	$\theta_0 + \theta_1$	-47.9145	$\theta_3 + \theta_4$	0.9061	-47.91	31.82	36.35
a=1	B ₂ (b=1)	$\theta_0 + \theta_1 + \theta_2$	-50.7987	$\theta_3 + \theta_4$	0.9061	-50.80	28.94	33.47

るデータが含まれる A_1B_2 と A_2B_1 の組み合わせについて散布図を作成し、2本の回帰直線および個別データの95%信頼区間を重ね書きする。

表 7.17 2本の回帰直線の個別データの95%信頼区間 (表 7.8 および表 7.9 の要約)

B	A	印	X	切片	a	b	x	ax	θ^{\wedge}	y^{\wedge}	$Var(y^{\wedge})$	L95%	U95%	個別 L95%	個別 U95%
B1	A2	●	87	1	1	0	87	87	-7.584	30.920	0.238	29.948	31.892	28.853	32.986
			89	1	1	0	89	89	-40.331	32.732	0.069	32.209	33.255	30.835	34.629
			91	1	1	0	91	91	-2.884	34.544	0.036	34.166	34.922	32.682	36.407
			93	1	1	0	93	93	0.441	36.356	0.140	35.612	37.101	34.387	38.326
B2	A1	△	87	1	0	1	87	0	0.466	27.862	0.138	27.123	28.602	25.895	29.830
			89	1	0	1	89	0		28.744	0.047	28.311	29.176	26.869	30.618
			91	1	0	1	91	0		29.625	0.037	29.241	30.009	27.761	31.488
			93	1	0	1	93	0		30.506	0.108	29.852	31.160	28.569	32.443
$t_{0.05}(80-5) =$												1.9921			

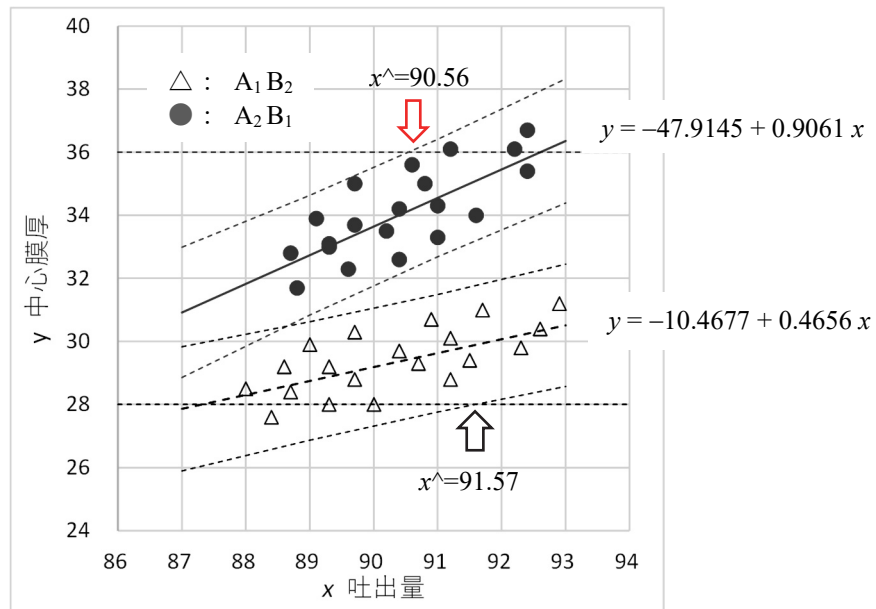


図 7.6 管理限界を超えるガンスプレーの吐出量 x の推定値

第 7.1 節では、各種の散布図に直線を重ね書きし、管理限界を超える事例について考察を行ってきたのであるが、図 7.6 に示すように、個別データの 95%信頼区間を重ね書きすることにより、統計的に管理限界を超える吐出量 x が、より低めであったり、より高めであったりが明確に数値として把握でき、現実の対策を考える際に役に立つ。

7.4. JMP のステップワイズ法による変数選択

表 7.13 に示した Excel による回帰分析によるモデル選択では、表 7.13 で「3 因子モデル vs. 2 因子モデル」、表 7.14 で「2 因子モデル vs. 主効果+交互作用モデル」、表 7.15 では、「主効果+交互作用モデル vs. 最終モデル」では、パラメータの p 値を見ながら手作業で段階的に変数選択を行った。

JMP では、「モデルのあてはめ」の中で「ステップワイズ法」がサポートされ、手動または自動的に変数を増加したり減少したりすることができる。量的変数のみならず質的変数も扱え、それらの交互作用も含めることができる。さらに、反応変数には、量的変数だけでなく質的変数（名義尺度、順序尺度）も扱えるようになっていが、本書では取り扱わない。また、変数の増減に際し、交互作用が対象となった場合には、主効果などの下位の変数も合わせて増減の判定を考慮する機能もある優れたものである。

手作業での変数選択

表 7.18 に示すのは、Excel での手作業での変数選択で用いた表 7.12 のデータを JMP にペーストして作成したデータである。

表 7.18 変数選択のための JMP データファイル

	No.	x0	a	b	c	x	ax	cx	bx	ab	ac	bc	abc	y
1	1	1	0	0	0	91.3	0.0	0.0	0.0	0	0	0	0	33.5
2	2	1	0	0	0	89.3	0.0	0.0	0.0	0	0	0	0	32.7
:														
9	9	1	0	0	0	87.7	0.0	0.0	0.0	0	0	0	0	31.2
10	10	1	0	0	0	89.7	0.0	0.0	0.0	0	0	0	0	31.5
11	11	1	0	0	1	89.0	0.0	89.0	0.0	0	0	0	0	31.1
:														
79	79	1	1	1	1	90.1	90.1	90.1	90.1	1	1	1	1	30.9
80	80	1	1	1	1	91.0	91.0	91.0	91.0	1	1	1	1	31.5

JMP の「ステップワイズ法」は、多機能で多彩であるが、ここでは手作業による変数選択の過程を示す。表 7.19 左（3 因子モデル）は、表 7.18 の全変数を追加した回帰分析の結果が示されている。3 因子の交互作用 abc の p 値が 0.631 なので、チェック・ボックスを off にし、2 因子交互作用モデルとする。表 7.19 中（2 因子モデル）の p 値を表 7.19 左（3 因子モデル）と比較すると p 値がかなり変化することが読み取れる。

Excel での変数選択は、変数の増減に伴い新たに回帰分析を行ない、結果をコピー&ペーストし、書式を整える必要があり、煩雑であった。JMP の場合は、チェック・ボックスをクリックするだけで変数を増減した結果が即時に表示される。表 7.19 中 (2 因子モデル) の 2 因子交互作用の p 値が大きいものから減少させると、交互作用 ax の p 値は揺れ動くが、小さいまま残り続ける。表 7.19 右 (最終モデル) に示したのは、主効果 c が $p=0.468$ であり、交互作用項にも含まれていないので、減少させた結果であり表 7.15 右の Excel で求めた最終モデルに一致する。

表 7.19 JMP のステップワイズ法を用いた手作業による変数減少法
(3 因子モデル) (2 因子モデル) (最終モデル)

SSE				DFE				MSE			
59.305				68				0.872			
現在の推定値				現在の推定値				現在の推定値			
追加	パラメータ	推定値	p値	追加	パラメータ	推定値	p値	追加	パラメータ	推定値	p値
<input checked="" type="checkbox"/>	切片	-12.077	1.000	<input checked="" type="checkbox"/>	切片	-12.552	1.000	<input checked="" type="checkbox"/>	切片	-7.584	1.000
<input checked="" type="checkbox"/>	a	-41.352	0.011	<input checked="" type="checkbox"/>	a	-41.303	0.011	<input checked="" type="checkbox"/>	a	-40.331	0.008
<input checked="" type="checkbox"/>	b	-10.703	0.475	<input checked="" type="checkbox"/>	b	-9.882	0.504	<input checked="" type="checkbox"/>	b	-2.884	0.000
<input checked="" type="checkbox"/>	c	16.105	0.296	<input checked="" type="checkbox"/>	c	16.142	0.292	<input type="checkbox"/>	c	0.000	0.468
<input checked="" type="checkbox"/>	x	0.489	0.003	<input checked="" type="checkbox"/>	x	0.494	0.002	<input checked="" type="checkbox"/>	x	0.441	0.000
<input checked="" type="checkbox"/>	ax	0.480	0.008	<input checked="" type="checkbox"/>	ax	0.481	0.008	<input checked="" type="checkbox"/>	ax	0.466	0.006
<input checked="" type="checkbox"/>	cx	-0.179	0.296	<input checked="" type="checkbox"/>	cx	-0.178	0.295	<input type="checkbox"/>	cx	0.000	0.458
<input checked="" type="checkbox"/>	bx	0.088	0.595	<input checked="" type="checkbox"/>	bx	0.080	0.626	<input type="checkbox"/>	bx	0.000	0.777
<input checked="" type="checkbox"/>	ab	-0.233	0.695	<input checked="" type="checkbox"/>	ab	-0.433	0.307	<input type="checkbox"/>	ab	0.000	0.347
<input checked="" type="checkbox"/>	ac	-0.163	0.789	<input checked="" type="checkbox"/>	ac	-0.367	0.395	<input type="checkbox"/>	ac	0.000	0.202
<input checked="" type="checkbox"/>	bc	0.151	0.799	<input checked="" type="checkbox"/>	bc	-0.049	0.907	<input type="checkbox"/>	bc	0.000	0.528
<input checked="" type="checkbox"/>	abc	-0.405	0.631	<input type="checkbox"/>	abc	0.000	0.631	<input type="checkbox"/>	abc	0.000	0.133

注) 切片の P 値が常に $p=1.00$ となっているのは、変数選択の対象外としているためである。

自動的な変数選択

回帰分析における変数の自動選択は、“最適なモデル”の探索に際し、恣意性を排除するための方法として多くの統計ソフトに組み込まれている。ただし、変数増加法、変数減少法、変数増減法、変数減増法、総当たり法などがあり、さらに、交互作用を含めた場合についての選択法など、恣意性を排除したいとの努力が続けられている。

表 7.19 に示した 3 因子モデルでは、吐出量 X との交互作用を 2 因子交互作用に限定していた。例えば、 $A \times B \times X$ といった 3 因子交互作用を含めていない。 $A \times B \times C \times X$ のような 4 因子交互作用をすべて含め、JMP のステップワイズによる変数の自動選択を試みる。

表 7.20 に示すのは、選択する変数として、主効果 (a, b, c, x) とし、交互作用を JMP の内部で自動的に設定するために「マクロ」の機能で「すべての組み合わせ」を選択すると、2

因子交互作用「 $a*b$ 」から4因子交互作用「 $a*b*c*x$ 」までが自動的に変数としてセットされた結果である。示されている自由度は、高次の交互作用をすべて積み上げた結果であり、変数 a の自由度8は、それぞれ自由度1の (a , $a*b$, $a*c$, $a*x$, $a*b*c$, $a*b*x$, $a*c*x$, $a*b*c*x$) を加えたものである。「閾値 p 値を」(0.1, 0.1) と設定し、すべての変数を「追加」し、「変数増減」, 「組み合わせ」のルールでの変数選択を実行する。

表 7.20 4次元交互作用まで含むフルモデル

ステップワイズ回帰の設定							
停止ルール:	<input type="text" value="閾値p値"/>						
	変数を追加するときのp値	0.1					
	変数を除去するときのp値	0.1					
方向:	<input type="text" value="変数増減"/>						
ルール:	<input type="text" value="組み合わせ"/>						
SSE	DFE	MSE	Cp	p	AICc	BIC	
57.153916	64	0.8930299	16	16	243.9988	274.6223	
現在の推定値							
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	p値
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	-5.6102	1	0.0000	0.0000	1.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a	-47.545	8	70.9514	9.9313	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b	-11.400	8	167.257	23.4115	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	c	4.9572	8	4.7610	0.6664	0.7189
<input type="checkbox"/>	<input checked="" type="checkbox"/>	x	0.4176	8	60.9531	8.5318	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b	-15.381	4	1.6514	0.4623	0.7631
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*c	4.3011	4	2.4386	0.6827	0.6065
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*x	0.5491	4	8.1424	2.2794	0.0703
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b*c	0.0527	4	1.2599	0.3527	0.8412
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b*x	0.0959	4	1.3638	0.3818	0.8209
<input type="checkbox"/>	<input checked="" type="checkbox"/>	c*x	-0.0550	4	3.1060	0.8695	0.4872
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b*c	61.2491	2	0.8165	0.4572	0.6351
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b*x	0.1674	2	0.8425	0.4717	0.6261
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*c*x	-0.0501	2	1.7276	0.9673	0.3856
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b*c*x	0.0008	2	1.1493	0.6435	0.5288
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*b*c*x	-0.6791	1	0.7704	0.8627	0.3565

自動的に変数選択された結果を表 7.21 に示す。結果は、表 7.19 示した手動での結果に一致する。示されている自由度は、選択されている変数の場合は、高次の交互作用の変数の自由度を合わせた結果であり、選択されていない場合は、低次の交互作用を含む結果となっている。

表 7.21 変数増減法により選択されたモデル

現在の推定値							
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	p値
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	-7.5835	1	0.0000	0.0000	1.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a	-40.331	2	67.6239	40.3508	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	b	-2.8842	1	164.831	196.708	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	c	0.0000	1	0.4481	0.5314	0.4683
<input type="checkbox"/>	<input checked="" type="checkbox"/>	x	0.4406	2	56.2733	33.5780	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	a*b	0.0000	1	0.7507	0.8946	0.3473
<input type="checkbox"/>	<input type="checkbox"/>	a*c	0.0000	2	1.4216	0.8447	0.4338
<input type="checkbox"/>	<input checked="" type="checkbox"/>	a*x	0.4656	1	6.7247	8.0252	0.0059
<input type="checkbox"/>	<input type="checkbox"/>	b*c	0.0000	1	0.0245	0.0288	0.8657
<input type="checkbox"/>	<input type="checkbox"/>	b*x	0.0000	1	0.0684	0.0806	0.7772
<input type="checkbox"/>	<input type="checkbox"/>	c*x	0.0000	2	1.6661	0.9940	0.3751
<input type="checkbox"/>	<input type="checkbox"/>	a*b*c	0.0000	5	2.3958	0.5548	0.7341
<input type="checkbox"/>	<input type="checkbox"/>	a*b*x	0.0000	1	0.0338	0.0398	0.8425
<input type="checkbox"/>	<input type="checkbox"/>	a*c*x	0.0000	4	3.4510	1.0313	0.3972
<input type="checkbox"/>	<input type="checkbox"/>	b*c*x	0.0000	5	2.3434	0.5422	0.7436
<input type="checkbox"/>	<input type="checkbox"/>	a*b*c*x	0.0000	11	5.6924	0.5795	0.8385

質的変数と量的変数の交互作用を含む変数選択

観察研究で、説明変数が多数ある場合に、目的とする反応 y に対し、最もあてはまりの良い線形モデルを自動的に選択したいとの要望に沿って様々な方法が研究され、多くの統計ソフトで実用化されてきた。恣意性を排除した「自動的」な変数選択が行われたことを強調する人達に対し、様々な選択方法があり一意に決めることはできない、変数選択ではなくすべての変数の組み合わせの中から選ぶ方が望ましいのではないか、などの反論もあり悩ましい問題でもある。

JMP では、「自動的」な変数選択だけでなく、「手動的」な変数選択が GUI 操作によって実現できる特徴がある。さらに、交互作用を自動的に変数リストに加える機能を有することも現実の課題解決に欠かせない。また、選択された主効果および交互作用を含んだ変数に対し「モデルのあてはめ」による「標準最小 2 乗法」に繋げる配慮もされており、きめ細かな対応が可能である。

これまで示してきた解析法は、質的変数をダミー変数により量的変数とした解析を行ってきたが、JMP の「ステップワイズ法」では、質的変数のままでも交互作用を含めて変数選択ができる。内部では、もちろん全て (1, -1) 対比型ダミー変数に置き換えられていて、「モデルのあてはめ」での結果の出力に際しては、質的変数として変換された表示となる。

表 7.22 に (0, 1) 型ダミー変数 (a, b, c) を元の 2 水準 (1, 2) の因子 (A, B, C) としたデータリストを示す. もちろん, Excel の回帰分析のデータとしては使うことができない.

表 7.22 質的変数のままのデータリスト

	No.	A	B	C	x	y
1	1	A1	B1	C1	91.3	33.5
2	2	A1	B1	C1	89.3	32.7
:						
9	9	A1	B1	C1	87.7	31.2
10	10	A1	B1	C1	89.7	31.5
11	11	A1	B1	C2	89.0	31.1
:						
79	79	A2	B2	C2	90.1	30.9
80	80	A2	B2	C2	91.0	31.5

表 7.23 に示したのは, JMP の「ステップワイズ法」を起動し, 表 7.20 に示したと同様に質的変数 (A, B, C) と共変量 x を選択し, 「マクロ」の「すべての組み合わせ」で 4 因子交互作用までを自動作成し, 「変数増減」による変数選択をした結果である. (平方和, F 値, p 値) は, 表 7.20 と同じ結果が得られているが, パラメータの推定値はまったく異なる. これは, JMP の内部で生成されるダミー変数が (1, -1) 対比型ダミー変数となっていることに起因する. 表示されている変数名 “A{A1-A2}” は, A1 に対し 1 が, A2 に対し -1 となる対比型ダミー変数と理解される. “B{B2-B1}” は, B1 に対し -1, B2 に対し 1 とする対比型ダミー変数が自動設定されている. また, 変数 x に関する交互作用において “A {A1-A2} * (x - 90.3662)” のように変数 x の総平均 90.3662 を差し引く中心化が行われている. これは, 多重共線性による致命的な計算誤差をなくすためである.

変数選択の結果をさらに深堀する

表 7.23 右上に, モデルの作成, モデルの実行 の選択ができるようになっていて, 選択されている交互作用を含む変数について「モデルのあてはめ」の「標準最小 2 乗法」を起動できるようになっている. 実際に実行すると, 表 7.24 に示すように分散分析表が出力される. 表 7.23 の $SSE=62.8463$ と表 7.24 の誤差の平方和 $S_e = 62.8463$ と一致することが確認される. ただし, 効果の検定の平方和は, 自由度が同じならば一致するが, 交互作用を含む主効果 A および x は一致しない.

表 7.23 質的変数と量的変数を込みにした変数選択の結果

yのステップワイズ

ステップワイズ回帰の設定

停止ルール:

変数を追加するときのp値 0.1

変数を除去するときのp値 0.1

方向:

ルール:

→

すべて追加

モデルの作成

←

すべて削除

モデルの実行

SSE	DFE	RMSE	R2乗	自由度調整R2乗	Cp	p	AICc	BIC
62.846289	75	0.9153964	0.8181	0.8084	0.374224	5	220.8741	234.0155

現在の推定値

ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	"p値"
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	-29.1911	1	0.0000	0.000	1.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	A{A1-A2}	-0.8701	2	67.6239	40.351	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}	0.0000	1	0.4481	0.531	0.4683
<input type="checkbox"/>	<input checked="" type="checkbox"/>	B{B2-B1}	-1.4421	1	164.8314	196.708	0.0000
<input type="checkbox"/>	<input checked="" type="checkbox"/>	x	0.6734	2	56.2733	33.578	0.0000
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}	0.0000	2	1.4216	0.845	0.4338
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}	0.0000	1	0.7507	0.895	0.3473
<input type="checkbox"/>	<input checked="" type="checkbox"/>	A{A1-A2}*(x-90.3662)	-0.2328	1	6.7247	8.025	0.0059
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*B{B2-B1}	0.0000	1	0.0245	0.029	0.8657
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*(x-90.3662)	0.0000	2	1.6661	0.994	0.3751
<input type="checkbox"/>	<input type="checkbox"/>	B{B2-B1}*(x-90.3662)	0.0000	1	0.0684	0.081	0.7772
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*B{B2-B1}	0.0000	5	2.3958	0.555	0.7341
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*(x-90.3662)	0.0000	4	3.4510	1.031	0.3972
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*B{B2-B1}*(x-90.3662)	0.0000	1	0.0338	0.040	0.8425
<input type="checkbox"/>	<input type="checkbox"/>	C{C1-C2}*B{B2-B1}*(x-90.3662)	0.0000	5	2.3434	0.542	0.7436
<input type="checkbox"/>	<input type="checkbox"/>	A{A1-A2}*C{C1-C2}*B{B2-B1}*(x-90.3662)	0.0000	11	5.6924	0.579	0.8385

表 7.24 選択された変数による「モデルのあてはめ」での分散分析表

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	4	282.7236	70.6809	84.3497
誤差	75	62.8463	0.8380	p値(Prob>F)
全体(修正済み)	79	345.5699		<.0001*

効果の検定					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
A	1	59.3629	59.3629	70.8430	<.0001*
B	1	164.8314	164.8314	196.7078	<.0001*
x	1	55.7923	55.7923	66.5819	<.0001*
A*x	1	6.7247	6.7247	8.0252	0.0059*

表 7.25 に示すのは、「モデルのあてはめ」で出力されたパラメータの推定値である。この推定値は、表 7.15 右に示した (0, 1) 型ダミー変数での推定値と全く異なるが、切片を除いて p 値は同じである。このように、用いるダミー変数が異なれば、推定値も全く異なるので、パラメータの推定値を用いた考察は、理解の妨げになる。

表 7.25 選択された変数による「モデルのあてはめ」でのパラメータ推定値

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	-29.1911	7.4616	-3.9122	0.0002*
A[A1]	-0.8701	0.1034	-8.4168	<.0001*
B[B1]	1.4421	0.1028	14.0253	<.0001*
x	0.6734	0.0825	8.1598	<.0001*
A[A1]*(x-90.3662)	-0.2328	0.0822	-2.8329	0.0059*

JMP の優れている点は、表 7.25 の推定値を使った各種の予測プロファイルの出力にある。特に交互作用がある場合に、交互作用プロットは、結果の解釈に役立つ。図 7.7 右上段に示したのは、 B_1 を選択した場合の A_1 および A_2 の回帰直線であり、図 7.7 右下段に示したのは、 B_2 を選択した場合の A_1 および A_2 の回帰直線である。

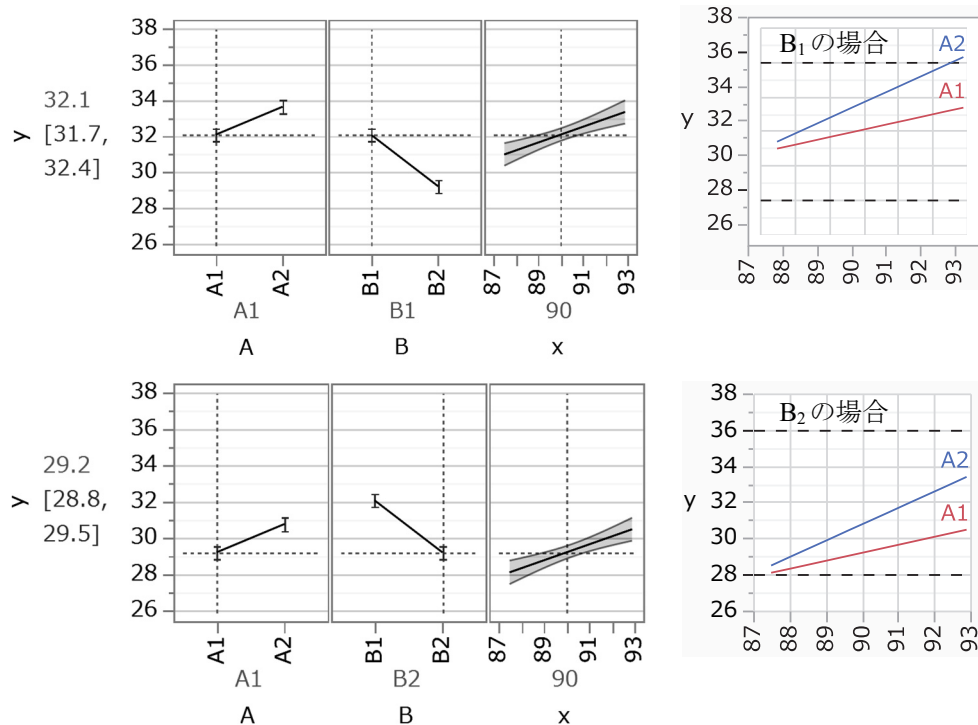


図 7.7 JMP の予測プロファイルと交互作用プロファイルの連動

8. 交絡変数と共変量を含む 2 群比較

医薬品あるいは農薬などを含む化合物の安全性の評価の一環として、実験動物を用いた反復投与による各種の毒性試験が行なわれている。竹内ら (1989), 「統計学辞典, IV 分野別応用, 4.3 毒性評価」に示されているように実験手順は、細かく規定されており、用いる統計的方法についても分野により細かな差異はあるもののおおむね規準化されている。反復投与試験の多くは、実験動物の成長期に行なわれ、化合物の実験動物に与える影響を多面的に評価するために数多くの変数が測定される。このために、化合物の直接的な作用と推定される反応だけではなく、交絡変数が間接的に作用する反応も拾い出してしまう。間接的な作用であることを識別するために、反応と交絡変数の比などを求めて、交絡変数の影響を取り除く工夫もされている。交絡変数の影響を統計的に取り除く方法として共分散分析が知られており、そのため交絡変数は、解析に際しては、“共変量”として混同されがちである。そこで、多面的な作用が内在する交絡変数を含めた探索的な解析法について例示する。

8.1. 体重の増加抑制は総摂餌量の影響だけか

体重の増加抑制

ある化合物に胎児に対する催奇形性が有るか否かの確認のために、動物を対象にした生殖発生毒性試験が行なわれている。表 8.1 に示すのは、吉村編著 (1987), 「毒性・薬効データの統計解析, 第 3.3 節 体重は増加量でみるべきか」で取り上げられているデータである。これは、ラットを対象に妊娠 7 日から 18 日まで化合物を反復投与し、胎仔に対する各種の影響を検討するために得られたデータの一部である。

このデータは、妊娠 7 日に体重 x_1 を測定した後に、化合物 A の反復投与を開始し、妊娠 18 日に体重 y_1 と総摂餌量 x_2 を測定した結果である。さらに、出生した仔数 y_2 が、示されている。化合物 A の反復投与によって、妊娠ラットの体重増加に何らかの抑制がみられるならば、出生仔数に影響を与えるかもしれない。また、データとして得られていないが、胎仔数 x_3 が妊娠ラットの体重 y_1 の増加に影響を与えているのかも知れない。化合物 A の反復投与が体重 y_1 の増加の抑制という影響をもたらしたとすれば、食欲行動抑制のために体重 y_1 の増加の抑制

表 8.1 妊娠ラットの体重増加データ [吉村編著 (1987), 表 3-6]

番号	対照群 A ₁							処置群 A ₂						
	——— 体重g ———				総摂餌量	飼料効率	出生仔数	——— 体重g ———				総摂餌量	飼料効率	出生仔数
	7日	18日	差	率%	x_2	x_2	$y_2(x_3)$	7日	18日	差	率%	x_2	x_2	$y_2(x_3)$
x_1	y_1	y_1-x_1	$差/x_1$	x_2	$差/x_2$	$y_2(x_3)$	x_1	y_1	y_1-x_1	$差/x_1$	x_2	$差/x_2$	$y_2(x_3)$	
1	248.7	318.1	69.4	27.9	203.3	0.341	13	257.9	309.1	51.2	19.9	199.2	0.257	13
2	261.3	324.0	62.7	24.0	223.0	0.281	13	260.7	305.4	44.7	17.1	167.9	0.266	13
3	265.6	339.9	74.3	28.0	244.9	0.303	12	277.0	331.5	54.5	19.7	185.9	0.293	14
4	284.0	356.1	72.1	25.4	236.5	0.305	14	283.8	323.2	39.4	13.9	195.6	0.201	13
5	254.9	328.6	73.7	28.9	204.5	0.360	15	281.2	332.1	50.9	18.1	203.6	0.250	12
6	267.4	338.1	70.7	26.4	223.0	0.317	15	292.3	339.6	47.3	16.2	196.7	0.240	13
7	271.0	349.3	78.3	28.9	228.5	0.343	14	270.1	330.2	60.1	22.3	207.7	0.289	13
8	257.4	308.3	50.9	19.8	221.4	0.230	8	265.8	318.7	52.9	19.9	186.3	0.284	11
9	282.2	340.0	57.8	20.5	221.1	0.261	17	269.9	317.1	47.2	17.5	184.1	0.256	11
10	268.3	335.4	67.1	25.0	210.4	0.319	14	269.8	328.7	58.9	21.8	198.8	0.296	16
11	236.8	296.0	59.2	25.0	191.1	0.310	12	267.1	335.4	68.3	25.6	219.3	0.311	15
12	288.0	367.2	79.2	27.5	258.4	0.307	13	259.4	332.0	72.6	28.0	217.0	0.335	14
13	246.9	313.4	66.5	26.9	197.8	0.336	16	246.3	309.9	63.6	25.8	188.6	0.337	14
14	236.6	304.5	67.9	28.7	215.4	0.315	12	230.1	282.5	52.4	22.8	164.5	0.319	12
15	254.8	342.7	87.9	34.5	232.0	0.379	16	223.6	275.4	51.8	23.2	166.7	0.311	13
16	251.5	320.6	69.1	27.5	225.1	0.307	10	277.1	329.5	52.4	18.9	199.3	0.263	14
17	270.8	345.9	75.1	27.7	227.9	0.330	15	269.1	329.1	60.0	22.3	209.9	0.286	14
18	254.4	326.6	72.2	28.4	215.5	0.335	14	231.4	295.7	64.3	27.8	171.9	0.374	15
19	240.2	310.1	69.9	29.1	206.7	0.338	14	239.0	306.5	67.5	28.2	180.8	0.373	15
20	245.8	310.7	64.9	26.4	216.0	0.300	12	257.4	316.0	58.6	22.8	193.0	0.304	14
平均	259.3	328.8	69.4	26.8	220.1	0.316	13.5	261.5	317.4	55.9	21.6	191.8	0.292	13.5
SD	15.2	18.8	8.18	3.2	16.00	0.033	2.1	18.8	17.6	8.57	4.1	16.07	0.043	1.3

が起きたのか、餌を食べたにもかかわらず化合物 A の投与による何らかの作用によって体重 y_1 の増加が抑制されたのか、を明らかにする必要がある。

図 8.1 に示すように、妊娠 7 日の体重 x_1 は、化合物 A の投与前の測定値であり、化合物の影響を受けない前値である。総摂餌量 x_2 は、化合物 A の投与によって直接的に抑制されるかも知れないし、体重 y_1 の増加の抑制による影響を受けることも考えられる。総摂餌量 x_2 は、妊娠 18 日の体重 y_1 に対する交絡変数であり、化合物 A の投与期間中の直接的な摂餌の抑制、あるいは体重 y_1 の抑制による間接的に相互の影響を受ける。化合物 A の投与により、胎仔が

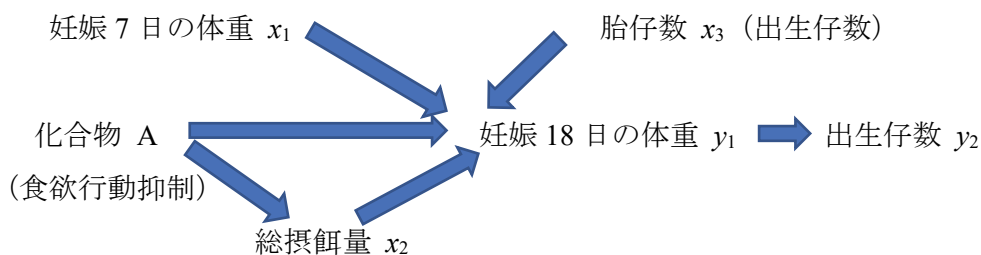


図 8.1 妊娠 18 日の体重に影響を与える交絡変数と共変量

発育不全などの原因で死亡していないことを確認するため、出生した仔数 y_2 も測定されている。なお、出生仔に対しては、さらに、催奇形性が有るか詳細に検討される。

2 群間の t 検定を使った検討

今から 30 年以上前の 1980 年代の交絡変数を含むデータに対する統計解析は、2 群間の t 検定を段階的に適用して考察することが一般的であった。計算には、各群の偏差平方和が必要であり、Excel の DevSq() 関数を使うことにより、元のデータから直接求めることができる。対照群 A_1 の妊娠 7 日の体重 x_1 の偏差平方和 $S_{x_1}^{(1)}$ は、

$$\left. \begin{aligned} S_{x_1}^{(1)} &= \sum_{i=1}^{20} (x_{1,i} - \bar{x}_1)^2 \\ &= \text{DevSq}(248.7, 261.3, \dots, 245.8) = 4380.40 \end{aligned} \right\} \quad (8.1)$$

に示すように、引数に DevSq(A_1 群の x_1 の範囲) と設定することにより求められる。表 8.2 に示すように 2 群間の t 検定は、平均値の差に対して差の SE で割った統計量であり、それぞれの群の偏差平方和を用いて容易に計算できる。なお、Excel には、 t 検定の p 値を直接計算する T.test() 関数もあるが、段階的な計算手順を示すために用いなかった。

表 8.2 対象群と処置群の差の t 検定

		対照群 A_1		処置群 A_2		t 検定				
		平均	平方和 $S^{(1)}$	平均	平方和 $S^{(2)}$	群間差	SE	t	p	
x_1	7日の体重	259.33	4380.40	261.45	6739.09	2.12	5.4094	0.3919	0.6973	NS
y_1	18日の体重	328.78	6715.10	317.38	5897.95	-11.40	5.7613	-1.9779	0.0552	+
$y_1 - x_1$	前後差	69.45	1271.55	55.93	1394.48	-13.52	2.6487	-5.1024	0.0000	***
$(y_1 - x_1)/x_1\%$	率%	26.82	189.78	21.58	317.72	-5.24	1.1557	-4.5369	0.0001	***
x_2	総摂餌量	220.13	4866.00	191.84	4908.93	-28.29	5.0718	-5.5769	0.0000	***
$(y_1 - x_1)/x_2$	飼料効率	0.3159	0.0211	0.2923	0.0355	-0.0236	0.0122	-1.9308	0.0610	+
y_2	出生仔数	13.45	84.95	13.45	32.95	0.00	0.5570	0.0000	1.0000	NS
偏差平方和 = DevSq (データの範囲)						$n = 2 \times 20$	$t = \text{群間差} / SE$			
標準誤差 $SE = \text{Sqrt} (((S^{(1)} + S^{(2)}) / (40 - 2)) * (1/20 + 1/20))$						$p = \text{T.dist.2T} (t , 40 - 2)$				

妊娠 7 日の体重 x_1 は、対照群 A_1 と処置群 A_2 に有意な差 ($p = 0.6973$) ではない。妊娠 18 日の体重 y_1 は、対照群 A_1 に比べて処置群 A_2 は -11.40 と体重の増加が少ないが、 t 検定では微妙 ($p = 0.0552$) な結果である。体重の前後差 ($y_1 - x_1$) および増加率 $(y_1 - x_1)/x_1\%$ にすると SE が減少し、統計的にも明らかな差 ($p < 0.0001$) となっている。体重の前後差 ($y_1 - x_1$) に対して総摂餌量 x_2 で割った飼料効率 $(y_1 - x_1)/x_2$ で比較すると、処置群がわずかに小さいが、微妙 ($p = 0.0610$) な結果である。出生仔数 y_2 には、差が全くないので化合物の妊娠中の胎仔の生死に関与する作用はないといえる。

体重の増加 y_1 の抑制が、化合物 A の直接作用なのか、総摂餌量 x_2 の抑制による間接的な作用なのか、化合物 A が摂餌量 x_2 の食欲行動抑制をしたことによる体重の増加 y_1 の抑制が起き

たのかは、これらの t 検定の結果からでは、はっきりしたことはわからないが、飼料効率に大きな差がないので、「体重の増加抑制は、主として餌を食べないことが影響しているようである」と吉村編著（1987）では述べられている。

体重の増加量および増加率

体重が増加するような場合には、前値 x_1 の影響を除くために前後差 $(y_1 - x_1)$ による群間比較が一般的に行なわれている。この事例は、交絡変数としての総摂餌量 x_2 、共変量としての妊娠7日の体重 x_1 を同時に含む場合に、どのような解析を行なったら良いのであろうか。

複数の交絡変数と共変量がある場合には、1変数ごとに散布図を作成し、相互の関連を概観することが欠かせない。図8.2左に示すように、妊娠7日の体重 x_1 と妊娠18日の体重 y_1 には強い相関関係があり、対照群 A_1 に比較し処置群 A_2 は、妊娠7日の体重が大きい場合に体重の増加抑制がより大きくなっているように思われる。図8.2右の体重の前後差 $(y_1 - x_1)$ をY軸とした散布図からは、より明確な増加抑制が示されている。表8.2に示した2群間の t 検定による検討では、対照群 A_1 の前後差69.45gに対し処置群 A_2 は55.93gであり、 $t=5.1024$ と明らかな体重抑制がある。散布図上の回帰直線からは、7日の体重 x_1 が大きいほど体重の前後差が拡大しているように思われる。

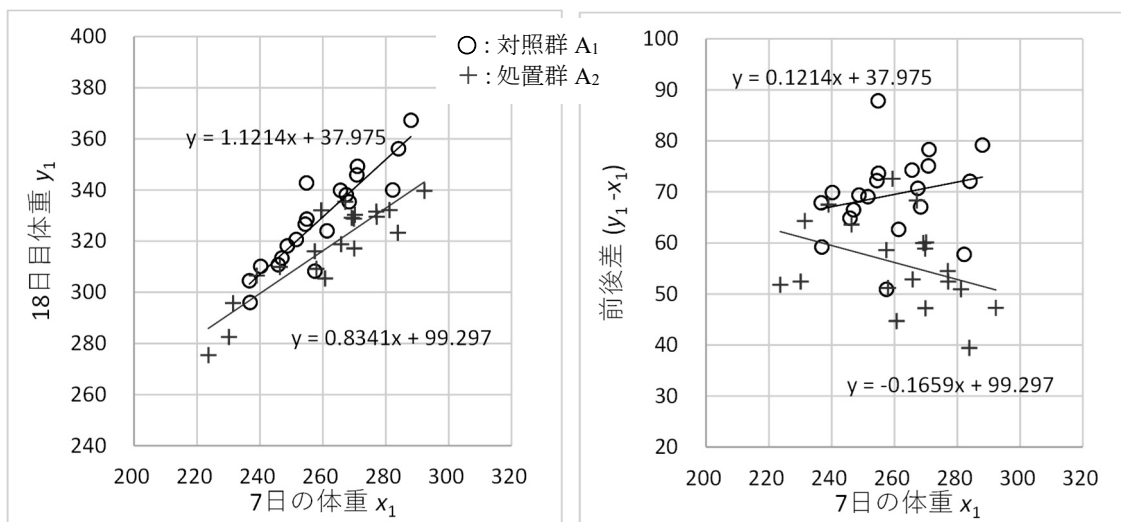


図8.2 妊娠7日の体重 x_1 と妊娠18日の体重 y_1 および前後差 $(y_1 - x_1)$ の関連

層別散布図の作成は、Excelシート上で対照群のX軸の変数、Y軸の変数を選択し散布図を作りマークを「○」印に変更する。「データの選択」により処置群の変数を「追加」しマークを「+」印に変更する。それぞれの群を選択し、「近似曲線の追加」で「線形近似」を選択し、回帰式を重ね書きし、書式を整えた結果である。

体重の前後差 $(y_1 - x_1)$ に対する t 検定では、図8.2右に示した前値 x_1 の体重による前後差の大きさの違いを無視して“平均的”な前後差についての結果であり、そのために、図8.2右

から推測される前値 x_1 の体重が大きいほど前後差が拡大，いわゆる交互作用が存在している可能性を無視していいのかとの疑問が生ずる。

前値 x_1 と後値 y_1 がある場合に，前後差 $(y_1 - x_1)$ での群間比較，あるいは，増加率 $100 \times (y_1 - x_1) / x_1$ での群間比較か，どちらを選ぶか，両方なのか，悩ましい．表 8.2 に示したように群間で t 検定の結果を示すだけであれば，両方の結果を示すことは簡単である．図 8.3 に示す増加率での散布図では，図 8.2 右に示した前後差に比べ，右下がりになるが，全体の位置関係はほとんど同じである．

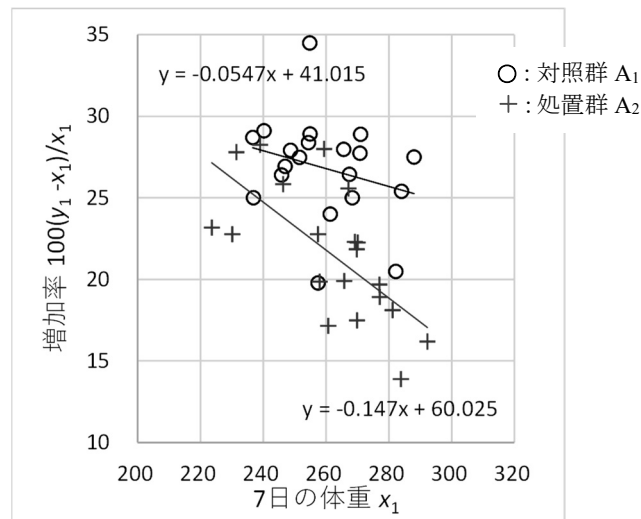


図 8.3 妊娠 7 日の体重 x_1 と投与後 18 日の体重 y_1 の増加率

前後差か増加率かの判断基準

前後差か，増加率か，どちらが望ましいのかについての選択は，図 8.4 に示すように前値 x_1 と後値 y_1 について個体ごとに線で結んだ線グラフを作成して判断できる．ほぼ平行であれば，前後差 $(y_1 - x_1)$ を選択すればよいのだが，前値 x_1 が広い範囲にあり，扇型に広がるような場合には，増加率 $(y_1 - x_1) / x_1$ を用いるべきである．なお，前値が相対的に狭い範囲にあり，扇型に広がるような場合には，増加率 $(y_1 - x_1) / x_1$ を取ることは，前値の誤差の変動が拡大するのでまったく推奨できないので，後値 y_1 そのものを使うべきである．

交絡変数・総摂餌量の影響

図 8.5 に示すのは，総摂餌量 x_2 と妊娠 18 日の体重 y_1 の関係，および，前後差 $(y_1 - x_1)$ との関係である．図 8.5 左に示すように，処置群 A_2 は，対照群 A_1 に比べ総摂餌量 x_2 が少ない方にシフトし，妊娠 18 日の体重 y_1 も下方にシフトしているにもかかわらず，同じ総摂餌量 x_2 で比較すると 18 日の体重 y_1 は，逆に大きくなっている．この原因は，図 8.5 右に示すように体重の前後差 $(y_1 - x_1)$ では，同じ総摂餌量 x_2 で比較すると処置群 A_2 は，前後差 $(y_1 - x_1)$ が小

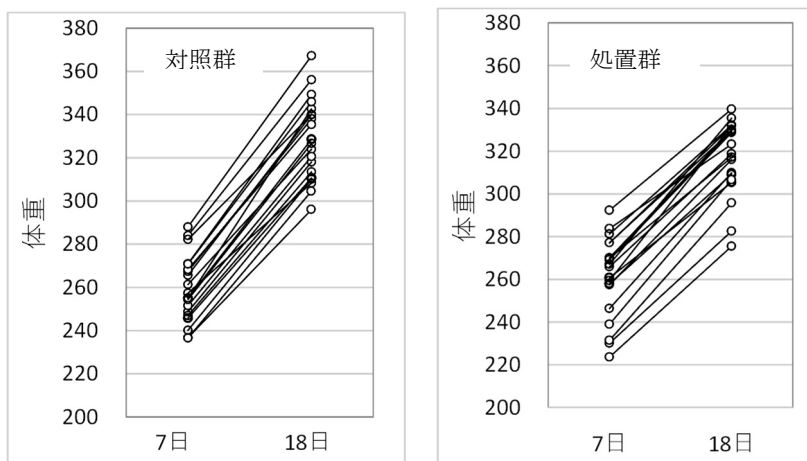


図 8.4 妊娠 7 日の体重 x_1 と投与後 18 日の体重 y_1 の線グラフ

Excel の折れ線グラフで、前値と後値の範囲を選択すると 2 本の折れ線が表示される。「データの選択」オプションで「行/列の切り替え」を選択すると 20 本の直線が表示されるので、書式を整えて作成した結果である。

さくなっている。このことから、妊娠 18 日の体重 y_1 に対して、体重の前値 x_1 と総摂餌量 x_2 が影響を与えていることになる。

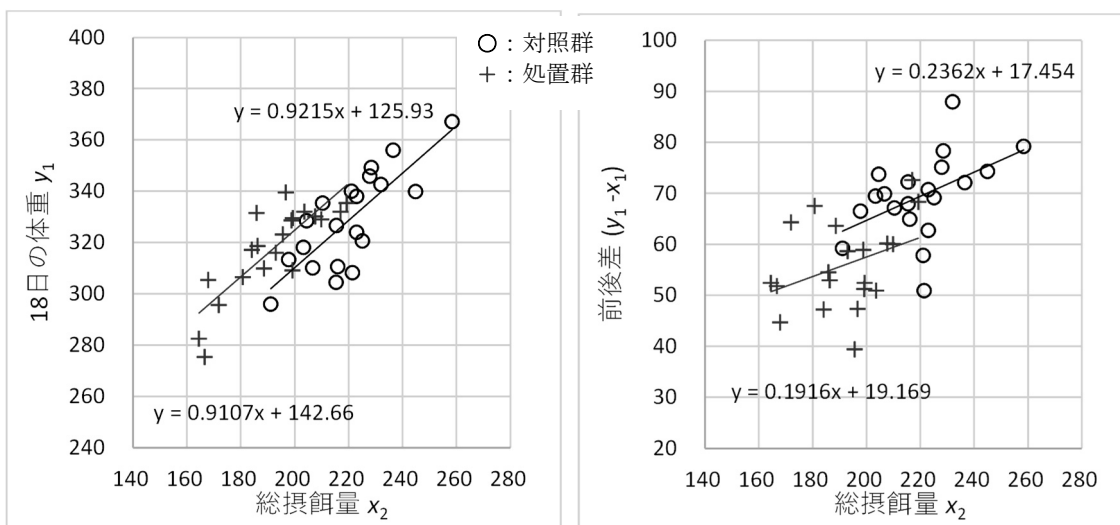


図 8.5 総摂餌量と妊娠 18 日の体重および前後差の関連

共変量・前値の影響

図 8.6 左に示すように妊娠 7 日の体重 x_1 に対する総摂餌量 x_2 は、処置群で体重に関連なく減少している。図 8.6 右に示す体重の前後差 ($y_1 - x_1$) を総摂餌量で割った摂餌効率 $(y_1 - x_1)/x_2$ では、処置群の体重が重い場合に摂餌効率が悪くなっているように思われる。以上の結果から、化合物の作用は、初期の体重 x_1 が大きい場合に餌の食べる量が減少し、その結果として体重の増加抑制が起きていると想定される。この実験データからでは、体重増加を抑制するような別の原因の探求はできないが、他の反復投与による毒性試験の結果から原因が示唆されるかも知れない。

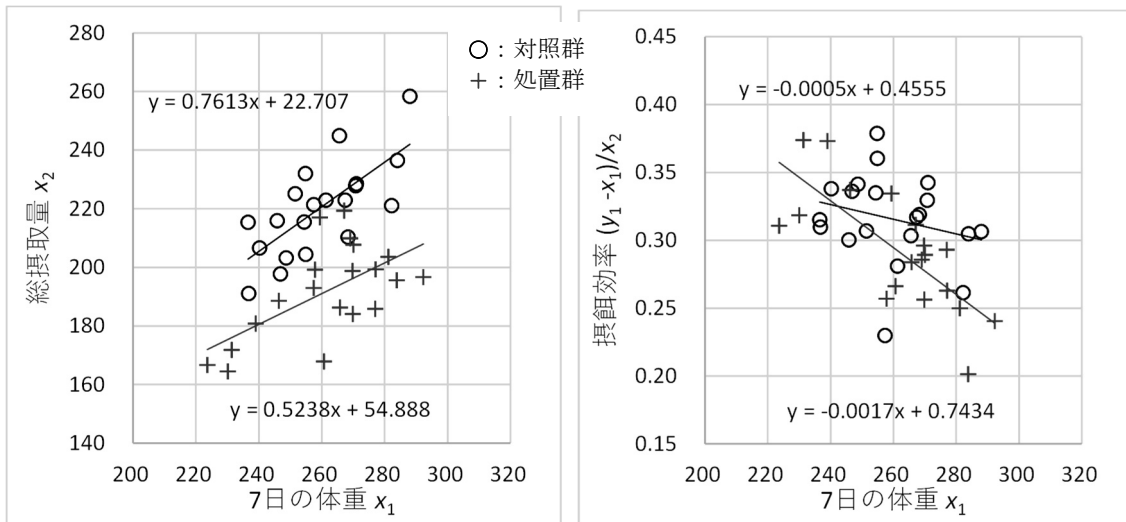


図 8.6 投与前 7 日の体重と総摂餌量および摂取効率

共変量・胎仔数の影響

出生仔数 y_2 には、表 8.2 で示したように群間で差が無いことから妊娠 7 日の胎仔数 x_3 とみなすことにする。図 8.7 に示すように、胎仔数 x_3 が体重の前後差 $(y_1 - x_1)$ に対してどのような影響があるか、散布図で確認する。胎仔数 x_3 の増加に伴い体重の前後差 $(y_1 - x_1)$ が増加していることが確認される。

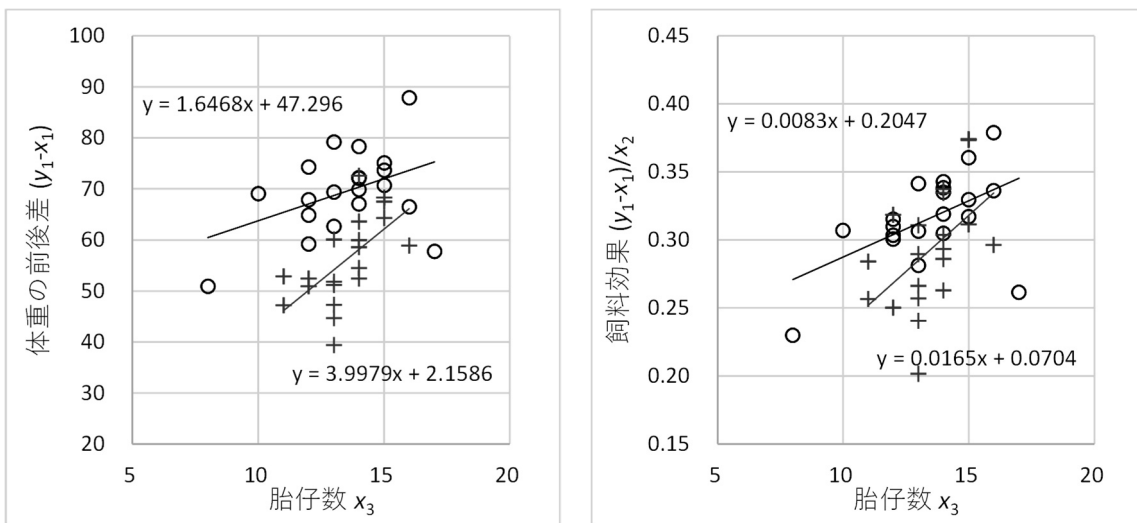


図 8.7 胎児数 x_3 : (y_2) と体重の前後差 $(y_1 - x_1)$ および飼料効率 $(y_1 - x_1) / x_2$ の関係

このように散布図に回帰直線を重ね書きすることにより、2 群間の t 検定の適用だけでは見えなかった関係を浮き彫りにすることができる。毒性試験で得られた膨大な測定変数に対して散布図を描いて考察することは現実的ではないが、何か疑問が生じた場合にその原因を掘り下げる時の最初の一步である。摂取効率 $(y_1 - x_1) / x_2$ に対する 2 群間の t 検定の結果は、 $p = 0.0610$ と微妙な結果であり、図 8.6 右に示すように、妊娠 7 日の体重 x_1 との交互作用が示唆されている。

8.2. Excelによる複数の共変量を含む探索的な回帰分析

前後差に対する共変量としての前値

2群間の t 検定を使った群間比較では、処置群 A_2 は、対照群 A_1 に対して体重の前後差 $(y_1 - x_1)$ から明らかな体重の増加抑制がある。しかし、飼料効率 $(y_1 - x_1)/x_2$ には明らかな差が2群間にないことから、「体重の増加抑制は、主として餌を食べないことが影響しているようである」との結論で良いのだろうか。第8.1節で示したように各種の層別散布図により検討を重ねても、なかなかすっきりとした結果が見えてこない。

そこで、対照群 A_1 に $a=1$ 、処置群 A_2 に対し $a=-1$ とする $(1, -1)$ 対比型ダミー変数を与え、前値の体重 x_1 (妊娠7日の体重)、総摂餌量 x_2 、胎仔数 x_3 (出生仔数 y_2 の読み替え) を共変量とする探索的な回帰分析を行なう。回帰分析に際しては、体重の後値 y_1 (妊娠18日の体重) についてのみを取り上げ、前後差 $(y_1 - x_1)$ については扱わない。これは、前後差 $(y_1 - x_1)$ を反応とした場合に、説明変数として前値 x_1 を含めた場合に、左辺の x_1 を右辺に移し整理すると

$$\left. \begin{aligned} y_1 - x_1 &= \theta_0 + \theta_1 a + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3 + \varepsilon \\ y_1 &= \theta_0 + \theta_1 a + (\theta_2 x_1 + x_1) + \theta_3 x_2 + \theta_4 x_3 + \varepsilon \\ &= \theta_0 + \theta_1 a + (\theta_2 + 1)x_1 + \theta_3 x_2 + \theta_4 x_3 + \varepsilon \end{aligned} \right\} \quad (8.2)$$

となり、前値 x_1 のパラメータが $(\theta_2 + 1)$ となるだけで、ダミー変数、他の共変量のパラメータの推定値が、後値 y_1 に対するパラメータの推定値と一致し、標準誤差も、 p 値も一致する。したがって、後値 y_1 に対する回帰分析の結果も、前後差 $(y_1 - x_1)$ による回帰分析の結果も共変量 x_1 に対する推定値が互いに ± 1 異なるだけで、どちらを用いても本質的な違いが無いためである。

複数の共変量との交互作用

交絡変数(以後、共変量として扱う)あるいは共変量を含む回帰分析を行なう際に、頭の痛い課題が存在する。それは、質的因子としての群 A と共変量間の交互作用が存在するかの確認である。質的因子と交絡変数および共変量との交互作用に対して、それぞれの平均からの差とする中心化を行ない

$$y_1 = \theta_0 + \theta_1 a + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3 + \theta_5 a(x_1 - \bar{x}_1) + \theta_6 a(x_2 - \bar{x}_2) + \theta_7 a(x_3 - \bar{x}_3) + \varepsilon \quad (8.3)$$

による検討を行う。なお、 $(1, -1)$ 対比型ダミー変数 a の期待平均は、 0 なので中心化を必要としない。交互作用が無視できるのであれば、主効果モデルとして

$$y_1 = \theta_0 + \theta_1 a + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3 + \varepsilon \quad (8.4)$$

による解析を行う。幸い質的因子は2水準なので交互作用モデルでは、Excelの回帰分析のパ

ラメータに対する t 検定の p 値が、分散分析の F 検定の p 値（タイプ III の平方和）に一致する。

表 8.3 に交互作用の検討のための Excel シートを示す。群 A については、(1, -1) 対比型ダミー変数とし、前値 x_1 、総摂取量を x_2 、胎仔数 x_3 とし、それらの総平均が (260.390, 205.983, 13.450) として示されている。交互作用についての中心化は、

$$\left. \begin{aligned} ax'_1 &= a(x_1 - \bar{x}_1) = a(x_1 - 260.390) \\ ax'_2 &= a(x_2 - \bar{x}_2) = a(x_2 - 205.983) \\ ax'_3 &= a(x_3 - \bar{x}_3) = a(x_3 - 13.450) \end{aligned} \right\} \quad (8.5)$$

によって計算した結果が示されている。

表 8.3 交互作用の検討のためのデザイン行列

番号	群 A	切片 x_0	群 a	前値	総摂	胎	交互作用			後値	前後差
				7日 x_1	餌量 x_2	胎仔数 x_3	(中心化)			18日 y_1	
							ax'_1	ax'_2	ax'_3		
1	A ₁	1	1	248.7	203.3	13	-11.69	-2.68	-0.45	318.1	69.4
2	A ₁	1	1	261.3	223.0	13	0.91	17.02	-0.45	324.0	62.7
:										:	
19	A ₁	1	1	240.2	206.7	14	-20.19	0.72	0.55	310.1	69.9
20	A ₁	1	1	245.8	216.0	12	-14.59	10.02	-1.45	310.7	64.9
21	A ₂	1	-1	257.9	199.2	13	2.49	6.78	0.45	309.1	51.2
22	A ₂	1	-1	260.7	167.9	13	-0.31	38.08	0.45	305.4	44.7
:										:	
39	A ₂	1	-1	239.0	180.8	15	21.39	25.18	-1.55	306.5	67.5
40	A ₂	1	-1	257.4	193.0	14	2.99	12.98	-0.55	316.0	58.6
総平均=				260.390	205.983	13.450					
				$x'_1=x_1-260.390$	$x'_2=x_2-205.983$	$x'_3=x_3-13.450$					

解析モデルに変数の積、あるいは、変数の 2 乗以上の変数を含める場合に用いる変数に対し、それらの平均値を差し引くことにより、多重共線性による行列計算の不具合を防ぐ必要がある。JMP の「モデルのあてはめ」でのデフォルトとなっている。この事例で、中心化を行なわないと、ダミー変数の推定値が正しく推定されない。これは、Excel の行列計算でも、回帰分析でも、JMP の「モデルのあてはめ」でも、SAS/GLM でも同様に不可解な結果となる。

表 8.3 に示した (40 行×8 列) のデザイン行列を用い、表 8.4 左に Excel の回帰分析による交互作用モデルの分散分析表およびパラメータの推定結果を示す。交互作用に関する変数 (ax'_1 , ax'_2 , ax'_3) について推定されたパラメータ (係数) に対する p 値は、(0.7734, 0.3941, 0.5272) であり、交互作用はまったくないとの結果である。表 8.4 右に 3 つの共変量との交互作用を除いた主効果モデルについて分散分析表およびパラメータの推定結果を示す。どの共変量も明らかな統計的な有意差があり、符号がプラスなので、妊娠 18 日の体重 y に対し、右上がりの関係となっている。対照群と処置群のダミー変数 a については、 $\hat{\theta}_1 = -0.0052$ とゼロに近く p 値も 1.0 に近く、妊娠 18 日の体重 y_1 に全く差が無いとの結果となっている。

表 8.4 交互作用および主効果モデルに対する分散分析表およびパラメータの推定

分散分析表 後値 y_1 交互作用モデル					分散分析表 後値 y_1 主効果モデル					
		自由度	変動	分散	分散比		自由度	変動	分散	分散比
	回帰	7	13104.31	1872.04	74.21	回帰	4	13064.19	3266.05	134.91
	残差	32	807.20	25.22		残差	35	847.32	24.2091	
	合計	39	13911.51			合計	39	13911.51		
		係数	標準誤差	t	P -値		係数	標準誤差	t	P -値
θ^{\wedge}_0	切片	30.6904	14.6666	2.0925	0.0444	切片	29.4769	13.4959	2.1841	0.0357
θ^{\wedge}_1	a	-0.3716	1.3904	-0.2673	0.7910	a	-0.0052	1.2513	-0.0042	0.9967
θ^{\wedge}_2	x_1	0.6175	0.0760	8.1277	0.0000	x_1	0.6437	0.0622	10.3406	0.0000
θ^{\wedge}_3	x_2	0.4754	0.0764	6.2212	0.0000	x_2	0.4515	0.0661	6.8258	0.0000
θ^{\wedge}_4	x_3	2.4319	0.5807	4.1882	0.0002	x_3	2.4529	0.4555	5.3852	0.0000
θ^{\wedge}_5	ax'_1	-0.0221	0.0760	-0.2904	0.7734					
θ^{\wedge}_6	ax'_2	0.0660	0.0764	0.8639	0.3941					
θ^{\wedge}_7	ax'_3	0.3712	0.5807	0.6393	0.5272					

複数の共変量を考慮した(2×2)格子状の散布図

共変量が一つの場合には、層別散布図に平行な直線を引くことにより、共分散分析の結果を可視化できる。複数の共変量がある場合には、どうしたら良いのであろうか。第 8.1 節では、様々な層別散布図に、それぞれの層に別々に回帰直線を重ね書きしたのであるが、このような方法では、複数の共変量による調整結果を図示できない。

そこで、3つの共変量の内、総摂餌量 x_2 を主たる共変量とし、前値（7日の体重） x_1 、胎仔数 x_3 を補助の共変量として、中央値（以下か、以上か）の 2 区分に分類し、さらに x_1 と x_3 の 2 区分内の平均値を求めて、後値（18日の体重） y_1 の推定値 \hat{y}_1 を重ね書きすることにより、可視化する。前値（7日の体重） x_1 の全 40 例の中央値は、261.0 なので、それ以下を $X_1^{(i)} = 1$ 、それ以上を $X_1^{(i)} = 2$ と 2 区分にし、 $X_1^{(i)} = 1$ となる区分内の x_1 を平均値は $\bar{x}_1^{(i)} = 246.7$ 、 $X_1^{(i)} = 2$ で $\bar{x}_1^{(i)} = 274.1$ に置き換える。胎仔数 x_3 の場合の中央値は、14.0 なので、それ未満を $X_3^{(i)} = 1$ 、それ以上を $X_3^{(i)} = 2$ とする。 $X_3^{(i)} = 1$ となる区分内の x_3 を $\bar{x}_3^{(i)} = 12.1$ に、 $X_3^{(i)} = 2$ の場合は $\bar{x}_3^{(i)} = 14.7$ に置き換える。

表 8.3 に示した元データに対し、 $X_1^{(i)}$ および $X_3^{(i)}$ の 2 区分変数を追加し、Excel のソート機能を使い、 $(X_1^{(i)}, X_3^{(i)}, A, x_2)$ の順番で並び替える。このような手順により、表 8.4 の結果が得られる。前値（7日の体重） x_1 は、 $X_1^{(i)}$ の (1, 2) に対応して、(246.7, 274.1) に置き換えられ、胎仔数 x_3 は、 $X_3^{(i)}$ の (1, 2) に対応して、(12.1, 14.7) に置き換えられている。置き換えられた変数を含めて変数名を別名 (l_0, l_1, l_2, l_3, l_4) とし与え、表 8.3 で推定されたパラメータ

$$\hat{\theta} = [29.4769 \quad -0.0052 \quad 0.6437 \quad 0.4515 \quad 2.4529]^T$$

を用いて、後値（18日の体重）の推定値 \hat{y}_1 を $\hat{y}_1 = l\hat{\theta}$ により推定している。

表 8.5 前値 x_1 ・胎仔数 x_3 の区分別の共変量を反映した後値 y_1 推定値

前値群	胎仔群	投与群	切片番号	投与群	前値7日の群平均	総摂餌量	胎仔数群平均	回帰パラメータ推定値	後値18日の推定値	後値18日の体重		
$X_1^{(i)}$	$X_3^{(i)}$	A	号	l_0	l_1	l_2	l_3	l_4	θ^{\wedge}	y_1^{\wedge}	y_1	
x_0	a	x_1	x_2	x_3	θ^{\wedge}	y_1^{\wedge}	y_1					
1	1	A ₁	11	1	1	246.7	191.1	12.1	29.4769	304.1	296.0	
			1	1	1	246.7	203.3	12.1	-0.0052	309.6	318.1	
			14	1	1	246.7	215.4	12.1	0.6437	315.1	304.5	
			20	1	1	246.7	216.0	12.1	0.4515	315.3	310.7	
			8	1	1	246.7	221.4	12.1	2.4529	317.8	308.3	
			16	1	1	246.7	225.1	12.1		319.5	320.6	
	A ₂	34	1	-1	246.7	164.5	12.1		292.1	282.5		
		35	1	-1	246.7	166.7	12.1		293.1	275.4		
		22	1	-1	246.7	167.9	12.1		293.6	305.4		
		21	1	-1	246.7	199.2	12.1		307.8	309.1		
		2	A ₁	13	1	1	246.7	197.8	14.7		313.6	313.4
	5	1		1	246.7	204.5	14.7		316.7	328.6		
	19	1		1	246.7	206.7	14.7		317.6	310.1		
	18	1		1	246.7	215.5	14.7		321.6	326.6		
	15	1		1	246.7	232.0	14.7		329.1	342.7		
	A ₂	38		1	-1	246.7	171.9	14.7		301.9	295.7	
39		1		-1	246.7	180.8	14.7		306.0	306.5		
33		1		-1	246.7	188.6	14.7		309.5	309.9		
40		1		-1	246.7	193.0	14.7		311.5	316.0		
32		1		-1	246.7	217.0	14.7		322.3	332.0		
2	1	A ₁	2	1	1	274.1	223.0	12.1		336.1	324.0	
			3	1	1	274.1	244.9	12.1		346.0	339.9	
			12	1	1	274.1	258.4	12.1		352.1	367.2	
			A ₂	29	1	-1	274.1	184.1	12.1		318.6	317.1
				28	1	-1	274.1	186.3	12.1		319.6	318.7
				24	1	-1	274.1	195.6	12.1		323.8	323.2
	26	1	-1	274.1	196.7	12.1		324.3	339.6			
	25	1	-1	274.1	203.6	12.1		327.4	332.1			
	27	1	-1	274.1	207.7	12.1		329.2	330.2			
	2	A ₁	10	1	1	274.1	210.4	14.7		337.0	335.4	
			9	1	1	274.1	221.1	14.7		341.8	340.0	
			6	1	1	274.1	223.0	14.7		342.7	338.1	
			17	1	1	274.1	227.9	14.7		344.9	345.9	
			7	1	1	274.1	228.5	14.7		345.2	349.3	
			4	1	1	274.1	236.5	14.7		348.8	356.1	
		A ₂	23	1	-1	274.1	185.9	14.7		326.0	331.5	
			30	1	-1	274.1	198.8	14.7		331.8	328.7	
			36	1	-1	274.1	199.3	14.7		332.0	329.5	
37			1	-1	274.1	209.9	14.7		336.8	329.1		
31			1	-1	274.1	219.3	14.7		341.0	335.4		

総摂餌量 x_2 が, $(X_1^{(i)}, X_3^{(i)}, A)$ の中で小さい順にソートしてあるのは, Excel の散布図上で推定値 \hat{y}_1 を回帰直線として描くためである. 区分 $(X_1^{(i)}=1, X_3^{(i)}=1, A=A_1)$ の最小値 $x_2^{(11)}=191.1$ に対して, $\mathbf{l}^{(11)}=[1 \ 1 \ 246.7 \ 191.1 \ 12.1]$ なので $\hat{y}_1^{(11)}=\mathbf{l}^{(11)}\hat{\boldsymbol{\theta}}=304.1$ が得られている. 同じ区分の最大値 $x_2^{(16)}=225.1$ に対しては, $\hat{y}_1^{(16)}=\mathbf{l}^{(16)}\hat{\boldsymbol{\theta}}=319.5$ が得られている.

表 8.4 の右端には, 後値の体重 y_1 が示されている. この区分の 6 個のデータ図 8.8 の左上側の散布上に書き込む. 後値の体重 y_1 は, デフォルトのマークを ○ 印とし, 推定値 \hat{y}_1 は, 回

帰直線としている. この直線の切片 $\hat{\beta}_0$ は, $x_2=0$ の場合であり, $\mathbf{l}_{(x_2=0)}^{(1)}=[1 \ 1 \ 246.7 \ 0 \ 12.1]$ なので $\hat{\beta}_0 = \mathbf{l}_{(x_2=0)}^{(1)} \hat{\boldsymbol{\theta}} = 217.83$ となる. 傾き $\hat{\beta}_1$ は, 表 8.5 から $\hat{\boldsymbol{\theta}}_3 x_2 = 0.4515x_2$ なので,

$$(X_1^{(i)}=1, X_3^{(i)}=1, A=A_1) : \hat{y}_1^{\text{対照}} = 217.83 + 0.4515x_2 \quad (8.6)$$

となる. 処置群 A_2 の場合は, $a=-1$ とした $\mathbf{l}_{(a=-1)}^{(1)}=[1 \ -1 \ 246.7 \ 191.1 \ 12.1]$ を用いて

$$(X_1^{(i)}=1, X_3^{(i)}=1, A=A_2) : \hat{y}_1^{\text{処置}} = 217.84 + 0.4515x_2 \quad (8.7)$$

となる. 切片が, わずか 0.01 異なるだけで, 2本の直線は, 完全に重なっているように見える.

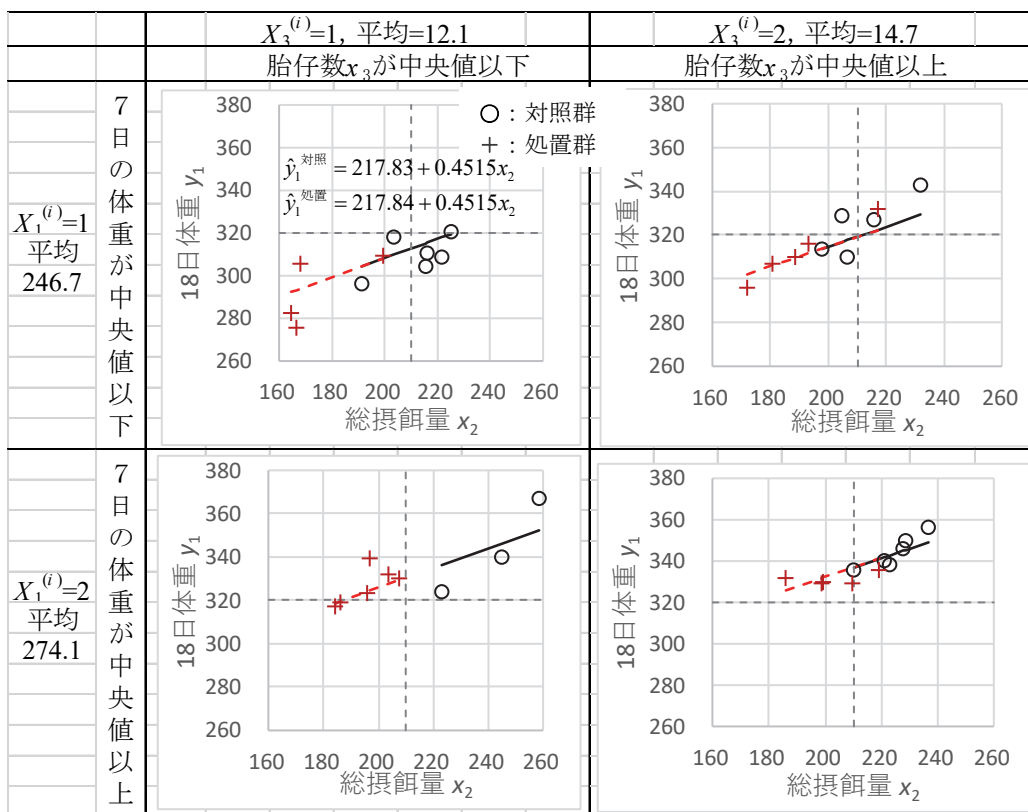


図 8.8 2×2 区分の総摂餌量に対する 18 日体重の散布図

図 8.8 右上側は, 胎仔数 x_3 が中央値以上の場合であり, 図 8.8 左上側に比べて, 分布が右上に少々シフトしている. 図 8.8 下側は, 前値 (7 日の体重) が中央値以上の場合であり, 後値 (18 日の体重) が増加する方向にシフトしていることが観察される.

このような 2×2 の格子状の散布図にすることにより, 複数の共変量が絡み合っている場合に, 処置群 A_2 に対する体重の増加抑制が, 体重に対する直接的な作用ではなく, 交絡変数である総摂餌量の抑制による間接的な体重減少と判断される.

8.3. JMP による複数の共変量を含む探索的な回帰分析

ステップワイズ法による変数選択

JMP の「モデルのあてはめ」には、「標準最小 2 乗法」に加えて「ステップワイズ法」も備わっており、(変数増加・変数減少・変数増減)を自動的に行なわせる機能だけでなく、 p 値を見ながら(モデルに含める変数・除く変数)について手動で取舍選択する機能を備えている。また、変数選択を自動的に行なわせる際に、第 7.4 節で示したように主効果と交互作用をセットにして扱う「組み合わせ」のルールがデフォルトで設定されている。

表 8.6 に示すのは、JMP ファイルに Excel のシート上のデータをペーストし、JMP 新規データファイル上に「列名と共に貼り付け」て変数名を整えた結果を示す。下段のそれぞれの変数に対する平均値の計算は、JMP の「表の作成」で計算したもので、Excel のピボットテーブルに対応している。

表 8.6 解析用の JMP データファイル

	番号	群A	前値7日 x_1	総摂餌量 x_2	胎仔数 x_3	後値18日 y_1
○	1	1 A1	248.7	203.3	13	318.1
○	2	2 A1	261.3	223.0	13	324.0
○	3	3 A1	265.6	244.9	12	339.9
+	38	38 A2	231.4	171.9	15	295.7
+	39	39 A2	239.0	180.8	15	306.5
+	40	40 A2	257.4	193.0	14	316.0

群A		前値7日 x_1	総摂餌量 x_2	胎仔数 x_3	後値18日 y_1
A1	A2	平均	平均	平均	平均
20	20	260.39	205.983	13.45	323.08

群 A と 3 つの共変量の交互作用を含むモデルを JMP の「モデルのあてはめ」で設定し、「手法」を「ステップワイズ法」に切り替える。モデルの設定で、主効果として(「群 A」, 「前値 7 日 x_1 」, 「総摂餌量 x_2 」, 「胎仔数 x_3 」), 交互作用として(「群 A*前値 7 日 x_1 」, 「群 A*総摂餌量 x_2 」, 「群 A*胎仔数 x_3 」)を設定する。ただし、表 8.7 に示すように、ステップワイズ法の設定の画面に現れる変数名は、(「群 A」 \rightarrow 「群 A{A2-A1}」)となり、何らかのダミー変数に変換されていることが示唆されている。「群 A」と共変量の交互作用については、それぞれの総平均を差し引いた“中心化”が、内部の変数として使われていることが示されている。

ステップワイズ法には、幾つかの方法があり、「停止ルール」を「閾値 p 値」, 「方向」を, 「変数増減」, 「ルール」には, 「組み合わせ」を選択している.

表 8.7 に示すのは, 「群 A $\{A_2 - A_1\}$ 」を追加し, ロックを掛け, 「前値 7 日 x_1 」を手動で追加した結果である. 推定された式は,

$$y_1 = 76.4157 - 6.7016 \times \text{「群 A } \{A_2 - A_1\} \text{」} + 0.9473 \times \text{「前値 7 日 } x_1 \text{」} \quad (8.8)$$

であり, 「群 A $\{A_2 - A_1\}$ 」の推定値がマイナスなので, 図 8.2 左の回帰直線の関係から内部で生成されたダミー変数は, $(-1, 1)$ 対比型ダミー変数が設定されていると判断される. また, 「群 A $\{A_2 - A_1\}$ 」の p 値が $1.36E-05$ と高度に有意であり, 共変量として「前値 7 日 x_1 」を含めた場合に, 対照群 A_1 と処置群 A_2 の「後値 18 日体重 y_1 」に明らかな差があるとの結論となる.

表 8.7 ステップワイズ回帰による「群 A」と共変量「前値 7 日 x_1 」の選択の結果

後値18日 y_1 のステップワイズ										
ステップワイズ回帰の設定										
停止ルール:	<input type="text" value="閾値p値"/>									
	変数を追加するときのp値	0.25	<input type="button" value="→"/>	<input type="button" value="すべて追加"/>	<input type="button" value="モデルの作成"/>					
	変数を除去するときのp値	0.25	<input type="button" value="←"/>	<input type="button" value="すべて削除"/>	<input type="button" value="モデルの実行"/>					
方向:	<input type="text" value="変数増減"/>									
ルール:	<input type="text" value="組み合わせ"/>									
SSE	DFE	RMSE	R2乗	自由度調整R2乗	Cp	p	AICc	BIC		
2635.1242	37	8.4391689	0.8106	0.8003	70.464931	3	290.1702	295.7828		
現在の推定値										
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	"p値(F)"			
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	76.4157	1	0.0000	0.0000	1			
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	群A $\{A_2-A_1\}$	-6.7016	1	1789.2339	25.1228	1.36e-5			
<input type="checkbox"/>	<input checked="" type="checkbox"/>	前値7日 x_1	0.9473	1	9977.9253	140.1009	3.8e-14			
<input type="checkbox"/>	<input type="checkbox"/>	総摂餌量 x_2	0.0000	1	1085.7412	25.2273	1.4e-5			
<input type="checkbox"/>	<input type="checkbox"/>	胎仔数 x_3	0.0000	1	659.8486	12.0259	0.00138			
<input type="checkbox"/>	<input type="checkbox"/>	群A $\{A_2-A_1\}$ *(前値7日 x_1 -260.39)	0.0000	1	219.0111	3.2633	0.07921			
<input type="checkbox"/>	<input type="checkbox"/>	群A $\{A_2-A_1\}$ *(総摂餌量 x_2 -205.983)	0.0000	1	2.7498	0.0376	0.84733			
<input type="checkbox"/>	<input type="checkbox"/>	群A $\{A_2-A_1\}$ *(胎仔数 x_3 -13.45)	0.0000	1	102.3277	1.4544	0.23568			

表 8.7 ステップワイズ回帰による「群 A」と共変量「前値 7 日 x_1 」の選択の結果「総摂餌量 x_2 」および「胎仔数 x_3 」の p 値が, $(1.4E-05, 0.00138)$ であることから, p 値の小さい「総摂餌量 x_2 」をモデルに追加する. 表 8.8 に示すように, 追加後「群 A $\{A_2 - A_1\}$ 」の推定値および p 値が大きく変化し, 有意な差ではなくなっている. さらに, 交互作用を含めて p 値が最も小さいのが, 「胎仔数 x_3 」であり, $p=4.99E-06$ であることが読み取れる.

表 8.8 ステップワイズ回帰による「総摂餌量 x2」を追加した結果結果

現在の推定値								
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	"p値(F)"	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	56.3873	1	0.0000	0.0000	1	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	群A{A2-A1}	-0.1493	1	0.3446	0.0080	0.92919	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	前値7日x1	0.6739	1	2860.4603	66.4630	1.08e-9	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	総摂餌量x2	0.4428	1	1085.7412	25.2273	1.4e-5	
<input type="checkbox"/>	<input type="checkbox"/>	胎仔数x3	0.0000	1	702.0632	28.9999	4.99e-6	
<input type="checkbox"/>	<input type="checkbox"/>	群A{A2-A1}*(前値7日x1-260.39)	0.0000	1	90.4582	2.1701	0.14965	
<input type="checkbox"/>	<input type="checkbox"/>	群A{A2-A1}*(総摂餌量x2-205.983)	0.0000	1	2.7498	0.0622	0.80447	
<input type="checkbox"/>	<input type="checkbox"/>	群A{A2-A1}*(胎仔数x3-13.45)	0.0000	1	14.7628	0.3367	0.56546	

表 8.9 に示すように「胎仔数 x_3 」をさらに加えると、交互作用の p 値は全て 0.25 以上となり、追加する「変数」ないと判断される。この結果は、表 8.4 で示した Excel での結果に一致する。なお、自由度がすべて 1なのは、交互作用がモデルに含まれていないためである。

表 8.9 ステップワイズ回帰による「胎仔数 x3」を追加した結果結果

SSE	DFE	RMSE	R2乗	自由度調整R2乗	Cp	p	AICc	BIC
847.31973	35	4.920278	0.9391	0.9321	3.5905223	5	250.1885	257.7763

現在の推定値								
ロック	追加	パラメータ	推定値	自由度	平方和	"F値"	"p値(F)"	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	切片	29.4769	1	0.0000	0.0000	1	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	群A{A2-A1}	0.0052	1	0.0004	0.0000	0.99669	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	前値7日x1	0.6437	1	2588.6242	106.9276	3.5e-12	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	総摂餌量x2	0.4515	1	1127.9559	46.5922	6.42e-8	
<input type="checkbox"/>	<input checked="" type="checkbox"/>	胎仔数x3	2.4529	1	702.0632	28.9999	4.99e-6	
<input type="checkbox"/>	<input type="checkbox"/>	群A{A2-A1}*(前値7日x1-260.39)	0.0000	1	9.3823	0.3807	0.54134	
<input type="checkbox"/>	<input type="checkbox"/>	群A{A2-A1}*(総摂餌量x2-205.983)	0.0000	1	28.1522	1.1685	0.28732	
<input type="checkbox"/>	<input type="checkbox"/>	群A{A2-A1}*(胎仔数x3-13.45)	0.0000	1	14.7628	0.6029	0.44285	

JMP の「ステップワイズ法」は、自動的に変数選択を行なう機能もあるが、ここに示したように手作業により変数を取捨選択する機能に優れている。ただし、この結果をもって探索的な解析が終わったわけではない。「ステップワイズ法」での出力は、(推定値, 平方和, F 値, p 値)のみであり、群 A の各水準の最小 2 乗平均などの推定値の計算が行なわれていないので、第 7 章に示したように、解析結果を考察するためには、種々の推定と結果のグラフ表示が欠かせない。JMP の「ステップワイズ法」は、変数選択に特化しているので、更なる解析のためには、表 8.7 右上の **モデルの作成** または **モデルの実行** を選択することにより、「モデルのあてはめ」の基本である「標準最小 2 乗」が自動的に起動される。

「標準最小2乗法」による予測プロファイルの活用

「標準最小2乗」による解析結果を表 8.10 に示す。分散分析表の「誤差の平方和=847.3197」が、表 8.9 の SSE に該当し、RMSE=4.9203 は、表 8.10 の「誤差の平均平方=24.2091」の平方根に一致している。「パラメータ推定値」の推定値と p 値が、表 8.9 に一致し、表 8.4 に示した Excel の回帰分析による主効果モデルにも一致している。細かいことであるが、群 A[A1] のパラメータ推定値は、-0.0052 と表 8.9 の 0.0052 と異なる。これは、「標準最小2乗法」でのタミー変数が、(-1, 1) 対比型ではなく (1, -1) 対比型ダミー変数に切り替わっていることによる。

表 8.10 主効果モデルによる解析結果

分散分析					
要因	自由度	平方和	平均平方	F値	p値(Prob>F)
モデル	4	13064.1900	3266.0475	134.9097	
誤差	35	847.3197	24.2091		
全体(修正済み)	39	13911.5098			<.0001*

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	29.4769	13.4959	2.1841	0.0357*
群A[A1]	-0.0052	1.2513	-0.0042	0.9967
前値7日x1	0.6437	0.0622	10.3406	<.0001*
総摂餌量x2	0.4515	0.0661	6.8258	<.0001*
胎仔数x3	2.4529	0.4555	5.3852	<.0001*

効果の検定					
要因	自由度	平方和	平均平方	F値	p値(F)
群A	1	0.0004	0.0004	0.0000	0.9967
前値7日x1	1	2588.6242	2588.6242	106.9276	<.0001*
総摂餌量x2	1	1127.9559	1127.9559	46.5922	<.0001*
胎仔数x3	1	702.0632	702.0632	28.9999	<.0001*

図 8.9 に示す予測プロファイルは、表 8.10 で推定された主効果モデルのパラメータの推定値に対する 95%信頼区間、推定された回帰直線とその 95%信頼区間が示されている。左端に示されている「323.1 [320.1, 326.1]」は、「群 A」の A1 水準の a=1、「前値 7 日 x₁」の平均値 260.39、「総摂餌量 x₂」の平均値 205.982、「胎仔数 x₃」の平均値 13.45 を式 (8.9) に代入した推定値（最小 2 乗平均）と 95%信頼区間である。

$$\left. \begin{aligned}
 \hat{y}_1 &= \hat{\theta}_0 x_0 + \hat{\theta}_1 a + \hat{\theta}_2 x_1 + \hat{\theta}_3 x_2 + \hat{\theta}_4 x_3 \\
 &= 29.4769 x_0 + (-0.0052) a + 0.6437 x_1 + 0.4515 x_2 + 2.4529 x_3 \\
 &= 323.1
 \end{aligned} \right\} \quad (8.9)$$

JMP の「予測プロファイル」は、GUI 操作により動的に変化する。群 A については「A1」

が選択されているが、マウスで「A2」を選択すると、95%信頼区間が微妙に変化することが実感される。共変量 (x_1, x_2, x_3) は、それぞれの平均値がセットされているが、マウス操作で自由に変更することができ、それに伴い、群 A の推定値も動的に変化する。

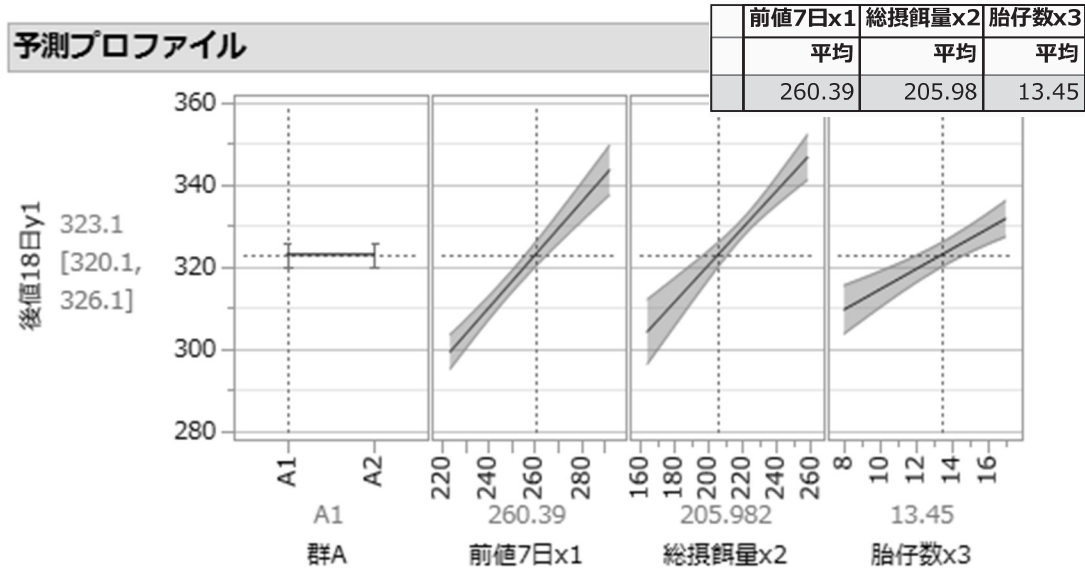


図 8.9 推定値に対する予測プロファイルによる結果の表示

表 8.11 に示すのは、共変量 (x_1, x_2, x_3) を平均値に固定したときの、群 A の各水準の最小 2 乗平均、「A₁ : 323.072, A₂ : 323.083」が推定されている。なお、この「最小 2 乗平均」は、図 8.9 左端に示される推定値にも対応している。表 8.11 の右端の「平均」は、群 A の各水準の算術平均であり、表 8.1 の「体重 18 日」についての「平均」に一致する。これに対し、「最小 2 乗平均」は、式 (8.9) に示したように表 8.10 の回帰分析で得られたパラメータ $\hat{\theta}$ と、係数ベクトル $l = [1 \ a \ \bar{x}_1 \ \bar{x}_2 \ \bar{x}_3]$ の線形和 $L = l\hat{\theta}$ により推定されている。線形和 $L = l\hat{\theta}$ の標準誤差 SE は、 $\Sigma(\hat{\theta})$ をパラメータの共分散行列したときに、 $SE = \sqrt{l\Sigma(\hat{\theta})l^T}$ によって求められている。

表 8.11 群 A についての最小 2 乗平均

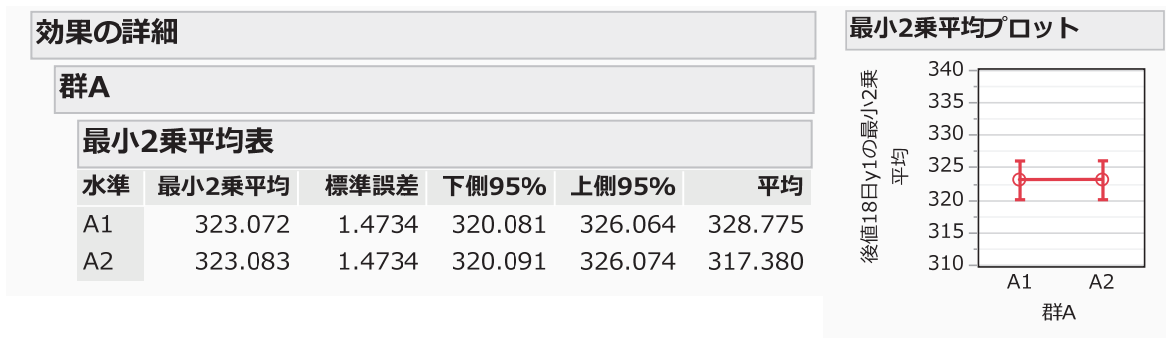


図 8.10 最小 2 乗平均プロット

8.4. 交絡変数の多彩な影響への対応

化合物の多面的な毒性を把握するために、反復投与毒性試験が行なわれている。医薬品の場合は、げっ歯類に対して1カ月、3カ月、6カ月の反復投与が規定されている。本節で取り上げるのはラットを対象にした3カ月の反復投与終了後の測定された体重および臓器重量に対する解析法を取り上げる。

吉村編著（1987）、「毒性・薬効データの統計解析，第3.3節 臓器重量の解析における体重データの利用」には、臓器重量と体重の関係の6通りの典型例が示されており、現実に得られたデータに対し、どのような解析法の選択が望ましいのかを前もって検討するための事例として示されている。さらに、高橋ら（1989）、「SASによる実験データの解析，第7章 共分散分析」では、SASを用いた詳細な解析方法が示されている。これらを踏まえて、ExcelおよびJMPを主体にした解析法を示す。

投与終了時の体重は交絡変数

図8.11に示すように反復投与終了時に体重 X の大きい個体は、一般的に臓器重量 Y も大きくなる。化合物Aが、体重の増加を抑制すると、それに伴い臓器の重量 Y の増加が抑制される。このときに体重 X を考慮しないで臓器の重量 Y だけを見ると、臓器に対して化合物Aの直接の作用がないのに、臓器に直接的な作用があったと誤解してしまう。このような場合に、本当の原因がどれであるか分からなくなる現象を原因の“交絡” **confounding** という。

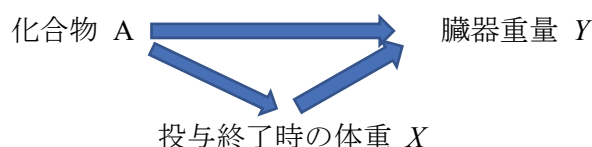


図 8.11 投与終了時の体重が臓器重量におよぼす影響

交絡があると、知りたいことについて正しい判断ができなくなる。図8.12に臓器重量 $y^{(i)}$ と体重 $x^{(i)}$ の種々の関連について人工データ（②組は心重量についての実データ）に基づいた6種類の散布図を例示する。

毒性試験における臓器重量の統計解析は、対照群と投与群間での臓器の絶対重量での比較、臓器重量を体重で除した体重比（相対重量）での比較が主体であり、共分散分析をどのような場合に使うかの検討がなされている。交絡変数としての体重 X と臓器重量 Y の關係に、

- i) 相関があるか
- ii) 2群の回帰直線が同一線上か

- iii) 直線が原点を通るか
- iv) 2群の直線の傾きは同じか

などを判断基準とし、表 8.12 に示すように ③ 組と ⑤ 組の場合に共分散分析が適用できると示唆している。⑥ 組のように、傾きが異なり、原点を通らない場合は、臓器の絶対重量と

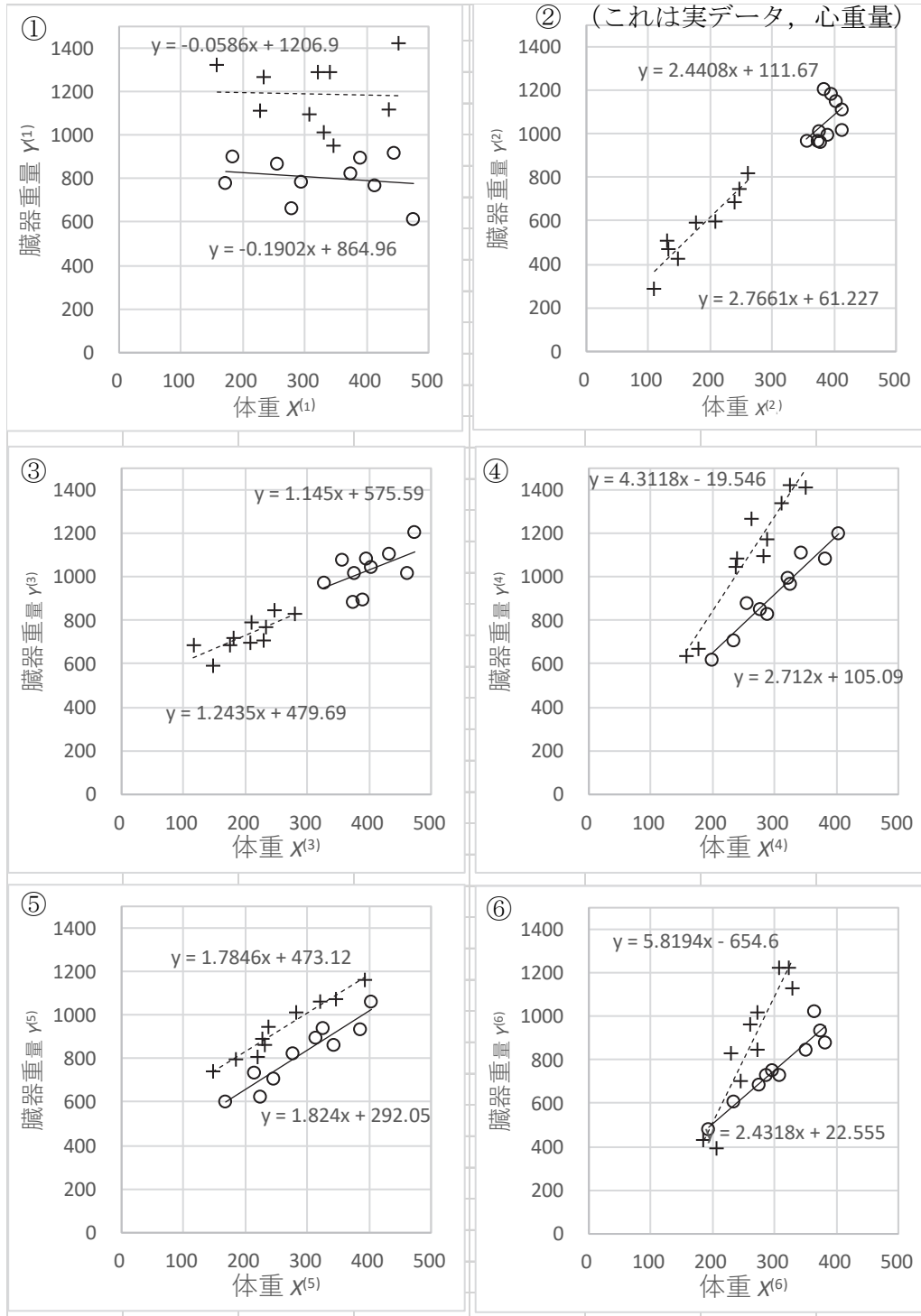


図 8.12 臓器重量と体重の関係の 6 通りの典型例

○印：対照群，＋印：処理群

回帰直線の差の比較を推奨している。このことから、共分散分析は、処置 A と共変量 X の交互作用がないことを前提にしている。② 組と ③ 組の場合は、絶対重量と体重比が、① 組の場合は、絶対重量と体重を別々に解析することが推奨されている。

表 8.12 体重の有意差が臓器重量の有意差に与える影響 [吉村編著 (1987)]

場合 図8.14 に対応	体重データ での有意差	臓器重量 での有意差	臓器への 直接影響	説明
1	無	有	有	臓器重量での解析結果が重要
2	有	有	無	体重比での解析結果が重要
3	有	有	無	共分散分析での解析が必要
4	無	無	有	体重比での解析結果が重要
5	無	無	有	共分散分析での解析が必要
6	無	無	有	回帰直線の差の検定が必要

さて、i) ~ iv) の設問での「相関があるか」については、 X と Y の相関係数を計算して判断すればよいこと分かるが、ii) ~ iv) の設問については、どのような計算をしたらよいのだろうか。図 8.12 から視覚的な判断に基づいて表 8.12 が作成されているが、Excel を用いた計算結果に基づいて検討し直したい。

図 8.12 の元になるデータは、高橋ら (1989)、図表 7.1 に掲載されているので、表 8.13 に引用する。図 8.12 に示したのは、表 8.13 のデータに基づき Excel の散布図によって作成し直し

表 8.13 臓器重量 X_i と体重 Y_i のデータリスト [高橋ら (1989)、図表 7.1]

処置 A	No.	① 組		② 組		③ 組		④ 組		⑤ 組		⑥ 組	
		$X^{(1)}$	$Y^{(1)}$	$X^{(2)}$	$Y^{(2)}$	$X^{(3)}$	$Y^{(3)}$	$X^{(4)}$	$Y^{(4)}$	$X^{(5)}$	$Y^{(5)}$	$X^{(6)}$	$Y^{(6)}$
1	1	255	869	355	969	355	1080	255	881	245	710	285	731
1	2	412	771	412	1111	432	1107	382	1082	224	627	382	882
1	3	443	917	413	1017	460	1015	320	995	342	865	350	845
1	4	183	904	383	1204	473	1204	403	1201	403	1064	193	481
1	5	374	826	374	966	374	886	234	707	214	736	234	607
1	6	475	613	375	1013	375	1020	325	970	325	940	275	685
1	7	294	785	394	1185	394	1085	198	621	384	935	364	1021
1	8	278	663	378	963	326	972	276	852	276	822	296	752
1	9	388	897	388	997	388	897	288	831	168	603	308	731
1	10	172	782	403	1150	403	1047	343	1112	313	897	373	932
2	11	347	953	247	744	247	844	287	1174	347	1074	322	1224
2	12	320	1290	110	290	210	790	240	1082	220	806	229	832
2	13	435	1117	260	817	280	827	282	1097	282	1013	272	847
2	14	158	1323	149	423	117	683	157	635	227	893	207	395
2	15	340	1288	240	688	175	688	325	1423	185	798	245	703
2	16	450	1424			181	721	311	1341	231	861	186	431
2	17	233	1269	133	469	233	769	263	1269	393	1162	273	1019
2	18	228	1111	178	591	148	591	178	671	320	1060	308	1221
2	19	330	1012	130	510	230	710	350	1410	148	741	260	960
2	20	308	1097	208	597	208	697	238	1047	238	947	328	1127

たものである。処置 A は、1 が対照群、2 が処置群、 $(X^{(1)}, Y^{(1)})$ が ① 組の臓器重量と体重であり、順次 ⑥ 組のデータが示されている。きれいな散布図を短時間で作成することは、統計解析の基礎の基礎でもある。

図 8.12 に示した ① 組の散布図は、対照群 $A_1^{(1)}$ の 10 例について散布図を描き、処置群 $A_1^{(1)}$ の 10 例を「データの選択」で重ね書きし書式を整え、さらに、「近似曲線の追加」で回帰直線を重ね書きする。②～⑥ 組は、① 組全体をコピーし、所定の位置にペーストする。散布図上の対照群 $A_1^{(1)}$ のマークを選択すると該当するデータ範囲がハイライトされるので所定の位置に範囲を移動し、 $A_2^{(1)}$ も同様に移動すると、回帰直線も自動的に更新される。

解析手法の選択のための決定樹

どのような解析手法を選択したら良いのか、図 8.13 に示す「決定樹」に表すと見通しがよくなる。

設問 1: X と Y に相関ある?

設問 3: 2 群の直線の傾きは同じか?

設問 2: 2 群が同一直線上か?

設問 4: 原点を通るか?

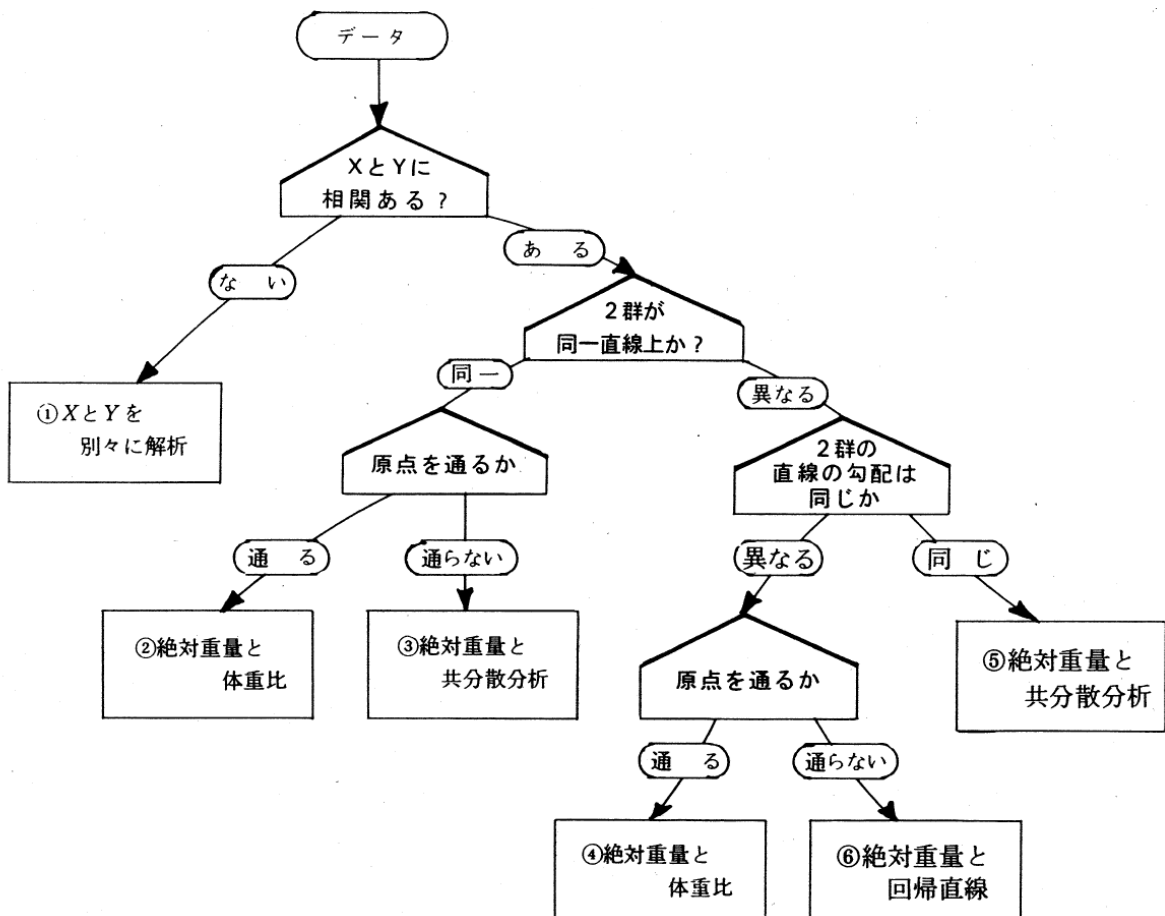


図 8.13 臓器重量の解析のための決定樹 [吉村編著 (1987), 図 3-5 を引用]

この決定樹からも、臓器重量と体重に対する解析は、常に探索的な解析になることが示されている。(相関があるか、原点を通るか)については、一般的な解析法であるが、(2群が同一直線上か、2群の直線の傾きは同じか)は、体重を共変量にした回帰分析が必要になる。しかし、吉村編著(1987)には、図の例示のみでデータよび解析結果は示されず、別の事例について、偏差平方和をベースにした伝統的な共分散分析の方法が示されているのみである。

1) XとYに相関があるか

視覚的には、「相関があるか」に対し①組が該当しないが、②組の対照群A₁も該当しないように思われる。表8.14に示すように、群ごとに2変数の相関係数を求めるのは、Excelシート上でCorrel()関数を使うのが効率的である。A₁の①組について、

$$r^{(1,A_1)} = \text{Correl}(X^{(1,A_1)} \text{の範囲}, Y^{(1,A_1)} \text{の範囲}) = -0.1976 \quad (8.10)$$

として計算し、その計算式をコピーし他の組み合わせ位置にペースする。(Xの範囲、Yの範囲)がずれるので、計算式をマウスで選択しハイライトされたデータ範囲を適切な位置にマウスで移動させし、表8.14の相関係数を得る。

表 8.14 データセットの組・群別の相関係数 r

	データセットの組					
	1	2	3	4	5	6
対処群A ₁	-0.1976	0.4711	0.5428 *	0.9628 **	0.9448 **	0.9532 **
処置群A ₂	-0.0350	0.9452 **	0.7924 *	0.9583 **	0.9739 **	0.9350 **
	無印: (r < 0.5)		*: (0.5 < r < 0.90)		**: (r > 0.90)	

JMPのグラフビルダーを用いて2×6の格子状に散布図を描き、確率楕円を重ね書きし、さらに、オプションで相関係数を重ね書きした結果を図8.14に示す。

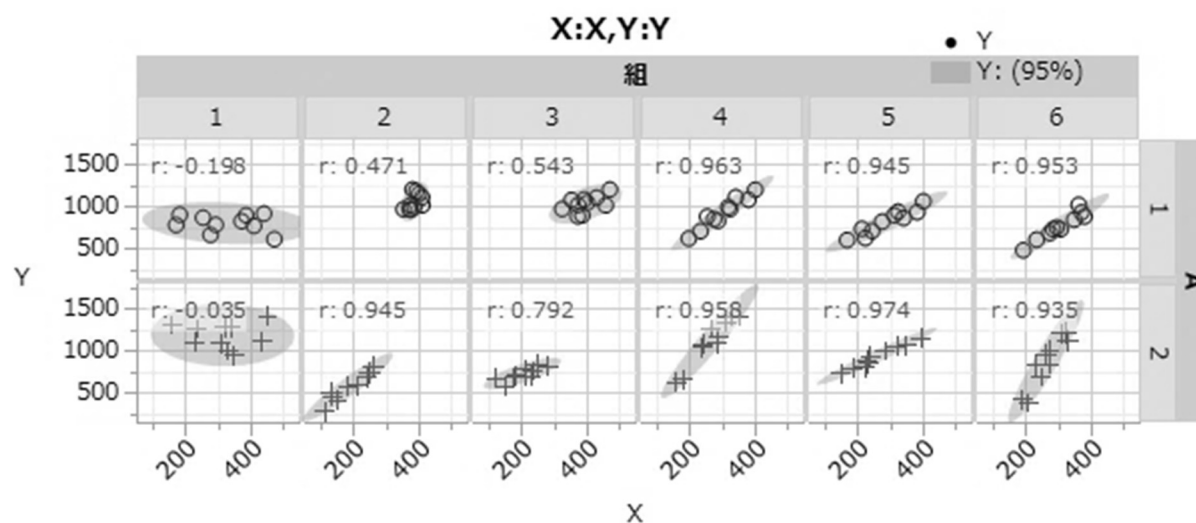


図 8.14 JMP グラフビルダーによる確率楕円と相関係数

2) 2 群の回帰直線が同一直線上か

視覚的には、② 組と ③ 組が該当するが、統計的には、臓器重量を Y とし、体重を共変量 X 、群をダミー変数 A とした回帰分析を行ない、群に関する p 値で判断する。例えば、「 $p > 0.15$ ならば、同一直線」などの基準を前もって設定しておく。② 組の場合に JMP の「モデルのあてはめ」による結果を表 8.15 に示す。JMP は、(1, -1) 対比型ダミー変数なので、対照群 A_1 と処置群 A_2 の回帰式は、

$$\textcircled{2} \text{ 組 } A_1 : \begin{cases} \hat{y}_i^{(2,A_1)} = 33.2462 - 34.3026 \times 1 + 2.7318 x_i^{(2,A_1)} \\ = -1.0564 + 2.7318 x_i^{(2,A_1)} \end{cases} \quad (8.11)$$

$$\textcircled{2} \text{ 組 } A_2 : \begin{cases} \hat{y}_i^{(2,A_2)} = 33.2462 - 34.3026 \times (-1) + 2.7318 x_i^{(2,A_2)} \\ = 67.5488 + 2.7318 x_i^{(2,A_2)} \end{cases} \quad (8.12)$$

となる。 A_1 に対応する $A[1]$ の検定結果は、 $p = 0.4802$ なので「同一直線上」と判断する。また、切片の推定値が $p = 0.7945$ と有意ではないので、原点を通るとみなせる。

表 8.15 ② 組に対する共分散分析

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	2	1340796	670398	123.01
誤差	16	87199	5450	p値
全体(修正済み)	18	1427995		<.0001*

パラメータ推定値				
項	推定値	標準誤差	t値	p値
切片	33.2462	125.5402	0.2648	0.7945
A[1]	-34.3026	47.4592	-0.7228	0.4802
X2	2.7318	0.4354	6.2742	<.0001*

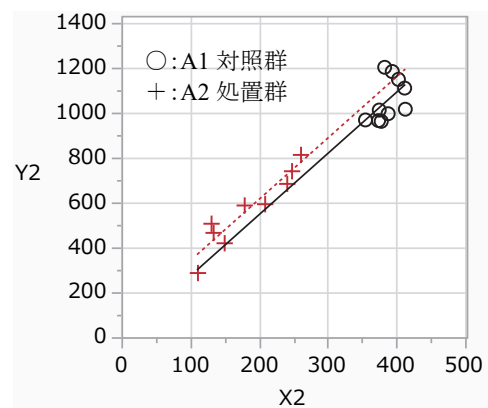


図 8.15 ② についての共分散分析

3) 2 本の直線の傾きは同じか

視覚的には、② 組と ⑤ 組が該当するのであるが、「傾きは同じか」について統計的には、臓器重量を Y とし、体重を共変量 X 、群をダミー変数 A 、さらに、 X と A の交互作用についての回帰分析を行ない、交互作用 $A \times X$ に関する p 値で判断する。例えば、「 $p > 0.15$ ならば、傾きが同じ」などの基準を前もって設定しておく。表 8.16 に ⑤ 組についての交互作用解析の結果を示す。対照群 A_1 と処置群 A_2 の回帰式は、

$$A_1 : \begin{cases} \hat{y}_i^{(5,A_1)} = 382.5823 - 90.5357 \times 1 + 1.8043 x_i^{(5,A_1)} + 0.0197 \times 1 x_i^{(5,A_1)} \\ = 292.0466 + 1.824 x_i^{(5,A_1)} \end{cases} \quad (8.13)$$

$$A_2 : \begin{cases} \hat{y}_i^{(5,A_2)} = 382.5823 - 90.5357 \times (-1) + 1.8043 x_i^{(5,A_2)} + 0.0197 \times (-1) \times x_i^{(5,A_2)} \\ = 473.118 + 1.7846 x_i^{(5,A_2)} \end{cases} \quad (8.14)$$

となり、交互作用「A[1]*X5」に対する検定結果は、 $p=0.8852$ なので傾きは同じと判断できる。

表 8.16 ⑤ 組に対する交互作用の検討

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	3	408190	136063	72.02
誤差	16	30226	1889	p値
全体(修正済み)	19	438416		<.0001*

パラメータ推定値				
項	推定値	標準誤差	t値	p値
切片	382.5823	38.1062	10.04	<.0001*
A[1]	-90.5357	38.1062	-2.38	0.0303*
X5	1.8043	0.1343	13.44	<.0001*
A[1]*X5	0.0197	0.1343	0.15	0.8852

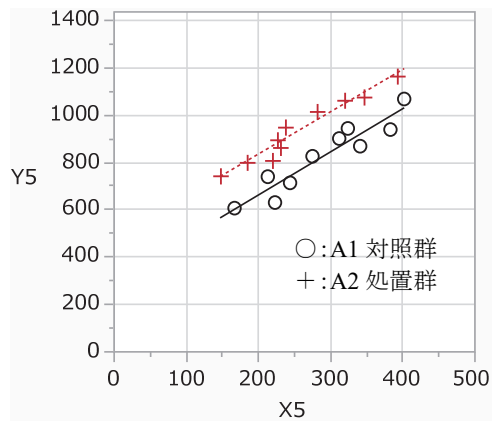


図 8.16 ⑤ 組に対する交互作用の検討

4) 回帰直線が原点を通るか、

視覚的には、② 組と ④ 組が該当するが、それぞれの群別に臓器重量を Y、体重を Xとした回帰分析で、切片の p 値で判断する。例えば、「 $p > 0.15$ ならば、原点を通る」などの基準を前もって設定しておく。④ 組の対照群 A₁ の場合に JMP の「二変量の関係」での結果を表 8.17 に示す。対照群 A₁、切片の推定値 83.0290 に対し、95%信頼区間が (-86.3777, 296.5526) と 0.0 を含んでいるので、原点を通ると判断する。

同様に、表 8.18 に示すように、処置群 A₂ についても切片の推定値 -19.5458 に対し、95%信頼区間が (-301.9799, 262.8839) と 0.0 を含んでいるので、原点を通ると判断される。

表 8.17 ④ 組の A₁ についての回帰分析

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	1	276949	276949	101.58
誤差	8	21811	2726	p値
全体(修正済み)	9	298760		<.0001*

パラメータ推定値				
項	推定値	標準誤差	下側95%	上側95%
切片	105.0874	83.0290	-86.3777	296.5526
X4	2.7120	0.2691	2.0915	3.3325

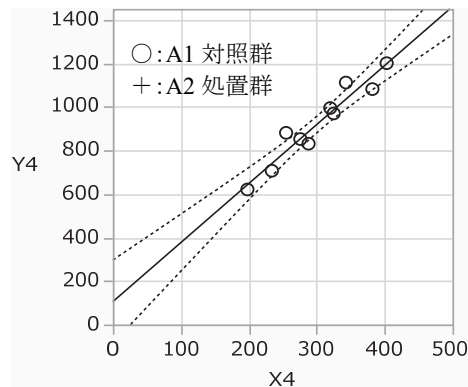


図 8.17 ④ A₁ についての回帰直線

表 8.18 ④ 組の A₂ についての回帰分析

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	1	637127	637127	90.04
誤差	8	56608	7076	p値)
全体(修正済み)	9	693735		<.0001*

パラメータ推定値				
項	推定値	標準誤差	下側95%	上側95%
切片	-19.5458	122.4777	-301.980	262.8884
X4	4.3118	0.4544	3.2640	5.3597

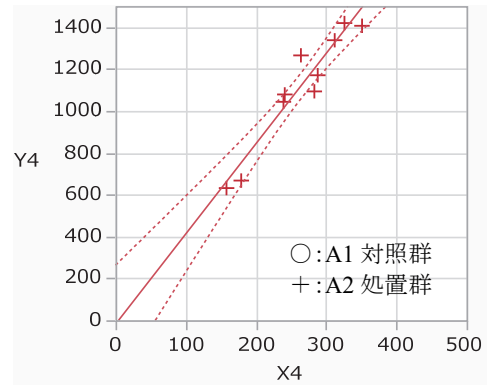


図 8.18 ④ A₂ についての回帰直線

5) 2本の直線が原点を通らず傾きが同じでない場合

視覚的には、⑥ 組が該当するが、JMP の「モデルのあてはめ」による解析結果を表 8.19 に示す。対照群 A₁ と処置群 A₂ の回帰式は、

$$A_1 : \begin{cases} \hat{y}_i^{(6,A_1)} = -316.0250 + 338.5796 \times 1 + 4.1256 x_i^{(6,A_1)} - 1.6938 \times 1 x_i^{(6,A_1)} \\ = 22.5546 + 2.4318 x_i^{(6,A_1)} \end{cases} \quad (8.15)$$

$$A_2 : \begin{cases} \hat{y}_i^{(6,A_2)} = -316.0250 + 338.5796 \times (-1) + 4.1256 x_i^{(6,A_2)} + (-1.6938) \times (-1) x_i^{(6,A_2)} \\ = -654.605 + 5.8194 x_i^{(6,A_2)} \end{cases} \quad (8.16)$$

となり、交互作用「A[1]*X6」に対する検定結果は、 $p = 0.0004$ であり、傾きに関して明らかな差がある。表 8.12 によれば、「回帰直線の差の検定が必要」とされているのだが、傾きの差

$$\hat{\beta}_1^{(6,A_2-A_1)} = 5.8194 - 2.4318 = 3.3876 \quad (8.17)$$

の検定結果は、交互作用の p 値に一致するので、 $p = 0.0004$ と有意な差があるとの結果になるが、これだけで良いのだろうか。Excel の回帰分析を用いて更なる解析方法について、次節で改めて取り上げる。

表 8.19 ⑥ 組に対する交互作用の検討

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	3	959842	319947	42.48
誤差	16	120494	7531	p値)
全体(修正済み)	19	1080336		<.0001*

パラメータ推定値				
項	推定値	標準誤差	t値	p値
切片	-316.025	108.5033	-2.91	0.0102*
A[1]	338.5796	108.5033	3.12	0.0066*
X6	4.1256	0.3818	10.81	<.0001*
A[1]*X6	-1.6938	0.3818	-4.44	0.0004*

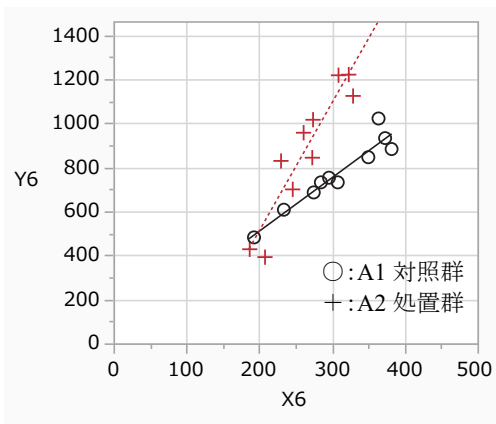


図 8.19 ⑥ に対する交互作用の検討

8.5. 体重を交絡変数とする回帰分析

反復毒性試験において化合物が臓器重量に与える影響を検討する場合に、体重が常に交絡変数となる可能性があるが、その影響はさまざまであり一筋縄では対応できない。探索的な解析は、体重を交絡変数（共変量）とする回帰分析が主体となる。この解析方法は、“共分散分析”と認識されており、そのために 2 本の回帰直線の傾きが異なる場合には、使えないとの“迷信”がはびこっている。そこで、“共分散分析”ではなく、“体重を交絡変数とする回帰分析”とすることにする。

前節の図 8.13 で示した決定樹は、探索的な解析を行うための考え方が示されている。最初の分岐である「 X と Y に相関がある？」について、(ない, ある) の判断基準は、明確に示されていない。その後の分岐でも同様に、判断基準は示されていない。反復毒性試験の統計解析は、あらかじめ定められた方法により自動化されているので、実験研究者に、後から探索的な解析を行うことを期待するのは非現実的である。図 8.13 の決定樹、および表 8.12 の解析方法の説明で、③ 組と ⑤ 組が共分散分析、⑥ 組が回帰直線の傾きの差、について新たな解析の必要性を求めている。これらの解析法については、あらかじめ定式化し、毒性試験終了後に必要に応じて速やかに結果が得られるように、解析システムをあらかじめ準備して置くことが望まれる。

Excel の分析ツールに含まれる回帰分析は、1 変数の単回帰分析のみならず 2 変数以上の“重回帰分析”にも対応している。質的変数としての“群”を量的変数としてのダミー変数とすることにより“共分散分析”にも対応できる。さらに、体重と群の交互作用は、群に関するダミー変数と体重の積を新たな変数として与えることで検定統計量を得ることができる。

群により回帰直線の傾きが異なる場合

表 8.13 に示す ⑥ 組のデータのように、傾きが異なると視覚的に判断される場合、傾きの差に関する統計量をどのように求めたら良いのであろうか。群 A について (0, 1) 型のダミー変数と体重 X の交互作用を含めた回帰分析を行なうことにより、交互作用のパラメータが傾きの差の推定値となり、その標準誤差 SE との比による t 値が検定統計量となり、 p 値も自動的に得られる。傾きの差に加えて、対照群 A_1 と処置群 A_2 について体重の総平均 $\bar{x}_{..} = 284.50$ における両群の臓器重量の推定値、いわゆる最小 2 乗平均による比較ができれば、結果の考察がしやすくなる。

表 8.20 左に示すように、切片を $x_0=1$ ，群 A に対して ($A_1 : a_2=0$ ， $A_2 : a_2=1$) とする (0, 1) 型のダミー変数，体重を x ，群 A と体重 x の交互作用を (a_2x) とする回帰式を

$$y_i = \theta_0 x_{0,i} + \theta_1 a_{2,i} + \theta_2 x_i + \theta_3 (a_{2,i} x_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (8.18)$$

とする。これらの変数をデザイン行列 X としてまとめ、Excel による回帰分析を行うと、表 8.20 に示すように分散分析表および回帰パラメータの推定値を得る。交互作用 (a_2x) のパラメータは、 $\hat{\theta}_3 = 3.3876$ で $p = 0.0004$ と有意な傾きの差となる。推定された回帰式は、

$$\hat{y}^{(6)} = \hat{\theta}_0 x_0 + \hat{\theta}_1 a_2 + \hat{\theta}_2 x + \hat{\theta}_3 (a_2 x) \quad \left. \vphantom{\hat{y}^{(6)}} \right\} \quad (8.19)$$

$$= 22.55 - 677.16 a_2 + 2.4318 x + 3.3876 (a_2 x)$$

であり、対照群 A_1 および処置群 A_2 の回帰式、

$$\text{対照群 } A_1 : \begin{cases} \hat{y}^{(6,A_1)} = 22.55 - 677.16 \times 0 + 2.4318 x + 3.3876 \times 0 \\ = 22.55 + 2.4318 x \end{cases} \quad (8.20)$$

$$\text{処置群 } A_2 : \begin{cases} \hat{y}^{(6,A_2)} = 22.55 - 677.16 \times 1 + 2.4318 x + 3.3876 \times (1 \times x) \\ = -654.60 + 5.8194 x \end{cases} \quad (8.21)$$

が導出される。これらは、図 8.12 の ⑥ 組の図に示されている回帰式に一致する。

表 8.20 傾きに差がある場合の交互作用モデルの適用

⑥組 No	群 A	— デザイン行列 —				$y^{(6)}$ y	分散分析表 (定数に 0 を使用) off					
		x_0	a_2	x	$a_2 x_0$		自由度	変動	分散	分散比		
1	1	1	0	285	0	731	回帰	3	959842	319947	42.4848	
2	1	1	0	382	0	882	残差	16	120494	7530.87	$=\sigma^2$	
3	1	1	0	350	0	845	合計	19	1080336			
4	1	1	0	193	0	481			$t_{0.05} =$	2.1199		
5	1	1	0	234	0	607		係数	標準誤差	t	P -値	
6	1	1	0	275	0	685	θ^{\wedge}_0 切片 x_0	22.55	144.52	0.1561	0.8779	
7	1	1	0	364	0	1021	θ^{\wedge}_1 a_2	-677.16	217.01	-3.1205	0.0066	
8	1	1	0	296	0	752	θ^{\wedge}_2 x	2.4318	0.4637	5.2444	0.0001	
9	1	1	0	308	0	731	θ^{\wedge}_3 $a_2 x$	3.3876	0.7635	4.4367	0.0004	
10	1	1	0	373	0	932						
11	2	1	1	322	322	1224		パラメータの共分散行列				
12	2	1	1	229	229	832		$\Sigma(\theta^{\wedge}) = (X^T X)^{-1} \sigma^2$				
13	2	1	1	272	272	847	θ^{\wedge}_0	20887	-20887	-65.80	65.80	
14	2	1	1	207	207	395	θ^{\wedge}_1	-20887	47092	65.80	-162.57	
15	2	1	1	245	245	703	θ^{\wedge}_2	-65.80	65.80	0.2150	-0.2150	
16	2	1	1	186	186	431	θ^{\wedge}_3	65.80	-162.57	-0.2150	0.5830	
17	2	1	1	273	273	1019		切片 x_0	a_2	x	$a_2 x$	
18	2	1	1	308	308	1221						
19	2	1	1	260	260	960						
20	2	1	1	328	328	1127						
総平均		$\bar{x}_{..} = 284.50$										

対照群 A_1 の切片の推定値は、 $\hat{\theta}_0 = 22.55$ ($p = 0.8779$) と有意ではないので、原点を通るとみなせる。処置群 A_2 の切片の推定値は、 $\hat{\theta}_0 + \hat{\theta}_1 = -654.60$ であり、 p 値は示されていないが、 $\hat{\theta}_1 = -677.16$ の $p = 0.0066$ からの類推で原点を通るとは見做せない。したがって、両群の傾き

の差 $\hat{\theta}_3 = 3.3876$ のみによる考察は、不十分である。これは、図 8.12 に例示した 傾きが異なる ⑥ 組の線グラフから、体重が 200 g 近傍で、両群間に臓器重量間に差がないが、体重の総平均 $\bar{x}_.. = 284.50$ g で両群の回帰直線上の推定値に明らかな差が読み取れる。したがって、臓器重量に対する何らかの要約統計量による比較が必要である。なお、絶対臓器重量での 2 群間の比較では、データの範囲が互いに重なり合い、有意な差とはならないことは明らかであり、不適切である。

体重の総平均での臓器重量の比較

体重の総平均における対照群 A_1 の臓器重量の推定値と、処置群 A_2 の臓器重量の推定値（最小 2 乗平均）を求め、統計的な比較をしたい。そのためには、これまでも繰返し示してきたように、パラメータの共分散行列 $\Sigma(\hat{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$ が必要となる。これは、表 8.20 右下段にデザイン行列 \mathbf{X} と分散分析表の残差の分散 $\hat{\sigma}^2 = 7530.87$ を用いて 4×4 の行列として示されている。

表 8.21 には、体重の総平均 $\bar{x}_.. = 284.50$ g における対照群 A_1 および処置群 A_2 の要約統計量としての推定値 ($y_{x=284.5}^{(6,A_1)} = 714.4$, $y_{x=284.5}^{(6,A_2)} = 1001.0$) および 95%信頼区間が計算されている。さらに、推定値の差 ($y_{x=284.5}^{(6,A_2)} - y_{x=284.5}^{(6,A_1)} = 286.6$) の 95%信頼区間 (197.3, 375.9) が 0 を含まないので有意な差となっている。

表 8.21 傾きに差がある場合の体重の総平均における臓器重量

	線形和	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
⑥ 組	$L^{(i)}$	x_0	a_2	x	$a_2 x$	$y^{\wedge} = \mathbf{l} \hat{\theta}$	$Var(y^{\wedge})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
A_1	$L^{(1)}$	1	0	284.5	0	714.4	852.5	61.9	652.5	776.3
A_2	$L^{(2)}$	1	1	284.5	284.5	1001.0	923.2	64.4	936.6	1065.4
$A_1 - A_1$	$L^{(3)}$	0	0	0	0	0.0	0.0	0.0	0.0	0.0
$A_2 - A_1$	$L^{(4)}$	0	1	0	284.5	286.6	1775.7	89.3	197.3	375.9
							$t_{0.05}(20-4) =$	2.1199		

図 8.20 左に示す折れ線グラフは、表 8.21 で計算された推定値と 95%信頼区間を描いた結果である。体重の総平均における対照群 A_1 と処置群 A_2 の推定値の 95%信頼区間が、互いに重なり合わないので、有意な差があることが視覚的に明らかである。図 8.20 右は、対照群 A_1 に対して処置群 A_2 の臓器重量の推定値が 286.6 mg 増大し、95%信頼区間が 0 を含まないので、有意な差であることが示されている。

これらの推定値を求めるために、変数 ($x_0, a_2, x, a_2 x$) の別名として (l_0, l_1, l_2, l_3) を与え、表 8.20 に示したパラメータの推定値 $\hat{\theta}$ との積和による線形和を $L^{(i)} = \mathbf{l}^{(i)} \hat{\theta}$ として定義する。線形和の係数ベクトル $\mathbf{l}^{(i)}$ を用いて対照群 A_1 の推定値 $y_{x=284.5}^{(6,A_1)}$ を得るために $\mathbf{l}^{(1)} = [1, 0, 284.5, 0]$ を与え、 $L^{(1)} = \mathbf{l}^{(1)} \hat{\theta}$ によって $y_{x=284.5}^{(6,A_1)} = 714.4$ が推定される。表 8.21 に示したの

は、係数ベクトル $\boldsymbol{l}^{(i)}$ に対する線形和 $L^{(i)}$ の推定値と、その分散、95%信頼区間を計算した結果となっている。

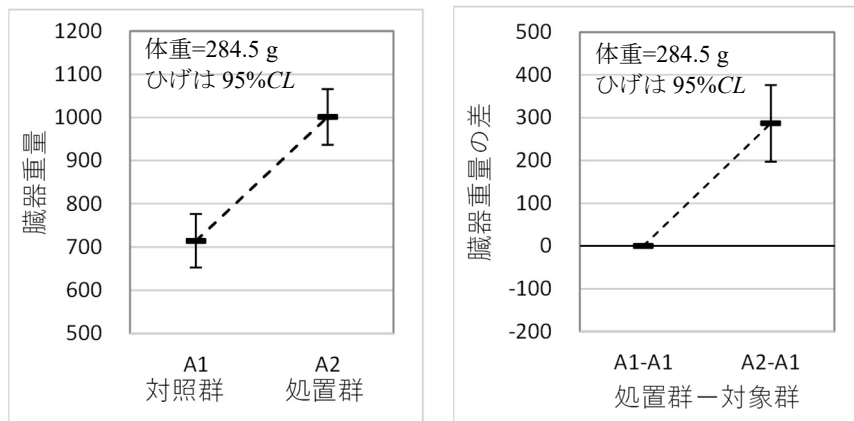


図 8.20 体重が 284.5 g の場合の臓器重量の推定値 (最小 2 乗平均)

実際の計算手順は、次のようである。Excel の行列の積の Mmult () 関数を用いて線形和 $L^{(i)} = \boldsymbol{l}^{(i)} \hat{\boldsymbol{\theta}}$ は、

$$y_{x=284.5}^{(6,A_1)} = L^{(i)} = \boldsymbol{l}^{(i)} \hat{\boldsymbol{\theta}} \tag{8.22}$$

=	1	0	284.5	0	22.55	=	714.4
					-677.16		
					2.4318		
					3.3876		

として計算されている。線形和の分散 $Var(\boldsymbol{l}^{(i)} \hat{\boldsymbol{\theta}})$ は、パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ に対する $\boldsymbol{l}^{(i)}$ の 2 次形式によって、

$$Var(\boldsymbol{l}^{(i)} \hat{\boldsymbol{\theta}}) = \boldsymbol{l}^{(i)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \boldsymbol{l}^{(i)T} \tag{8.23}$$

=	1	0	284.5	0	20887	-20887	-65.80	65.80	1	=	852.5
					-20887	47092	65.80	-162.57	0		
					-65.80	65.80	0.2150	-0.2150	284.5		
					65.80	-162.57	-0.2150	0.5830	0		

として求められ、推定値 $y_{x=284.5}^{(6,A_1)} = 714.4$ に対する 95%信頼区間は、

$$\left. \begin{aligned}
 95\%CL &= y_{x=284.5}^{(6,A_1)} \pm t(0.05,16) \sqrt{Var(\boldsymbol{l}^{(i)} \hat{\boldsymbol{\theta}})} \\
 &= 714.4 \pm 2.1199 \sqrt{852.48} \\
 &= (652.5, 776.3)
 \end{aligned} \right\} \tag{8.24}$$

によって計算されている。ただし、非平行な回帰直線なので、体重を変化させることにより 2 群間の差は、変化することに注意が必要である。

予測プロファイル作成

表 8.22 に 2 本の回帰直線の推定値に対する 95%信頼区間が計算されている。計算方法は、表 8.21 に示した体重が $\bar{x}_{..} = 284.50$ g の場合の対照群 A_1 および処置群 A_2 の推定値と 95%信

区間の計算と同様であるが、体重を $x = 0, \dots, 284.5, \dots, 500$ と変化させている。推定値 $y_x^{(6,A_1)}$ および $y_x^{(6,A_2)}$ ，それらの分散，95%信頼区間の計算法は，表 8.21 と全く同じである。

表 8.22 傾きに差がある場合の回帰直線の 95%信頼区間

	線形和	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
⑥組	$L^{(i)}$	x_0	a_2	x	a_2x	$y^\wedge = \theta$	$Var(y^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
A ₁	$L^{(5)}$	1	0	0	0	22.6	20887	306.4	-283.8	328.9
回	$L^{(6)}$	1	0	100	0	265.7	9878	210.7	55.0	476.4
帰	$L^{(7)}$	1	0	200	0	508.9	3169	119.3	389.6	628.3
	$L^{(8)}$	1	0	284.5	0	714.4	852	61.9	652.5	776.3
	$L^{(9)}$	1	0	400	0	995.3	2653	109.2	886.1	1104.5
	$L^{(10)}$	1	0	500	0	1238.5	8846	199.4	1039.1	1437.9
A ₂	$L^{(11)}$	1	1	0	0	-654.6	26205	343.2	-997.8	-311.4
回	$L^{(12)}$	1	1	100	100	-72.7	10530	217.5	-290.2	144.9
帰	$L^{(13)}$	1	1	200	200	509.3	2214	99.7	409.5	609.0
	$L^{(14)}$	1	1	284.5	284.5	1001.0	923	64.4	936.6	1065.4
	$L^{(15)}$	1	1	400	400	1673.2	7660	185.5	1487.6	1858.7
	$L^{(16)}$	1	1	500	500	2255.1	21422	310.3	1944.8	2565.4

表 8.23 には，2本の回帰直線の差の推定値と 95%信頼区間が示されている。 $L^{(20)}$ の $l_3^{(20)}$ が， $a_2x = 233.8$ に対し，95%信頼区間の下限 L95% が 0 となっているので，体重が 233.8 g において，95%信頼区間の下限 L95% が 0.0 となり，有意な差となることが示されている。これは，Excel のソルバーで，L95% が 0 となるように， $l_3^{(20)} = 233.8$ を変化させた結果である。

表 8.23 回帰直線の差についての 95%信頼区間

	線形和	l_0	l_1	l_2	l_3	推定値	分散	幅	95%信頼区間	
⑥組	$L^{(i)}$	x_0	a_2	x	a_2x	$y^\wedge = \theta$	$Var(y^\wedge)$	$t_{0.05} \times SE$	L95%	U95%
A ₂ -A ₁	$L^{(17)}$	0	1	0	0	-677.2	47092	460.0	-1137.2	-217.1
回	$L^{(18)}$	0	1	0	100	-338.4	20407	302.8	-641.2	-35.6
帰	$L^{(19)}$	0	1	0	200	0.4	5383	155.5	-155.2	155.9
	$L^{(20)}$	0	1	0	233.8	114.9	2940	114.9	0.0	229.9
	$L^{(21)}$	0	1	0	400	677.9	10313	215.3	462.6	893.1
	$L^{(22)}$	0	1	0	500	1016.6	30267	368.8	647.8	1385.4

図 8.21 左は，図 8.12 の ⑥組のデータで示した体重と臓器重量の散布図に表 8.22 で計算した回帰直線および 95%信頼区間を，Excel の散布図の「データの選択」オプションを使って順次追加し，書式を整えたもので，体重がおおよそ 250 g 以上の場合に，A₂ の L95% と A₁ の U95% が完全に分離し，統計的な差が明らかにあることが読み取れる。

図 8.21 右は，表 8.20 に示されているパラメータの推定値 $\hat{\theta}_1$ と $\hat{\theta}_3$ についての回帰直線である。図 8.21 左は，2本の回帰直線の差の推定直線となっている。

$$\begin{aligned} \hat{y}^{(6,A_2-A_1)} &= \hat{\theta}_1 a_2 + \hat{\theta}_3 (a_2 x) \\ &= -677.16 + 3.3876x \end{aligned} \tag{8.25}$$

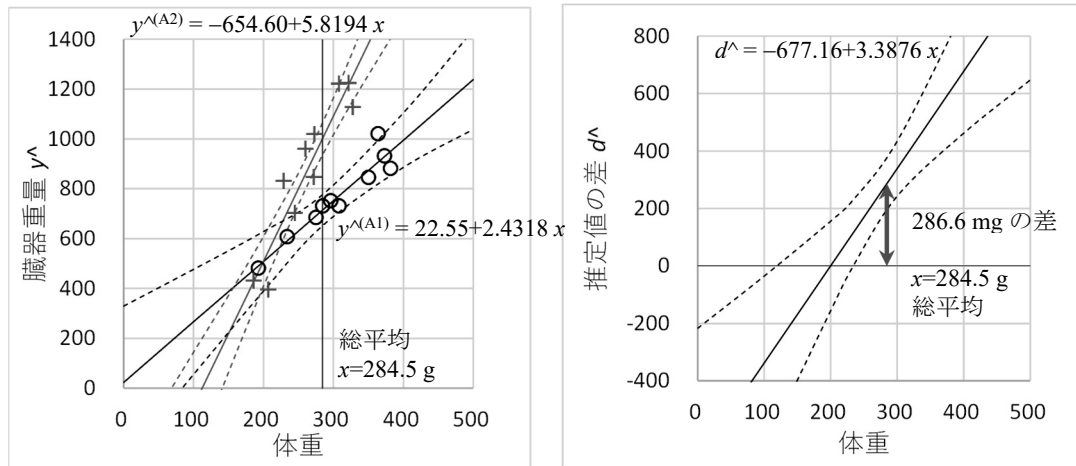


図 8.21 非平行な 2 本の回帰直線および差の回帰直線

また、体重の総平均 $\bar{x} = 284.50$ g に対する直線の推定値 286.6 mg は、図 8.21 左の 2 本の直線間の差となっている。

回帰直線が同一直線上と見なせる場合

③ 組のデータは、図 8.12 に示したように、対照群 A_1 に対し、処置群 A_2 の体重 $x^{(3)}$ および臓器重量 $y^{(3)}$ が共に、明らかに小さくなっているが、それぞれについて回帰直線を重ね書きすると、多少のずれはあるものの同一直線上にあるとみなせ、原点は通らない事例である。

「2 群が同一直線上か」を統計的に判断するためには、群 A と体重 x に交互作用 $A \times x$ がないことを確認し、交互作用 $A \times x$ を除いた（群 A、体重 x ）を変数とする回帰分析（主効果モデル）で、群間に統計的な差がないことを示す必要がある。③ 組のデータについて、Excel の回帰分析を主体にした解析方法を示す。

表 8.24 左に示すように、群 A を (0, 1) 型ダミー変数とすると処置群 A_2 が 1 となるので変数名としては a_2 としている。体重 X_3 は x とし、交互作用を $a_2x = a_2 \times x$ として計算している。切片に対する変数を x_0 として明示的に与え、 (x_0, a_2, x, a_2x) の 20 行分のデータが“デザイン行列 X ”となる。臓器重量 Y_3 は y として示す。表 8.24 右上は、交互作用 a_2x を含めたモデルの結果で、表 8.24 右下段は、交互作用の変数 a_2x を除いた主効果モデルの結果である。

交互作用モデルで得られた推定式は、

$$\left. \begin{aligned} \hat{y}^{(3)} &= \hat{\theta}_0 x_0 + \hat{\theta}_1 a_2 + \hat{\theta}_2 x + \hat{\theta}_3 (a_2 x) \\ &= 575.59 x_0 - 95.90 a_2 + 1.1450 x + 0.0985 (a_2 x) \end{aligned} \right\} \quad (8.26)$$

であり、 A_1 および A_2 の回帰式は、

表 8.24 回帰直線が同一直線上の場合の交互作用モデル vs. 主効果モデル

③ 組	群	— デザイン行列 X —				$y^{(3)}$	交互作用モデル 推定値, 分散分析表は省略						
		x_0	a_2	x	a_2x			係数	標準誤差	t	P -値		
1	1	1	0	355	0	1080	θ^{\wedge}_0	切片 x_0	575.59	204.37	2.8164	0.0124	
2	1	1	0	432	0	1107	θ^{\wedge}_1	a_2	-95.90	227.60	-0.4214	0.6791	
3	1	1	0	460	0	1015	θ^{\wedge}_2	x	1.1450	0.5105	2.2431	0.0394	
4	1	1	0	473	0	1204	θ^{\wedge}_3	a_2x	0.0985	0.7016	0.1404	0.8901	
5	1	1	0	374	0	886							
6	1	1	0	375	0	1020	分散分析表 主効果モデル						
7	1	1	0	394	0	1085		自由度	変動	分散	分散比		
8	1	1	0	326	0	972		回帰	2	505458	252729	54.4466	
9	1	1	0	388	0	897		残差	17	78910	4641.78		
10	1	1	0	403	0	1047		合計	19	584369			
11	2	1	1	247	247	844				$t_{0.05} =$	2.1098		
12	2	1	1	210	210	790			係数	標準誤差	t	P -値	
13	2	1	1	280	280	827	θ^{\wedge}_0	切片 x_0	554.84	137.01	4.0496	0.0008	
14	2	1	1	117	117	683	θ^{\wedge}_1	a_2	-65.74	72.99	-0.9006	0.3804	
15	2	1	1	175	175	688	θ^{\wedge}_2	x	1.1971	0.3400	3.5213	0.0026	
16	2	1	1	181	181	721	パラメータの共分散行列						
17	2	1	1	233	233	769	$\Sigma(\beta^{\wedge}) = (X^T X)^{-1} \sigma^{\wedge 2}$						
18	2	1	1	148	148	591	θ^{\wedge}_0		18772.4	-9438.9	-46.0007		
19	2	1	1	230	230	710	θ^{\wedge}_1		-9438.9	5327.8	22.5496		
20	2	1	1	208	208	697	θ^{\wedge}_2		-46.0007	22.5496	0.1156		
総平均		$\bar{x}_{..} = 300.5$											

$$\text{対照群 } A_1 : \begin{cases} \hat{y}^{(3,A_1)} = \hat{\theta}_0 \times 1 + \hat{\theta}_1 \times 0 + \hat{\theta}_2 x + \hat{\theta}_3 (0 \times x) \\ = \hat{\theta}_0 + \hat{\theta}_2 x \\ = 575.59 + 1.1450x \end{cases} \quad (8.27)$$

$$\text{処置群 } A_2 : \begin{cases} \hat{y}^{(3,A_2)} = \hat{\theta}_0 x_0 + \hat{\theta}_1 \times 1 + \hat{\theta}_2 x + \hat{\theta}_3 (1 \times x) \\ = (\hat{\theta}_0 + \hat{\theta}_1) + (\hat{\theta}_2 + \hat{\theta}_3) x \\ = 479.69 + 1.2435x \end{cases} \quad (8.28)$$

であり、傾きの差は $\hat{\theta}_3 = 0.0985$ ($p = 0.8901$) と有意な差ではない。

表 8.24 右下の交互作用項 (a_2x) を除いた主効果モデルによる回帰式は、

$$\left. \begin{aligned} \hat{y}^{(3)} &= \hat{\theta}_0 x_0 + \hat{\theta}_1 a_2 + \hat{\theta}_2 x \\ &= 554.84 \times 1 + (-65.74) \times a_2 + 1.1971x \end{aligned} \right\} \quad (8.29)$$

であり、群別の回帰式は

$$\text{対照群 } A_1 : \begin{cases} \hat{y}^{(3,A_1)} = \hat{\theta}_0 \times 1 + \hat{\theta}_1 \times 0 + \hat{\theta}_2 x \\ = 554.84 + 1.1971x \end{cases} \quad (8.30)$$

$$\text{処置群 } A_2 : \begin{cases} \hat{y}^{(3,A_2)} = \hat{\theta}_0 \times 1 + \hat{\theta}_1 \times 1 + \hat{\theta}_2 x \\ = 489.10 + 1.1971x \end{cases} \quad (8.31)$$

であり、Y 軸方向の回帰直線の差は $\hat{\theta}_1 = -65.74$ ($p = 0.3804$) と有意な差ではなく、 p 値も 0.4 に近いので、2 本の回帰直線は、「同一直線上」にあると統計的に判断する。

推定された統計量について考察する際には、折れ線グラフ表示が有効である。有意な差がない場合でも、どの程度なのか視覚的に判断することも必要である。表 8.25 に示すように、両群の体重の総平均 $\bar{x}_{..} = 300.5$ g における直線上の推定値を求め、その 95%信頼区間を推定

表 8.25 回帰直線が同一直線上の場合の体重の総平均 300.5 g における推定値

線形和	l_0	l_1	l_2	推定値	分散	幅	95%信頼区間	
③組 $L^{(i)}$	x_0	a_2	x	$y^{\wedge} = l_0 + l_1 x$	$Var(y^{\wedge})$	$t_{0.05} \times SE$	L95%	U95%
A ₁ $L^{(1)}$	1	0	300.5	914.5	1564.0	83.4	831.1	998.0
A ₂ $L^{(2)}$	1	1	300.5	848.8	1564.0	83.4	765.3	932.2
A ₁ -A ₁ $L^{(3)}$	0	0	0	0.0	0.0	0.0	0.0	0.0
A ₂ -A ₁ $L^{(4)}$	0	1	0	-65.7	5327.8	154.0	-219.7	88.3
					$t_{0.05}(20-3) =$	2.1098		

し、図 8.22 左に図示する。平行な直線間の差は、体重 x と関係なく -65.7 mg と推定され、その SE は、表 8.24 で推定されたダミー変数 a_2 の $SE = 72.99$ に一致する。差 ($A_2 - A_1$) に対する 95%信頼区間を図 8.22 右に図示する。このように、有意な差ではないと結論付けるだけでなく、視覚的に示すことも必要である。

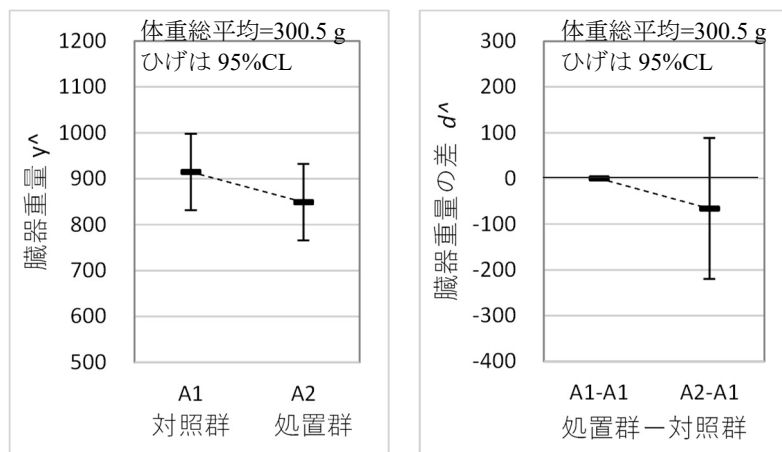


図 8.22 体重の総平均 $\bar{x}_{..} = 300.5$ g における臓器重量および差の推定値

体重の総平均 $\bar{x}_{..} = 300.5$ g における臓器重量および差について、散布図上に平行な直線と 95%信頼区間を重ね書きし、図 8.22 の結果と比較することにより、結果の解釈に際し要約統計量として役立つ。表 8.26 に示すのは、表 8.25 の体重の総平均 $\bar{x}_{..} = 300.5$ g における推定のための計算シートを活用し、0 g から 500 g まで 100 g ごとに引き伸ばした結果である。

表 8.26 回帰直線が同一直線上の場合の 95%信頼区間

③ 組	$L^{(i)}$	l_0	l_1	l_2	推定値 $y^{\wedge}=\boldsymbol{l}\boldsymbol{\theta}$	分散 $Var(y^{\wedge})$	幅 $t_{0.05}\times SE$	95%信頼区間	
		x_0	a_2	x				L95%	U95%
A ₁	$L^{(5)}$	1	0	0	554.8	18772.4	289.1	265.8	843.9
回	$L^{(6)}$	1	0	100	674.6	10728.1	218.5	456.0	893.1
帰	$L^{(7)}$	1	0	200	794.3	4995.4	149.1	645.2	943.4
	$L^{(8)}$	1	0	300.5	914.6	1562.9	83.4	831.2	998.0
	$L^{(9)}$	1	0	400	1033.7	464.6	45.5	988.2	1079.2
	$L^{(10)}$	1	0	500	1153.4	1666.7	86.1	1067.3	1239.5
A ₂	$L^{(11)}$	1	1	0	489.1	5222.4	152.5	336.6	641.6
回	$L^{(12)}$	1	1	100	608.8	1688.0	86.7	522.1	695.5
帰	$L^{(13)}$	1	1	200	728.5	465.2	45.5	683.0	774.0
	$L^{(14)}$	1	1	300.5	848.8	1565.2	83.5	765.4	932.3
	$L^{(15)}$	1	1	400	968.0	4954.3	148.5	819.5	1116.5
	$L^{(16)}$	1	1	500	1087.7	10666.2	217.9	869.8	1305.6

図 8.23 に示すのは、表 8.24 の ③ 組のデータについて層別散布図を描き、表 8.26 の推定結果をさらに重ね書きした結果である。体重の総平均 $\bar{x}_{..} = 300.5$ g における回帰直線の推定値と 95%信頼区間が、図 8.22 左に対応する。

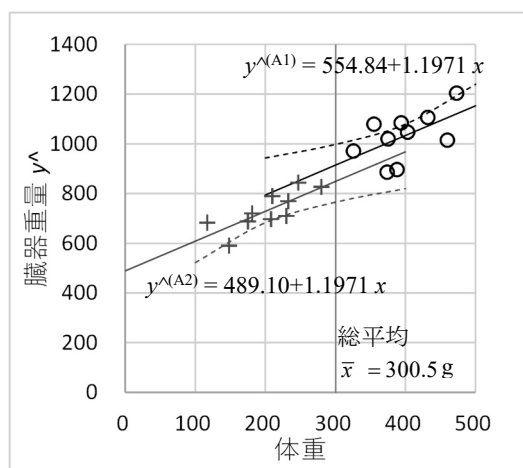


図 8.23 同一直線上の平行な 2 本の回帰直線の 95%信頼区間

回帰直線間の差および 95%信頼区間は、体重 x が変化しても一定であるので、図示するまでもない。このことは、表 8.25 に示したように、処置 A₂ と対照 A₁ の回帰直線の差の推定に対し、体重 x が 0 となっており、体重に関係なく変数 a_2 のパラメータ $\hat{\theta}_1 = -65.74$ のみによる推定であることから明らかである。さて、ここに示した解析法は、いわゆる“共分散分析”の手順と同じであるが、共変量と見なした交絡変数としての体重の範囲が、処置 A₂ と対照 A₁ で全く異なり、いわゆる“共分散分析”の適用外である。このように交絡変数の反応に与える影響は、多様であり“明らかな交互作用”がある場合も含め、柔軟な対応が求められる。

2本の回帰直線が原点を通る場合

2本の回帰直線が原点を通るとの判定は、別々に回帰直線をあてはめ、切片の95%信頼区間が原点を含むことで対応してきたのであるが、(1, 1) 標示型ダミー変数を用いることにより一括して対応でき、原点を通る2本の回帰直線のあてはめにも応用できる。(1, 1) 標示型ダミー変数は、対照A₁に対し a₁=1, A₁以外は a₁=0, 処置A₂に対し a₂=1, A₂以外は a₂=0とするダミー変数である。

表 8.27 左に示すように、デザイン行列は対照群A₁と処置群A₂それぞれに対する回帰分析の変数を単に並列的に並べた (a₁, a₂, a₁x, a₂x) 形式となっている。ここには、いつもの“切片”のための共通の変数x₀が含まれていないことに注意してもらいたい。ここでは、x₀に代えて (a₁, a₂) が2本の回帰直線のそれぞれの切片を求めるための変数となっている。表 8.27 右上は、Excel の回帰分析で (定数に 0 を使用) オプションを on として求めたパラメータ推定値となっている。推定された回帰式は、

$$\text{対照群 } A_1 \begin{cases} \hat{y}^{(4,A_1)} = \hat{\theta}_{0,1} \times a_1 + \hat{\theta}'_1 \times (a_1x) \\ = 105.09 + 2.7120x \end{cases} \quad (8.32)$$

$$\text{処置群 } A_2 \begin{cases} \hat{y}^{(4,A_2)} = \hat{\theta}_{0,2} \times a_2 + \hat{\theta}'_2 \times (a_2x) \\ = -19.55 + 4.3118x \end{cases} \quad (8.33)$$

表 8.27 2本の回帰直線が原点を通る場合の (1, 1) 標示型ダミー変数による回帰分析

④組	群	— デザイン行列 X —				y ⁽⁴⁾	(定数に 0 を使用) on, 分散分析表を除く				
No	A	a ₁	a ₂	a ₁ x	a ₂ x	y		係数	標準誤差	t	P-値
1	1	1	0	255	0	881	切片	0.0000	#N/A	#N/A	#N/A
2	1	1	0	382	0	1082	θ [^] _{0,1}	105.09	111.32	0.9440	0.3592
3	1	1	0	320	0	995	θ [^] _{0,2}	-19.55	101.93	-0.1918	0.8503
4	1	1	0	403	0	1201	θ [^] ₁	2.7120	0.3608	7.5171	0.0000
5	1	1	0	234	0	707	θ [^] ₂	4.3118	0.3782	11.4015	0.0000
6	1	1	0	325	0	970					
7	1	1	0	198	0	621	(定数に 0 を使用) on, (a ₁ とa ₂)を除くモデル				
8	1	1	0	276	0	852		自由度	変動	分散	分散比
9	1	1	0	288	0	831	回帰	2	21899498	10949749	2172
10	1	1	0	343	0	1112	残差	18	82967	4609	=σ [^] 2
11	2	0	1	0	287	1174	合計	20	21982465		
12	2	0	1	0	240	1082			t _{0.05} =	2.1009	
13	2	0	1	0	282	1097		係数	標準誤差	t	P-値
14	2	0	1	0	157	635	切片	0.0000	#N/A	#N/A	#N/A
15	2	0	1	0	325	1423	θ [^] ₁	3.0458	0.0696	43.7751	0.0000
16	2	0	1	0	311	1341	θ [^] ₂	4.2411	0.0797	53.2441	0.0000
17	2	0	1	0	263	1269	パラメータの共分散行列				
18	2	0	1	0	178	671	Σ(β [^])=(X ^T X) ⁻¹ σ [^] 2				
19	2	0	1	0	350	1410	θ [^] ₁	0.004841	0.000000		
20	2	0	1	0	238	1047	θ [^] ₂	0.000000	0.006345		
総平均		x̄ _{..} =		282.8							

であり、図 8.12 の ④ 組の散布図で、Excel の線形近似により重ね書きした回帰式に一致する。それぞれの切片の推定値に対する p 値から有意な差ではなく、95%信頼区間が原点を含むことが示されている。

原点を通る 2 本の回帰直線の同時推定は、変数 (a_1x , a_2x) に対し、(定数に 0 を使用) オプションを on として行うことができる。結果は、表 8.27 右下に示されているように、

$$\text{対照群 } A_1 : \hat{y}^{(4,A_1)} = \hat{\theta}_1 x = 3.0458x \quad (8.34)$$

$$\text{処置群 } A_2 : \hat{y}^{(4,A_2)} = \hat{\theta}_2 x = 4.2411x \quad (8.35)$$

が得られる。共に原点を通る場合には、臓器重量 y^4 を体重 x^4 で割った“体重比”での比較が正当性を持つのであるが、推定された原点を通る回帰直線を用い、対照群 A_1 に対し処置群 A_2 の臓器重量の肥大の程度を量的に示したい。

表 8.28 および図 8.24 に示すように、体重の総平均 $\bar{x}_{..} = 282.8$ g における各群の臓器重量は、対照 A_1 群で 861.2 mg, 処置 A_2 群で 1199.2 mg であり、その差は+338.0 mg であり、95% 信頼区間は (275.1~400.8) と明らかな臓器重量を肥大させている。

表 8.28 2 本の回帰直線が原点を通る場合の体重の総平均 $\bar{x}_{..} = 282.8$ g における推定値

線形和	l_1	l_2	推定値	分散	幅	95%信頼区間	
④ 組 $L^{(i)}$	a_1x	a_2x	$y^4 = l\theta$	$Var(y^4)$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
A_1 $L^{(1)}$	282.8	0	861.2	387.0	41.3	819.9	902.5
A_2 $L^{(2)}$	0	282.8	1199.2	507.2	47.3	1151.8	1246.5
$A_1 - A_1$ $L^{(3)}$	0	0	0.0	0.0	0.0	0.0	0.0
$A_2 - A_1$ $L^{(4)}$	-282.8	282.8	338.0	894.3	62.8	275.1	400.8
				$t_{0.05}(20-2) =$	2.1009		

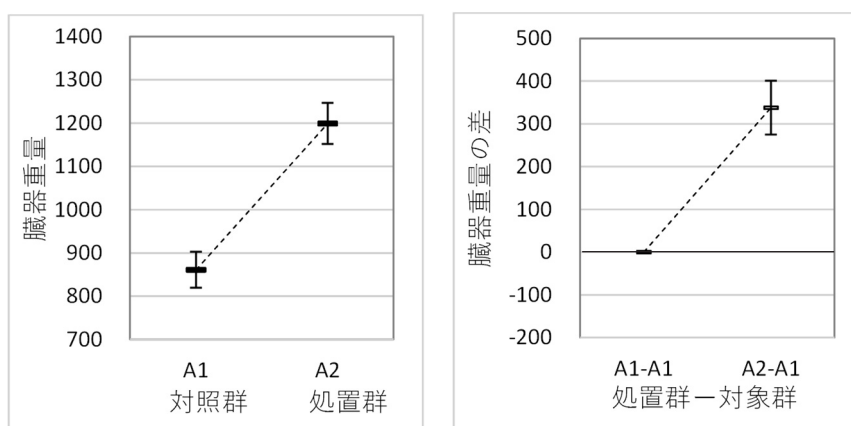


図 8.24 体重の総平均 $\bar{x}_{..} = 282.8$ g の場合の臓器重量および差の推定値

さらに、結果を吟味するために原点を通る回帰直線に 95%信頼区間を重ね書きし、体重の総平均 $\bar{x}_{..} = 282.8$ g における臓器重量の推定値を示したい。表 8.29 に示すのが、体重 x を 0 ~ 500 g まで 100 g ごとに变化させ場合の推定値と 95%信頼区間の計算結果である。一般的な

回帰直線の 95%信頼区間は、データの範囲への外挿に際し広がる曲線なのであるが、原点を通る直線の場合は、まったく異なり 95%信頼区間は、原点を通る直線となっている。奇異な感じがするが、体重が 100 g の場合の対照群の推定値は、 $\hat{y}_{x=100}^{A_1} = 2.7120 \times 100$ であり、分散は、 $Var(a_1x) = x^2 Var(a_1)$ 、標準誤差は、 $SE(a_1x) = x \cdot SE(a_1)$ となるので、95%信頼区間も当然 x に比例し、原点を通る直線となる。

表 8.29 2 本の回帰直線が原点を通る場合の 95%信頼区間

線形和		l_1	l_2	推定値	分散	幅	95%信頼区間	
④ 組	$L^{(i)}$	a_1x	a_2x	$y^{\wedge}=l\theta$	$Var(y^{\wedge})$	$t_{0.05} \times SE$	L95%	U95%
A ₁	$L^{(5)}$	0	0	0.0	0.0	0.0	0.0	0.0
回	$L^{(6)}$	100	0	304.6	48.4	14.6	290.0	319.2
帰	$L^{(7)}$	200	0	609.2	193.6	29.2	579.9	638.4
	$L^{(8)}$	282.8	0	861.2	387.0	41.3	819.9	902.5
	$L^{(9)}$	400	0	1218.3	774.6	58.5	1159.8	1276.8
	$L^{(10)}$	500	0	1522.9	1210.3	73.1	1449.8	1596.0
A ₂	$L^{(11)}$	0	0	0.0	0.0	0.0	0.0	0.0
回	$L^{(12)}$	0	100	424.1	63.4	16.7	407.4	440.8
帰	$L^{(13)}$	0	200	848.2	253.8	33.5	814.7	881.7
	$L^{(14)}$	0	282.8	1199.2	507.2	47.3	1151.8	1246.5
	$L^{(15)}$	0	400	1696.4	1015.1	66.9	1629.5	1763.4
	$L^{(16)}$	0	500	2120.5	1586.2	83.7	2036.9	2204.2

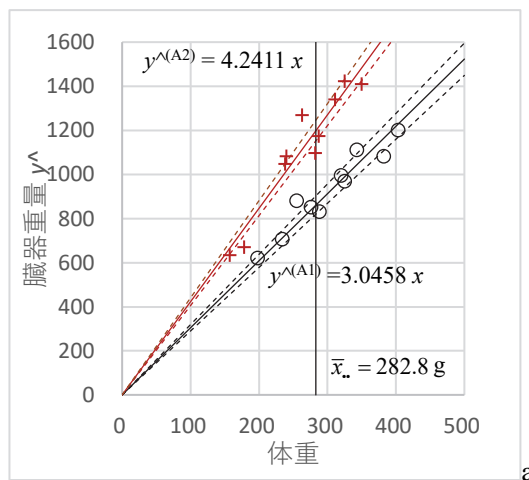


図 8.25 原点を通る 2 本の回帰直線の 95%信頼区間

交絡変数を常に念頭にした現実的な対応

反復毒性試験の解剖時の体重と臓器重量を事例とし、交絡変数を常に念頭にした場合の探索的な解析について例示してきた。現実に行なわれている反復毒性試験では、雄と雌それぞれに対照群に対し最低でも 4 用量の投与群が設定され、対象となる臓器は、約 15 種類あり、

前もって準備されたコンピュータ上の解析システムによって試験終了後速やかに解析報告書が出力され、実験研究者の元に届けられる。このような現実の体制の中で、本章で示した探索的な解析を実験研究者に求めるのは、非現実的である。

図 8.13 に示した「臓器重量の解析のための決定樹」に沿って、解析方法を示してきたのであるが、現実の毒性試験の現場で運用を目指すためには、どうしたら良いのであろうか。通常の解析システムでは、性別・臓器別に対照群と投与群間で絶対重量および体重比での相対重量の平均値に対する多重比較を行っている。着眼点は、対処群に対し有意な差となる投与（用量）群がどこからなのかを知ることである。

これまでの検討から、「臓器重量の解析のための決定樹」に沿った解析法は、煩雑で不十分であることを間接的に示してきた。Excel による解析事例として示してきた⑥、③、および、④の事例について共通するのは、体重と臓器重量について層別散布図を描き、交互作用があることを前提とした回帰直線を重ね書きし、体重の総平均におけるそれぞれの回帰直線の推定値および 95%信頼区間、体重の総平均における対処群と処置群の差および 95%信頼区間の計算結果であろう。

SAS の GLM プロシジャを用いることにより、交互作用を含む“共分散分析”であっても群に対する lsmeans ステートメントを用いた最小 2 乗平均が、体重の総平均 \bar{x} における推定値と 95%信頼区間を容易に求めることができる。もちろん、JMP の「モデルのあてはめ」によって、SAS と GLM プロシジャと同等の結果が得られる。

いわゆる“共分散分析”の適用に際し、群と共変量の交互作用がないことが、古くから“妄想”のごとく受け継がれている。交互作用がなければ、切片の差が群間の差と見なせるのであるが、交互作用がある場合の群間差と 95%信頼区間の計算は、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を必要とし、第 6 章で示した“手計算”による定式化の範疇外であることに起因している。

これまでに示してきたように、交絡変数の影響は多彩である。交絡変数がある場合には、“共分散分析”を適用すればいいとの短絡的な発想は、“共分散分析”の“交互作用がない場合”に適用できるとの前提条件に絡め取れてしまいがちになる。“交絡変数”は、いわゆる“共変量”ではないことを重ねて注意したい。本節のタイトルにあるように「何々を交絡変数とする回帰分析」、あるいは、「交絡変数を含む回帰分析」と区別する必要がある。

9. 前後差の前値は常に共変量

経時的に測定できる反応は、何らかの処置を行う直前にも測定することができる。処置の直前の測定値を前値としたときに、前値が実験の開始前に得られ、幾つかの群ごとに前値の平均と範囲が均一になるような群分けができる場合もある。前値による群分けができない場合、「平均への回帰」という厄介な現象への対応が必要となる。前値を含む経時データに対しては、前値からの前後差（変化量）、あるいは、変化率に対する群間の比較が一般的に行なわれている。前後差に対して有意差検定をするまでもない明らかな平均値の差がある場合には、「平均への回帰」が結果に与える影響は相対的に小さい。だが、統計的に微妙な前後差となる場合に、「平均への回帰」による影響を無視できなくなる。本章では、前値と後値の関係に潜む「平均への回帰」が、群間の比較に際し、どのように影のような影響をおよぼしているか明らかにする。前値を共変量した場合の回帰分析を適用することにより、厄介な「平均への回帰」の影響を排除できることを示す。なお、第 1.3 節では、前値を共変量とする 2 群間比較について示し、第 1.4 節では、前値を共変量とする（対照群+2×2 群、全体で 5 群）の解析事例を示した。

9.1. 「平均への回帰」をめぐる葛藤

多様な揺らぎ

人間を含む動物の血圧、血液中のコレステロール値などの変数は、経時的に測定することが可能であり、多様な揺らぎを伴い一定の値が得られないが、長期的には安定していることが知られている。人間の血圧は、加齢と共に高くなり、食生活の違いの影響を受けて高くなる場合もあるだろうし、内因的な何らかの原因により高くなる場合もある。血圧が高めである場合に、脳梗塞あるいは心疾患による死亡リスクが高くなること、コレステロールも、ある一定の範囲を超えた状態が続く場合に、血圧と同様に脳梗塞あるいは心疾患などの発症リスクが高くなることも知られている。

血圧が高めの人たちを対象に、血圧を下げると期待される薬物をある一定の期間反復投与する臨床試験を想定しよう。有効成分が含まれていないプラセボと実薬を用い、そのどちらかが投与されているか被検者はもちろんのこと、実験の関係者にも分からないように第 3 者

がブラインド化する。投与開始前の血圧測定もブラインド化された状態であり、全てのデータが揃った後に第3者から、それぞれの被験者がプラセボか実薬なのかが知らされる。このような状況下で得られた血圧の前値と投与終了時の後値に対し、プラセボと実薬の血圧がどのくらい下がったのかを統計的に比較評価したい。

前後差での群間比較

プラセボを A_1 、実薬を A_2 とし、前値を X 、後値を Y とする。血圧には個人差があるので、被検者ごとに後値から前値を引いた前後差 $D=Y-X$ について薬剤群間 (A_1, A_2) について対応のない t 検定で有意差検定をすればいいのではないかと直感的に思われている。しかし、前後差 D についての群間比較には、無視できない欠陥が内在する。前値 X について各群の平均値が同じとなるような群分けがされていないので、前値 X の各群の平均値がほぼ等しいとみなせることはまれであり、統計的に有意な平均値の差が出てしまう場合もある。この結果として、前値 X の平均値が相対的に大きい群では、後値 Y が小さ目になるバイアスが混入し、前値 X の平均値が相対的に小さい群には、後値 Y が大きめになるバイアスが混入する。これが「平均への回帰」現象であり、前後差 D による有意差検定の結果にバイアスが影のように忍び込み、信頼することができなくなる。

前後差 D での群間比較に際し、前値 X の群平均の違いによるバイアスの混入について指摘したが、更なる問題点が内在する。前値 X と後値 Y との時点間の相関が 0.5 以下の場合には、前値 X を無視した後値 Y での群間比較の方が前後差 D の場合よりも統計的に優れること（有意な差が出やすい）も知られており、いったいどうすればよいのだろうか。

癖のある共変量

生物統計の名著であるスネデガー・コクラン著、畑村・奥野・津村訳（1972）、「統計的方法 原著第6版」の「第14章 共分散分析」に前後差 $D=Y-X$ ではなく、後値 Y に対して前値を共変量 X とした共分散分析を適用すべきであると強調されている。だが、前値を共変量にした共分散分析にも厄介な問題が発生する。第1.3節で取り上げた事例は、前値 X を共変量とした場合に、薬剤 A との間の交互作用 $A \times X$ が有意な差となり、単純に共変量 X の影響を除いて薬剤 A の群間比較ができない場合への対応方法を示した。

いずれにしても一筋縄では解決できない課題であるが、経時データの解析における前値 X と後値 Y の統計的な関連を正確に把握することが、この問題に対する適切な対応を可能にする。前値 X と後値 Y にどのような統計的な関係があるのかを抽象的ではなく、Excel の様々な機能を使い可視化し、内在する統計的な課題を的確に把握できるように多面的な切り口を示す。

9.2. 前後差 D に潜む「平均への回帰」の可視化

Excel による正規乱数の発生

仮想的な実験を Excel で設定し、薬剤 (A_1 , A_2) の収縮期血圧の前値の揺らぎを実感してみよう。対象とする被検者をそれぞれ 10 人、2 群で 20 人とし、収縮期血圧の 2 群の平均が 155 mmHg、標準偏差を 10 mmHg とする。表 9.1 左に示すように、正規乱数を 20 個発生させて、20 人分の前値とし、2 群の平均値、 SD 、 t 値および p 値を求める。正規乱数 20 個を何回かセットで発生し直し、対応のない t 検定の結果のバリエーションを実感することが最初の一步である。

Excel で正規乱数を発生する方法は、2 通りある。一つは、「分析ツール」の「乱数」の中から「正規乱数」を選択し、Excel シート上に矩形状の正規乱数を発生させる方法である。ただし、新たな「正規乱数」を発生させるためには、同じ操作を手作業で繰り返す必要があり煩わしい。他の方法は、一様乱数 (0 から 1) を発生させる `Rand ()` 関数を用い、正規分布の逆関数 `Norm.Inv (Rand (), 0, 1)` の引数とする方法である。この方法により、平均が 0、標準偏差が 1 の標準正規乱数を Excel シートのセル上に 20 個を一括して得ることができる。ただし、Excel では、自動的に再計算する方法がデフォルトとして設定されているので、何か操作するたびに新たな正規乱数が目まぐるしく変化してしまい使い物にならない。そのために、「数式タブ」でデフォルトの自動計算を「手動」に変更する。その結果として新たな正規乱数のセットを発生するために「シート再計算 (Shift+F9)」を実行する必要がある。

表 9.1 に示すのは、平均が 0、標準偏差が 1 となる (20 行×1 列) の正規乱数 Z_{ij} を発生させ、2 群の共通の平均を $\mu=155$ 、標準偏差を $\sigma=10$ とする前値 X_{ij} を

$$\begin{aligned} Z_{ij} &= \text{Norm.Inv}(\text{Rand}(), 0, 1) \\ X_{ij} &= \mu + \sigma Z_{ij} = 155 + 10 Z_{ij} \end{aligned} \quad , \quad i=1,2, \quad j=1,2, \dots, 10 \quad (9.1)$$

によって計算している。さらに、`Round ()` 関数により小数点以下 1 桁で四捨五入した結果を示している。数式タブの「シート再計算」の実行により、正規乱数 Z_{ij} は全て変化し、もちろん前値 X_{ij} も変化する。 A_1 薬および A_2 薬の平均 μ_i を `Average (A_i の X 範囲)` 関数、標準偏差 σ_i を `Stdev.S (A_i の X 範囲)` 関数で計算する。平均値の差を $\mu_2 - \mu_1$ 、分散は 2 群で共通と仮定し、それぞれの σ_i から分散 $\sigma_{\text{共通}}^2$

$$\sigma_{\text{共通}}^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2} \quad (9.2)$$

を計算し、差の標準誤差 SE_{diff} を

表 9.1 正規乱数を用いた2群の前値Xの4通りの揺らぎ

2群共通の平均=		155		左の正規乱数を変化させた			
個体間SD=		10		場合の結果			
群	乱数を変化			特徴的な事例のスナップショット			
A	No	正規	前値	$\mu^{\wedge}_1 >> \mu^{\wedge}_2$	$\mu^{\wedge}_1 > \mu^{\wedge}_2$	$\mu^{\wedge}_1 < \mu^{\wedge}_2$	$\mu^{\wedge}_1 << \mu^{\wedge}_2$
i	j	乱数 Z	X	$X^{(1)}$	$X^{(2)}$	$X^{(3)}$	$X^{(4)}$
1	1	-0.40	151	148	158	160	151
1	2	0.18	157	154	163	163	156
1	3	-1.27	142	167	151	154	157
1	4	1.26	168	165	159	152	156
1	5	0.94	164	177	146	150	158
1	6	0.71	162	155	154	154	153
1	7	0.17	157	140	168	153	152
1	8	-1.23	143	160	169	155	132
1	9	1.35	169	154	158	137	145
1	10	-0.27	152	184	156	163	164
2	1	-1.10	144	140	156	142	168
2	2	-0.35	152	141	166	159	153
2	3	-0.54	150	134	135	165	156
2	4	0.66	162	162	160	168	168
2	5	-0.38	151	157	141	153	145
2	6	0.70	162	134	153	172	171
2	7	-0.33	152	150	165	159	154
2	8	1.18	167	161	154	144	180
2	9	-1.39	141	143	163	161	165
2	10	0.08	156	164	143	174	166
A ₁	平均 μ^{\wedge}_1		156.5	160.4	158.2	154.1	152.4
	SD σ^{\wedge}_1		9.5	13.2	7.1	7.5	8.7
A ₂	平均 μ^{\wedge}_2		153.7	148.6	153.6	159.7	162.6
	SD σ^{\wedge}_2		8.2	11.7	10.7	10.9	10.4
	平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$		-2.80	-11.80	-4.60	5.60	10.20
	差のSE		3.9730	5.5885	4.0770	4.1753	4.2880
	t値		-0.7048	-2.1115	-1.1283	1.3412	2.3787
	両側p値		0.4900	0.0490	0.2740	0.1965	0.0286
	再計算により常に変化する			*	NS	NS	*

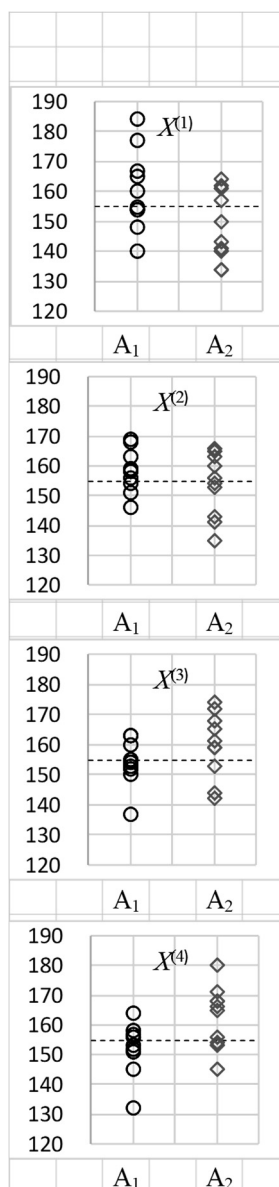


図 9.1 前値の平均値の揺らぎ

$$SE_{diff} = \sqrt{\frac{\sigma_{共通}^2}{n_1} + \frac{\sigma_{共通}^2}{n_2}} \quad (9.3)$$

によって求めている。対応のないt検定のt値を

$$t = \frac{\mu_2 - \mu_1}{SE_{diff}} \quad (9.4)$$

とし、両側p値をExcelの両側t分布のT.dist.2T()関数

$$p = T.dist.2T(Abs(t値), (n_1 + n_2 - 2)) \quad (9.5)$$

を用いて求めている。

前値 X の揺らぎの可視化

新しい乱数を「シート再計算」により発生し、表 9.1 左「前値 X 」の欄の「両側 p 値」に着目し、 A_1 薬の平均と A_2 薬の平均にかなり違いがある事例、若干の違いがある事例が得られるまで乱数の発生を繰り返す。適当な事例が現れた場合に「前値 X 」の列のデータ全体をスナップショットし、表 9.1 右の $X^{(1)}$ 、 $X^{(2)}$ 、 $X^{(3)}$ 、および、 $X^{(4)}$ の欄に「値」のみをペーストする。それぞれの $X^{(k)}$ に対し、図 9.1 に示すように Excel の散布図に 2 群の共通の期待平均値 155 の点線を重ね書きしている。 $X^{(1)}$ と $X^{(4)}$ は、前値 X について 2 群間で有意な差 ($p < 0.05$) があつた事例で、 $X^{(2)}$ と $X^{(3)}$ は、若干の違いがある事例である。前値をあらかじめ知った上で 2 群間の平均値と標準偏差がほぼ等しくなるような群分けが、実験の手順によりできない場合は、ここに示したような前値の比較で有意な差となつた場合でも、適切な群間比較ができるような解析手順を前もって定めておく必要がある。

前値 X について正規乱数を用い 2 群間の平均値の差のバリエーションを可視化した但、個体間の揺らぎの大きさを相対的に示すため CV (*Coefficient of Variation*, 変動係数)

$$CV = \frac{SD}{\text{平均値}} \times 100 \quad (9.6)$$

が用いられている。ここでのシミュレーションでは、

$$CV = \frac{SD_{\text{個体間}}}{\text{2群の共通の平均値}} = \frac{10}{155} \times 100 = 6.45\% \quad (9.7)$$

としている。

個体内の揺らぎに起因する時点間の相関

血圧などのバイタルサインは、同一の個体内で常に揺らいでいる。また、測定に伴う誤差も付きまとうので、複数回の測定結果の平均値などを用いることも一般的に行なわれている。試験終了時に測定した結果から、同一個体内で前値 X と後値 Y がセットで得られる。表 9.1 では、個体間の揺らぎに注目し、前値に対する揺らぎを可視化したのだが、実際には、個体間の揺らぎにさらに個体内の揺らぎが加わつた前値 X_{ij} および後値 Y_{ij} が得られる。

Excel シート上で正規乱数を発生させて得られたデータを用い、前値 X の 2 群間の平均値の揺らぎを図 9.1 で可視化し、群間で有意な差が時として出ることを例示した。個体間の揺らぎと個体内の揺らぎの大きさの割合は、目的とする反応変数によって大きく異なる。そこで、全体の揺らぎの分散を 100 (標準偏差で 10 mmHg) とした場合に、個体間の分散を 90、個体内の分散を 10、あるいは、(70 対 30) などのように比率を変えることにより、時点間の相関関係を変化させることができる。

表 9.2 に示すのは、分散を (50 対 50) の割合で設定した場合で、期待される時点間の相関 ρ は、

$$\rho = \frac{\sigma_{\text{個体間}}^2}{\sigma_{\text{個体間}}^2 + \sigma_{\text{個体内}}^2} = \frac{50}{50 + 50} = 0.50 \quad (9.8)$$

となり、個体間用の Z_0 、前値用の Z_1 、後値用の Z_2 とする正規乱数を発生させた結果が示されている。第 1 行目の個体間の正規乱数 $Z_{0,1} = 1.12$ を用い四捨五入のための Round () 関数により整数化された $(XY)_1 = 163$ 、個体内の正規乱数 $Z_{1,1} = -0.47$ を加味し、Round () 関数により整数化された前値 $X_1 = 160$ 、正規乱数 $Z_{2,1} = -0.28$ を加味した後値 $Y_1 = 161$ が示されている。

表 9.2 相関の期待値が 0.5 の場合の前値と後値の例示

	平均=	155	SD	相関の期待値			
個体間 分散=	50		7.07	0.50			
個体内 分散=	50		7.07				
	個体間		個体内の乱数		正規乱数を加味		
	正規 乱数 Z_0	個体間 (XY)	正規 Z_1	正規 Z_2	前値 X	後値 Y	前後差 $Y-X$
No							
1	1.12	163	-0.47	-0.28	160	161	1
2	-0.13	154	0.19	0.28	155	156	1
3	2.36	172	-1.20	-0.01	164	172	8
4	0.47	158	0.19	-0.05	159	158	-1
5	-1.13	147	0.40	0.52	150	151	1
6	0.27	157	-1.68	0.28	145	159	14
7	-0.59	151	0.20	0.85	152	157	5
8	1.36	165	1.18	-0.52	173	161	-12
9	1.06	162	0.15	-1.43	163	152	-11
10	-0.39	152	-1.50	-0.50	141	148	7
				平均=	156.20	157.50	1.30
				SD=	9.58	6.69	8.07
				相関 $r =$	0.5568		

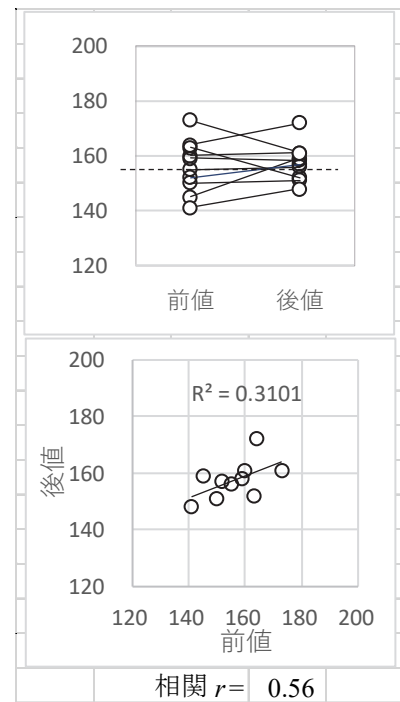


図 9.2 相関が 0.56 の場合の線グラフおよび散布図

これらの正規乱数は、Norm.Inv (Rand (), 0, 1) 関数により平均が 0.0、標準偏差が 1.0 の標準正規分布に従うように引数を設定している。個体間の標準偏差 $\sqrt{\sigma_{\text{個体間}}^2} = \sqrt{50}$ 、および、個体内の標準偏差 $\sqrt{\sigma_{\text{個体内}}^2} = \sqrt{50}$ に対し、

$$\left. \begin{aligned} (XY)_1 &= \mu + Z_{0,1} \sqrt{50} = 155 + 1.12 \times 7.07 = 163 \\ X_1 &= (XY)_1 + Z_{1,1} \sqrt{50} = 163 - 0.47 \times 7.07 = 160 \\ Y_1 &= (XY)_1 + Z_{2,1} \sqrt{50} = 163 - 0.28 \times 7.07 = 161 \end{aligned} \right\} \quad (9.9)$$

のように計算されている。残りの 9 例分を同様に計算し、Excel の Correl (X の範囲, Y の範囲) 関数により時点間の相関係数が $r = 0.5568$ が計算されている。図 9.2 には、前値と後値に関する線グラフおよび散布図により前値と後値の関係が示されている。なお、表 9.2 に示したのは、正規乱数を何回か発生させ、時点間の相関係数の期待値 0.50 に近い相関係数となっ

た場合についてスナップショットし、正規乱数をコピーし「値」のみで同じセル上にペーストすることにより、固定化した結果である。

図 9.3 に示すのは、全体の分散が 100 となるように（個体間 対 個体内）の分散を、（90 対 10），（70 対 30），（30 対 70），（10 対 90）に変更し、それぞれについて正規乱数を何回か発生した中から、期待される時点間の相関に近くなった線グラフと散布図をスナップショットし、図（メタファイル）として固定化した結果である。時点間の相関が小さくなるにつれて前値と後値を結ぶ線が互いにクロスする頻度が多くなることが観察される。また、散布図と線グラフは、連動しているので前値と後値の時点間の相関の違いによる形状の変化を実感してもらいたい。

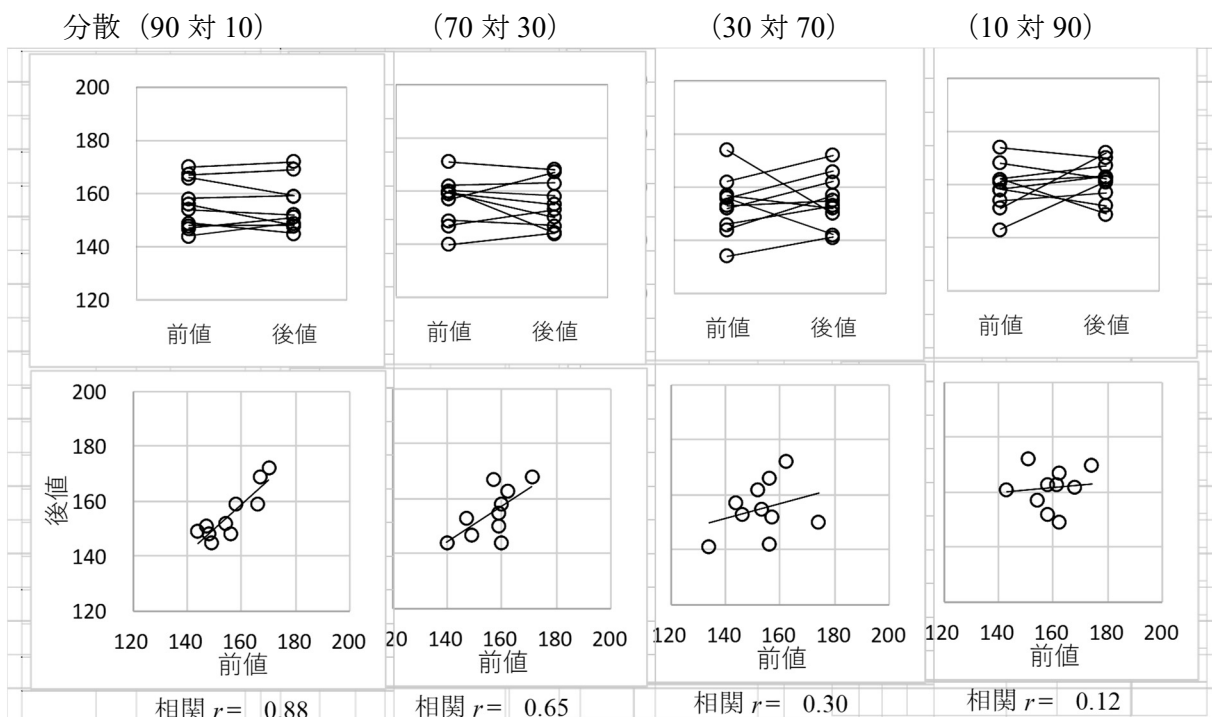


図 9.3 各種の時点間の相関に対応する線グラフおよび散布図

前後差 D の分散のインフレーション

前値 X と後値 Y の時点間の相関に関連して、前後差 $D = Y - X$ の分散 $Var(Y - X)$ が大きく変化する。前後差の相関が 1.0 に近づけば前後差 D の分散は、後値 Y の分散 $Var(Y)$ に対して相対的に小さくなり、前後差 D の分散は 0.0 に近づく。前後差 D の分散は、合成分散の一般式から、

$$Var(Y - X) = Var(Y) - 2Cov(X, Y) + Var(X) \quad (9.10)$$

として求められる。ここで、共分散 $Cov(X, Y)$ は、

$$\rho = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (9.11)$$

で定義されており， $\sigma_X = \sigma_Y$ なので，

$$Cov(X, Y) = \rho \cdot \sigma_X \cdot \sigma_Y = \rho \cdot \sigma^2 \tag{9.12}$$

のように相関係数 ρ と分散 σ^2 の積となる。

式 (9.12) を使い，時点間の相関が $\rho = 0.50$ の場合は，前後差 D の分散は，

$$\left. \begin{aligned} Var(Y - X) &= Var(Y) - 2Cov(Y, X) + Var(X) \\ &= \sigma_Y^2 - 2 \times 0.50 \times \sigma^2 + \sigma_X^2 = \sigma^2 \end{aligned} \right\} \tag{9.13}$$

と後値 Y の分散と同じになる。時点間の相関が 0.0 に近づけば後値 Y の分散に比べ前後差 D の分散は相対的に大きくなる。前値 X と後値 Y の時点間の相関が 0.0 の場合は，前後差の分散は，分散の加法性により

$$\left. \begin{aligned} Var(Y - X) &= Var(Y) + Var(X) \\ &= \sigma_Y^2 + \sigma_X^2 = 2\sigma^2 \end{aligned} \right\} \tag{9.14}$$

なので後値 Y の分散の 2 倍となる。このことは，前後差 D での群間比較について有意差検定を行うことが，いつでも有意な差が出やすくなるとは限らないことが示唆されている。

時点間の相関の期待値が $\rho = 0.3$ の場合

表 9.3 に示すのは，時点間の相関係数の期待値が 0.3 となるように個体間の分散を 30 ，個体内の分散を 70 とし，正規乱数を何回か発生し，相関が $\rho = 0.30$ に近くなった場合をスナップショットし，正規乱数を固定化した結果である。

表 9.3 時点間の相関の期待値が 0.3 の場合

		平均=	155	SD	相関の期待値			
個体間 分散=		30		5.48				
個体内 分散=		70		8.37	0.30			
		個体間		個体内の乱数		正規乱数を加味		
		正規	個体間	正規	正規	前値	後値	前後差
No	乱数 Z_0	(XY)	Z_1	Z_2	X	Y	$Y - X$	
1	-0.86	150	1.70	1.46	164	162	-2	
2	1.48	163	-1.03	0.44	154	167	13	
3	1.19	161	-0.04	0.06	161	162	1	
4	-0.63	152	-0.47	0.37	148	155	7	
5	-1.47	147	1.60	1.59	160	160	0	
6	0.76	159	0.20	1.06	161	168	7	
7	0.86	160	0.22	0.44	162	164	2	
8	0.29	157	0.91	-0.40	165	154	-11	
9	-0.97	150	1.66	1.14	164	160	-4	
10	0.98	160	-1.25	-0.96	150	152	2	
					平均=	158.90	160.40	1.50
					SD=	6.06	5.38	6.62
					相関 $r =$	0.3352		

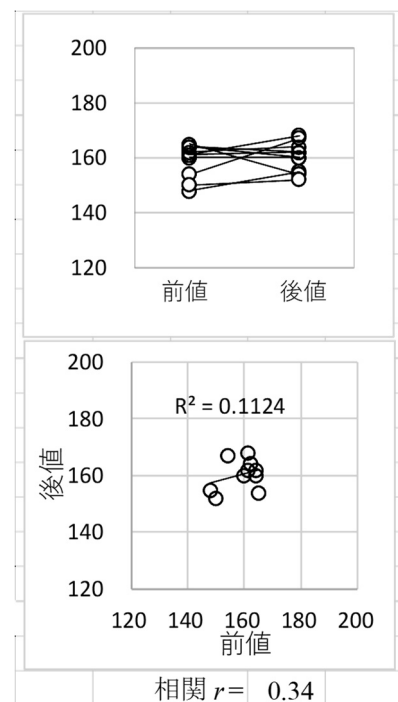


図 9.4 時点間の相関が 0.34 の場合の散布図

前後差 $D=Y-X$ の SD を計算すると、 $SD_D=6.62$ となり、後値の $SD_Y=5.32$ に比べ大きくなっている。これらのことから、前後差 D による統計解析が、後値 Y のみの解析に対し、前値 X の影響を取り除くための方法として適しているとは、必ずしも言い切れないことを示唆している。

表 9.3 に示したのは、時点間の相関係数が $\rho=0.30$ に近くなった場合のスナップショットの一例である。表 9.4 に示すのは、正規乱数を連続的に発生させ、表 9.3 で計算させた統計量を 1 行に要約した 10 回分の結果である。前後の相関係数の 10 回分の平均は、0.26 となり、おおよそ 0.30 となっている。前値 X と後値 Y の平均の平均は、(155.36, 155.32) であり、シミュレーションで設定した 155.0 に近い値となっている。

表 9.4 相関係数の期待値が 0.3 の場合の SD_Y と SD_D の比較

正規乱数 発生順	前後 相関	前値 平均 X	後値 平均 Y	前後差 平均 D	前値 SD_X	後値 SD_Y	前後差 SD_D	SD_Y vs. SD_D 大小
1	0.34	158.9	160.4	1.5	6.06	5.38	6.62	<
2	0.53	156.3	154.4	-1.9	10.58	7.15	9.12	<
3	0.51	155.2	150.0	-5.2	5.20	9.30	7.98	>
4	-0.06	156.2	155.1	-1.1	7.24	9.22	12.04	<<
5	0.63	153.6	151.8	-1.8	10.80	10.11	9.07	>
6	-0.22	154.7	156.2	1.5	10.95	10.71	16.91	<<<
7	0.72	153.0	155.8	2.8	7.80	11.68	8.08	>>
8	-0.56	152.3	155.2	2.9	9.09	6.84	14.11	<<<
9	0.08	155.0	156.1	1.1	8.76	12.49	14.66	<<
10	0.62	158.4	158.2	-0.2	9.02	12.66	9.95	>>
平均	0.26	155.36	155.32	-0.04	8.55	9.55	10.85	<

前値 X と後値 Y の SD のそれぞれの平均は、(8.55, 9.55) であり、個体内の SD に個体間の SD を加味した結果となっている。さて、前後差の SD_D は 10.85 となり、後値の $SD_Y=9.55$ に比べ大きくなっている。正規乱数の発生順に後値の SD_Y と前後差の SD_D を比較した場合の大小関係は、一定はしないが 6 対 4 で前後差の SD_D が大きく、4.0 以上の差 (<<<) となる場合も 2 回あり、全 10 回のみシミュレーションではあるが、後値の SD_Y に比べて前後差の SD_D が大きくなる傾向を実感できるであろう。

表 9.4 に示したように時点間の相関係数の期待値を 0.30 とした場合、シミュレーションの結果の相関係数 r は (-0.56 から 0.72) とかなり広い範囲に揺らいでいる。したがって、1 回の実験で得られた相関係数の大きさにだけにより、後値 Y での解析か前後差 D での解析かを選択することは慎まなければならない。ここに示したように、同じ条件で 10 回程度の繰り返し実験でデータを収集し、それらの時点間の相関係数の平均値を算出することで、ようやく相関の期待値 0.30 に近づくことが認識できる。

9.3. 後値 Y および前後差 D に対する群間比較

前値 X の揺らぎが正規分布に従い、20 症例を 2 群にランダムに割り当てた場合に、2 群間の平均値が、有意な差となる場合もあることを表 9.1 に示した。同一個体内で前値 X と後値 Y がセットで観察されるような場合に、人間を含む動物を対象にした血圧などのバイタルサイン、コレステロールなどの血液生化学検査では、 X と Y に正の相関があることが経験的に知られている。

「平均への回帰」現象

前値 X にいくばくかの群間差（例えば、両側検定で $p < 0.40$ ）であるような場合に、前値の平均が大きい群は、後値の平均値が前値 X よりさらに大きくなることはまれで、確率的に小さくなる。逆に前値 X の平均が小さい群は、後値 Y の平均値が前値よりさらに小さくなることはまれで、平均値が確率的に大きくなる。

いわゆる「平均への回帰」現象で、身長の高い親から生まれる子供が親よりもさらに大きくなるよりも小さくなる確率が大きく、逆に身長の小い親から生まれる子供が親よりもさらに小さくなるよりも大きくなる確率が大きくなることが知られている。親の身長の平均値に向かって子供の身長が回帰する現象であり、前値 X と後値 Y の関係にも同様な「平均への回帰現象」が内在する。より詳しくは、岩崎（2002）、「『処置前—処置後』データの解析と平均への回帰」を参照のこと。

処置の効果が全くない2群における前値 X と後値 Y を想定しよう。2群に共通の平均を 155、時点間の相関が 0.5 となるように個体間の分散を $\sigma_{\text{個体間}}^2 = 50$ 、個体内の分散を $\sigma_{\text{個体内}}^2 = 50$ とする。正規乱数を用いたシミュレーションにより、前値 X 、後値 Y および前後差 D について 2 群間の差について t 検定を行ない「平均への回帰現象」を可視化する。

1) 前後差 D にのみ偶然に有意な差が出る場合

表 9.5 に示すのは、表 9.1 と表 9.2 とを併合し、後値 Y および前後差 D について 2 群間の t 検定を追加し、 $2 \times 10 \times 3$ の正規乱数を逐次発生させ、前後差 D にのみ偶然に有意な差が出る場合をスナップショットした結果である。ただし、スナップショットした結果を固定するために、正規乱数の関数が入っているセル全体をコピーし、同じ場所に「値のみ」をペーストし固定化した結果である。

前値 X の平均は A₁ で 158.2, A₂ で 151.2 と群間差が -7.0 あるが, 検定の結果は $p=0.1974$ である. 後値 Y の t 検定の結果は $p=0.8111$ であり, 線グラフからも平均値が重なって見える. ただし, 前後差 D では, A₁ で -3.0, A₂ で +5.3 と平均値の差が 8.30 に拡大し, t 検定の結果は $p=0.0263^*$ と有意な差となっている. これは, 典型的な「平均への回帰」現象が現れた場合である. 後値 Y では, 有意な差がないことを示さずに, 前後差 D で有意な群間差が得られたことのみを強調したい誘惑にかられる.

表 9.5 前後差 D にのみ偶然に有意な差が出現する一例

2群共通の平均=		155		SD					
個体間 分散=		50		7.07		A ₁ 群効果		0.0	
個体内 分散=		50		7.07		A ₂ 群効果		0.0	
個体間				個体内の乱数					
A	No	正規	個体間	正規	正規	前値	後値	差	
i	j	乱数 Z ₀	XY	Z ₁	Z ₂	X	Y	D	
1	1	2.41	172	0.79	1.20	178	180	2	
1	2	-0.66	150	0.65	0.34	155	152	-3	
1	3	2.01	169	1.42	0.78	179	174	-5	
1	4	-0.71	150	0.31	0.96	152	157	5	
1	5	-1.58	144	0.42	-1.44	147	134	-13	
1	6	1.03	162	0.24	-1.37	164	152	-12	
1	7	0.14	156	-0.24	-0.37	154	153	-1	
1	8	0.02	155	0.87	-0.78	161	149	-12	
1	9	0.19	156	-1.10	-1.09	148	148	0	
1	10	-1.29	146	-0.21	0.94	144	153	9	
2	1	-2.04	141	0.72	0.37	146	144	-2	
2	2	-1.74	143	-0.78	0.91	137	149	12	
2	3	0.45	158	-0.06	2.00	158	172	14	
2	4	1.26	164	-0.52	0.81	160	170	10	
2	5	-0.75	150	-1.59	0.66	139	155	16	
2	6	1.34	164	-0.17	-0.82	163	158	-5	
2	7	-1.15	147	-0.67	-1.03	142	140	-2	
2	8	1.33	164	0.82	0.44	170	167	-3	
2	9	-1.36	145	0.29	1.05	147	152	5	
2	10	-0.17	154	-0.61	0.61	150	158	8	
A ₁				平均 μ^{\wedge}_1		158.2	155.2	-3.0	
				SD σ^{\wedge}_1		12.3	13.1	7.5	
A ₂				平均 μ^{\wedge}_2		151.2	156.5	5.3	
				SD σ_2		11.0	10.8	7.8	
				平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$		-7.0	1.3	8.3	
				差の SE		5.23	5.36	3.43	
				t 値		-1.3386	0.2425	2.4207	
				両側 p 値		0.1974	0.8111	0.0263	
						NS	NS	*	

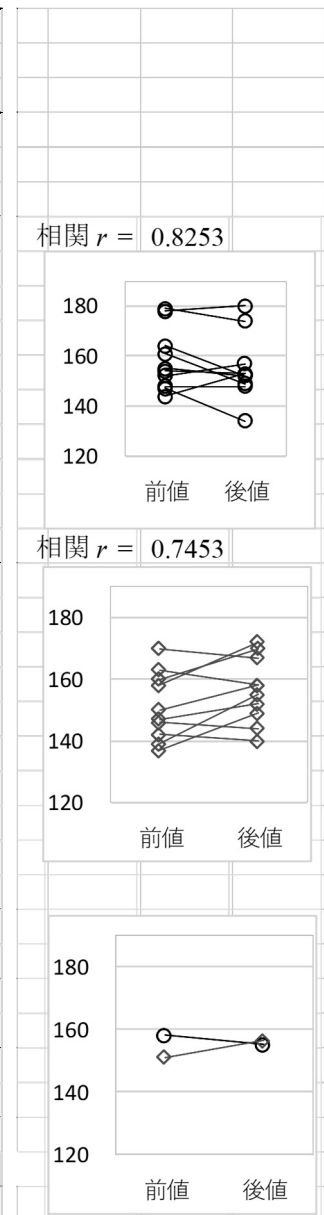


図 9.5 前後差 D にのみ有意な差

2) 後値 Y にのみ偶然に有意な差が出る場合

表 9.6 に示すのは、後値 Y に対し A₁ の平均が 154.2, A₂ の平均が 160.9 と差が 6.7 であり、後値 Y にのみ有意な差が得られた場合である。前後差 D は、A₁ の平均が -3.1, A₂ の平均が +2.5 と平均値の差が 5.6 であり、前後差 D が後値 Y よりも 5.6 - 6.7 = -1.1 とわずかに小さいために有意な差にならなかった場合である。前後差 D で結果を評価しようと考えていた場合に、それをやめて後値 Y で有意な群間差があったと言いたい誘惑にかられる。誘惑に負けた場合には、後ろめたさが、常に付きまとうことになる。

表 9.6 後値 Y のみ有意な差が出る場合

2群共通の平均=		155		SD					
個体間 分散=		50		7.07		A ₁ 群効果		0.0	
個体内 分散=		50		7.07		A ₂ 群効果		0.0	
		個体間		個体内の乱数					
A	No	正規	個体間	正規	正規	前値	後値	差	
i	j	乱数 Z ₀	XY	Z ₁	Z ₂	X	Y	D	
1	1	-0.81	149	1.12	0.27	157	151	-6	
1	2	-0.66	150	1.75	-0.19	162	149	-13	
1	3	-0.38	152	0.14	-0.36	153	149	-4	
1	4	-0.12	154	-0.38	1.09	151	162	11	
1	5	0.84	161	1.07	-0.14	169	160	-9	
1	6	-1.29	146	0.49	0.32	149	148	-1	
1	7	0.37	158	-0.05	-0.54	158	154	-4	
1	8	0.17	156	-0.12	-0.63	155	152	-3	
1	9	1.02	162	-1.05	0.39	155	165	10	
1	10	-0.67	150	1.92	0.33	164	152	-12	
2	1	1.14	163	0.98	-0.23	170	161	-9	
2	2	-1.11	147	0.43	1.19	150	155	5	
2	3	0.18	156	-1.13	1.14	148	164	16	
2	4	0.50	159	-1.11	0.21	151	160	9	
2	5	1.64	167	0.41	0.06	170	167	-3	
2	6	-0.58	151	0.68	1.83	156	164	8	
2	7	0.12	156	0.24	-0.18	158	155	-3	
2	8	-0.46	152	-0.37	0.12	149	153	4	
2	9	1.36	165	0.77	0.58	170	169	-1	
2	10	1.21	164	-0.33	-0.44	162	161	-1	
A ₁				平均 μ^{\wedge}_1		157.3	154.2	-3.1	
				SD σ^{\wedge}_1		6.2	6.0	8.1	
A ₂				平均 μ^{\wedge}_2		158.4	160.9	2.5	
				SD σ_2		9.1	5.3	7.3	
				平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$		1.1	6.7	5.6	
				差のSE		3.47	2.54	3.46	
				t 値		0.3166	2.6427	1.6185	
				両側 p 値		0.7552	0.0165	0.1229	
						NS	*	NS	

相関 $r = 0.1034$

相関 $r = 0.5955$

図 9.6 後値 Y にのみ有意な差

3) 後値 Y および前後差 D 共に有意な差が出る場合

表 9.7 に示すのは、後値 Y および前後差 D 共に有意な差が出る場合であり、このような結果が得られたならば、前後差 D で有意な差が出たと結果を報告したとしても、後日、第 3 者から「後値 Y ではどうか」と尋ねられる可能性もあるが、安心していただけるのであろうか。では、いつでも後値 Y および前後差 D 共に有意な差が出る場合のみ、有意な差があったとする保守的な対応が適切なのだろうか。

表 9.7 後値 Y および前後差 D 共に有意な差が出る場合

2群共通の平均=		155		SD					
個体間 分散=		50		7.07		A ₁ 群効果		0.0	
個体内 分散=		50		7.07		A ₂ 群効果		0.0	
		個体間		個体内の乱数					
A	No	正規	個体間	正規	正規	前値	後値	差	
i	j	乱数 Z ₀	XY	Z ₁	Z ₂	X	Y	D	
1	1	0.31	157	-1.69	-0.57	145	153	8	
1	2	-0.07	154	-0.46	0.48	151	157	6	
1	3	0.80	161	-1.16	0.10	153	162	9	
1	4	-0.05	155	-0.65	1.18	150	163	13	
1	5	-1.85	142	0.63	0.66	146	147	1	
1	6	-0.19	154	-1.38	0.18	144	155	11	
1	7	1.63	167	0.03	-0.31	167	165	-2	
1	8	-0.06	155	-0.40	-0.64	152	150	-2	
1	9	1.71	167	-0.03	-0.56	167	163	-4	
1	10	-0.90	149	-0.62	0.01	145	149	4	
2	1	-1.03	148	-1.05	0.06	141	148	7	
2	2	-0.79	149	-0.11	-1.35	148	139	-9	
2	3	0.05	155	0.63	-1.47	159	145	-14	
2	4	0.60	159	-0.11	-1.52	158	148	-10	
2	5	1.29	164	0.52	-1.56	168	153	-15	
2	6	-0.69	150	-0.80	-0.54	144	146	2	
2	7	-0.38	152	-0.14	-0.99	151	145	-6	
2	8	-0.46	152	-0.72	0.96	147	159	12	
2	9	0.22	157	-0.25	-0.43	155	154	-1	
2	10	-0.70	150	0.53	-0.26	154	148	-6	
A ₁					平均 μ ¹	152.0	156.4	4.4	
					SD σ ¹	8.5	6.6	5.9	
A ₂					平均 μ ²	152.5	148.5	-4.0	
					SD σ ₂	8.0	5.6	8.9	
					平均の差 μ ² - μ ¹	0.50	-7.90	-8.40	
					差の SE	3.70	2.73	3.38	
					t 値	0.1351	-2.8892	-2.4825	
					両側 p 値	0.8940	0.0098	0.0231	
						NS	*	*	

相関 r = 0.7184

前値 後値

相関 r = 0.1843

前値 後値

前値 後値

図 9.7 後値 Y と前後差 D に有意な差

9.4. 前値 X を共変量とした後値 Y および前後差 D の解析

前値を共変量とする解析

各種のシミュレーションを通じ、前値がある場合の経時データの解析の困難さについて示してきた。前値 X と後値 Y がある場合の群間比較に際しては、前値 X を共変量とした共分散分析が適していると生物統計の名著、スネデカーら（1972）で指摘されているが、他の成書ではほとんど見かけない。一般的に経時データの解析では、前後差 D での群間比較が簡便かつ容易に行なえるためでもあり、前値の揺らぎを消すための合理的な方法と思われる。

他方、反応に影響があると分かっている変数であるが、実験開始時に制御できない変数を共変量とする共分散分析が知られている。第6章で示したように、共分散分析は、単回帰分析を重層的に使う伝統的な計算方法が連綿と引き継がれ、さらに、交互作用の検討など探索的な手順が定式化されており、計算が難しい厄介な方法として認識されている。また、前値 X を共変量、前後差 D を反応とする共分散分析については、その振る舞いについて十分に知られてこなかった。

前値 X を共変量とする解析の有用性は、Frison, Pocock (1992) の「Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and its Implications for Design」によって示された。この論文は、2群の投与前値を含む経時データについて、投与後値のみを使う群間比較 POST, 投与前から投与後の差についての群間比較 CHANGE, 投与前値を共変量とした ANCOVA について、時点間の相関が変化した場合に、POST vs ANCOVA, および CHANGE vs ANCOVA をそれぞれ比較し、ANCOVA が時点間の相関にかかわらず常に群間比較の際の分散が小さくなるので、標準的に使うことを薦めている。

そこで、第9.3節で取り上げた3種の事例について前値 X を共変量とする解析方法を追加し、対応のない t 検定の結果と対比する。共分散分析の適用に際し、因子 A に (0, 1) 型ダミー変数 a を用いることにより、Excel の回帰分析のパラメータ（係数）についての t 検定が、 A_1 群と A_2 群の差についての対応のない t 検定と同様の結果となる。

1) 前後差 D にのみ偶然に有意な差が出る場合

表 9.5 に示したのは、後値 Y では有意な差ではなく、前後差 D にのみ有意な差が出る場合であった。この事例に対し、表 9.8 に示すように、切片、因子 A を (0, 1) 型ダミー変数 a ,

前値 X を共変量とする (20 行×3 列) のデザイン行列 X を Excel シート上に設定する. 反応変数として後値 Y と前後差 D を並べて示してある. 次の 2 つの線形モデル

$$\left. \begin{aligned} \text{後値 } Y_{ij} &= \beta_0 x_0 + \beta_1 a_{ij} + \beta_2 X_{ij} + \varepsilon_{ij} \\ \text{前後差 } D_{ij} &= \beta'_0 x_0 + \beta'_1 a_{ij} + \beta'_2 X_{ij} + \varepsilon_{ij} \end{aligned} \right\} \quad (9.15)$$

を適用し, 因子 A の群間差 β_1 および β'_1 についての結果を比較する. 反応変数が Y と D で異なるが, デザイン行列 X は全く同じである. なお, Excel の回帰分析では, 変数 x_0 を必要としないが, パラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算のために必要なので明示的に示している.

表 9.8 後値 Y と前後差 D に対する前値 X を共変量とした Excel の回帰分析

A	No	デザイン行列 X			後値	前後差						
i	j	x_0	a	X	Y	D	分散分析表 反応変数: <<後値 Y >>					
							自由度	変動	分散	分散比		
1	1	1	0	178	180	2						
1	2	1	0	155	152	-3	回帰	2	1624.91	812.4546	14.2442	
1	3	1	0	179	174	-5	残差	17	969.64	57.0377		
1	4	1	0	152	157	5	合計	19	2594.55			
1	5	1	0	147	134	-13						
1	6	1	0	164	152	-12						
1	7	1	0	154	153	-1	β_0^{\wedge}	切片	26.9919	24.2013	1.1153	0.2802
1	8	1	0	161	149	-12	β_1^{\wedge}	a	6.9729	3.5416	1.9688	0.0655
1	9	1	0	148	148	0	β_2^{\wedge}	X	0.8104	0.1522	5.3236	0.0001
1	10	1	0	144	153	9						
2	1	1	1	146	144	-2						
2	2	1	1	137	149	12	分散分析表 反応変数: <<前後差 D >>					
2	3	1	1	158	172	14						
2	4	1	1	160	170	10	回帰	2	432.91	216.4546	3.7949	
2	5	1	1	139	155	16	残差	17	969.64	57.0377		
2	6	1	1	163	158	-5	合計	19	1402.55			
2	7	1	1	142	140	-2						
2	8	1	1	170	167	-3						
2	9	1	1	147	152	5	β'_0^{\wedge}	切片	26.9919	24.2013	1.1153	0.2802
2	10	1	1	150	158	8	β'_1^{\wedge}	a	6.9729	3.5416	1.9688	0.0655
1		平均 μ^{\wedge}_1		158.2	155.2	-3.0	β'_2^{\wedge}	X	-0.1896	0.1522	-1.2453	0.2299
		$SD \ \sigma^{\wedge}_1$		12.3	13.1	7.5						
2		平均 μ^{\wedge}_2		151.2	156.5	5.3	後値 Y		前後差 D			
		$SD \ \sigma^{\wedge}_2$		11.0	10.8	7.8	a		a			
		平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$		-7.0	1.3	8.3	係数		6.9729	6.9729		
		差の SE		5.23	5.36	3.43	標準誤差		3.5416	3.5416		
		t 値		-1.339	0.2425	2.4207	t		1.9688	1.9688		
		両側 p 値		0.1974	0.8111	0.0263	p - 値		0.0655	0.0655		
				NS	NS	*			+	+		
		2群込みの平均		154.70								

なお, 前後差 $D = Y - X$ の場合は, 左辺の X_{ij} を右辺に移項し整理すると

$$\left. \begin{aligned} (Y_{ij} - X_{ij}) &= \beta'_0 x_0 + \beta'_1 a_{ij} + \beta'_2 X_{ij} + \varepsilon_{ij} \\ Y_{ij} &= \beta'_0 x_0 + \beta'_1 a_{ij} + \beta'_2 X_{ij} + X_{ij} + \varepsilon_{ij} \\ Y_{ij} &= \beta_0 x_0 + \beta_1 a_{ij} + (\beta'_2 + 1)X_{ij} + \varepsilon_{ij} \end{aligned} \right\} \quad (9.16)$$

となり、

$$\hat{\beta}'_0 = \hat{\beta}_0, \quad \hat{\beta}'_1 = \hat{\beta}_1, \quad \hat{\beta}'_2 = \hat{\beta}_2 - 1 \quad (9.17)$$

の関係となることは自明である。したがって、後値 Y あるいは前後差 D に対し、前値 X を共変量とした共分散分析において、因子 A の群間差のダミー変数 a のパラメータ $\hat{\beta}_1$ は、全く同じ推定値となる。

表 9.8 右上段に示したのは、後値 Y に対する Excel の回帰分析の結果で、分散分析表およびデザイン行列 X の変数に対するパラメータの推定値 $\hat{\beta}_0 = 26.9919$, $\hat{\beta}_1 = 6.9729$, $\hat{\beta}_2 = 0.8104$ (係数と表示) が示されている。中段には、同様に前後差 D に対する分散分析表と係数の推定値 $\hat{\beta}'_0 = 26.9919$, $\hat{\beta}'_1 = 6.9729$, $\hat{\beta}'_2 = -0.1896$ が示されている。前後差 D の傾きに関するパラメータ $\hat{\beta}'_2$ は、後値 Y のパラメータ $\hat{\beta}_2$ から 1 を引いて

$$\hat{\beta}'_2 = \hat{\beta}_2 - 1 = 0.8104 - 1 = -0.1896 \quad (9.18)$$

として求められる。

表 9.8 右下段には、後値 Y および前後差 D に関する因子 A の変数 a の列方向の (係数, 標準誤差, t , p -値) が、回帰のパラメータに対する t 検定の (平均の差 $\mu_1 - \mu_2$, 差の SE , t 値, p 値) に対応しているので、列方向の並びを行方向に並べ替えて貼り付けた結果である。前値 X を共変量とした場合に、後値 Y および前後差 D の検定統計量が、まったく同じ結果となっていることが確認できる。

表 9.8 に示した結果を解釈するために、図 9.8 に示すように A_1 群および A_2 群を合わせた散布図上に推定されたそれぞれの群の回帰直線を重ね書きする。回帰直線の始点と終点は、それぞれの群の前値 X の最小値と最大値に対し、推定されたパラメータ (係数) を用いて後値 Y の推定値、および、前後差 D の推定値を別途計算し、それらの 2 点を結ぶ直線を Excel の散布図上に重ね書きした結果である。Excel の散布図の良さは、必要に応じて座標を別途設定することにより自由に (点・直線・曲線) を重ね書きできることにある。

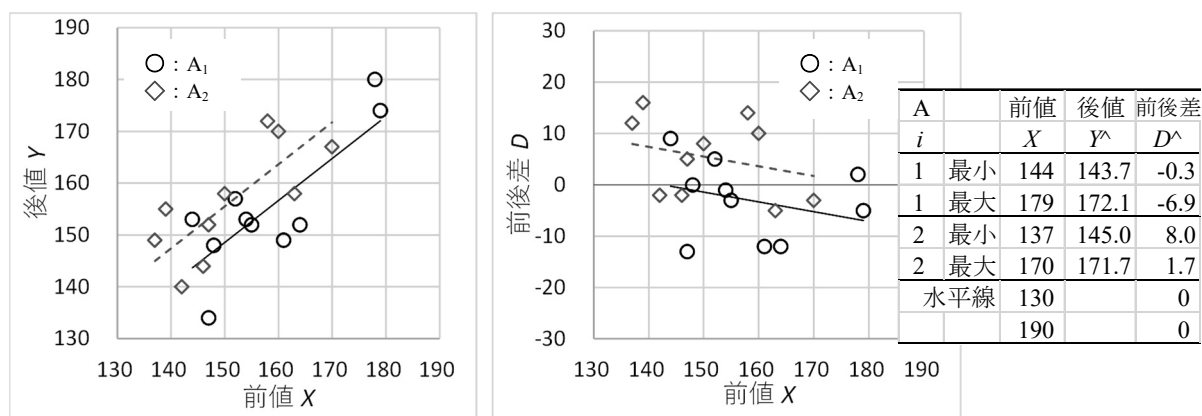


図 9.8 前値 X を共変量とした後値 Y と前後差 D の A_1 群および A_2 群の回帰直線

後値 Y に対する前値 X を共変量とした場合の推定された回帰直線は、表 9.8 右上段の結果から

$$A_1 \text{ 群} : \begin{cases} \hat{y}_1 = 26.9919 + 0 \times 6.9729 + 0.8104x \\ = 26.9919 + 0.8104x \end{cases} \quad (9.19)$$

$$A_2 \text{ 群} : \begin{cases} \hat{y}_2 = 26.9919 + 1 \times 6.9729 + 0.8104x \\ = 33.9648 + 0.8104x \end{cases} \quad (9.20)$$

となり、直線間の差は、因子 A のダミー変数 a のパラメータの推定値 $\hat{\beta}_1 = 6.9729$ が、 A_1 群および A_2 群の差の推定値となり、 $\hat{\beta}_1$ の $SE = 3.5416$ で除した、

$$t = \frac{\text{推定値}\hat{\beta}_1}{\hat{\beta}_1 \text{ の } SE} = \frac{6.9729}{3.5416} = 1.9688 \quad (9.21)$$

t 値が、分散分析表の誤差の自由度 $df = 17$ の t 分布に従うことから

$$\left. \begin{aligned} p &= t_{0.05}(t, 17) \\ &= \text{T.dist.2T}(\text{Abs}(1.9688), 17) = 0.0655 \end{aligned} \right\} \quad (9.22)$$

が得られ、 $p < 0.10$ との結果が得られている。

さて、前後差 D について前値 X を共変量とした場合の推定された回帰直線は、

$$A_1 \text{ 群} : \begin{cases} \hat{d}_1 = 26.9919 + 0 \times 6.9729 - 0.1896x \\ = 26.9919 - 0.1896x \end{cases} \quad (9.23)$$

$$A_2 \text{ 群} : \begin{cases} \hat{d}_2 = 26.9919 + 1 \times 6.9729 - 0.1896x \\ = 33.9648 - 0.1896x \end{cases} \quad (9.24)$$

となり、前値 X を固定した場合の直線間の差は、後値 Y の場合と同様に因子 A のダミー変数 a の推定値 $\hat{\beta}_1 = 6.9729$ が、共分散分析における A_1 群および A_2 群の差の推定値となり、同じ t 値となり、もちろん p 値も等しくなる。なお、この関係は、どのような前値 X と後値 Y に対しても一般的に成り立つ。

共分散分析における最小 2 乗平均

前値 X を共変量とした A_1 群と A_2 群の差は、図 9.8 に示した 2 本の直線の Y 軸方向の差として表されることを示した。後値 Y でも前後差 D でも A_1 群および A_2 群について平均と SD または SE が要約統計量として使われているが、前値 X を共変量とした場合の各群に対する要約統計量はどのように表したらよいのであろうか。SAS および JMP では、 A_1 群および A_2 群の“平均”に代わるものとして、“最小 2 乗平均 (lsmeans)” が用いられている。図 9.9 左に示すように最小 2 乗平均は、 A_1 群および A_2 群を併合した前値 X の総平均 $\bar{x}_{..} = 154.70$ に対する 2 本の回帰直線

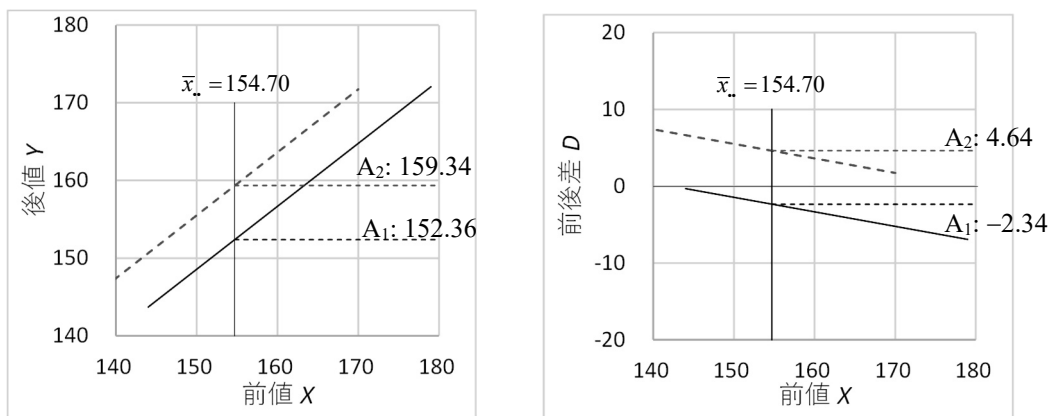


図 9.9 最小 2 乗平均の例示

$$A_1 \text{ 群} : \begin{cases} \hat{y}_1 = \hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 \bar{x}_{..} \\ = 26.9919 + 6.9729 \times 0 + 0.8104 \times 154.70 = 152.36 \end{cases} \quad (9.25)$$

$$A_2 \text{ 群} : \begin{cases} \hat{y}_2 = \hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 \bar{x}_{..} \\ = 26.9919 + 6.9729 \times 1 + 0.8104 \times 154.70 = 159.34 \end{cases} \quad (9.26)$$

の推定値として計算される。

図 9.9 右に示すように前後差 D についても同様に，“最小 2 乗平均”を

$$A_1 \text{ 群} : \begin{cases} \hat{d}_1 = \hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2' \bar{x}_{..} \\ = 26.9919 + 6.9729 \times 0 - 0.1896 \times 154.70 = -2.34 \end{cases} \quad (9.27)$$

$$A_2 \text{ 群} : \begin{cases} \hat{d}_2 = \hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2' \bar{x}_{..} \\ = 26.9919 + 6.9729 \times 1 - 0.1896 \times 154.70 = 4.64 \end{cases} \quad (9.28)$$

として求めることができる。

最小 2 乗平均の 95%信頼区間

これらの推定値“最小 2 乗平均”の 95%信頼区間は，どのようにして求めるのだろうか。必要なのは，それぞれの推定値の分散 $Var(\hat{y}_{x=\bar{x}_{..}}^{(1)})$ ， $Var(\hat{y}_{x=\bar{x}_{..}}^{(2)})$ ， $Var(\hat{d}_{x=\bar{x}_{..}}^{(1)})$ ，および， $Var(\hat{d}_{x=\bar{x}_{..}}^{(2)})$ である。 $Var(\hat{y}_{x=\bar{x}_{..}}^{(i)})$ の場合であれば，合成分散の一般式から

$$\left. \begin{aligned} Var(\hat{y}_{x=\bar{x}_{..}}^{(i)}) &= Var(\hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 \bar{x}_{..}) \\ &= Var(\hat{\beta}_0) + a^2 Var(\hat{\beta}_1) + \bar{x}_{..}^2 Var(\hat{\beta}_2) \\ &\quad + 2a Cov(\hat{\beta}_0, \hat{\beta}_1) + 2\bar{x}_{..} Cov(\hat{\beta}_0, \hat{\beta}_2) + 2a\bar{x}_{..} Cov(\hat{\beta}_1, \hat{\beta}_2) \end{aligned} \right\} \quad (9.29)$$

のように，パラメータ $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ に関する分散と共分散を必要とする。これまでも示してきたが，デザイン行列 X と分散分析表の誤差分散の推定値 $\hat{\sigma}^2$ から，パラメータの共分散行列 $\Sigma(\hat{\beta})$ は，

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2 \quad (9.30)$$

として求められることを、繰返し示してきた。そして各パラメータに掛かる係数を

$$\left. \begin{aligned} \mathbf{l} &= (l_0, l_1, l_2) \\ &= (1, a, \bar{x}_..) = (1, a, 154.70) \end{aligned} \right\} \quad (9.31)$$

とベクトルとしたときに、線形和 $L = \mathbf{l}\hat{\boldsymbol{\beta}}$ の分散は、 \mathbf{l} に関する 2 次形式

$$\left. \begin{aligned} \text{Var}(L) &= \text{Var}(\mathbf{l}\hat{\boldsymbol{\beta}}) \\ &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 a + \hat{\beta}_2 \bar{x}_..) \\ &= \mathbf{l}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\mathbf{l}^T \end{aligned} \right\} \quad (9.32)$$

となり、この式 (9.32) を展開すると式 (9.29) になる。

表 9.9 に表 9.8 のデザイン行列 \mathbf{X} を用いて、パラメータの共変量行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の計算過程を示す。

表 9.9 パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の計算手順

デザイン行列 \mathbf{X}					
x_0	a	X	$\mathbf{X}^T\mathbf{X}$		
1	0	178	20	10	3094
1	0	155	10	10	1512
1	0	179	3094	1512	481348
1	0	152	=Mmult (Transpose (\mathbf{X} の範囲), \mathbf{X} の範囲)		
1	0	147			
1	0	164	$(\mathbf{X}^T\mathbf{X})^{-1}$		
1	0	154	10.2687	-0.5499	-0.0643
1	0	161	-0.5499	0.2199	0.0028
1	0	148	-0.0643	0.0028	0.0004
1	0	144	=Minverse ($\mathbf{X}^T\mathbf{X}$ の範囲)		
1	1	146			
1	1	137	$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^{2\wedge}$		
1	1	158	$\hat{\beta}_0$	585.7038	-31.3675
1	1	160	$\hat{\beta}_1$	-31.3675	12.5431
1	1	139	$\hat{\beta}_2$	-3.6662	0.1622
1	1	163	x_0	a	X
1	1	142	= $[(\mathbf{X}^T\mathbf{X})^{-1}$ の範囲] $\times \sigma^{2\wedge}$		
1	1	170			
1	1	147	$\sigma^{2\wedge}$		
1	1	150	57.0377	分散分析表の残差の分散	

表 9.10 に A_1 群と A_2 群の後値 Y と前後差 D に対するに最小 2 乗平均 95% 信頼区間の計算結果を示す。後値 Y に対する A_1 群の最小 2 乗平均の分散は、

$$\text{Var}(L^{(1)}) = \mathbf{l}^{(1)}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})(\mathbf{l}^{(1)})^T \quad (9.33)$$

	$\mathbf{l}^{(1)}$		$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$		$(\mathbf{l}^{(1)})^T$				
=	1	0	154.70	585.7038	-31.3675	-3.6662	1	=	5.9877
				-31.3675	12.5431	0.1622	0		
				-3.6662	0.1622	0.0232	154.70		

として計算される．分散の 5.9877 の平方根 $SE=2.4470$ を用いて，95%信頼区間（147.20, 157.53）を求めることができる． A_2 群と A_1 群の差の分散の平方根 $SE=3.5416$ は，表 9.8 に示した回帰分析のダミー変数 a の推定値 $\hat{\beta}_1$ の $SE=3.5416$ に一致することが確かめられる．前後差 D についても同様に求められている．

表 9.10 A_1 群と A_2 群に対する最小 2 乗平均と 95%信頼区間

$L^{(k)}$	A	l_0	l_1	l_2	β^{\wedge}	線形和	分散	標準誤差	幅	95%信頼区間	
		x_0	a	X_{ii}		Y^{\wedge}	$Var(Y^{\wedge})$	SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
$L^{(1)}$	A_1	1	0	154.70	26.9919	152.36	5.9877	2.4470	5.1627	147.20	157.53
$L^{(2)}$	A_2	1	1	154.70	6.9729	159.34	5.9877	2.4470	5.1627	154.17	164.50
$L^{(3)}$	A_1-A_1	0	0	0.00	0.8104	0.00	0	0	0	0	0
$L^{(4)}$	A_2-A_1	0	1	0.00		6.97	12.5431	3.5416	7.4722	-0.50	14.45
$t_{0.05}(20-3)= 2.1098$											
$L^{(k)}$	A	μ	a	X_{ii}	β^{\wedge}	D^{\wedge}	$Var(D^{\wedge})$	SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
		$L^{(5)}$	A_1	1		0	154.70	26.9919	-2.34	5.9877	2.4470
$L^{(6)}$	A_2	1	1	154.70	6.9729	4.64	5.9877	2.4470	5.1627	-0.53	9.80
$L^{(7)}$	A_1-A_1	0	0	0.00	-0.1896	0.00	0	0	0	0	0
$L^{(8)}$	A_2-A_1	0	1	0.00		6.97	12.5431	3.5416	7.4722	-0.50	14.45

表 9.10 で推定された最小 2 乗平均と 95%信頼区間を，Excel の折れ線グラフを用いて，図 9.10，図 9.11 に示す．この図からも明らかなように，前値 X を共変量とした場合に，後値 Y の最小 2 乗平均の群間差は，前後差 D の最小 2 乗平均の群間差に完全に一致している．

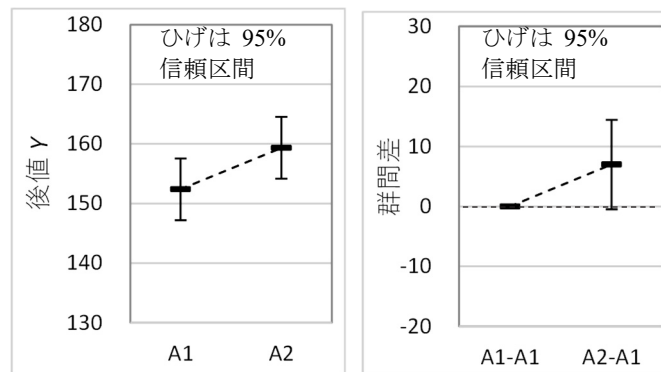


図 9.10 後値 Y および群間差についての最小 2 乗平均の図示

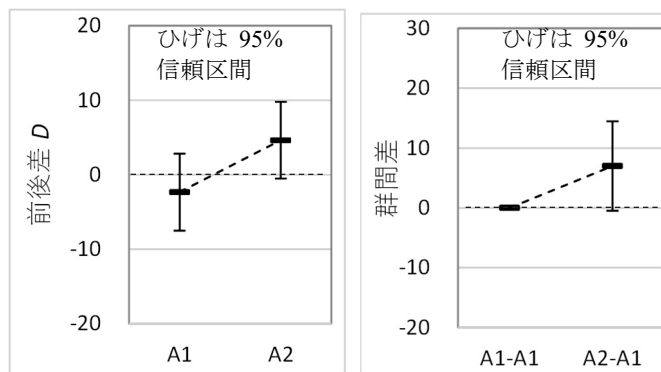


図 9.11 前後差 D および群間差についての最小 2 乗平均の図示

2) 後値 Y にのみ偶然に有意な差が出る場合

前値 X を共変量とした後値 Y に対する解析を行う際に、因子 A との交互作用がないことが前提条件であるとの“長年の慣習”の壁が立ちはだかる。困ったことに、交互作用があった場合に、どのような解析を行ったら良いのかとの手順は、残念ながら定式化されていない。その原因は、第 6 章で示した伝統的な共分散分析の解析手順に起因する。重層的な単回帰分析を適用し、交互作用のないことを分散分析表で確認し、交互作用を含まない主効果モデルの分散分析表により、因子 A に対する F 値により有意差検定がようやく行える。この解析方法では、交互作用がないことが絶対条件であり、交互作用があった場合には、「分散分析表での因子 A に対する有意差検定が正しく行えない」との理由により諦めざるをえなかったのである。

前値 X を共変量として、因子 A との交互作用がある場合に、前値 X の大きさに関連して回帰直線間の差の変化に対する解析方法をこれまでも示してきた。表 9.6 に示した「後値 Y にのみ有意な差が出る場合」を例にし、前値 X を共変量とした場合の前後差 D に対する回帰分析を行ない、薬剤 A について前値 X の総平均 $\bar{x}_.$ における A_1 、 A_2 および $(A_2 - A_1)$ について最小 2 乗平均と 95%信頼区間を求めて比較検討を行なう。

表 9.11 左に、共変量 X と薬剤 A との交互作用を含む (20 行×4 列) のデザイン行列 X を示す。表 9.11 左下段には、後値 Y に A_1 薬と A_2 薬間に有意な差 ($p=0.0165$) があることが示されている。表 9.11 右上段には、前後差 D について交互作用 $a \times X$ を含まない主効果モデルに対する回帰分析の分散分析表およびパラメータの推定値が示されている。(0, 1) 型ダミー変数 a のパラメータの推定値は、 $\hat{\beta}_1 = 6.4025$ ($p=0.0174$) と有意な差となり、前後差 D での 2 群間比較での差 $(\bar{D}_2 - \bar{D}_1) = 5.60$ ($p=0.1229$) では、有意な差でなかったが、前値 X を共変量とした場合に、有意な差となる。後値 Y について前値 X を共変量とする結果は省略するが、まったく同じ結果となる。

交互作用を含む共分散分析

さて、表 9.11 右中段に示すのは、前後差 D について交互作用 $a \times X$ を含む交互作用モデルに対する回帰分析の分散分析表およびパラメータの推定値が示されている。(0, 1) 型ダミー変数 a のパラメータの推定値は、 $\hat{\beta}'_1 = -32.6801$ ($p=0.5726$) と理解に苦しむ結果が示されている。自由度が 1 なので、ここには示されていないが分散分析表にまとめ直した場合も p 値は、同じ結果となるので、この結果をどのように受け止めたら良いのだろうか。幸い、交互作用は、($p=0.5726$) と有意な差ではないので、悩むことなく交互作用なしの主効果モデルの結果が使えばよい。もしも、交互作用が有意であったとしたら、どのように対処したら

表 9.11 前後差 D に対する交互作用 $A \times X$ 無しの場合、有りの場合の比較

A	No	デザイン行列 X				後値	前後差	分散分析表 反応変数:前後差 D 交互作用なし					
i	j	x_0	a	X	$a \times X$	Y	D		自由度	変動	分散	分散比	
1	1	1	0	157	0	151	-6	回帰	2	735.15	367.573	8.2485	
1	2	1	0	162	0	149	-13	残差	17	499.05	29.3561		
1	3	1	0	153	0	149	-4	合計	19	1234.20			
1	4	1	0	151	0	162	11		係数	標準誤差	t	p -値	
1	5	1	0	169	0	160	-9	β^{\wedge}_0	切片	111.6646	25.9128	4.3092	0.0005
1	6	1	0	149	0	148	-1	β^{\wedge}_1	a	6.4025	2.4298	2.6350	0.0174 *
1	7	1	0	158	0	154	-4	β^{\wedge}_2	X	-0.7296	0.1644	-4.4386	0.0004
1	8	1	0	155	0	152	-3	パラメータの共分散行列					
1	9	1	0	155	0	165	10	$\Sigma(\beta^{\wedge})=(X^T X)^{-1} \sigma^2 \wedge$					
1	10	1	0	164	0	152	-12	β^{\wedge}_0	671.4735	1.7395	-4.2501		
2	1	1	1	170	170	161	-9	β^{\wedge}_1	1.7395	5.9039	-0.0297		
2	2	1	1	150	150	155	5	β^{\wedge}_2	-4.2501	-0.0297	0.0270		
2	3	1	1	148	148	164	16						
2	4	1	1	151	151	160	9						
2	5	1	1	170	170	167	-3	分散分析表 反応変数:前後差 D 交互作用あり					
2	6	1	1	156	156	164	8		自由度	変動	分散	分散比	
2	7	1	1	158	158	155	-3	回帰	3	749.55	249.851	8.2485	
2	8	1	1	149	149	153	4	残差	16	484.65	30.2905		
2	9	1	1	170	170	169	-1	合計	19	1234.20			
2	10	1	1	162	162	161	-1		係数	標準誤差	t	p -値	
1		平均 μ^{\wedge}_1	157.3			154.2	-3.1	β^{\wedge}_0'	切片	138.3826	46.8388	2.9544	0.0093
		$SD \sigma^{\wedge}_1$	6.2			6.0	8.1	β^{\wedge}_1'	a	-32.6801	56.7262	-0.5761	0.5726 NS
2		平均 μ^{\wedge}_2	158.4			160.9	2.5	β^{\wedge}_2'	X	-0.8994	0.2976	-3.0227	0.0081
		$SD \sigma^{\wedge}_2$	9.1			5.3	7.3	β^{\wedge}_3'	$a \times X$	0.2479	0.3595	0.6896	0.5003
平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$		1.10				6.70	5.60	パラメータの共分散行列					
差の SE		3.47				2.54	3.46	$\Sigma(\beta^{\wedge})=(X^T X)^{-1} \sigma^2 \wedge$					
		t 値	0.3166			2.6427	1.6185	β^{\wedge}_0'	2126.19	-2126.19	-13.498	13.498	
		p 値	0.7552			0.0165	0.1229	β^{\wedge}_1'	-2126.19	3118.60	13.498	-19.745	
			NS			*	NS	β^{\wedge}_2'	-13.498	13.498	0.086	-0.086	
		総平均	157.85					β^{\wedge}_3'	13.498	-19.745	-0.086	0.125	

よいのだろうか。交互作用を含めたモデルで、薬剤 A の群間比較を適切に行う方法は、あるのだろうか。

交互作用がある場合には、しかるべきグラフを作成し、結果をよく見て考察をすることが必要と言われている。交互作用がない場合に、共変量である前値 X の総平均 $\bar{x}_{..} = 157.85$ に対する 2 本の直線の推定値（最小 2 乗平均）、2 本の直線間の差の推定値（最小 2 乗平均）により、薬剤 A の群間比較を行ってきた。交互作用がある場合にも同様の比較方法は、グラフを見て適切に考察することに対応する。デザイン行列 X を用いた回帰分析の場合には、交互作用のある非平行な直線であっても前値 X の総平均 $\bar{x}_{..} = 157.85$ に対する 2 本の直線の推定値を得ることは、これまでも示したきたように、パラメータに関する線形和 $L^{(k)} = I^{(k)} \hat{\beta}$ によって可能である。

図 9.12 左に示すのは、表 9.11 で推定された交互作用なしのパラメータ $\hat{\beta}$ により推定された平行な直線であり、図 9.12 右は、交互作用ありのパラメータ $\hat{\beta}'$ により推定された非平行な直線である。散布図の前値 X の中央の縦線は、 A_1 薬および A_2 薬の前値 X の総平均 $\bar{x}_{..} = 157.85$ を示している。平行な直線間の差は、表 9.11 右上の変数 a のパラメータ推定値 $\hat{\beta}_1 = 6.4025$ ($p = 0.0174$) から有意な差であることが分かる。

図 9.12 右の交互作用あり場合の非平行な直線は、総平均 $\bar{x}_{..} = 157.85$ を中心にそれぞれの直線が微妙に回転しているように観察される。したがって、交互作用ありとした非平行な 2 本の直線の場合でも総平均 $\bar{x}_{..} = 157.85$ における直線間の差は、ほとんど同じと判断される。ただし、変数 a のパラメータ推定値 $\hat{\beta}'_1 = -32.6801$ ($p = 0.5726$) は、切片における群間の差であり、新たに推定する必要がある。

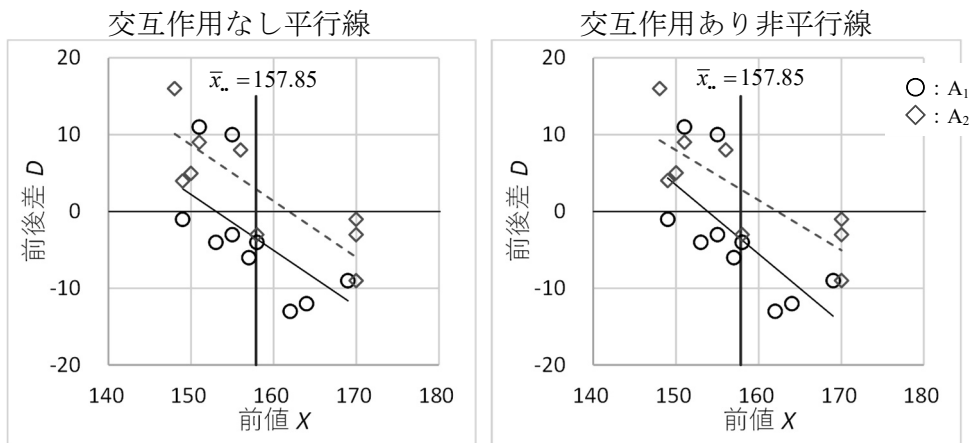


図 9.12 前値 X を共変量とした前後差 D の A_1 薬および A_2 薬の回帰直線

図 9.12 の散布図に重ね書きした直線は、表 9.12 に示すように推定されたパラメータ $\hat{\beta}$ および $\hat{\beta}'$ を用い、 A_1 薬および A_2 薬それぞれの前値 X の最小値と最大値に対する推定値を求め、Excel の散布図に対する「データの選択」の「追加」によって直線の始点と終点を結ぶ直線を引いて調整した結果である。総平均も同様に始点と終点を $(157.85, 15)$ 、 $(157.85, -20)$ を Excel シート上に設定し、重ね書きしている。

表 9.12 2本の直線の最小値と最大値の推定

A	l_0	l_1	l_2		$l\beta^{\wedge}$
i	μ	a	X		D^{\wedge}
1	1	0	149		111.66 = 2.96
1	1	0	169		6.4025 = -11.64
2	1	1	148		-0.7296 = 10.09
2	1	1	170		-5.96
交互作用なし平行線					

A	l_0	l_1	l_2	l_3		$l\beta'^{\wedge}$
i	μ	a	X	$a \times X$		D'^{\wedge}
1	1	0	149	0		138.38 = 4.37
1	1	0	169	0		-32.68 = -13.62
2	1	1	148	148		-0.8994 = 9.28
2	1	1	170	170		0.2479 = -5.06
交互作用あり非平行線						

交互作用がある場合の最小2乗平均

交互作用を含まない平行な直線間の差は、表 9.11 右上段の変数 a のパラメータ $\hat{\beta}_1 = 6.4025$ ($SE = 2.4298$) である。交互作用を含む場合は、表 9.13 に示すように、 A_1 薬、 A_2 薬、差 ($A_2 - A_1$) に対し、総平均 $\bar{x}_{..} = 157.85$ における非平行な直線上の推定値を求め、その差を平行な場合と比較する。 A_1 薬の線形和 $L^{(3)} = l^{(3)}\hat{\beta}' = 6.453$ 、 A_2 薬の線形和 $L^{(4)} = l^{(4)}\hat{\beta}' = 2.858$ 、その差は、 $L^{(6)} = l^{(6)}\hat{\beta}' = 6.453$ であり、平行の場合とほぼ同等である。また、差の分散は、 $Var(L^{(6)}) = l^{(6)}\Sigma(\hat{\beta}')(l^{(6)})^T = 5.909$ と推定され、その平方根は、 $SE(L^{(6)}) = 2.431$ と平行の場合とほぼ同等である。

表 9.13 交互作用を含む場合の前値 X の総平均に対する最小2乗平均と 95%信頼区間

$L^{(k)}$	A	l_0	l_1	l_2	l_3	推定値 $\hat{\beta}'$	線形和 L^\wedge	分散 $Var(L^\wedge)$	標準誤差 SE	幅 $t_{0.05} \times SE$	95%信頼区間	
		μ	a	X	$a \times X$						$L_{95\%}$	$U_{95\%}$
$L^{(3)}$	A_1	1	0	157.85	0.00	138.38	-3.595	2.962	1.721	3.648	-7.243	0.053
$L^{(4)}$	A_2	1	1	157.85	157.85	-32.68	2.858	2.948	1.717	3.640	-0.781	6.498
$L^{(5)}$	$A_1 - A_1$	0	0	0.00	0.00	-0.8994	0.000	0.000	0.000	0.000	0.000	0.000
$L^{(6)}$	$A_2 - A_1$	0	1	0.00	157.85	0.24791	6.453	5.909	2.431	5.153	1.300	11.606
$t_{0.05}(16) =$										2.1199		

図 9.13 左は、図 9.12 右の再掲で、図 9.13 中は、 A_1 薬と A_2 薬の推定値（最小2乗平均）、図 9.13 右が、薬剤間の差と 95%信頼区間が示されている。このような対応を付けることにより、結果の解釈が容易になる。このように、交互作用を含めた場合でも、傾きが異なる直線上の推定値に対し、前値 X の総平均 $\bar{x}_{..} = 157.85$ における群間比較が行え、交互作用がないと仮定した平行な直線の場合とほぼ同様な群間比較となる。

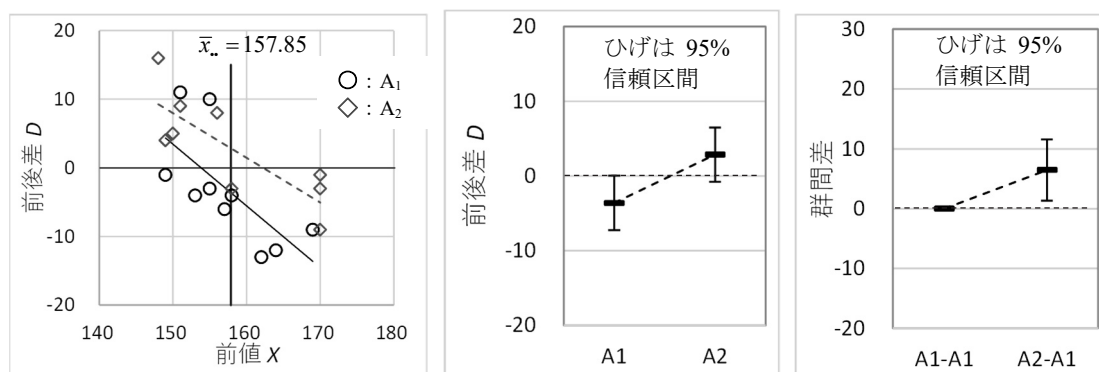


図 9.13 群間差についての最小2乗平均と 95%信頼区間

薬剤 A と前値 X の間に交互作用が有るか否かの判断は、 p 値の大小による統計的な判断にゆだねられている。交互作用を気にする場合は、 $p < 0.15$ などの基準、交互作用を否定したい場合は $p < 0.05$ のような 2 重基準となりやすい。ここに示したように、交互作用の p 値に関わらず薬剤 A の水準間の最小2乗平均による p 値が、2 重基準からの脱却となる。

3) 後値 Y および前後差 D 共に有意な差が出る場合

表 9.7 に示した後値 Y および前後差 D 共に有意な差が出る場合に対しては、結果のグラフ表示にすぐれた JMP の「モデルのあてはめ」を用いた結果を示す。Excel を用いた回帰分析では、因子 A の A_1 群に対し $a=0$ 、 A_2 群に対し $a=1$ とする $(0, 1)$ 型ダミー変数を用いたが、JMP で因子 A を文字型変数とすると内部で $(1, -1)$ 対比型ダミー変数が自動生成される。そのために、パラメータの推定値が、 $(0, 1)$ 型ダミー変数を用いた場合と異なるので注意が必要である。表 9.14 に解析用の JMP データセットを示す。

表 9.14 後値 Y および前後差 D 共に有意な差が出る場合の JMP データファイル (表 9.7)

	観測値	A	X	Y	D
○	1	A1	145	153	8
○	2	A1	151	157	6
⋮					
◇	19	A2	155	154	-1
◇	20	A2	154	148	-6

$\bar{x}_{..} = 152.25$

JMP による主効果モデル

前値 X を共変量とした前後差 D について モデル $D=A+X$ についての結果を表 9.15 に示す。分散分析表に引き続きパラメータの推定値が示されている。 A_1 薬に対する $A[A1]$ の推定値が 4.0387 となり、 $p=0.0045$ と有意な差であることが分かる。図 9.14 に前値 X を X 軸に前後差 D を Y 軸とする散布図上に、平行な 2 本の直線

$$A_1 \text{ 群 : } \begin{cases} \hat{D} = \hat{\beta}_0 + \hat{\beta}_1 \times A[A1] + \hat{\beta}_2 X \\ = 98.4059 + 4.0387 \times 1 - 0.6450 X \\ = 102.4446 - 0.6450 X \end{cases} \quad (9.34)$$

表 9.15 前値 X を共変量とした前後差 D についての解結果

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	2	865.5992	432.7996	14.2148
誤差	17	517.6008	30.4471	p値(Prob>F)
全体(修正済み)	19	1383.200		0.0002*

パラメータ推定値				
項	推定値	標準誤差	t値	p値
切片	98.4059	23.9615	4.1068	0.0007*
A[A1]	4.0387	1.2345	3.2717	0.0045*
X	-0.6450	0.1572	-4.1039	0.0007*

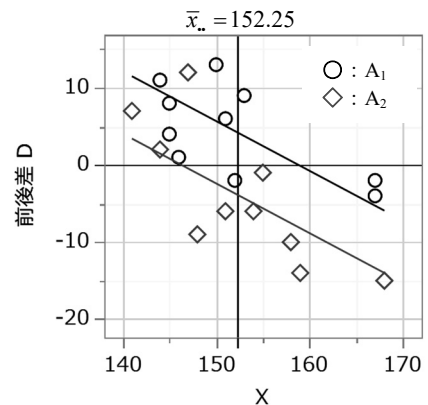


図 9.14 散布図上での平行な直線

$$A_2 \text{ 群 : } \begin{cases} \hat{D} = \hat{\beta}_0 + \hat{\beta}_1 \times A[A2] + \hat{\beta}_2 X \\ = 98.4059 + 4.0387 \times (-1) - 0.6450X \\ = 94.3671 - 0.6450X \end{cases} \quad (9.35)$$

JMPの「パラメータ推定値」では、因子Aの第2水準A[A2]が示されていない。
 (1, -1) 対比型ダミー変数が自動生成されているので、A[A2]=-1となる。

が示されている。総平均 $\bar{x}_{..} = 152.25$ における平行な直線間の差は、(1, -1) 対比型ダミー変数なので表 9.15 の変数 A[A1] のパラメータ推定値 4.0387 の2倍の 8.0775 となり、標準誤差も2倍の 4.2690 となるので、t値とp値は変わらないので有意な差と判定される。

JMPによる交互作用モデル

主効果モデルに引き続き、交互作用 $A \times X$ を含む交互作用モデル $D = A + X + (A \times X)$ についての結果を表 9.16 に示す。交互作用の変数が、 $A[A1] \times (X - 152.25)$ となっている。JMPでは、量的変数を交互作用に含めるとデフォルトで“中心化”が設定され、交互作用に用いる前値 X は、総平均 $\bar{x}_{..} = 152.25$ との差（中心化と言う）が内部のデザイン行列に自動的にセットされる。中心化を行なうと、回帰式が複雑になるので避けたい気持ちがいつもあったが、ここでは、中心化を行なうことが、結果の解釈に有益であることに気が付かされた。

表 9.16 交互作用を含む前後差 D についての解結果

分散分析				
要因	自由度	平方和	平均平方	F値
モデル	3	921.3759	307.1253	10.6404
誤差	16	461.8241	28.8640	p値(Prob>F)
全体(修正済み)	19	1383.2000		0.0004*

パラメータ推定値				
項	推定値	標準誤差	t値	p値
切片	100.4469	23.3764	4.2969	0.0006*
A[A1]	4.0355	1.2019	3.3575	0.0040*
X	-0.6581	0.1533	-4.2922	0.0006*
A[A1]*(X-152.25)	0.2131	0.1533	1.3901	0.1835

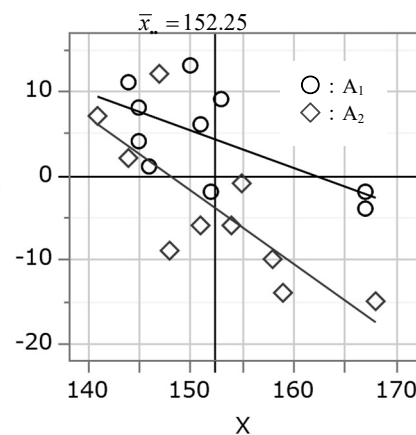


図 9.15 散布図上での非平行な直線

図 9.15 に示す散布図は、JMPの「モデルのあてはめ」によりデフォルトで出力されたものに格子線および前値 X の総平均 152.15 の参照線を書き入れたものである。重ね書きされた非平行な直線の形状は、切片と傾きによって理解されるのであるが、表 9.16 の推定されたパラメータからでは見当が付かない。そこで、得られたパラメータを整理し、

$$A_1 \text{ 群 : } \begin{cases} \hat{D}^{(A_1)} = \hat{\beta}_0 + \hat{\beta}_1 A[A1] + \hat{\beta}_2 x + \hat{\beta}_3 \{A[A1] \times (X - 152.25)\} \\ = 100.4469 + 4.0355 \times 1 - 0.6581x + 0.2131 \times \{1 \times (X - 152.25)\} \\ = 72.03 - 0.4450X \end{cases} \quad (9.36)$$

$$A_2 \text{ 群 : } \begin{cases} \hat{D}^{(A_2)} = \hat{\beta}_0 + \hat{\beta}_1(-A[A1]) + \hat{\beta}_2 x + \hat{\beta}_3 \{-A[A1] \times (X - 152.25)\} \\ = 100.4469 + 4.0355 \times (-1) - 0.6581x + 0.2131 \times \{(-1) \times (X - 152.25)\} \\ = 128.86 - 0.8712X \end{cases} \quad (9.37)$$

が得られる。また、2本の直線が重なる血圧は、式(9.36)と式(9.37)の差を x について解くと $x_{\text{交点}} = 133.3158 \text{ mmHg}$ が得られる。

非平行な2本の回帰直線のパラメータを推定したいのであれば、 A_1 薬と A_2 薬に対して単回帰をすれば、簡単に推定することができる。ただし、推定したいのは、前値 X の総平均 $\bar{x}_{..} = 152.25$ における非平行な2本の直線間の差とその95%信頼区間を共通の誤差分散を用いて推定したいので役に立たない。

中心化された場合の最小2乗平均

表9.17に因子Aについての最小2乗平均と95%信頼区間を示す。最小2乗平均は、図9.15の前値 X の総平均 $\bar{x}_{..} = 152.25$ に対する各直線の推定値なので、式(9.36)および式(9.37)に $X = 152.25$ を代入すると交互作用の X が中心化されているので0となり、差を取ると

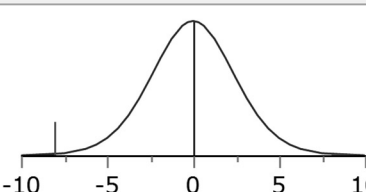
$$\begin{aligned} A_2 \text{ 薬 : } \hat{D}_{X=152.25}^{(A_2)} &= \hat{\beta}_0 + \hat{\beta}_1(-A[A1]) + \hat{\beta}_2 \times 152.25 \\ A_1 \text{ 薬 : } \hat{D}_{X=152.25}^{(A_1)} &= \hat{\beta}_0 + \hat{\beta}_1(+A[A1]) + \hat{\beta}_2 \times 152.25 \\ \hline A_2 - A_1 : \hat{D}_{X=152.25}^{(A_2-A_1)} &= -2\hat{\beta}_1 A[A1] \\ &= -2 \times 4.0355 = -8.0710 \end{aligned}$$

のように $A[A1]$ の推定値 $\hat{\beta}_1 = 4.0355$ の(-2)倍の-8.0710となる。したがって、交互作用のある非平行な直線の総平均 $x_0 = 152.25$ における直線間の差を推定されたパラメータから直接推定することができる。このことが、前値 X を含む交互作用に対し、中心化する最大のメリットである。

表 9.17 交互作用を含む前後差 D についての最小2乗平均

A					
最小2乗平均表					
水準	最小2乗平均	標準誤差	下側95%	上側95%	平均
A1	4.2888	1.6998	0.6854	7.8921	4.4000
A2	-3.7822	1.6999	-7.3857	-0.1787	-4.0000

A2をA1と比較			
差	-8.0710	t値	-3.3575
差の標準誤差	2.4039	自由度	16.0000
差の上側信頼限界	-2.9749	p値(Prob> t)	0.0040
差の下側信頼限界	-13.1670	p値(Prob>t)	0.9980
信頼率	0.9500	p値(Prob<t)	0.0020



交互作用に中心化 ($X - 152.25$) した変数を用いることは、総平均 $\bar{x}_{..} = 152.25$ において A_1 薬と A_2 薬の共通の傾き $\hat{\beta}_2 = -0.6581$ に、推定された交互作用の推定値 $\hat{\beta}_3 = 0.2131$ をプラス・マイナスさせることと解される。このような解釈は、抽象的になりがちな交互作用の解釈に光明を与える。すなわち、交互作用を含まない平行な直線間の差を「2 群間の平均的な効果」としたときに、交互作用を含む場合に総平均 $\bar{x}_{..} = 152.25$ における非平行な直線間の差が「2 群間の平均的な効果」を表し、中心化した交互作用の推定値をプラス・マイナスすることにより、前値 X の大小に関連する効果の違いを浮き彫りにする。図 9.15 に即して言えば、前値の血圧 X が小さい場合には、効果の差が縮小し、高い場合には、効果が拡大するといったごく常識的な結果と説明できる。

JMP の解析結果を Excel でグラフ化

交互作用を考慮した図 9.15 の散布図上の非平行な直線に対し、それらの交点 133.3158 mmHg、血圧の前値が 145 mmHg と低い場合と 165 mmHg と高い場合の非平行な直線間の差と 95%信頼区間を算出して、折れ線グラフに示したい。表 9.18 に示すのは、JMP の「最小 2 乗平均の対比」の機能を用いて、前値が 145 mmHg と 165 mmHg の場合の非平行な直線間の差と 95%信頼区間を算出し、前値が 145 mmHg の場合は 2 群間の差が -4.9805 で、165 mmHg の場合は -13.5058 との結果が得られている。

表 9.18 「最小 2 乗平均の対比」によって求めた非平行な直線間の差と 95%信頼区間

検定の詳細		検定の詳細		検定の詳細	
A1	-1.0000	A1	-1.0000	A1	-1.0000
A2	1.0000	A2	1.0000	A2	1.0000
推定値	0.0000	推定値	-4.9805	推定値	-13.5058
標準誤差	6.2883	標準誤差	3.2775	標準誤差	4.5856
t値	0.0000	t値	-1.5196	t値	-2.9453
p値(Prob> t)	1.0000	p値(Prob> t)	0.1481	p値(Prob> t)	0.0095
平方和	0.0000	平方和	66.6539	平方和	250.3860
下側95%	-13.3306	下側95%	-11.9285	下側95%	-23.2268
上側95%	13.3307	上側95%	1.9674	上側95%	-3.7848
X	133.3158	X	145	X	165

これらの統計量を Excel に取り込み、結果を整理し折れ線グラフを作成する。表 9.19 に示すのが、JMP の異なる形式の出力結果を Excel に取り込み整理した結果である。表 9.19 の右端の $t_{0.05} \times SE$ は、Excel の折れ線グラフで 95%信頼区間の幅を付けるために必要なもので、追加したものである。

表 9.19 JMP の出力結果を統一した形式でまとめた結果

前値 X	水準	最小2乗		t 値	p 値	下側95%	上側95%	$t_{0.05} \times SE$ 幅
		平均	標準誤差					
153.25	A_1	4.2888	1.6998			0.6854	7.8921	3.6034
153.25	A_2	-3.7822	1.6999			-7.3857	-0.1787	3.6036
133.3158	A_2-A_1	0	6.2883	0.0000	1.0000		-13.3306	13.3307
145	A_2-A_1	-4.9810	3.2775	-1.5200	0.1481	-11.9300	1.9674	6.9480
152.25	A_2-A_1	-8.0710	2.4039	-3.3575	0.0020	-13.1670	-2.9749	5.0960
165	A_2-A_1	-13.5058	4.5856	-2.9453	0.0095	-23.2268	-3.7848	9.7210
			$t_{0.05}(16)=$	2.1199				

図 9.16 左は、図 9.15 に示した JMP の出力結果の再掲であり、図 9.16 中は、 $\bar{x}_{..} = 152.25$ における A_1 群と A_2 群の非平行な直線の推定値であり、図 9.16 右は、前値 X の (145, 152.25, 165) における非平行な直線の間差の推定値である。この結果から、前値 X が小さい場合には群間差が小さく、大きくなるにつれ有意な群間差となることが、きちっと読み取れるようになる。

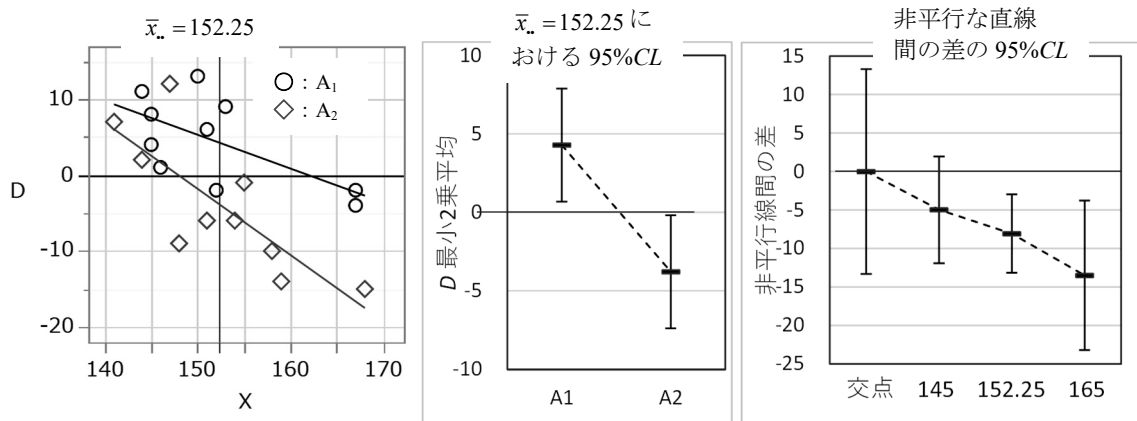


図 9.16 交互作用を非平行な直線間の最小 2 乗平均と 95%信頼区間

交互作用を含める解析の妥当性

交互作用に与える前値 X について前値 X に対して総平均 $\bar{x}_{..} = 152.25$ からの差にする中心化が、結果の解釈を容易にし、かつ実践的な解釈が行えるようになることは、画期的な解析方法である。「平均への回帰」の影を消すために、前値 X を共変量にした解析をすべきであるとの提言に、常に交互作用に関する予備検定の必要性が問われ続けている。その結果とし、交互作用が有意な差となった場合に、共分散分析が適用できない、どうしたら良いのか、これまで葛藤の連続であった。ここに示した様に、前値 X との交互作用を常に含め、前値 X の総平均における 2 本の回帰直線上の最小 2 乗平均による解析法が、新たな解析法として結果の解釈に役立つことを示した。

9.5. 前値 X と後値 Y をめぐる各種の群間比較の第1種の過誤 (α エラー)

前値 X と後値 Y の前後差 D に対する群間比較の統計的な諸問題に対処するために、前値 X を共変量とする回帰分析を適用し、典型的な事例に対する統計的な性質を示してきた。後値 Y 、および、前後差 D について前値 X を共変量とした群間比較について統計的性能を明らかにするためには、幾つかの条件設定を行った上で、統計的なシミュレーションにより、第1種の過誤 (α エラー) と次節で取り上げる検出力 ($1-\beta$ エラー) が、前値 X と後値 Y の相関に関連し、どのように変化するかを把握する必要がある。

第1種の過誤 (α エラー) は、因子 A の A_1 群をプラセボ、 A_2 群を実薬としたときに、両群の後値 Y が同程度の減少効果であった場合に、有意な差となる両側確率 $p < 0.05$ をカウントして α エラーを算出する。各種の条件下で $\alpha = 0.05$ が保たれているか、統計的なシミュレーションにより、有意な差となる回数の割合を算出して評価する。

表 9.20 に示すのは、Excel シートの表頭部分で、 A_1 群、 A_2 群共に共通の2群の平均 155.0 に対し、後値 Y の期待値がそれぞれ -10.0 減少して $Y_{A_1} = 145.0$ 、 $Y_{A_2} = 145.0$ となるように設定してある。実際に正規乱数を加味したデータセットを何組も発生させた場合に対応のない t 検定の p 値が、 $p < 0.05$ となる回数の割合を第1種の過誤 (α エラー) とする。時点間の相関を変化させたときに A_1 群、 A_2 群の前値 X の平均値が、(1: $\bar{X}_{A_1} > \bar{X}_{A_2}$)、(2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$)、(3: $\bar{X}_{A_1} < \bar{X}_{A_2}$) であった場合に、後値 Y 、前後差 D 、前値 X を共変量とする共分散分析における群間差の p 値がどのように変化するかを概観する。

表 9.20 シミュレーションで設定する各群の条件設定

2群の平均=		155	SD	相関の	A_1 群効果	-10.0					
個体間分散=		50	7.07	期待値	A_2 群効果	-10.0					
個体内分散=		50	7.07	0.5							
個体間乱数			個体内乱数		デザイン行列			前値X共変量			
A	No	正規	個体間	正規	正規	前値	後値	差	推定値	推定値	
i	j	乱数Z	XY	Z	Z	μa	X	Y	D	Y^{\wedge}	D^{\wedge}

安定した α エラーを得るためには、最低でも 1,000 回の繰り返し、通常は 10,000 回程度の検定を行う必要がある。このようなシミュレーションを Excel で実行することは極めて困難であり、SAS などのプログラミング機能がある統計ソフトを用いる必要がある。実際に SAS で実施したシミュレーション結果を取りまとめた高橋 (2012)、「平均への回帰を考慮した投与前値の調整」を引用する。

1): ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合の α エラー

表 9.21 にすように, Excel によるシミュレーションの一例は, 両群の減少効果が共に -10.0 であった場合に, 前値 X の平均が A_1 群で $\bar{X}_{A_1} = 158.5$, A_2 群で $\bar{X}_{A_2} = 153.2$ と A_1 群が大きく有意な差ではないが $p = 0.105$ となったスナップショットである. このような場合に前値 X を共変量にした場合に, 名目の $\alpha = 0.05$ が保たれているのだろうか. 後値 Y での 2 群間の比較で前値 X が A_1 群で高めの場合に

1. 後値 Y も高めで有意な差が出やすい
2. 平均への回帰現象で有意な差が出にくい

と相反することが想定されるが, 実際の α エラーは, 0.05 で保たれるのであろうか. その際に, 時点間の相関の大きさの変化との実際の α エラーの関連はどのようなものなのだろうか.

前後差 D の場合は,

1. 平均への回帰現象が強く出れば有意な差が出やすい
2. 平均への回帰現象が弱ければ有意な差が出にくい

のように, 後値 Y の場合と同様に相反することが想定されるが, 前値 X からの差なので, 名目の $\alpha = 0.05$ が保たれるのだろうか

表 9.21 には, 後値 Y , 前後差 D に対する対応のない t 検定の形式に合わせ, 前値 X を共変量とした場合の後値 Y , および, 前後差 D に対する共分散分析の結果が追加されている. 表 9.8, および, 表 9.11 では, Excel の回帰分析を用いて回帰パラメータおよび誤差分散を求めていた. しかし, シミュレーションに際して, 正規乱数のセットを逐次的に発生した場合に, Excel の回帰分析では, 改めて回帰分析を行ない, 結果をコピー&ペーストする煩わしさが伴い, 使い物にならない. そこで, 新たな正規乱数のセットに対し, 自動的に共分散分析の回帰パラメータおよび誤差分散の推定値が得られるように, Excel の行列関数を用いて

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\beta}' = (X^T X)^{-1} X^T D$$

回帰パラメータを推定し, 推定値 \hat{Y} および \hat{D} を

$$\hat{Y} = X\hat{\beta}$$

$$\hat{D} = X\hat{\beta}'$$

によって推定し, 誤差分散の推定値を

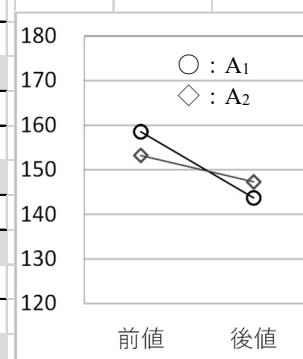
$$\hat{\sigma}_Y^2 = (Y - \hat{Y})^T (Y - \hat{Y}) / df, \quad df = 20 - 3$$

$$\hat{\sigma}_D^2 = (D - \hat{D})^T (D - \hat{D}) / df, \quad df = 20 - 3$$

により, 自動計算ができるようにしている. さらに各群の最小 2 乗平均の計算のために必要

表 9.21 前値 X で A_1 群の平均値がやや大きい場合のシミュレーションの一例

2群の平均=		155	SD	7.07	相関の期待値	0.5	A_1 群効果	-10.0	A_2 群効果	-10.0										
個体間 分散=		50	個体内 分散=		50	個体内乱数		デザイン行列 <th colspan="2">前値 X 共変量</th> <th colspan="2"></th> <th colspan="2"></th> <th colspan="2"></th>		前値 X 共変量										
A	No	正規乱数	個体間 Z_0	正規 Z_1	正規 Z_2	μ	a	前値 X	後値 Y	差 D	推定値 \hat{Y}	推定値 \hat{D}	$(X^T X)^{-1}$							
1	1	-0.53	151	0.54	0.49	1	0	155	144	-11	141.0	-14.0	29.106	-1.070	-0.183					
1	2	-0.92	148	1.53	0.45	1	0	159	141	-18	144.1	-14.9	-1.070	0.232	0.006					
1	3	0.05	155	0.91	-0.56	1	0	161	141	-20	145.6	-15.4	-0.183	0.006	0.001					
1	4	0.72	160	0.77	1.60	1	0	165	161	-4	148.6	-16.4								
1	5	-1.93	141	1.30	-0.51	1	0	150	127	-23	137.3	-12.7								
1	6	-0.85	149	-0.08	-1.02	1	0	148	132	-16	135.7	-12.3	23.484		23.484					
1	7	-0.37	152	0.85	1.74	1	0	158	154	-4	143.3	-14.7	7.620		7.620					
1	8	1.12	163	0.20	-0.65	1	0	164	148	-16	147.9	-16.1	0.758		-0.242					
1	9	0.39	158	-0.30	-0.84	1	0	156	142	-14	141.8	-14.2								
1	10	0.48	158	1.61	-0.09	1	0	169	147	-22	151.7	-17.3								
2	1	0.83	161	-0.06	0.44	1	1	161	154	-7	153.2	-7.8					$Y:\sigma^2$	$D:\sigma^2$		
2	2	-0.67	150	0.29	0.07	1	1	152	140	-12	146.4	-5.6	64.469		64.469					
2	3	0.65	160	-0.93	0.50	1	1	153	154	1	147.1	-5.9								
2	4	0.00	155	0.88	-0.84	1	1	161	139	-22	153.2	-7.8								
2	5	-0.72	150	1.08	1.64	1	1	158	152	-6	150.9	-7.1								
2	6	0.11	156	-0.32	-0.28	1	1	154	144	-10	147.9	-6.1								
2	7	0.49	158	-2.20	0.88	1	1	142	154	12	138.8	-3.2								
2	8	-0.93	148	-0.13	0.09	1	1	147	139	-8	142.6	-4.4								
2	9	-1.12	147	-0.58	-0.30	1	1	143	135	-8	139.6	-3.4								
2	10	0.51	159	0.25	1.78	1	1	161	162	1	153.2	-7.8								
						1	0	155.9	$\bar{x}_.$	最小2乗平均										
						平均 μ^1		158.5	143.7	-14.8	141.7	-14.16								
						SD σ^1		6.6	9.8	6.7										
						1	1	155.9	$\bar{x}_.$	最小2乗平均										
						平均 μ^2		153.2	147.3	-5.9	149.3	-6.54								
						SD σ^2		7.3	9.0	9.1										
						平均の差 $\mu^2 - \mu^1$		-5.30	3.60	8.90	7.62	7.62								
						差のSE		3.10	4.21	3.57	3.87	3.87								
						t 値		-1.708	0.855	2.494	1.968	1.968								
						両側 p 値		0.105	0.404	0.023	0.066	0.066								
								NS	*	+	+									



注) Excelシート上での計算式については、第9.6節の表9.26および表9.27を参照のこと

な、パラメータの共分散行列の計算

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}_Y^2$$

$$\Sigma(\hat{\beta}') = (X^T X)^{-1} \hat{\sigma}_D^2$$

も同一シート上で行っている。ただし、 $\Sigma(\hat{\beta}) = \Sigma(\hat{\beta}')$ なので、 $\Sigma(\hat{\beta}')$ は省略している。

表 9.21 に示すのは、個体内の分散が 50、個体間の分散が 50 に設定した場合で、前値 X と後値 Y の相関係数の期待値が $\rho = 0.50$ となる場合である。前値 X の平均値が、適当な (X_{A_1}

X_{A_2}) となるまで乱数を発生させ、別のシート上にスナップショットした結果である。スナップショットしたままでは、セルに Norm.Inv (Rand () ,0, 1) 関数が残っているために結果が揺らぐので、乱数の列をコピーし、スナップショットしたシート上に「値」のみをペーストし固定した結果である。対応のない t 検定の結果、後値 Y では、 $p = 0.404$ と NS であるが、前後差 D では、 $p = 0.023$ で有意な差となっている。両群の平均値のプロットを見れば、「平均への回帰」が起きた結果として、検定結果に大きな違いが出たと理解される。前値 X を共変量とし後値 Y および前後差 D に対する最小2乗平均の差に関する t 検定では $p = 0.066$ と微妙な結果となり、プロットされた平均値から、妥当な結果と思われる。

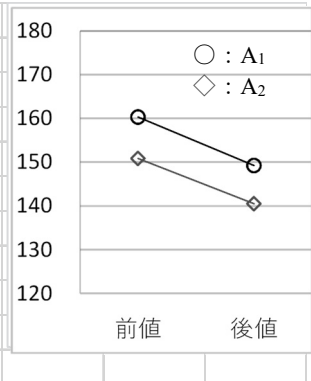
2): ($\bar{X}_{A_1} \gg \bar{X}_{A_2}$; 相関が大) 場合の α エラー

時点間の相関が 1.0 に近づくと後値 Y の場合には、平均への回帰現象は相対的に弱まり、前値 X の群平均での大小関係がそのまま後値 Y に引き継がれ、前値 X の群間差の p 値の影響がそのまま反映され、有意な差が出やすくなり、実質の α エラーが増大すると考えられる。前後差 D の場合は、前値 X の p 値の影響を受けにくくなり、名目の α エラーに接近することと思われる。

表 9.22 に示すは、前値 X が ($\bar{X}_{A_1} \gg \bar{X}_{A_2}$) と明らかに A_1 群が大きい場合で、前後差 D では、 $p = 0.640$ と有意ではなく、後値 Y では、 $p = 0.020$ と有意な差となっている。前値 X を

表 9.22 前値 X で A_1 群の平均値が大きく相関が 0.9 の場合のシミュレーションの一例

2群の平均=		155	SD	相関の	A_1 群効果	-10.0															
個体間 分散=		90	9.49	期待値	A_2 群効果	-10.0															
個体内 分散=		10	3.16	0.9																	
個体間		個体内乱数		デザイン行列		前値 X 共変量															
A	No	正規	個体間	正規	正規		前値	後値	差	推定値	推定値										
i	j	乱数 Z_0	XY	Z_1	Z_2	μ	a	X	Y	D	Y^{\wedge}	D^{\wedge}	$(X^T X)^{-1}$								
1	1	0.89	163	0.89	1.33	1	0	166	157	-9	155.1	-10.9	34.745	-2.153	-0.216						
1	2	-0.04	155	-1.49	-0.12	1	0	150	145	-5	138.6	-11.4	-2.153	0.322	0.013						
1	3	-0.70	148	1.89	-1.12	1	0	154	134	-20	142.7	-11.3	-0.216	0.013	0.001						
:																					
						1	0	155.6	$\bar{x}_{..}$			最小2乗平均									
1				平均 μ^{\wedge}_1		160.3	149.2	-11.1	144.3		-11.24										
				SD σ^{\wedge}_1		6.0	7.5	3.9													
						1	1	155.6	$\bar{x}_{..}$			最小2乗平均									
2				平均 μ^{\wedge}_2		150.8	140.5	-10.3	145.4		-10.16										
				SD σ^{\wedge}_2		6.8	7.7	3.6													
				平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$		-9.50	-8.70	0.80	1.08		1.08										
				差の SE		2.87	3.40	1.68	2.19		2.19										
				t 値		-3.309	-2.559	0.475	0.491		0.491										
				両側 p 値		0.004	0.020	0.640	0.629		0.629										
						**	*	NS	NS		NS										



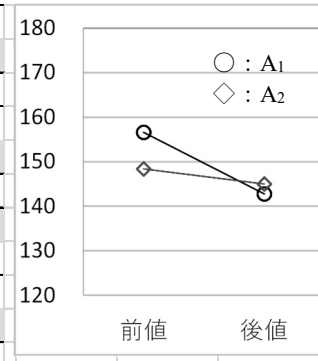
共変量とした場合の前後差 D の最小 2 乗平均は、 A_1 群で $\hat{\mu}_1 = -11.24$ ， A_2 群で $\hat{\mu}_2 = -10.16$ であり，その差は 1.08 であり， $p = 0.629$ と有意ではない．表 9.22 の右下の線グラフをみれば，後値 Y でのみ有意な差があるという結果は，まさに前値 X での有意な差が，そのまま後値 Y に反映した結果と読み取れる．

3): ($\bar{X}_{A_1} \gg \bar{X}_{A_2}$; 相関が小) 場合の α エラー

表 9.23 に示すように，時点間の相関が 0.0 に近づくと，前値 X の群平均での大小関係が後値 Y に影響を及ぼしにくくなりなり，後値 Y の p 値は，名目の α エラーに接近することが期待される．前後差 D の場合は，前値 X の群平均での大小関係の影響が薄れることにより，逆に前後差 D が増大する方向となり，実質の α エラーが増大すると思われる．

表 9.23 前値 X で A_1 群が大きく相関が 0.1 の場合のシミュレーションの一例

2群の平均=		155	SD	相関の	A_1 群効果	-10.0											
個体間分散=		10	3.16	期待値	A_2 群効果	-10.0											
個体内分散=		90	9.49	0.1													
個体間			個体内乱数		デザイン行列			前値 X 共変量									
A	No	正規	個体間	正規	正規		前値	後値	差	推定値	推定値						
i	j	乱数 Z_0	XY	Z_1	Z_2	μ	a	X	Y	D	Y^{\wedge}	D^{\wedge}	$(X^T X)^{-1}$				
1	1	0.36	156	-0.27	-0.63	1	0	153	140	-13	142.2	-10.8	19.187	-1.099	-0.122		
1	2	-0.28	154	-0.91	-1.34	1	0	145	131	-14	140.8	-4.2	-1.099	0.252	0.006		
1	3	-0.02	155	-0.09	1.28	1	0	154	157	3	142.4	-11.6	-0.122	0.006	0.001		
:																	
						1	0	152.50	$\bar{x}_{..}$		最小2乗平均						
1								平均 μ^{\wedge}_1	156.6	142.8	-13.8	142.1	-10.40				
								SD σ^{\wedge}_1	10.1	10.3	11.5						
						1	1	152.50	$\bar{x}_{..}$		最小2乗平均						
2								平均 μ^{\wedge}_2	148.4	145.0	-3.4	145.7	-6.80				
								SD σ^{\wedge}_2	6.4	11.1	13.8						
								平均の差 $\mu^{\wedge}_2 - \mu^{\wedge}_1$	-8.20	2.20	10.40	3.60	3.60				
								差のSE	3.78	4.79	5.69	5.48	5.48				
								t 値	-2.170	0.459	1.829	0.656	0.656				
								両側 p 値	0.044	0.651	0.084	0.520	0.520				
									*	NS	+	NS	NS				



前値 X の群間で有意な差 $p < 0.044$ であるが，時点間の相関が小さいために後値 Y に前値 X の大小の影響が極めて弱くなり，後値 Y では $p = 0.651$ と全く差がないものの，前後差 D で $p = 0.084$ と微妙な結果となる．前値 X を共変量とした場合に $p = 0.520$ と NS となる．

4): ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) の場合の α エラー

前値 X が，ほぼ等しい ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) の場合の第 1 種の過誤 (α エラー) について考えてみよう．時点間の相関が $\rho = 0.5$ に近い場合に実施の α エラーについて推論することは難儀である．ただし， $r = 1.0$ に近い場合は，前値 X と後値 Y も同程度の群間差となるので，有

意な差となる場合が極端に減り、実質の α エラーよりも小さめになることが推論される。時点間の相関が $\rho=0.0$ に近づいた場合には、後値 Y は前値 X の影響を受けないので、実質の α エラーは名目の α エラーに近づくと推測される。

前後差 D の実質の α エラーは、時点間の相関 ρ によってどのような影響を受けるのであろうか。相関が $\rho=0.0$ に近づくとつれて、前後差 D の分散は、分散の加法性により、前値 X の分散と後値 Y の分散の和に漸近するので、差の SE は $\sqrt{2}=1.41$ 倍となり、 t 値は小さ目となり、実質の α エラーよりも小さめになることが推論される。逆に、相関が $\rho=1.0$ に近づいた場合は、どうなるのだろうか。後値 Y の分散よりも前後差 D の分散が小さくなるために、前値 X で、2 群間にきわどいところで有意な差でない場合、有意な差となる可能性が高くなるのではないかと思うと、名目の α エラーを保持すると推論される。

5): ($\bar{X}_{A_1} < \bar{X}_{A_2}$) 場合の α エラー

1): ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合の逆であるが、両群共に -10.0 の減少を仮定しているので、どちらの群が大きいか小さいかを区別することなく、実質の α エラーは同様と判断される。

SAS によるシミュレーションによる第 1 種の過誤の検討

表 9.21 に示した Excel のシミュレーションと同様の設定を SAS のプログラミング機能を用いて設定した結果を示す。あらかじめ設定した時点間の相関ごとに 10,000 回の試行を行い前値 X に対する対応のある t 検定の両側確率 p 値を計算する。表 9.24 に示すのは、両側確率が $p \leq 0.40$ の場合に、 A_1 群の平均値 \bar{x}_1 と A_2 群の平均値 \bar{x}_2 と比較し、 $\bar{x}_1 > \bar{x}_2$ の場合を (1: $\bar{X}_{A_1} > \bar{X}_{A_2}$)、逆に $\bar{x}_1 < \bar{x}_2$ の場合を (3: $\bar{X}_{A_1} < \bar{X}_{A_2}$) と分類する。両側確率が $p > 0.40$ の場合を (2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$) のように 3 区分する。

次に、後値 Y および前後差 D について t 検定の両側の p 値、前値 X を共変量とした後値 Y についての共分散分析による群間差についての両側の p 値を求める。これを 10,000 回繰返し、

表 9.24 時点間の相関が 0.5 の場合の前値 X の大小による実質の α エラー

相関 ρ	前値 X p 値の比較	回数 n	後値 Y		前後差 D		共分散 Cov	
			n_Y	$\alpha_Y \%$	n_D	$\alpha_D \%$	n_{Cov}	$\alpha_{Cov} \%$
0.50	1: $\bar{X}_{A_1} > \bar{X}_{A_2}$	2,024	172	8.5	147	7.3	94	4.6
0.50	2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$	6,055	163	2.7	191	3.2	297	4.9
0.50	3: $\bar{X}_{A_1} < \bar{X}_{A_2}$	1,921	143	7.4	154	8.0	108	5.6
	全体	10,000	478	4.78	492	4.92	499	4.99

(1: $\bar{X}_{A_1} > \bar{X}_{A_2}$), (2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$), (3: $\bar{X}_{A_1} < \bar{X}_{A_2}$) における $p < 0.05$ の回数をカウントした結果が示されている。

表 9.24 の全体=10,000 に対する, 後値 Y で $p < 0.05$ となる回数は $n_Y = 478$, 前後差 D で $n_D = 492$, 共分散分析で $n_{Cov} = 499$ となり, 名目の α エラー 5.0% に対し, それぞれ (4.78%, 4.92%, 4.99%) であり, ミュレーションが適切に行なわれていることが示されている。

前値 X の群間の有意差検定の p 値で 3 通りに分けると (1: $\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合は, 10,000 回中 2,024 回で後値 Y が有意となるのは 172 回で実質の α エラーは 8.5% と増加する. (2: $\bar{X}_{A_1} = \bar{X}_{A_2}$) の場合は 6,055 回中, 実質の α エラーは 2.7% と減少する. このことは, 前値 X の両群の平均値の大小関係により α エラーが変化し, 適切な有意差検定が行なわれなくなっていることが示されている. 前後差 D も同様の結果で, 前値 X の両群の平均値の大小関係により (7.3%, 3.2%, 8.0%) と変化し名目の α エラーが担保されていないことが示されている. 前値 X を共変量とした場合は (4.6%, 4.9%, 5.6%) とほぼ名目の α エラーが保持されているとの結果となっている。

さて, 時点間の相関を (0.05, 0.10, ..., 0.50, ..., 0.95, 0.98) のように変化させた場合は実質の α エラーはどのように変化するのであろうか, 表 9.25 (図 9.17) に結果を示す. 図 9.17 中 (2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$) の場合は, 相関が低くなるにつれ後値 Y の場合は, 実質の α エラーは減少し, 前後差 D の場合は名目の α エラー 5% に近づき, 相関が高くなる場合は, 逆転する. 前値 X を共変量とした場合は, 相関の変化に関わらず実質の α エラーは, 名目の $\alpha = 5\%$ を保っていることが分かる。

表 9.25 時点間の相関を変化の場合の前値 X の大小による実質の α エラー

時点間 相関	1: $\bar{X}_{A_1} > \bar{X}_{A_2}$				2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$				3: $\bar{X}_{A_1} < \bar{X}_{A_2}$			
	Y	D	Cov		Y	D	Cov		Y	D	Cov	
ρ	n_1	α_{1Y}	α_{1D}	α_{1Cov}	n_2	α_{2Y}	α_{2D}	α_{2Cov}	n_3	α_{3Y}	α_{3D}	α_{3Cov}
0.05	2,044	5.4	10.5	4.9	5,928	5.1	1.4	5.0	2,028	5.4	10.7	5.6
0.10	1,912	5.1	9.9	4.8	6,123	4.8	1.4	4.9	1,965	5.4	8.8	4.8
0.20	1,969	6.1	9.7	5.3	6,028	4.2	2.1	4.7	2,003	5.1	8.4	4.8
0.30	2,034	6.2	8.9	5.3	6,004	4.4	2.4	5.0	1,962	5.9	10.3	4.8
0.40	1,949	6.1	8.2	5.0	6,023	3.7	2.6	5.2	2,028	8.2	7.7	4.8
0.50	2,024	8.5	7.3	4.6	6,055	2.7	3.2	4.9	1,921	7.4	8.0	5.6
0.60	1,916	9.4	8.6	5.6	6,074	2.2	3.8	5.2	2,010	9.2	6.8	5.5
0.70	2,001	9.3	7.8	4.3	6,008	1.2	3.8	4.9	1,991	10.9	6.3	4.7
0.80	1,991	11.0	6.2	4.8	5,969	0.6	3.9	4.6	2,040	12.0	6.9	5.3
0.85	2,005	12.0	6.6	5.5	5,982	0.4	4.6	5.1	2,013	12.1	5.4	5.3
0.90	1,957	12.6	5.3	5.2	6,083	0.0	4.6	4.9	1,960	11.2	5.8	4.6
0.95	2,032	13.5	5.5	5.0	6,037	0.0	4.8	4.7	1,931	12.3	4.6	4.1
0.98	1,922	12.6	5.5	4.9	6,042	0.0	4.8	4.9	2,036	12.5	5.3	5.2

図 9.17 左 (1: $\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合は、相関が低くなるにつれ後値 Y の場合は、実質の α エラーは増大し、前後差 D の場合は名目の α エラー 5% に近づき、相関が高くなる場合は、逆転する。前値 X を共変量とした場合は、相関の変化に関わらず実質の α エラーは、名目の α エラー 5% を保っていることが分かる。図 9.17 右 (3: $\bar{X}_{A_1} < \bar{X}_{A_2}$) の場合は、図 9.17 左 (1: $\bar{X}_{A_1} > \bar{X}_{A_2}$) と同様の結果となっている。

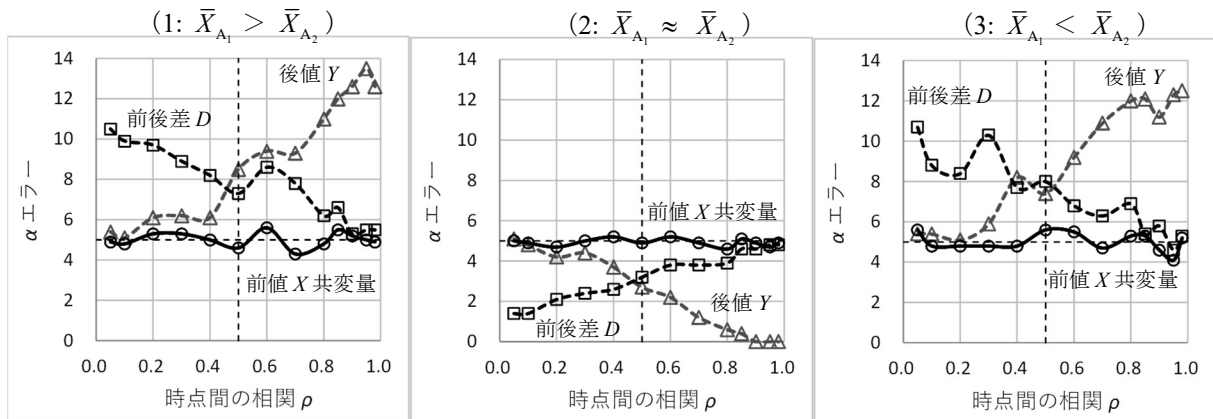


図 9.17 時点間相関を変化させた場合の実質の α エラー

SAS によるシミュレーションの結果、時点間の相関の大小に関連し、前値 X でも前後差 D でも、実質の α エラーが大きく変化するのであるが、前値 X を共変量とした場合は、相関の変化に関わらず実質の α エラーは、名目の α エラー 5% を常に保っていることが分かる。

前値 X と後値 Y が得られるような 2 群の実験データにおける群間比較をする際に、後値 Y に対する対応のない t 検定か、前後差 $D = Y - X$ による対応のない t 検定か、前値 X を共変量とする共分散分析か、の選択が常に付きまとう。厄介なのは、前値 X と後値 Y に相関の大小により、後値 Y と前後差 D の群間比較の検出力が大きく変化すが、前値 X を共変量とする共分散分析による群間比較が、前値 X と後値 Y に相関の大小に関わらず、後値 Y と前後差 D の群間比較よりも検出力が高くなること知られている。しかし、共分散分析は、難しい手法であり、「群と前値 X に交互作用があると使えない」ないので、前後差 D の群間比較の適用が望ましいとの主張に、しばしば遭遇する。

このような不毛の議論に終止符を打つために、事前に群分けがされていない場合の前値 X の群平均の大きさに関連し、後値 Y と前後差 D の群間比較に際し、第 1 種の過誤 (α エラー) が揺れ動き、比較可能性の前提が成り立たっていないことを示し、後値 Y と前後差 D どちらの場合であっても前値 X を共変量とすることにより、 α エラーを一定に保つことができる事を示した。

9.6. 前値 X と後値 Y をめぐる各種の群間比較の検出力 ($1 - \beta$ エラー)

統計的方法の性能を見極めるための方法として前節では、名目の α エラーに対し実質の α エラーの乖離の程度を可視化した。本節では、第2種の過誤として知られている β エラーに代えて検出力 ($1 - \beta$ エラー) を用いて、時点間の相関に関連して後値 Y 、前後差 D 、前値 X を共変量とした場合の統計的な性能比較を行う。

表 9.26 に示したのは、 A_1 群、 A_2 群共に共通の2群の平均 $\mu_0 = 155.0$ に対し、後値 Y の期待値が A_1 群の効果で -10.0 減少して $\mu_1 = 145.0$ 、 A_2 群の効果で -20.0 減少して $\mu_2 = 135.0$ となるように設定している。このように基準とする A_1 群に対し A_2 群の平均を一定数ずらした場合には、正規乱数を加味した後値 Y および前後差 D の計算過程、前値 X を共変量とする後値 Y の推定値 \hat{Y} および前後差 D の推定値 \hat{D} の算出過程を示している。

表 9.26 検出力を可視化するための Excel シートの準備

2群の平均=	155	SD	相関の	A_1 群効果	-10.0					
個体間分散=	50	7.07	期待値	A_2 群効果	-20.0					
個体内分散=	50	7.07	0.5							
	個体間乱数	個体内乱数	デザイン行列						前値X共変量	
A No	正規 個体間	正規 正規		前値	後値	差		推定値	推定値	
i j	乱数Z	XY	$Z_{前}$ $Z_{後}$	μ a	X	Y	D	\hat{Y}	\hat{D}	

正規乱数： $(Z, Z_{前}, Z_{後}) = \text{Norm.Inv}(\text{Rand}(), 0, 1)$, $XY = \text{Round}((155 + Z * 7.07), 0)$,
 デザイン行列： $[\mu_{切片} = 1, (a_{A1} = 0, a_{A2} = 1)]$, $X = \text{Round}((XY + Z_{前} * 7.07), 0)$,
 $Y_{A1} = \text{Round}((XY - 10.0 + Z_{後} * 7.07), 0)$, $Y_{A2} = \text{Round}((XY - 20.0 + Z_{後} * 7.07), 0)$, $D = Y - X$,
 $\hat{\beta} = (X^T X)^{-1} X^T Y$, $\hat{Y} = X \hat{\beta}$, $\hat{\beta}' = (X^T X)^{-1} X^T D$, $\hat{D} = X \hat{\beta}'$

表 9.20 に示した α エラーのための Excel シートでは、 A_1 群と A_2 群で、等しい効果とし、前値 X に対し後値 Y が -10.0 減少すると設定していたが、検出力の検討のための表 9.26 では、 A_2 群の効果として -20.0 減少するように変更している。

α エラー：	<table border="1"><tr><td>A_1群効果</td><td>-10.0</td></tr><tr><td>A_2群効果</td><td>-10.0</td></tr></table>	A_1 群効果	-10.0	A_2 群効果	-10.0	検出力：	<table border="1"><tr><td>A_1群効果</td><td>-10.0</td></tr><tr><td>A_2群効果</td><td>-20.0</td></tr></table>	A_1 群効果	-10.0	A_2 群効果	-20.0
A_1 群効果	-10.0										
A_2 群効果	-10.0										
A_1 群効果	-10.0										
A_2 群効果	-20.0										

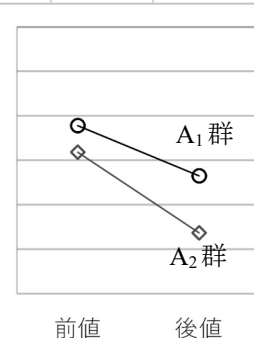
SAS で実施したシミュレーション結果を示す前に、Excel シート上で典型的な事例を概観し、どのような検出力になるのかを推論することにより、シミュレーション結果の理解の助になるようにしたい。 A_1 群、 A_2 群の前値 X の平均の大小関係を 1： $(\bar{X}_{A1} > \bar{X}_{A2})$ 、2： $(\bar{X}_{A1} \approx \bar{X}_{A2})$ 、3： $(\bar{X}_{A1} < \bar{X}_{A2})$ と分類し、時点間の相関を変化させた場合に、検出力がどのように変化するのかを推論する。

1: 前値 X の差が ($\bar{X}_{A_1} > \bar{X}_{A_2}$) と大きい場合の検出力

表 9.27 に示すのは、前値 X と後値 Y の相関の期待値を $\rho=0.50$ とし、前値 X の平均値が 1: ($\bar{X}_{A_1} > \bar{X}_{A_2}$) と大きい場合に、後値 Y で有意な差が出やすく、前後差 D で有意な差が出にくくなると想定される。数式タブの「シート再計算」により、20×3 の正規乱数の発生を繰り返し、想定通りの結果が得られた場合のスナップショットした結果である。

表 9.27 $\rho=0.5$ の場合の検出力の計算手順の Excel シートによる可視化

2群の平均=		155		SD	相関の	A ₁ 群効果	-10.0													
個体間分散=		50		7.07	期待値	A ₂ 群効果	-20.0													
個体内分散=		50		7.07	0.5															
		個体内乱数		個体内乱数		デザイン行列			前値X共変量											
A	No	正規	個体間	正規	正規		前値	後値	差	推定値	推定値									
i	j	乱数Z	XY	Z _前	Z _後	μ	a	X	Y	D	Y [^]	D [^]	(X ^T X) ⁻¹							
1	1	-1.55	144	-0.84	0.16	1	0	138	135	-3	135.3	-2.7	14.876	-0.662	-0.094					
1	2	-0.45	152	-0.63	-1.79	1	0	148	129	-19	141.0	-7.0	-0.662	0.221	0.004					
1	3	1.90	168	-0.48	1.24	1	0	165	167	2	150.6	-14.4	-0.094	0.004	0.001					
1	4	-1.64	143	0.95	1.34	1	0	150	142	-8	142.1	-7.9								
1	5	1.04	162	1.34	0.68	1	0	171	157	-14	153.9	-17.1	β [^] =(X ^T X) ⁻¹ X ^T Y							
1	6	-0.30	153	2.64	0.07	1	0	172	143	-29	154.5	-17.5	57.599			57.599				
1	7	-0.08	154	0.56	-0.89	1	0	158	138	-20	146.6	-11.4	-9.420			-9.420				
1	8	0.56	159	-0.13	2.19	1	0	158	164	6	146.6	-11.4	0.563			-0.437				
1	9	-0.27	153	-0.44	-0.28	1	0	150	141	-9	142.1	-7.9	β [^] =(X ^T X) ⁻¹ X ^T D							
1	10	-0.39	152	2.33	0.95	1	0	168	149	-19	152.2	-15.8								
2	1	-0.60	151	-0.88	-0.43	1	1	145	128	-17	129.9	-15.1	Y:σ ^{^2}		D:σ ^{^2}					
2	2	1.28	164	-0.98	-1.62	1	1	157	133	-24	136.6	-20.4	86.337			86.337				
2	3	0.66	160	-1.40	0.27	1	1	150	142	-8	132.7	-17.3								
2	4	-1.90	142	1.03	-0.27	1	1	149	120	-29	132.1	-16.9	Σ(β [^])=(X ^T X) ⁻¹ σ ² [^]							
2	5	-0.72	150	1.18	0.05	1	1	158	130	-28	137.2	-20.8	1284.4	-57.141	-8.085					
2	6	0.55	159	-1.99	0.68	1	1	145	144	-1	129.9	-15.1	-57.141	19.112	0.307					
2	7	0.42	158	-0.93	-0.76	1	1	151	133	-18	133.2	-17.8	-8.085	0.307	0.051					
2	8	0.21	156	1.78	1.10	1	1	169	144	-25	143.4	-25.6	パラメータの共分散行列							
2	9	-1.10	147	-0.63	-0.18	1	1	143	126	-17	128.7	-14.3								
2	10	0.42	158	-0.93	-0.17	1	1	151	137	-14	133.2	-17.8								
						I ⁽¹⁾ =	1	0	154.80	$\bar{x}_{..}$	最小2乗平均									
1							平均 μ [^] ₁	157.8	146.5	-11.3	144.81	-9.99								
						SD σ [^] ₁	11.3	12.5	10.9											
						I ⁽²⁾ =	1	1	154.80	$\bar{x}_{..}$	最小2乗平均									
2							平均 μ [^] ₂	151.8	133.7	-18.1	135.39	-19.41								
						SD σ [^] ₂	7.8	8.1	8.9											
						平均の差 μ [^] ₂ -μ [^] ₁	-6.00	-12.80	-6.80	-9.42	-9.42									
						差のSE	4.33	4.72	4.46	4.37	4.37									
						t値	-1.387	-2.714	-1.525	-2.155	-2.155									
						両側 p値	0.183	0.014	0.145	0.046	0.046									
							NS	*	NS	*	*									



差の SE = Sqrt(((9*σ₁²)+(9*σ₂²))/18*(1/10+1/10))

σ_y² = (Y - Ŷ)^T(Y - Ŷ) / 17 = 86.337, σ_D² = (Y - Ŷ)^T(Y - Ŷ) / 17 = 86.337, Σ(β[^]) = (X^TX)⁻¹σ_y² = (X^TX)⁻¹σ_D²

最小 2 乗平均: β[^] = (X^TX)⁻¹X^TY, ŷ₁^{lsmean} = I⁽¹⁾β[^] = 144.81, ŷ₂^{lsmean} = I⁽²⁾β[^] = 135.39

最小 2 乗平均: β[^] = (X^TX)⁻¹X^TD, d₁^{lsmean} = I⁽¹⁾β[^] = -9.99, d₂^{lsmean} = I⁽²⁾β[^] = -19.41

ŷ₂^{lsmean} - ŷ₁^{lsmean} = d₂^{lsmean} - d₁^{lsmean} = -9.42, 差の SE^{lsmean} = √Σ(β[^])_{2,2} = √19.112 = 4.37}

表 9.27 の脚注に表 9.26 の続きの計算式が示されている。差の SE は、それぞれの t 検定の分母の計算である。さらに、前値 X を共変量とした解析の続きで、前値 X を共変量とした場合の推定値 \hat{Y} および \hat{D} に対する誤差分散 $\hat{\sigma}_Y^2$ および $\hat{\sigma}_D^2$ の結果が同じ 86.337 となり、パラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算、デザイン行列の変数に対応する変数ベクトル $\mathbf{I}^{(1)}$ および $\mathbf{I}^{(2)}$ を用いた最小 2 乗平均、それらの差の SE が $\Sigma(\hat{\beta})$ の 2 行 2 列目の平方根となることが示される。

前値 X の A_1 群の平均値は $\bar{X}_{A_1} = 157.8$ 、 A_2 群は $\bar{X}_{A_2} = 151.8$ で $p = 0.183$ となり、後値 Y で $p = 0.014^*$ 、前後差 D で $p = 0.145$ となっている。表 9.28 に示すように、 A_1 群の前後差 D は -11.3 であり、 A_1 群の効果 -10.0 よりも大きい。これは、前値 X が 157.8 と設定した平均値 155.0 よりも大きいので期待した差 -10.0 よりも平均への回帰により大きな -11.3 の減少となったと解される。 A_2 群の平均の差は -18.1 であり、設定した -20.0 よりも小さくなり、平均への回帰と解される。共分散分析では、 $p = 0.046^*$ と後値 Y と前後差 D の p 値の間となっている。

表 9.28 「平均への回帰」現象の典型的な事例

	前値	後値	差	
	X	Y	D	
A_1 : 平均	157.8	146.5	-11.3	-10 を超えている
A_2 : 平均	151.8	133.7	-18.1	-20 にとどかない
平均の差	-6.00	-12.80	-6.80	差が 10 以下

相関の期待値: $\rho = 0.90$

表 9.29 に示すのは、前値と後値の相関の期待値を $\rho = 0.90$ とした場合である。時点間の相関が高くなるにつれて「平均への回帰」が弱まり、前値 X で 2 群間に $1 : (\bar{X}_{A_1} > \bar{X}_{A_2})$ の関係があったとしても、前値 X での群間差が後値 Y で縮まることが少なくなる。そのため、後値 Y の群間差では有意な差が極めて出やすくなる。前後差 D では、前値 X の群間差の影響が相殺される、平均への回帰により群間差が設定した効果 -10.0 よりやや小さくなる。そのために、後値 Y の場合に比べ検出力 $(1 - \beta)$ はやや小さくなると推測される。

後値 Y の平均の差は、前値 X の差が加わり設定した効果 -10.0 に対し -12.40 と拡大し、 $p = 0.014^*$ と有意な差となっている。前後差 D では、平均の差 $= -6.30$ と平均への回帰により設定した効果 -10.0 よりも平均への回帰により縮小しているが、 SE が 2.71 と小さくなり、結果として $p = 0.033^*$ と有意な差となっている。共分散分析では、 $p = 0.042$ と後値 Y と前後差 D の p 値よりも大きくなっている。

表 9.29 ρ = 0.9 の場合の検出力の可視化

2群の平均=		155		SD	3.16	相関の期待値	0.9	A ₁ 群効果	-10.0									
個体間分散=		90		SD	9.49	相関の期待値		A ₂ 群効果	-20.0									
個体内分散=		10		SD	3.16	相関の期待値	0.9											
		個体間乱数		個体内乱数		デザイン行列			前値 X 共変量									
A	No	正規	個体間	正規	正規			前値	後値	差	推定値	推定値						
i	j	乱数 Z	XY	Z	Z	μ	a	X	Y	D	Y [^]	D [^]	(X ^T X) ⁻¹					
1	1	1.09	165	-2.26	0.62	1	0	158	157	-1	145.8	-12.2	18.708	-0.820	-0.118			
1	2	0.26	157	1.16	-0.89	1	0	161	144	-17	148.6	-12.4	-0.820	0.228	0.005			
1	3	-1.31	143	1.94	-1.61	1	0	149	128	-21	137.2	-11.8	-0.118	0.005	0.001			
:																		
				I ⁽¹⁾ =		1	0	154.65	$\bar{x}_{..}$	最小2乗平均								
1				平均 μ [^] ₁		157.7	145.5	-12.2	142.6	-12.1								
				SD σ [^] ₁		5.3	9.4	8.0										
				I ⁽²⁾ =		1	1	154.65	$\bar{x}_{..}$	最小2乗平均								
2				平均 μ [^] ₂		151.6	133.1	-18.5	136.0	-18.6								
				SD σ [^] ₂		11.0	11.0	3.2										
				平均の差 μ [^] ₂ - μ [^] ₁		-6.10	-12.40	-6.30	-6.59	-6.59								
				差の SE		3.85	4.57	2.73	2.99	2.99								
				t 値		-1.58	-2.71	-2.31	-2.20	-2.20								
				両側 p 値		0.131	0.014	0.033	0.042	0.042								
						NS	*	*	*	*								

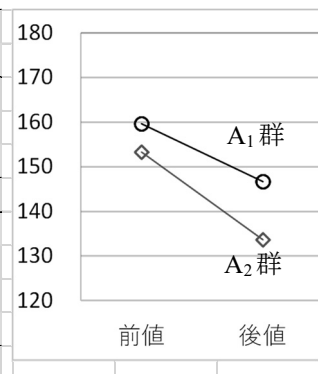


相関の期待値: ρ = 0.10

表 9.30 に示すのは、前値と後値の相関の期待値を ρ = 0.10 とした場合である。前値 X で 2 群間に 1 : ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の関係でも場合、前値 X での群間差が後値 Y に影響することが、ほ

表 9.30 ρ = 0.1 の場合の検出力の可視化

2群の平均=		155		SD	3.16	相関の期待値	0.1	A ₁ 群効果	-10.0									
個体間分散=		10		SD	9.49	相関の期待値		A ₂ 群効果	-20.0									
個体内分散=		90		SD	3.16	相関の期待値	0.1											
		個体間乱数		個体内乱数		デザイン行列			前値 X 共変量									
A	No	正規	個体間	正規	正規			前値	後値	差	推定値	推定値						
i	j	乱数 Z	XY	Z	Z	μ	a	X	Y	D	Y [^]	D [^]	(X ^T X) ⁻¹					
1	1	-1.14	151	1.39	-0.01	1	0	164	141	-23	147.5	-16.5	23.984	-1.043	-0.150			
1	2	0.46	156	0.89	-1.13	1	0	164	135	-29	147.5	-16.5	-1.043	0.237	0.006			
1	3	-1.53	150	-0.32	0.21	1	0	147	142	-5	144.4	-2.6	-0.150	0.006	0.001			
:																		
				I ⁽¹⁾ =		1	0	156.45	$\bar{x}_{..}$	最小2乗平均								
1				平均 μ [^] ₁		159.6	146.7	-12.9	146.1	-10.31								
				SD σ [^] ₁		7.3	11.9	14.4										
				I ⁽²⁾ =		1	1	156.45	$\bar{x}_{..}$	最小2乗平均								
2				平均 μ [^] ₂		153.3	133.6	-19.7	134.2	-22.29								
				SD σ [^] ₂		8.1	11.7	12.1										
				平均の差 μ [^] ₂ - μ [^] ₁		-6.30	-13.10	-6.80	-11.97	-11.97								
				差の SE		3.44	5.27	5.95	5.87	5.87								
				t 値		-1.83	-2.49	-1.14	-2.04	-2.04								
				両側 p 値		0.084	0.023	0.268	0.057	0.057								
						NS	*	NS	+	+								



ばなくなり群間の効果の差 -10.0 に対する検出力に接近すると想定される。前後差 D は、前値 X の群間差が相殺され、前値 X の揺らぎと後値 Y の揺らぎが加わり差の SE が増大し、さらに平均への回帰により群間差が縮まる方向となり、検出力が極端に減弱すると予想される。

後値 Y の場合は、表 9.29 に示した $\rho=0.90$ の場合と同様に $p=0.023^*$ となっているが、前後差 D では、差の SE が、 2.73 から 5.95 と増大した結果、 $p=0.268$ と有意な差となくなっている。共分散分析では、 $p=0.057$ と後値 Y と前後差 D の p 値の間となっている時点間の相関が $\rho=0.10$ のような場合に前後差 D での検出力の落ち込みについては、SAS によるシミュレーションで明らかにする。

2: 前値 X が ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) と小さい場合の検出力

前値 X の群間差が 2: ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) と小さい場合に後値 Y の群間差は、時点間の相関が大きくなった場合にどのようなのであろうか。1: ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合のように前値 X の群間差の上積みは期待できないので、後値 Y の群間差の検出力 ($1-\beta$) が大きくなることはないので、設定した群間差 -10.0 に対する検出力と想定される。時点間の相関が小さい場合には、1: ($\bar{X}_{A_1} > \bar{X}_{A_2}$) と同様と予測される。これらの推測から 2: ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) 場合に、後値 Y の検出力は、時点間の相関の大小に関わらず一定と想定される。

前後差 D の場合は、どうなるのであろうか。前後差 D の場合は、前値 X の群間差の影響を受けづらく、平均への回帰の影響も小さくなり、1: ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合に比べて群間差はやや大きくなり、検出力が増加方向になると想定される。

3: 前値 X の差が ($\bar{X}_{A_1} < \bar{X}_{A_2}$) と逆に大きい場合の検出力

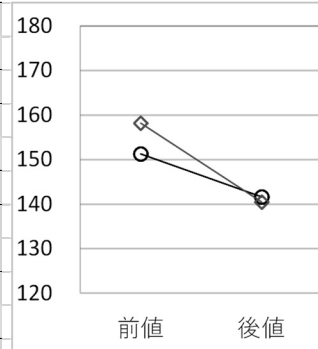
相関の期待値: $\rho=0.90$

時点間の相関が 0.9 の場合の Excel によるシミュレーション結果を表 9.31 に示す。前値 X で A_2 群が A_1 群より 6.90 大きくなっている。時点間の相関が大きい場合には、後値 Y は、それぞれの群の減少効果を反映することになり、群間差が小さくなり検出力 ($1-\beta$) は、激減すると想定される。

前後差 D では、前値 X の A_2 群が A_1 群より大きい場合に平均への回帰により、 A_2 群の減少効果に上積みされ、逆に A_1 群は、上昇方向に働くことになり、前後差 D での検出力は 2: ($\bar{X}_{A_1} < \bar{X}_{A_2}$) に比べて増大が見込まれる。

表 9.31 ρ = 0.9 の場合の検出力の可視化

2群の平均=		155	SD	相関の	A ₁ 群効果	-10.0									
個体間分散=		90	9.49	期待値	A ₂ 群効果	-20.0									
個体内分散=		10	3.16	0.9											
		個体間乱数		個体内乱数		デザイン行列			前値 X 共変量						
A	No	正規	個体間	正規	正規	前値	後値	差	推定値	推定値					
i	j	乱数 Z	XY	Z	Z	μ	a	X	Y	D	Y [^]	D [^]	(X ^T X) ⁻¹		
1	1	0.77	162	-0.46	-0.17	1	0	161	151	-10	152.2	-8.8	14.891	0.575	-0.098
1	2	-0.43	151	-0.23	1.57	1	0	150	146	-4	140.2	-9.8	0.575	0.231	-0.004
1	3	-1.83	138	0.86	-2.49	1	0	141	120	-21	130.3	-10.7	-0.098	-0.004	0.001
:															
				I ⁽¹⁾ =		1	0	154.75	$\bar{x}_{..}$	最小2乗平均					
1			平均 μ [^] ₁		151.3	141.6	-9.7	145.4	-9.37						
				SD σ [^] ₁		7.2	10.3	5.1							
				I ⁽²⁾ =		1	1	154.75	$\bar{x}_{..}$	最小2乗平均					
2			平均 μ [^] ₂		158.2	140.5	-17.7	136.7	-18.03						
				SD σ [^] ₂		11.0	11.9	4.1							
				平均の差 μ [^] ₂ - μ [^] ₁		6.90	-1.10	-8.00	-8.65	-8.65					
				差の SE		4.15	4.97	2.07	2.25	2.25					
				t 値		1.66	-0.22	-3.86	-3.85	-3.85					
				両側 p 値		0.113	0.827	0.001	0.001	0.001					
						NS	NS	**	**	**					

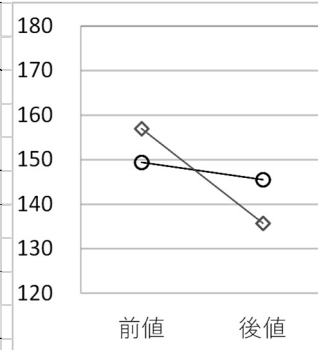


相関の期待値: ρ = 0.10

時点間の相関が 0.1 の場合の Excel によるシミュレーション結果を表 9.32 に示す。前値 X で A₂ 群が A₁ 群より 7.6 大きくなっている。時点間の相関が低いこともあり平均への回帰が強めに働き、後値 Y の群間差に p = 0.015* との結果になっている。そのため、名目の群間差

表 9.32 ρ = 0.1 の場合の検出力の可視化

2群の平均=		155	SD	相関の	A ₁ 群効果	-10.0									
個体間分散=		10	3.16	期待値	A ₂ 群効果	-20.0									
個体内分散=		90	9.49	0.1											
		個体間乱数		個体内乱数		デザイン行列			前値 X 共変量						
A	No	正規	個体間	正規	正規	前値	後値	差	推定値	推定値					
i	j	乱数 Z	XY	Z _前	Z _後	μ	a	X	Y	D	Y [^]	D [^]	(X ^T X) ⁻¹		
1	1	-2.05	149	-0.71	0.14	1	0	142	140	-2	142.2	0.2	14.423	0.629	-0.096
1	2	-1.26	151	-0.47	-0.57	1	0	147	136	-11	144.4	-2.6	0.629	0.237	-0.005
1	3	0.43	156	-1.05	0.18	1	0	146	148	2	144.0	-2.0	-0.096	-0.005	0.001
:															
				I ⁽¹⁾ =		1	0	153.20	$\bar{x}_{..}$	最小2乗平均					
1			平均 μ [^] ₁		149.4	145.5	-3.9	147.2	-6.03						
				SD σ [^] ₁		7.0	7.7	6.6							
				I ⁽²⁾ =		1	1	153.20	$\bar{x}_{..}$	最小2乗平均					
2			平均 μ [^] ₂		157.0	135.7	-21.3	134.0	-19.17						
				SD σ [^] ₂		11.1	8.7	10.6							
				平均の差 μ [^] ₂ - μ [^] ₁		7.60	-9.80	-17.40	-13.15	-13.15					
				差の SE		4.16	3.66	3.93	3.55	3.55					
				t 値		1.83	-2.68	-4.42	-3.70	-3.70					
				両側 p 値		0.084	0.015	0.000	0.002	0.002					
						+	*	***	**	**					



-10.0が保持しやすいとも考えられ、そこそこの検出力 $(1-\beta)$ が見込まれる。前後差 D の場合は、平均への回帰により差が拡大する方向となり、検出力はかなりの増加が見込まれる。

SASによるシミュレーションによる検出力 $(1-\beta)$ の検討

これまで示したきた Excel シートにより、時点間に相関がある前値 X と後値 Y に関連する統計的な諸問題の典型例について可視化し、検出力 $(1-\beta)$ について想定してきた。図 9.17 と表 9.25 に SAS のプログラミング機能を用いて各種の条件下での 10,000 回の試行を行い、第 1 種の過誤 α エラーについて検討結果を示した。検出力 $(1-\beta)$ は、同じ SAS のプログラムで、 A_2 群の減少効果を -10 から -20 へ変更することにより得られる。

前値 X について A_1 群と A_2 群間の平均値の差について対応のない t 検定を行い、両側 p 値を求め、次の 3 区分

1 : 前値 X ($\bar{X}_{A_1} > \bar{X}_{A_2}$) : 両側 $p \leq 0.40$

2 : 前値 X ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) : 両側 $p > 0.40$

3 : 前値 X ($\bar{X}_{A_1} < \bar{X}_{A_2}$) : 両側 $p \leq 0.40$

について、時点間の相関の期待値 ρ を

0.05, 0.10, 0.20, ..., 0.80, 0.85, 0.90, 0.95, 0.98

のように、0.10 刻みに加え、0.0 に近い 0.05 を加え、1.0 に近い 0.85, 0.95, 0.98 を加えた。なお、前節の α エラーの場合も同じ条件を設定した。

それぞれの ρ ごとに、後値 Y 、前後差 D 、前値 X を共変量とした有意差検定の両側確率で $p < 0.05$ となる回数をカウントする。さらに、前値 X の検定結果に基づいて 3 区分について検出力 $(1-\beta)$ を計算する。表 9.33 に示したのは、 $\rho = 0.50$ の場合について、全体で 10,000 回の繰り返し検定を行った結果である。1 : ($\bar{X}_{A_1} > \bar{X}_{A_2}$) となったのは 2,024 回、2 : ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) は 6,055 回、3 : ($\bar{X}_{A_1} < \bar{X}_{A_2}$) は 1,921 回で、計 10,000 回となり、おおよそ (2 割, 6 割, 2 割) の頻度となっており、前値 X の検定結果による区分分けが適切に行なわれたことが確認される。

表 9.33 $\rho = 0.50$ の場合の検出力 (表 9.34 の $\rho = 0.50$ 行を抜粋)

相関 ρ	前値 X 比較	回数 n	後値 Y		前後差 D		共分散 Cov	
			n_Y	$(1-\beta)_Y$	n_D	$(1-\beta)_D$	n_{Cov}	$(1-\beta)_{Cov}$
0.50	1: $\bar{X}_{A_1} > \bar{X}_{A_2}$	2,024	1,409	69.6	355	17.5	939	46.4
0.50	2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$	6,055	2,471	40.8	2,492	41.2	3,170	52.4
0.50	3: $\bar{X}_{A_1} < \bar{X}_{A_2}$	1,921	380	19.8	1,377	71.7	953	49.6
	全体	10,000	4,260	42.60	4,224	42.24	5,062	50.62

全体 10,000 回での検出力 ($1-\beta$) では、後値 Y では 42.60%、前後差 D で 42.24% とほぼ同等であるが、共変量 Cov では、50.62% と検出力が高くなっている。Frison・Pocock (1992) で指摘されている分散の期待値の違いが検出力に反映されている。

ただし、前値 X の平均値の 2 群間の大小関係を加味した場合に驚くべき検出力の差が明らかにされている。すなわち、後値 Y の場合であれば、

$$[1 : (\bar{X}_{A_1} > \bar{X}_{A_2}) \text{ vs. } 3 : (\bar{X}_{A_1} > \bar{X}_{A_2})] : [69.6\% \text{ vs. } 19.8\%]$$

であり、前後差 D の場合であれば、

$$[1 : (\bar{X}_{A_1} > \bar{X}_{A_2}) \text{ vs. } 3 : (\bar{X}_{A_1} > \bar{X}_{A_2})] : [17.5\% \text{ vs. } 71.7\%]$$

のような、後値 Y とは逆の検出力の差となっている。2 : ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) の場合は、全体とほぼ同じ検出力となるが、残りの 4 割については、このような驚くべき検出力の差となることは、まったく予想もできなかった。

表 9.34 時点間の相関を変化させた場合の区分別の検出力 ($1-\beta$) の比較

時点間 相関	1: $\bar{X}_{A_1} > \bar{X}_{A_2}$			2: $\bar{X}_{A_1} \approx \bar{X}_{A_2}$			3: $\bar{X}_{A_1} < \bar{X}_{A_2}$					
	Y	D	Cov	Y	D	Cov	Y	D	Cov			
ρ	n_1	$(1-\beta)_{1Y}$	$(1-\beta)_{1D}$	$(1-\beta)_{1Cov}$	n_2	$(1-\beta)_{2Y}$	$(1-\beta)_{2D}$	$(1-\beta)_{2Cov}$	n_3	$(1-\beta)_{3Y}$	$(1-\beta)_{3D}$	$(1-\beta)_{3Cov}$
0.05	2,044	43.8	3.3	37.5	5,928	43.1	19.8	42.2	2,028	38.0	60.3	37.3
0.10	1,912	46.0	4.6	39.0	6,123	42.2	21.5	42.0	1,965	36.7	62.0	37.4
0.20	1,969	53.4	5.6	41.5	6,028	42.0	24.8	42.8	2,003	33.0	65.0	39.9
0.30	2,034	59.1	7.7	41.4	6,004	41.9	27.8	44.4	1,962	29.4	64.6	42.7
0.40	1,949	65.4	12.1	43.8	6,023	41.6	34.5	48.4	2,028	23.4	66.7	44.2
0.50	2,024	69.6	17.5	46.4	6,055	40.8	41.2	52.4	1,921	19.8	71.7	49.6
0.60	1,916	74.6	27.5	53.5	6,074	41.6	50.7	59.4	2,010	14.5	74.8	55.6
0.70	2,001	82.1	42.3	62.6	6,008	39.2	61.7	67.5	1,991	8.4	80.3	61.7
0.80	1,991	89.4	68.1	78.1	5,969	38.2	80.0	82.9	2,040	5.4	89.1	76.0
0.85	2,005	92.3	81.8	86.7	5,982	38.0	90.3	91.1	2,013	3.3	95.2	88.3
0.90	1,957	96.9	96.5	97.0	6,083	36.9	97.6	97.8	1,960	1.9	98.8	95.8
0.95	2,032	98.5	100.0	99.9	6,037	37.4	100.0	100.0	1,931	1.3	100.0	99.9
0.98	1,922	99.7	100.0	100.0	6,042	37.5	100.0	100.0	2,036	0.5	100.0	100.0

図 9.18 は、表 9.34 に示した時点間の相関を変化させた場合の区分別の検出力 ($1-\beta$) の変化を Excel の散布図を用いて描いたものである。表 9.34 で示した $\rho=0.50$ の場合の区分別の検出力の比較は、3 区分別の図の X 軸の中央部の点線上での比較を行ったものである。1 : ($\bar{X}_{A_1} > \bar{X}_{A_2}$) の場合に前後差 D の検出力 ($1-\beta$) の低下は、 $\rho=0.80$ から際立ち始める。逆に 3 : ($\bar{X}_{A_1} < \bar{X}_{A_2}$) の場合には、前後差 D の検出力 ($1-\beta$) の高さは共分散分を凌駕している。このように、前値 X の 2 群間の平均値の大小関係により、検出力 ($1-\beta$) が劇的に揺れ動くことから統計的方法として、後値 Y および前後差 D による群間比較を全く推奨できないことの証拠である。

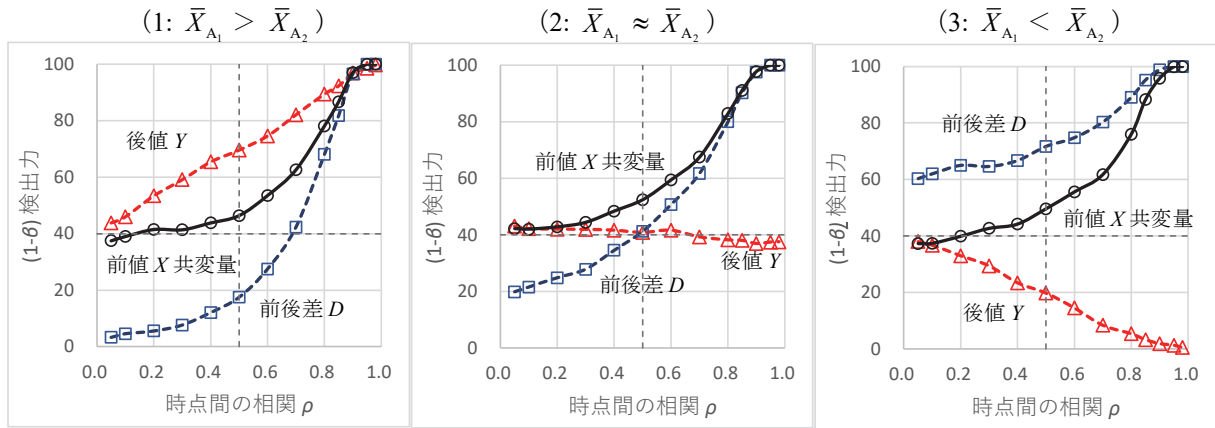


図 9.18 時点間の相関を変化させた場合の区別の検出力 $(1-\beta)$ の変化

これまでは、前値 X を共変量とした場合については、その性能について論評することができなかったのであるが、図 9.18 の各区分の時点間の相関の関わらずほぼ一定の検出力であることは、平均への回帰の影響をほとんど受けない優れた方法であることが判明した。図 9.19 に前値 X を共変量とした場合の区分ごとの検出力 $(1-\beta)$ について重ね合わせ図を示す。どの区分でも時点間の相関 ρ が 1.0 に近くなるにつれ検出力 $(1-\beta)$ は、100%に収束し、 ρ が 0.0 に近づくにつれ検出力 $(1-\beta)$ は、40%に収束している。

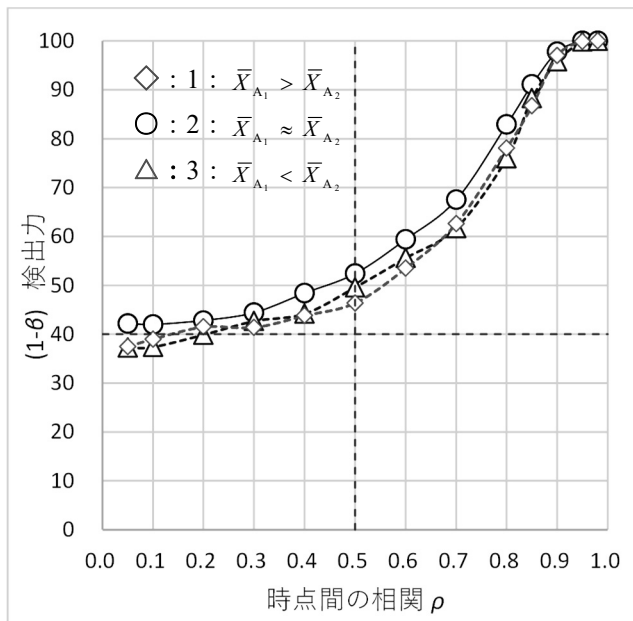


図 9.19 前値 X を共変量とした検出力 $(1-\beta)$ の比較

2: ($\bar{X}_{A_1} \approx \bar{X}_{A_2}$) 区分を基準とした場合に検出力 $(1-\beta)$ は、1: ($\bar{X}_{A_1} > \bar{X}_{A_2}$) および 3: ($\bar{X}_{A_1} < \bar{X}_{A_2}$) でおおよそ 5%程度減少する程度で、後値 Y および前後差 D のように検出力 $(1-\beta)$ が飛び跳ねるようなことがなく安定している。この結果は、前値 X を共変量とした後値 Y および前後差 D に対する共分散分析でも全く同じ結果になることは、第 9.4 節で示した。

10. 層別因子を含むロジスティック曲線のあてはめ

ロジスティック曲線のあてはめは、反応が 2 値の場合の用量反応関係に対して定式化され、複数の説明変数に対応するロジスティック回帰分析が、2 値反応に対する標準的な解析方法として広く使われている。本章では、2 値反応ではなく、反応が量的変数の場合のロジスティック曲線のあてはめを扱う。ロジスティック分布の (0 から 1) の範囲を (最小値から最大値) に拡張する。層別因子を含む回帰分析を行なった際に、交互作用があり平行な直線のあてはめがためられるような場合が少なからずある。このような場合に、ロジスティック曲線をあてはめることにより、より現実の問題に即した対応が可能となる場合がある。ロジスティック曲線のあてはめは、1 つの変数に対し 1 つのパラメータが付随するような線形モデルに該当しないので、デザイン行列 X を用いた線形モデルによるパラメータの推定が行えない。デザイン行列 X の代わりに、ロジスティック曲線式をパラメータで偏微分して得られる微係数行列 Z をデザイン行列 X と同様に扱う解析方法を示す。

10.1. カドミウムガス曝露後の肺活量の減少

アーミテージら著、椿・椿訳 (2001)、「医学研究のための統計的方法、第 9.4 節 群内の回帰」には、カドミウム産業で働いている 84 人の男性を、(A_1 : 曝露されていない群, A_2 : 少なくとも 10 年はカドミウムガスに曝露されている群, A_3 : 10 年はカドミウムガスに曝露されている群) の 3 群に分け、年齢および肺活量 (リットル) のデータが示されている。ただし、 A_3 群の男性は、 A_1 群や A_2 群に属する男性よりも年齢が高くなっている。肺活量は、加齢と共に衰えることが知られており、群間の平均年齢の違いを考慮した比較が必要となる。

アーミテージら (2001) には、「第 9.4 節 群内の回帰」に引き続き「第 9.5 節 共分散分析」でこの事例が取り上げられている。「共分散分析」は、共変量と層別因子に交互作用がない場合に適用でき、交互作用がある場合については、回帰直線の傾きの差の検討が必要となると「第 9.4 節 群内の回帰」で強調されている。原著での計算方法は、第 6 章で示した単回帰分析を重層的に使う伝統的な解析方法が示されているので、Excel の回帰分析を用いた線形モデルによる計算法により、回帰直線の傾きの差の検討を行い、ロジスティック曲線のあてはめの必要性について導入する。

表 10.1 に示すのは、群ごとの年齢 x と肺活量 y についての A_1 群（非曝露）44 人、 A_2 群（曝露 < 10 年）28 人、 A_3 群（曝露 > 10 年）12 人のデータである。 A_3 群の平均年齢は、49.75 歳と他の群より 10 歳ほど高くなっている。

表 10.1 カドミウムガス曝露による肺活量の減少（ x ：年齢， y ：肺活量リットル）

A ₁ 群 非曝露						A ₂ 群 曝露 < 10年				A ₃ 群 曝露 > 10年	
x	y	x	y	x	y	x	y	x	y	x	y
27	5.29	35	4.09	48	3.86	29	5.21	43	4.61	39	4.62
25	3.67	35	4.24	47	4.68	29	5.17	39	4.73	40	5.29
24	5.82	41	3.88	53	4.74	33	4.88	38	4.58	41	5.52
32	4.77	38	4.85	49	3.76	32	4.50	42	5.12	41	3.71
23	5.71	41	4.79	54	3.98	31	4.47	43	3.89	45	4.02
25	4.47	36	4.36	48	5.00	29	5.12	43	4.62	49	5.09
32	4.55	36	4.02	49	3.31	29	4.51	37	4.30	52	2.70
18	4.61	41	3.77	47	3.11	30	4.85	50	2.70	47	4.31
19	5.86	41	4.22	52	4.76	21	5.22	50	3.50	61	2.70
26	5.20	37	4.94	58	3.95	28	4.62	45	5.06	65	3.03
33	4.44	42	4.04	62	4.60	23	5.07	48	4.06	58	2.73
27	5.52	39	4.51	65	4.83	35	3.64	51	4.51	59	3.67
33	4.97	41	4.06	62	3.18	38	3.64	46	4.66		
25	4.99	43	4.02	59	3.03	38	5.09	58	2.88		
42	4.89	41	4.99								
				$n =$	44			$n =$	28	$n =$	12
			年齢	x 平均 =	39.80			x 平均 =	37.79	x 平均 =	49.75
			肺活量	y 平均 =	4.46			y 平均 =	4.47	y 平均 =	3.95
アーミテージら(2001), 表9.4 より引用, ただし, 群のコードは付け替えている。											

デザイン行列 X を用いた 3 本の回帰直線の同時あてはめ

図 10.1 に示すように、Excel で作成した層別散布図に対し、「近似曲線の追加」で群ごとに「線形近似」を用いて回帰直線を重ね書きしている。回帰直線は、それぞれの群の年齢の範囲（ A_1 ：18～65 歳， A_2 ：21～58 歳， A_3 ：39～65 歳）を反映している。 A_3 群の傾きが A_1 群および A_2 群に比べて大きいようなので、有意な差があるのか検討したい。そのために、 A_1 群を基準とする（0，1）型ダミー変数（ a_2 ， a_3 ）を用いた線形モデル

$$y_i = \theta_0 x_{0,i} + \theta_1 a_{2,i} + \theta_2 a_{3,i} + \theta_3 x_i + \theta_4 a_{2,i} x_i + \theta_5 a_{3,i} x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (10.1)$$

を適用する。このモデルから得られるパラメータの推定値 $\hat{\theta}_4$ が A_1 群と A_2 群の傾きの差の推定値となり、 $\hat{\theta}_5$ が A_1 群と A_3 群の傾きの差の推定値となる。なお、 $\hat{\theta}_0$ は A_1 群の切片、 $\hat{\theta}_1$ は A_1 群と A_2 群の切片の差、 $\hat{\theta}_2$ は A_1 群と A_3 群の切片の差、 $\hat{\theta}_3$ は A_1 群の傾きの推定値となる。

表 10.2 に示すのは、線形モデル式 (10.1) に対応する（84 行×5 列）のデザイン行列 X を用いた回帰分析の結果である。 A_1 群の傾きは、 $\hat{\beta}_1^{(A_1)} = \hat{\theta}_3 = -0.0306$ 、 A_1 群と A_2 群の傾きの差は

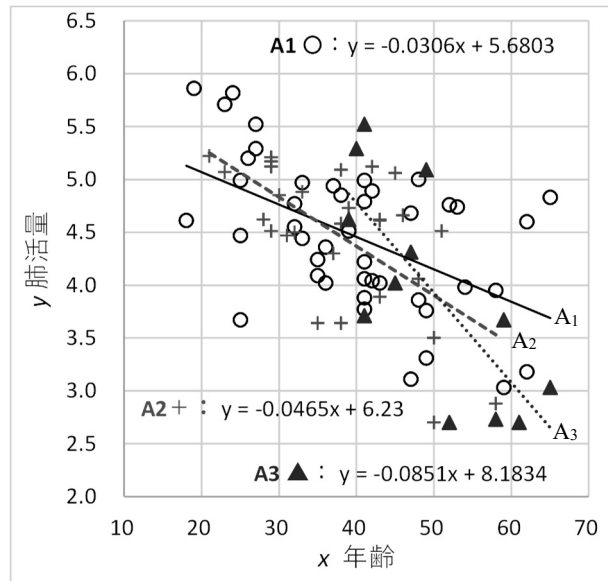


図 10.1 Excel の層別散布図に 3 本の回帰直線の重ね書き

$\hat{\beta}_1^{(A_2-A_1)} = \hat{\theta}_4 = -0.0159$ なので, A_2 群の傾きは,

$$\left. \begin{aligned} \hat{\beta}_1^{(A_2)} &= \hat{\theta}_3 + \hat{\theta}_4 \\ &= -0.0306 + (-0.0159) = -0.0465 \end{aligned} \right\} \quad (10.2)$$

であり, A_3 群の傾きは,

$$\left. \begin{aligned} \hat{\beta}_1^{(A_3)} &= \hat{\theta}_3 + \hat{\theta}_5 \\ &= -0.0306 + (-0.0545) = -0.0851 \end{aligned} \right\} \quad (10.3)$$

表 10.2 Excel の回帰分析による 3 本の回帰直線の同時あてはめ

No	A	デザイン行列 X						y	分散分析表 (定数に 0 を使用) off						
		x_0	a_2	a_3	x	a_2x	a_3x		自由度	変動	分散	分散比			
1	A ₁	1	0	0	27	0	0	5.29	回帰	5	20.1057	4.0211	11.3909		
2	A ₁	1	0	0	25	0	0	3.67	残差	78	27.5352	0.3530	$=\sigma^2$		
3	A ₁	1	0	0	24	0	0	5.82	合計	83	47.6410				
:															
43	A ₁	1	0	0	62	0	0	3.18							
44	A ₁	1	0	0	59	0	0	3.03	θ^0 切片 x_0	5.6803	0.3134	18.1232	0.0000		
45	A ₂	1	1	0	29	29	0	5.21	θ^1 a_2	0.5497	0.5759	0.9546	0.3427		
46	A ₂	1	1	0	29	29	0	5.17	θ^2 a_3	2.5031	1.0418	2.4026	0.0187		
47	A ₂	1	1	0	33	33	0	4.88	θ^3 x	-0.0306	0.0075	-4.0560	0.0001		
:									θ^4 a_2x	-0.0159	0.0145	-1.0943	0.2772		
71	A ₂	1	1	0	46	46	0	4.66	θ^5 a_3x	-0.0545	0.0211	-2.5866	0.0116		
72	A ₂	1	1	0	58	58	0	2.88							
73	A ₃	1	0	1	39	0	39	4.62	パラメータの共分散行列 $\Sigma(\theta^{\wedge}) = (X^T X)^{-1} \sigma^{\wedge 2}$						
74	A ₃	1	0	1	40	0	40	5.29	θ^0	0.0982	-0.0982	-0.0982	-0.0023	0.0023	0.0023
75	A ₃	1	0	1	41	0	41	5.52	θ^1	-0.0982	0.3316	0.0982	0.0023	-0.0081	-0.0023
:									θ^2	-0.0982	0.0982	1.0854	0.0023	-0.0023	-0.0215
83	A ₃	1	0	1	58	0	58	2.73	θ^3	-0.0023	0.0023	0.0023	0.0001	-0.0001	-0.0001
84	A ₃	1	0	1	59	0	59	3.67	θ^4	0.0023	-0.0081	-0.0023	-0.0001	0.0002	0.0001
									θ^5	0.0023	-0.0023	-0.0215	-0.0001	0.0001	0.0004
									切片 x_0	a_2	a_3	x	a_2x	a_3x	

と推定される．これらは，図 10.1 に示した Excel の近似直線で求めた回帰直線の傾きに一致する．

3 本の回帰直線の傾きの差に対する t 検定

表 10.2 右上段に示した Excel の回帰分析の結果から， A_1 群に対する A_2 群の傾きの差は， $\hat{\theta}_4 = \hat{\beta}_1^{(A_2)} - \hat{\beta}_1^{(A_1)} = -0.0159$ であり， t 検定の結果は $p_4 = 0.2772$ と有意な差ではない． A_1 群に対する A_3 群の傾きの差は， $\hat{\theta}_5 = \hat{\beta}_1^{(A_3)} - \hat{\beta}_1^{(A_1)} = -0.0545$ であり， t 検定の結果は $p_5 = 0.0116^*$ と有意な差であることが示されている． A_3 群と A_2 群の傾きの差は，

$$\begin{aligned} \hat{\beta}_1^{(A_3)} - \hat{\beta}_1^{(A_2)} &= (\hat{\theta}_3 + \hat{\theta}_5) - (\hat{\theta}_3 + \hat{\theta}_4) \\ &= \hat{\theta}_5 - \hat{\theta}_4 \\ &= -0.0545 - (-0.0159) = -0.0386 \end{aligned} \tag{10.4}$$

として求められる．その分散 $Var(\hat{\theta}_5 - \hat{\theta}_4)$ は，表 10.2 右下段のパラメータの共分散行列 $\Sigma(\hat{\theta})$ を抜粋し有効数字を増やした結果を用いて合成分散の一般式により

	$\Sigma(\hat{\theta})$ の抜粋	
$\hat{\theta}_4$	0.0002116	0.0000570
$\hat{\theta}_5$	0.0000570	0.0004439
	a_{2x}	a_{3x}

$$\begin{aligned} Var(\hat{\theta}_5 - \hat{\theta}_4) &= Var(\hat{\theta}_5) - 2Cov(\hat{\theta}_5, \hat{\theta}_4) + Var(\hat{\theta}_4) \\ &= 0.0004439 - 2 \times 0.0000570 + 0.0002116 \\ &= 0.00054162 \end{aligned} \tag{10.5}$$

が得られる． A_3 群と A_2 群の傾きの差の t 値は，

$$\begin{aligned} t_{(A_3-A_2)} &= \frac{\hat{\theta}_5 - \hat{\theta}_4}{\sqrt{Var(\hat{\theta}_5 - \hat{\theta}_4)}} \\ &= \frac{-0.0386}{\sqrt{0.00054162}} = -1.6577 \end{aligned} \tag{10.6}$$

となり，有意な差ではないことは自明である．なお，このような計算を行わなくとも， A_2 群を基準とした $(0, 1)$ 型ダミー変数 (a_1, a_3) を用いれば，Excel の回帰分析で A_3 群と A_2 群の傾きの差を直接推定することができる．

残差平方和の差から交互作用の平方和の算出

表 10.2 右上段に示した分散分析表には，3 本の直線の傾きの差の全体に関する交互作用項が含まれていないので，主効果モデルと交互作用モデルの残差平方和の差から交互作用の平方和を算出する．表 10.3 左に示した主効果モデルは，交互作用を含まない分散分析表であり，表 10.3 中に示したのは，交互作用モデルの分散分析表である．この 2 つの分散分析表の残差平方和の差分を用いて，表 10.3 右に示すように交互作用 $A \times x$ を分離する．主効果モデルの残

差の平方和（変動） $S_e^{(主)} = 30.0347$ と交互作用モデルの残差の平方和 $S_e^{(交)} = 27.5352$ の差から交互作用 $A \times x$ の平方和（変動） $S_{A \times x}$

$$\left. \begin{aligned} S_{A \times x} &= S_e^{(主)} - S_e^{(交)} \\ &= 30.0347 - 27.5352 = 2.4995 \end{aligned} \right\} \quad (10.7)$$

を求めることができる。

表 10.3 残差平方和(変動)の差による交互作用 $A \times x$ の分離

主効果モデル			交互作用モデル			交互作用項 分離					
	自由度	変動		自由度	変動		自由度	変動	分散	F 値	p 値
回帰	3	17.6063	回帰	5	20.1057	(A+x)	3	17.6063	5.8688	16.6247	0.0000
残差	80	30.0347	残差	78	27.5352	$A \times x$	2	2.4995	1.2497	3.5402	0.0338
合計	83	47.6410	合計	83	47.6410	残差	78	27.5352	0.3530	1	
							合計	83	47.6410		

$S_{A \times x} = \text{主効果モデルの残差変動} - \text{交互作用モデルの残差変動} = 30.0347 - 27.5352 = 2.4995$

また、モデル全体の平方和 $S_{回帰}^{(交)} = 20.1057$ ， $S_{回帰}^{(主)} = 17.6063$ の差からも交互作用の平方和を

$$\left. \begin{aligned} S_{A \times x} &= S_{回帰}^{(交)} - S_{回帰}^{(主)} \\ &= 20.1057 - 17.6063 = 2.4995 \end{aligned} \right\} \quad (10.8)$$

として求めることもできる。これらを表 10.3 右の分散分析表にまとめると交互作用の F 値が 3.5402 ($p=0.0338^*$) と有意であることが確認される。なお、表 10.3 右の交互作用を分離した分散分析表では、因子 A および年齢 x については分離せずに、合算したままにしている。これらの主効果を分離することは、交互作用が有意なので、無意味であるだけでなく、判断を錯誤させる危険性がある。交互作用を含む場合の因子 A の主効果は、年齢が $x=0$ 歳、すなわち図 10.1 における A_1 群と A_2 群の切片の差 $\hat{\theta}_1$ ，および、 A_1 群と A_3 群の切片の差 $\hat{\theta}_2$ を合わせた平方和から計算されるので、結果の解釈にまったく役立たない。なお、因子 A および年齢 x の平方和の分離法については、第 3 章を参照のこと。

回帰直線の傾きの差を用いた検討

A_1 群(非曝露) と A_3 群(曝露 > 10 年) の傾きに統計的に差があることが判明したので、その差異をどのように説明したらよいのであろうか。図 10.1 の 2 本の回帰直線を見ながら考察してみよう。 A_1 群は、20 歳で肺活量が約 5 リットルあり、加齢と共に肺活量が減少し、65 歳で約 3.8 リットルと 1 リットル強の減少となる。 A_3 群は、カドミウムガスの曝露が 10 年以上の人達なので、最少年齢が 39 歳と A_1 群の 18 歳に比べ高年齢となっている。 A_3 群の 40 歳の肺活量は 5 リットル弱であり、 A_1 群の 20 歳の肺活量と同程度であるが、加齢と共に肺活量の減少は大きく、65 歳で約 2.6 リットルに減少し、 A_1 群に比べ肺活量が約 1 リットル少ない。その後、さらに加齢を重ねた場合の推定は、外挿となるので、さらに肺活量に差が広がるのかは、データがないので更なる観察が必要である。

交互作用がある場合の考察は、このような説明をするのが精いっぱいであり、傾きの差を用いた説明とはなっていない。あえて用いれば、A₁群と A₃群の回帰直線がクロスする 45 歳あたりから 65 歳における 20 歳分の傾きの差は、

$$20 \times \hat{\theta}_5 = 20 \times (-0.0545) = -1.0900 \text{ リットル}$$

と推定される。グラフから読みとり考察した同様の結果であり、代わり映えない。なお、第 1.3 節で例示したように、A₁群の回帰直線と A₃群の回帰直線について、同一年齢で差の推定および 95%信頼区間を計算することにより、統計量を用いた説明ができるようになる。

回帰直線の差の 95%信頼区間

曝露 10 年以上の A₃群と非曝露の A₁群の回帰直線間の肺活量の差の推定値について 95%信頼区間を描き、何歳から肺活量が有意な差になるかを検討する。交互作用モデルの式 (10.1) において、A₃群と A₁群の場合の回帰式を特定し、A₃群と A₁群の回帰式の差を求める。式 (10.1) のダミー変数 a_2 および a_3 に (0, 1) を反映させると、その差が、

$$\begin{array}{r} A_3 : \hat{y}^{(A_3)} = \hat{\theta}_0 + \hat{\theta}_2 + \hat{\theta}_3 x + \hat{\theta}_5 x \\ -) A_1 : \hat{y}^{(A_1)} = \hat{\theta}_0 + \hat{\theta}_3 x \\ \hline (A_3 - A_1) : \hat{d} = \hat{\theta}_2 + \hat{\theta}_5 x \end{array} \quad (10.9)$$

のように求められる。差 \hat{d} の分散 $Var(\hat{d})$ は、合成分散の一般式から

$$Var(\hat{\theta}_2 + \hat{\theta}_5 x) = Var(\hat{\theta}_2) + 2Cov(\hat{\theta}_2, \hat{\theta}_5)x + Var(\hat{\theta}_5)x^2 \quad (10.10)$$

で求められる。これらのパラメータの分散および共分散は、表 10.4 に示すように表 10.2 のパラメータの共分散行列 $\Sigma(\hat{\theta})$ から該当するセルを抜き出した 2×2 の行列となる。

表 10.4 パラメータの共分散行列 $\Sigma(\hat{\theta})$ の抜粋

	$\Sigma(\hat{\theta})$ の抜粋	
$\hat{\theta}_2$	1.0854347	-0.0215188
$\hat{\theta}_5$	-0.0215188	0.0004439
	a_3	$a_3 x$

小数点以下の桁数を増やして表示している

表 10.5 に傾きが異なる A₃群と A₁群の 2 本の回帰直線の推定値、それらの差、差の分散と SE、さらに 95%信頼区間および t 検定の結果を示す。65 歳における A₁群と A₃群の回帰直線の差の 95%信頼区間を示す。A₁群の推定値は $\hat{y}_{x=65}^{(A_1)} = 3.6905$ 、A₃群の推定値は $\hat{y}_{x=65}^{(A_3)} = 2.6512$ 、その差は $\hat{d}_{x=65}^{(A_3-A_1)} = -1.0392$ として求められている。

$$\left. \begin{array}{l} \hat{y}_{x=65}^{(A_1)} = \hat{\theta}_0 + \hat{\theta}_3 x \\ \quad = 5.6803 + (-0.0306) \times 65 = 3.6905 \\ \hat{y}_{x=65}^{(A_3)} = \hat{\theta}_0 + \hat{\theta}_2 + \hat{\theta}_3 x + \hat{\theta}_5 x \\ \quad = 5.6803 + 2.5031 + (-0.0306) \times 65 + (-0.0545) \times 65 = 2.6512 \\ \hat{d}_{x=65}^{(A_3-A_1)} = \hat{y}_{x=65}^{(A_3)} - \hat{y}_{x=65}^{(A_1)} \\ \quad = 2.6512 - 3.6905 = -1.0392 \end{array} \right\} \quad (10.11)$$

表 10.5 傾きが異なる 2 本の回帰直線の差に関する 95%信頼区間と t 検定

年齢	A_3	A_1	A_3-A_1						
x	\hat{y}_{A_3}	\hat{y}_{A_1}	\hat{d}	$Var(\hat{d})$	$SE(\hat{d})$	$L95\%$	$U95\%$	t	p
35	5.2046	4.6088	0.5957	0.1229	0.3506	-0.1023	1.2938	1.6990	0.0933
40	4.7790	4.4558	0.3232	0.0742	0.2725	-0.2192	0.8656	1.1863	0.2391
45	4.3534	4.3027	0.0507	0.0477	0.2184	-0.3842	0.4856	0.2322	0.8170
50	3.9279	4.1497	-0.2218	0.0434	0.2083	-0.6365	0.1930	-1.0646	0.2904
55	3.5023	3.9966	-0.4943	0.0613	0.2475	-0.9871	-0.0014	-1.9967	0.0493 *
60	3.0768	3.8435	-0.7668	0.1014	0.3184	-1.4006	-0.1330	-2.4085	0.0184 *
65	2.6512	3.6905	-1.0392	0.1636	0.4045	-1.8445	-0.2339	-2.5692	0.0121 *
70	2.2257	3.5374	-1.3117	0.2481	0.4981	-2.3034	-0.3201	-2.6335	0.0102 *
75	1.8001	3.3843	-1.5842	0.3548	0.5956	-2.7700	-0.3984	-2.6598	0.0095 *
# 45.93	4.2742	4.2742	0.0000	0.0452	0.2127	-0.4234	0.4234	0.0000	1.0000
# 54.95	3.5063	3.9980	-0.4917	0.0610	0.2470	-0.9835	0.0000	-1.9908	0.0500 *
# ソルバーを用いた推定値					$t_{0.05}(84-6)=$	1.9908			* $P<0.05$

差の分散は、式 (10.10) の合成分散の一般式から $Var(\hat{\theta}_2 + \hat{\theta}_5 x) = 0.1636$ と計算され、95%信頼区間は、

$$\left. \begin{aligned}
 Var(\hat{\theta}_2 + \hat{\theta}_5 x) &= Var(\hat{\theta}_2) + 2Cov(\hat{\theta}_2, \hat{\theta}_5)x + Var(\hat{\theta}_5)x^2 \\
 &= 1.0854347 + 2 \times (-0.0215186) \times 65 + 0.0004439 \times 65^2 = 0.1636 \\
 (L95\%, U95\%) &= \hat{d}_{x=65}^{(A_3-A_1)} \pm t_{0.05}(84-6) \sqrt{Var(\hat{\theta}_2 + \hat{\theta}_5 x)} \\
 &= -1.0392 \pm 1.9908 \times \sqrt{0.1636} \\
 &= (-1.8445, -0.2339)
 \end{aligned} \right\} (10.12)$$

とゼロを含まないので有意な差となっている。もちろん $t = -2.5692$ に対し $p = 0.0121$ と有意な差であることが示されている。表 10.5 の計算結果を図 10.2 に示す。

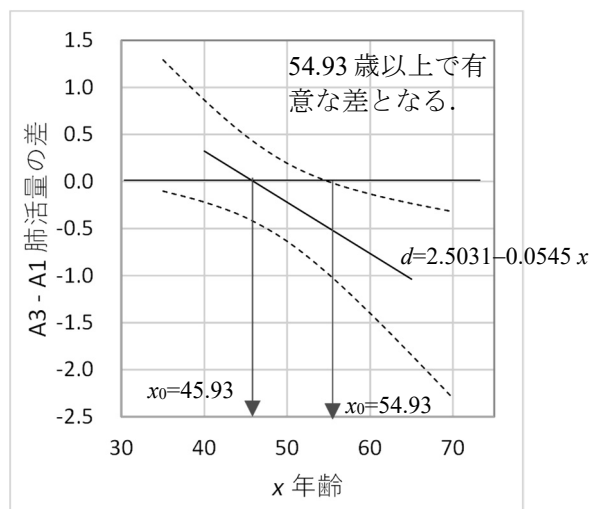


図 10.2 A_3 群と A_1 群の回帰直線の差に対する 95%信頼区間

年齢が 45 歳以下の場合、 A_3 群は A_1 群よりも推定値は大きく、55 歳以上では逆に推定値が有意に小さくなり、さらに、差が拡大することが読み取れる。年齢が 65 歳の場合 A_1 群は

3.6905 リットル, A_3 群は 2.6512 リットルと, 1.0392 リットルの差となり, $p=0.0121$ と有意な差がある. まとめると, カドミウムガスに 10 年は曝露されていた A_3 群の 40 歳前後では, 非曝露の A_1 群と同程度の肺活量であったが, それ以後に急激に開くことが示されている.

表 10.5 の下から 2 行目は, Excel のソルバーを起動し, 「目的セル」を 2 群の差「 A_3-A_1 」のセルを「目標値」として「指定値 = 0」と設定し, 「変化セル」に「年齢 x 」のセルを設定して求めた結果である. この結果から, 年齢が 45.93 歳で, A_3 群と A_1 群の肺活量が等しくなることが読み取れる. また, 回帰直線の上側 95%信頼区間が「0」となる年齢を探索することもできる. Excel のソルバーで「U95%」のセルが「0」となるように「変化セル」に「年齢 x 」のセルを設定して求めた結果が 54.95 歳であることが示されている. その場合の肺活量の差は 0.4917 リットルで, t 検定の結果は, $p=0.05$ と有意な差があることが確認される.

交互作用がある場合に, このような細かな考察を行なうことになるのだが, 表層的な考察としか言いようがない. 加齢と共に肺活量は, 減少するのであるが直線的ではなく, シグモイド曲線状になると思われるが, 得られたデータの範囲からでは確証が持てない.

スプライン曲線による直線のあてはめの妥当性チェック

図 10.3 に示すのは, JMP のグラフビルダーを用いてスプライン曲線を因子 A の 3 群に別々にあてはめた結果である. 曲線の滑らかさを調整する λ スライダーで変更し, 直線のあてはめの妥当性を視覚的に確認する. 群ごとに見た場合には, はっきりとしたシグモイド曲線とは言い難いが, 3 群全体から直線のあてはめではなくシグモイド曲線があてはめられそうである.

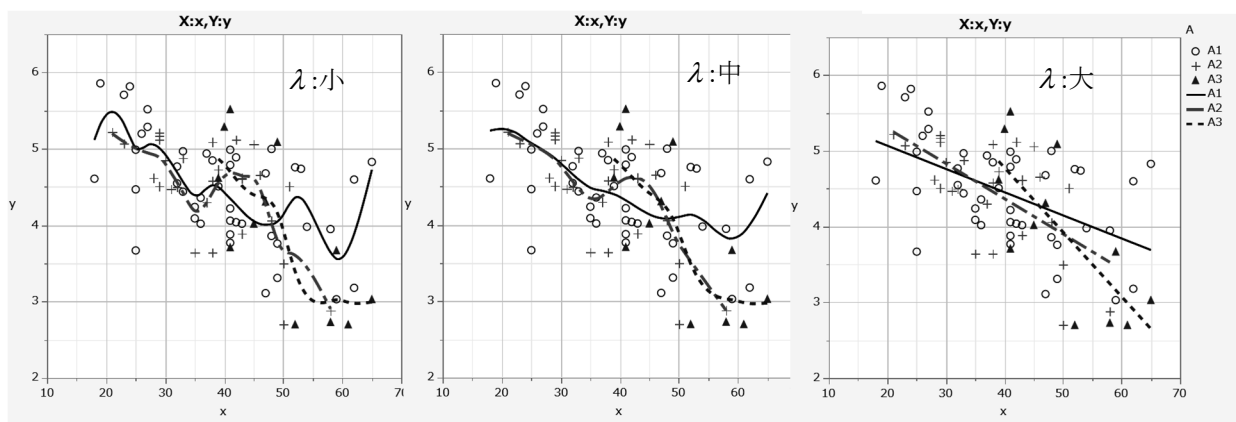


図 10.3 JMP のグラフビルダーによるスプライン曲線のあてはめ

10.2. 複数のロジスティック曲線の同時あてはめ

図 10.1 に示したように直線のあてはめは、得られたデータの範囲内での推定の問題に広く使われている。しかし、年齢と肺活量の関連は、成人となる年齢で最大化し、加齢と共に徐々に低下し、ある最小値に収束するようなシグモイド曲線をあてはめることが自然である。表 10.5 から加齢と共にカドミウムガスに 10 年以上曝露された作業員 A₃ 群の肺活量は、非曝露 A₁ 群に比べて 55 歳で 0.4943 リットル、65 歳で 1.0392 リットルと少なくなっている。外挿は問題があることは承知しているが、75 歳で 1.5 リットルの減少が予測されると言っているのだろうか。いずれにしても回帰直線をあてはめることは、便宜的な方法であると認識しなければならない。代表的なシグモイド曲線は、ロジスティック曲線であるが、他にも第 11 章で示すようにプロビット曲線、ワイブル曲線、ゴンペルツ曲線などもあるが、ここでは立ち入らない。

ロジスティック曲線のあてはめの基礎

ロジスティック曲線のあてはめは、2 値反応に対するロジスティック回帰を連想するが、量的な反応に対しても拡張できる。図 10.1 の散布図から、非曝露 A₁ 群の 18 歳～30 歳の肺活量の平均は、5.0 リットル程度で、60 歳で約 3.8 リットルである。肺活量が、年齢と共に直線的に減少するのではなく、図 10.3 に示したスプライン曲線のあてはめからも、ある最小値に漸近するロジスティック曲線となると想定される。

データ数 n_i に対して反応数 y_i の割合を p_i とした場合に、ロジスティック曲線は、 x_i に対し、

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} + \varepsilon_i, \quad \varepsilon_i \sim 2\text{項分布} \quad (10.13)$$

として定義され、最小の反応が 0.0、最大の反応が 1.0 となり、誤差に 2 項分布を仮定する。また、分母および分子を $\exp(\beta_0 + \beta_1 x_i)$ で除した式

$$p_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim 2\text{項分布} \quad (10.14)$$

も広く使われている。

式 (10.13) を $(\beta_0 + \beta_1 x_i)$ について誤差を無視して解くと、

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_i \quad (10.15)$$

いわゆるロジット $\ln[p_i / (1-p_i)]$ に関する線形式が得られる。この式だけを見るとロジットに対して回帰分析が適用できるように思われるが、式 (10.15) の誤差に正規分布を仮定する

ことができないので、いわゆる反復重み付き回帰による最尤法を適用する必要がある。詳しくは、高橋 (2017), 「一般化線形モデルを Excel で極め活用する ープロビット法・ロジット法・補 2 重対数法ー」, 高橋 (2018), 「二値データの用量反応関係における新たな解析 ープロビット・ロジット・補 2 重対数の使い分けー」を参照のこと。

式 (10.14) で $n_i=1$ とし、反応が (0, 1) ではない計量値 y_i のロジスティック曲線は、最大値を θ_0 , 最小値を θ_1 として

$$y_i = \theta_1 + \frac{\theta_0 - \theta_1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (10.16)$$

と定義される。さらに、式 (10.16) を拡張し、3 群共通の最大値を θ_0 , 各群の最小値を θ_k とするロジスティック曲線

$$y_i = \theta_k + \frac{\theta_0 - \theta_k}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (10.17)$$

として拡張できる。ただし、 $\theta_k = (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3)$, a_k は A_k 群の場合 1, それ以外は 0 とする (1, 1) 標示型ダミー変数とする。

Excel の回帰分析は、1 つの変数 (2 乗あるいは 3 乗なども 1 つの変数) に 1 つのパラメータを持つ線形和

$$y_i = \sum_{k=0}^p \theta_k x_{k,i} + \varepsilon_i \quad (10.18)$$

として定義された線形モデル式にしか適用できない。式 (10.17) のように線形モデル式とならない場合に対しては、推定したいパラメータ $\hat{\theta}^{(0)}$ におおよその初期値を設定し、推定値 $\hat{y}_i^{(0)}$ を計算する。反応 y_i と推定値 $\hat{y}_i^{(0)}$ の残差 $\hat{\varepsilon}_i^{(0)} = y_i - \hat{y}_i^{(0)}$ の平方和を $S_e^{(0)}$ とし、何らかの方法で残差平方和 $S_e^{(0)}$ を小さくするような微小なパラメータ増減値 $\hat{\delta}^{(0)}$ を推定し、元のパラメータに加え $\hat{\theta}^{(1)} = \hat{\theta}^{(0)} + \hat{\delta}^{(0)}$ とする。再度、推定値 $\hat{y}_i^{(1)}$ を計算し、反応 y_i との残差 $\hat{\varepsilon}_i^{(1)} = \hat{y}_i^{(1)} - y_i$ の残差平方和を $S_e^{(1)}$ とし、さらに、残差平方和 $S_e^{(1)}$ を小さくするような微小なパラメータ増減値 $\hat{\delta}^{(1)}$ を推定する。線形モデルでない場合には、このように逐次的に残差平方和 $S_e^{(m)}$ を減少させ、増減値 $\hat{\delta}^{(m)}$ がほとんどゼロとなるまで、パラメータを変化させる方法が非線形回帰モデルとして定式化されている。本章の末尾の「Wikipedia のガウス・ニュートン法」に示した、非線形のみカエリス・メンテン式のあてはめが、典型的な事例である。

厄介な方法であるが、多くの統計ソフトに、このような「線形モデルではない」という意味での「非線形モデル」のパラメータを推定する関数あるいはアプリケーションが組み込まれている。SAS では「NLIN プロシジャ」、JMP では「非線形回帰」、R では「nls 関数」、Excel でも「ソルバー」というアドイン・アプリケーションを使うことにより、残差平方和 $S_e^{(m)}$ を最小化するようなパラメータの推定が容易にできる。線形モデル式で表せない式 (10.17) の 6

パラメータを Excel の「ソルバー」を用いて推定することから始め、次節でロジスティック曲線のパラメータの 95%信頼区間の推定方法について導入する。なお、芳賀 (2016), 「医薬品開発のための統計解析, 第 3 部 非線形モデル 改訂版」が、非線形モデルに関する必読の書である。

ソルバーによる残差平方和の最小化

表 10.6 に示すのは、表 10.1 に示した「カドミウムガス曝露による肺活量の減少」データに対し、式 (10.17) を適用し、Excel のソルバーを用いて推定したパラメータ $\hat{\theta}$ である。この推定値を得るため、表 10.6 左上段に示す初期値を与えて Excel のソルバーで、残差平方和 S_e が最小になるように 6 個のパラメータを変化させて得られた結果である。

各群の最小値を求めるために (1, 1) 型ダミー変数を $\mathbf{a}_k = [a_1, a_2, a_3]$ とし、対応するパラメータ $\hat{\theta}_k = [\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3]^T$ との積和、

$$\hat{\theta}_k = \mathbf{a}_k \hat{\theta}_k = \text{Mmult}(\mathbf{a}_k \text{ の範囲}, \hat{\theta}_k \text{ の範囲}) \tag{10.19}$$

によって各群の $\hat{\theta}_k$ を求める。これにより、表 10.6 の $\hat{\theta}_k$ の欄に、A₁ 群の 44 個のデータに対し $\hat{\theta}_1 = 3.4947$, A₂ 群に対し $\hat{\theta}_2 = 2.9612$, A₃ 群に対し $\hat{\theta}_3 = 2.8613$ が設定される。

表 10.6 Excel ソルバーによる 3 本のロジスティック曲線の同時あてはめ

初期値		パラメータの推定値					誤差分散の推定値						
5.0000	4.0000	3.5000	3.0000	4.0000	-0.1000	変化させるセル	$\hat{\sigma}^2 = \frac{29.2457}{84-6} = 0.3749$						
最大値		θ^{\wedge}_0	4.9924										
最小値		a_1	θ^{\wedge}_1	3.4947									
		a_2	θ^{\wedge}_2	2.9621									
		a_3	θ^{\wedge}_3	2.8613									
4.0000		回帰パラメータ		$\theta^{\wedge}_4: \beta^{\wedge}_0$	6.0867								
-0.1000				$\theta^{\wedge}_5: \beta^{\wedge}_1$	-0.1273								
		残差平方和		S_e	29.2457	最小化セル		ロジスティック曲線の推定値					
No	A	変数			肺活量	最小値	推定値	$y - y^{\wedge}$	ε^{\wedge}	x	非曝露	曝露<10	曝露>10
		a_1	a_2	a_3	x	y	θ^{\wedge}_k	y^{\wedge}			$y^{\wedge(A1)}$	$y^{\wedge(A2)}$	$y^{\wedge(A3)}$
1	A ₁	1	0	0	27	5.29	3.4947	4.8936	0.3964	15	4.9698		
2	A ₁	1	0	0	25	3.67	3.4947	4.9147	-1.2447	20	4.9503	4.9353	
:										25	4.9147	4.8870	
43	A ₁	1	0	0	62	3.18	3.4947	3.7065	-0.5265	30	4.8520	4.8021	
44	A ₁	1	0	0	59	3.03	3.4947	3.7858	-0.7558	35	4.7476	4.6605	4.6440
45	A ₂	0	1	0	29	5.21	2.9621	4.8229	0.3871	40	4.5885	4.4448	4.4176
46	A ₂	0	1	0	29	5.17	2.9621	4.8229	0.3471	45	4.3769	4.1580	4.1165
:										50	4.1407	3.8378	3.7805
71	A ₂	0	1	0	46	4.66	2.9621	4.0948	0.5652	55	3.9237	3.5436	3.4717
72	A ₂	0	1	0	58	2.88	2.9621	3.3988	-0.5188	60	3.7571	3.3179	3.2347
73	A ₃	0	0	1	39	4.62	2.8613	4.4695	0.1505	65	3.6461	3.1673	3.0766
74	A ₃	0	0	1	40	5.29	2.8613	4.4176	0.8724	70	3.5788	3.0761	2.9809
:										75	3.5404	3.0241	2.9263
83	A ₃	0	0	1	58	2.73	2.8613	3.3196	-0.5896	∞	3.4947	2.9621	2.8613
84	A ₃	0	0	1	59	3.67	2.8613	3.2755	0.3945		θ^{\wedge}_1	θ^{\wedge}_2	θ^{\wedge}_3

A₁ 群の年齢 $x_1 = 27$ における推定値 \hat{y}_1 は,

$$\left. \begin{aligned} \hat{y}_1^{(A_1)} &= \hat{\theta}_1 + \frac{\hat{\theta}_0 - \hat{\theta}_1}{1 + \exp[-(\hat{\beta}_0 + \hat{\beta}_1 x_1)]} \\ &= 3.4947 + \frac{4.9924 - 3.4947}{1 + \exp[-(6.0867 + (-0.1273) \times 27)]} = 4.8936 \end{aligned} \right\} \quad (10.20)$$

として計算されている. 残差 $\hat{\varepsilon}_1$ は,

$$\left. \begin{aligned} \hat{\varepsilon}_1 &= y_1 - \hat{y}_1^{(A_1)} \\ &= 5.29 - 4.8936 = 0.3964 \end{aligned} \right\} \quad (10.21)$$

である. これらの計算を $i=1, 2, \dots, 84$ まで繰り返し, 残差平方和 S_e が,

$$\left. \begin{aligned} S_e &= \sum_i \hat{\varepsilon}_i^2 \\ &= \text{SumSq}(\hat{\varepsilon} \text{ の範囲}) = 29.2457 \end{aligned} \right\} \quad (10.22)$$

として計算されている.

表 10.6 右は, 推定されたパラメータ $\hat{\theta}$ を用いて A₁ 非曝露群では, 15 歳から 5 歳刻みで 75 歳までの推定値 $\hat{y}^{(A_1)}$ を計算し, A₂ 曝露<10 年群は 20 歳から, A₃ 曝露>10 年群では 35 歳からと各群の最小年齢を考慮して $\hat{y}^{(A_2)}$ および $\hat{y}^{(A_3)}$ を計算している. これは, 図 10.4 示す Excel の散布図上に推定曲線を重ね書きする際に, 年齢 x が小さい順となっていないと連続した曲線がきれいに描けないためである.

図 10.4 は, 反応 y_i の層別散布図に群ごとのロジスティック曲線を散布図上に重ね書きした結果である. 回帰直線をあてはめた図 10.1 で苦し紛れの解釈をしたのであるが, より現実的な解釈が可能となる. A₁ 非曝露群では, 20 歳で 4.95 リットル, 40 歳で 4.59 リットル, 65 歳で 3.65 リットルとシグモイド曲線状に減少し, 最小値 $\hat{\theta}_1 = 3.4947$ リットルに漸近する.

図 10.1 に示した A₂ 曝露<10 年群の直線回帰では, A₁ 非曝露群の直線の傾きに比べ傾きが同程度で, カドミウムガスの曝露の影響が少ないとの印象を持つ. これは, 20 歳代での曝露年数が少ない人達も数多く含まれていることも影響しているのかも知れないが, 図 10.4 のロジスティック曲線のあてはめでは, A₃ 曝露>10 年群にほぼ一致した結果になっていて, 同じデータでの分析にもかかわらず, 推定の様子が全く異なる. これは, 年齢が 40 歳ぐらいまでは, カドミウムガスの曝露による肺活量に与える影響が相対的に少ないことの表れであり, A₂ 曝露<10 年群の 50 歳以上の 4 症例の肺活量の分布は, A₃ 曝露>10 年群に類似していることも, 同様のロジスティック曲線の推定になったと思われる.

A₃ 曝露>10 年群の最少年齢は 39 歳であり, 回帰直線のあてはめでは, 肺活量が 5 リットルから 65 歳まで 10 歳あたりの肺活量の減少は, 式 (10.3) で推定された傾き $\hat{\beta}_1^{(A_3)} = -0.0851$ に

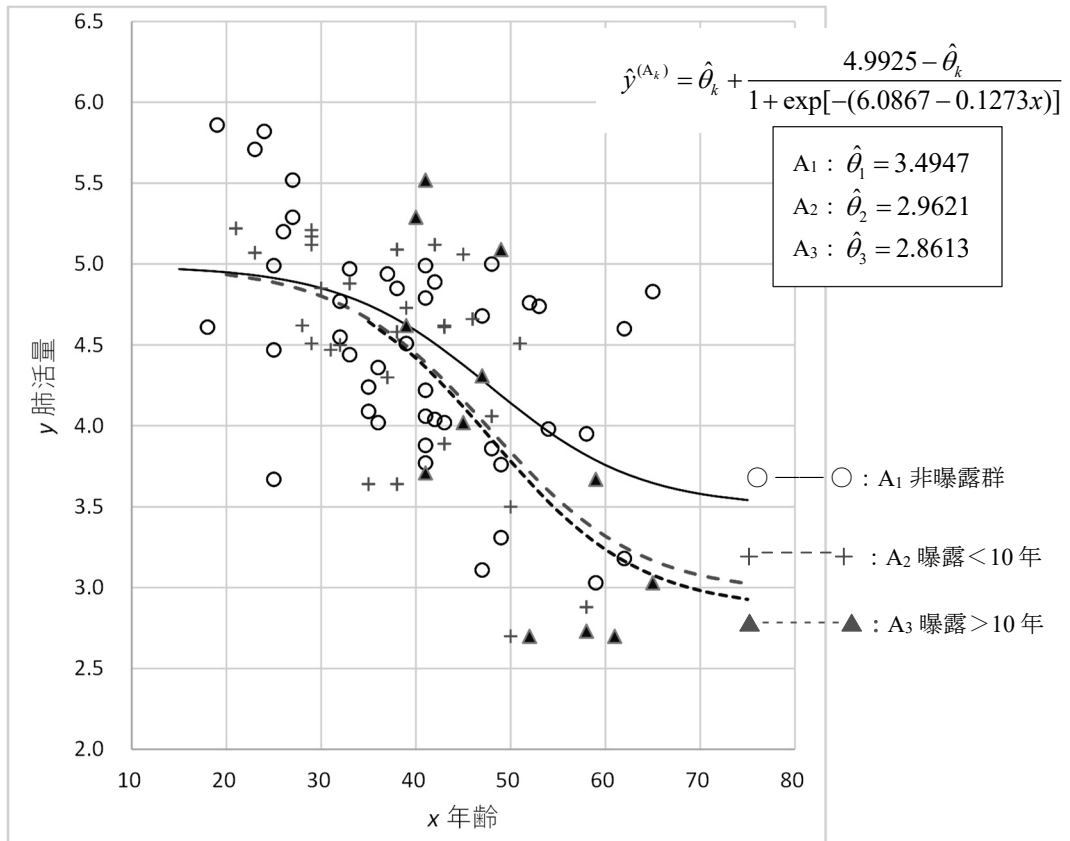


図 10.4 非曝露・曝露群に対するロジスティック曲線のあてはめ

10歳を掛けて -0.851 リットルの減少と推定されている。このこと自体は、データの存在範囲なので問題はないが、どこまで肺活量が落ちるのかについて外挿による推定は、危険である。ロジスティック曲線によるあてはめでは、パラメータとして最小値が $\hat{\theta}_3 = 2.8613$ リットルとして推定され、65歳で3.08リットル、75歳で2.93リットルと徐々に下限 $\hat{\theta}_3 = 2.8613$ リットルに漸近する。

これらの結果を総合して、カドミウムガスの曝露を受けた場合には、加齢と共に肺活量の減少が非曝露群に比べ大きくなり、

$$\left. \begin{aligned} \hat{\theta}_3 - \hat{\theta}_1 &= 2.8613 - 3.4947 = -0.6335 \\ \hat{\theta}_3 - \hat{\theta}_2 &= 2.8613 - 2.9621 = -0.5326 \end{aligned} \right\} \quad (10.23)$$

高年齢になった際に、約0.5リットル強の肺活量の減少と推測される。このような考察は、図10.1の回帰直線のあてはめではなく、元々の現象に忠実な統計モデルの適用によって可能となる。Excelのソルバーは、3本のロジスティック曲線のあてはめによるパラメータの推定が手軽に行えるのであるが、推定値の95%信頼区間を求めるために必要なパラメータの共分散行列の出力がない。そこで、第10.3節で、パラメータの共分散行列の算出方法、および、各種の95%信頼区間の推定方法について示す。

パラメータの初期値の探索的な推定

残差平方和 S_e をソルバーで最小にするため、「変化させるセル」には、何らかの「初期値」を必要とする。そのためには、図 10.1 をよく観察し、最大値 θ_0 は 5.0 ぐらいと推定し、 A_1 群の最小値 θ_1 は、4.0 ぐらい、 A_2 群は 3.5 ぐらい、 A_3 群は 3.0 ぐらい、とあたりが付けられる。

さて、 $\exp()$ 関数の中の線形式のパラメータ $\hat{\beta}_0$ とパラメータ $\hat{\beta}_1$ の初期値は、どのように設定したら良いのだろうか。最初は、 $\hat{\beta}_0 = 1$ 、 $\hat{\beta}_1 = 1$ から初めて、どちらかを前後に動かしてみる。形状パラメータ $\hat{\beta}_1$ を大きくしても $\hat{y}^{(A)}$ は、フラットのまま変化しない。 $\hat{\beta}_1 = -1$ とすると下方に直線のまま移動、 $\hat{\beta}_1 = -0.5$ にしても同じ、 $\hat{\beta}_1 = -0.1$ にするとわずかにシグモイド曲線状となる。ソルバーでパラメータを変化させると、求める解が得られる。解が得られた状態で、パラメータを前後に動かしてシグモイド曲線がどの様に変化するか観察することにより、理解が深まるのであるが、場当たりのであり推奨できない。

パラメータの初期値の設定の正攻法

初期値設定の正攻法は、各パラメータがロジスティック曲線の形状にどのように変化を与えるかを洞察することである。回帰分析における ($\hat{\beta}_0$ が切片、 $\hat{\beta}_1$ が傾き) にとられることなく、 $\exp[-(\beta_0 + \beta_1 x_i)]$ のふるまいを指数関数の引数として認識するために、次式のように μ_{LGS} を位置パラメータ、 σ_{LGS} を形状パラメータに変換することを考える。次式、

$$\left. \begin{aligned} -(\beta_0 + \beta_1 x_i) &= \frac{x_i - \mu_{LGS}}{\sigma_{LGS}} \\ -\beta_0 - \beta_1 x_i &= -\frac{\mu_{LGS}}{\sigma_{LGS}} + \frac{1}{\sigma_{LGS}} x_i \end{aligned} \right\} \quad (10.24)$$

の関係から、

$$\left. \begin{aligned} -\beta_0 &= -\frac{\mu_{LGS}}{\sigma_{LGS}}, & -\beta_1 &= \frac{1}{\sigma_{LGS}} \\ \beta_0 &= \frac{\mu_{LGS}}{\sigma_{LGS}}, & \beta_1 &= -\frac{1}{\sigma_{LGS}} \end{aligned} \right\} \quad (10.25)$$

が導き出される。ここで、 μ_{LGS} は、ロジスティック曲線の X 軸方向の中心位置のパラメータであり、 σ_{LGS} は正規分布の標準偏差と同様の形状パラメータである。

表 10.1 に示されている年齢 x のおおよその平均から $\hat{\mu}_{LGS} = 40$ とし、図 10.1 の散布図から、X 軸方向のデータの広がり約 4 分の 1 から $\hat{\sigma}_{LGS} = 10$ と見積れば、

$$\hat{\beta}_0 = \frac{\hat{\mu}_{LGS}}{\hat{\sigma}_{LGS}} = \frac{40}{10} = 4, \quad \hat{\beta}_1 = -\frac{1}{\hat{\sigma}_{LGS}} = -\frac{1}{10} = -0.10 \quad (10.26)$$

が得られる。これらの推測された初期値を設定した結果を図 10.5 に示す。まずまずのあてはめとなっている。

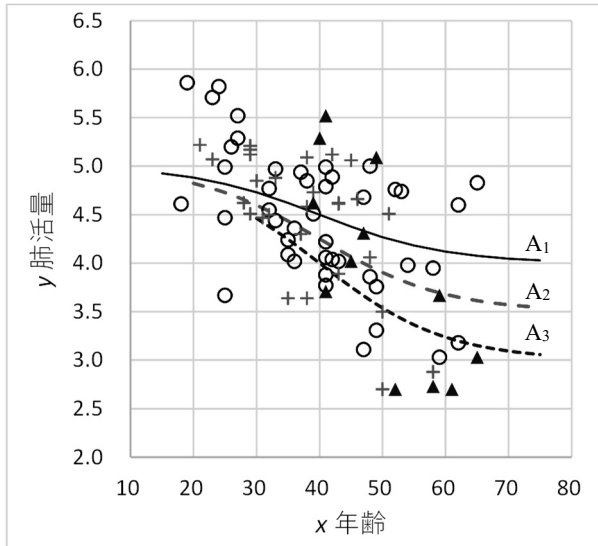


図 10.5 初期パラメータに対するロジスティック曲線

表 10.7 パラメータの初期値

		推定値	
最大値	θ_0^{\wedge}	5.0000	変化させるセル
最小値	A_1	4.0000	
	A_2	3.5000	
	A_3	3.0000	
回帰パラメータ	$\theta^{\wedge}_4 : \beta^{\wedge}_0$	4.0000	最小化セル
	$\theta^{\wedge}_5 : \beta^{\wedge}_1$	-0.1000	
残差平方和	S_e	32.8667	

ソルバーで S_e を最小にするように $\hat{\theta}$ を変化させると表 10.6 の結果が得られる

Excel のソルバーの活用法

Excel ソルバーは、各種の統計的な課題の解決に役に立つ。表 10.7 に示したのは、ある初期値に対し、取り敢えず計算された残差平方和 S_e が“最小”になるようにパラメータを変化させ、非線形モデルに対する“最小 2 乗解”を求めている。Excel の「回帰分析」を使わなくとも、残差平方和 S_e の計算式 $S_e = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x)]^2$ を Excel シート上で設定すれば、どのような初期値 ($\hat{\beta}_0, \hat{\beta}_1$) を与えても、ソルバーで最小 2 乗解を得ることができる。第 12 章の表 12.5, 表 12.7, 表 12.9 に示すのは、Excel の回帰分析の代わりにソルバーで、パラメータの推定を行っている事例である。表 12.2 では、非線形モデルの典型例として知られているミカエリス・メンテン式のパラメータ推定での事例である。

Excel ソルバーは、第 1 章でも使われている。表 1.27 では、回帰直線の 95%信頼区間が、X 軸を横切る x の推定にソルバーが用いられている。これは、 x を与えて回帰の推定値の上側 95%信頼区間の推定値が“0”となるように“ x ”を変化させた結果である。第 5 章の表 5.10 では、“0”ではなく“目標値 6”としている。第 7 章の表 7.9, 表 7.17 も同様の事例である。

10.3. ロジスティック曲線の 95%信頼区間

線形モデルに対する回帰分析に際して、各種の 95%信頼区間の推定に必要なパラメータの共分散行列 $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ を用いた活用事例を数多く示してきた。Excel の回帰分析から得られる誤差分散の推定値 $\hat{\sigma}^2$ とデザイン行列 \mathbf{X} を用いて、パラメータの共分散行列

$$\Sigma(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2 \quad (10.27)$$

を求めることができた。このような行列計算でパラメータの共分散行列を推定できるのは、式 (10.18) で示した線形モデルの場合に限定されている。式 (10.17) で示したロジスティック曲線は、線形モデルではないので、式 (10.27) に示したデザイン行列 \mathbf{X} を用いたパラメータの共分散行列 $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ を求めることは、残念ながらできない。

ロジスティック曲線式のパラメータに関する偏微分

線形モデルで表わすことができない非線形モデルの場合には、与えられたモデル式に対し、各パラメータで偏微分した微分係数を新たな変数 z_k とし、 z_k に関する微係数行列を \mathbf{Z} とする。これをデザイン行列 \mathbf{X} と同様に扱い、

$$\Sigma(\hat{\theta}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2 \quad (10.28)$$

により、パラメータの共分散行列 $\Sigma(\hat{\theta})$ が得られる。線形モデルの $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ と同様に非線形の $\Sigma(\hat{\theta})$ を、ロジスティック曲線式のパラメータの共分散行列として用いることにより、各種の 95%信頼区間の推定を行なうことができる。と言うのは簡単だが、実際に Excel で実行するためにはロジスティック曲線の式 (10.17) をパラメータで偏微分をしなければならないし、参考になる書籍も見出すことができない。最近の統計ソフトは、式を与えれば、内部で自動的に偏微分する機能が備わっており、偏微分した結果を出力することもできる。なお、JMP と SAS を用いた場合については、第 10.4 節および第 10.5 節で例示する。

ソルバーであてはめたロジスティック曲線の式 (10.17) にダミー変数を加えた式は、

$$y_i = (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3) + \frac{\theta_0 - (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3)}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i \quad (10.29)$$

である。この式に対し、全てのパラメータ ($\theta_0, \theta_1, \theta_2, \theta_3, \beta_0, \beta_1$) で偏微分する。このままでは、見通しが悪いので、

$$f(x_i) = \theta_k + (\theta_0 - \theta_k)[1 + g(x_i)]^{-1} \quad (10.30)$$

$$\text{ただし, } \theta_k = \theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3$$

$$g(x_i) = \exp[-(\beta_0 + \beta_1 x_i)]$$

とすると見通しが良くなる。

パラメータに関する偏微分式を以下に示す.

$$\left. \begin{aligned} z_1 &= \frac{\partial f(x_i)}{\partial \theta_0} = [1 + g(x_i)]^{-1} \\ z_2 &= \frac{\partial f(x_i)}{\partial \theta_1} = a_1 - a_1 [1 + g(x_i)]^{-1} \\ z_3 &= \frac{\partial f(x_i)}{\partial \theta_2} = a_2 - a_2 [1 + g(x_i)]^{-1} \\ z_4 &= \frac{\partial f(x_i)}{\partial \theta_3} = a_3 - a_3 [1 + g(x_i)]^{-1} \end{aligned} \right\} \quad (10.31)$$

$$\left. \begin{aligned} z_5 &= \frac{\partial f(x_i)}{\partial \beta_0} = (\theta_0 - \theta_k) g(x_i) [1 + g(x_i)]^{-2} \\ z_6 &= \frac{\partial f(x_i)}{\partial \beta_1} = \frac{\partial f(x_i)}{\partial \beta_0} x_i \end{aligned} \right\} \quad (10.32)$$

これらの偏微分式に対する数値計算は、表 10.8 に示すように Excel を用いればストレスなく計算でき、また計算された微係数行列 \mathbf{Z} を用いた $(\mathbf{Z}^T \mathbf{Z})^{-1}$ も容易に計算することができる.

表 10.8 に示した計算シートは、込み入っているが、列方向に表の下部に示した計算式を第 1 行目に順次入力し、第 84 行目までフィルハンドルで計算式をコピーする. そのためには、 $(\theta_1, \theta_2, \theta_3)$ などダミー変数 a_k のパラメータの参照には、セル位置が変わらないように

$$\hat{\theta}_k : = \text{Mmult} (\text{D14:F14}, \$\text{J\$6:}\text{J\$8}) \\ a_1 \sim a_3, \quad \hat{\theta}_1 \sim \hat{\theta}_3$$

のように参照するセル位置にファンクションキー「F4」により「\$」を付ける. この計算は、 $\hat{\theta}_k$ 変数とし A_1 群の場合に $\hat{\theta}_1 = 3.495$ が、 A_2 群の場合に $\hat{\theta}_2 = 2.962$ が、 A_3 群の場合に $\hat{\theta}_3 = 2.861$ が、設定される. 引き続き、 \hat{y}_1 を

$$g(x_1) = \exp[-(\hat{\beta}_0 + \hat{\beta}_1 x_1)] = 0.071,$$

$$\text{推定値 } \hat{y}_1 = \hat{\theta}_k + (\hat{\theta}_0 - \hat{\theta}_k) / [1 + g(x_1)] = 4.894,$$

として推定し、残差 $\hat{\varepsilon}_1 = y_1 - \hat{y}_1 = 0.396$ を計算する. 式 (10.31) および式 (10.32) の偏微分式は、Excel の表示上の制約から「 ∂_{θ_0} 」のように接頭語「 ∂_{θ} 」にパラメータ名をつけて偏微分した結果の変数名としている.

微係数行列 \mathbf{Z} は、サイズが (84 行 \times 6 列) であり、(6 \times 6) の積和行列の逆行列 $(\mathbf{Z}^T \mathbf{Z})^{-1}$ は、Excel の行列関数を用いて次式

$$(\mathbf{Z}^T \mathbf{Z})^{-1} : = \text{Minverse} (\text{Mmult} (\text{Transpose} (\mathbf{Z} \text{ の範囲}), \mathbf{Z} \text{ の範囲}))$$

で計算する.

表 10.8 Excel による偏微分式の計算および積和行列の逆数

		θ^{\wedge}		推定値			$(Z^T Z)^{-1}$									
最大値		θ^{\wedge}_0		4.9925	変化 させる セル		0.202	-0.278	-0.319	-0.322	-2.461	0.055				
最小値		A_1	θ^{\wedge}_1	3.4947			-0.278	1.055	1.307	1.166	5.057	-0.131				
		A_2	θ^{\wedge}_2	2.9621			-0.319	1.307	2.237	1.658	6.277	-0.170				
		A_3	θ^{\wedge}_3	2.8613			-0.322	1.166	1.658	1.726	6.333	-0.165				
回帰パラメータ		$\theta^{\wedge}_4; \beta^{\wedge}_0$		6.0862			-2.461	5.057	6.277	6.333	39.989	-0.936				
		$\theta^{\wedge}_5; \beta^{\wedge}_1$		-0.1273		0.055	-0.131	-0.170	-0.165	-0.936	0.023					
残差平方和		S_e		29.2457	最小化セル	$\sigma^{\wedge 2} = \mathbf{0.3749} = S_e / (84 - 6)$										
群	ダミー	年齢	肺活量	中間	推定値	残差	微係数行列 Z									
No	A	a_1	a_2	a_3	x_i	y_i	θ_k^{\wedge}	$g(x_i)$	y_i^{\wedge}	ε_i^{\wedge}	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \beta_0$	$\partial \beta_1$
1	A_1	1	0	0	27	5.29	3.495	0.071	4.894	0.396	0.934	0.066	0.000	0.000	0.092	2.492
2	A_1	1	0	0	25	3.67	3.495	0.055	4.915	-1.245	0.948	0.052	0.000	0.000	0.074	1.843
:																
44	A_1	1	0	0	59	3.03	3.495	4.145	3.786	-0.756	0.194	0.806	0.000	0.000	0.235	13.838
45	A_2	0	1	0	29	5.21	2.962	0.091	4.823	0.387	0.917	0.000	0.083	0.000	0.155	4.506
46	A_2	0	1	0	29	5.17	2.962	0.091	4.823	0.347	0.917	0.000	0.083	0.000	0.155	4.506
:																
72	A_2	0	1	0	58	2.88	2.962	3.649	3.399	-0.519	0.215	0.000	0.785	0.000	0.343	19.880
73	A_3	0	0	1	39	4.62	2.861	0.325	4.469	0.151	0.755	0.000	0.000	0.245	0.395	15.392
74	A_3	0	0	1	40	5.29	2.861	0.369	4.418	0.872	0.730	0.000	0.000	0.270	0.420	16.792
:																
84	A_3	0	0	1	59	3.67	2.861	4.145	3.276	0.394	0.194	0.000	0.000	0.806	0.334	19.690
$\theta_k^{\wedge} = \text{Mmult}(\mathbf{a}$ の範囲, θ^{\wedge} の範囲)											$\partial \theta_0 = 1 / (1 + g(x_i))$		$\partial \theta_3 = a_3 - a_2 / (1 + g(x_i))$			
$g(x_i) = \exp(-(\beta_0^{\wedge} + \beta_1^{\wedge} x_i))$											$\partial \theta_1 = a_1 - a_1 / (1 + g(x_i))$		$\partial \beta_0 = (\theta_0^{\wedge} - \theta_k^{\wedge}) g(x_i) / (1 + g(x_i))^2$			
$y_i^{\wedge} = \theta_k^{\wedge} + (\theta_0^{\wedge} - \theta_k^{\wedge}) / (1 + g(x_i))$											$\partial \theta_2 = a_2 - a_2 / (1 + g(x_i))$		$\partial \beta_1 = (\partial \beta_0) * x_i$			

パラメータに関する 95%信頼区間

表 10.9 に示すようにロジスティック曲線の 95%信頼区間の推定のためのパラメータの共分散行列 $\Sigma(\hat{\theta})$ は、微係数行列 Z を用いた $(Z^T Z)^{-1}$ に誤差分散の推定値 $\hat{\sigma}^2$ を掛けて得られる。表 10.8 に示した残差 $\hat{\varepsilon}_i$ の平方和 $S_e = 29.2457$ を自由度 $84 - 6 = 78$ で割ると誤差分散の推定値 $\hat{\sigma}^2 = 0.3749$ が得られる。また、 $\Sigma(\hat{\theta})$ の対角セルが、パラメータの分散の推定値となるので平方根を取り SE が得られ、各パラメータの 95%信頼区間を求めることができる。

表 10.9 パラメータの共分散行列 $\Sigma(\hat{\theta})$ および 95%信頼区間

θ	推定値	$\Sigma(\hat{\theta}) = (Z^T Z)^{-1} \sigma^{\wedge 2}$						SE	$t_{0.05} \times SE$	$L95\%$	$U95\%$
θ^{\wedge}_0	4.9925	0.0758	-0.1042	-0.1196	-0.1209	-0.9228	0.0204	0.2753	0.5480	4.4444	5.5405
θ^{\wedge}_1	3.4947	-0.1042	0.3957	0.4899	0.4373	1.8961	-0.0491	0.6290	1.2523	2.2424	4.7470
θ^{\wedge}_2	2.9621	-0.1196	0.4899	0.8387	0.6215	2.3535	-0.0636	0.9158	1.8232	1.1389	4.7853
θ^{\wedge}_3	2.8613	-0.1209	0.4373	0.6215	0.6472	2.3746	-0.0619	0.8045	1.6016	1.2597	4.4628
$\theta^{\wedge}_4; \beta^{\wedge}_0$	6.0862	-0.9228	1.8961	2.3535	2.3746	14.9937	-0.3511	3.8722	7.7089	-1.6227	13.7951
$\theta^{\wedge}_5; \beta^{\wedge}_1$	-0.1273	0.0204	-0.0491	-0.0636	-0.0619	-0.3511	0.0084	0.0919	0.1830	-0.3102	0.0557
		$\sigma^{\wedge 2} = 0.3749$						$t_{0.05}(84-6) = 1.9908$			

図 10.6 左にシグモイド曲線の共通の最大値および各群の最小値の推定値と 95%信頼区間を示す。最大値は、各群で共通であり 4.99 (4.44, 5.54) リットルと推定されている。各群の最小値の推定値の 95%信頼区間は、極めて広く A₁ 非曝露群と A₃ 曝露>10 年群の間に全く差が見いだされない。これは、高年齢層のデータが少なく、最大 65 歳までなので、最小値は外挿された結果となり 95%信頼区間が大きく広がっているためである。図 10.6 右には、A₁ 非曝露群と A₂ 群および A₃ 群の差についての 95%信頼区間を示している。

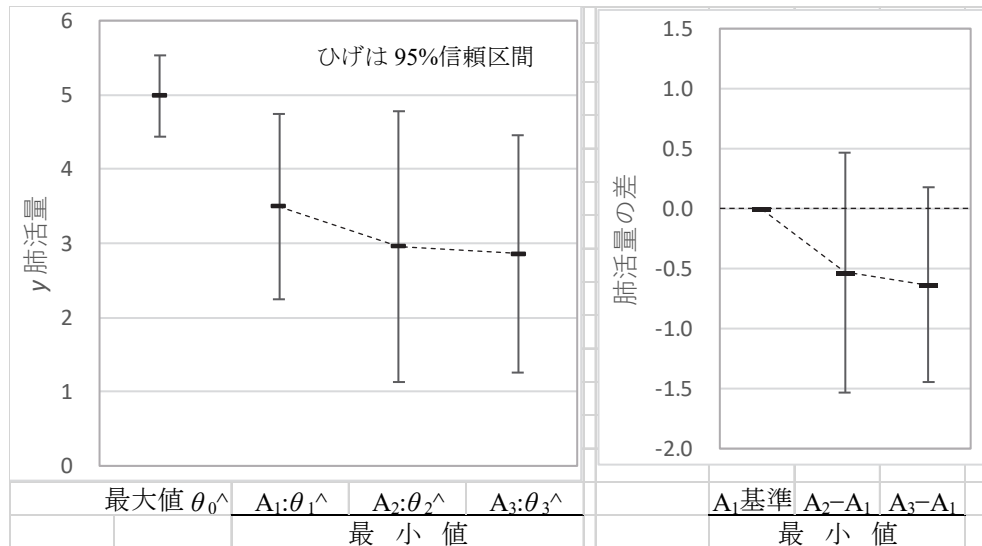


図 10.6 シグモイド曲線の共通の最大値および各層の最小値

パラメータ間の差の推定値は、簡単に求められるが、分散は、表 10.9 に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ を用いて合成分散の一般式

$$\begin{aligned} \text{Var}(\hat{\theta}_2 - \hat{\theta}_1) &= \text{Var}(\hat{\theta}_2) - 2\text{Cov}(\hat{\theta}_2, \hat{\theta}_1) + \text{Var}(\hat{\theta}_1) \\ &= 0.8387 - 2 \times 0.4899 + 0.3957 = 0.2548 \end{aligned} \tag{10.33}$$

$$\text{Var}(\hat{\theta}_3 - \hat{\theta}_1) = \text{Var}(\hat{\theta}_3) - 2\text{Cov}(\hat{\theta}_3, \hat{\theta}_1) + \text{Var}(\hat{\theta}_1) = 0.1681 \tag{10.34}$$

によって求める。これらの推定された分散を用いて、表 10.10 に示すように 95%信頼区間を計算し、この結果を図 10.6 右のひげとして示した。

表 10.10 パラメータの差の 95%信頼区間

θ 間の差	推定値	分散	SE	$t_{0.05} \times SE$	L95%	U95%
$\theta_1^{\wedge} - \theta_1^{\wedge}$	0	0	0	0.0000	0.0000	0.0000
$\theta_2^{\wedge} - \theta_1^{\wedge}$	-0.5326	0.2545	0.5045	1.0043	-1.5369	0.4718
$\theta_3^{\wedge} - \theta_1^{\wedge}$	-0.6334	0.1681	0.4100	0.8163	-1.4497	0.1829

ロジスティック曲線の最小値の群間差の 95%信頼区間は、群ごとの分散に対し相対的に小さくなり、A₁ 非曝露群に対し有意な差とは言えないものの、さらなるデータの収集により、有意な差となることが期待される。

ロジスティック曲線の 95%信頼区間

回帰直線における 95%信頼区間の表示と同様にロジスティック曲線の 95%信頼区間に挑戦する。図 10.4 に示した 3 本のロジスティック曲線は、式 (10.30) に表 10.6 で推定されたパラメータを代入すると次式となり、

$$y^{(A_k)} = \hat{\theta}_k + \frac{4.9925 - \hat{\theta}_k}{1 + \exp[-(6.0867 - 0.1273x)]} \tag{10.35}$$

$$A_1 : \hat{\theta}_1 = 3.4947, \quad A_2 : \hat{\theta}_2 = 2.9621, \quad A_3 : \hat{\theta}_3 = 2.8613$$

年齢 x を適当な範囲と刻みで設定し肺活量 \hat{y} を推定し、散布図上に重ね書きして描く。

推定値 \hat{y} の分散 $Var(\hat{y})$ は、表 10.8 で示した微係数行列 \mathbf{Z} から求められたパラメータの共分散行列 $\mathbf{\Sigma}(\hat{\theta})$ に関する \hat{y} についての微係数ベクトル \mathbf{z} による 2 次形式

$$Var(\hat{y}) = \mathbf{z} \mathbf{\Sigma}(\hat{\theta}) \mathbf{z}^T \tag{10.36}$$

によって求めることができる。

表 10.11 に A_1 群に対するロジスティック曲線の 95%信頼区間を推定するために、年齢 x を 10 歳から 5 歳刻みで 80 歳まで変化させた場合の推定値 \hat{y} と分散 $Var(\hat{y})$ を式 (10.36) によって

表 10.11 各種のロジスティック曲線の 95%信頼区間

群	—変数—				推定値			z_0	z_1	z_2	z_3	z_4	z_5	$Var(\hat{y}^{\wedge})$	SE	L95%	U95%	
	a_1	a_2	a_3	x	θ_k	$g(x)$	y^{\wedge}											
A ₁	1	0	0	10	3.49	0.01	4.98	0.99	0.01	0.00	0.00	0.01	0.12	0.06	0.24	4.50	5.46	
	1	0	0	15	3.49	0.02	4.97	0.98	0.02	0.00	0.00	0.02	0.33	0.05	0.22	4.54	5.40	
				:														
	1	0	0	80	3.49	60.0	3.52	0.02	0.98	0.00	0.00	0.02	1.93	0.29	0.54	2.45	4.59	
A ₂	0	1	0	10	2.96	0.01	4.98	0.99	0.00	0.01	0.00	0.02	0.16	0.05	0.23	4.52	5.43	
	0	1	0	15	2.96	0.02	4.96	0.98	0.00	0.02	0.00	0.03	0.45	0.04	0.20	4.56	5.36	
				:														
	0	1	0	80	2.96	60.0	3.00	0.02	0.00	0.98	0.00	0.03	2.62	0.65	0.80	1.40	4.60	
A ₃	0	0	1	10	2.86	0.01	4.98	0.99	0.00	0.00	0.01	0.02	0.17	0.05	0.23	4.52	5.43	
	0	0	1	15	2.86	0.02	4.96	0.98	0.00	0.00	0.02	0.03	0.48	0.04	0.20	4.57	5.36	
				:														
	0	0	1	80	2.86	60.0	2.90	0.02	0.00	0.00	0.98	0.03	2.75	0.46	0.68	1.54	4.25	
A ₂	-1	1	0	0	-0.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.01	
-A ₁	-1	1	0	10	-0.53	0.01	0.00	0.00	0.0	0.01	0.00	0.00	0.04	0.00	0.01	-0.03	0.02	
				:														
	-1	1	0	100	-0.53	765	-0.53	0.00	-1.0	1.00	0.00	0.00	0.07	0.25	0.50	-1.53	0.47	
A ₃	-1	0	1	0	-0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.01	
	-A ₁	-1	0	1	10	-0.63	0.01	-0.01	0.00	0.0	0.00	0.01	0.01	0.05	0.00	0.02	-0.04	0.03
					:													
	-1	0	1	100	-0.63	765	-0.63	0.00	-1.0	0.00	1.00	0.00	0.08	0.17	0.41	-1.44	0.18	

計算し、95%信頼区間を求めた結果である。計算には、表 10.9 に示したパラメータの推定値および共分散行列を用い、表 10.8 の下段に示した計算式を用いている。

表 10.11 で計算されたロジスティック曲線の推定値 \hat{y} 、および、 $L95\%$ 、 $U95\%$ を A_1 非曝露群の散布図上に重ね書きした結果を図 10.7 左に示す。年齢を 10 歳に外挿した場合の 95%信頼区間は (4.50, 5.46) であり、図 10.6 の最大値の 95%信頼区間に漸近し、年齢を 80 歳に外挿した場合の (2.445, 4.593) は、最小値の A_1 群に漸近する。高年齢への外挿に対して、95%信頼区間の増大が著しいことがわかる。同様に、 A_2 曝露<10年群および A_3 曝露>10年群についても 95%信頼区間を計算しそれぞれの散布図上に重ね書きした結果を示す。

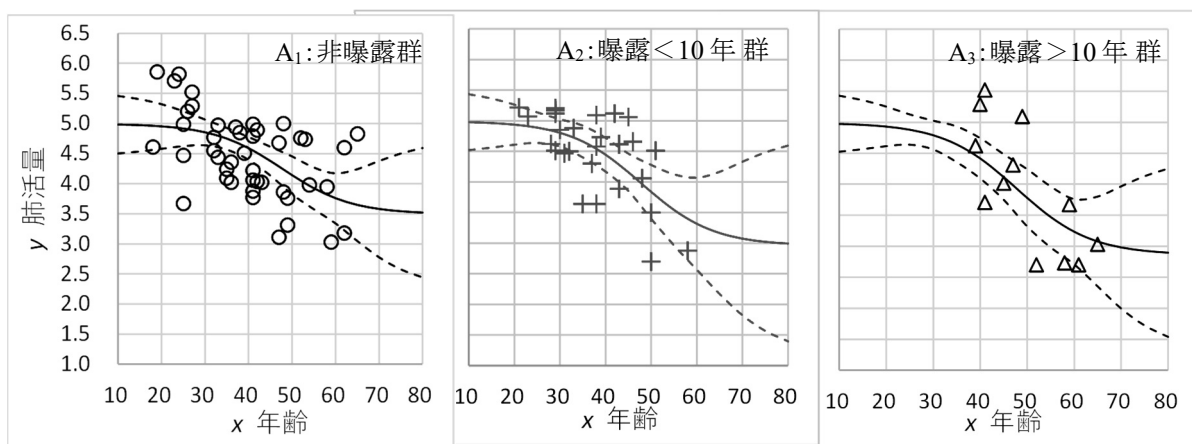


図 10.7 ロジスティック曲線 95%信頼区間

ロジスティック曲線の差の 95%信頼区間

A_1 非曝露群のロジスティック曲線は、表 10.8 から $\hat{\theta}_1 = 3.4947$ に収束し、その分散は、表 10.9 に示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素から $Var(\hat{\theta}_1) = 0.3957$ であり、95%信頼区間は (2.2423, 4.7470) となる。 A_3 群の場合は、 $\hat{\theta}_3 = 2.8612$ (1.2597, 4.4628) であり、 A_1 非曝露群と A_3 曝露>10年群のロジスティック曲線の収束値の差は、表 10.10 から

$$\hat{\theta}_3 - \hat{\theta}_1 = 2.8612 - 3.4947 = -0.6334$$

である。その分散は、

$$Var(\hat{\theta}_3 - \hat{\theta}_1) = Var(\hat{\theta}_3) + 2Cov(\hat{\theta}_3, \hat{\theta}_1) + Var(\hat{\theta}_1) = 0.1681$$

なので、95%信頼区間は (-1.4497, 0.1829) と 0 を超え、有意な差とは言えない。

A_1 群と A_2 群および A_3 群のロジスティック曲線の差は、表 10.11 下段に示すように、(1, 1) 標示型ダミー変数の差として求められる。 A_2 群の場合は、 $(a_1 \ a_2 \ a_3) = (0 \ 1 \ 0)$ で、 A_1

群の場合は、 $(a_1 \ a_2 \ a_3) = (1 \ 0 \ 0)$ なので、それらの差 $(A_2 - A_1)$ は、 $(a_1 \ a_2 \ a_3) = (-1 \ 1 \ 0)$ となり、

$$y^{(A_2 - A_1)} = \hat{\theta}_2 + \frac{4.9925 - \hat{\theta}_2}{1 + \exp[-(6.0867 - 0.1273x)]} - \hat{\theta}_1 - \frac{4.9925 - \hat{\theta}_1}{1 + \exp[-(6.0867 - 0.1273x)]} \quad (10.37)$$

$$= (\hat{\theta}_2 - \hat{\theta}_1) + \frac{0.0000 - (\hat{\theta}_2 - \hat{\theta}_1)}{1 + \exp[-(6.0867 - 0.1273x)]}$$

$$A_1 : \hat{\theta}_1 = 3.4947, \quad A_2 : \hat{\theta}_2 = 2.9621, \quad A_3 : \hat{\theta}_3 = 2.8613$$

が Excel シート上で計算されている。このロジスティック曲線は、最大値が 0.0000、最小値が $(\hat{\theta}_2 - \hat{\theta}_1) = -0.5326$ のロジスティック曲線となることが理解される。A₃群と A₁群の場合は、最大値は同様に 0.0000、最小値は、 $(\hat{\theta}_3 - \hat{\theta}_1) = -0.6334$ となる。

表 10.11 下段に年齢 x に 0 歳から 10 歳ごとに 100 歳までの差の推定値と 95%信頼区間が計算されているので、図 10.8 に重ね書きした結果を示す。これらから、カドミウムガスの曝露により、肺活量は加齢と共に約 0.6 リットル減少すると推定されるが、高齢者のデータが少ないために有意な差とは言い切れない。図 10.2 の直線回帰の差からは、55 歳以上で有意な肺活量の減少となっていたのと対照的である。

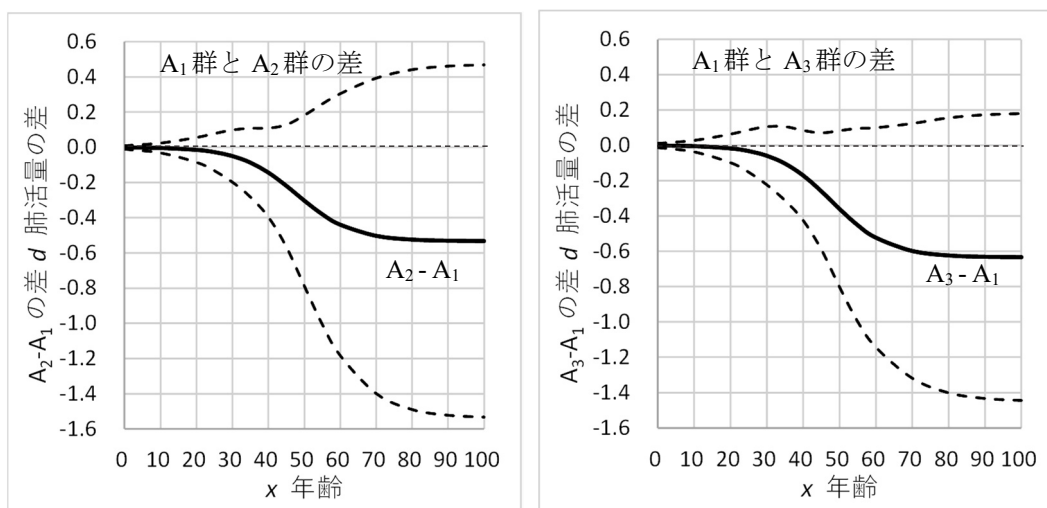


図 10.8 群間の差に対するロジスティック曲線のおよび 95%信頼区間

表 10.11 に示した各群の最大値は、 $\hat{\theta}_0 = 4.9925$ を用い、 $\hat{\theta}_0$ による偏微分式は、式 (10.31) の最初の行 $\partial f(x_i) / \partial \theta_0 = [1 + g(x_i)]^{-1}$ を用いている。ところが、群間差 $(A_2 - A_1)$ および $(A_3 - A_1)$ の場合の最大値は、 $\hat{\theta}_0 = 0.0$ なので式 (10.31) に代えて $\partial f(x_i) / \partial \theta_0 = 0.0$ としている。

10.4. JMPによる3本のロジスティック曲線の同時あてはめ

JMPによるロジスティック曲線のあてはめ

JMPによる3本のロジスティック曲線のあてはめは、「非線形回帰」を用いる。表10.12に示すJMPデータセットの形式は、表10.6に示したExcelのファイルの形式と同様である。JMPの場合も表10.8に示したようにパラメータに対し初期値を設定し、あらかじめ肺活量 y_i に対する予測値 \hat{y}_i をJMPファイル上に設定する必要がある。予測値 \hat{y}_i の設定は、計算式エディタのパラメータの新規作成によって設定し、それらを引用しつつ予測値 \hat{y}_i の計算式をGUIにより設定する。パラメータは太字で、変数はイタリックで区別して表示されている。JMPの非線形回帰の詳しい使い方については、芳賀(2016)を参照のこと。

表 10.12 JMPデータファイルおよび予測値 \hat{y} の計算式の設定

	No	A	a1	a2	a3	x	y	初期値 \hat{y}
○	1	A1	1	0	0	27	5.29	4.7858
○	2	A1	1	0	0	25	3.67	4.8176
:								
○	44	A1	1	0	0	59	3.03	4.1301
+	45	A2	0	1	0	29	5.21	4.6254
+	46	A2	0	1	0	29	5.17	4.6254
:								
+	72	A2	0	1	0	58	2.88	3.7128
▲	73	A3	0	0	1	39	4.62	4.0500
▲	74	A3	0	0	1	40	5.29	4.0000
:								
▲	83	A3	0	0	1	58	2.73	3.2837
▲	84	A3	0	0	1	59	3.67	3.2602

初期値 \hat{y} :

パラメータ

パラメータの新規作成...

$\theta_0 = 5$

$\theta_1 = 4$

$\theta_2 = 3.5$

$\theta_3 = 3$

$\beta_0 = 4$

$\beta_1 = -0.1$

$$\theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3 + \frac{\left(\theta_0 - \left(\theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3 \right) \right)}{\left(1 + \text{Exp} \left(- \left(\beta_0 + \beta_1 \cdot x \right) \right) \right)}$$

JMP による非線形回帰

図 10.9 に示すように JMP の「非線形回帰」を起動し、「Y, 応答変数」に「y」を、「X, 予測式列」に「 \hat{y} 」をセットし、実行する。デフォルトの反復回数（60 回）に達しても収束しないので、「実行」を 3 回繰り返す。

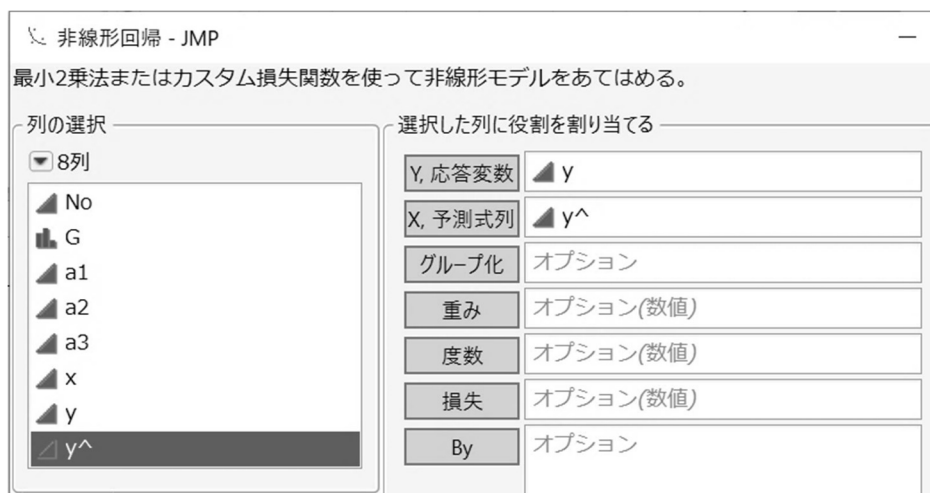


図 10.9 JMP の「非線形回帰」の起動

表 10.13 に JMP の非線形回帰で収束した結果を示す。パラメータの推定値は、表 10.6 に示した Excel のソルバーで得られた結果に少数以下 3 桁まで一致し、標準誤差 SE が出力されている。JMP の非線形回帰で出力された SE は、表 10.9 に示した微分係数行列 Z を用いて導出されたパラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根を取ることにより求められた SE に少数以下 3 桁まで一致する。少数以下 4 桁まで一致しないのは、表 10.14 に示すようにパラメータ間の相関が 1.0 に近い $Corr(\hat{\beta}_0, \hat{\beta}_1) = -0.9866$ などがあり、残差平方和 S_e を最小化す

表 10.13 JMP の非線形回帰による 3 本のロジスティック回帰のパラメータ推定

解					
	SSE	DFE	MSE	RMSE	R2乗
	29.2457	78	0.3749	0.6123	0.3861
パラメータ	推定値	近似標準誤差	推定値	SE	
θ_0	4.9925	0.2753	4.9925	0.2753	
θ_1	3.4947	0.6291	3.4947	0.6290	
θ_2	2.9621	0.9159	2.9621	0.9158	
θ_3	2.8612	0.8045	2.8613	0.8045	
β_0	6.0859	3.8721	6.0862	3.8722	
β_1	-0.1272	0.0919	-0.1273	0.0919	
解法: 解析 Gauss-Newton					表 10.9 抜粋

るパラメータがなかなか確定できないためであること、また、統計ソフトによるパラメータ探索の打ち切り条件が完全に一致しないためでもある。

表 10.14 パラメータ（推定値）の相関行列

推定値の相関						
	θ_0	θ_1	θ_2	θ_3	β_0	β_1
θ_0	1.0000	-0.6018	-0.4743	-0.5458	-0.8657	0.8082
θ_1	-0.6018	1.0000	0.8505	0.8643	0.7785	-0.8487
θ_2	-0.4743	0.8505	1.0000	0.8437	0.6637	-0.7559
θ_3	-0.5458	0.8643	0.8437	1.0000	0.7623	-0.8376
β_0	-0.8657	0.7785	0.6637	0.7623	1.0000	-0.9866
β_1	0.8082	-0.8487	-0.7559	-0.8376	-0.9866	1.0000

表 10.15 に示すように、表 10.9 で示したパラメータの共分散行列 $\Sigma(\hat{\theta})$ も JMP のバージョン 16（2021 年リリース）から出力されるようになったので、これらの結果を Excel に取り込むことにより、各種の推定が容易に行えるようになったことは喜ばしい。表 10.14 に示したパラメータの相関行列は、多重共線性の確認のために必要である。しかし、パラメータの共分散行列は、内部の計算のためには大活躍しているが、それ自体を見ても役立つ情報に欠けているために、出力されてこなかったと思われる。

表 10.15 パラメータの共分散行列 $\Sigma(\hat{\theta})$

推定値の共分散						
	θ_0	θ_1	θ_2	θ_3	β_0	β_1
θ_0	0.0758	-0.1042	-0.1196	-0.1209	-0.9228	0.0204
θ_1	-0.1042	0.3957	0.4900	0.4374	1.8962	-0.0491
θ_2	-0.1196	0.4900	0.8388	0.6216	2.3537	-0.0636
θ_3	-0.1209	0.4374	0.6216	0.6473	2.3748	-0.0619
β_0	-0.9228	1.8962	2.3537	2.3748	14.9930	-0.3511
β_1	0.0204	-0.0491	-0.0636	-0.0619	-0.3511	0.0084

JMP に限ったわけではないが、現代の統計ソフトでの「非線形回帰」は、厄介なパラメータに関する偏微分式の設定が、内部に組み込まれた自動微分の機能により、必要としなくなっている。JMP では、「予測プロファイル」により、図 10.10 に示すようにシグモイド曲線の 95%信頼区間の表示も手軽に行える。予測プロファイルには、任意の年齢に対する推定値および 95%信頼区間の表示ができるようになっている。A₁ 群の 50 歳での予測値は、($a_1=1$, $a_2=0$, $a_3=0$, $x=50$) と設定することにより、図 10.10 左に予測値が 4.14 と表示され、さらに

95%信頼区間が (3.83, 4.45) と示されている。この JMP の「予測プロファイル」で作成したロジスティック曲線の 95%信頼区間は、Excel で作成した図 10.7 左に対応する。

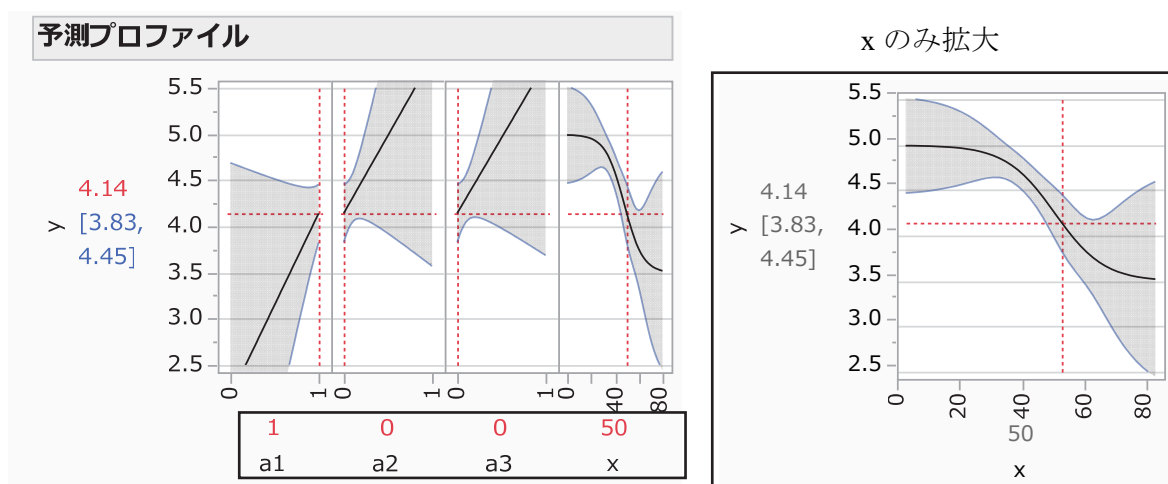


図 10.10 A_1 群のシグモイド曲線と 95%信頼区間の表示

ロジスティック曲線の 95%信頼区間

「非線形回帰」がサポートしているグラフ表示は、他にも沢山あるが、図 10.4 および図 10.7 で示したような元のデータ y と予測値 \hat{y} を重ね書きする機能はないので、予測値 \hat{y} 、 SE 、および、95%信頼区間を元の JMP データファイルに追加して書き出し、「グラフビルダー」などで別途作成しても良いし、Excel に予測値 \hat{y} を取り出して、編集機能が豊富な Excel の散布図を用いることもできる。図 10.11 に示すように予測値 \hat{y} などの JMP ファイルへの出力に際し、予測式（計算式）付きを選択することにより、内部の計算式を JMP ファイルで可視化できる。これにより、JMP 内部での計算方法を知ることができる。

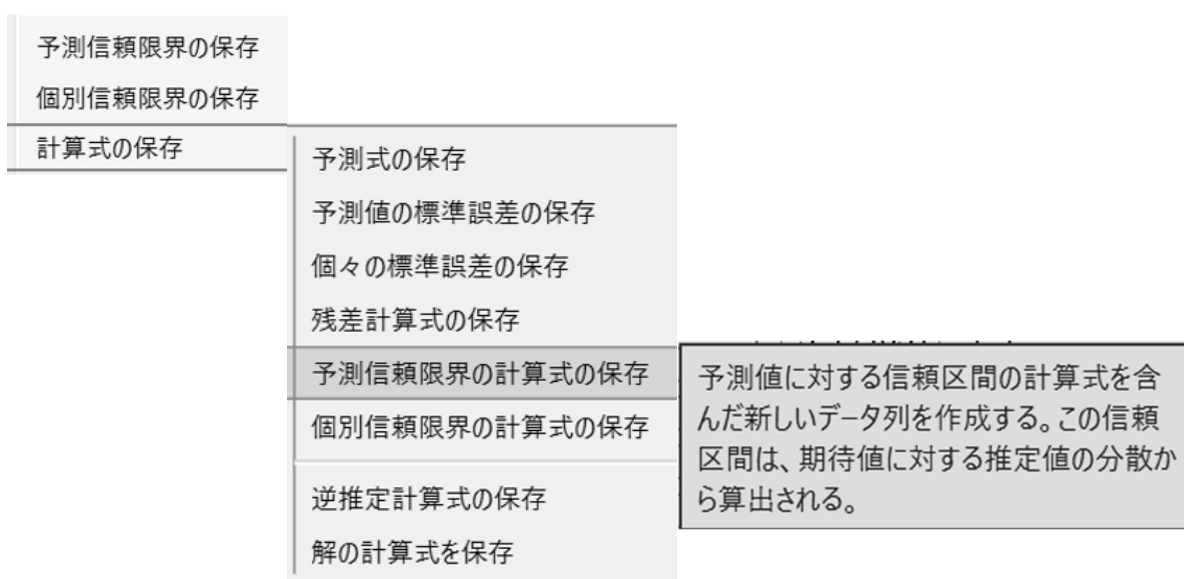


図 10.11 JMP の計算式を含む予測値の 95%信頼区間の保存

表 10.16 JMP データファイルに出力された予測値, SE, 95%信頼区間

	No	A	a1	a2	a3	x	y	初期値 y [^]	予測値 y [^]	予測値の標 準誤差 y [^]	下限M y [^]	上限M y [^]
○	1	1 A1	1	0	0	27	5.29	4.7858	4.8937	0.1293	4.6363	5.1510
○	2	2 A1	1	0	0	25	3.67	4.8176	4.9147	0.1459	4.6243	5.2052
:												
○	44	44 A1	1	0	0	59	3.03	4.1301	3.7858	0.1970	3.3937	4.1779
+	45	45 A2	0	1	0	29	5.21	4.6254	4.8230	0.1110	4.6020	5.0439
+	46	46 A2	0	1	0	29	5.17	4.6254	4.8230	0.1110	4.6020	5.0439
:												
+	72	72 A2	0	1	0	58	2.88	3.7128	3.3988	0.3385	2.7250	4.0727
▲	73	73 A3	0	0	1	39	4.62	4.0500	4.4695	0.1609	4.1491	4.7898
▲	74	74 A3	0	0	1	40	5.29	4.0000	4.4176	0.1631	4.0930	4.7423
:												
▲	83	83 A3	0	0	1	58	2.73	3.2837	3.3197	0.2488	2.8243	3.8150
▲	84	84 A3	0	0	1	59	3.67	3.2602	3.2755	0.2543	2.7693	3.7818

新たな列として出力された「予測値 y[^]」について、計算式エディタを起動して内容を確認する。表 10.16 に示されている数値の裏には、図 10.12 に示すような計算式が埋め込まれていて、最初のデータ (a₁=1, a₂=0, a₃=0, x=27) を式に代入した結果とし、予測値 $\hat{y}_1 = 4.8937$ が表示されている。

a1 • 3.4946507199
+ a2 • 2.9620974482
+ a3 • 2.8612283464

$$+ \frac{\left(4.9924878393 - \left(a1 \cdot 3.4946507199 + a2 \cdot 2.9620974482 + a3 \cdot 2.8612283464 \right) \right)}{\left(1 + \text{Exp} \left(- \left(6.0859239471 + -0.127248413 \cdot x \right) \right) \right)}$$

$$\hat{y}_i = (\hat{\theta}_1 a_{1,i} + \hat{\theta}_2 a_{2,i} + \hat{\theta}_3 a_{3,i}) + \frac{\hat{\theta}_0 - (\hat{\theta}_1 a_{1,i} + \hat{\theta}_2 a_{2,i} + \hat{\theta}_3 a_{3,i})}{1 + \exp[-(\hat{\beta}_0 + \hat{\beta}_1 x_i)]}$$

図 10.12 「予測値 y[^]」に埋め込まれている計算式

「予測値の標準誤差 y[^]」に含まれている計算式を確認した結果を図 10.13 に示す。ページ全体をおおう目も眩むような計算式となっている。左端に√を縦に長く引き伸ばした表示があるので、予測値 y[^] の分散の平方根を取り、SE の計算がされている。「Vec Quadratic ([6 × 6], Matrix(...))」となっているのは、以下の 2 次形式による計算を、パラメータの共分散行列 $\Sigma(\hat{\theta})$ を用い、予測値 y[^] についてのパラメータによる偏微分式に推定されたパラ

<p>Vec Quadratic [6 × 6], Matrix</p>	<p>2 次形式の計算の関数 [6×6] は, 表 10.15 に示した パラメータの共分散行列</p>	$\frac{\partial f(x_i)}{\partial \theta_0} = [1 + g(x_i)]^{-1}$ $\frac{\partial f(x_i)}{\partial \theta_1} = a_{1,i} - a_{1,i}[1 + g(x_i)]^{-1}$ $\frac{\partial f(x_i)}{\partial \theta_2} = a_{i,2} - a_{i,2}[1 + g(x_i)]^{-1}$ $\frac{\partial f(x_i)}{\partial \theta_3} = a_{i,3} - a_{i,3}[1 + g(x_i)]^{-1}$ $\theta_k = \theta_1 a_{1,i} + \theta_2 a_{2,i} + \theta_3 a_{3,i}$ $g(x_i) = \exp[-(\beta_0 + \beta_0 x_i)]$
$\begin{pmatrix} 1 \\ \left(1 + \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right) \\ -a1 \\ a1 + \left(1 + \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right) \\ -a2 \\ a2 + \left(1 + \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right) \\ -a3 \\ a3 + \left(1 + \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right) \\ -1 \\ \cdot \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \\ - \left(4.992487734 - \left(a1 \cdot 3.4946511645 + a2 \cdot 2.9620980914 + a3 \cdot 2.8612289403 \right) \right) \\ \cdot \left(1 + \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right)^{-2} \\ \left(\text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right) \\ \cdot -x \\ - \left(4.992487734 - \left(a1 \cdot 3.4946511645 + a2 \cdot 2.9620980914 + a3 \cdot 2.8612289403 \right) \right) \\ \cdot \left(1 + \text{Exp} \left(- \left(6.0859263112 + -0.127248475 \cdot x \right) \right) \right)^{-2} \end{pmatrix}$		
$\frac{\partial f(x_i)}{\partial \beta_0} = (\theta_0 - \theta_k) g(x_i) [1 + g(x_i)]^{-2}$ $\frac{\partial f(x_i)}{\partial \beta_1} = \frac{\partial f(x_i)}{\partial \beta_0} x_i$		

図 10.13 「予測値の標準誤差 y^」に埋め込まれている計算式

メータ代入した計算式が示されている．実際には内部での計算なので，表 10.17 に示す Excel での計算結果と対比されたい．

表 10.17 JMP データファイルに出力された予測値，SE，95%信頼区間（表 10.8 抜粋）

A ₁	—変数—				推定値		z ₀ z ₁ z ₂ z ₃ z ₄ z ₅					Var(y [^])	SE	L95%	U95%		
	No	a ₁	a ₂	a ₃	x	θ _k	g(x)	y [^]	∂ θ ₀	∂ θ ₁	∂ θ ₂					∂ θ ₃	∂ θ ₄
1	1	0	0	27	3.49	0.07	4.89	0.93	0.07	0.00	0.00	0.09	2.49	0.02	0.13	4.64	5.15
θ _k [^] =Mmult(aの範囲, θ [^] の範囲)							∂ θ ₀ =1/(1+g(x _i))			∂ θ ₃ =a ₃ -a ₃ /(1+g(x _i))							
g(x _i) = exp(-(β ₀ [^] +β ₁ [^] x _i))							∂ θ ₁ =a ₁ -a ₁ /(1+g(x _i))			∂ β ₀ =(θ ₀ [^] -θ _k [^])g(x _i)/(1+g(x _i)) ²							
y _i [^] =θ _k [^] +(θ ₀ [^] -θ _k [^])/(1+g(x _i))							∂ θ ₂ =a ₂ -a ₂ /(1+g(x _i))			∂ β ₁ =(∂ β ₀)*x _i							

標準誤差 SE の計算は，偏微分した結果を

$$z=[0.93 \quad 0.07 \quad 0.00 \quad 0.00 \quad 0.09 \quad 2.49]$$

と微分係数ベクトル z とし，表 10.15 に示したパラメータの共分散行列 Σ(θ̂) を用い

$$SE = \sqrt{z \Sigma(\hat{\theta}) z^T} = 0.1293 \tag{10.38}$$

として計算している．

表 10.16 に示した予測値を用い，図 10.4 を JMP のグラフビルダーを用いて再現してみよう．図 10.14 に示すように Y 軸に肺活量 y，X 軸に年齢 x をセットし，予測値 y[^] を Y 軸にさらに加える．因子 A を「重ね合わせ」にセットし，記号を小さなドットに変更する．予測値を外挿するために，表 10.16 に手作業でデータを追加し，予測値 y[^] の計算結果を得ている．

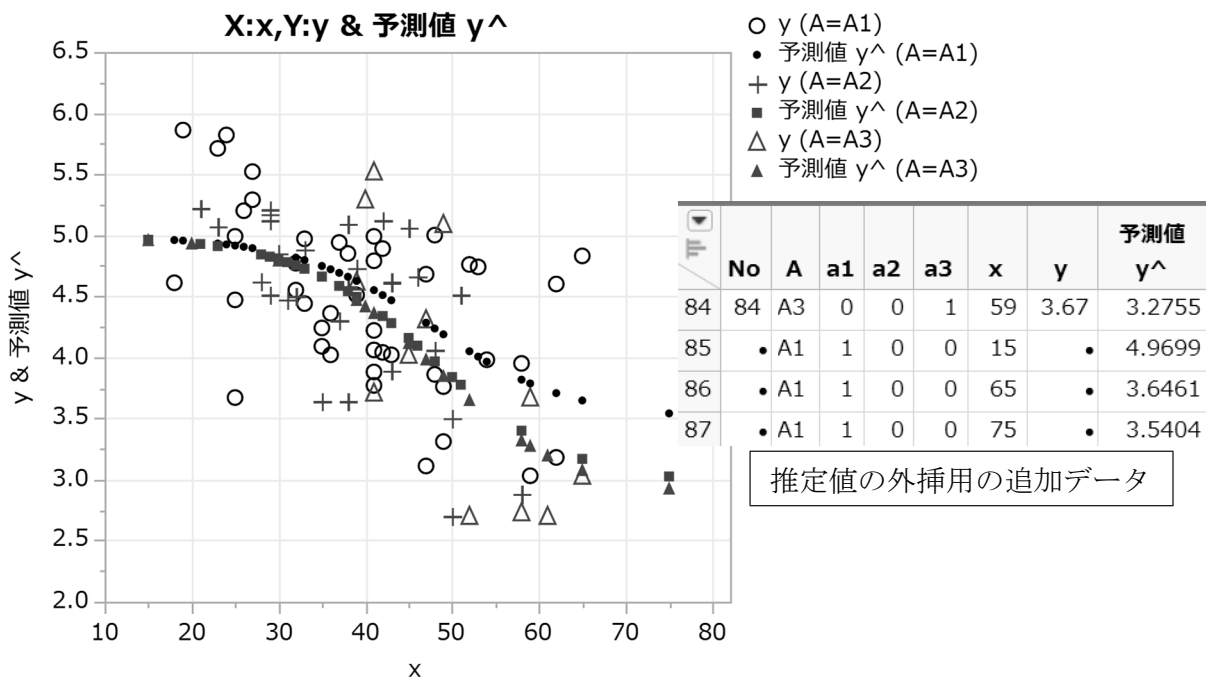


図 10.14 散布図へロジスティック曲線の重ね合わせ

図 10.14 で 95%信頼区間を重ね書きしても互いに重なり合っ、判別がつかなくなるので、図 10.15 に示すように、X 軸方向で因子 A の 3 群を分割表示し、95%信頼区間を重ね書きした結果を示す。

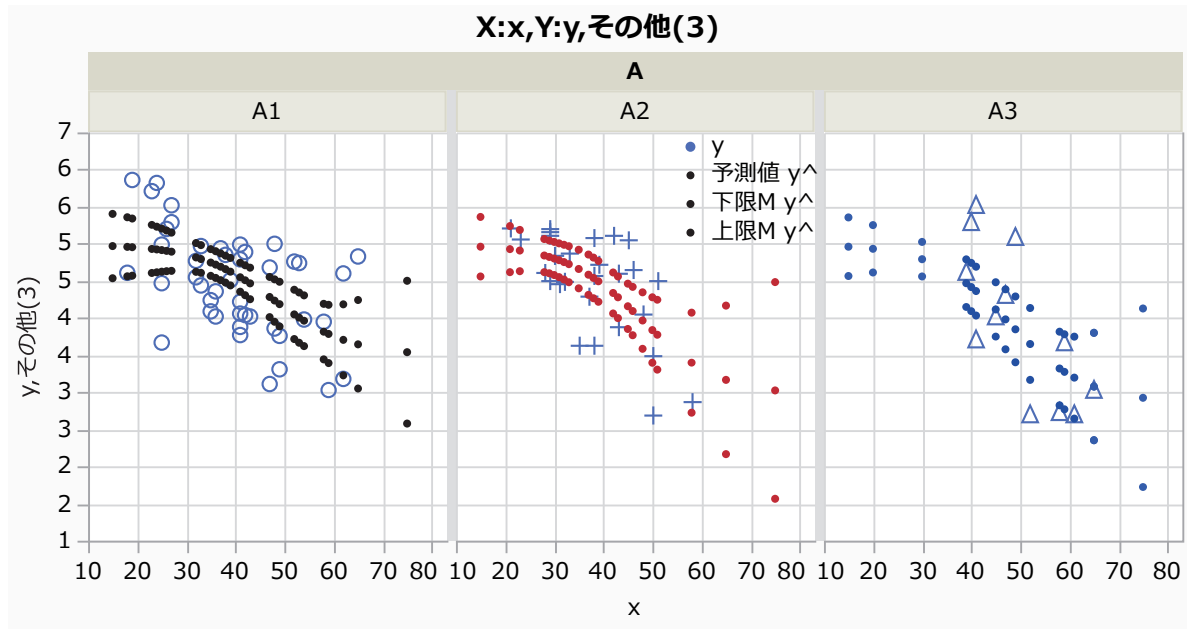


図 10.15 グラフビルダーによる 95%信頼区間入りロジスティック曲線のあてはめ

JMP によるロジスティック曲線式のパラメータに関する偏微分

最近の統計ソフトでは、非線形回帰による解析に不可欠なパラメータに関する偏微分を自動的に行う機能が組み込まれており、非線形回帰が、線形回帰と同様に気楽に使えるようになったことはうれしい限りである。JMP の「非線形回帰」では、どのように偏微分が行われているかをログに出力する機能があるので、以下に示す。

予測式

$$\theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3 + (T\#54 = \theta_0 - (\theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3)) /$$

$$\text{First}(T\#51 = 1 + \text{First}(T\#52 = \text{Exp}(-(\beta_0 + \beta_1 \cdot x)),$$

$$T\#55 = T\#52 \cdot -x), T\#53 = -1 \cdot T\#52)$$

予測式をパラメータで微分した式

$$\{1 / T\#51, \quad \rightarrow \theta_0 \text{ で偏微分した式}$$

$$:a_1 + -:a_1 / T\#51, \quad \rightarrow \theta_1 \text{ で偏微分した式}$$

$$:a_2 + -:a_2 / T\#51, \quad :$$

$$:a_3 + -:a_3 / T\#51, \quad :$$

- (T#53 * T#54 * T#51 ^ (-2)), → β_0 で偏微分した式
- (T#55 * T#54 * T#51 ^ (-2)) } → β_1 で偏微分した式

ログに出力された予測式および偏微分式は、内部計算の過程を示しているのので、難解であり、数式としての微分式にするのは煩わしい。そこで、JMP データファイル上で GUI による偏微分する機能を使うことを推奨する。

- 手順 1) 偏微分したいパラメータについて新たな変数 ∂_{β_1} を列に追加する。
- 手順 2) 予測式 \hat{y} の計算式をコピーする。
- 手順 3) 変数 ∂_{β_1} のブランクの計算式を開き、ペーストする。
- 手順 4) ペーストされた予測式 \hat{y} のパラメータ β_1 を選択する。
- 手順 5) プルダウンメニューから「微分した式」を選択する。
- 手順 6) 計算式エディターに微分した式が表示され、変数 ∂_{β_1} 列に微分した計算結果が埋め込まれる。

The image shows the JMP 'Differentiated Formula' menu on the left, with '微分した式' (Differentiated Formula) selected. To the right, the mathematical formula for the partial derivative is displayed:

$$\frac{\partial f(x_i)}{\partial \beta_1} = - \left(\begin{array}{l} \text{Exp} \left(- \left(\beta_0 + \beta_1 \cdot x \right) \right) \\ \cdot - x \\ \cdot \left(\theta_0 - \left(\theta_1 \cdot a_1 + \theta_2 \cdot a_2 + \theta_3 \cdot a_3 \right) \right) \\ \cdot \left(1 + \text{Exp} \left(- \left(\beta_0 + \beta_1 \cdot x \right) \right) \right)^{-2} \end{array} \right)$$

$$= \frac{[\theta_0 - (\theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3)] \cdot \exp[-(\beta_0 + \beta_1 x)] \cdot x}{\{1 + \exp[-(\beta_0 + \beta_1 x)]\}^2}$$

図 10.16 JMP の「微分した式」で得た $\hat{y} = f(x_i)$ の β_1 に関する偏微分式

図 10.16 に示した $\hat{y} = f(x_i)$ の β_1 に関する偏微分式は、図 10.13 に示した「予測値の標準誤差 \hat{y} 」の計算式の中で用いられている。

10.5. SAS の非線形 NLIN プロシジャによる解析

SAS には、非線形回帰のための NLIN プロシジャが古くからサポートされていたが、シグモイド曲線の式に対する偏微分式も明示的に自ら与える必要があり、手軽に使える状況ではなかった。その後、JMP と同様に自動偏微分の機能が加わり、気楽に使えるようになった。ここに示すのは、無償で継続的に使える OnDemand SAS による結果である。

非線形モデルの解析法については、自由にダウンロードできる SAS の NLIN プロシジャの解説書 SAS Institute (2020), 「The NLIN Procedure」が最も優れている。全 118 ページで、「Getting Started: NLIN Procedure」節に、線形モデルと非線形モデルの識別方法に引き続き、サンプルプログラムと実行結果が詳細に示されている。また、非線形モデルの解析法についても丁寧な説明があり、非線形モデルについて学習するために適している。

SAS/NLIN プロシジャによる解析

表 10.1 で示した Excel ファイル上のデータをスペース区切りのテキストを SAS の DATA ステップで、SAS データセット化する。データセット化は、一行ごとに A₁ 群の 3 組の (x_i, y_i) , A₂ 群の 2 組, A₃ 群の 1 組の順に読込、15 行目でデータリストが尽きるまで読み込む。13 行目以後に欠測値「.」もデータセットに書き込まれるが、解析に際しては無視される。

SAS プログラム 10.1

```
TITLE1 '層別シグモイド.SAS << カドミウムガスの曝露 >>';

DATA d01 ;
  input n @ ;
  A ="A1"; a1=1; a2=0; a3=0;
  do i=1,2,3 ;
    input x y @ ; output ; end;

  A ="A2"; a1=0; a2=1; a3=0;
  do i=1,2 ;
    input x y @ ; output ; end;

  A ="A3"; a1=0; a2=0; a3=1;
  input x y @ ; output ;

/* ----- A1 ----- ----- A2 ----- -- A3 -- */
/* x y x y x y x y x y x y x y */
datalines ;
1 27 5.29 35 4.09 48 3.86 29 5.21 43 4.61 39 4.62
2 25 3.67 35 4.24 47 4.68 29 5.17 39 4.73 40 5.29
3 24 5.82 41 3.88 53 4.74 33 4.88 38 4.58 41 5.52
4 32 4.77 38 4.85 49 3.76 32 4.50 42 5.12 41 3.71
5 23 5.71 41 4.79 54 3.98 31 4.47 43 3.89 45 4.02
```

```

6  25  4.47  36  4.36  48  5.00  29  5.12  43  4.62  49  5.09
7  32  4.55  36  4.02  49  3.31  29  4.51  37  4.30  52  2.70
8  18  4.61  41  3.77  47  3.11  30  4.85  50  2.70  47  4.31
9  19  5.86  41  4.22  52  4.76  21  5.22  50  3.50  61  2.70
10 26  5.20  37  4.94  58  3.95  28  4.62  45  5.06  65  3.03
11 33  4.44  42  4.04  62  4.60  23  5.07  48  4.06  58  2.73
12 27  5.52  39  4.51  65  4.83  35  3.64  51  4.51  59  3.67
13 33  4.97  41  4.06  62  3.18  38  3.64  46  4.66  .  .
14 25  4.99  43  4.02  59  3.03  38  5.09  58  2.88  .  .
15 42  4.89  41  4.99  .  .  .  .  .  .  .  .
;
proc print data=d01 ; run;

```

NLIN プロシジャの設定は、DATA ステップで作成された SAS データセット「d01」を使い、最大の反復回数 maxiter を 200 回、list オプションで偏微分式の出力を設定している。図 10.5 で設定したパラメータの初期値の設定を parms ステートメントで設定している。モデル式は、式 (10.30) に準じて model ステートメントで設定している。

SAS プログラム 10.2 3本のロジスティック曲線の同時あてはめ

```

proc nlin data=d01 maxiter=200 list outest=cov1 ;
  parms      theta0=5 theta1=4 theta2=3.5 theta3=3 beta0=4 beta1=-0.1 ;
            theta_k = (theta1*a1 + theta2*a2 + theta3*a3) ;
            g_x = exp(-(beta0 + beta1*x)) ;
  model      y = theta_k + (theta0 - theta_k) * (1 + g_x)**(-1) ;
run ;

proc print data=cov1 ; run;

```

表 10.18 に分散分析表とパラメータの推定値を示す。JMP で計算した表 10.13 に一致することが確かめられる。表 10.21 示すパラメータ間の相関行列から、 $\hat{\beta}_0$ と $\hat{\beta}_1$ の相関が 0.9866 となり、多重共線性すれすれであり、このことがなかなか収束しない理由でもある。OnDemand SAS の標準出力は、HTML 形式なので、結果を Excel シートに張り付け、書式を整え Word 上に「図 (拡張メタファイル)」形式でコピー&ペーストした結果である。

表 10.18 SAS/NLIN プロシジャによるパラメータの推定結果

要因	自由度	平方和	平均平方	F 値	Pr > F
Model	5	18.3953	3.6791	9.81	<.0001
Error	78	29.2457	0.3749		
C. Total	83	47.6410			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		
theta0	4.9925	0.2753	4.4444	5.5406	
theta1	3.4946	0.6291	2.2421	4.7471	
theta2	2.9621	0.9159	1.1386	4.7855	
theta3	2.8612	0.8046	1.2594	4.4630	
beta0	6.0858	3.8720	-1.6228	13.7944	
beta1	-0.1272	0.0919	-0.3102	0.0557	

NLIN プロシジャでは、パラメータの相関行列がデフォルトで出力される。ただし、パラメータの共分散行列は、「OUTEST=cov1」オプションで外部ファイルに出力し、その結果を PRINT プロシジャで出力する必要がある。表 10.19 に結果を示すが、反復計算の過程が出力され、「FINAL」が、最終的に得られたパラメータであり、「COVB」の行にパラメータ共分散行列が出力されている。

表 10.19 パラメータの共分散行列 $\Sigma(\hat{\theta})$

OBS	_TYPE_	_STATUS_	_NAME_	θ_0	θ_1	θ_2	θ_3	β_0	β_1
125	COVB	0 Converged	θ_0	0.0758	-0.1042	-0.1196	-0.1209	-0.9229	0.0205
126	COVB	0 Converged	θ_1	-0.1042	0.3958	0.4901	0.4375	1.8964	-0.0491
127	COVB	0 Converged	θ_2	-0.1196	0.4901	0.8389	0.6217	2.3538	-0.0636
128	COVB	0 Converged	θ_3	-0.1209	0.4375	0.6217	0.6474	2.3749	-0.0619
129	COVB	0 Converged	β_0	-0.9229	1.8964	2.3538	2.3749	14.9926	-0.3511
130	COVB	0 Converged	β_1	0.0205	-0.0491	-0.0636	-0.0619	-0.3511	0.0085

表 10.20 に示すように NLIN プロシジャの「list」オプションにより、内部で計算されている偏微分した結果を出力することができる。「@MODEL.y/@ β_0 」と「@MODEL.y/@ β_1 」の式は組み入っているが、式 (10.32) に示したように、

$$f(x_i) = \theta_k + (\theta_0 - \theta_k)[1 + g(x_i)]^{-1}$$

ただし、 $\theta_k = \theta_1 a_{i,1} + \theta_2 a_{i,2} + \theta_3 a_{i,3}$

$$g(x_i) = \exp[-(\beta_0 + \beta_1 x_i)]$$

とおくと、 θ_0 から θ_4 までの偏微分は、式 (10.31) と同じであるが、 β_0 と β_1 の偏微分式には、他の偏微分式が引用されているが、整理すると式 (10.32) に一致する。

$$z_4 = \frac{\partial f(x_i)}{\partial \beta_0} = (\theta_0 - \theta_k) \frac{\partial g(x_i)}{\partial \beta_0} \{-[1 + g(x_i)]^{-2}\}$$

$$= (\theta_0 - \theta_k) g(x_i) [1 + g(x_i)]^{-2}$$

表 10.20 内部で計算された偏微分式の結果

Listing of Compiled Program Code		
Stmt	Line:Col	Statement as Parsed
1	71:08:00	$\theta_k = \theta_1 * a_1 + \theta_2 * a_2 + \theta_3 * a_3;$
2	72:08:00	$g_x = \text{EXP}(-(\beta_0 + \beta_1 * x));$
2	72:08:00	$@g_x/@\beta_0 = -g_x;$
2	72:08:00	$@g_x/@\beta_1 = -x * g_x;$
3	73:05:00	$\text{MODEL.y} = \theta_k + (\theta_0 - \theta_k) * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_0 = (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_1 = a_1 + - a_1 * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_2 = a_2 + - a_2 * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\theta_3 = a_3 + - a_3 * (1 + g_x) ** -1;$
3	73:05:00	$@\text{MODEL.y}/@\beta_0 = (\theta_0 - \theta_k) * @g_x/@\beta_0 * -(1 + g_x) ** -2;$
3	73:05:00	$@\text{MODEL.y}/@\beta_1 = (\theta_0 - \theta_k) * @g_x/@\beta_1 * -(1 + g_x) ** -2;$

$$\begin{aligned}
 z_5 &= \frac{\partial f(x_i)}{\partial \beta_1} = (\theta_0 - \theta_k) \frac{\partial g(x_i)}{\partial \beta_0} \{-[1 + g(x_i)]^{-2}\} \\
 &= (\theta_0 - \theta_k) \cdot x_i \cdot g(x_i) [1 + g(x_i)]^{-2} \\
 &= \frac{\partial f(x_i)}{\partial \beta_0} x_i
 \end{aligned}$$

式 (10.32) に一致する.

パラメータの共分散行列とパラメータの相関行列は、表 10.21 の Excel シート示すように、定義に従った簡単な行列計算で片方が得れば、もう片方を計算することができる。なお、SE と示したのは、表 10.18 に示したパラメータの推定結果の「近似標準誤差」であり、有効数字が少数定以下 4 桁なので、若干の計算誤差が出ている。

表 10.21 パラメータの相関行列からパラメータの共分散行列の求め方

近似相関行列							
	θ_0	θ_1	θ_2	θ_3	β_0	β_1	SE
θ_0	1	-0.6018	-0.4743	-0.5458	-0.8657	0.8082	0.2753
θ_1	-0.6018	1	0.8505	0.8643	0.7785	-0.8487	0.6291
θ_2	-0.4743	0.8505	1	0.8437	0.6637	-0.7559	0.9159
θ_3	-0.5458	0.8643	0.8437	1	0.7623	-0.8376	0.8046
β_0	-0.8657	0.7785	0.6637	0.7623	1	-0.9866	3.8720
β_1	0.8082	-0.8487	-0.7559	-0.8376	-0.9866	1	0.0919
Excel の行列計算によるパラメータの共分散行列の算出 $\Sigma(\hat{\theta})$							
	θ_0	θ_1	θ_2	θ_3	β_0	β_1	
θ_0	0.0758	-0.1042	-0.1196	-0.1209	-0.9228	0.0204	
θ_1	-0.1042	0.3958	0.4901	0.4375	1.8963	-0.0491	
θ_2	-0.1196	0.4901	0.8389	0.6217	2.3538	-0.0636	
θ_3	-0.1209	0.4375	0.6217	0.6474	2.3750	-0.0619	
β_0	-0.9228	1.8963	2.3538	2.3750	14.9924	-0.3511	
β_1	0.0204	-0.0491	-0.0636	-0.0619	-0.3511	0.0084	
=(相関行列の範囲)*(SEの範囲)*Transpose (SEの範囲)							

NLIN プロシジャの出力結果を Excel の折れ線グラフで

NLIN プロシジャの出力結果を Excel に取り込むことにより、これまで示してきた Excel の線グラフおよび散布図を用いた各種のグラフ表示に繋げることができる。表 10.18 に示した 3 本のロジスティック曲線の最大値および最小値に関して 95 信頼区間のひげ付きグラフ、 A_1 非曝露群と A_2 および A_3 曝露群の差についても同様に髭付きグラフに示したい。すでに、図 10.6 で示してあるが、SAS/NLIN プロシジャの出力結果から、同様のグラフ表示をする方法を示すことにより、SAS と Excel の連携法の例示とする。

表 10.22 に示すのは、表 10.18 のパラメータの推定結果から $\hat{\theta}_0 \sim \hat{\theta}_3$ を抜き出し、さらに、表 10.21 からパラメータの共分散行列から $\hat{\theta}_1 \sim \hat{\theta}_3$ を抜き出した結果である。Excel の線グラフで 95%信頼区間の幅を付けるためには、推定値に対し $t_{0.05} \times SE$ の上下幅が必要なので、(上限 - 下限) / 2 による計算を付け加えてある。

A₁ 非曝露群との差は、推定値の差から求められが、その分散は

$$\begin{aligned} \text{Var}(\hat{\theta}_2 - \hat{\theta}_1) &= \text{Var}(\hat{\theta}_2) - 2\text{Cov}(\hat{\theta}_2, \hat{\theta}_1) + \text{Var}(\hat{\theta}_1) \\ &= 0.3958 - 2 \times 0.4901 + 0.8389 = 0.2545 \end{aligned}$$

によって計算されている。図 10.17 は、図 10.6 と同じ線グラフを別途作成した表 10.22 の結果を用いて作成したものである。

表 10.22 パラメータの推定値および共分散行列から信頼区間の幅の算出

パラメータ	推定値		近似標準誤差	近似 95% 信頼限界		t*SE 幅
θ₀	4.9925		0.2753	4.4444	5.5406	0.5481
θ₁	3.4946		0.6291	2.2421	4.7471	1.2525
θ₂	2.9621		0.9159	1.1386	4.7855	1.8235
θ₃	2.8612		0.8046	1.2594	4.4630	1.6018
差	推定値	分散	SE	L95%	U95%	t*SE 幅
θ ₁ - θ ₁	0					0
θ ₂ - θ ₁	-0.5325	0.2545	0.5045	-1.5369	0.4719	1.0044
θ ₃ - θ ₁	-0.6334	0.1682	0.4101	-1.4498	0.1830	0.8164
パラメータの共分散行列				t _{0.05(84-6)} =	1.9908	
	θ ₁	θ ₂	θ ₃			
θ ₁	0.3958	0.4901	0.4375			
θ ₂	0.4901	0.8389	0.6217			
θ ₃	0.4375	0.6217	0.6474			

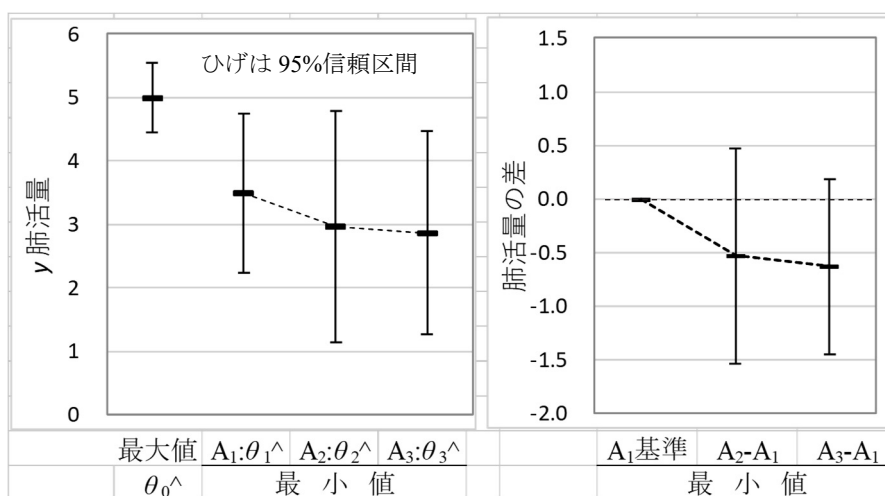


図 10.17 シグモイド曲線の共通の最大値および各層の最小値 (図 10.6 再掲)

10.6. 残差線形化法によるロジスティック曲線のパラメータ推定

第 10.2 節では、ロジスティック曲線のパラメータ推定をするために Excel のソルバーを用いた。第 10.3 節では、ロジスティック曲線式をパラメータで偏微分し、微分係数行列 \mathbf{Z} からパラメータの共分散行列 $\Sigma(\hat{\theta})$ を求めることにより、各種の推定値に対する 95%信頼区間が得られることを示した。ただし、初期値からどのような手順でパラメータの推定値を得るかについては、“ソルバー”まかせであった。また、第 10.4 節の JMP の「非線形回帰」、第 10.5 節の SAS の NLIN プロシジャを用いた場合についても、同様にブラックボックスのままであった。そこで、本節では、残差 $\hat{\varepsilon}_i = y_i - \hat{y}_i$ に対し、微分係数行列 \mathbf{Z} をデザイン行列と見なした残差線形化法（ガウス・ニュートン法）によるパラメータの推定値法を示す。

残差を反応とした回帰分析の反復

ロジスティック曲線のパラメータ推定は、パラメータの初期値によって推定された $\hat{y}_i^{(0)}$ と y_i の残差 $\hat{\varepsilon}_i^{(0)} = y_i - \hat{y}_i^{(0)}$ を新たな反応とみなし、デザイン行列 \mathbf{X} に変えて微分係数行列 \mathbf{Z} を用いて回帰分析を繰り返し行うことから、残差線形化法とも言われている。表 10.23 に示すのは、表 10.6 に示した初期値 ($\hat{\theta}_0^{(0)} = 5.0$, $\hat{\theta}_1^{(0)} = 4.0$, $\hat{\theta}_2^{(0)} = 3.5$, $\hat{\theta}_3^{(0)} = 3.0$, $\hat{\beta}_0^{(0)} = 4.0$, $\hat{\beta}_1^{(0)}$)

表 10.23 初期値に対する残差 $\hat{\varepsilon}_i^{(0)}$ および微分係数行列 $\mathbf{Z}^{(0)}$ の計算結果

		θ				$(\mathbf{Z}^T \mathbf{Z})^{-1}$						$\mathbf{Z}^T \hat{\varepsilon}$				
		θ_0	θ_1	θ_2	θ_3	$\delta \wedge^{(m-1)}$	$\theta \wedge^{(m-1)}$	$\theta \wedge^{(m)}$								
最大値		θ_0	5.000	0.046	5.046	0.683	-0.447	-0.449	-0.536	-5.194	0.120	5.343				
最小A ₁		θ_1	4.000	-0.647	3.353	-0.447	0.606	0.696	0.796	4.301	-0.117	-2.368				
A ₂		θ_2	3.500	-0.674	2.826	-0.449	0.696	1.126	1.097	4.806	-0.140	1.262				
A ₃		θ_3	3.000	-0.360	2.640	-0.536	0.796	1.097	1.429	6.013	-0.171	2.006				
切片		β_0	4.000	1.124	5.124	-5.194	4.301	4.806	6.013	47.645	-1.165	2.976				
傾き		β_1	-0.100	0.006	-0.094	0.120	-0.117	-0.140	-0.171	-1.165	0.030	103.8				
残差平方和		S_e	32.867													
----- 微分係数行列 \mathbf{Z} -----																
群	ダミー	年齢	肺活量	推定値	残差	z_0	z_1	z_2	z_3	z_4	z_5					
No	A	a_1	a_2	a_3	x	y	θ_k	$g(x)$	y^{\wedge}	ε^{\wedge}	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \theta_4$	$\partial \theta_5$
1	A ₁	1	0	0	27	5.29	4.000	0.273	4.786	0.504	0.786	0.214	0.000	0.000	0.168	4.544
2	A ₁	1	0	0	25	3.67	4.000	0.223	4.818	-1.148	0.818	0.182	0.000	0.000	0.149	3.729
:																
44	A ₁	1	0	0	59	3.03	4.000	6.686	4.130	-1.100	0.130	0.870	0.000	0.000	0.113	6.678
45	A ₂	0	1	0	29	5.21	3.500	0.333	4.625	0.585	0.750	0.000	0.250	0.000	0.281	8.151
46	A ₂	0	1	0	29	5.17	3.500	0.333	4.625	0.545	0.750	0.000	0.250	0.000	0.281	8.151
:																
72	A ₂	0	1	0	58	2.88	3.500	6.050	3.713	-0.833	0.142	0.000	0.858	0.000	0.183	10.59
73	A ₃	0	0	1	39	4.62	3.000	0.905	4.050	0.570	0.525	0.000	0.000	0.475	0.499	19.45
74	A ₃	0	0	1	40	5.29	3.000	1.000	4.000	1.290	0.500	0.000	0.000	0.500	0.500	20.00
:																
84	A ₃	0	0	1	59	3.67	3.000	6.686	3.260	0.410	0.130	0.000	0.000	0.870	0.226	13.36

= -0.10) を用いて, 式 (10.29) により $\hat{y}_i^{(0)}$ を推定し, さらに, 残差 $\hat{\varepsilon}_i^{(0)}$ を計算している. 式 (10.31) および式 (10.32) に示したパラメータで偏微分した式 z_k を計算し, (84 行 \times 6 列) の微係数行列 $\mathbf{Z}^{(0)}$ を計算した結果である.

残差 $\hat{\varepsilon}_i^{(0)}$ を反応として, 微係数行列 $\mathbf{Z}^{(0)}$ をデザイン行列 \mathbf{X} と見なした回帰分析により, パラメータ $\delta_i^{(0)}$ を線形モデル $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ と同様に

$$\hat{\boldsymbol{\delta}}^{(0)} = (\mathbf{Z}^{(0)T} \mathbf{Z}^{(0)})^{-1} \mathbf{Z}^{(0)T} \hat{\boldsymbol{\varepsilon}}^{(0)} \quad (10.39)$$

によりパラメータ $\hat{\boldsymbol{\delta}}^{(0)}$ を推定する. 表 10.23 右上段に, $(\mathbf{Z}^{(0)T} \mathbf{Z}^{(0)})^{-1}$ および $\mathbf{Z}^{(0)T} \hat{\boldsymbol{\varepsilon}}^{(0)}$ の計算結果が示され, 表 10.23 右上段にパラメータの推定値としての増減値 $\hat{\boldsymbol{\delta}}^{(0)}$ の推定結果が示されている. さらに, 推定された増減値 $\hat{\boldsymbol{\delta}}^{(0)}$ を初期パラメータ $\hat{\boldsymbol{\theta}}^{(0)}$ に加え

$$\hat{\boldsymbol{\theta}}^{(0)} + \hat{\boldsymbol{\delta}}^{(0)} = \hat{\boldsymbol{\theta}}^{(1)} \quad (10.40)$$

次のように, パラメータ $\hat{\boldsymbol{\theta}}^{(0)}$ を更新した $\hat{\boldsymbol{\theta}}^{(1)}$

$$m=1 : \begin{array}{|c|} \hline \hat{\boldsymbol{\theta}}^{(m-1)} \\ \hline 5.000 \\ 4.000 \\ 3.500 \\ 3.000 \\ 4.000 \\ -0.100 \\ \hline \end{array} + \begin{array}{|c|} \hline \hat{\boldsymbol{\delta}}^{(m-1)} \\ \hline 0.046 \\ -0.647 \\ -0.674 \\ -0.360 \\ 1.124 \\ 0.006 \\ \hline \end{array} = \begin{array}{|c|} \hline \hat{\boldsymbol{\theta}}^{(m)} \\ \hline 5.046 \\ 3.353 \\ 2.826 \\ 2.640 \\ 5.124 \\ -0.094 \\ \hline \end{array}$$

が計算されている. 推定された $\hat{\boldsymbol{\theta}}^{(1)}$ をコピーし, 第 2 回目の反復 $m=2$ とし, $\hat{\boldsymbol{\theta}}^{(m-1)}$ に値のみをペーストする. すると, Excel の自動計算機能により, パラメータを $\hat{\boldsymbol{\theta}}^{(1)}$ として全ての計算が更新される. 更新された増減値 $\hat{\boldsymbol{\delta}}^{(1)}$ をパラメータ $\hat{\boldsymbol{\theta}}^{(1)}$ に加えパラメータ $\hat{\boldsymbol{\theta}}^{(2)}$ が

$$m=2 : \begin{array}{|c|} \hline \hat{\boldsymbol{\theta}}^{(m-1)} \\ \hline 5.046 \\ 3.353 \\ 2.826 \\ 2.640 \\ 5.124 \\ -0.094 \\ \hline \end{array} + \begin{array}{|c|} \hline \hat{\boldsymbol{\delta}}^{(m-1)} \\ \hline -0.210 \\ 1.708 \\ 2.093 \\ 2.305 \\ 2.883 \\ -0.098 \\ \hline \end{array} = \begin{array}{|c|} \hline \hat{\boldsymbol{\theta}}^{(m)} \\ \hline 4.836 \\ 5.062 \\ 4.919 \\ 4.946 \\ 8.007 \\ -0.192 \\ \hline \end{array}$$

得られる. このような繰り返して, 増減値 $\hat{\boldsymbol{\delta}}^{(m-1)}$ 全体が 0 とみなせるまで繰り返す.

経験的には, 何回かの繰り返して増減値 $\hat{\boldsymbol{\delta}}^{(m-1)}$ が 0 となるのだが, ここに示した事例では途中で残差平方和 S_e が減少するのではなく, 逆に増大し解が求められなくなってしまう. このような場合には, 無理をせずソルバーで残差平方和 S_e を最小にするように, パラメータを変化させつつ増減値 $\hat{\boldsymbol{\delta}}^{(m-1)}$ が 0 となることを確認するとよい. ソルバーには, 残差平方和 S_e が増大しないような計算上の工夫がされていると思われるが, その方法についてはブラックボックスである. ただし, とんでもない初期値に対しては, 計算不能となってしまうこともあ

る。また、なかなか収束しない場合は、「実行を一時中止しました」とのメッセージが出るが、残差平方和 S_e が減少していれば、「続行」することにより、表 10.24 に示すように増減値 $\hat{\delta}^{(m-1)}$ が 0.0 となる。

表 10.24 3本のシグモイド曲線の同時推定の結果

		θ		$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$	$(Z^T Z)^{-1}$						$Z^T \varepsilon^{\wedge}$			
	最大値	θ_0	4.992	0.000	4.992		0.202	-0.278	-0.319	-0.322	-2.461	0.055	0.000			
	最小	$A_1 \theta_1$	3.495	0.000	3.495		-0.278	1.055	1.307	1.166	5.057	-0.131	0.000			
		$A_2 \theta_2$	2.962	0.000	2.962		-0.319	1.307	2.237	1.658	6.277	-0.170	0.000			
		$A_3 \theta_3$	2.861	0.000	2.861		-0.322	1.166	1.658	1.726	6.333	-0.165	0.000			
	切片	β_0	6.087	0.000	6.087		-2.461	5.057	6.277	6.333	39.992	-0.936	0.000			
	傾き	β_1	-0.127	0.000	-0.127		0.055	-0.131	-0.170	-0.165	-0.936	0.023	0.007			
	平方和	S_e	29.246													
----- 微分係数行列 Z -----																
群	ダミー	年齢	肺活量			推定値	残差	z_0	z_1	z_2	z_3	z_4	z_5			
No	A	a_1	a_2	a_3	x	y	θ_k	$g(x)$	y^{\wedge}	ε^{\wedge}	$\partial \theta_0$	$\partial \theta_1$	$\partial \theta_2$	$\partial \theta_3$	$\partial \theta_4$	$\partial \theta_5$
1	A_1	1	0	0	27	5.29	3.495	0.071	4.894	0.396	0.934	0.066	0.000	0.000	0.092	2.491
2	A_1	1	0	0	25	3.67	3.495	0.055	4.915	-1.245	0.948	0.052	0.000	0.000	0.074	1.842
:																
44	A_1	1	0	0	59	3.03	3.495	4.145	3.786	-0.756	0.194	0.806	0.000	0.000	0.235	13.84
45	A_2	0	1	0	29	5.21	2.962	0.091	4.823	0.387	0.917	0.000	0.083	0.000	0.155	4.504
46	A_2	0	1	0	29	5.17	2.962	0.091	4.823	0.347	0.917	0.000	0.083	0.000	0.155	4.504
:																
72	A_2	0	1	0	58	2.88	2.962	3.649	3.399	-0.519	0.215	0.000	0.785	0.000	0.343	19.88
73	A_3	0	0	1	39	4.62	2.861	0.325	4.469	0.151	0.755	0.000	0.000	0.245	0.395	15.39
74	A_3	0	0	1	40	5.29	2.861	0.369	4.418	0.872	0.730	0.000	0.000	0.270	0.420	16.79
:																
84	A_3	0	0	1	59	3.67	2.861	4.145	3.275	0.395	0.194	0.000	0.000	0.806	0.334	19.69

推定されたパラメータを用いて各群の推定値を求め、散布図上に重ね書きした結果が、すでに示した図 10.4 である。カドミウムガス非曝露 A_1 群は、最大肺活量の $\hat{\theta}_0=4.9925$ リットルから徐々に肺活量が減少し $\hat{\theta}_1=3.4947$ リットルに収束している。曝露 A_2 群は、 $\hat{\theta}_2=2.9621$ リットルと曝露 A_3 群の $\hat{\theta}_3=2.8612$ リットルに近い推定値が得られている。これらから、カドミウムガスの曝露により、徐々に肺活量が低下し、最終的に非曝露群に対し、約 0.6 リットルの減少となると推測される。

Wikipedia のガウス・ニュートン法

ロジスティック曲線のあてはめにかぎらず、簡単ではあるが典型的な事例を示し、きちっとした解析法の考え方を導入するのが、段階的な進め方である。この章では、いきなり高度な応用事例から初めてしまった。本書のテーマである「層別因子を含む探索的な回帰分析入門」の趣旨として、「層別因子を含むロジスティック曲線」を最初に取り上げたのは、推定結果の統計的な解釈にどのように寄与するのかを提示したいと考えたからである。

非線形回帰の典型例は、第 12 章で取り上げるミカエリス・メンテン式である。Web 上の Wikipedia「非線形回帰」にも登場する。さらに、Wikipedia「ミカエリス・メンテン式」を参照すると、「線形化によるパラメータ推定は、非線形回帰に劣る」とあるが、非線形回帰の数値例が見当たらない。このような状況を改善するために適切な計算事例の提示が必要と考えている。

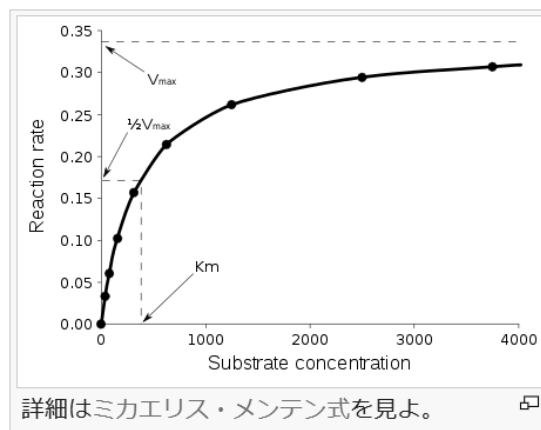
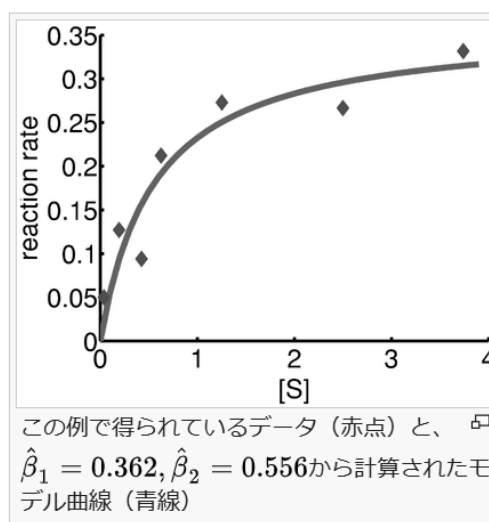


図 10.18 Wikipedia「非線形回帰」

数値例は、Wikipedia「ガウス・ニュートン法」で見出すことができる。掲載されている数値を使い残差線形化法（ガウス・ニュートン法の計算方法）を用いた事例を表 10.25 に示す。Wikipedia で示されている初期値 ($\hat{\beta}_1 = 0.9$, $\hat{\beta}_2 = 0.5$) を与え、反復を繰り返すと、5 回目で Wikipedia に示されている ($\hat{\beta}_1 = 0.362$, $\hat{\beta}_2 = 0.556$) が得られる。



ただし、推定されてパラメータ $\hat{\beta}_1 = 0.362$ は、最大値 V_{max} であり、 $\hat{\beta}_2 = 0.556$ は最大値の 1/2 となる濃度（ミカエリス定数） K_m である。さらなる解析については、第 12 章を参照のこと。

図 10.19 Wikipedia「ガウス・ニュートン法」

表 10.25 Wikipedia の「ガウス・ニュートン法」のデータ・Excel により 5 反復で収束

初期値		$\beta^{(m)}$	δ	$\beta^{(m+1)}$	$(Z^T Z)^{-1}$	
0.9	$\beta_1^{\wedge} = V_{max}^{\wedge} =$	0.362	0.000	0.362	1.521	6.341
0.2	$\beta_2^{\wedge} = K_m^{\wedge} =$	0.556	0.000	0.556	6.341	36.172
	$S_e =$	0.008				
	[S]	V	V^{\wedge}	$\varepsilon^{\wedge} = V - V^{\wedge}$	$\partial V_{max}^{\wedge}$	∂K_m^{\wedge}
No	x	y	y^{\wedge}	$\varepsilon^{\wedge} = y - y^{\wedge}$	z_1	z_2
1	0.038	0.050	0.023	0.027	0.064	-0.039
2	0.194	0.127	0.094	0.033	0.259	-0.125
3	0.425	0.094	0.157	-0.063	0.433	-0.160
4	0.626	0.2122	0.192	0.021	0.530	-0.162
5	1.253	0.2729	0.251	0.022	0.693	-0.139
6	2.500	0.2665	0.296	-0.029	0.818	-0.097
7	3.740	0.3317	0.315	0.017	0.871	-0.073
$V^{\wedge} = V_{max}^{\wedge} [S] / ([S] + K_m^{\wedge})$		$z_1 = [S] / ([S] + K_m^{\wedge})$		$z_2 = -V_{max}^{\wedge} [S] / ([S] + K_m^{\wedge})^2$		
		$\delta = (Z^T Z)^{-1} Z^T \varepsilon^{\wedge}$		$\beta^{(m+1)} = \beta^{(m)} + \delta$		

11. 各種のシグモイド曲線を用いた逆推定

人間を含む生物に対する化合物の作用の多くは、シグモイド曲線状の反応になることが知られており、シグモイド曲線となる関数として累積分布関数が使われてきた。化合物の急性毒性データの 50 パーセント致死量 (LD_{50}) を推定するためのプロビット法は、正規分布の累積分布関数が用いられている。正規分布の累積分布関数に代え、ロジスティック分布の累積分布関数も広く使われている。累積分布関数は、0.0 から 1.0 の範囲で単調増加する関数であるが、前章で例示したように、最小反応から最大反応の範囲となるようなシグモイド曲線に拡張することも容易である。反応が 2 値の場合には、ロジスティック回帰分析として多くの説明変数含む標準的な解析法として活用されている。しかしながら、反応が連続量の場合には、一般的な解析法として普及していない。本章では、シグモイド曲線のあてはめの基本に戻り、典型的な事例について例示する。

11.1. 累積分布関数のシグモイド曲線としての活用

正規分布

正規分布の確率密度関数 $f_{NOR}(x)$ は、位置パラメータ μ 、形状パラメータ^{注)} σ を用いて

$$f_{NOR}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (11.1)$$

として定義される。その累積分布関数 $F_{NOR}(x)$ は、

$$F_{NOR}(x) = \int_{-\infty}^x f_{NOR}(x) dx \quad (11.2)$$

と積分で定義され、シグモイド曲線の代表的な存在であり、プロビット法として 50 パーセント致死量の推定に用いられてきた。ただし、累積分布関数は、一般的な関数を用いて表すことができないこともあり、ロジスティック分布が広く使われるようになってきた。ただし、表 11.1 に示すように Excel の Norm.dist () 関数により、 $f_{NOR}(x)$ および $F_{NOR}(x)$ の計算が簡単に行なえるようになり、シグモイド曲線として手軽に使えるようになった。なお、シグモイド曲線は、単調増加だけでなく、単調減少となる場合もあり、1 から $F_{NOR}(x)$ を差し引いた

$$S_{NOR}(x) = 1 - F_{NOR}(x) \quad (11.3)$$

のように、いわゆる生存関数 $S_{NOR}(x)$ となる場合もあるので、使い分けが必要となる。

表 11.1 に示すのは、位置パラメータを $\mu=5.0$ 、形状パラメータを $\sigma=1.5$ とした場合に、 x を 0 から 1 刻みで 10 まで変化させ、正規分布の確率密度関数 $f_{NOR}(x)$ 、累積分布関数 $F_{NOR}(x)$ 、生存関数 $S_{NOR}(x)$ を Excel の関数を用いて計算した結果である。正規分布のパラメータは、期待値 μ (平均値) と分散 σ^2 (標準偏差 σ) であるが、各種のシグモイド曲線となる分布関数のパラメータも μ と σ で表わすことができる。ただし、平均と標準偏差に一致しない場合もあるので、正規分布の場合に位置パラメータを μ_{NOR} 、形状パラメータを σ_{NOR} と明示する必要もある。図 11.1 に示すのは、Excel の散布図で 3 種の関数を重ね書きし、「データ系列の書式設定」で「スムージング」を適用して滑らかな曲線とした結果である。

表 11.1 正規分布の 3 種の関数

$\mu_{NOR} =$	5	位置パラメータ	
$\sigma_{NOR} =$	1.5	形状パラメータ	
	密度	累積	生存
x	$f_{NOR}(x)$	$F_{NOR}(x)$	$1-F_{NOR}(x)$
0.0	0.0010	0.0004	0.9996
1.0	0.0076	0.0038	0.9962
2.0	0.0360	0.0228	0.9772
3.0	0.1093	0.0912	0.9088
4.0	0.2130	0.2525	0.7475
5.0	0.2660	0.5000	0.5000
6.0	0.2130	0.7475	0.2525
7.0	0.1093	0.9088	0.0912
8.0	0.0360	0.9772	0.0228
9.0	0.0076	0.9962	0.0038
10.0	0.0010	0.9996	0.0004
$f_{NOR}(x)=\text{Norm.dist}(x, 5, 1.5, \text{false})$			
$F_{NOR}(x)=\text{Norm.dist}(x, 5, 1.5, \text{true})$			

図 11.1 正規分布の 3 種の関数の例示

反応の有無のような 2 値データに対し、半数の個体に反応が起きる化合物の 50%反応量 D_{50} を推定したい。幾つかの固定用量 $x_i = \log_{10}(\text{dose}_i)$ に対し、 n_i 回の試行に対する反応があった場合を y_i 回とする。各用量 x_i における出現率 $p_i = y_i / n_i$ は、用量 x_i に関してシグモイド曲線状になることが経験的に知られている。

$$\frac{y_i}{n_i} = F_{NOR}\left(\frac{x_i - \mu_{NOR}}{\sigma_{NOR}}\right) + \varepsilon_i, \quad \varepsilon_i \sim \text{二項分布} \quad (11.4)$$

誤差分布に二項分布を仮定するので、尤度関数 L

$$L = \left. \begin{aligned} & \prod_i \text{Binomial}(y_i, n_i, \hat{\pi}_i) \\ & \text{ただし, } \hat{\pi}_i = F_{NOR}\left(\frac{x_i - \hat{\mu}_{NOR}}{\hat{\sigma}_{NOR}}\right)_i \end{aligned} \right\} \quad (11.5)$$

が最大になるような $\hat{\mu}_{NOR}$ および $\hat{\sigma}_{NOR}$ を推定したい。歴史的には、式 (11.4) を標準正規分布の逆関数 $F_{NOR}^{-1}()$ を用い

$$F_{NOR}^{-1}\left(\frac{y_i}{n_i}\right) = \left. \begin{aligned} &= -\frac{\mu_{NOR}}{\sigma_{NOR}} + \frac{1}{\sigma_{NOR}}x_i + \varepsilon'_i \\ &= \beta_0 + \beta_1 x_i + \varepsilon'_i \end{aligned} \right\} \quad (11.6)$$

と線形化し、反復重み付き回帰による最尤法で解く方法がプロビット法として定式化されている。求めたい 50%反応量 D_{50} は、 $F_{NOR}^{-1}(0.50) = 0$ なので、用量 x_i について解いて、推定された回帰直線の切片 $\hat{\beta}_0$ と傾き $\hat{\beta}_1$ の比によって

$$D_{50} = -\frac{\hat{\beta}_0}{\hat{\beta}_1} = \hat{\mu}_{NOR} \quad (11.7)$$

として推定される。詳しくは、吉村編著 (1987)、「毒性・薬効データの統計解析、5.4 節 LD_{50} の推定法」、ドブソン著・田中ら訳 (2008)、「一般化線形モデル入門 原著第 2 版、7.3 節 用量反応モデル」、Dobson ら (2018)、「An Introduction to Generalized Linear Model 4th Ed. Section 7 Dose response models」、高橋 (2017)、「第 6 回 続高橋セミナー：一般化線形モデルを Excel で極め活用するープロビット法・ロジット法・補 2 重対数法ー」を参照のこと。

注) 「 σ 」は、尺度パラメータとも言われるが、信頼性工学でワイブル分布 [式 (11.37)] の「尺度 (scale) パラメータ η 」は、実際には「位置パラメータ」であり、「形状 (shape) パラメータ m 」の逆数が、 σ であり、確率密度関数の形状を変化させる σ として、形状パラメータとすることにする。

ロジスティック分布

ロジスティック分布の累積分布関数 $F_{LGS}(x)$ は、 μ_{LGS} を位置パラメータ、 σ_{LGS} を形状パラメータとした場合に、

$$F_{LGS}(x) = \frac{1}{1 + \exp\left[-\frac{(x - \mu_{LGS})}{\sigma_{LGS}}\right]} \quad (11.8)$$

と定義され、確率密度関数 $f_{LGS}(x)$ は、 $F_{LGS}(x)$ を x で微分して

$$f_{LGS}(x) = \frac{\exp\left[-\frac{(x - \mu_{LGS})}{\sigma_{LGS}}\right]}{\sigma_{LGS} \left\{1 + \exp\left[-\frac{(x - \mu_{LGS})}{\sigma_{LGS}}\right]\right\}^2} \quad (11.9)$$

として求められる。正規分布のパラメータ σ_{NOR} の推定値は、算術平均 $\hat{\mu}$ からの偏差平方和を自由度 $(n-1)$ で除した不偏分散 $\hat{\sigma}^2$ の平方根を取った標準偏差 $\hat{\sigma}$

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}, \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{\mu})^2}{n-1}} \quad (11.10)$$

に一致する。しかし、ロジスティック分布の形状パラメータ σ_{LGS} は、正規分布の σ_{NOR} と異なり、

$$\sigma_{NOR}^2 = \frac{\pi^2}{3} \sigma_{LGS}^2 \tag{11.11}$$

との関係が知られており、 σ_{LGS} は、

$$\sigma_{LGS}^2 = \frac{3}{\pi^2} \sigma_{NOR}^2 = 0.3040 \sigma^2, \quad \sigma_{LGS} = 0.5513 \sigma_{NOR} \tag{11.12}$$

と σ_{NOR} の 0.5513 倍となる。ただし、 σ_{LGS} は、 σ_{NOR} と同様にロジスティック分布の広がりやを規定するパラメータである。

表 11.2 に示すのは、表 11.1 の正規分布の σ_{NOR} と同じ $\sigma_{LGS} = 1.5$ の場合、正規分布と同様にするための $\sigma_{LGS} = 0.5513 \times 1.5 = 0.8270$ とした場合の 2 通りについての計算結果である。図 11.1 と図 11.2 を比べると、 $\sigma_{LGS} = 0.8270$ の場合には、正規分布と完全には一致しないがほぼ同様の形状となっている。このことが、正規分布をシグモイド曲線として使うのではなく、ロジスティック分布が広く使われるようになった理由である。

表 11.2 ロジスティック分布の 2 種の関数

$\mu_{LGS} =$	5		5	
	$\sigma_{LGS} = 1.5$		0.8270	
x	密度	累積	密度	累積
	$f_{LGS}(x)$	$F_{LGS}(x)$	$f_{LGS}(x)$	$F_{LGS}(x)$
0.0	0.0222	0.0344	0.0028	0.0024
1.0	0.0405	0.0650	0.0094	0.0079
2.0	0.0700	0.1192	0.0305	0.0259
3.0	0.1101	0.2086	0.0908	0.0818
4.0	0.1494	0.3392	0.2141	0.2298
5.0	0.1667	0.5000	0.3023	0.5000
6.0	0.1494	0.6608	0.2141	0.7702
7.0	0.1101	0.7914	0.0908	0.9182
8.0	0.0700	0.8808	0.0305	0.9741
9.0	0.0405	0.9350	0.0094	0.9921
10.0	0.0222	0.9656	0.0028	0.9976

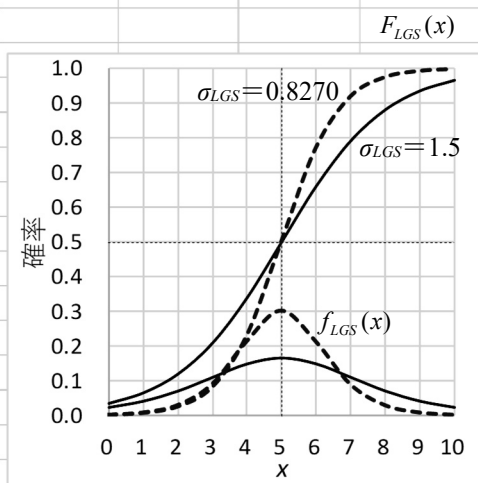


図 11.2 ロジスティック分布の 2 種の関数の例示

パラメータの線形化

式 (11.8) の、分母の指数関数 $\exp()$ の中の分母を分子を σ_{LGS} で除し、

$$F_{LGS}(x) = \frac{1}{1 + \exp\left[-\left(\frac{\mu_{LGS}}{\sigma_{LGS}} + \frac{1}{\sigma_{LGS}} x\right)\right]} \tag{11.13}$$

さらに, β_0 と β_1 で

$$\beta_0 = -\frac{\mu_{LGS}}{\sigma_{LGS}}, \quad \beta_1 = \frac{1}{\sigma_{LGS}} \quad (11.14)$$

と置き換えることにより,

$$F_{LGS}(x) = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x)]} \quad (11.15)$$

元のパラメータ (μ_{LGS} , σ_{LGS}) に代え線形式として表すことができる. ただし, このような線形式へのパラメータの置き換えによって, 得られたデータからロジスティック分布の位置 (中心位置) と形状 (分布の広がり) についてイメージできにくくなる. ただし, $(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$ のように多変数の線形式への拡張が容易に行なえるメリットがある. なお, 図 11.2 の $\sigma_{LGS} = 1.5$ 場合であれば, 線形化した場合のパラメータは,

$$\beta_0 = -\frac{\mu_{LGS}}{\sigma_{LGS}} = -\frac{5}{1.5} = -3.3333, \quad \beta_1 = \frac{1}{\sigma_{LGS}} = \frac{1}{1.5} = 0.6667$$

となるが, この数値からでは, どのような累積分布関数になるか, 見当が付きにくい. なお, β_1 の逆数が σ_{LGS} , 負の (β_0 / β_1) が μ_{LGS} となることから換算することが容易にできる.

$$\sigma_{LGS} = \frac{1}{\beta_1} = 1.5, \quad \mu_{LGS} = -\frac{\beta_0}{\beta_1} = 5.0 \quad (11.16)$$

反応が 2 値の場合に対するロジスティック回帰分析は, 式 (11.15) を, n_i 回の試行に対する反応ありの回数を y_i とし, 出現率を $p_i = y_i / n_i$ とした場合に累積分布関数 $F_{LGS}(x)$ は,

$$p_i = \frac{1}{1 + \exp[-(\beta_0 + \beta_1 x_i)]} + \varepsilon_i, \quad \varepsilon_i \sim 2\text{項分布} \quad (11.17)$$

となる. 誤差 ε_i を無視し両辺の逆数を取り, 1 を左辺に移行し, 対数を取ると

$$\left. \begin{aligned} \ln\left(\frac{1-p_i}{p_i}\right) &= \ln\{\exp[-(\beta_0 + \beta_1 x_i)]\} + \varepsilon'_i \\ \ln\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 x_i + \varepsilon'_i \end{aligned} \right\} \quad (11.18)$$

のように, いわゆる, ロジット $\ln[p_i / (1-p_i)]$ に関する線形式となり, ロジスティック回帰回帰分析として多変数への拡張が容易になり, 反応が 2 値データの場合の探索的な解析法として広く使われている. 厄介なのは, 誤差 ε'_i が, 位置パラメータ $\mu_{LGS} = -(\beta_0 / \beta_1)$ で最小となるが, それよりも小さい場合でも大きい場合でも発散してしまい, 誤差分布を特定できない. そのために反復重み付き回帰を適用する最尤法によるパラメータの推定が必要となる. 詳細は, ドブソン (2008), 高橋 (2017) を参照のこと.

累積ロジスティック分布関数の量的な反応への拡張

式 (11.17) を量的な反応に拡張する. 最小の反応を θ_{min} , 最大の反応を θ_{max} とし, 出現率 (0~1) の範囲の累積分布関数を $[0.0 \sim (\theta_{max} - \theta_{min})]$ の範囲に広げ, 最小の反応 θ_{min} の下駄をはかせて

$$y^{(LGS)} = \theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + \exp[-(\beta_0 + \beta_1 x)]} \quad (11.19)$$

のように拡張する. この式は, θ_{min} から θ_{max} の範囲のロジスティック曲線となり, 4パラメータモデルとも言われている. 最小の反応が 0.0 の場合は, θ_{min} がなくなり 3パラメータモデルとなる. 前章での式 (10.17) は, 最小値 θ_{min} に対し, (1, 1) 型のダミー変数 (a_1, a_2, a_3) を加え,

$$\theta_{min} = \theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3 \quad (11.20)$$

とし, 4パラメータモデルをさらに拡張した事例である.

式 (11.8) の μ_{LGS} を $\mu_{LGS}^{(k)}$ とし,

$$y^{(LGS)} = \theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + \exp\left[-\frac{(x - \mu_{LGS}^{(k)})}{\sigma_{LGS}}\right]} \quad (11.21)$$

$\mu_{LGS}^{(0)}$ を式 (11.20) と同様に (1, 1) 標示型ダミー変数を用いて

$$\mu_{LGS}^{(k)} = \mu_{LGS}^{(1)} a_1 + \mu_{LGS}^{(2)} a_2 + \mu_{LGS}^{(3)} a_3 \quad (11.22)$$

と拡張することにより, 3本の位置が異なる平行なロジスティック曲線の同時あてはめが行なえる. 第 13.4 節では, 2本の平行なロジスティック曲線の位置の差を求める事例を導入する.

酵素反応速度論のヒルの式

酵素反応速度論のヒルの式は,

$$V = \frac{V_{max} [S]^h}{[S]^h + K_{0.5}^h} \quad (11.23)$$

ただし, $[S]$ は基質濃度, V_{max} は基質濃度が無限大のときの反応速度

h はヒル係数, $K_{0.5}$ は (最大速度 V_{max} の半分の速度) を与える基質濃度

として定義されている. 詳しくは, 芳賀 (2016), 「医薬品開発のための統計解析, 第 3 部 非線形モデル 改訂版」の「第 1.4 節 E_{max} モデルとロジスティック回帰」を参照されたい.

ヒルの式を統計的な観点から, $V = y$, $V_{max} = \theta_{max}$, $[S] = x$, $K_{0.5} = \mu$, $h = \beta_1$ と置き換え,

$$V = \frac{V_{max} [S]^h}{[S]^h + K_{0.5}^h}, \quad y = \frac{\theta_{max} x^{\beta_1}}{x^{\beta_1} + \mu^{\beta_1}} \quad (11.24)$$

分子と分母を x^{β_1} で除し、指数と対数をセットで加え、整理すると、

$$\begin{aligned}
 y &= \frac{\theta_{max}}{1 + \frac{\mu^{\beta_1}}{x^{\beta_1}}} \\
 &= \frac{\theta_{max}}{1 + \exp\left[\ln\left(\frac{\mu^{\beta_1}}{x^{\beta_1}}\right)\right]} \\
 &= \frac{\theta_{max}}{1 + \exp\{-\beta_1 [\ln(x) - \ln(\mu)]\}}
 \end{aligned}
 \tag{11.25}$$

基質濃度の対数 $\ln(x)$ に対する 3 パラメータ・ロジスティック曲線式が得られる。この式は、 $K_{0.5} = \mu$, $V_{max} = \theta_{max}$ など、ヒルの式のパラメータを置き変えるだけなので、ヒルの式になじみがある人達には好まれている。

基質濃度 $[S]$ を 0.125 から倍々に、 $V_{max} = 10$, $K_{0.5} = 5$ を設定し、ヒル係数 $h = 0.5, 1.0, 2.0$ と変化させた場合の反応を V とした場合を図 11.3 に示す。基質濃度 $[S]$ を対数目盛にすると式 (11.25) に示すロジスティック回帰線となることが確認できる。

表 11.3 ヒルの式の形状の変化

V_{max}	20	20	20
$K_{0.5}$	5	5	5
h	0.5	1.0	2.0
$[S]$	V_1	V_2	V_3
0.001	0.279	0.004	0.000
0.125	2.731	0.488	0.012
0.250	3.655	0.952	0.050
0.50	4.805	1.818	0.198
1.0	6.180	3.333	0.769
2.0	7.749	5.714	2.759
4.0	9.443	8.889	7.805
8.0	11.170	12.308	14.382
16.0	12.829	15.238	18.221
32.0	14.334	17.297	19.523
64.0	15.631	18.551	19.879
128.0	16.699	19.248	19.970

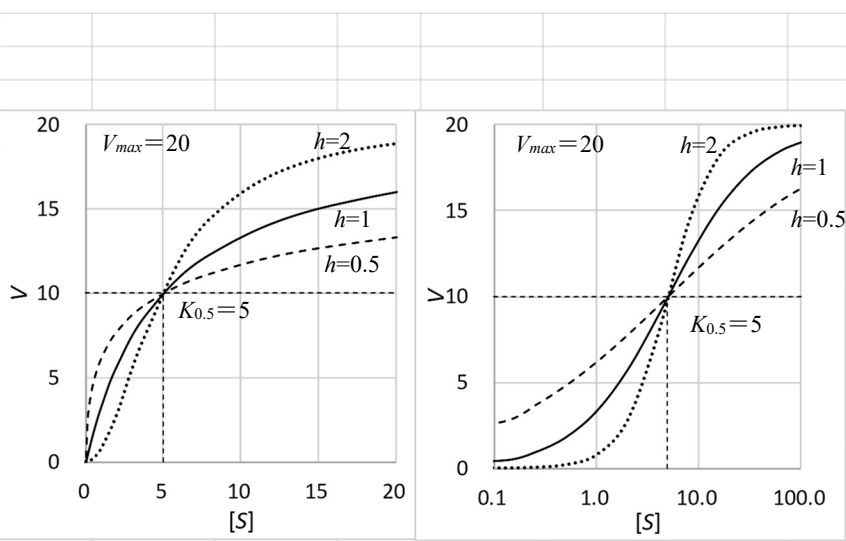


図 11.3 ヒルの式に対する対数目盛によるロジスティック曲線

線形化にこだわりがある人達には、式 (11.25) の 2 つのパラメータの積を $\beta_1 \ln(\mu) = -\beta_0$ に置き換えると、

$$y = \frac{\theta_{max}}{1 + \exp[-(\beta_0 + \beta_1 \ln(x))]}
 \tag{11.26}$$

x の対数 $\ln(x)$ に関して最小値を 0.0, 最大値を θ_{max} とする 3 パラメータ・ロジスティック回帰モデルに帰着する。

ロジスティック曲線を統計的に理解しやすくするためには、 $\beta_1 = 1/\sigma_{LGS}$ 、 $\beta_0 = -\ln(\mu_{LGS})/\sigma_{LGS}$ と置き換えることにより、

$$y = \frac{\theta_{max}}{1 + \exp\left[-\left(\frac{\ln(\mu_{LGS})}{\sigma_{LGS}} + \frac{1}{\sigma_{LGS}} \ln(x)\right)\right]} = \frac{\theta_{max}}{1 + \exp\left[-\left(\frac{\ln(x) - \ln(\mu_{LGS})}{\sigma_{LGS}}\right)\right]} \quad (11.27)$$

位置パラメータ $\ln(\mu_{LGS})$ と形状パラメータ $\ln(\sigma_{LGS})$ を用いた 3 パラメータモデルで表わすことができ、得られたデータの散布図などから、パラメータについておおよその見当が付けられるなどの利点がある。

酵素反応速度論でのヒルの式は、基質濃度 $[S]$ に対し対数を取ることで、ロジスティック曲線になることを示した。各種の薬理試験で用いられている *E_{max}* モデル式、受容体結合実験で用いられる式

$$y = \frac{E_{max} \cdot x^\gamma}{x^\gamma + D_{50}^\gamma}, \quad L_B = \frac{B_{max}[F_L]^n}{[F_L]^n + K_D^n} \quad (11.28)$$

なども式 (11.23) で示したヒルの式と同様にロジスティック曲線となる。ただし、式を見ただけでは、統計的な観点から、変数はどれなのか、推定したいパラメータはどれなのかが、統計的には判読がしにくい。ただし、よく見れば、どれもヒルの式に類似していることが認識されるであろう。ミカエリス・メンテン式

$$V = \frac{V_{max}[S]}{[S] + K_m} \quad (11.29)$$

は、式 (11.23) のヒル係数が $h=1.0$ の場合に帰着することも自明である。ミカエリス・メンテン式は、非線形モデルの典型例としても知られており、第 10.6 節で取り上げた Wikipedia の「ガウス・ニュートン法」事例でも取り上げられている。

ゴンペルツ・最大極値分布

生物の成長の経年変化、新製品の売り上げ数の年次変化、新しく作成された大規模なソフトウェアの累積バグ数の経時変化などは、ゴンペルツ曲線に従うことが経験的に知られている。ゴンペルツ曲線は、ある時点から急激に伸び始め、それが直線的になる時期を経て、次第に伸び悩みが始め、最終的には天井を打つようなシグモイド曲線である。

ゴンペルツ曲線の式には、様々な表記法

$$1) y = Kb^{e^{-cx}}, \quad 2) y = Ka^{b^x}, \quad 3) y = \alpha \exp[-\beta e^{-kx}], \quad 4) y = \exp[-\exp(a - bx)] \quad (11.30)$$

などが混在している。これは、異なる学問分野で定式化されたためであるが、これまで示してきた統計的な観点からは、分布の位置パラメータ μ_{GMP} 、形状パラメータ σ_{GMP} を用いた式、あるいは、線形式 ($\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$) を用いた式としたい。

これらの式 (11.30) は、互いに変換可能である。

- 1) 式で $e^{-c} = b$ とすれば 2) 式となり、
- 1) 式で $b e^{-cx} = \exp[\ln(b e^{-cx})]$ とすれば、3) 式となり、
- 3) 式で $\beta = e^a$ 、 $\alpha = 1$ とすれば、4) 式となる。

統計的な観点からは、4) 式に対し最小値を $a = \beta_0$ 、 $b = -\beta_1$ として、

$$y = \theta_{mim} + (\theta_{max} - \theta_{mim}) \cdot \exp[-\exp(\beta_0 + \beta_1 x)] \quad (11.31)$$

のように 2 重指数関数として扱える。または、これまでと同様に形状パラメータ σ_{GMP} を用いて $\beta_1 = 1/\sigma_{GMP}$ 、位置パラメータ μ_{GMP} に対し $\beta_0 = -\mu_{GMP}/\sigma_{GMP}$ と置き換え、

$$y^{GMP} = \theta_{mim} + (\theta_{max} - \theta_{mim}) \cdot \exp\left[-\exp\left(-\frac{x - \mu_{GMP}}{\sigma_{GMP}}\right)\right] \quad (11.32)$$

とすることで、パラメータの推定に関して見通しが良くなる。

ゴンペルツ曲線の最小値を $\theta_{mim} = 0$ 、最大値を $\theta_{max} = 1$ とすることで、最大極値分布 (Maximam Extreme Value Distribution) となり、累積分布関数 $F_{MEV}(x)$ は、 $\mu_{MEV} = \mu_{GMP}$ 、 $\sigma_{MEV} = \sigma_{GMP}$ と置き換え、

$$F_{MEV}(x) = \exp\left[-\exp\left(-\frac{x - \mu_{MEV}}{\sigma_{MEV}}\right)\right] \quad (11.33)$$

となる。累積分布関数 $F_{MEV}(x)$ を x で微分し、確率密度関数 $f_{MEV}(x)$ は、

$$f_{MEV}(x) = \frac{1}{\sigma_{MEV}} \exp\left(-\frac{x - \mu_{MEV}}{\sigma_{MEV}}\right) \exp\left[-\exp\left(-\frac{x - \mu_{MEV}}{\sigma_{MEV}}\right)\right] \quad (11.34)$$

のように導出される。

表 11.1 に示した正規分布、表 11.2 に示したロジスティック分布と同様に、表 11.4 にゴンペルツ・最大極値分布について、確率密度関数と累積分布関数を示す。明らかに異なるのは、累積分布の変曲点が、正規分布およびロジスティック分布の ($x=5$, $y=0.5$) から ($x=5$, $y=0.368$) となる。そのために確率密度関数 $f_{MEV}(x)$ が、右に大きく裾を引き、それに伴い累積分布関数が、急速に立ち上がり、徐々に最大値に収束するシグモイド曲線となっている。変曲点は、式 (11.33) で $x = \mu_{MEV}$ とした場合で、

$$F_{MEV}(x = \mu_{MEV}) = \exp[-\exp(-0)] = 0.368 \quad (11.35)$$

となる。

表 11.4 ゴンペルツ・最大極値分布

$\mu_{MEV} =$		5		5		
$\sigma_{MEV} =$		1.5		1.0		
		密度	累積		密度	累積
x	x'	$f_{MEV}(x)$	$F_{MEV}(x)$	x'	$f_{MEV}(x)$	$F_{MEV}(x)$
0.0	-3.33	0.000	0.000	-5.00	0.000	0.000
1.0	-2.67	0.000	0.000	-4.00	0.000	0.000
2.0	-2.00	0.003	0.001	-3.00	0.000	0.000
3.0	-1.33	0.057	0.023	-2.00	0.005	0.001
4.0	-0.67	0.185	0.143	-1.00	0.179	0.066
5.0	0.00	0.245	0.368	0.00	0.368	0.368
6.0	0.67	0.205	0.598	1.00	0.255	0.692
7.0	1.33	0.135	0.768	2.00	0.118	0.873
8.0	2.00	0.079	0.873	3.00	0.047	0.951
9.0	2.67	0.043	0.933	4.00	0.018	0.982
10.0	3.33	0.023	0.965	5.00	0.007	0.993

$$x' = (x - \mu_{MEV}) / \sigma_{MEV}$$

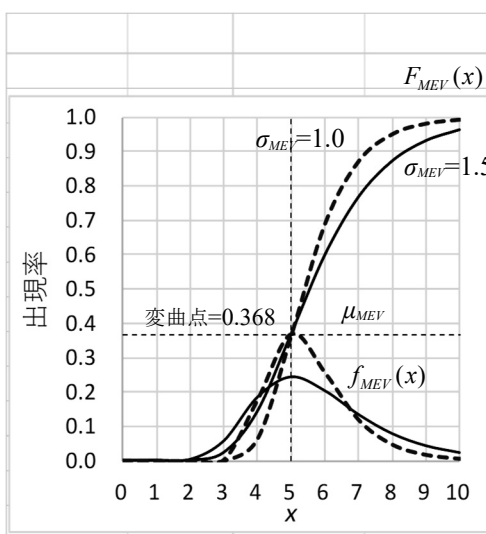


図 11.4 最大極値分布

ワイブル・最小極値分布

ワイブル分布は、故障あるいは寿命データの解析で標準的に用いられているが、一般化線形モデルでの 2 値反応に対する解析法として、プロビット法、ロジット法などと同様に「補 2 重対数法」のための分布としても知られている。ワイブル分布に基づくシグモイド曲線は、ゴンペルツ曲線とは対照的に、 x が小さい場合には反応がゆっくりと上昇し、 x が大きくなるにつれて急激に上昇するシグモイド曲線となる。

ワイブル分布の累積分布関数 $F_{WBL}(x)$ は、 x を寿命時間とし、

$$F_{WBL}(x) = 1 - \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\} \quad (11.36)$$

で与えられる。ゴンペルツ曲線の場合と同様に、関数の形式はまちまちであり、Excel の関数 Weibull.dist () のように、パラメータ (α, β) が逆となる場合もある。日本の信頼性工学分野では、 x を t (time) として

$$F_{WBL}(t) = 1 - \exp\left\{-\left(\frac{t}{\eta}\right)^m\right\} \quad (11.37)$$

のようにパラメータに m と η が使われ、 m を形状パラメータ、 η を尺度パラメータとしている。さらに (尺度・形状・位置など) のパラメータも乱れて使われているので、ここでは深入りしない。累積分布関数 $F_{WBL}(x)$ を x で微分して確率密度関数 $f_{WBL}(x)$

$$f_{WBL}(x) = \frac{\beta}{\alpha^\beta} x^{\beta-1} \exp\left\{-\left(\frac{x}{\alpha}\right)^\beta\right\} \quad (11.38)$$

が得られる。

ワイブル分布の $F_{WBL}(x)$ の α を $\alpha = \exp(\mu_{SEV})$, β を $\beta = 1/\sigma_{SEV}$ とすることにより, 最小極値分布 (SEV: Smallest Extreme Value Distribution) の関数形式 $F_{SEV}(\ln(x))$ に変換することができる. さらに, 指数と対数をセットで加えることにより,

$$\begin{aligned}
 F_{SEV}(\ln(x)) &= 1 - \exp \left\{ -\exp \left[\ln \left(\frac{x}{\exp(\mu_{SEV})} \right)^{1/\sigma_{SEV}} \right] \right\} \\
 &= 1 - \exp \left[-\exp \left(\frac{\ln(x) - \mu_{SEV}}{\sigma_{SEV}} \right) \right]
 \end{aligned}
 \tag{11.39}$$

を得る.

ワイブル分布と最小極値分布の関係は, 対数正規分布と正規分布の関係と同様である. このような対数変換を行なうことにより, ワイブル分布を正規分布と同様に, 位置パラメータ μ_{SEV} を“平均値”として, 形状パラメータ σ_{SEV} を“標準偏差”として用いて表すことができる.

最小極値分布の確率密度関数 $f_{SEV}(y)$ は, 累積分布関数 $F_{SEV}(x)$ を $\ln(x)$ で微分することにより,

$$f_{SEV}(\ln(x)) = \frac{1}{\sigma_{SEV}} \cdot \exp \left(\frac{\ln(x) - \mu_{SEV}}{\sigma_{SEV}} \right) \cdot \exp \left\{ -\exp \left(\frac{\ln(x) - \mu_{SEV}}{\sigma_{SEV}} \right) \right\}
 \tag{11.40}$$

として得られる. 位置パラメータ μ_{SEV} を 5, 形状パラメータ σ_{SEV} を (1.5, 1.0) と変えたときの確率密度関数および累積分布関数を表 11.5 に示す.

表 11.5 ワイブル・最小極値分布

$\mu_{SEV} =$		5		5		
$\sigma_{SEV} =$		1.5		1.0		
		密度	累積	密度	累積	
$\ln(x)$	$\ln x'$	$f_{SEV}(x)$	$F_{SEV}(x)$	$\ln x'$	$f_{SEV}(x)$	$F_{SEV}(x)$
0.0	-3.33	0.023	0.035	-5.00	0.007	0.007
1.0	-2.67	0.043	0.067	-4.00	0.018	0.018
2.0	-2.00	0.079	0.127	-3.00	0.047	0.049
3.0	-1.33	0.135	0.232	-2.00	0.118	0.127
4.0	-0.67	0.205	0.402	-1.00	0.255	0.308
5.0	0.00	0.245	0.632	0.00	0.368	0.632
6.0	0.67	0.185	0.857	1.00	0.179	0.934
7.0	1.33	0.057	0.977	2.00	0.005	0.999
8.0	2.00	0.003	0.999	3.00	0.000	1.000
9.0	2.67	0.000	1.000	4.00	0.000	1.000
10.0	3.33	0.000	1.000	5.00	0.000	1.000

$$\ln(x') = [\ln(x) - \mu_{SEV}] / \sigma_{SEV}$$

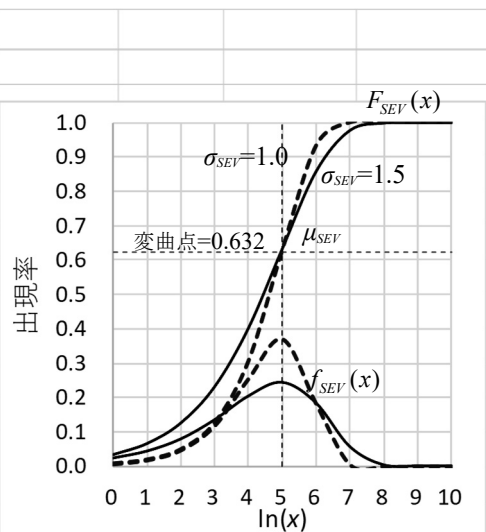


図 11.5 最小極値分布

式 (11.38) のワイブル分布の確率密度関数の形状は、形状パラメータ $\beta \leq 1$ の場合に単調減少、 $\beta > 1.0$ の場合は右に裾を引く分布で一山型ではあるがその位置が一定ではない。なお、ワイブル分布の累積分布関数は、 $\beta \leq 1$ の場合に上に凸の単調増加、 $\beta > 1.0$ の場合に S 字型となる。図 11.6 に示すのは、式 (11.38) のワイブル分布の確率密度関数 $f_{WBL}(x)$ 、式 (11.36) の累積分布関数 $F_{WBL}(x)$ で $\alpha = 10.0$ と固定し、 β を (0.5, 1.0, 2, 4, 6) と変化させた場合の例示である。この様な多彩な形状の変化が、寿命データの解析に貢献している。

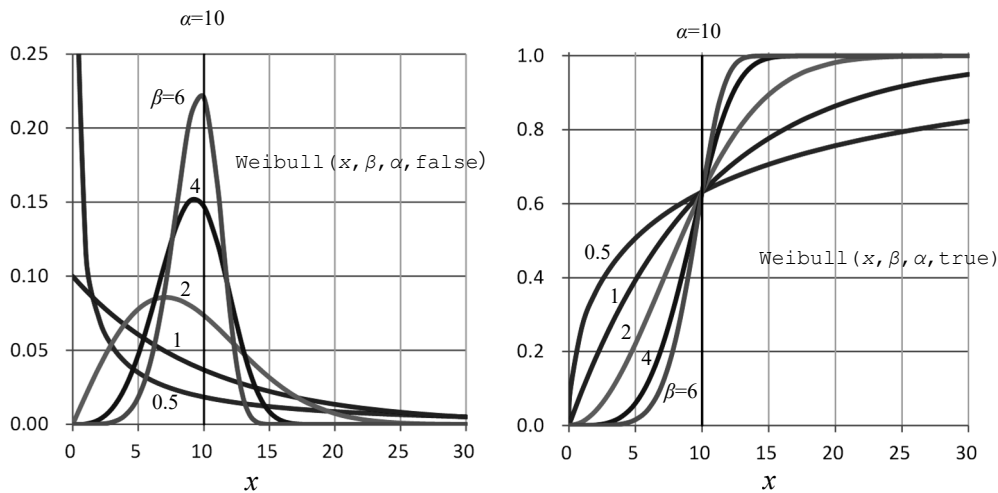


図 11.6 ワイブル分布の確率密度関数 (左) と累積分布関数 (右)

ワイブル分布に対し、最小極値分布は、パラメータ μ_{SEV} および σ_{SEV} に関して正規分布と同様の性質を持つために、寿命データの解析に際してワイブル分布のままでは推定が困難な各種の信頼区間の推定に際して黒子的な役割を果している。図 11.5 の最小極値分布の $\sigma_{SEV} = 1.0$ の場合の累積分布関数は、シグモイド曲線となっているが、 $\beta = 1/\sigma_{SEV}$ の関係から $\beta = 1.0$ であり、図 11.6 右のワイブル分布の累積分布関数は、単調増加となっていて形状が異なる。図 11.6 の X 軸を対数目盛にすると、確率密度関数は、 $\mu = \ln(\alpha)$ をピークとする一山型となり、累積分布関数は、 $\mu = \ln(\alpha)$ を変曲点とするシグモイド曲線となる。詳しくは、ネルソン著・奥野監訳 (1988)、「寿命データの解析」、高橋 (2015)、「第 4 回 続高橋セミナー：寿命試験データの統計解析」を参照されたい。

11.2. ロジスティック・ゴンペルツ・ワイブル曲線のあてはめ

山田・吉田・高橋 (2003), 「陰性および陽性対照があるシグモイド曲線—ダミー変数を持つ非線型回帰モデルの応用—」の事例で表 11.6 に示すような「環境ホルモン EE (ethinyl estradiol) 投与後のラット子宮重量」が取り上げられている。データは、共同研究の施設ごとの平均値が示されている。

表 11.6 環境ホルモン EE (ethinyl estradiol) 投与後のラットの子宮重量 (blotted uterine)

施設 番号	Vehicle	ethinyl estradiol (EE), $\mu\text{g}/\text{kg}$						
		0.01	0.03	0.1	0.3	1	3	10
1	102.35	95	105	112.22	190.45	319.78	373.72	382.00
2	120.82	115	115	123.47	217.48	351.32	384.72	404.32
3	115.92	115	120	144.42	213.95	326.07	378.37	354.37
7	121.62	120	125	131.25	220.83	317.52	387.43	391.67
8	79.22	90	80	105.08	211.13	287.68	262.20	273.73
9	108.47	115	115	123.60	211.37	357.57	353.82	362.05
11	82.45	100	100	113.38	191.23	297.67	307.60	312.40
18	89.25	90	90	91.80	193.07	334.95	334.48	366.20
19	99.17	100	100	83.17	104.67	135.17	234.17	332.67
平均	102.14	104.44	105.56	114.27	194.91	303.08	335.17	353.27

0.01, 0.03群はグラフから読み取り, 0.03, 0.3, 3 群は, 0.0316, 0.316, 3.16 の略表示

Kano, J., Onyon, L., Haseman, J., et al. (2001). The OECD program to validate the rat uterotrophic bioassay to screen compounds for in vivo estrogenic responses: phase 1, Environmental Health Perspectives, 109(8);(785-794). Table 5.

この事例に対し、山田ら (2003) では、累積ロジスティック分布関数を拡張したロジスティック曲線が適用されているが、図 11.7 に示すように用量反応関係を平均値で概観すると、 $0.1\mu\text{g}/\text{kg}$ から $0.3\mu\text{g}/\text{kg}$ にかけて子宮重量が急激に増加し、 $1.0\mu\text{g}/\text{kg}$ 以上では緩やかな子宮重

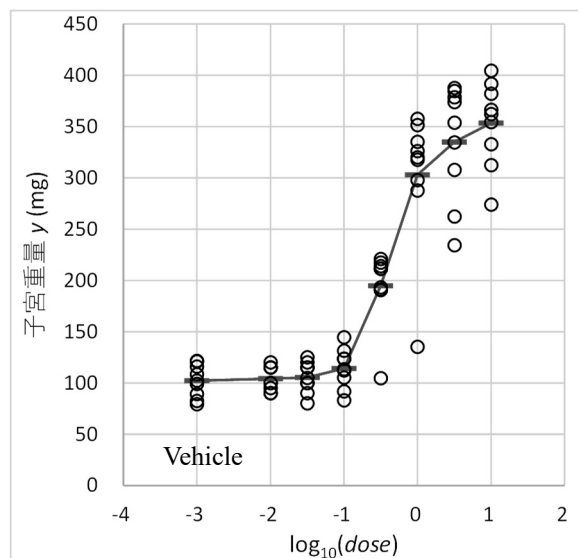


図 11.7 子宮重量の変化

量の増加となっており，ゴンペル曲線（最大極値分布）のあてはめが適切と思われる．Excel のソルバーを用いて 3 種のシグモイド曲線をあてはめ，それぞれの残差平方和の大きさで，あてはまりの性能評価を行なう．

3 種のシグモイド曲線のあてはめ

図 11.7 から，Vehicle 群の平均値は，低用量群の平均値の延長線上にあると判断されるので，用量を 2 段階落とした $dose = 0.001 \mu\text{g}/\text{kg}$ として解析する．常用対数用量 $x = \log_{10}(dose)$ に対して，3 種のシグモイド曲線をあてはめ，性能評価を行なう．

$$\text{ロジスティック曲線: } y_i^{LGS} = \theta_{min}^{LGS} + \frac{\theta_{max}^{LGS} - \theta_{min}^{LGS}}{1 + \exp\left[-\frac{(x_i - \mu_{LGS})}{\sigma_{LGS}}\right]} + \varepsilon_i \quad (11.41)$$

$$\text{ゴンペルツ曲線: } y_i^{MEV} = \theta_{min}^{MEV} + (\theta_{max}^{MEV} - \theta_{min}^{MEV}) \cdot \exp\left[-\exp\left(-\frac{x_i - \mu_{MEV}}{\sigma_{MEV}}\right)\right] + \varepsilon_i \quad (11.42)$$

最大極値

$$\text{対数ワイブル曲線: } y_i^{SEV} = \theta_{min}^{SEV} + (\theta_{max}^{SEV} - \theta_{min}^{SEV}) \cdot \left\{1 - \exp\left[-\exp\left(\frac{x_i - \mu_{SEV}}{\sigma_{SEV}}\right)\right]\right\} + \varepsilon_i \quad (11.43)$$

最小極値

表 11.7 左には，表 11.6 の Vehicle を含め 8 用量群×9 施設=72 のデータを行方向に展開し，表 11.7 右には，グラフ表示のために $dose$ 間隔を細かく設定し，滑らかなシグモイド曲線が

表 11.7 子宮重量データに対する 3 種のシグモイド曲線のあてはめ

			MEV	LGS	SEV						
初期値	350	$\theta_{max}^{\wedge} =$	355.3908	348.4339	343.6261						
	100	$\theta_{min}^{\wedge} =$	105.1279	101.9131	99.5568	MEV	ゴンペルツ・最大極値曲線				
	-0.5	$\mu^{\wedge} =$	-0.5005	-0.3677	-0.2209	LGS	ロジスティック曲線				
	0.5	$\sigma^{\wedge} =$	0.3657	0.2460	0.3510	SEV	ワイブル・最小極値曲線				
残差平方和		$S_e =$	93259.73	93602.41	95712.09						
No.	dose	x	y	$y^{\wedge MEV}$	$y^{\wedge LGS}$	$y^{\wedge SEV}$	dose	x	$y^{\wedge MEV}$	$y^{\wedge LGS}$	$y^{\wedge SEV}$
1	0.001	-3.0	102.35	105.13	101.92	99.65	0.001	-3.00	105.13	101.92	99.65
2	0.001	-3.0	120.82	105.13	101.92	99.65	0.01	-2.00	105.13	102.24	101.09
:							0.0316	-1.50	105.13	104.36	105.85
9	0.001	-3.0	99.17	105.13	101.92	99.65		-1.25	105.23	108.55	112.22
10	0.01	-2.0	95.00	105.13	102.24	101.09	0.1	-1.00	110.10	119.43	124.69
11	0.01	-2.0	115.00	105.13	102.24	101.09		-0.75	139.74	144.93	148.05
:							0.316	-0.50	197.24	192.73	188.17
:								-0.25	256.29	254.12	246.40
63	3	0.5	234.17	339.66	341.39	343.53	1	0.00	299.16	303.28	306.24
64	10	1.0	382.00	351.29	347.49	343.63		0.25	325.22	329.93	338.30
65	10	1.0	404.32	351.29	347.49	343.63	3.16	0.50	339.66	341.39	343.53
:								0.75	347.33	345.84	343.63
71	10	1.0	366.20	351.29	347.49	343.63	10	1.00	351.29	347.49	343.63
72	10	1.0	332.67	351.29	347.49	343.63		1.50	354.34	348.31	343.63
dose の 0.03, 0.3, 3 は, 0.0316, 0.316, 3.16 の略表示								2.00	355.12	348.42	343.63

得られるように 3 種のシグモイド曲線の推定値を計算した結果が示されている。初期値は、図 11.7 から推定されるパラメータとして共通の $\hat{\theta}_{max} = 350$, $\theta_{min} = 100$, $\hat{\mu} = -0.5$, $\hat{\sigma} = 0.5$ とした。この初期値をそれぞれの曲線のパラメータの欄にコピー&ペーストし、Excel のソルバーで残差平方和 S_e が最小になるようにパラメータを変化させた結果が示されている。残差平方和 S_e は、 $\{=\text{SumSq}(\mathbf{y}$ の範囲 $-\mathbf{y}^{\wedge}$ の範囲) $\}$ のように反応変数 y の列ベクトルとそれぞれの曲線の推定値 \hat{y} の列ベクトルの差を引数にして求めている。なお、計算式の外側の $\{ \}$ は、行列計算として扱われているために自動的に付与される。

Excel のソルバーで誤差平方和 S_e を最小化した結果、

<i>MEV</i>	ゴンペルツ・最大極値曲線	$S_e^{MEV} = 93,259.73$	-342.69
<i>LGS</i>	ロジスティック曲線	$S_e^{LGS} = 93,602.41$	規準
<i>SEV</i>	ワイブル・最小極値曲線	$S_e^{SEV} = 95,712.09$	+2109.68

に示すように、ゴンペルツ・最大極値曲線が、ロジスティック曲線に対し、残差平方和 S_e が -342.69 と減少し、あてはまりが良くなり、ワイブル・最小極値曲線は、逆に +2109.68 と増加し、あてはまりが悪くなっている。

図 11.8 に示すのは、表 11.7 の対数用量 x と子宮重量 y の散布図を描き、その上に 3 本のシグモイド曲線を重ね書きした結果である。ゴンペルツ・最大極値曲線は、 $x = \log_{10}(0.10) = -1$ を過ぎてから急速に立ち上がり、 $x = \log_{10}(1.0) = 0$ からは、他の曲線に比べ緩やかに上昇している。この実験は、環境ホルモンが生態系に与える影響が、どのくらいの *dose* から起きるのかを見極めて規制のための *dose* を推定することを目的としている。あてはめたシグモイド曲線から、最大反応と最小反応 ($\theta_{max} - \theta_{min}$) の 10 分の 1 の推定値の 95%信頼区間の下限などが、規制用量の目安とされている。図 11.8 では、おおよそ $y = 120$ mg あたりで 3 本のシグモイド曲線の位置が大きく開いている。このことから、実験結果に最もあてはまるシグモイド曲線を選択する必要がある、統計的にもロジスティック曲線ではなく、ゴンペルツ・最大極値曲線が、尤もあてはまりが良いとの推論される。

10 パーセント反応の用量 D_{10} の逆推定

シグモイド曲線をあてはめて最大反応の半分となる用量 *dose* の推定は、急性毒性試験の 2 値反応データに対するプロビット法による 50%致死量 LD_{50} (*Lethal Dose 50*) の推定が、古くから行なわれてきた。表 11.3 に示したように、酵素反応速度論で使われているヒルの式には、パラメータとして 50%反応となる用量が $K_{0.5}$ として直接推定される。また、ヒルの式を

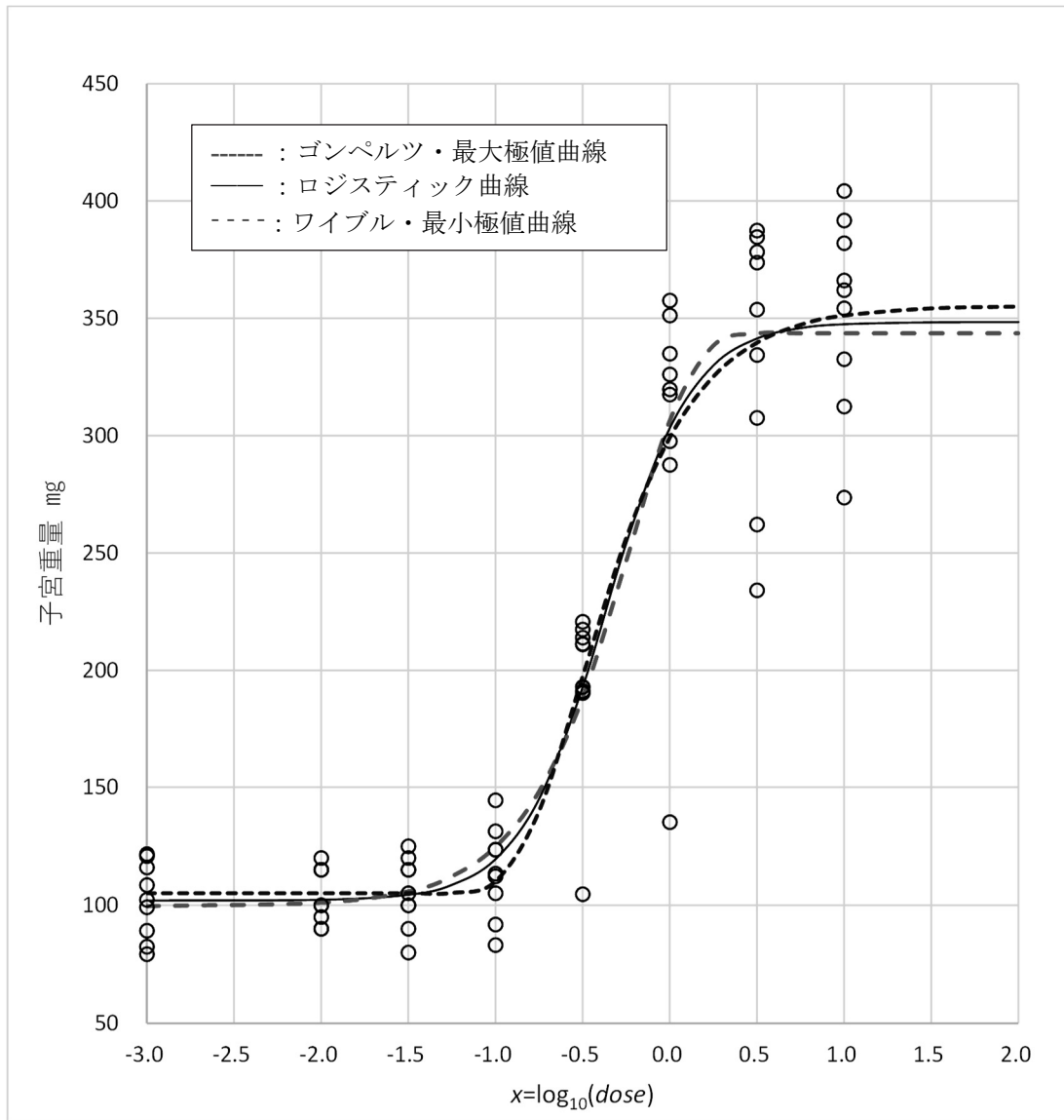


図 11.8 子宮重量の変化に対する各種のシグモイド曲線のあてはめ

ロジスティック曲線に変換した場合にも 50%反応となる用量 D_{50} が表 11.7 上段の *LGS* 欄に示すように位置パラメータの推定値 $\hat{\mu}_{LGS} = -0.3677$ ($dose = e^{-0.3677} = 0.69$ mg) として得られる。環境ホルモンなど環境汚染物質の場合は、規制用量の設定のために 10%反応となる用量 D_{10} などの推定が求められている。詳しくは、広瀬ら (2021), 「ベンチマークドース手法の適用の現状と課題 —動物実験データへの適用を中心に—」を参照のこと。

D_{10} を推定するためには、式 (11.41), 式 (11.42), 式 (11.43) において, $\pi_0^{\text{分布}} = 0.10$ となる反応を

$$y_{D_{10}}^{\text{分布}} = \theta_{\min}^{\text{分布}} + (\theta_{\max}^{\text{分布}} - \theta_{\min}^{\text{分布}}) \times 0.10 \quad (11.44)$$

とにおいて、用量 x について解くことによって推定式が得られる。

$$\text{ロジスティック曲線: } x_{D_{10}}^{LGS} = -\ln\left(\frac{\theta_{max}^{LGS} - \theta_{min}^{LGS}}{y_{D_{10}}^{LGS} - \theta_{min}^{LGS}} - 1\right) \sigma_{LGS} + \mu_{LGS} \quad (11.45)$$

$$\text{ゴンペルツ曲線: } x_{D_{10}}^{MEV} = -\ln\left(-\ln\frac{y_{D_{10}}^{MEV} - \theta_{min}^{MEV}}{\theta_{max}^{MEV} - \theta_{min}^{MEV}}\right) \sigma_{MEV} + \mu_{MEV} \quad (11.46)$$

最大極値

$$\text{対数ワイブル曲線: } x_{D_{10}}^{SEV} = \ln\left\{-\ln\left[-\left(\frac{y_{D_{10}}^{SEV} - \theta_{min}^{SEV}}{\theta_{max}^{SEV} - \theta_{min}^{SEV}} - 1\right)\right]\right\} \sigma_{SEV} + \mu_{SEV} \quad (11.47)$$

最小極値

これらの逆推定の式を用いて Excel シート上で D_{10} を計算した結果を表 11.8 に示す。元の用量に 10 の冪乗で戻した結果は、あてはまりの順にゴンペルツ曲線が $0.156 \mu\text{g/kg}$ 、ロジスティック曲線が $0.124 \mu\text{g/kg}$ 、ワイブル・最小極値曲線が $0.098 \mu\text{g/kg}$ となる。この実験の目的は、環境ホルモン EE (ethinyl estradiol) の規制用量の推定であり、3 種のシグモイド曲線のあてはめにより、 D_{10} の推定用量が大きく異なる。図 11.8 から 50%反応 D_{50} ならば、違いは線の幅程度と小さいので気にすることはないが、 D_{10} の場合には、残差平方和の最も小さいゴンペルツ曲線・最大極値分布の $D_{10}^{(MEV)} = 0.156 \mu\text{g/kg}$ を選ぶべきである。図 11.7 の元のデータの平均値の推移の形状からもゴンペルツ・最大極値曲線の選択の妥当性は揺るがない。

表 11.8 3 種のシグモイド曲線に対する 10 パーセント反応となる用量の推定

	最大	最小	位置	形状	D_{10}	D_{10} となる $\log_{10}(\text{dose})$			dose ($\mu\text{g/kg}$)
	θ_{max}	θ_{min}	μ^{\wedge}	σ^{\wedge}	$y^{\wedge}_{D_{10}}$	x^{\wedge}_{MEV}	x^{\wedge}_{LGS}	x^{\wedge}_{SEV}	
ゴンペルツ・最大極値 MEV	355.4	105.1	-0.500	0.366	130.2	-0.806			0.156
ロジスティック LGS	348.4	101.9	-0.368	0.246	126.6		-0.908		0.124
ワイブル・最小極値 SEV	343.6	99.6	-0.221	0.351	124.0			-1.011	0.098
個別の $y^{\wedge}_{D_{10}} = \theta_{min}^{\wedge} + (\theta_{max}^{\wedge} - \theta_{min}^{\wedge}) * 0.10$				$\pi_0 =$	0.10				

50 パーセント反応の用量 D_{50} の逆推定

ロジスティック曲線のあてはめでは、 $\hat{\mu}_{LGS} = -0.3677$ が D_{50} の推定値として得られるのであるが、ゴンペルツ・最大極値曲線の場合は、 $\hat{\mu}_{MEV}$ の推定値が 36.8%点となり、対数ワイブル・最小極値曲線の場合は、 $\hat{\mu}_{SEV}$ の推定値が 63.2%点となるので、表 11.8 に示したように別途計算する必要がある。表 11.9 に示すように、それぞれの曲線に対する $\hat{y}_{D_{50}}^{\text{分布}}$ を推定し、式

表 11.9 3 種のシグモイド曲線に対する 50 パーセント反応となる用量の推定

	最大	最小	位置	形状	D_{50}	D_{50} となる $\log_{10}(\text{dose})$			dose ($\mu\text{g/kg}$)
	θ_{max}	θ_{min}	μ^{\wedge}	σ^{\wedge}	$y^{\wedge}_{D_{50}}$	x^{\wedge}_{MEV}	x^{\wedge}_{LGS}	x^{\wedge}_{SEV}	
ゴンペルツ・最大極値 MEV	355.4	105.1	-0.500	0.366	230.3	-0.366			0.430
ロジスティック LGS	348.4	101.9	-0.368	0.246	225.2		-0.368		0.429
ワイブル・最小極値 SEV	343.6	99.6	-0.221	0.351	221.6			-0.350	0.447
個別の $y^{\wedge}_{D_{50}} = \theta_{min}^{\wedge} + (\theta_{max}^{\wedge} - \theta_{min}^{\wedge}) * 0.50$				$\pi_0 =$	0.50				

(11.45), 式 (11.46), 式 (11.47) の $\hat{y}_{D_{10}}^{\text{分布}}$ を $\hat{y}_{D_{50}}^{\text{分布}}$ に代えて D_{50} の推定値 $\hat{x}_{D_{50}}^{\text{分布}}$ を求めている. ロジスティック曲線の場合は, 計算結果と $\hat{\mu}_{LGS}$ と $\hat{y}_{D_{50}}^{(LGS)} = -0.368$ と一致していることが確かめられる.

10 パーセント反応量に対する逆推定値の用量 D_{10} は, あてはめるシグモイド曲線の種類によって $\hat{D}_{10}^{MEV} = 0.156 \text{ mg/kg}$, $\hat{D}_{10}^{LGS} = 0.124 \text{ mg/kg}$, $\hat{D}_{10}^{SEV} = 0.098 \text{ mg/kg}$ とかなり異なる. ただし, 50 パーセント反応量に対する逆推定値の用量 D_{50} は, $\hat{D}_{50}^{MEV} = 0.430 \text{ mg/kg}$ と $\hat{D}_{50}^{LGS} = 0.429 \text{ mg/kg}$ は, ほとんど同じ用量であり, $\hat{D}_{50}^{SEV} = 0.447 \text{ mg/kg}$ についても 0.018 mg/kg の差があるが, 相対的には D_{10} の場合に比べて小さい. 図 11.8 のおおよその 50 パーセント反応である $y = 225 \text{ mg}$ における各シグモイド曲線の位置の違いは, 線の太さ程度の違いである. このように 50 パーセント反応の D_{50} の推定ならば, あてはめるシグモイド曲線の選択について神経質になる必要なく, ロジスティック曲線を標準的に使えばよいことがわかる.

正規分布 vs. ロジスティック分布

図 11.1 に示した正規分布と図 11.2 に示したロジスティック分布を比較すると若干の差異があるかのように見える. 表 11.7 に正規(分布)曲線のあてはめを追加し, ロジスティック曲線を重ね書きした結果を図 11.9 に示す. ほとんど全ての範囲で線の太さ程度の違いであり, 正規(分布)曲線とロジスティック曲線どちらを使っても本質的な差はない.

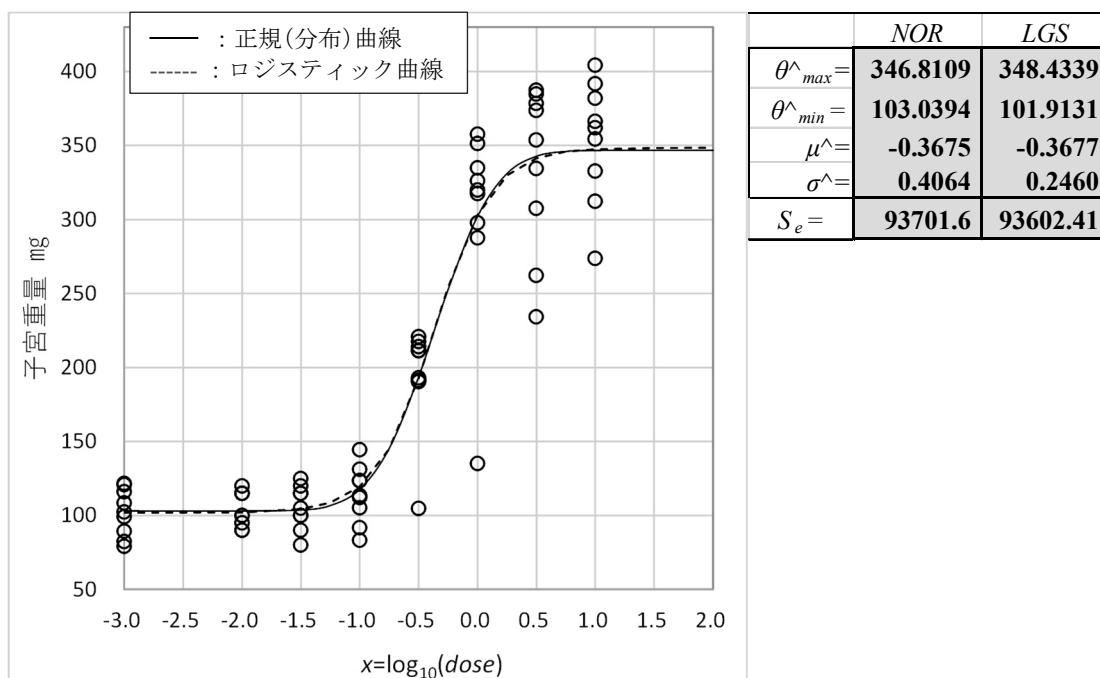


図 11.9 子宮重量の変化に対する正規(分布)曲線とロジスティック曲線のあてはめ

11.3. ゴンペルツ・最大極値曲線に対する各種の95%信頼区間

残差線形化法によるパラメータの推定

前節では、子宮重量データに3種のシグモイド曲線をあてはめ、ゴンペルツ・最大極値曲線が最もあてはまりが良いことを示した。10%反応量 D_{10} の95%信頼区間を推定するためには、第10.3章でロジスティック曲線に対する95%信頼区間の推定で示したように、ゴンペルツ・最大極値曲線の式のパラメータについて偏微分した式による微分係数の行列 \mathbf{Z} が必要となる。式(11.42)で示したゴンペルツ・最大極値曲線の式は、

$$y_i^{MEV} = \theta_{min}^{MEV} + (\theta_{max}^{MEV} - \theta_{min}^{MEV}) \cdot \exp\left[-\exp\left(-\frac{x_i - \mu_{MEV}}{\sigma_{MEV}}\right)\right] + \varepsilon_i$$

であり、偏微分を簡単化するために

$$\left. \begin{aligned} g(x_i) &= \exp\left(-\frac{x_i - \mu_{MEV}}{\sigma_{MEV}}\right) \\ \frac{\partial g(x_i)}{\partial \mu_{MEV}} &= \frac{g(x_i)}{\sigma_{MEV}} \\ \frac{\partial g(x_i)}{\partial \sigma_{MEV}} &= -\frac{(x_i - \mu_{MEV}) \cdot g(x_i)}{\sigma_{MEV}^2} \end{aligned} \right\} \quad (11.48)$$

とすると見通しがよくなり、パラメータに関する偏微分式、

$$\left. \begin{aligned} z_{1,i} &: \frac{\partial y}{\partial \theta_{max}^{MEV}} = \exp(-g(x_i)) \\ z_{2,i} &: \frac{\partial y}{\partial \theta_{min}^{MEV}} = 1 - \exp(-g(x_i)) \\ z_{3,i} &: \frac{\partial y}{\partial \mu_{MEV}} = -\frac{g(x_i)}{\sigma_{MEV}} \cdot (\theta_{max}^{MEV} - \theta_{min}^{MEV}) \cdot \exp(-g(x_i)) \\ z_{4,i} &: \frac{\partial y}{\partial \sigma_{MEV}} = -\frac{(x_i - \mu_{MEV}) \cdot g(x_i)}{\sigma_{MEV}^2} \cdot (\theta_{max}^{MEV} - \theta_{min}^{MEV}) \cdot \exp(-g(x_i)) \end{aligned} \right\} \quad (11.49)$$

が得られる。

表11.10に示すのは、表11.7に示した3種のシグモイド曲線のあてはめからゴンペルツ・最大極値曲線の式 \hat{y}^{MEV} を抜き出し、残差 $\varepsilon_i = y_i - \hat{y}_i^{MEV}$ 、 $g(x_i)$ 、 $z_{1,i}$ 、 $z_{2,i}$ 、 $z_{3,i}$ 、 $z_{4,i}$ を加えている。表11.10右上段に、残差線形化法によるパラメータの推定に必要な微分係数の行列 \mathbf{Z} による $(\mathbf{Z}^T \mathbf{Z})^{-1}$ の計算結果が示され、残差の列ベクトル $\hat{\varepsilon}$ を用い、増減パラメータ $\hat{\delta}^{(m-1)}$ の推定のための回帰式

$$\hat{\delta}^{(m-1)} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\varepsilon} \quad (11.50)$$

表 11.10 Excel による残差線形化法によるゴンペルツ・最大極値曲線のパラメータ推定

		$\xi^{(0)}$			$\xi^{(m-1)}$	$\delta^{(m-1)}$	$\xi^{(m)}$	$(Z^T Z)^{-1}$		
	初期値	350	$\theta^{max} =$	355.3920	0.0000	355.3920	0.1338	-0.0095	0.0002	0.0006
		100	$\theta^{min} =$	105.1279	0.0000	105.1279	-0.0095	0.0322	0.0001	-0.0001
		-0.5	$\mu^{MEV} =$	-0.5005	0.0000	-0.5005	0.0002	0.0001	2.08E-06	-1.81E-07
		0.5	$\sigma^{MEV} =$	0.3657	0.0000	0.3657	0.0006	-0.0001	-1.81E-07	5.66E-06
	残差平方和	$S_e =$		93259.73	-----微係数行列 Z-----					
No.	dose	x	y	y^{MEV}	ε^{\wedge}	$g(x)$	$\partial_{\theta^{max}}$	$\partial_{\theta^{min}}$	$\partial_{\mu^{MEV}}$	$\partial_{\sigma^{MEV}}$
1	0.001	-3.0	102.35	105.13	-2.78	929.144	0.0000	1.0000	0.0000	0.0000
2	0.001	-3.0	120.82	105.13	15.69	929.144	0.0000	1.0000	0.0000	0.0000
:										
9	0.001	-3.0	99.17	105.13	-5.96	929.144	0.0000	1.0000	0.0000	0.0000
10	0.01	-2.0	95.00	105.13	-10.13	60.342	0.0000	1.0000	0.0000	0.0000
11	0.01	-2.0	115.00	105.13	9.87	60.342	0.0000	1.0000	0.0000	0.0000
:										
:										
63	3	0.5	234.17	339.66	-105.49	0.065	0.9371	0.0629	-41.627	-113.84
64	10	1.0	382.00	351.29	30.71	0.017	0.9836	0.0164	-11.125	-45.641
65	10	1.0	404.32	351.29	53.03	0.017	0.9836	0.0164	-11.125	-45.641
:										
71	10	1.0	366.20	351.29	14.91	0.017	0.9836	0.0164	-11.125	-45.641
72	10	1.0	332.67	351.29	-18.62	0.017	0.9836	0.0164	-11.125	-45.641

が、埋め込まれている。さらに、パラメータ $\hat{\xi}^{(m-1)}$ に $\hat{\delta}^{(m-1)}$ を加えた $\hat{\xi}^{(m)} = \hat{\xi}^{(m-1)} + \hat{\delta}^{(m-1)}$ の結果も示されている。パラメータのベクトル名とし、これまで、 $\hat{\beta}$ あるいは $\hat{\theta}$ を用いてきたのであるが、パラメータに μ と σ も含まれているので、 $\hat{\theta}$ ではなく $\hat{\xi}$ (グザイ、クサイ) を使うことにする。

表 11.11 左上の初期値 $\hat{\xi}^{(0)}$ をコピーし $\hat{\xi}^{(m-1)}$ に値のみをペーストすると、表 11.11 に示すように、 $\hat{\varepsilon}_i$ に関する残差平方和 $S_e = 106023.5$ が計算され、残差 $\hat{\varepsilon}_i = y_i - \hat{y}_i^{MEV}$ に対し式 (11.50) を用いて増減パラメータ $\hat{\delta}^{(m-1)}$ が推定され、新たなパラメータの推定値として $\hat{\xi}^{(m)}$ が計算されている。

表 11.11 初期値に対する残差線形化による計算結果 ($m=1$)

		$\xi^{(0)}$			$\xi^{(m-1)}$	$\delta^{(m-1)}$	$\xi^{(m)}$
	初期値	350	$\theta^{max} =$	350.0000	3.9643	353.9643	
		100	$\theta^{min} =$	100.0000	4.7949	104.7949	
		-0.5	$\mu^{MEV} =$	-0.5000	-0.0179	-0.5179	
		0.5	$\sigma^{MEV} =$	0.5000	-0.1369	0.3631	
	残差平方和	$S_e =$		106023.53			

表 11.12 に示すのは、ソルバーで残差平方和 S_e を最小にするように $\hat{\xi}^{(m-1)}$ を変化させた結果である。残差平方和が $S_e = 93259.73$ と減少し、増減パラメータ $\hat{\delta}^{(m-1)}$ は、 $\hat{\theta}_{max}^{MEV}$ を除いて 0.0000 となっている。 $\hat{\theta}_{max}^{MEV}$ の増減パラメータが 0.0011 なので、表 11.12 の $\hat{\xi}^{(m)}$ をコピーし、 $\hat{\xi}^{(m-1)}$ に値のみをペーストすると表 11.13 に示すように $\hat{\delta}^{(m-1)}$ がすべて 0.0000 となる。ただし、

表 11.12 ソルバーを用いた残差平方和 S_e の最小化

	$\xi^{(0)}$		$\xi^{(m-1)}$	$\delta^{(m-1)}$	$\xi^{(m)}$
初期値	350	$\theta^{max} =$	355.3908	0.0011	355.3919
	100	$\theta^{min} =$	105.1279	0.0000	105.1279
	-0.5	$\mu^{MEV} =$	-0.5005	0.0000	-0.5005
	0.5	$\sigma^{MEV} =$	0.3657	0.0000	0.3657
残差平方和		$S_e =$	93259.73		

小数点以下 5 桁目での四捨五入のために $\hat{\xi}^{(m-1)}$ と $\hat{\xi}^{(m)}$ の $\hat{\theta}_{min}$ が完全に一致しないので、さらにコピー&値のみペーストした結果を表 11.13 に示す。残差平方和 S_e は、表 11.12 と見た目は変化しないが、小数点以下の表示を増やすと 5 桁目で小さくなっていることが確認できる。

表 11.13 増減パラメータ $\delta^{(m-1)}$ の更なる最小化

	$\xi^{(0)}$		$\xi^{(m-1)}$	$\delta^{(m-1)}$	$\xi^{(m)}$
初期値	350	$\theta^{max} =$	355.3920	0.0000	355.3920
	100	$\theta^{min} =$	105.1279	0.0000	105.1279
	-0.5	$\mu^{MEV} =$	-0.5005	0.0000	-0.5005
	0.5	$\sigma^{MEV} =$	0.3657	0.0000	0.3657
残差平方和		$S_e =$	93259.73		

推定されたパラメータの 95%信頼区間およびゴンペルツ・最大極値曲線の 95%信頼区間を求めるためにパラメータに関する偏微分係数行列 \mathbf{Z} を必要とするのであるが、Excel シート上で正しく設定することは、簡単ではない。正しく設定されていれば、表 11.11 から表 11.13 で示した計算が可能であるが、少しでも数式にミスがあると直ぐに発散し計算不能になってしまう。冷静に特に式のプラス・マイナスなどにも注意を払う必要がある。

ゴンペルツ・最大極値曲線の 95%信頼区間

表 11.14 にゴンペルツ・最大極値曲線の 95%信頼区間の計算結果を示す。そのためにパラメータの共分散行列 $\Sigma(\hat{\xi})$ の算出が必要となる。線形モデルの場合は、デザイン行列 \mathbf{X} に対しパラメータの推定のために用いた $(\mathbf{X}^T \mathbf{X})^{-1}$ に誤差分散 $\hat{\sigma}^2$ を掛けることによりパラメータ $\hat{\theta}$ について共分散行列 $\Sigma(\hat{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$ が求められた。これと同様に、微分係数行列 \mathbf{Z} をデザイン行列とみなした $\Sigma(\hat{\xi}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$ が、線形モデルの $\Sigma(\hat{\theta})$ に相当する。

第 10 章でロジスティック曲線について活用事例を示してきたので、ここでは、ゴンペルツ・最大極値曲線の 95%信頼区間について示す。表 11.14 上段に示すように、表 11.10 でパラメータ推定のために用いた $(\mathbf{Z}^T \mathbf{Z})^{-1}$ に残差平方和 $S_e = 93259.73$ を自由度 $df = 72 - 4 = 68$ で除し、誤差分散 $\hat{\sigma}^2 = 93259.73 / 68 = 1371.47$ を掛けたパラメータの共分散行列 $\Sigma(\hat{\xi}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$ 示されている。この $\Sigma(\hat{\xi})$ の対角要素がパラメータ $\hat{\xi}$ の分散となるので、その

表 11.14 ゴンペルツ・最大極値曲線の 95%信頼区間の推定

		ξ	$\xi^{\wedge(m-1)}$	$\Sigma(\xi^{\wedge})=(Z^T Z)^{-1}\sigma^{\wedge^2}$					SE	L95%	U95%
	$\theta^{\wedge}_{max} =$	355.392		183.50	-13.02	0.220	0.887		13.546	328.36	382.42
	$\theta^{\wedge}_{min} =$	105.128		-13.02	44.190	0.134	-0.153		6.6476	91.86	118.39
	$\mu^{\wedge}_{MEV} =$	-0.5005		0.220	0.134	0.003	0.000		0.0534	-0.6071	-0.3938
	$\sigma^{\wedge}_{MEV} =$	0.3657	σ^{\wedge^2}	0.887	-0.153	0.000	0.008	$t_{0.05}(68)$	0.0881	0.1900	0.5415
残差平方和 $S_e =$		93259.73	1371.47	z_1	z_2	z_3	z_4	1.9955			
dose	x	y^{\wedge}_{MEV}	$g(x)$	$\partial_{\theta^{\wedge}_{max}}$	$\partial_{\theta^{\wedge}_{min}}$	$\partial_{\mu^{\wedge}_{MEV}}$	$\partial_{\sigma^{\wedge}_{MEV}}$	Var(y^{\wedge})	SE	L95%	U95%
0.001	-3.00	105.13	929.14	0.000	1.000	0.000	0.000	44.190	6.648	91.86	118.39
0.01	-2.00	105.13	60.342	0.000	1.000	0.000	0.000	44.190	6.648	91.86	118.39
0.0316	-1.50	105.13	15.391	0.000	1.000	-0.002	0.006	44.188	6.647	91.86	118.39
	-1.25	105.23	7.763	0.000	1.000	-2.259	4.629	42.310	6.505	92.25	118.21
0.1	-1.00	110.10	3.919	0.020	0.980	-53.27	72.754	59.379	7.706	94.72	125.48
	-0.75	139.74	1.978	0.138	0.862	-187.2	127.73	214.72	14.653	110.50	168.98
0.316	-0.50	197.24	1.000	0.368	0.632	-251.7	-0.120	133.81	11.568	174.16	220.32
	-0.25	256.29	0.504	0.604	0.396	-208.4	-142.7	121.58	11.026	234.29	278.30
1	0.00	299.16	0.255	0.775	0.225	-135.0	-184.8	116.74	10.805	277.60	320.72
	0.25	325.22	0.128	0.879	0.121	-77.31	-158.7	71.87	8.478	308.30	342.14
3.16	0.50	339.66	0.065	0.937	0.063	-41.63	-113.8	57.94	7.612	324.47	354.85
	0.75	347.33	0.033	0.968	0.032	-21.68	-74.13	78.29	8.848	329.68	364.99
10	1.00	351.29	0.017	0.984	0.016	-11.12	-45.64	109.10	10.445	330.45	372.13
	1.50	354.34	0.004	0.996	0.004	-2.870	-15.70	154.79	12.441	329.51	379.17
	2.00	355.12	0.001	0.999	0.001	-0.734	-5.016	174.06	13.193	328.80	381.45
	-0.5005	197.19	1.0000	0.3679	0.6321	-251.7	0.000	133.88	11.571	174.11	220.28

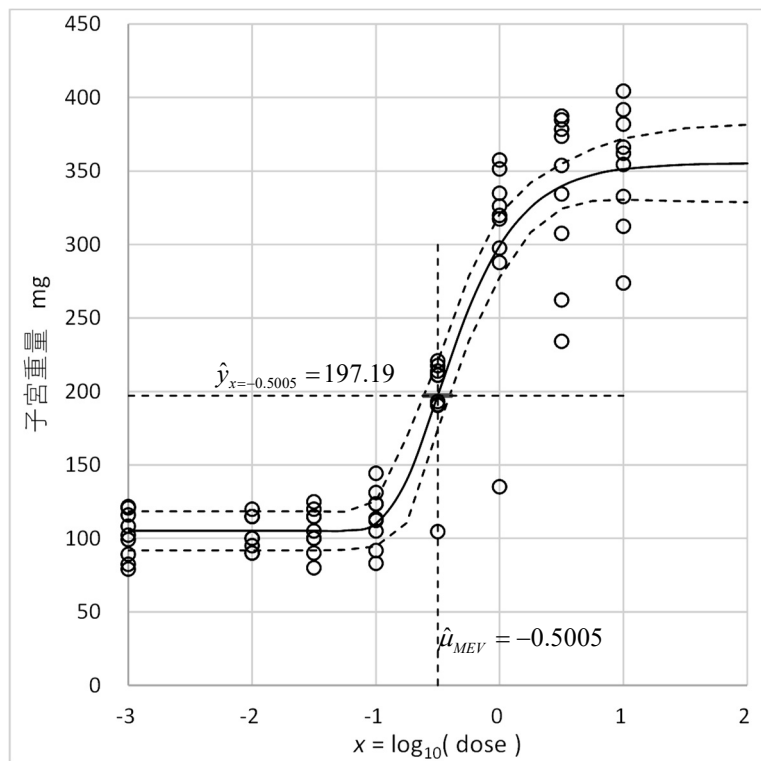


図 11.10 ゴンペルツ・最大極値曲線の 95%信頼区間

平方根によって標準誤差 SE と 95%信頼区間が求められている。最大値 $\hat{\theta}_{max}^{MEV} = 355.392$ に対する 95%信頼区間が (328.36, 382.42) と推定され、最小値 $\hat{\theta}_{min}^{MEV} = 105.128$ に対する 95%信頼区間が (91.86, 118.39) と推定されていて図 11.10 に反映されている。

推定された位置パラメータは、 $\hat{\mu}_{MEV} = -0.5005$ 、95%信頼区間は (-0.6071, -0.3938) であり、子宮重量 $\hat{y}_{x=-0.5005}$ の推定値は、表 11.14 の最終行に $x = -0.5005$ を追加することにより、

$$\begin{aligned} \hat{y}_{x=-0.5005} &= \theta_{min}^{(MEV)} + (\theta_{max}^{(MEV)} - \theta_{min}^{(MEV)}) \cdot \exp\left[-\exp\left(-\frac{x_i - \mu_{MEV}}{\sigma_{MEV}}\right)\right] \\ &= 105.1279 + (355.3920 - 105.1279) \times \exp\left[-\exp\left(-\frac{-0.5005 - (-0.5005)}{0.3657}\right)\right] \\ &= 105.1279 + 250.2641 \times \exp[-\exp(0)] \\ &= 105.1279 + 250.2641 \times 0.3679 = 197.19 \end{aligned} \quad (11.51)$$

として計算されている。

表 11.14 に示すように、ギンペルツ・最大極値曲線の 95%信頼区間は、 $x = \log_{10}(dose)$ に対し -3.00 から 2.00 までの範囲で適当な間隔で変化させ、式 (11.51) と同様に \hat{y}_x を計算している。その分散 $Var(\hat{y}_x)$ は、偏微分式に x を代入した $\mathbf{z}_x = [z_{1,x} \ z_{2,x} \ z_{3,x} \ z_{4,x}]$ に対し、パラメータの共分散行列 $\Sigma(\hat{\xi})$ に関する 2 次形式

$$Var(\hat{y}_x) = \mathbf{z}_x \Sigma(\hat{\xi}) \mathbf{z}_x^T \quad (11.52)$$

によって求められる。

投与量 $dose = 0.316$ に対する推定値は、 $\hat{y}_{x=-0.5} = 197.24$ と計算され、その微分係数のベクトルは、

$$\mathbf{z}_{x=-0.5} = [0.368 \quad 0.632 \quad -251.7 \quad -0.120] \quad (11.53)$$

として計算されている。推定値 $\hat{y}_{x=-0.5}$ に対する分散 $Var(\hat{y}_{x=-0.5})$ は、表 11.15 に示す $\mathbf{z}_{x=-0.5}$ の $\Sigma(\hat{\xi})$ に対する 2 次形式によって $Var(\hat{y}_{x=-0.5}) = 133.81$ として求められ、その 95%信頼区間は、(174.16, 220.32) が示されている。

表 11.15 ギンペルツ・最大極値曲線のパラメータの分散の計算

$\mathbf{z}_{x=-0.5}$								$\mathbf{z}_{x=-0.5}^T$	
z_1	z_2	z_3	z_4	$\Sigma(\hat{\xi}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \sigma^2$				$Var(\hat{y}_{x=-0.5})$	
0.3681	0.6319	-251.73	-0.1196	183.50	-13.02	0.2204	0.8870	0.3681	= 133.81
$\partial \theta_{max}^M$	$\partial \theta_{min}^M$	$\partial \mu_{MEV}$	$\partial \sigma_{MEV}$	-13.02	44.19	0.1344	-0.1528	0.6319	
				0.2204	0.1344	0.0029	-0.0002	-251.73	
				0.8870	-0.1528	-0.0002	0.0078	-0.1196	

10 パーセント反応の逆推定値の 95%信頼区間

表 11.14 に残差線形化法を用いて推定した推定値 \hat{y}^{MEV} の 95%信頼区間の計算結果を示し、図 11.10 にこれらを重ね書きした結果を示した。表 11.7 では、Excel のソルバーを用いて、残差平方和を最小にするようにパラメータを変化させてきたのであるが、Excel のソルバーには、更に優れた機能が備わっている。第 10 章の図 10.2 で示したのは、2 本の直線の差の回帰直線 $d = 2.5031 - 0.0545x$ の 95%信頼区間の上限が、 $d = 0.0$ となる x をソルバーを用いて推定した。これにより、傾きが異なる 2 本の回帰直線に対し、有意な差となる x を推定することができた。

式 (11.44) に示した 10 パーセント反応量の計算式から

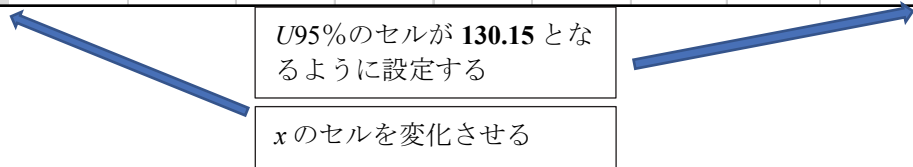
$$\left. \begin{aligned} y_{D10} &= \theta_{min} + (\theta_{max} - \theta_{min}) \times 0.10 \\ &= 105.1279 + (355.3920 - 105.1279) \times 0.10 = 130.15 \end{aligned} \right\} \quad (11.54)$$

$y_{D10}^{MEV} = 130.15$ が推定され、式 (11.46) から表 11.8 に示すように、逆推定の対数用量 $x_{D10}^{MEV} = -0.806$ および $dose_{D10}^{MEV} = 10^{-0.806} = 0.156$ を求めた。この 10 パーセント反応量に対する 95%信頼区間を求めたい。これは、図 11.10 に示したゴンペルツ・最大極値曲線の 95%信頼区間に対し、 $y_{D10}^{MEV} = 130.15$ を通る水平線が、L95% と U95% とクロスする対数用量 x をソルバーで求めることで実現できる。

表 11.16 は、表 11.7 の任意の 3 行を取り出し、第 1 行目の \hat{y}^{MEV} を Excel のソルバーの「目的セルの設定」で選択設定し「目標値」の「指定値」で、 $\hat{y}_{D10}^{MEV} = 130.15$ を入力、「変数セルの変更」に変数 x セルを設定して実行した結果である。その結果、 $\hat{y}_{D10}^{MEV} = 130.15$ となるように変数 x が変化し、 $\hat{x}_{D10}^{MEV} = -0.8055$ が推定されている。

表 11.16 ソルバーによるゴンペルツ・最大極値曲線の 10 パーセント点の推定

dose	x	\hat{y}^{MEV}	$g(x)$	$\partial_{\theta_{max}}$	$\partial_{\theta_{min}}$	$\partial_{\mu^{MEV}}$	$\partial_{\sigma^{MEV}}$	Var(\hat{y})	SE	L95%	U95%
0.156	-0.8055	130.15	2.303	0.100	0.900	-157.5	131.41	192.54	13.876	102.46	157.84
0.220	-0.6582	158.82	1.539	0.215	0.785	-226.0	97.46	206.44	14.368	130.15	187.49
0.109	-0.9622	112.43	3.534	0.029	0.971	-70.57	89.10	78.85	8.880	94.71	130.15



引き続き、第 2 行目の L95% のセルが 130.15 となるように変数 x セルを変化させると $\hat{x}_{D10(L95\%)}^{MEV} = -0.6582$ が推定される。同様に、第 3 行目の U95% のセルが 130.15 となるように変数 x セルを変化させると $\hat{x}_{D10(U95\%)}^{MEV} = -0.9622$ が推定される。求められた 10 パーセント反応となる 95%信頼区間 $(-0.9622, -0.6582)$ を図 11.11 に重ね書する。

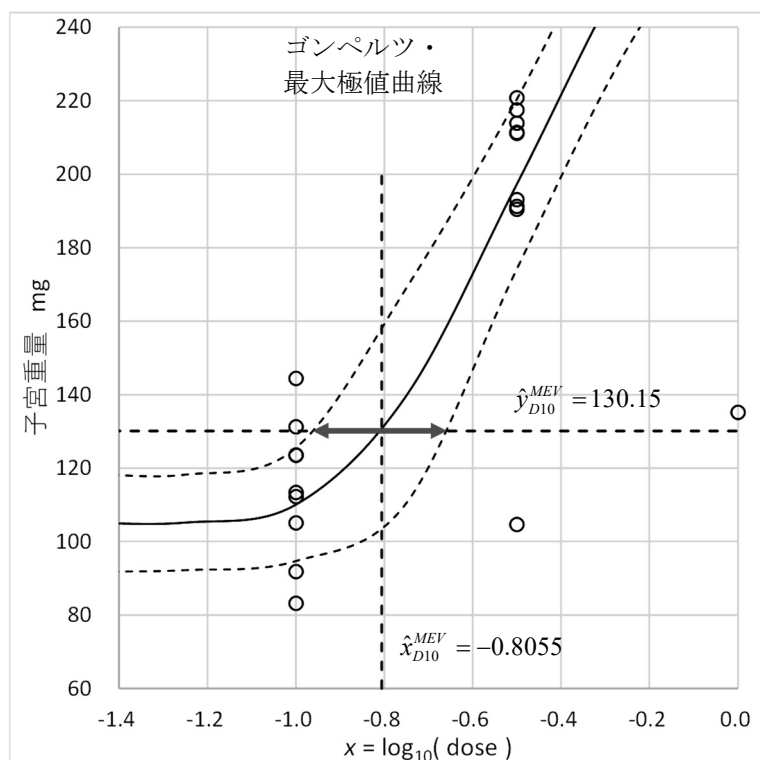


図 11.11 10 パーセント反応の逆推定値に対する 95%信頼区間

Excel および統計ソフトの脆弱性

このようにソルバーを用いて、任意のパーセント点の逆推定値の 95%信頼区間が手軽に求められることを示したのであるが、何段階もの計算過程を経ており、推定された結果が正しいかは、しかるべき統計ソフトによる検証が必要である。また、ここに示した結果が正しいと検証されたとしても、新たなデータセットに対しては、Excel の計算シートに何らかの変更が必要であり、このことが、正しくセットされた計算手順が、どこかで知らないうちに変更されてしまう可能性があり、また、それを防止するすべもない。これが、Excel による解析の本質的な脆弱性であり、Excel を統計解析の業務として用いることは推奨できない。ただし、統計ソフトで出力された結果を整理し、ここに示したようなグラフ表示には、Excel が適している。

Excel の本質的な脆弱性について指摘したのであるが、統計ソフトの結果が常に信頼できるとも限らない。典型的な事例に対し結果が再現できれば、使い方に間違いはないと判断できるが、応用事例に対して結果が正しいと自信をもって判断できるのであろうか。同じ統計ソフトを用いて別の人が同じ結果を出したと言っても、どこかで同じミスを犯しているかも知れない。別の人が別の統計ソフトを使って、同じ結果が得られれば、ほぼ正しいと言えるかもしれないが、微妙に異なる結果が得られた場合に、どれが正しいか苦しむことになる。

11.4. JMP によるゴンペルツ・最大極値曲線の推定および逆推定

各種のシグモイド曲線のパラメータ推定が、表 11.7 に示したように Excel のソルバーにより手軽にでき、Excel の散布図にシグモイド曲線を重ね書きした結果も図 11.8 に示した。また、得られたデータに、どのようなシグモイド曲線のあてはめが適しているかを、残差平方和の比較によってゴンペルツ曲線が適していることを示した。

JMP の「曲線のあてはめ」

JMP によるゴンペルツ・最大極値曲線のあてはめは、簡易版と精密版がある。まず、簡易版の「曲線のあてはめ」を用いて、ロジスティック曲線およびゴンペルツ・最大極値曲線をあてはめて、残差平方和の比較を行なう。表 11.7 に示した *dose* および *y* を JMP データファイルとし、 $x = \log_{10}(dose)$ を新たな列として追加する。「曲線のあてはめ」を起動し、プルダウンメニューから、「ロジスティック 4P のあてはめ」、および、「Gompertz 4P のあてはめ」を選択すると、即座に図 11.12 が得られる。「Gompertz 4P」の残差平方和 $SSE=93259.73$ が「ロジスティック 4P」の $SSE=93602.41$ に比べ 342.68 小さいことが確認でき、表 11.7 の結果に一致している。

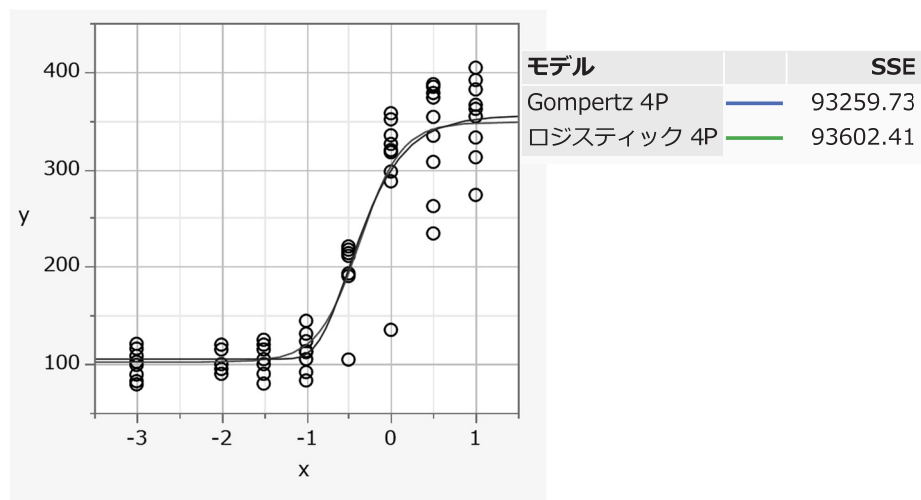


図 11.12 JMP の「曲線のあてはめ」の結果の表示

表 11.17 に示すのは、「ロジスティック 4P」のパラメータ推定であり、変数 a (増加率) の逆数 $1/a = 1/4.0655$ が、表 11.7 に示した $\hat{\sigma}_{LGS} = 0.2460$ に一致し、変曲点 $b = -0.3677$ が位置パラメータに一致する。下側漸近線は最小値 $\hat{\theta}_{min}$ ，上側漸近線は最大値 $\hat{\theta}_{max}$ を意味する。

表 11.17 「ロジスティック 4P」によるパラメータの推定

a = 増加率	$c + \frac{(d-c)}{\left(1 + \text{Exp}(-a \cdot (x-b))\right)}$	パラメータ	推定値	標準誤差
b = 変曲点		増加率	4.0655	0.8133
c = 下側漸近線		変曲点	-0.3677	0.0563
d = 上側漸近線		下側漸近線	101.9132	7.2030
		上側漸近線	348.4343	10.5196

表 11.18 に示すのは、「Gompertz 4P」のパラメータ推定であり増加率 c の逆数 $1/c = 1/2.7342$ が、式 (11.42) の $\hat{\sigma}_{MEV} = 0.3657$ に一致する。

表 11.18 「Gompertz 4P」によるパラメータの推定

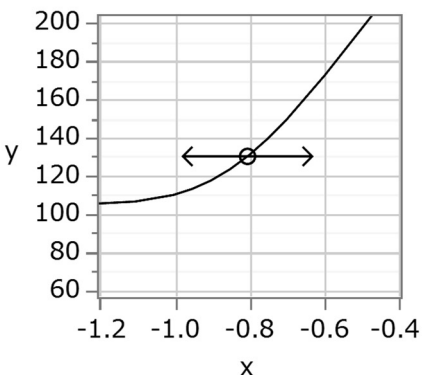
a = 下側漸近線	$a + (b-a) \cdot \text{Exp}\left(-\text{Exp}(-c \cdot (x-d))\right)$	パラメータ	推定値	標準誤差
b = 上側漸近線		下側漸近線	105.1279	6.6476
c = 増加率		上側漸近線	355.3920	13.5461
d = 変曲点		増加率	2.7342	0.6584
		変曲点	-0.5005	0.0534

前節で Excel のソルバーを用いた 10 パーセント反応に対する逆推定の方法を示したが、JMP の「曲線のあてはめ」の「カスタム逆推定」で 10%点に相当する $\hat{y}_{D10}^{MEV} = 130.15$ を入力すると表 11.19 に示すように逆推定値 $\hat{x}_{D10}^{MEV} = -0.8055$ が得られ、標準誤差と 95%信頼区間の推定値が得られる。

表 11.19 「Gompertz 4P」による 10%反応量の逆推定

予測値				
指定されたy	xの予測値	標準誤差	下側95%	上側95%
130.1500	-0.8055	0.0881	-0.9782	-0.6329

95%信頼区間が、Excel のソルバーで求めたプロフィアル法による (-0.9622, -0.6582) と異なるのは、逆推定の計算式を用いた線形近似計算による標準誤差 0.0881 を用い、正規分布の 95%点 1.96 による幅を計算しているためである。



このように手軽に「ロジスティック 4P」および「Gompertz 4P」のあてはめができることは、精密版の「非線形モデル」を使うための準備として適している。ただし、「曲線のあてはめ」で用いている計算式は、これまで用いてきた位置パラメータと形状パラメータを用いた計算式および変数名が異なることに注意が必要である。

「非線形回帰」によるゴンペルツ・最大極値曲線のあてはめと逆推定

表 11.20 に示したように JMP データファイルにゴンペルツ・最大極値曲線の推定値のための変数「y^_MEV」を新たに追加し、「計算式」の設定を行なう。

表 11.20 ゴンペルツ・最大極値曲線のあてはめのための JMP データファイル

	No.	dose	x	y	y^_MEV	
○	1	1	0.001	-3.00	102.35	100.00
○	2	2	0.001	-3.00	120.82	100.00
○	3	3	0.001	-3.00	115.92	100.00
⋮						
○	71	71	10	1.00	366.2	337.86
○	72	72	10	1.00	332.67	337.86

表 11.21 に示すように計算式エディタ上で、パラメータの初期値を設定する。JMP は 3 バイトコードなので、「シーた」と入力し、変換された「θ」を選択し、「θmax_M」のようなパラメータ名設定が行なえる。GUI 形式の計算式エディタでゴンペルツ・最大極値曲線式を完成させると、表 11.20 に y^_MEV に初期値を適用した計算結果が表示されている。

表 11.21 JMP の計算式エディタ上でのパラメータの初期値の設定

パラメータ

パラメータの新規作成...

θmax_M = 350

θmin_M = 100

μ_MEV = -0.5

σ_MEV = 0.5

$$\theta_{\min_M} + (\theta_{\max_M} - \theta_{\min_M}) \cdot \text{Exp} \left(- \text{Exp} \left(- \left(\frac{x - \mu_{\text{MEV}}}{\sigma_{\text{MEV}}} \right) \right) \right)$$

JMP の「非線形回帰」の設定画面で、「Y 応答変数」に変数 Y を、「X 予測式列」に y^_MEV 変数をセットし「実行」すると、表 11.22 に示すようにパラメータの推定が行なわれ、表 11.14 に示した Excel での結果と一致することが確認される。

表 11.22 JMP によるゴンペルツ曲線のあてはめ結果

解				
	SSE	DFE	MSE	RMSE
	93259.73	68	1371.4666	37.0333
パラメータ	推定値	近似標準誤差	下側信頼限界	上側信頼限界
θmax_M	355.3920	13.5461	330.6990	390.3705
θmin_M	105.1279	6.6476	91.1320	118.2310
μ_MEV	-0.5005	0.0534	-0.6170	-0.3927
σ_MEV	0.3657	0.0881	0.1911	0.5846

解法: 解析 Gauss-Newton

JMP の信頼区間はプロファイル法によるので、表 11.14 に示した Excel の結果と異なる。

図 11.13 左に推定されたゴンペルツ曲線を重ね書きした結果を示し、図 11.13 右には、予測プロファイルの機能によりゴンペルツ曲線に対する 95%信頼区間を示す。X 軸には、位置パラメータ $x_0 = \hat{\mu}_{MEV} - 0.5005$ を設定した結果として、 $y_0 = \hat{y}_{MEV} = 197.1915$ におよび 95%信頼区間 (140.10, 220.28) が示されている。

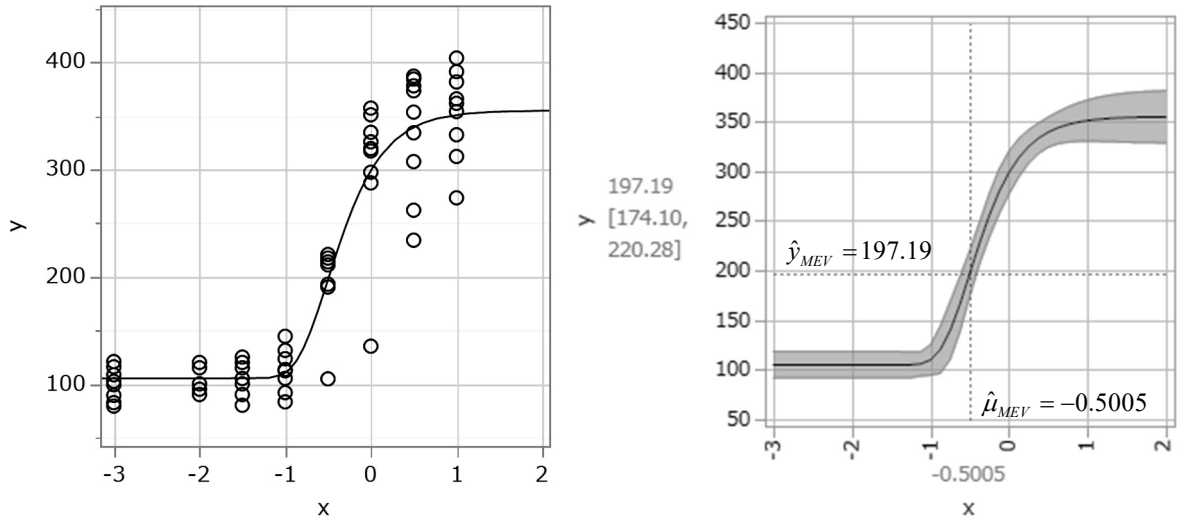


図 11.13 ギンペルツ曲線に対する予測プロファイルによる 95%信頼区間の表示

表 11.23 に示すのは、JMP の「非線形回帰」の逆推定の結果で、式 (11.46) に示した逆推定の計算式も示され、 $\hat{x}_{D10}^{MEV} = -0.8055$ が得られている。

表 11.23 ギンペルツ曲線の 10%反応量の逆推定

逆推定				
-Log(-Log((130.15 - θ_{min_M}) / (θ_{max_M} - θ_{min_M}))) * σ_{MEV} + μ_{MEV}				
指定されたy	xの予測値	標準誤差	下側0.95	上側0.95
130.1500	-0.8055	0.0881	-0.9813	-0.6298

逆推定値 $\hat{x}_{D10}^{MEV} = -0.8055$ の標準誤差は、0.0881 は、表 11.19 に示した「曲線のあてはめ」に一致するが、95%信頼区間が一致しない。これは、信頼区間の計算に自由度 (72-4) の t 分布の両側 5%点 1.9955 を用いているのに対し、「曲線のあてはめ」では、正規分布の両側 5%点 1.96 を使用しているために起きた不整合である。

シグモイド曲線の 95%信頼区間

JMP の「非線形回帰」では、図 11.13 右に示すようにゴンペルツ・最大極値曲線の 95%信頼区間が予測プロファイルとして図示できることを示した。この結果を図 11.13 右の散布図上に重ね書きしたいのであるが、残念ながらできない。そのために、「予測式の保存」およ

表 11.24 JMP データファイルに書き込まれた予測値および 95%信頼区間

	No.	dose	x	y	y^_MEV	予測値 y^_MEV	分散	予測値の 標準誤差 y^_MEV	下限M y^_MEV	上限M y^_MEV	
○	70	10.000	1.00	312.40	337.86	351.29	109.097	10.445	330.45	372.13	
○	71	10.000	1.00	366.20	337.86	351.29	109.097	10.445	330.45	372.13	
○	72	10.000	1.00	332.67	337.86	351.29	109.097	10.445	330.45	372.13	
	73	• 0.001	-3.00	•	100.00	105.13	44.190	6.648	91.86	118.39	
	74	• 0.010	-2.00	•	100.00	105.13	44.190	6.648	91.86	118.39	
	75	• 0.032	-1.49	•	100.17	105.13	44.187	6.647	91.86	118.39	
:											
	86	• 31.623	1.50	•	345.46	354.34	154.786	12.441	329.51	379.17	
	87	• 100.000	2.00	•	348.32	355.12	174.059	13.193	328.80	381.45	

元のデータは、72 行目までで、73 行目～87 行目に表 11.14 に示した 95%信頼区間の作図用の *dose* と *x* が追加されている。この追加データも含めて、予測値および 95%信頼区間の推定値が書き込まれている。JMP では、「非線形回帰」に限らず、95%信頼区間の推定式が、埋め込まれているので、図 10.13 に示したと同様の「計算式」を表示することができる。ここでは、省略するが、表 11.15 に示した Excel の行列計算式と同じ計算式であることが確認できる。

び「予測信頼限界の計算式の保存」をプルダウンメニューから選択し、表 11.24 に示すように元の JMP データファイルに書き出し、図 11.14 に示すようにグラフ・ビルダーで作図することにより可能となる。

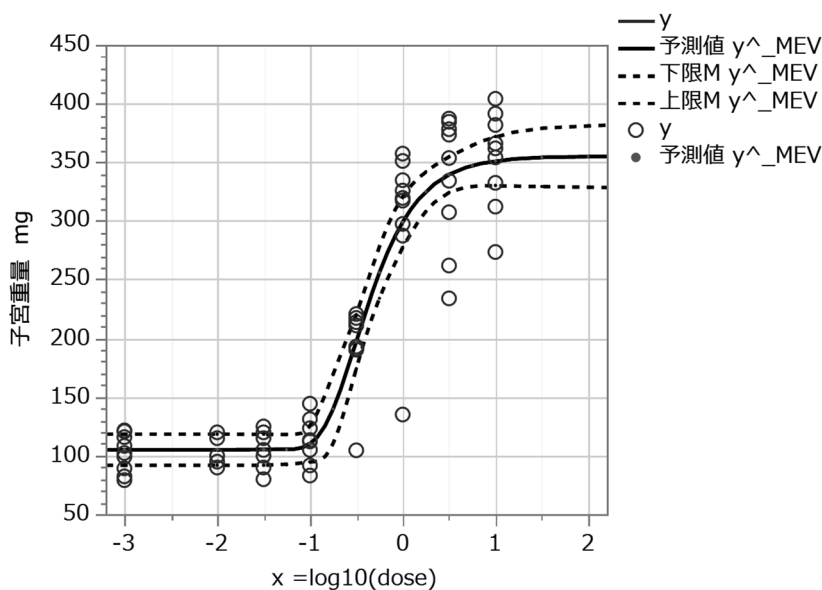


図 11.14 JMP のグラフ・ビルダーによるゴンペルツ曲線と 95%信頼区間

11.5. オフセットを活用したシグモイド曲線に対する逆推定

得られたデータに各種のシグモイド曲線をあてはめ、パラメータを推定しつつ残差平方和を算出し、最適なシグモイド曲線を選択する目的であれば、Excel のソルバーを用いた方法が簡便である。10 パーセントの反応となるような用量 x を逆推定し、その 95% 信頼区間を推定したいとの目的のためには、逆推定の機能が組み込まれている JMP が最強である。ただし、有償であり誰でも手軽に使えるわけではない。そこで、無償で継続的に使える OnDemand SAS の NLIN プロシジャにより、オフセットを活用した逆推定値の 95% 信頼区間の推定法を導入する。

オフセットを活用した逆推定

累積分布関数を活用したシグモイド曲線のあてはめのパラメータとして、 $\theta_{max}^{分布}$ 、 $\theta_{min}^{分布}$ 、 $\mu_{分布}$ 、 $\sigma_{分布}$ の 4 つのパラメータを用いてきた。シグモイド曲線について、位置パラメータと形状パラメータを用いて次のような規準化

$$\eta^{LGS} = \frac{x - \mu_{LGS}}{\sigma_{LGS}}, \quad \eta^{MEV} = \frac{x - \mu_{MEV}}{\sigma_{MEV}}, \quad \eta^{SEV} = \frac{x - \mu_{SEV}}{\sigma_{SEV}} \quad (11.55)$$

を行なってきた。この基準化した変数 $\eta^{分布}$ を用いることにより、 $\theta_{max}^{分布} = 1$ 、 $\theta_{min}^{分布} = 0$ とするシグモイド曲線は、

$$LGS \text{ ロジスティック曲線:} \quad \pi^{LGS} = \frac{1}{1 + \exp(-\eta^{LGS})} \quad (11.56)$$

$$MEV \text{ ゴンペルツ・最大極値曲線:} \quad \pi^{MEV} = \exp[-\exp(-\eta^{MEV})] \quad (11.57)$$

$$SEV \text{ 対数ワイグル・最小極値曲線:} \quad \pi^{SEV} = 1 - \exp[-\exp(\eta^{SEV})] \quad (11.58)$$

のように簡単化することができる。基準化した変数 $\eta^{分布}$ が 0.0 となるのは、それぞれ ($\hat{\mu}_{LGS}$ 、 $\hat{\mu}_{MEV}$ 、 $\hat{\mu}_{SEV}$) であることは式 (11.55) から自明である。

これらのシグモイド曲線のパーセント点 $\pi^{分布}$ を $\eta^{分布}$ に関して解くと

$$\text{ロジスティック曲線:} \quad \eta_{\pi}^{LGS} = \ln\left(\frac{\pi^{LGS}}{1 - \pi^{LGS}}\right) \quad (11.59)$$

$$\text{ゴンペルツ・最大極値曲線:} \quad \eta_{\pi}^{MEV} = -\ln[-\ln(\pi^{MEV})] \quad (11.60)$$

$$\text{対数ワイグル・最小極値曲線:} \quad \eta_{\pi}^{SEV} = \ln[-\ln(1 - \pi^{SEV})] \quad (11.61)$$

が得られる。これらの変数 $\eta_{\pi}^{分布}$ が、任意の反応のパーセント点 $\pi^{分布}$ (0.0~1.0) に対するシグモイド曲線の X 軸の位置となっている。これらの変数 $\eta_{\pi}^{分布}$ を式 (11.55) に加えることによ

り，推定される位置パラメータ $\mu_{\text{分布}}$ が元の位置からオフセットされたパラメータ $\mu_{\text{分布}}^{\text{offset}}$ として推定される。

実際に反応が $\pi = 0.10$ となる オフセット変数 $\eta_{\pi=0.10}^{\text{分布}}$ を式 (11.55) に加えたシグモイド曲線は，

$$\text{ロジスティック曲線：} \quad \pi^{LGS} = \frac{1}{1 + \exp\left(-\frac{x - \mu_{LGS}^{\text{offset}}}{\sigma_{LGS}} + \eta_{\pi=0.10}^{LGS}\right)} \quad (11.62)$$

$$\text{ゴンペルツ・最大極値曲線：} \quad \pi^{MEV} = \exp\left[-\exp\left(-\frac{x - \mu_{MEV}^{\text{offset}}}{\sigma_{MEV}} + \eta_{\pi=0.10}^{MEV}\right)\right] \quad (11.63)$$

$$\text{対数ワイグル・最小極値曲線：} \quad \pi^{SEV} = 1 - \exp\left[-\exp\left(\frac{x - \mu_{SEV}^{\text{offset}}}{\sigma_{SEV}} + \eta_{\pi=0.10}^{SEV}\right)\right] \quad (11.64)$$

のように設定することができる。

このような式で，推定された位置パラメータ $\mu_{\text{分布}}^{\text{offset}}$ が，どうして 10 パーセント反応量となるか，謎めいており，なかなか理解しづらい。実際のデータ y_i に対しシグモイド曲線のあてはめる場合を想定する。残差線形化法によりシグモイド曲線の推定値 \hat{y}_i に対する位置パラメータ $\hat{\mu}_{\text{分布}}$ が得られたとしたときに，オフセット変数が $\eta_{\pi}^{\text{分布}} = 0.0$ となっていれば，なんにも悪さはしないので通常的位置パラメータが推定される。シグモイド曲線の 10 パーセント点は $\eta_{\pi=0.10}^{\text{分布}} \neq 0.0$ なので，実際のデータ y_i にシグモイド曲線をあてはめようとしたときに，オフセット値 $\eta_{\pi=0.10}^{\text{分布}}$ が邪魔をするので，本来の位置パラメータ $\hat{\mu}_{\text{分布}}$ ではなく，邪魔された分だけ位置を変えて $\hat{\mu}_{\text{分布}}^{\text{offset}}$ が，残差平方和を最小にするパラメータとして推定される。したがって，オフセット変数 $\eta_{\pi}^{\text{分布}}$ を適切に調整することにより，任意のパーセント点の推定が可能となる。

シグモイド曲線のパーセント点についての逆推定を行なうためのオフセット変数 $\eta_{\pi}^{\text{分布}}$ がどのように変化するかを概観する。表 11.25 に示すのは，幾つかのパーセント点に対し，各シグモイド曲線のオフセット値を計算した結果である。

表 11.25 各シグモイド曲線に対応するオフセット変数 $\eta_{\pi}^{\text{分布}}$ の推定値

パーセント点		η_{π}^{LGS}	η_{π}^{MEV}	η_{π}^{SEV}
π	変曲点	$=\ln(\pi/(1-\pi))$	$=-\ln(-\ln(\pi))$	$=\ln(-\ln(1-\pi))$
0.9000		2.1972	2.2504	0.8340
0.6321	$=\exp(-\exp(-0))$	0.5413	0.7794	0
0.5000	$=1/(1+\exp(-0))$	0	0.3665	-0.3665
0.3679	$=1-\exp(-\exp(0))$	-0.5413	0	-0.7794
0.1000		-2.1972	-0.8340	-2.2504

図 11.15 に示すのは、ゴンペルツ・最大極値曲線 π^{MEV} に対し表 11.25 に示した各パーセント点に対するオフセット値を重ね書きした結果である。これだけを見れば、単にゴンペルツ・最大極値曲線のパーセント点 π^{MEV} に対する η^{MEV} の位置を示しているに過ぎない。ただし、オフセット値を含む式 (11.63) を用いて実際のデータに対し残差線形化法により推定した場合に、位置パラメータ $\hat{\mu}_{MEV}^{offset}$ の 95%信頼区間を得ることができる。

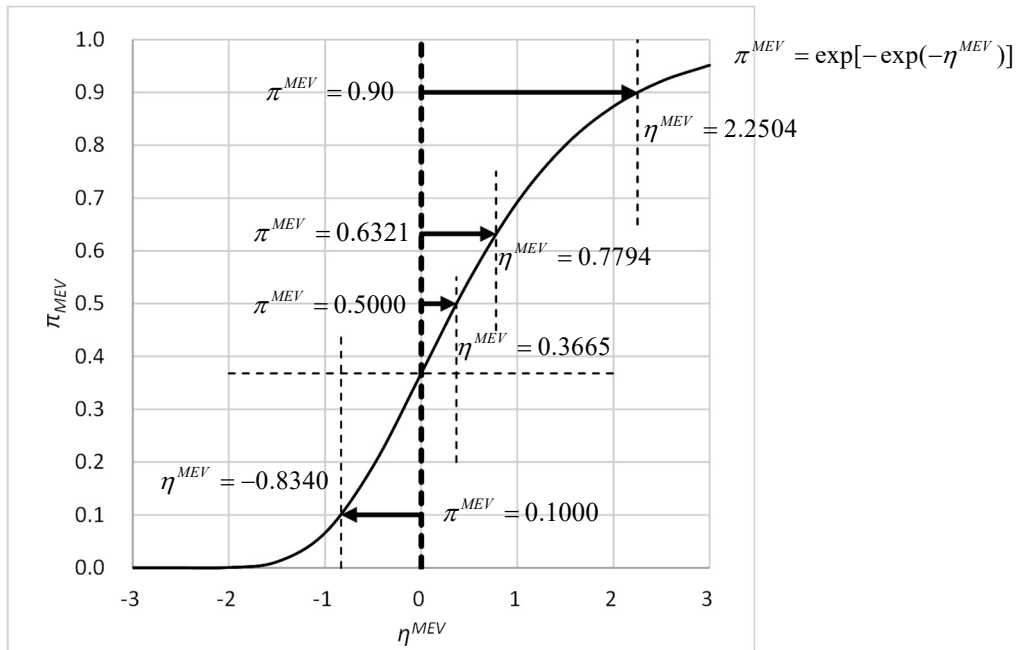


図 11.15 ゴンペルツ・最大極値曲線に対するオフセット

SAS の非線形 NLIN プロシジャによる解析

無償で継続的に使える OnDemand SAS の NLIN プロシジャにより、ゴンペルツ・最大極値曲線をあてはめ、10 パーセント点の逆推定と 95%信頼区間を求める。NLIN プロシジャでは、パラメータに関する偏微分を自動的に行なう機能が含まれており、式 (11.63) に θ_{max}^{MEV} 、 θ_{min}^{MEV} を付加した計算式を与えるだけで μ_{MEV}^{offset} と 95%信頼区間を求めてくれる。なお、NLIN プロシジャの詳細については、SAS Institute (2014)、「The NLIN Procedure」を参照のこと。

SAS による解析を行なうためには、データを SAS データセット化する必要がある。Excel 形式でも、CSV 形式のテキストデータを読み込むことも自在であるが、解析に用いるデータが表 11.6 で示した表で与えられている場合には、そのままの形式のテキストデータとして SAS プログラムの中で明示することが、解析結果の再現性にとって望ましい。ただし、表 11.6 に示されている *dose* をそのまま読み込んでデータとして扱うことは難しいので、表 11.6

全体を Excel 上で転置して SAS に取り込むことにする。ただし、SAS でのデータの区切りは、スペースなので、Excel のデータを「メモ帳」などのエディタ上にペースとし、標準の区切りコードである Tab をスペースに全置換して SAS のプログラムエディタに取り込む。

ラットの子宮重量の SAS データセット化

データの SAS への読み込みは、DATA ステップで行なう。SAS データセット名は「d01」とし、input ステートメントで、1 行目の *dose* を読み込み、常用対数変換で *x* とし、do ステートメントで、9 個のデータを読み込みつつ、output ステートメントで逐次「d01」に書き出す。読み込むデータなくなると自動的に終了となる。作成された「d01」の変数は、(*dose*, *x*, *i*, *y*) の 4 変数であり、 $8 \times 9 = 72$ 行のデータセットとなり、表 11.26 に示すように print プロシジャで確認することができる。

SAS プログラム 11.1

```
TITLE1 '環境ホルモン.SAS << ラットの子宮重量 >>';
DATA d01;
  input dose @;
  x = log10(dose);
  do i=1 to 9;
    input y @; output;
  end;
/* dose 1 2 3 7 8 9 11 18 19 */
datalines;
0.0010 102.35 120.82 115.92 121.62 79.22 108.47 82.45 89.25 99.17
0.0100 95 115 115 120 90 115 100 90 100
0.0316 105 115 120 125 80 115 100 90 100
0.100 112.22 123.47 144.42 131.25 105.08 123.60 113.38 91.80 83.17
0.316 190.45 217.48 213.95 220.83 211.13 211.37 191.23 193.07 104.67
1.00 319.78 351.32 326.07 317.52 287.68 357.57 297.67 334.95 135.17
3.16 373.72 384.72 378.37 387.43 262.20 353.82 307.60 334.48 234.17
10.00 382.00 404.32 354.37 391.67 273.73 362.05 312.40 366.20 332.67
;
proc print data=d01; run;
```

表 11.26 Excel に取り込んだ SAS データセット「d01」の出力

OBS	dose	x	i	y
1	0.001	-3	1	102.35
2	0.001	-3	2	120.82
3	0.001	-3	3	115.92
:				
71	10	1	8	366.2
72	10	1	9	332.67

オフセット無しの解析・オフセットを用いた逆推定

オフセットを含む解析に先立ち、NLIN プロシジャによるゴンペルツ・最大極値曲線のあてはめを行いパラメータの推定結果を、Excel および JMP の結果と照合する。NLIN プロシジャ

は、初期値設定のための `parms` ステートメントでの初期値設定に引き続き、`model` ステートメントでの式 (11.42) で示した式

$$y_i^{MEV} = \theta_{min}^{MEV} + (\theta_{max}^{MEV} - \theta_{min}^{MEV}) \cdot \exp \left[-\exp \left(-\frac{x_i - \mu_{MEV}}{\sigma_{MEV}} \right) \right] + \varepsilon_i$$

を設定する.

SAS プログラム 11.2 /* ギンペルツ・最大極値曲線の推定 */

```
proc nlin data=d01 list ;
  parms    theta_maxM=350 theta_minM=100 mu_MEV=-0.5 sigma_MEV=0.5 ;
  g = Exp( -(x-mu_MEV)/sigma_MEV ) ;
  model y = theta_minM + ( theta_maxM - theta_minM ) * Exp( - g ) ;
run ;
```

表 11.27 にギンペルツ・最大極値曲線のパラメータの推定結果を示す. 残差平方和 $S_e = 93259.7$ と表 11.14 に示した Excel での結果に一致し, 推定されたパラメータの表示桁数が少なくなっているが, 表 11.14 と一致することが確認できる.

表 11.27 NLIN プロシジャによるギンペルツ・最大極値曲線のパラメータの推定

要因	自由度	平方和	平均平方	F 値	近似 Pr > F
Model	3	785681	261894.00	190.96	<.0001
Error	68	93259.7	1371.50		
Corrected Total	71	878941			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		(注)
theta_maxM	355.4	13.5461	328.4	382.4	t 分布の
theta_minM	105.1	6.6476	91.8628	118.4	両側5%点を
mu_MEV	-0.5005	0.0534	-0.6071	-0.3938	用いている
sigma_MEV	0.3657	0.0881	0.1900	0.5415	

オフセット無しでの結果の再現性が確認できたので, 反応の 10 パーセント点における逆推定を式 (11.60), および, 式 (11.63) で示したオフセット変数 η_{π}^{MEV} を用いて NLIN プロシジャに組み込み推定する. NLIN プロシジャの中で「 $\pi=0.1$ 」を与え, 式 (11.63) により, `model` 式の中で「 $(-\log(-\log(\pi)))$ 」として与えている. ここで, 「 $\pi=0.9$ 」とすれば, 90 パーセント点を推定することができる.

SAS プログラム 11.3 * ギンペルツ・最大極値曲線 D10% 用量の推定 */

```
proc nlin data=d01 list ;
  pi=0.1 ;
  offset = -log(-log(pi)) ;
  parms    theta_maxM=350 theta_minM=100 mu_MEV_offset=-0.5 sigma_MEV=0.5 ;
  g = exp(- ( x - mu_MEV_offset ) / sigma_MEV + offset ) ;
  model y = theta_minM + ( theta_maxM - theta_minM ) * exp( - g ) ;
run ;
```

表 11.28 に示すようにパラメータ「 μ_MEV_offset 」の行に 10 パーセント点の推定値 $\hat{x}_{y=D_{10}}^{MEV} = -0.8055$ が得られ、近似標準誤差が $SE(\hat{x}_{y=D_{10}}^{MEV}) = 0.0931$ となる。この結果は、表 11.23 に示した JMP の逆推定で求めた $SE = 0.0881$ とは異なる。JMP では、逆推定式をパラメータに関して偏微分した式を用いた 2 次形式での近似計算であるのに対し、NLIN プロシジャでの逆推定は、非線形モデルとして直接推定しているためである。なお、表 11.10 に示した Excel の計算シートにオフセットを加えれば、SAS の NLIN プロシジャの結果と同じ結果が得られる。

表 11.28 ギンペルツ・最大極値曲線の 10 パーセント点に対する逆推定

要因	自由度	平方和	平均平方	F 値	近似 Pr > F
Model	3	785681	261894	190.96	<.0001
Error	68	93259.7	1371.5		
Corrected Total	71	878941			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		
θ	355.4	13.5460	328.4	382.4	
θ_{minM}	105.1	6.6476	91.8629	118.4	
μ_MEV_offset	-0.8055	0.0931	-0.9913	-0.6197	
σ_MEV	0.3657	0.0881	0.19	0.5415	

ギンペルツ・最大極値曲線の 95%信頼区間

NLIN プロシジャで計算された 10 パーセント点の推定値 $\hat{x}_{y=D_{10}}^{MEV} = -0.8055$ 、および、95%信頼区間 $(-0.9913, -0.6197)$ が適切に推定されているか自己検証するには、どうしたら良いのであろうか。JMP のように手軽に結果のグラフ表示ができれば良いのだが、SAS ではなかなか思うようにできないので、Excel の力を借るのが現実的である。表 11.14 で示したと同様の出力を NLIN プロシジャで再現することに挑戦する。

表 11.14 には、元の $dose$ 、常用対数変換した x 、推定値 \hat{y}^{MEV} 、分散 $Var(\hat{y}^{MEV})$ 、 SE 、 $L95\%$ 、 $U95\%$ などが示され、これらを元に図 11.10 が作成されている。また、図 11.11 には、10 パーセント点の拡大表示もなされている。SAS により表 11.14 と同じ内容のものを作成するために解析用の SAS データセットに、新たな推定用のデータを結合する。結合データセットには、 $dose = 10^x$ または $x = \log_{10}(dose)$ だけが含まれていれば、推定値 \hat{y}^{MEV} などが自動的に計算されるので、それを Excel シートに取り出すことにする。

SAS の DATA ステップで「d02」データセット作成を宣言し、do ステートメントを用いて表 11.14 に示されている x の並びを再現するために、do ステートメントを 3 回繰り返し、 $dose = 10^x$ も同時に計算する。作成された「d02」SAS データセットを元の「d01」に set ステートメントで行方向に結合し「d03」データセットを作成する。この「d03」に対し NLIN プロ

シジヤで解析しつつ、output ステートメントで ($y_{\hat{}}$, SE, L95, U95) を「Out03」データセットに書き出す。表 11.29 に Excel に取り込んだ結果を示す。

```
SAS プログラム 11.4 /* 95% 信頼区間 */
  ddata d02 ;
    retain dose ;
    do x = -3, -2 ;          dose=10**x; output; end;
    do x = -1.5 to 1 by 0.25 ; dose=10**x; output; end;
    do x = 1.5, 2 ;          dose=10**x; output; end;

  data d03 ;
    set d01 d02 ;

  proc nlin data=d03 ;
    parms  theta_maxM=350 theta_minM=100 mu_MEV=-0.5 sigma_MEV=0.5 ;
          g = exp(-(x - mu_MEV)/sigma_MEV) ;
    model y = theta_minM + (theta_maxM - theta_minM) * exp(-g) ;
    output out=Out03 predicted=y_hat STDP=SE L95M=L95 U95M=U95 ;
  run ;
  proc print data=out03 ; run;
```

表 11.29 に示すのは、NLIN プロシジヤの output ステートメントで出力された SAS データセット「do3」の出力結果を Excel に取り込んだ結果である。OBS 変数の 1 から 72 は、元々のデータ y について推定値 \hat{y} などの計算結果が示され、OBS 変数の 73 から 87 が、表 11.14 の内容に対応する。

表 11.29 ゴンペルツ・最大極値曲線の推定値および 95%信頼区間の計算結果

OBS	dose	x	i	y	y_hat	SE	L95	U95
1	0.001	-3	1	102.35	105.13	6.6476	91.86	118.39
2	0.001	-3	2	120.82	105.13	6.6476	91.86	118.39
:								
71	10	1	8	366.2	351.29	10.4449	330.45	372.13
72	10	1	9	332.67	351.29	10.4449	330.45	372.13
OBS	dose	x	i	y	y_hat	SE	L95	U95
73	0.001	-3	.	.	105.13	6.648	91.86	118.39
74	0.01	-2	.	.	105.13	6.648	91.86	118.39
75	0.032	-1.5	.	.	105.13	6.647	91.86	118.39
76	0.056	-1.25	.	.	105.23	6.505	92.26	118.21
77	0.1	-1	.	.	110.10	7.706	94.72	125.48
78	0.178	-0.75	.	.	139.74	14.654	110.50	168.98
79	0.316	-0.5	.	.	197.32	11.563	174.24	220.39
80	0.562	-0.25	.	.	256.29	11.027	234.29	278.30
81	1	0	.	.	299.16	10.805	277.60	320.72
82	1.778	0.25	.	.	325.22	8.478	308.30	342.14
83	3.162	0.5	.	.	339.68	7.612	324.49	354.87
84	5.623	0.75	.	.	347.33	8.848	329.68	364.99
85	10	1	.	.	351.29	10.445	330.45	372.13
86	31.623	1.5	.	.	354.34	12.441	329.51	379.17
87	100	2	.	.	355.12	13.193	328.80	381.45

図 11.16 に示すのは、表 11.29 に示した SAS の出力を Excel に取り込んだデータに基づいて作成している。10 パーセント反応量における逆推定は、表 11.28 のパラメータ $\mu_offsetM$ についての「近似 95%信頼限界」が重ね書きされている。よく見ると曲線の 95%信頼区間の幅よりも広めになっているが、計算方法に依存し微妙な差異となるためである。

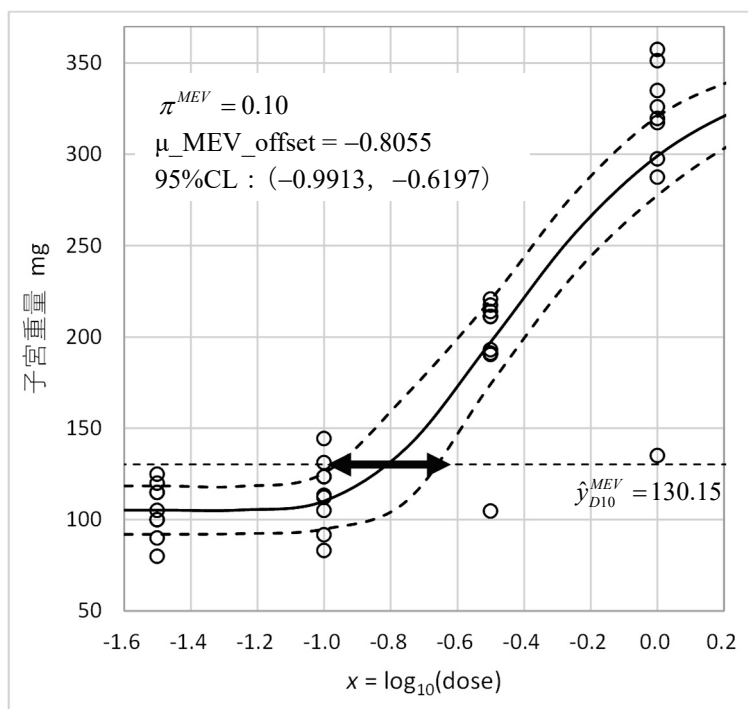


図 11.16 SAS の結果によるゴンペルツ・最大極値曲線に対する 10%点の推定

SAS の結果は正しいのか？

表 11.28 に示した SAS/NLIN プロシジャによる 10 パーセント点の推定値 -0.8055 、95%信頼区間 $(-0.9913, -0.6197)$ は、正しく計算されているのであろうか。表 11.19 に示した JMP の「曲線のあてはめ」では、推定値は一致するが 95%信頼区間 $(-0.9782, -0.6329)$ と一致しない、表 11.23 の JMP の「非線形回帰」の逆推定では、推定値は一致するが 95%信頼区間 $(-0.9813, -0.6298)$ と一致しない。その理由は、JMP では、表 11.23 に示されている逆推定式を用いたデルタ法による近似計算、SAS/NLIN プロシジャでは、非線形回帰による直接推定の違いと推測される。

表 11.10 に示した Excel による残差線形化表の計算シートに、オフセットを組み込んで計算

すると

ξ	$\xi^{(m-1)}$	SE	L95%	U95%
μ^{MEV}	-0.8055	0.0931	-0.9913	-0.6197

との結果が得られ、SAS/NLIN プロシジャ

による計算結果が Excel で再現できる。

12. ミカエリス・メンテン式をめぐる新たな統計解析

ミカエリス・メンテン (*Michaelis-Menten*) 式は、酵素反応速度論における代表的な式であり、統計的にも非線形モデルの典型的な例としても広く知られている。また、ミカエリス・メンテン式の基質濃度 $[S]$ に対し対数を取ると 2 パラメータ・ロジスティック曲線になることも良く知られている。ミカエリス・メンテン式については、知名・岡田 (2014)、「原典からの酵素反応速度論」に歴史的背景も含め、線形化による各種のパラメータの推定法のみならず、非線形回帰によるパラメータの推定のための Excel も含む各種のソフトウェアも紹介されている。ミカエリス・メンテン式に対する数多くの線形化によるパラメータの推定方法は、実験結果を方眼紙にプロットし、回帰直線を目の子であてはめ Y 軸の切片、あるいは、X 軸の切片を方眼紙から読み取り、逆数などの変換で手軽に得られることの特徴が強調されている。したがって、得られたパラメータの標準誤差を求めることは無視されてしまう。そこで、Excel を用いて、各種の線形化によるパラメータの推定値に対し、標準誤差を求める方法を示し、非線形モデルにより推定されたパラメータの標準誤差と比較検討を行なう。さらに、酵素阻害実験データに対する線形化によるパラメータ推定の脆弱性を例示しつつ、非線形モデルによるパラメータ推定の頑健性について例示する。

12.1. ミカエリス・メンテン式によるパラメータ推定

線形化によるパラメータの推定精度

医学研究に関する統計の名著, Armitage ら (2002) , 「Statistical Methods in Medical Research, 4th ed.」 20.5 節の「*Michaelis-Menten assay*」に各種の線形化によるパラメータ推定値の比較検討に加えて、非線形回帰による推定結果も示されている。また、12.4 節の「*Non-linear regression*」にもミカエリス・メンテン式の線形化による推定法と非線形回帰によるパラメータの推定値について比較検討が詳細に示されている。

原著には、ミカエリス・メンテン式に対し、非線形回帰によって得られたミカエリス定数 K_m と最大反応 V_{max} の推定結果が示されている。さらに、ミカエリス・メンテン式に対する各種の線形化によって推定されたパラメータ K_m と V_{max} を比較し、線形化によるパラメータの推定精度が良くないことが示されている。ミカエリス・メンテン式は、式 (12.1) に示すよ

うに分子に基質濃度 $[S]$ と最大反応 V_{max} との積, 分母に基質濃度 $[S]$ に最大反応 V_{max} の 2 分の 1 となる基質濃度であるミカエリス定数 K_m との和で構成されている. 基質濃度 $[S]$ に対する反応速度 V データが得られたときに, 最大反応 V_{max} とミカエリス定数 K_m を推定したい.

a) 非線形・ミカエリス・メンテン式

$$V = \frac{V_{max}[S]}{[S] + K_m}, \quad y = \frac{\theta_{max} x}{x + \mu} \quad (12.1)$$

b) 両逆数・Lineweaver-Burk 変換

$$\frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m}{V_{max}} \frac{1}{[S]}, \quad \frac{1}{y} = \frac{1}{\theta_{max}} + \frac{\mu}{\theta_{max}} \frac{1}{x} \quad (12.2)$$

c) Hanes-Woolf 変換

$$\frac{[S]}{V} = \frac{K_m}{V_{max}} + \frac{1}{V_{max}} [S], \quad \frac{x}{y} = \frac{\mu}{\theta_{max}} + \frac{1}{\theta_{max}} x \quad (12.3)$$

d) Scatchard・Eadie-Hofstee 変換

$$V = V_{max} - K_m \frac{V}{[S]}, \quad y = \theta_{max} - \mu \frac{y}{x} \quad (12.4)$$

b) 両逆数・Lineweaver-Burk 変換は, 式 (12.1) の両辺に逆数を取り, 式を整理して反応速度の逆数 ($1/V$) に対する基質濃度の逆数 ($1/[S]$) に関する線形式となっている. 切片 β_0 が推定したい最大反応の逆数 ($1/V_{max}$), 傾き β_1 がミカエリス定数 K_m と V_{max} の比 (K_m/V_{max}) となり, ミカエリス定数 K_m は, $K_m = \beta_1 / \beta_0$ で推定することができる.

c) Hanes-Woolf 変換は, 式 (12.2) の両逆数・Lineweaver-Burk 変換の両辺に基質濃度 $[S]$ を掛けて整理したものである. 基質濃度 $[S]$ を反応速度 V で除した ($[S]/V$) に対し, 基質濃度 $[S]$ による回帰式であり, 切片 β_0 が (K_m/V_{max}), 傾き β_1 が ($1/V_{max}$) となっている. ミカエリス定数 K_m は, $K_m = \beta_0 / \beta_1$ で推定することができる.

d) Scatchard・Eadie-Hofstee 変換は, さらに技巧的な線形化であり, 式 (12.2) の両逆数変換の両辺に $V \cdot V_{max}$ を掛け

$$\left. \begin{aligned} \frac{\cancel{K} \cdot V_{max}}{\cancel{K}} &= \frac{V \cdot \cancel{V_{max}}}{\cancel{V_{max}}} + \frac{K_m}{\cancel{V_{max}}} \frac{V \cdot \cancel{V_{max}}}{[S]} & \frac{\cancel{x} \cdot \theta_{max}}{\cancel{x}} &= \frac{y \cdot \cancel{\theta_{max}}}{\cancel{\theta_{max}}} + \frac{\mu}{\cancel{\theta_{max}}} \frac{y \cdot \cancel{\theta_{max}}}{x} \\ V_{max} &= V + K_m \frac{V}{[S]} & \theta_{max} &= y + \mu \frac{y}{x} \\ V &= V_{max} - K_m \frac{V}{[S]} & y &= \theta_{max} - \mu \frac{y}{x} \end{aligned} \right\} \quad (12.5)$$

反応速度 V について整理した式となっていて, 切片 β_0 の推定値が V_{max} , 傾き β_1 の推定値が $-K_m$ となり, 回帰のパラメータとして直接得られることに特徴がある.

これらの線形化によって、非線形回帰ではなく身近な回帰分析を用いて最大反応 V_{max} とミカエリス定数 K_m が推定できることは、画期的なことである。数学的には、どのように変換してもパラメータの推定に関して全く問題はないのであるが、実験データには、何らかの誤差がつきまとう。そのために、非線形の基質濃度 $[S]$ に対するミカエリス・メンテン式の反応（反応速度 V ）に対し、誤差が正規分布に従うと仮定した場合に、逆数変換 ($1/V$)、または、 $([S]/V)$ を反応とする回帰分析においては誤差が不均一となる。そのために V_{max} と K_m の推定値が不安定となる。Scatchard・Eadie-Hofstee 変換は、通常回帰分析が前提としている「変数 x には、誤差が含まれない」を無視し、誤差が含まれる y を変数 x に含め $x' = (y/x)$ による回帰分析をするため、推定精度の著しい劣化が懸念される。

Armitage ら(2002) には、a) から d) に基づく解析結果のサマリーに加えて散布図に回帰直線を重ね書きした結果が示されている。ただし、元のデータが示されていない。そのために、拡大コピーした複数枚の散布図の点の位置を計測し、元々の変数である V と $[S]$ に換算し、幾つかの散布図から計測されたデータの平均を用いることにした。表 12.1 に復元したデータ、図 12.1 に散布図を示す。

表 12.1 Armitage ら(2002) の散布図から復元したデータ

	基質濃度	反応速度
	$[S]$	V
	x	y
1	2	65
2	5	362
3	10	581
4	20	686
5	80	985

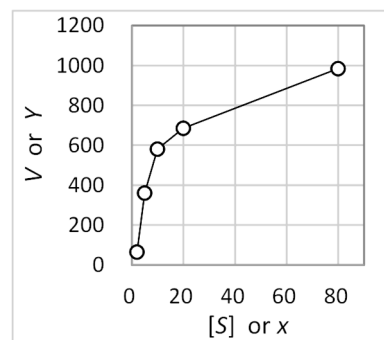


図 12.1 基質濃度 $[S]$ に対する反応速度 V との関係

a) 非線形回帰によるミカエリス・メンテン式のあてはめ

式 (12.1) に示したミカエリス・メンテン式で、反応の V は、Velocity 反応速度であり、 $[S]$ は、Substrate concentration 基質濃度である。パラメータとしての V_{max} は、反応速度 V の最大値、 K_m は V_{max} の2分の1となる基質濃度 $[S]$ であり、ミカエリス定数として知られている。統計的には、 $V = y$, $[S] = x$, $V_{max} = \theta_{max}$, $K_m = \mu$ すると理解しやすい。

非線形モデルであるミカエリス・メンテン式を表 12.1 に示したデータにあてはめ、 V_{max} および K_m を推定し、それらの標準誤差の導出するためには、ミカエリス・メンテン式のパラメータに関する偏微分式が必要となる。式 (12.1) に示したミカエリス・メンテン式をパラメータ V_{max} および K_m で偏微分した式は、

$$z_1 = \frac{\partial V}{\partial V_{max}} = \frac{[S]}{[S] + K_m}, \quad z_1 = \frac{\partial y}{\partial \theta_{max}} = \frac{x}{x + \mu} \quad (12.6)$$

$$z_2 = \frac{\partial V}{\partial K_m} = \frac{-V_{max}[S]}{([S] + K_m)^2}, \quad z_2 = \frac{\partial y}{\partial \mu} = \frac{-\theta_{max} x}{(x + \theta_{max})^2} \quad (12.7)$$

として得られる。

表 12.2 に示すように、Excel のソルバーにより推定された \hat{V}_{max} および \hat{K}_m を使い、 \hat{V} (V ハット) が推定されている。さらに、残差を $\hat{\varepsilon} = V - \hat{V}$ とし、残差 $\hat{\varepsilon}$ に対する残差平方和 $S_e = 14904.51$ 、残差平方和 S_e を自由度 $df = [(データ数 n) - (推定に用いたパラメータ数 m)]$ で除した誤差分散の推定値が、 $\hat{\sigma}^2 = 14904.51 / (5 - 2) = 4968.17$ として計算されている。これらは、パラメータの初期値を ($\hat{V}_{max} = 1000$, $\hat{K}_m = 10$) とし、残差平方和 S_e が最小になるように、Excel のソルバーで、2つのパラメータ (\hat{V}_{max} , \hat{K}_m) を変化させて得られた結果である。推定されたパラメータ ($\hat{V}_{max} = 1130.00$, $\hat{K}_m = 11.64$) の標準誤差 SE を計算するためには、ミカエリス・メンテン式をパラメータで偏微分した式 (12.6), および、式 (12.7) による微係数 (z_1 , z_2) についての (5 行×2 列) の微係数行列 \mathbf{Z} が必要となる。

表 12.2 ミカエリス・メンテン式を非線形モデルとした解析結果

初期値				θ^{\wedge}	パラメータの共分散行列		分散
1000		$\theta^{\wedge}_{max} = V^{\wedge}_{max} =$		1130.00	10885.37	262.53	10885.37
10		$\mu = K^{\wedge}_m =$		11.64	$\sigma^{\wedge 2} = S_e / 3$	262.53	9.1359
		$S_e =$		14904.51	4968.17	$\Sigma(\theta^{\wedge}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \sigma^{\wedge 2}$	
	[S]	V	V^{\wedge}	$\varepsilon^{\wedge} = V - V^{\wedge}$	$\partial_{V^{\wedge}_{max}}$	$\partial_{K^{\wedge}_m}$	
No	x	y	y^{\wedge}	$\varepsilon^{\wedge} = y - y^{\wedge}$	z_1	z_2	
1	2	65	165.69	-100.69	0.1466	-12.147	
2	5	362	339.54	22.46	0.3005	-20.405	
3	10	581	522.18	58.82	0.4621	-24.130	
4	20	686	714.29	-28.29	0.6321	-22.575	
5	80	985	986.47	-1.47	0.8730	-10.765	
		$V^{\wedge} = V^{\wedge}_{max} [S] / ([S] + K^{\wedge}_m)$		$z_1 = [S] / ([S] + K^{\wedge}_m)$	$z_2 = -V^{\wedge}_{max} [S] / ([S] + K^{\wedge}_m)^2$		

得られた (5 行×2 列) の微係数行列 \mathbf{Z} に対し Excel の行列関数、Mmult (), Transpose (), Minverse () 関数を用いて (2×2) のパラメータの共分散行列 $\Sigma(\hat{\theta})$ が

$$\left. \begin{aligned} \Sigma(\hat{\theta}) &= (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2 \\ &= \text{Minverse}(\text{Mmult}(\text{Transpose}(\mathbf{Z} \text{ の範囲}), \mathbf{Z} \text{ の範囲})) * 4968.17 \\ &= \begin{bmatrix} 10885.37 & 262.53 \\ 262.53 & 9.1359 \end{bmatrix} \end{aligned} \right\} \quad (12.8)$$

として求められている。このパラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素が、(\hat{V}_{max} , \hat{K}_m) の分散であり、 $Var(\hat{V}_{max}) = 10885.37$, $Var(\hat{K}_m) = 9.1359$ が得られている。

表 12.2 で得られた推定値と分散を用いて、パラメータの 95%信頼区間を求めた結果を表 12.3 示す。最大反応 $\hat{V}_{max}=1130.00$ の分散は、パラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素から $Var(\hat{V}_{max})=10885.37$ ，標準誤差は、分散の平方根で $SE(\hat{V}_{max})=104.33$ ，95%信頼区間が (797.97, 1462.04) と推定されている。同様に、ミカエリス定数は、 $\hat{K}_m=11.64$ で、95%信頼区間は、自由度 (5-2) の t 分布の両側 5 パーセント点 $T.inv.2T(0.05, 5-2)=3.1824$ を用いて (2.02, 21.26) と推定されている。

表 12.3 非線形回帰によるパラメータの 95%信頼区間

	推定値	分散	SE	L95%	U95%
$\theta^{\wedge}_{max}=V^{\wedge}_{max}=\mu^{\wedge}=K^{\wedge}_m=$	1130.00	10885.37	104.33	797.97	1462.04
	11.64	9.1359	3.02	2.02	21.26
		$t_{0.05}(5-2)=$	3.1824		

推定されたパラメータを用いて、ミカエリス・メンテン曲線の推定値および 95%信頼区間を計算する。表 12.4 に示すように滑らかな曲線が描けるように基質濃度 $[S]$ を設定し、反応速度の推定値を \hat{V} を推定し、パラメータに関する微係数ベクトル $\mathbf{z}=[z_1 \ z_2]$ を求める。推定値 \hat{V} の分散 $Var(\hat{V})$ は、パラメータの共分散行列 $\Sigma(\hat{\theta})$ に関する微係数ベクトル \mathbf{z} の 2 次形式 $Var(\hat{V})=\mathbf{z}\Sigma(\hat{\theta})\mathbf{z}^T$ によって求められている。さらに分散 $Var(\hat{V})$ の平方根を取った標準誤差 $SE(\hat{V})$ から 95%信頼区間が推定されている。最後の行に $[S]=100000$ を設定し、表 12.3 に示した \hat{V}_{max} の推定値と 95%信頼区間 (797.97, 1462.04) に漸近することを確かめている。

表 12.4 ミカエリス・メンテン曲線の 95%信頼区間

		θ^{\wedge}	$\Sigma(\theta^{\wedge})=(\mathbf{Z}^T\mathbf{Z})^{-1}\sigma^{\wedge 2}$				
$V^{\wedge}_{max}=\mu^{\wedge}=K^{\wedge}_m=$		1130.00	10885.37	262.53			
		11.64	262.53	9.1359			
x	y^{\wedge}	$\partial_{V^{\wedge}_{max}}$	$\partial_{K^{\wedge}_m}$	分散	標準誤差	95%信頼区間	
$[S]$	V^{\wedge}	z_1	z_2	$Var(V^{\wedge})$	$SE(V^{\wedge})$	L95%	U95%
0.1	9.6	0.0085	-0.820	3.3	1.8	3.9	15.4
1.0	89.4	0.0791	-7.073	231.3	15.2	41.0	137.8
5.0	339.5	0.3005	-20.405	1567.4	39.6	213.5	465.5
10.0	522.2	0.4621	-24.130	1789.2	42.3	387.6	656.8
20.0	714.3	0.6321	-22.575	1512.7	38.9	590.5	838.1
50.0	916.6	0.8112	-14.870	2849.1	53.4	746.7	1086.5
100.0	1012.2	0.8957	-9.066	5220.6	72.3	782.2	1242.1
300.0	1087.8	0.9626	-3.491	8434.4	91.8	795.5	1380.1
100000	1129.9	0.9999	-0.011	10876.9	104.3	798.0	1461.8
		$Var(V^{\wedge})=\mathbf{z}\Sigma(\theta^{\wedge})\mathbf{z}^T$		$t_{0.05}(5-2)=$	3.1824		

ミカエリス・メンテン式で推定された反応速度 \hat{V} を計算し、散布図に重ね書きすること自体は容易なのであるが、推定値 \hat{V} の 95%信頼区間も含めて図示するためには、表 12.4 に示したように推定値 \hat{V} の分散 $Var(\hat{V})=\mathbf{z}\Sigma(\hat{\theta})\mathbf{z}^T$ を計算する必要がある。表 12.2 上段にさりげなく (2×2) のパラメータの共変量 $\Sigma(\hat{\theta})$ の計算結果を示したのであるが、これを求めるた

めに (5行×2列) の微係数行列 \mathbf{Z} を用いた行列計算 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$ を必要としている。行列計算の式に対し自ら計算するすべがない時代では、拒絶反応が起きることは当然であり、一般的な統計関連の成書では徹底的に避けられてきた。その結果として、回帰直線の場合を除いて、各種の推定曲線に対する 95%信頼区間の計算方法が示されることが皆無となっているのが現状である。

現代の算盤である Excel の計算シート上では、統計解析で必要となる t 分布などの各種の分布関数の計算も手軽にでき、さらに、行列計算のための関数も用意されているので、Excel の計算シートを用いた統計解析を前提にすることにより、伝統的な統計解析の軛（くびき）を乗り越えることが可能となった。さて、推定値 \hat{V} の分散 $Var(\hat{V}) = \mathbf{z} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{z}^T$ の計算手順について段階を追って示す。

表 12.4 に示した $[S]=10.0$ に対する反応速度の推定値 \hat{V} は、

$$\hat{V}_{[S]=10} = \frac{\hat{V}_{max}[S]}{[S] + \hat{K}_m} = \frac{1130.00 \times 10.0}{10.0 + 11.64} = 522.2 \quad (12.9)$$

として計算されている。 \hat{V} を \hat{V}_{max} および \hat{K}_m で偏微分した式は、

$$z_1 = \frac{\partial \hat{V}}{\partial \hat{V}_{max}} = \frac{[S]}{[S] + \hat{K}_m} = \frac{10.0}{10.0 + 11.64} = 0.4621 \quad (12.10)$$

$$z_2 = \frac{\partial \hat{V}}{\partial \hat{K}_m} = \frac{-V_{max}[S]}{([S] + K_m)^2} = \frac{-1130.00 \times 10.0}{(10.0 + 11.64)^2} = -24.130 \quad (12.11)$$

であり、 z_1 と z_2 をベクトル $\mathbf{z} = [z_1 \ z_2]$ とする。推定値 \hat{V} の分散 $Var(\hat{V}) = \mathbf{z} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{z}^T$ は、以下に示すように、Excel の行列関数 Mmult () 関数、および、Transpose () 関数を使用し、

[S]=10.0 の場合		パラメータの共分散		転置
z_1	z_2	$\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$		$[z_1 \ z_2]^T$
0.4621	-24.130	10885.37	262.53	0.4621
		262.53	9.14	-24.1305
		$\mathbf{z} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$		
		-1304.84	-99.135	0.4621
		-1304.84	-99.135	-24.1305
		$Var(\hat{V}) = \mathbf{z} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{z}^T$		1789.188
=Mmult (Mmult (\mathbf{z} の範囲, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の範囲), Transpose (\mathbf{z} の範囲))				

として計算されている。

図 12.2 左は、表 12.4 で求められた 95%信頼区間をミカエリス・メンテン曲線に重ね書きした結果である。図 12.2 右は、Excel の X 軸の書式設定で「対数目盛」を選択した結果で、2 パラメータ・ロジスティック曲線となっている。

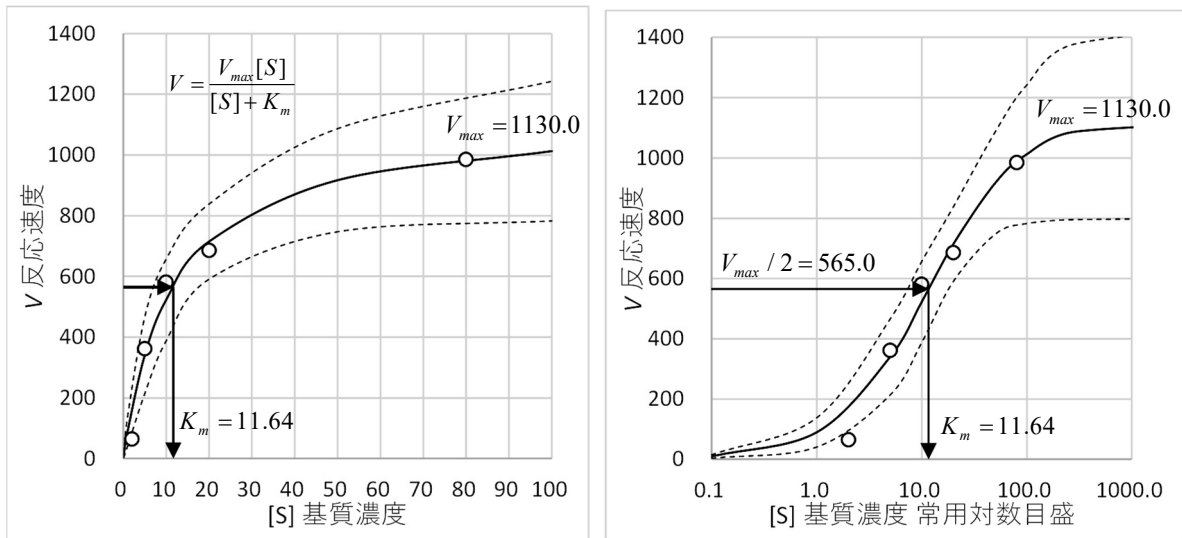


図 12.2 推定されたミカエリス・メンテン曲線に対する 95%信頼区間

式 (12.1) のミカエリス・メンテン式に対し、分子と分母を $[S]$ で除し、分母の $(K_m/[S])$ に対し指数と対数を $\exp[\ln(\cdot)]$ のようにセットで適用し、整理すると対数濃度 $\ln([S])$ に関して 2 パラメータ・ロジスティック曲線式が得られる。

$$\begin{aligned}
 V &= \frac{V_{max}[S]}{[S] + K_m} \\
 &= \frac{V_{max}}{1 + \exp\left[\ln\left(\frac{K_m}{[S]}\right)\right]} \\
 &= \frac{V_{max}}{1 + \exp\{-[\ln([S]) - \ln(K_m)]\}}
 \end{aligned}
 \quad , \quad
 \begin{aligned}
 y &= \frac{\theta_{max} x}{x + \mu} \\
 &= \frac{\theta_{max}}{1 + \exp\left[\ln\left(\frac{\mu}{x}\right)\right]} \\
 &= \frac{\theta_{max}}{1 + \exp\{-[\ln(x) - \ln(\mu)]\}}
 \end{aligned}
 \quad \left. \vphantom{\begin{aligned} V &= \frac{V_{max}[S]}{[S] + K_m} \right.} \right\} \quad (12.13)$$

b) 両逆数・Lineweaver-Burk 変換

ミカエリス・メンテン式を線形化するために、測定誤差が含まれる反応速度 V について逆数とするので、基質濃度 $[S]$ が低い場合に反応速度 V は大きくなり、 $(1/V)$ の誤差も相対的に大きくなる。その結果、切片として推定される $(1/V_{max})$ は、必然的に回帰直線の外挿となり、95%信頼区間も広がり、推定結果も不安定になる。

$$\text{式 (12.2) : } \quad \frac{1}{V} = \frac{1}{V_{max}} + \frac{K_m}{V_{max}} \frac{1}{[S]}, \quad \frac{1}{y} = \frac{1}{\theta_{max}} + \frac{\mu}{\theta_{max}} \frac{1}{x}$$

表 12.5 に示すのは、両逆数・Lineweaver-Burk 変換によるパラメータ推定のための Excel シートで、基質濃度 $[S]$ の逆数、反応速度 V の逆数に対する回帰分析の結果である。回帰分析は、Excel の分析ツールの「回帰分析」、または、行列計算で行ってきたのであるが、ここでは、必要とするパラメータをコンパクトに推定できる Excel の「ソルバー」を用いて残差平方 S_e を最小になるように $(\hat{\beta}_0, \hat{\beta}_1)$ を変化させた結果である。なお、線形モデルなので、

初期値にどんな数値を設定しても解が求まる．パラメータの共分散行列 $\Sigma(\hat{\beta})$ は， $[x_0, 1/x]$ の (5行×2列) のデザイン行列 X を用いて， $\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$ によって求められている． $\Sigma(\hat{\beta})$ の対角要素が，パラメータの分散となり．その平方根の標準誤差 SE を用いてパラメータの 95%信頼区間を求めている．

表 12.5 両逆数・Lineweaver-Burk 変換によるパラメータ推定

ソルバーで S_e を最小化		$\beta_0 \hat{=}$	-0.00074	$\sigma^2 \hat{=}$	$3.39E-06$	$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$		分散	SE			
		$\beta_1 \hat{=}$	0.03021	$S_e \hat{=}$	$1.02E-05$	$1.33E-06$	$-3.80E-06$		$1.33E-06$	0.0012		
[S]	V	1/[S]	1/V	(1/V) ²	残差	$-3.80E-06$	$2.20E-05$		$2.20E-05$	0.0047		
x	y	x_0	1/x	1/y	(1/y) ²	$\epsilon \hat{=}$		推定値	SE	L95%	U95%	
2	65	1	0.5000	0.0154	0.0144	0.0010		$\beta_0 \hat{=} 1/V_{max} =$	-0.0007	$1.15E-03$	-0.0044	0.0029
5	362	1	0.2000	0.0028	0.0053	-0.0025		$\beta_1 \hat{=} K_m/V_{max} =$	0.0302	$4.69E-03$	0.0153	0.0451
10	581	1	0.1000	0.0017	0.0023	-0.0006		$V_{max} = 1/\beta_0 \hat{=}$	-1345.7	2091.6	-8002.0	5310.5
20	686	1	0.0500	0.0015	0.0008	0.0007		$K_m = \beta_1/\beta_0 \hat{=}$	-40.7	58.9	-228.2	146.9
80	985	1	0.0125	0.0010	-0.0004	0.0014		$t_{0.05}(5-2) =$	3.1824			

回帰分析で推定されたパラメータ ($\hat{\beta}_0, \hat{\beta}_1$) を用いて，最大反応 \hat{V}_{max} とミカエリス定数 \hat{K}_m が，表 12.5 右下段に

$$\hat{V}_{max} = 1/\hat{\beta}_0 = 1/(-0.00074) = -1345.7 \quad (12.14)$$

$$\hat{K}_m = \frac{\hat{K}_m/\hat{V}_{max}}{1/\hat{V}_{max}} = \frac{\hat{\beta}_1}{\hat{\beta}_0} = \frac{0.03021}{-0.00074} = -40.7 \quad (12.15)$$

として推定されている．最大反応 \hat{V}_{max} の分散 $Var(\hat{V}_{max})$ は，切片 $\hat{\beta}_0$ の逆数 ($1/\hat{\beta}_0$) を $\hat{\beta}_0$ および $\hat{\beta}_1$ で偏微分し，

$$d_1 = \frac{\partial(1/\hat{\beta}_0)}{\partial \hat{\beta}_0} = -\frac{1}{\hat{\beta}_0^2} = -\frac{1}{(-0.00074)^2} = -1810969, \quad d_2 = \frac{\partial(1/\hat{\beta}_0)}{\partial \hat{\beta}_1} = 0 \quad (12.16)$$

パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する $[d_1 \ d_1]$ の 2 次形式により

$$\begin{aligned} Var(V_{max}) &= [d_1 \ d_1] \Sigma(\hat{\beta}) [d_1 \ d_2]^T \\ &= [-1810969 \ 0] \begin{bmatrix} 1.33E-06 & -3.80E-06 \\ -3.80E-06 & 2.20E-05 \end{bmatrix} \begin{bmatrix} -1810969 \\ 0 \end{bmatrix} \\ &= (2091.6)^2 \end{aligned} \quad (12.17)$$

として求められている．推定値は， $\hat{V}_{max} = -1345.7$ であり，奇異な結果となっている．また 95%信頼区間は， $(-8002.4, 5311.0)$ と凄まじい範囲となっており，非線形回帰で求めた $\hat{V}_{max}^{(非線形)} = 1130.0$ の推定値を含んでいるので，統計的には許容範囲であるが，まったく使い物にならない．

ミカエリス定数 K_m の推定値は， $(\hat{\beta}_1/\hat{\beta}_0)$ なので， $\hat{\beta}_0$ および $\hat{\beta}_1$ で偏微分し

$$d_1 = \frac{\partial(\hat{\beta}_1/\hat{\beta}_0)}{\partial \hat{\beta}_0} = -\frac{\hat{\beta}_1}{\hat{\beta}_0^2} = -54713.7, \quad d_2 = \frac{\partial(\hat{\beta}_1/\hat{\beta}_0)}{\partial \hat{\beta}_1} = \frac{1}{\hat{\beta}_0} = -1345.766 \quad (12.18)$$

\hat{V}_{max} と同様にパラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する $[d_1 \ d_1]$ の2次形式により

$$\left. \begin{aligned} \text{Var}(\hat{K}_m) &= [d_1 \ d_2] \Sigma(\hat{\beta}) [d_1 \ d_2]^T \\ &= [-54710 \quad -1345.72] \begin{bmatrix} 1.33E-06 & -3.80E-06 \\ -3.80E-06 & 2.20E-05 \end{bmatrix} \begin{bmatrix} -54710 \\ -1345.72 \end{bmatrix} \\ &= (58.9)^2 \end{aligned} \right\} \quad (12.19)$$

によって求められている。ミカエリス定数の推定値は $\hat{K}_m = -40.7$ と基質濃度が負となる不可思議な結果であるが、95%信頼区間 $(-228.2, 146.9)$ の中に非線形モデルで求めた $\hat{K}_m^{(\text{非線形})} = 11.64$ が含まれているので、 \hat{V}_{max} と同様に統計的には許容範囲であるが、実験結果として提示することは不可能である。

表 12.6 に示すのは、推定された $(\hat{\beta}_0, \hat{\beta}_1)$ を用いて、回帰の推定値 $(\widehat{1/y}) = \hat{\beta}_0 + \hat{\beta}_1(1/x)$ を求め、推定値の分散を $\Sigma(\hat{\beta})$ に関する $[1 \ (1/x)_i]$ の2次形式により

$$\text{Var}(\widehat{1/y}) = [1 \ (1/x)_i] \Sigma(\hat{\beta}) [1 \ (1/x)_i]^T \quad (12.20)$$

によって計算し、95%信頼区間を求めている。図 12.3 に示すのは、与えられたデータに対し回帰直線および95%信頼区間を重ね書きした結果であり、妥当な結果と思われる。ただし、切片 $\hat{\beta}_0$ が、きわどいところでマイナスとなっているために、式 (12.14) に示したように最大反応 $\hat{V}_{max} = -1345.7$ が、不可解な推定値となってしまう。

表 12.6 両逆数・Lineweaver-Burk 変換による回帰直線の95%信頼区間

	1/[S]	1/Y	分散	SE	95%信頼区間	
x_0	1/x	(1/y)^	[(1/y)^]	[(1/y)^]	L 95%	U 95%
1	0.000	-0.0007	1.33E-06	0.0012	-0.0044	0.0029
1	0.050	0.0008	1.01E-06	0.0010	-0.0024	0.0040
1	0.100	0.0023	7.94E-07	0.0009	-0.0006	0.0051
1	0.200	0.0053	6.95E-07	0.0008	0.0026	0.0080
1	0.300	0.0083	1.04E-06	0.0010	0.0051	0.0116
1	0.400	0.0113	1.82E-06	0.0013	0.0070	0.0156
1	0.500	0.0144	3.04E-06	0.0017	0.0088	0.0199
1	0.600	0.0174	4.71E-06	0.0022	0.0105	0.0243
			$t_{0.05}(5-2) =$	3.1824		

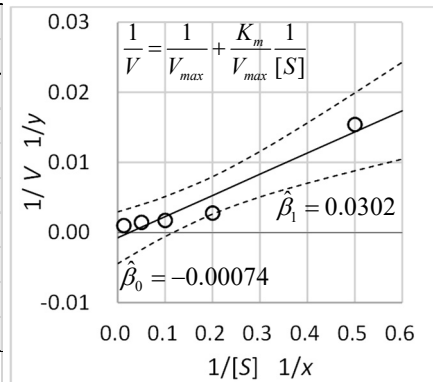


図 12.3 両逆数・Lineweaver-Burk 変換による回帰直線

両逆数・Lineweaver-Burk 変換による線形化によるパラメータの推定値が不安定であることは、実際に使用した化学者・技術者たちも実感したに違いない。そのために新たな線形変換式が、次から次へと考案されたと理解される。ただし、両逆数・Lineweaver-Burk 変換は、ミカエリス・メンテン式のパラメータ推定のための代表的な線形変換式として、消えることなく使い続けられている。

c) Hanes-Woolf 変換

Hanes-Woolf 変換は、両逆数・Lineweaver-Burk 変換の両辺に基質濃度 [S] を掛けて整理した線形変換式である。反応速度 V の逆数に基質濃度 [S] を掛けた ($[S]/V$) を新たな反応とし、基質濃度 [S] に対する回帰式となっていて、切片が $\beta_0 = K_m / V_{max}$ 、傾きが $\beta_1 = 1/V_{max}$ となっている。これにより、両逆数・Lineweaver-Burk 変換で、切片の推定値が、 $(1/V_{max})$ であることの不安定さの解消を狙ったものと思われる。

$$\text{式 (12.3) : } \frac{[S]}{V} = \frac{K_m}{V_{max}} + \frac{1}{V_{max}}[S], \quad \frac{x}{y} = \frac{\mu}{\theta_{max}} + \frac{1}{\theta_{max}}x$$

表 12.7 に示すのは、Hanes-Woolf 変換によるパラメータ推定のための Excel シートで、基質濃度 [S] に対し、[S] を反応速度 V で除した反応 ($[S]/V$) に対する回帰分析の結果である。回帰パラメータの推定は、「ソルバー」を用いて残差平方 S_e を最小にするように ($\hat{\beta}_0, \hat{\beta}_1$) を変化させた手順を用いている。パラメータの共分散行列 $\Sigma(\hat{\beta})$ は、変数 (x_0, x) に対する (5 行×2 列) のデザイン行列 X とし、 $\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$ によって求め、 $\Sigma(\hat{\beta})$ の対角要素の平方根をパラメータの標準誤差 SE と 95% 信頼区間を求めている。

表 12.7 Hanes-Woolf 変換によるパラメータ推定

ソルバーで		$\hat{\beta}_0 =$	1.57E-02	$\hat{\sigma}^2 =$	8.90E-05	$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$		分散	SE			
S_e を最小化		$\hat{\beta}_1 =$	8.00E-04	$S_e =$	2.67E-04	2.94E-05	-4.97E-07	2.94E-05	5.42E-03			
[S]	V	行列 X		[S]/V	推定値	残差	-4.97E-07	2.12E-08	2.12E-08	1.46E-04		
x	y	x_0	x	(x/y)	(x/y) [^]	ϵ^{\wedge}	推定値		SE	L95%	U95%	
2	65	1	2	0.0308	0.0173	0.0135	$\hat{\beta}_0 = K_m / V_{max} =$	0.0157	5.42E-03	-0.0016	0.0330	
5	362	1	5	0.0138	0.0197	-0.0059	$\hat{\beta}_1 = 1/V_{max} =$	0.0008	1.46E-04	0.0003	0.0013	
10	581	1	10	0.0172	0.0237	-0.0065	$V_{max} = 1/\hat{\beta}_1 =$	1249.38	227.45	525.53	1973.24	
20	686	1	20	0.0292	0.0317	-0.0026	$K_m = \hat{\beta}_0 / \hat{\beta}_1 =$	19.62	9.44	-10.42	49.66	
80	985	1	80	0.0812	0.0797	0.0015	$t_{0.05}(5-2) =$	3.1824				

最大反応の推定値 \hat{V}_{max} およびミカエリス定数の推定値 \hat{K}_m は、回帰分析で求められたパラメータ ($\hat{\beta}_0, \hat{\beta}_1$) を用いて、

$$\hat{V}_{max} = \frac{1}{\hat{\beta}_1} = \frac{1}{0.0008004} = 1249.4 \tag{12.21}$$

$$\hat{K}_m = \frac{\hat{K}_m / \hat{V}_{max}}{1 / \hat{V}_{max}} = \frac{\hat{\beta}_0}{\hat{\beta}_1} = \frac{0.0157}{0.0008004} = 19.62 \tag{12.22}$$

として推定されている。最大反応 \hat{V}_{max} の分散 $Var(\hat{V}_{max}) = Var(1/\hat{\beta}_1)$ は、 $(1/\hat{\beta}_1)$ を $\hat{\beta}_0$ および $\hat{\beta}_1$ で偏微分し、

$$d_1 = \frac{\partial(1/\hat{\beta}_1)}{\partial\hat{\beta}_0} = 0, \quad d_2 = \frac{\partial(1/\hat{\beta}_1)}{\partial\hat{\beta}_1} = -\frac{1}{\hat{\beta}_1^2} = -\frac{1}{0.0008004^2} = -1560960 \tag{12.23}$$

$\Sigma(\hat{\beta})$ に関する $[d_1 \ d_1]$ の 2 次形式によって

$$\begin{aligned} \text{Var}(\hat{V}_{max}) &= [d_1 \ d_1] \Sigma(\hat{\beta}) [d_1 \ d_2]^T \\ &= [0 \ -1560960] \begin{bmatrix} 2.94E-05 & -4.97E-07 \\ -4.97E-07 & 2.12E-08 \end{bmatrix} \begin{bmatrix} 0 \\ -1560960 \end{bmatrix} \\ &= (227.45)^2 \end{aligned} \quad (12.24)$$

として求められている. 最大反応 \hat{V}_{max} の 95%信頼区間は, (525.5, 1973.3) となっているので, 非線形モデルで求めた $\hat{V}_{max}^{(非線形)} = 1130.00$ の推定値を含んでいる

ミカエリス定数 K_m の推定値は, $\hat{K}_m = \hat{\beta}_0 / \hat{\beta}_1$ なので, $\hat{\beta}_0$ および $\hat{\beta}_1$ で偏微分し

$$d_1 = \frac{\partial(\hat{\beta}_0 / \hat{\beta}_1)}{\partial \hat{\beta}_0} = \frac{1}{\hat{\beta}_1} = 1249.38, \quad d_2 = \frac{\partial(\hat{\beta}_0 / \hat{\beta}_1)}{\partial \hat{\beta}_1} = -\frac{\hat{\beta}_0}{\hat{\beta}_1^2} = -24513.36 \quad (12.25)$$

最大反応 \hat{V}_{max} と同様にパラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する $[d_1 \ d_1]$ の 2 次形式によって

$$\begin{aligned} \text{Var}(\hat{K}_m) &= [d_1 \ d_2] \Sigma(\hat{\beta}) [d_1 \ d_2]^T \\ &= [1249.38 \ -24513.36] \begin{bmatrix} 2.94E-05 & -4.97E-07 \\ -4.97E-07 & 2.12E-08 \end{bmatrix} \begin{bmatrix} 1249.38 \\ -24513.36 \end{bmatrix} \\ &= (9.44)^2 \end{aligned} \quad (12.26)$$

によって求められている. ミカエリス定数の推定値は $\hat{K}_m = 19.6$ であり, 95%信頼区間 (-10.42, 49.66) の中に非線形回帰で求めた $\hat{K}_m^{(非線形)} = 11.64$ が含まれている.

表 12.8 に示すのは, *Hanes-Woolf* 変換によって求められた回帰直線および 95%信頼区間を示したものである. 分散の計算は, $[1 \ x_i]$ のパラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式によって計算している.

表 12.8 *Hanes-Woolf* 変換による回帰直線の 95%信頼区間

	[S]	[S]/Y	分散	SE	95%信頼区間	
x_0	x	(x/y)^ [^]	(x/y)^ [^]	(x/y)^ [^]	L 95%	U 95%
1	0	0.0157	2.94E-05	0.0054	-0.0016	0.0330
1	10	0.0237	2.16E-05	0.0046	0.0089	0.0385
1	20	0.0317	1.80E-05	0.0042	0.0182	0.0452
1	30	0.0397	1.87E-05	0.0043	0.0259	0.0535
1	50	0.0557	3.28E-05	0.0057	0.0375	0.0740
1	60	0.0637	4.62E-05	0.0068	0.0421	0.0854
1	70	0.0717	6.39E-05	0.0080	0.0463	0.0972
1	80	0.0797	8.58E-05	0.0093	0.0503	0.1092
1	90	0.0877	1.12E-04	0.0106	0.0541	0.1214
1	100	0.0957	1.42E-04	0.0119	0.0578	0.1337
			$t_{0.05}(5-2) =$	3.1824		

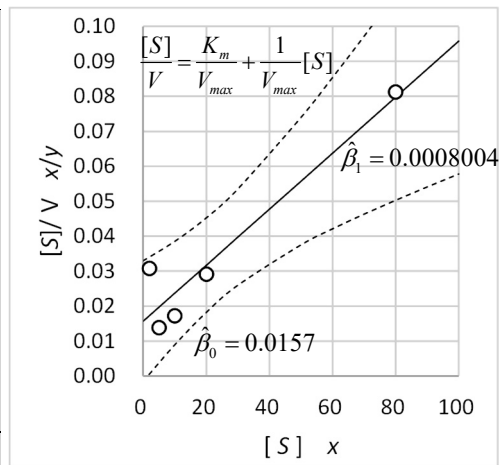


図 12.4 *Hanes-Woolf* 変換による回帰直線

d) *Scatchard・Eadie-Hofstee* 変換

Hanes-Woolf 変換により，両逆数・*Lineweaver-Burk* 変換で反応速度の最大反応 V_{max} の逆数 ($1/V_{max}$) が，切片の推定値 $\hat{\beta}_0$ から傾きの推定値 $\hat{\beta}_1$ に変更され， V_{max} の推定が安定したのであるが，ミカエリス定数 K_m の推定に $\hat{K}_m = \hat{\beta}_0 / \hat{\beta}_1$ のようにパラメータの比で求める必要があり，扱いにくかったのであろう．さらなる改良が *Scatchard・Eadie-Hofstee* 変換であり，

$$\text{式 (12.4) : } V = V_{max} - K_m \frac{V}{[S]}, \quad y = \theta_{max} - \mu \frac{y}{x}$$

切片の推定値が V_{max} ，傾きの推定値が $-K_m$ となり，回帰分析のパラメータとして得られることに特徴がある．

ただし，測定誤差が含まれる反応速度 V （反応 y ）を変数に含んだ回帰式となっているので，一般的な回帰分析の前提「説明変数 x に測定誤差を含まない」を満たしていないことに注意が必要である．測定誤差が小さく精度の高い実験データであれば，*Scatchard・Eadie-Hofstee* 変換は，魅力的な方法であるが，測定誤差が大きい場合には，パラメータの推定値が不安定になり，使用に耐えないのではないかと危惧される．

表 12.9 に示したのは，これまでと同様の計算シートであり，*Scatchard・Eadie-Hofstee* 変換に合わせたデータがセットされている．切片の推定値が最大反応 $\hat{V}_{max} = 811.76$ ，95%信頼区間が $(-320.74, 1944.25)$ と *Hanes-Woolf* 変換に比べてかなり拡大している．傾きの推定値が負のミカエリス定数 $-K_m$ となるので $\hat{K}_m = 6.58$ ，95%信頼区間が $(-17.56, 30.73)$ と推定されている．

表 12.9 *Scatchard・Eadie-Hofstee* 変換によるパラメータ推定

ソルバーで		$\beta^{\wedge}_0 =$	811.76	$\sigma^{\wedge 2} =$	1.27E+05	$\Sigma(\beta^{\wedge}) = (X^T X)^{-1} \sigma^{\wedge 2}$		分散	SE		
S_e を最小化		$\beta^{\wedge}_1 =$	-6.5826	$S_e =$	3.82E+05	126634	-2412.7	126634	355.86		
[S]	V	行列 X		V	推定値	残差					
x	y	x_0	y/x	y	y^{\wedge}	ϵ^{\wedge}		推定値	SE	L95%	U95%
2	65	1	32.50	65	597.82	-532.82	$\beta^{\wedge}_0 = V_{max} =$	811.76	355.86	-320.74	1944.25
5	362	1	72.40	362	335.18	26.82	$\beta^{\wedge}_1 = -K_m =$	-6.58	7.59	-30.73	17.56
10	581	1	58.10	581	429.31	151.69	$V_{max} = \beta^{\wedge}_0 =$	811.76	355.86	-320.74	1944.25
20	686	1	34.30	686	585.98	100.02	$K_m = -\beta^{\wedge}_1 =$	6.58	7.59	-17.56	30.73
80	985	1	12.31	985	730.71	254.29		$t_{0.05}(5-2) =$	3.1824		

表 12.10 に示すのは，推定されたパラメータを用い，散布図に *Scatchard・Eadie-Hofstee* 変換データを用いた回帰直線と 95%信頼区間を重ね書きした結果である．X 軸の変数は， x ではなく (y/x) であり，そのために反応 y に対する測定誤差の影響を受け 95%信頼区間が大きく広がっている．なお，反応の推定値 \hat{V} に対する分散 $Var(\hat{V})$ は，これまでと同様に $\Sigma(\hat{\beta})$ に関する $[x_0 \ y/x]$ の 2 次形式によって求めている．

表 12.10 *Scatchard・Eadie-Hofstee* 変換による回帰直線の 95%信頼区間

	$V/[S]$	V	分散	標準誤差	95%信頼区間	
x_0	y/x	y^{\wedge}	$Var(y^{\wedge})$	$SE(y^{\wedge})$	$L95\%$	$U95\%$
1	0	811.8	1.27E+05	355.9	-320.7	1944.3
1	10	745.9	8.41E+04	290.1	-177.2	1669.0
1	20	680.1	5.31E+04	230.5	-53.6	1413.8
1	30	614.3	3.37E+04	183.5	30.3	1198.2
1	50	482.6	2.92E+04	171.0	-61.6	1026.8
1	60	416.8	4.43E+04	210.5	-253.0	1086.6
1	70	351.0	7.09E+04	266.2	-496.2	1198.1
1	80	285.2	1.09E+05	330.0	-765.2	1335.5
1	90	219.3	1.59E+05	398.1	-1047.7	1486.4
1	100	153.5	2.20E+05	468.6	-1337.9	1644.9
			$t_{0.05}(5-2)=$	3.1824		

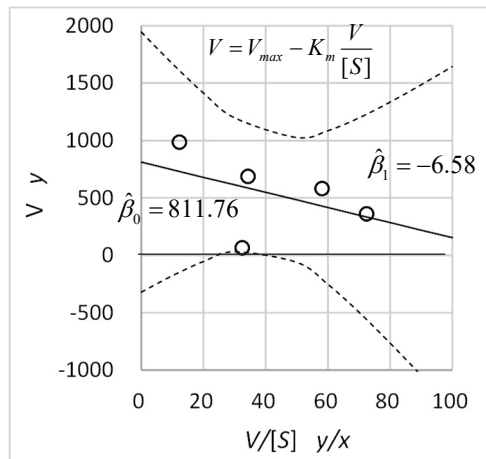


図 12.5 線形化 *Scatchard・Eadie-Hofstee* 変換

非線形回帰に対する線形化のパラメータの推定精度の比較

代表的な非線形モデルであるミカエリス・メンテン式のパラメータ推定のために各種の線形変換が考案され実用化されてきた。コンピュータがない時代には、手計算に代えて方眼紙上で散布図を作成し、目の子で回帰直線を引く、切片および傾きを読み取ることによりパラメータの推定値が得る方法は、広く活用されていたと推測する。PC が普及している現代では、方眼紙に代え、コンピュータを用いた回帰分析が手軽に行えるようになった。他方、非線形回帰によるミカエリス・メンテン式のパラメータ推定は、何らかの統計ソフトの使用が推奨されている。その結果として計算方法がブラック・ボックスのままとなっているために、いくつもの線形化の方法が、現在でも消えることなく綿々と伝承されている。

非線形回帰によるミカエリス・メンテン式のパラメータ推定は、表 12.2 に示したように Excel シート上で計算式を設定し、おおよその初期パラメータに対する推定値 \hat{V}_i を設定し、ソルバーで残差平方和 $S_e = \sum_i (V_i - \hat{V}_i)^2$ を最小にするように、初期パラメータを変化させることにより手軽に推定できることを示した。また、推定されたパラメータの 95%信頼区間を求めるためには、式 (12.6) および式 (12.7) に示した偏微分式を用い、Excel シート上で計算することも示した。

線形化によるパラメータの推定にも Excel の「ソルバー」を用いて、非線形回帰の場合と同様に、おおよその初期値を与え、残差平方和 S_e を最小にする方法を示し、パラメータの共分散行列 $\Sigma(\hat{\beta})$ から V_{max} および K_m の推定値に対する 95%信頼区間を求める方法を示した。線形化した場合の V_{max} および K_m の推定は、推定されたパラメータの逆数あるいは比により求められるが、逆数あるいは比に対する 95%信頼区間の算出のために必要な分散は、式

(12.16) および式 (12.18) に示したように推定したパラメータに関して偏微分した式に対し、パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式による計算が必要であることも示した。

表 12.11 に示すのは、これまで別々に推定結果を示してきた反応速度 V に対する最大反応 V_{max} およびミカエリス定数 K_m についてまとめたものである。線形化した場合の切片の推定値、傾きの推定値を示し、 V_{max} および K_m に戻した推定値を示す。両逆数・Lineweaver-Burk 変換を除いて、そこそこの推定値が得られているが、非線形回帰の場合の推定値に比べ乖離が大きい。ただし、いずれにしても“推定値”であり、どれが真の値に近いかは、なんとも言い難いのであるが、推定値の 95%信頼区間の大きさを比較することにより、推定法の良し悪しを評価することが可能となる。

表 12.11 非線形回帰および線形化により推定された V_{max} および K_m の比較

推定方法	切片 β_0^{\wedge} の推定		傾き β_1^{\wedge} の推定		推定値	
	意味	値	意味	値	V_{max}	K_m
(a) 非線形回帰	—	—			1130.0	11.64
(b) 両逆数・Lineweaver-Burk	$1/V_{max}$	-0.00074	K_m/V_{max}	0.0302	-1345.7	-40.65
(c) Hanes-Woolf	K_m/V_{max}	0.0157	$1/V_{max}$	0.0008	1249.4	19.62
(d) Scatchard・Eadie-Hofstee	V_{max}	811.8	$-K_m$	-6.58	811.8	6.58

非線形回帰による最大反応 V_{max} およびミカエリス定数 K_m の推定が優れていることを示すためには、線形化による V_{max} および K_m の標準誤差 SE を算出し、95%信頼区間を比較することが必須である。表 12.12 に最大反応 V_{max} について、すでに計算されている標準誤差 SE と 95%信頼区間をまとめて示す。

表 12.12 非線形回帰および線形化により推定された最大反応 V_{max} の 95%信頼区間

推定方法	推定値	標準誤差	95%信頼区間		幅
	V_{max}	SE	$L_{95\%}$	$U_{95\%}$	$t_{0.05} \times SE$
(a) 非線形回帰	1130.0	104.3	798.0	1462.0	332.0
(b) 両逆数・Lineweaver-Burk	-1345.7	2091.7	-8002.4	5311.0	6656.7
(c) Hanes-Woolf	1249.4	227.5	525.5	1973.3	723.9
(d) Scatchard・Eadie-Hofstee	811.8	355.9	-320.7	1944.3	1132.5
	$t_{0.05}(5-2)=$	3.1824			

両逆数・Lineweaver-Burk 変換では、非線形の $SE^{(a)} = 104.3$ に比べ、 $SE^{(b)} = 2091.7$ と桁違いに大きい。Hanes-Woolf 変換では、 $SE^{(c)} = 227.5$ と 2 倍強、Scatchard・Eadie-Hofstee 変換では、 $SE^{(d)} = 355.9$ と 3.5 倍弱となっている。線形変換それぞれの 95%信頼区間は、非線形モデルの推定値 $V_{max}^{(a)} = 1130.0$ を含んでいることから、統計的には、おかしい推定値ではないとも言える。ただし、精度の極めて悪い標準誤差 SE に基づく V_{max} を実験データの要約として用いることは、まったく薦められない。

表 12.13 に示すのは、ミカエリス定数 K_m について最大反応 V_{max} と同様にまとめた結果である。それぞれの標準誤差 SE を非線形回帰の $SE^{(a)} = 3.02$ と比較すると、*Hanes-Woolf* 変換では、 $SE^{(c)} = 9.44$ と 3 倍強の違い、*Scatchard・Eadie-Hofstee* 変換では、 $SE^{(d)} = 7.59$ と 2.5 倍強の違いとなっている。この程度の違いならば、良しとするのであろうか。ただし、ここに示したのは、1 つの事例に過ぎないので、他の事例についても検討する必要があるとも言えるが、非線形回帰による推定が Excel により手軽にできるようになった時代に、常にセットで扱われる 3 種の線形変換式にこだわるのは、いかがなものか。

表 12.13 非線形および線形化により推定されたミカエリス定数 K_m の 95%信頼区間

	推定方法	推定値	標準誤差	95%信頼区間		幅
		K_m	SE	$L_{95\%}$	$U_{95\%}$	$t_{0.05} \times SE$
(a)	非線形回帰	11.64	3.02	2.02	21.26	9.62
(b)	両逆数・ <i>Lineweaver-Burk</i>	-40.65	58.93	-228.21	146.90	187.55
(c)	<i>Hanes-Woolf</i>	19.62	9.44	-10.42	49.66	30.04
(d)	<i>Scatchard・Eadie-Hofstee</i>	6.58	7.59	-17.56	30.73	24.14
		$t_{0.05}(5-2) =$	3.1824			

図 12.6 に 4 種の方法で推定された最大反応 V_{max} およびミカエリス定数 K_m の 95%信頼区間を示す。表に示した結果からでは、はっきりしなかったが、(a) 非線形回帰の推定値が、(c) *Hanes-Woolf* 変換、および、(d) *Scatchard・Eadie-Hofstee* 変換の中間となっていること、95%信頼区間が、他に比べて極めて小さいことが浮き彫りにされている。

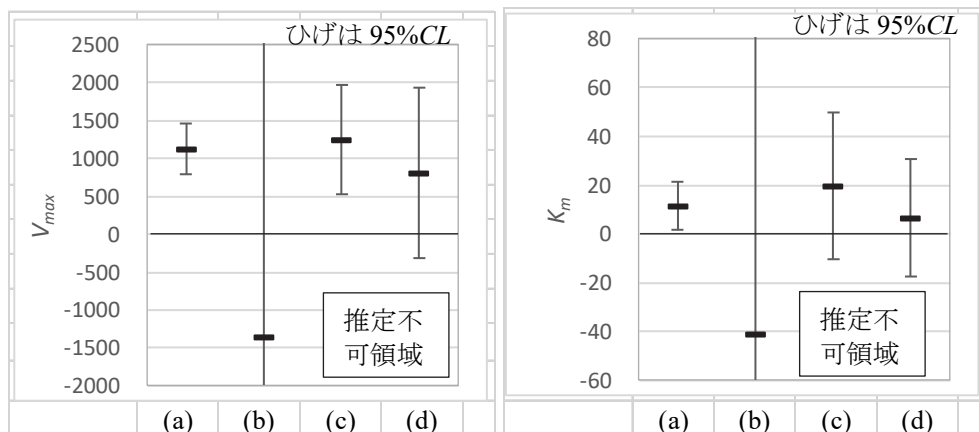


図 12.6 非線形回帰に対する線形化の V_{max} および K_m の 95%信頼区間の比較

(a) 非線形回帰, (b) 両逆数・*Lineweaver-Burk* 変換, (c) *Hanes-Woolf* 変換, (d) *Scatchard・E-H* 変換

データ数が 5 点と少ないために 95%信頼区間の計算に自由度 $df = 5 - 2 = 3$ の $t_{0.05}(5-2) = 3.1824$ を使っているとはいえ、(c) *Hanes-Woolf* 変換によるミカエリス定数 K_m が“推定不可領域”まで伸びていること、(d) *Scatchard・Eadie-Hofstee* 変換では、 V_{max} および K_m が共に“推定不可領域”まで伸びていることから、推定値そのものが不安定であることの証である。

各種の線形化による V_{max} の推定値, K_m の推定値に対する 95%信頼区間を付けて比較することにより, 線形化の方法で推定された V_{max} と K_m の 95%信頼区間が, 非線形回帰の適用に比べ, あまりにも大きいことである. 表 12.2 に示したように非線形回帰が, Excel の 1 枚のシート上で簡単に行えるようになった時代に, 推定誤差の大きい各種の線形化によるパラメータの推定に固執する理由は何であろうか.

文献の推定値との差異

Armitage ら(2002) に示されている V_{max} および K_m の推定値と, これまでに示した推定値に若干の差異が生じている. 表 12.14 に示すように, 差異が大きいのは (b) 両逆数・Lineweaver-Burk 変換の場合に非現実的な推定値ではあるが, 最大反応 V_{max} で -11.2% と小さくなっているものの, 他は 5%前後の差異におさまっている. 文献には, 標準誤差が示されていないので, Excel シート上で計算した結果を示し, 非線形モデルの線形化に対する優位性を定量化して明らかにした.

表 12.14 文献値との差異

推定方法	Armitage 文献値		推定値		差異	
	V_{max}	K_m	V_{max}	K_m	V_{max}	K_m
(a) 非線形回帰	1114	11.3	1130.0	11.64	1.4%	3.0%
(b) 両逆数・Lineweaver-Burk	-1515	-43.9	-1345.7	-40.65	-11.2%	-7.4%
(c) Hanes-Woolf	1225	18.6	1249.4	19.62	2.0%	5.5%
(d) Scatchard・Eadie-Hofstee	802	6.2	811.8	6.58	1.2%	6.2%
(e) Direct linear plot	979	7.7				
(f) 両逆数GLM/gamma errors	1395	20.2				

なお, Armitage ら(2002) には, (e) Direct linear plot, (f) 両逆数 GLM/gamma errors, による V_{max} および K_m の推定値が示されている. (e) Direct linear plot は, 知名ら (2014) では, Cressie-Keightley プロットとして示されているが, 段階的に複数の直線をあてはめるような手順であり, 標準誤差の算出が困難なので割愛した. (f) 両逆数 GLM/gamma errors は, 一般化線形モデル (GLM) で gamma errors を用いた方法であるが, 線形化の範囲を超えるので割愛した. どちらの方法も非線形回帰での推定値との乖離が大きいので, 魅力的ではないことも割愛の理由である. 知名ら (2014) には, さらに Cleland-Wilkinson 法, Tseng-Hsu 法なども紹介されているが, 手順が煩雑なので割愛した.

第 10.6 節 の末尾で Wikipedia での日本語版の「ガウス・ニュートン法」に非線形モデルによるミカエリス・メンテン式の計算事例を紹介したが, 事例の出典は示されていない. 念のために英語版の「Gauss-Newton algorithm」を確認すると日本語版と同じ事例が示され, External links として「Probability, Statistics and Estimation. The algorithm is detailed and applied to the biology experiment discussed as an example in this article (page 84 with the uncertainties on the estimated values).」, Rouaud (2017), <https://www.incertitudes.fr/book.pdf> に, Wikipedia に引用されていないミカエリス・メンテン式の信頼区間の図が示されている.

12.2. 統計ソフト SAS によるミカエリス・メンテン式のパラメータ推定

現代の算盤である Excel を用い、ミカエリス・メンテン式に派生する各種のあてはめ結果を示してきたのであるが、様々な操作ミス、計算式の設定ミスをしてでも発見できにくい本質的な脆弱性を内在している。そのために、結果の再現性に強い統計ソフトによる検証が欠かせない。そこで、無償で継続的に使える OnDemand SAS の NLIN プロシジャを用い、これまでの Excel の結果の検証を行うと共に、非線形回帰の統計ソフトの使い方を例示する。

非線形回帰のパラメータ推定値の SAS による検証

SAS は、データを入力し、SAS データセットを作成する DATA ステップ、PROC ステップ、から構成され交互に実行することができる。表 12.1 に示したデータを読み込む DATA ステップを次に示す。

SAS プログラム 12.1 データの入力

```
TITLE1 'ミカエリス.SAS 2023/02/26 <<アーミテイジ>>' ;
DATA d01 ;
  input No S V ;
datalines ;
  1 2 65
  2 5 362
  3 10 581
  4 20 686
  5 80 985
;
proc print data=d01 ; run;
```

PRINT プロシジャにより、DATA ステップで作成された「d01」SAS データセットを出力する。OnDemand SAS の標準出力は HTML 形式なので出力結果をコピーし Excel シートに貼り付け書式を整えた結果をワードに「図 (拡張メタファイル)」形式でペーストした結果を表 12.15 に示す。

表 12.15 SAS の PRINT プロシジャの出力

OBS	No	S	V
1	1	2	65
2	2	5	362
3	3	10	581
4	4	20	686
5	5	80	985

非線形回帰のための NLIN プロシジャでは、用いるデータ「d01」を指定し、parms ステートメントで初期値 $\hat{V}_{max} = 1000$ および $\hat{K}_m = 10$ を設定し、model ステートメントでミカエリス・メンテン式を設定する。

SAS プログラム 12.2 ミカエリス・メンテン式

```
proc nlin data=d01 ;
  parms      Vmax = 1000  Km = 10  ;
  model  V = Vmax * S / (S + Km)  ;
run;
```

表 12.16 に NLIN プロシジャの出力結果を示す。最初の分散分析表で「Error」行の「平方和」列の 14904.5 が、表 12.2 に示した Excel の「ソルバー」で最小化した残差平方和 S_e に一致する。パラメータの推定値 $\hat{V}_{max} = 1130$ 、 $\hat{K}_m = 11.64$ は、「ソルバー」で残差平方和 S_e を最小化するために変化させたセル (\hat{V}_{max} , \hat{K}_m) に一致する。Excel による計算結果で統計ソフトによる検証が欠かせないのが、(\hat{V}_{max} , \hat{K}_m) の標準誤差である。NLIN プロシジャの出力では、「近似標準誤差」が該当し (104.3, 3.0226) が得られる。表 12.3 に示した SE に一致し、もちろん 95%信頼区間も一致している。

表 12.16 SAS のミカエリス・メンテン式による NLIN プロシジャの出力

要因	自由度	平方和	平均平方	F 値	近似Pr > F
Model	2	1898746	949373	191.09	0.0007
Error	3	14904.5	4968.2		
Uncorrected Total	5	1913651			
パラメータ					
	推定値	近似標準誤差	近似 95% 信頼限界		
Vmax	1130	104.3	798	1462	
Km	11.64	3.0226	2.0208	21.2591	
近似相関行列					
	Vmax	Km			
Vmax	1	0.8325			
Km	0.8325	1			

パラメータの相関行列とパラメータの共分散行列の相互関係

NLIN プロシジャの標準出力に「パラメータの共分散行列 $\Sigma(\hat{\theta})$ 」がなく、「パラメータの相関行列 $R(\hat{\theta})$ 」が出力されている。これらは、推定されたパラメータの標準誤差の SE ベクトルを用いることにより、

$$\begin{aligned}
 R(\hat{\theta}) &= \Sigma(\hat{\theta}) / SE / SE^T \\
 \Sigma(\hat{\theta}) &= R(\hat{\theta}) \times SE \times SE^T
 \end{aligned}
 \tag{12.27}$$

のように相互に互換性がある. 表 12.2 のパラメータの共分散行列 $\Sigma(\hat{\theta})$ と推定されたパラメータのベクトル $SE = [104.33 \ 3.02]^T$ を用いて, Excel で

$$R(\hat{\theta}) = [\Sigma(\hat{\theta}) \text{の範囲}] / [SE \text{の範囲}] / [SE^T \text{の範囲}]$$

	$\Sigma(\hat{\theta})$			SE			$R(\hat{\theta})$	
	10885.37	262.53	÷	104.33	=	1	0.8325	
	262.53	9.1359		3.02		0.8325	1	
	÷							
SE^T	104.33	3.02						

$$\Sigma(\hat{\theta}) = [R(\hat{\theta}) \text{の範囲}] \times [SE \text{の範囲}] \times [SE^T \text{の範囲}]$$

	$R(\hat{\theta})$			SE			$\Sigma(\hat{\theta})$	
	1	0.8325	×	104.33	=	10885.37	262.53	
	0.8325	1		3.02		262.53	9.1359	
	×							
SE^T	104.33	3.02						

のように対応するセル同士の四則演算により計算できる. 表 12.16 に示した「近似相関係数」から計算したパラメータの共分散行列 $\Sigma(\hat{\theta})$ は, Excel で計算した表 12.2 の結果に一致する.

ミカエリス・メンテン曲線の 95%信頼区間

図 12.2 に示したミカエリス・メンテン曲線を SAS で再現する. そのため表 12.4 に示した推定値とその 95%信頼区間の計算を必要とする. このような推定値を統計ソフトで計算するための基本は, 元のデータファイルに, 推定したい基質濃度 $[S]$ のみを追加し, 元のデータも込みにして反応速度 V の推定値 \hat{V} , SE , $L95\%$, および, $U95\%$ を計算させる.

基質濃度 $[S]$ の横並びのリストを読み込むために, input ステートメントで, 変数 S の後に「@@」により, 1 つずつデータが尽きるまで繰り返せと指示している. その後の output ステートメントで, 読み込んだ変数 S を SAS データファイル「d02」に出力させている. データが尽きると, 次の DATA ステップで, すでにある「d01」に「d02」を行方向に併合し, 「d03」データファイルが作成され, print プロシジャで「d03」を出力する.

SAS プログラム 12.3 追加データファイルの作成

```

data d02 ;
  input S @@ ; output ;
datalines;
  0.1 1 5 10 20 40 60 100 300 ;
  ;
data d03 ;
  set d01 d02 ;

proc print data=d03 ; run;

```

表 12.17 併合した「d03」データファイル

OBS	No	S	V
1	1	2	65
2	2	5	362
3	3	10	581
4	4	20	686
5	5	80	985
6	.	0.1	.
7	.	1	.

8	.	5	.
9	.	10	.
10	.	20	.
11	.	40	.
12	.	60	.
13	.	100	.
14	.	300	.

NLIN プロシジャで、推定された内部計算で使われているパラメータの共分散行列を出力するために `outest=cov1` オプションを設定し、`list` オプションでミカエリス・メンテン式をパラメータで偏微分した式を出力するように設定している。Output ステートメントで、`y_hat`, `SE`, `L95`, `U95` 変数に推定結果を保存するように設定している。

SAS プログラム 12.4 予測値および 95%信頼区間の計算

```
proc nlin data=d03 outest=cov1 list ;
  parms Vmax=1000 Km=10 ;
  model V = Vmax * S / (S + Km) ;
  output out=out1 predicted=y_hat stdp=SE L95M=L95 U95M=U95 ;
run ;
proc print data=cov1 ; run ;
proc print data=out1 ; run ;
```

表 12.18 に示すのは、`list` オプションでミカエリス・メンテン式を `MODEL.V` としたときに、最大反応 `Vmax` での偏微分式が $@MODEL.V/@Vmax = S / (S + Km)$ 、ミカエリス定数 `Km` での偏微分式が $@MODEL.V/@Km = - MODEL.V / (S + Km)$ として内部で使用されているとの表示である。これは、パラメータに関して偏微分した式 (12.6) および式 (12.7) に対応している。

表 12.18 偏微分式の出力

Listing of Compiled Program Code		
Stmnt	Line:Col	Statement as Parsed
1	71:05:00	MODEL.V = (Vmax * S) / (S + Km);
1	71:05:00	@MODEL.V/@Vmax = S / (S + Km);
1	71:05:00	@MODEL.V/@Km = - MODEL.V / (S + Km);

$$V = \frac{V_{max}[S]}{[S] + K_m}, \quad z_1 = \frac{\partial V}{\partial V_{max}} = \frac{[S]}{[S] + K_m}, \quad z_2 = \frac{\partial V}{\partial K_m} = -\frac{V_{max}[S]}{[S] + K_m} \cdot \frac{1}{[S] + K_m} = \frac{-V_{max}[S]}{([S] + K_m)^2}$$

表 12.19 に示すのは、 (V_{max}, K_m) を変化させて残差平方和 S_e を最小化させる反復計算の過程である。FINAL 行が推定値で、その下にパラメータの共分散行列 $\Sigma(\hat{\theta})$ が出力されている。

表 12.19 反復の過程とパラメータの共分散行列 $\Sigma(\hat{\theta})$ の出力

OBS	_TYPE_	_STATUS	_NAME_	_ITER_	_SSE_	Vmax	Km
1	ITER	2 Iteration		0	27330.01	1000	10
2	ITER	2 Iteration		1	14925.99	1130.93	11.771
3	ITER	2 Iteration		2	14904.55	1129.76	11.631
4	ITER	2 Iteration		3	14904.51	1130.02	11.641
5	ITER	2 Iteration		4	14904.51	1130	11.64
6	FINAL	0		.	14904.51	1130	11.64
7	COVB	0	Vmax	.	14904.51	10885.37	262.534
8	COVB	0	Km	.	14904.51	262.53	9.136

表 12.20 左に示すのは、PRINT プロシジャで出力された予測値と 95%信頼区間を Excel に張り付けた結果である。図 12.7 Excel によるミカエリス・メンテン曲線と 95%信頼区間を散布図に重ね書きした結果である。このような統計ソフトの出力と Excel の作図機能のコラボレーションが、スムーズにできるようになったことは、嬉しいかぎりである。

表 12.20 予測値と 95%信頼区間の推定結果

OBS	No	S	V	y_hat	SE	L95	U95
1	1	2	65	165.7	25.4	84.7	246.6
2	2	5	362	339.5	39.6	213.6	465.5
3	3	10	581	522.2	42.3	387.6	656.8
4	4	20	686	714.3	38.9	590.5	838.1
5	5	80	985	986.5	66.5	774.9	1198.1
6	.	0.1	.	9.6	1.8	3.9	15.4
7	.	1	.	89.4	15.2	41.0	137.8
8	.	5	.	339.5	39.6	213.6	465.5
9	.	10	.	522.2	42.3	387.6	656.8
10	.	20	.	714.3	38.9	590.5	838.1
11	.	40	.	875.3	47.6	723.9	1026.7
12	.	60	.	946.4	58.5	760.3	1132.5
13	.	100	.	1012.2	72.3	782.2	1242.1
14	.	300	.	1087.8	91.8	795.5	1380.1

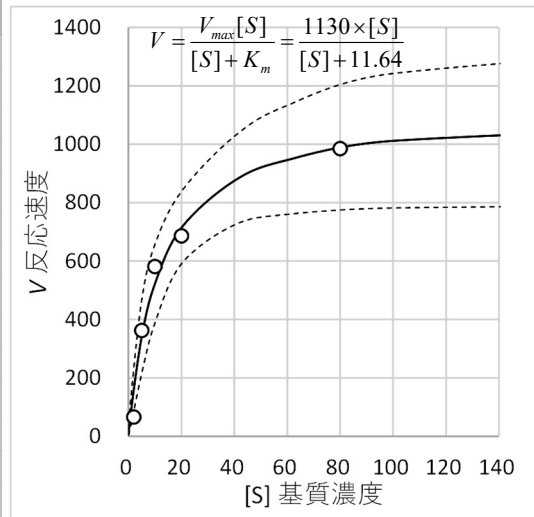


図 12.7 Excel によるミカエリス・メンテン曲線の作図例

非線形回帰による Hanes-Woolf 変換によるパラメータの直接推定

非線形モデルであるミカエリス・メンテン式を Hanes-Woolf 変換により線形化し、

$$\text{式 (12.2) : } \frac{[S]}{V} = \frac{K_m}{V_{max}} + \frac{1}{V_{max}}[S], \quad \frac{x}{y} = \frac{\mu}{\theta_{max}} + \frac{1}{\theta_{max}}x$$

回帰パラメータ ($\hat{\beta}_0, \hat{\beta}_1$) の推定を行い、最大反応 \hat{V}_{max} およびミカエリス定数 \hat{K}_m を Excel により

$$\text{式 (12.21) : } \hat{V}_{max} = \frac{1}{\hat{\beta}_1} = 1249.4, \quad \text{式 (12.22) : } \hat{K}_m = \frac{K_m / V_{max}}{1 / V_{max}} = \frac{\hat{\beta}_0}{\hat{\beta}_1} = 19.62$$

を推定した。最大反応 \hat{V}_{max} のミカエリス定数 \hat{K}_m の標準誤差を求めるために、それぞれの推定式を回帰パラメータで偏微分し、パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式で分散

を推定し、表 12.8 に標準誤差の推定結果を示した。推定に際し、Excel の行列関数を用いたのであるが、結果が正しいと言えるのだろうか。そのためのバリデーション（検証）をどのように行ったらよいのだろうか。計算事例が公表されている場合には、用いられているデータに対する計算結果を参照すればよいのだが、見いだすことができない。

そこで、*Hanes-Woolf* 変換式を回帰モデルとして扱うのではなく、最大反応 V_{max} とミカエリス定数 K_m をそのままパラメータとして用い、非線形回帰を適用することにより、推定値と標準誤差を直接推定する。SAS の DATA ステップで表 12.1 のデータを読み込みつつ、変換式 $([S]/V)$ を計算し、変数名として S_V とする。表 12.21 に示すのは、PRINT プロシジャによって、読みこんだデータが出力された結果である。

SAS プログラム 12.5 線形変換

```
TITLE1 'ミカエリス Woolf. SAS 2023/02/27';
DATA d01;
  input No S V;
  S_V = S / V;
datalines;
1 2 65
2 5 362
3 10 581
4 20 686
5 80 985
;
proc print data=d01; run;
```

表 12.21 *Hanes-Woolf* 変換による解析用データファイルの出力

OBS	No	S	V	S_V
1	1	2	65	0.0308
2	2	5	362	0.0138
3	3	10	581	0.0172
4	4	20	686	0.0292
5	5	80	985	0.0812

非線形回帰による解析では、*Hanes-Woolf* 変換で用いられている回帰パラメータ $\beta_0 = K_m / V_{max}$ 、 $\beta_1 = 1 / V_{max}$ ではなく、式 (12.2) に示されている V_{max} および K_m を model ステートメントでパラメータとして直接用いる。

SAS プログラム 12.6 *Hanes-Woolf* 変換を非線形回帰で解く

```
proc nlin data=d01;
  parms Vmax = 1000 Km = 10;
  model S_V = Km / Vmax + (1 / Vmax) * S;
run;
```

表 12.22 に示す NLIN プロシジャの結果は、表 12.7 に示した *Hanes-Woolf* 変換式によるパラメータ推定の結果と完全に一致することが確かめられる。分散分析表の「Error」の平方和と

表 12.22 SAS の NLIN プロシジャによるの *Hanes-Woolf* 変換式によるパラメータ推定

要因	自由度	平方和	平均平方	F 値	近似Pr > F
Model	1	0.00268	0.00268	30.17	0.0119
Error	3	0.000267	0.000089		
Correcte	4	0.00295			
パラメー	推定値	近似標準	近似 95% 信頼限界		
Vmax	1249.4	227.5	525.5 1973.3		
Km	19.6205	9.4405	-10.4233 49.6643		
近似相関行列					
	Vmax	Km			
Vmax	1	0.829621			
Km	0.829621	1			

平均平方は、残差平方和 $S_e=2.67E-04$ ，誤差分散の推定値 $\sigma^2=8.90E-05$ のように一致する。 \hat{V}_{max} および \hat{K}_m も

表 12.7 右下段の再掲

	推定値	SE	L95%	U95%
$V_{max}=1/\beta^{\wedge}_1=$	1249.38	227.45	525.53	1973.24
$K_m=\beta^{\wedge}_0/\beta^{\wedge}_1=$	19.62	9.44	-10.42	49.66

のように完全に一致する。

線形化のための *Hanes-Woolf* 変換式に対して非線形回帰を適用すること自体はナンセンスなのであるが、線形回帰で推定されたパラメータの逆数 ($1/\hat{\beta}_1$) あるいは比 ($\hat{\beta}_0/\hat{\beta}_1$) の標準誤差を求める方法は、「知る人ぞ知る、合成分散の一般式 (デルタ法)」方法なので、Excel による計算結果が正しいのか検証するために例示した。

誤差の伝播式を用いた傾き $\hat{\beta}_1$ の逆数の標準誤差

表 12.7 で *Hanes-Woolf* 変換式で求めた回帰係数の傾き $\hat{\beta}_1$ の逆数が、 \hat{V}_{max} となることを示した。さて、表 12.22 では、 \hat{V}_{max} の推定値は 1249.4、標準誤差は 227.5 となっている。表 12.7 および式 (12.21) から、傾き $\hat{\beta}_1$ の推定値は 0.0008004、標準誤差は 0.000146 が読み取れる。推定値の逆数は、 $\hat{V}_{max}=1/0.0008004=1249.4$ と SAS の結果に一致する。さて、 \hat{V}_{max} の標準誤差と $\hat{\beta}_1$ の関係はどのように説明したらよいのであろうか。式 (12.24) を整理すると誤差の伝播式を適用することにより、 $\hat{V}_{max}=1/\hat{\beta}_1$ の標準誤差を

$$\left. \begin{aligned} SE(1/\hat{\beta}_1) &= \sqrt{\left(\frac{\partial(1/\hat{\beta}_1)}{\partial\hat{\beta}_1}\right)^2 \cdot SE(\hat{\beta}_1)^2} = \sqrt{\left(\frac{-1}{\hat{\beta}_1^2}\right)^2 \cdot SE(\hat{\beta}_1)^2} \\ &= \sqrt{\left(\frac{-1}{0.0008004^2}\right)^2 \times 0.0001457^2} = 227.45 \end{aligned} \right\} \tag{12.28}$$

のように求めることができる。

12.3. ミカエリス・メンテン式に基づく阻害定数 K_i の推定

ミカエリス・メンテン式で示される酵素反応曲線は、基質濃度 $[S]$ の増加に伴い単調増加し、反応速度 V の最大値 V_{max} 、および、 V_{max} の 2 分の 1 となる基質濃度 $[S]$ である K_m (ミカエリス定数) によって定義されている。酵素阻害薬 $[I]$ を加えた場合に、基質 $[S]$ に対する酵素反応曲線にどのような変化が現れるか見極め定量化したい。

阻害定数 K_i の推定方法による不一致

山崎ら (2008) は、「非線形回帰法によって薬理研究の精度は向上する - 酵素阻害薬の阻害定数の推定の事例 -」で、ミカエリス・メンテン式の最大値 V_{max} に影響しないが、ミカエリス定数 K_m にのみ影響を与える拮抗阻害薬について、a) SAS の NLIN プロシジャを用いた非線形回帰による阻害定数の推定値、b) 両逆数・Lineweaver-Burk 変換を用いた 2 段階による推定値、の比較を行っている。表 12.23 は、4 種の実験系で 4~5 回の実験データについて a) および b) で推定された阻害定数 K_i の比 (b/a) を示している。同じ実験データを用いた統計量としての阻害定数 K_i が、1/4 から 3 倍弱と大きく異なり、どちらの阻害定数 K_i が正しいのか困惑している。方法 b) は、各種の阻害薬の濃度 $[I]$ ごとに異なる V_{max} を推定しているのに対し、方法 a) は、阻害薬の各濃度で共通の V_{max} を使っていることから、方法 a) が良い推定値が得られると確信している。しかし、どちらがより良い推定値なのかを判定する決め手がないので、モンテカルロ法による定量的な性能比較を行っている。その結果、方法 b) には、Michaelis-Menten モデルに整合しない推定値が得られることがあり、推定精度が一般に方法 a) に劣り、さらに b) は、極端に偏って推定された値が得られることがあることを結論付けている。

表 12.23 2 種類の推定方法で得られた阻害定数 K_i の比

実験番号	実験系 1	実験系 2	実験系 3	実験系 4
1	0.666	0.663	1.49	0.264
2	0.776	0.946	2.15	1.41
3	0.589	0.816	1.92	0.712
4	0.835	1.08	1.53	1.67
5		1.34	2.85	
平均	0.7165	0.969	1.988	1.014
総平均	1.2059			
	山崎ら(2008)の表2の「推定値の比」を引用 比 = b)両逆数 / a)非線形			

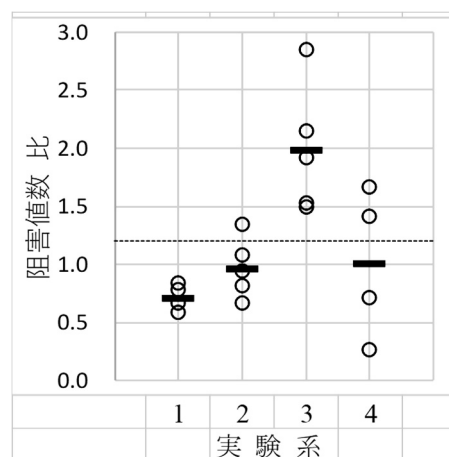


図 12.8 2 種類の推定方法で得られた阻害定数 K_i の比

ミカエリス・メンテン式によるパラメータの推定には、各種の線形化による方法よりも非線形回帰による推定が望ましいと言われているが、現実には各種の線形化によるミカエリス定数 K_m の推定が使われ続けている。非線形回帰を適用したいと思っても、何らかの統計ソフトの使用が前提とされていることが、敬遠される原因であるに違いない。第 12.1 節では、統計ソフトに頼らずとも身近な Excel で、非線形回帰によるミカエリス定数 K_m の推定および酵素反応曲線の 95%信頼区間の推定が簡単にできることを示した。

酵素阻害実験データ

表 12.24 に示すのは、阻害薬 A の濃度 $[I]_i$ を設定し、基質濃度 $[S]_j$ を変化させた場合の反応速度 V_{ijk} のデータである。実験計画的には、繰り返しが 3 の量×量の 2 因子実験データとなっているが、一般的な分散分析の適用ではなく、ミカエリス・メンテン式の適用が求められている。伝統的には、解析の第 1 段階とし、阻害薬 A の濃度 $[I]_i$ ごとにミカエリス・メンテン式の各種の線形化式により、最大値 $(V_{max})_i$ 、および、最大値の 2 分の 1 となる基質の濃度であるミカエリス定数 $(K_m)_i$ を推定する。第 2 段階として、阻害薬 A がいない場合 $[I]_0$ のミカエリス定数 $(K_m)_0$ の 2 倍となる阻害薬 A の濃度 $[I]$ を阻害定数 (K_i) として推定したい。

表 12.24 阻害薬 A の濃度別の反応速度 V の変化

阻害薬 A		基質 $[S]$ mM		
i	$[I]$ nM	0.6	0.8	1.2
0	0	19.488	22.922	27.364
	0	17.910	21.071	26.094
	0	23.168	29.148	32.158
1	8	15.039	17.896	22.449
	8	13.204	15.997	20.957
	8	17.061	22.603	25.253
2	14	12.338	15.076	19.199
	14	10.568	13.065	17.875
	14	13.364	18.218	21.287
3	20	10.179	12.718	16.559
	20	8.640	10.653	15.030
	20	10.841	14.871	18.145
4	28	7.635	9.979	13.461
	28	6.415	8.070	12.232
	28	8.044	11.364	14.020

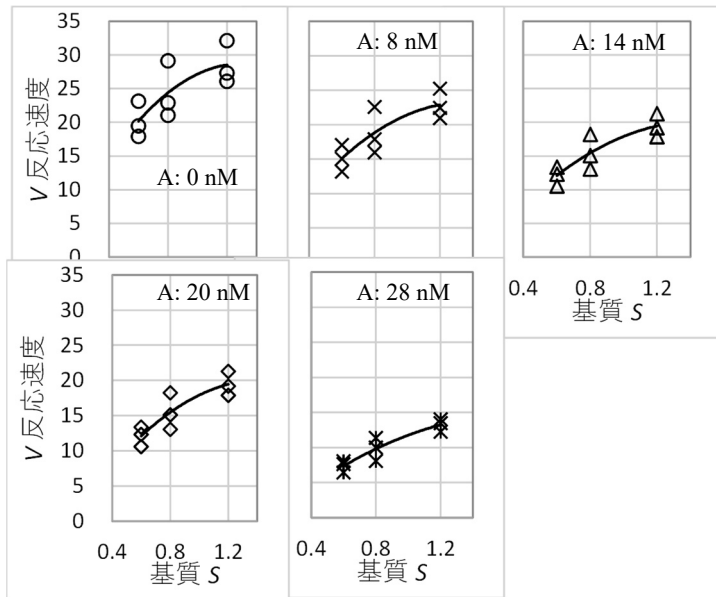


図 12.9 阻害薬 A の濃度別の反応速度 V の変化

阻害薬の酵素反応に対する作用は、(1) V_{max} は変化しないが K_m を増加させる、(2) K_m は変化しないが V_{max} を減少させる、(3) V_{max} と K_m を同じ割合で減少させる、(4) K_m の増加と共に V_{max} が減少する、の 4 種類の阻害様式に分類できることが知られている。図 12.9 の阻害薬 A の濃度 $[I]$ を増加させた場合のデータのプロットからは、 V_{max} と K_m がどのよ

うに変化するか、はっきりしないので、それぞれの濃度別に、 $(Vmax)_i$ と $(Km)_i$ を推定することにより、(1) から (4) で示した阻害様式となるかの見当が付けられる。

阻害様式が、(1) の拮抗阻害であると確認されれば、表 12.25 に示すように、阻害薬 A の濃度 $[I]_i$ 別の $(Vmax)_i$ ではなく、共通の $(Vmax)^{共通}$ とした場合の $(Km)'_i$ の推定を行なうことにより推定精度の向上が期待できる。また、阻害薬 A の濃度に比例して $(Km)'_i$ の増加が確認できれば、阻害薬 A が不在の場合 $[I]_0$ の $(Km)'_0$ を基準とし、阻害定数 $(Ki)^{共通}$ を用いて阻害薬 A の濃度 $[I]_i$ に比例する式に置き換えることができる。

表 12.25 拮抗阻害薬 A の添加によるミカエリス・メンテン式によるパラメータ推定

No	阻害薬 A (nM)	基質 [S] mM		基質 [S] mM		基質 [S] mM	
		個別に推定		共通 $Vmax$		共通 $Vmax$, (Km) が濃度 $[I]$ に比例	
i	濃度 $[I]$	$(Vmax)_i$	$(Km)_i$	$(Vmax)^{共通}$	$(Km)'_i$	$(Vmax)^{共通}$	$(Km)'_i$
0	0	$(Vmax)_0$	$(Km)_0$	$(Vmax)^{共通}$	$(Km)'_0$	$(Vmax)^{共通}$	$(Km)'_0$
1	8	$(Vmax)_1$	$(Km)_1$	$(Vmax)^{共通}$	$(Km)'_1$	$(Vmax)^{共通}$	$(Km)'_0 + (Km)'_0 \frac{[I]_1}{(Ki)^{共通}}$
2	14	$(Vmax)_2$	$(Km)_2$	$(Vmax)^{共通}$	$(Km)'_2$	$(Vmax)^{共通}$	$(Km)'_0 + (Km)'_0 \frac{[I]_2}{(Ki)^{共通}}$
3	20	$(Vmax)_3$	$(Km)_3$	$(Vmax)^{共通}$	$(Km)'_3$	$(Vmax)^{共通}$	$(Km)'_0 + (Km)'_0 \frac{[I]_3}{(Ki)^{共通}}$
4	28	$(Vmax)_4$	$(Km)_4$	$(Vmax)^{共通}$	$(Km)'_4$	$(Vmax)^{共通}$	$(Km)'_0 + (Km)'_0 \frac{[I]_4}{(Ki)^{共通}}$

阻害定数 Ki の推定方法

一般的な実験データの解析においては、制御変数 X に対する反応 Y の変化の推定が主体となるが、酵素阻害の実験では、ある一定の反応 Y_0 となる濃度 X_0 を逆推定することが目的となる。ミカエリス定数 $(Km)_i$ の推定も制御変数である基質濃度 $[S]$ に対する逆推定値であり、まず、阻害薬 A が不在 $[I]_0=0$ 場合のミカエリス定数 $(Km)_0$ を推定する。次に、ある阻害薬 A の濃度 $[I]_i$ に対するミカエリス定数 $(Km)_i$ を推定する。これらの関係から、 $(Km)_0$ の 2 倍となる阻害薬 A の濃度 $[I]_i$ における阻害定数 $(Ki)_i$ を逆推定する。

阻害定数 $(Ki)_i$ の推定方法を手順化するために、簡単な数値を与えて例示する。図 12.10 に示すように阻害薬 A の濃度が $[I]_0 = 0$ nM の場合にミカエリス定数として $(\widehat{Km})_0 = 0.50$ mM、阻害薬 A の濃度が $[I]_1 = 10$ nM の場合にミカエリス定数として $(\widehat{Km})_1 = 0.80$ mM が得られたとする。ミカエリス定数 $(\widehat{Km})_0$ を 2 倍にする阻害薬 A の濃度 $[I]_1$ に対応する阻害定数 $(Ki)_1$ を推定したい。 $(\widehat{Km})_0$ と $(\widehat{Km})_1$ の 2 点を結ぶ回帰直線の式は、切片を $(Km)_0$ 、傾きを $\{(Km)_1 - (Km)_0\} / [I]_1$ とする回帰式

$$(Km) = (Km)_0 + \frac{(Km)_1 - (Km)_0}{[I]_1} \cdot [I] \tag{12.29}$$

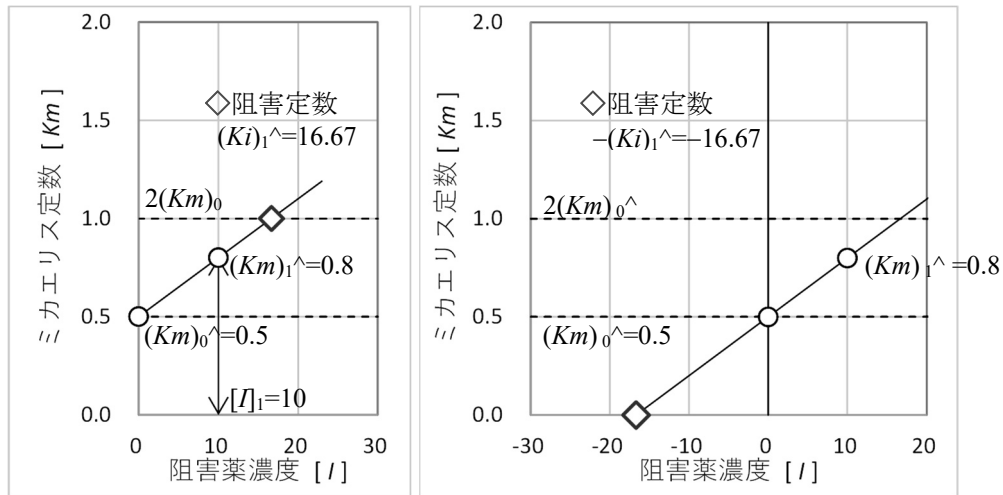


図 12.10 逆推定による阻害定数 K_i の推定および X 切片による負の阻害定数 K_i の推定

となる。この式で、ミカエリス定数 $(Km)_1$ が $(Km)_0$ の 2 倍となる酵素阻害薬の濃度 $[I]$ を逆推定したい。そこで、 $(Km)_1$ に $2 \times (Km)_0$ を代入して $[I]$ について解き、 $[I]$ を $(Ki)_1$ とおくと

$$\left. \begin{aligned}
 2 \times (Km)_0 &= (Km)_0 + \frac{(Km)_1 - (Km)_0}{[I]_1} \cdot [I] \\
 (\widehat{Ki})_1 &= \frac{(\widehat{Km})_0 \cdot [I]_1}{(\widehat{Km})_1 - (\widehat{Km})_0} \\
 &= \frac{0.5 \times 10.0}{0.8 - 0.5} = 16.67 \text{ nM}
 \end{aligned} \right\} \quad (12.30)$$

阻害定数 $(\widehat{Ki})_1 = 16.67 \text{ nM}$ が得られる。阻害定数 $(Ki)_1$ をグラフ上の目盛りを読んで推定したい場合には、図 12.10 右に示すように X 軸の切片が $-(Ki)_1$ となる。式 (12.29) で $(Km) = 0$ を代入し、式 (12.30) を適用すれば、

$$[I] = -\frac{(Km)_0 \cdot [I]_1}{(Km)_1 - (Km)_0} = -(Ki)_1 \quad (12.31)$$

負の $(\widehat{Ki})_1$ が推定される。この方法は、厄介な計算をせずとも方眼紙の目盛りから阻害定数 $(\widehat{Ki})_1$ 読み取ることができる利便性がある。

式 (12.29) でミカエリス定数 $(Km)_i$ が $2 \times (Km)_0$ となる阻害薬の濃度を $[I] = (Ki)_i$ と定義し、阻害薬の濃度を $[I]_i$ におけるミカエリス定数を $(Km)_i$ とする。この式を $(Km)_i$ で解くと

$$\left. \begin{aligned}
 2 \times (Km)_0 &= (Km)_0 + \frac{(Km)_i - (Km)_0}{[I]_i} \cdot (Ki)_i \\
 (Km)_i &= (Km)_0 + \frac{(Km)_0 \cdot [I]_i}{(Ki)_i} \\
 (Km)_i &= (Km)_0 \left(1 + \frac{[I]_i}{(Ki)_i} \right)
 \end{aligned} \right\} \quad (12.32)$$

が得られる。この式は、ミカエリス定数 $(Km)_i$ を阻害薬の濃度 $[I]_i$ および阻害定数 $(Ki)_i$ を用いた式として定義できる。

阻害薬の濃度 $[I]_i$ ごとにミカエリス・メンテン式をあてはめた場合には、推定されるパラメータ数は $(2 \times 5 = 10)$ であり、 V_{max} を共通にした場合には $(1 + 5 = 6)$ となる。阻害薬 A の濃度 $[I]_i$ に対し共通の阻害定数 $(K_i)^{共通}$ をパラメータとして用いた場合には、3 パラメータ $\{(V_{max})^{共通}, (K_m)_0^{共通}, (K_i)^{共通}\}$ により、表 12.24 に示した阻害薬 A に関する実験データを要約することができる。第 12.5 節を参照のこと。

両逆数・Lineweaver-Burk 変換によるミカエリス定数 $(K_m)_i$ の推定

表 12.24 に示したデータについて、第 1 ステップとして阻害薬濃度 $[I]_i$ ごとに両逆数変換による $(V_{max})_i$ および $(K_m)_i$ の推定を行い、第 2 ステップとして阻害定数 $(K_i)^{回帰}$ の推定を行なう。両逆数変換による推定は、表 12.5 に示したように、基質濃度の逆数 $1/[S]$ 、反応速度の逆数 $1/V$ を Excel の計算シート上で計算し、推定された回帰パラメータ $\hat{\beta}_0$ と $\hat{\beta}_1$ から、 $(\widehat{Vmax})_i = 1/\hat{\beta}_0$ 、 $(\widehat{Km})_i = \hat{\beta}_1/\hat{\beta}_0$ を推定する。

表 12.26 に阻害薬の添加がない $[I]_0 = 0$ 場合についての両逆数変換による回帰パラメータ推定のための計算シートを示す。基質濃度 $[S]$ の逆数、反応速度 V の逆数を計算し、回帰パラメータの初期値 $\hat{\beta}_0 = 0.01$ および $\hat{\beta}_1 = 0.01$ を用いて推定値 $(1/y) = \hat{\beta}_0 + \hat{\beta}_1(1/x_i)$ を計算する。さらに、残差 $\hat{\varepsilon}_i = (1/y_i) - (1/y_i)$ から、残差平方和 S_e を計算する。Excel のソルバーを用いて、 S_e が最小になるように $(\hat{\beta}_0, \hat{\beta}_1)$ を変化させた結果が示されている。

$$\begin{aligned} (1/y) &= \hat{\beta}_0 + \hat{\beta}_1(1/x_i) \\ &= 0.0202 + 0.0177 \cdot (1/x_i) \end{aligned} \tag{12.33}$$

推定されたパラメータから、

$$\left. \begin{aligned} (\widehat{Vmax})_0 &= 1/\hat{\beta}_0 \\ &= 1/0.0202 = 49.4209 \\ SE[(\widehat{Vmax})_0] &= \sqrt{d_{(Vmax)_0} \Sigma(\hat{\beta}) d_{(Vmax)_0}^T} = 16.7360 \end{aligned} \right\} \tag{12.34}$$

表 12.26 阻害薬 A なし $[I]_0 = 0$ の場合のミカエリス定数の推定

初期値 $\beta^{\wedge}_0 =$				$\beta^{\wedge}_0 =$	0.0202	$\sigma^2 =$	2.91E-05	$\Sigma(\beta^{\wedge}) = (X^T X)^{-1} \sigma^2 \wedge$		分散	
$\beta^{\wedge}_1 =$				$\beta^{\wedge}_1 =$	0.0177	$S_e =$	0.0002	$\beta^{\wedge}_0 =$	4.70E-05	-3.50E-05	4.70E-05
A薬	基質	反応		$1/[S]$	$1/V$	推定値	残差	$\beta^{\wedge}_1 =$	-3.50E-05	2.80E-05	2.80E-05
$[I]$	$[S]$	V	x_0	$1/x$	$1/y$	$(1/y)^{\wedge}$	ε^{\wedge}				
0	0.6	19.488	1	1.6667	0.0513	0.0498	0.0015			推定値	SE
0	0.6	17.910	1	1.6667	0.0558	0.0498	0.0060	$\beta^{\wedge}_0 = 1/V_{max} =$		0.0202	0.0069
0	0.6	23.168	1	1.6667	0.0432	0.0498	-0.0066	$\beta^{\wedge}_1 = Km/V_{max} =$		0.0177	0.0053
0	0.8	22.922	1	1.2500	0.0436	0.0424	0.0012			$d_1: \partial \beta^{\wedge}_0$	$d_2: \partial \beta^{\wedge}_1$
0	0.8	21.071	1	1.2500	0.0475	0.0424	0.0051	$V_{max} = 1/\beta^{\wedge}_0 =$		-2442.51	0
0	0.8	29.148	1	1.2500	0.0343	0.0424	-0.0081	$Km = \beta^{\wedge}_1/\beta^{\wedge}_0 =$		-43.3273	49.4218
0	1.2	27.364	1	0.8333	0.0365	0.0350	0.0015			推定値	SE
0	1.2	26.094	1	0.8333	0.0383	0.0350	0.0033	$V_{max} = 1/\beta^{\wedge}_0 =$		49.4218	16.7366
0	1.2	32.158	1	0.8333	0.0311	0.0350	-0.0039	$Km = \beta^{\wedge}_1/\beta^{\wedge}_0 =$		0.8767	0.5534
										$SE = \sqrt{(d \Sigma(\beta^{\wedge}) d^T)}$	

$$\left. \begin{aligned} (\widehat{Km})_0 &= \hat{\beta}_1 / \hat{\beta}_0 \\ &= 0.0177 / 0.0202 = 0.8767 \\ SE [(\widehat{Km})_0] &= \sqrt{\mathbf{d}_{(Km)_0} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{d}_{(Km)_0}^T} = 0.5534 \end{aligned} \right\} \quad (12.35)$$

が計算されている。標準誤差 SE の計算方法については、ここでは割愛するので表 12.5 に対する式 (12.14) から式 (12.19) の解説を参照のこと。

表 12.27 に示すのは、表 12.26 で計算された回帰係数 $\hat{\beta}_0$ および $\hat{\beta}_1$ による回帰直線の 95%信頼区間を計算した結果である。図 12.11 に示すのは、両逆数・Lineweaver-Burk 変換での回帰直線と 95%信頼区間を重ね書きし、Y 切片が $1/(\widehat{Vmax})_0 = 0.0202$ 、X 切片が $-1/(\widehat{Km})_0 = -1.1406$ となることを示している。

表 12.27 阻害薬 A なし $[I]_0=0$ の 95%信頼区間

x_0	X [1/S]	Y [1/V]^	分散	$t_{0.05}SE$	L 95%	U 95%
1	-2.0	-0.0152	2.99E-04	0.0409	-0.0561	0.0256
1	-1.0	0.0025	1.45E-04	0.0285	-0.0260	0.0310
1	0.0	0.0202	4.70E-05	0.0162	0.0040	0.0364
1	0.5	0.0291	1.90E-05	0.0103	0.0188	0.0394
1	1.0	0.0380	4.99E-06	0.0053	0.0327	0.0433
1	1.5	0.0468	4.99E-06	0.0053	0.0416	0.0521
1	2.0	0.0557	1.90E-05	0.0103	0.0454	0.0660
1	2.5	0.0646	4.70E-05	0.0162	0.0484	0.0808
1	3.0	0.0735	8.89E-05	0.0223	0.0512	0.0957
			$t_{0.05}(9-2)=$	2.3646		

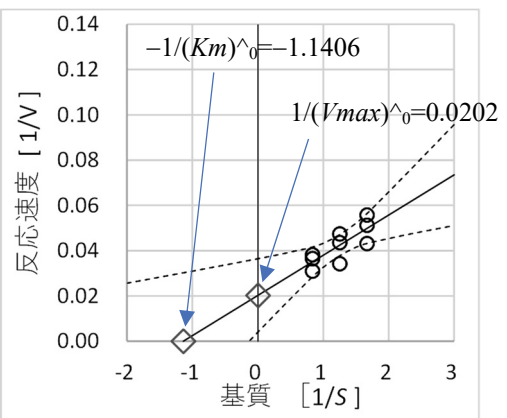


図 12.11 両逆数変換による $[I]_0=0$ の場合の 95%信頼区間

両逆数・Lineweaver-Burk 変換によるミカエリス定数 $(Km)_0$ の推定を方眼紙上で推定するような場合に、Y 軸の 0 となる点 (X 切片) を読んで、その点が $-1/(Km)_0$ となることは、式 (12.2) で $(1/V)=0$ と置いて、 $(1/S)$ について解くと

$$\left. \begin{aligned} \frac{1}{V} &= \frac{1}{V_{max}} + \frac{K_m}{V_{max}} \frac{1}{[S]} \\ \frac{1}{[S]} &= -\frac{1}{V_{max}} \cdot \frac{V_{max}}{K_m} = -\frac{1}{K_m} \end{aligned} \right\} \quad (12.36)$$

となることから簡単に確認できる。

表 12.26 は、Excel で作成してあるので、阻害薬 A の他の濃度の基質 $[S]$ および反応速度 V に入れ替え、ソルバーで残差平方和を S_e を最小化させると、全ての計算が自動的に行なわれる。表 12.28 には、阻害薬 A の濃度 $[I]_2=14$ の場合についての計算結果が示されている。

表 12.28 阻害薬 A の $[I]_2=14$ の場合のミカエリス定数の推定

初期値 $\beta^{\wedge}_0=$		$\beta^{\wedge}_0=$		$\sigma^{\wedge 2}=$		$\Sigma(\beta^{\wedge})=(X^T X)^{-1} \sigma^{\wedge 2}$		分散	
$\beta^{\wedge}_1=$		$\beta^{\wedge}_1=$		$S_e=$		$\beta^{\wedge}_0=$		$\beta^{\wedge}_1=$	
A薬	基質	反応		$1/[S]$	$1/V$	推定値	残差		
$[I]$	$[S]$	V	x_0	$1/x$	$1/y$	$(1/y)^{\wedge}$	ε^{\wedge}	推定値	SE
14	0.6	12.338	1	1.6667	0.0811	0.0829	-0.0019		
14	0.6	10.568	1	1.6667	0.0946	0.0829	0.0117	$\beta^{\wedge}_0=1/Vmax=$	0.0193
14	0.6	13.364	1	1.6667	0.0748	0.0829	-0.0081	$\beta^{\wedge}_1=Km/Vmax=$	0.0382
14	0.8	15.076	1	1.2500	0.0663	0.0670	-0.0007	$d_1: \partial \beta^{\wedge}_0$	$d_2: \partial \beta^{\wedge}_1$
14	0.8	13.065	1	1.2500	0.0765	0.0670	0.0095	$Vmax=1/\beta^{\wedge}_0=$	-2689.57
14	0.8	18.218	1	1.2500	0.0549	0.0670	-0.0121	$Km=\beta^{\wedge}_1/\beta^{\wedge}_0=$	-102.737
14	1.2	19.199	1	0.8333	0.0521	0.0511	0.0010	推定値	SE
14	1.2	17.875	1	0.8333	0.0559	0.0511	0.0048	$Vmax=1/\beta^{\wedge}_0=$	51.8611
14	1.2	21.287	1	0.8333	0.0470	0.0511	-0.0041	$Km=\beta^{\wedge}_1/\beta^{\wedge}_0=$	1.9810

推定結果は、 $(\widehat{Vmax})_2 = 1/0.0193 = 51.8611$ 、 $(\widehat{Km})_2 = 0.0382/0.0193 = 1.9810$ であり、さらに、標準誤差 $SE[(\widehat{Vmax})_2] = 28.4336$ 、 $SE[(\widehat{Km})_2] = 1.4986$ も計算されている。表 12.29 に示すのは、表 12.28 で計算された回帰係数 $\hat{\beta}_0$ および $\hat{\beta}_1$ による回帰直線の 95%信頼区間を計算した結果である。図 12.12 に示すのは、両逆数・Lineweaver-Burk 変換での回帰直線と 95%信頼区間を重ね書きし、Y切片が $1/(Vmax)_2$ 、X切片が $-1/(Km)_2$ となることを示している。

表 12.29 阻害薬 A $[I]_2=14$ nM の 95%信頼区間

x_0	X	Y	分散	$t_{0.05}SE$	L95%	U95%
1	-2.0	-0.0571	7.11E-04	0.0631	-0.1202	0.0059
1	-1.0	-0.0189	3.45E-04	0.0439	-0.0628	0.0250
1	0.0	0.0193	1.12E-04	0.0250	-0.0057	0.0443
1	0.5	0.0384	4.52E-05	0.0159	0.0225	0.0543
1	1.0	0.0575	1.19E-05	0.0081	0.0493	0.0656
1	1.5	0.0766	1.19E-05	0.0081	0.0684	0.0847
1	2.0	0.0957	4.52E-05	0.0159	0.0798	0.1116
1	2.5	0.1148	1.12E-04	0.0250	0.0898	0.1398
1	3.0	0.1339	2.12E-04	0.0344	0.0995	0.1683
			$t_{0.05}(9-2)=$	2.3646		

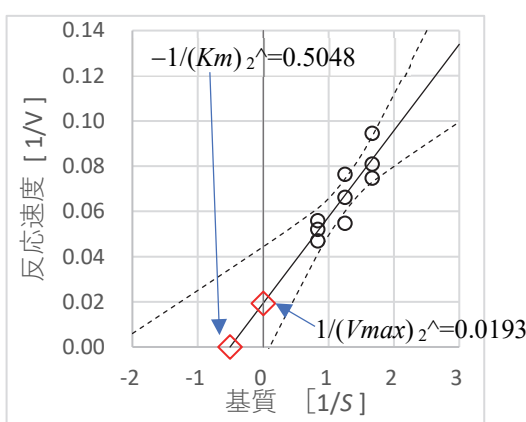


図 12.12 両逆数変換による $[I]_2=14$ の場合の 95%信頼区間

同様に、他の阻害薬の濃度に対しても計算を繰り返し、 $(Vmax)_i$ と $(Km)_i$ を Excel シート上の別の表にコピー&ペーストし、その結果を表 12.30 に示す。推定された $(Km)_i$ それぞれに対し $(Km)_0$ の 2 倍となる阻害定数 $(Ki)_i$ を式 (12.30) を用いて推定することができる。

表 12.30 に示すように反応速度 V の最大値 $Vmax$ は、阻害薬 A の濃度が 20 nM までは、おおよそ 50 で安定した推定値になっているが、28 nM の場合は、50 を大幅に超える 72.8 となっている。ただし、その SE は 91.4 であり、推定精度は極めて悪いが、全体的には、最大値

表 12.30 阻害薬 A の濃度別の反応速度 $(Vmax)_i$ およびミカエリス定数 $(Km)_i$ の推定結果

No. i	A薬 [I]	最大反応速度		ミカエリス定数		阻害定数 (K_i)
		$(Vmax)$	SE [$(Vmax)$]	(Km)	SE [(Km)]	
0	0	49.4209	16.7360	0.8767	0.5534	—
1	8	48.6683	21.3948	1.3429	0.9221	15.04
2	14	51.8610	28.4336	1.9810	1.4986	11.11
3	20	53.2052	36.5593	2.6430	2.3321	9.93
4	28	72.8533	91.4534	5.3564	7.6631	5.48
	平均	55.2017				10.39
	SD	10.0353				3.94
$(K_i)_i = (Km)_i * [I]_i / [(Km)_i - (Km)_0]$						

$Vmax$ は一定と仮定できそうである。ミカエリス定数 $(\widehat{Km})_i$ は、阻害薬の濃度におおよそ比例しているとみなせるので、拮抗阻害と判断することができる。ただし阻害薬の濃度ごとの推定された阻害定数 $(\widehat{K_i})_i$ は、一定せず“定数”として受け入れ難い結果である。

阻害定数 K_i の揺らぎの実際

図 12.13 に示すのは、阻害薬の3段階の濃度 $\{ [I]_0=0, [I]_2=14, [I]_4=28 \}$ を選択し、推定された $(Vmax)_i$ と $(Km)_i$ を用い、基質 $[S]$ の濃度を与えて反応速度 V を推定した結果を対数目盛によるミカエリス・メンテン曲線 (2パラメータ・ロジスティック曲線) である。

表 12.31 阻害薬 A の濃度 $[I]$ 別の反応速度 V の推定値

基質 [S]	反応 V		
	$[I]_0=0$	$[I]_2=14$	$[I]_4=28$
0.01	0.56	0.26	0.14
0.03	1.64	0.77	0.41
0.1	5.06	2.49	1.34
0.3	12.60	6.82	3.86
1	26.33	17.40	11.46
3	38.25	31.24	26.15
10	45.44	43.29	47.44
30	48.02	48.65	61.82
100	48.99	50.85	69.15
1000	49.38	51.76	72.47
10000	49.42	51.85	72.81

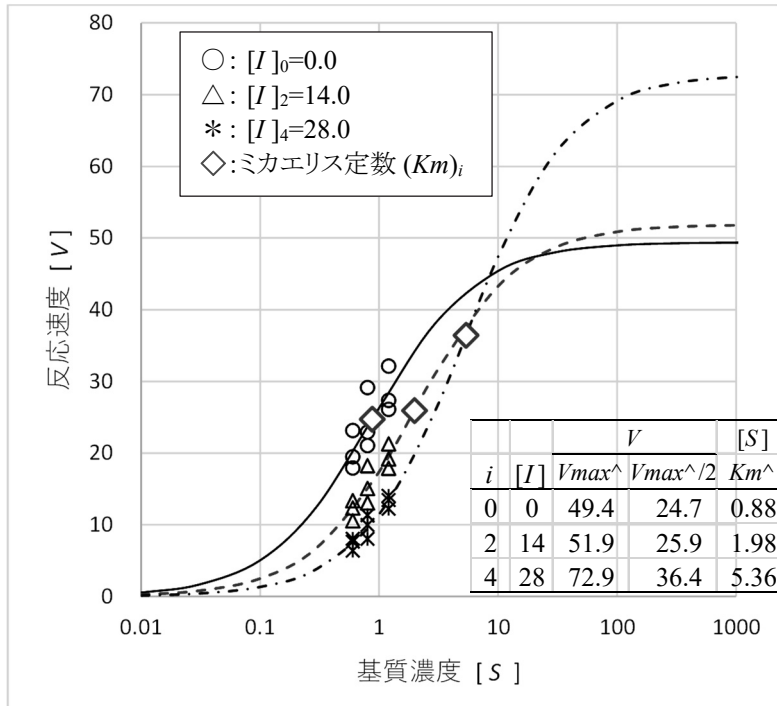


図 12.13 阻害薬 A の濃度 $[I]$ 別の対数目盛によるミカエリス・メンテン曲線

表 12.30 で計算された阻害薬 A の濃度 $[I]_i$ ごとの阻害定数 $(\widehat{K}i)_i$ は、図 12.14 左に示すように、推定された $(\widehat{K}m)_0 = 0.8767$ とそれぞれの $(\widehat{K}m)_i$ を結ぶ直線を引き、 $(\widehat{K}m)_0$ の 2 倍となる参照線を横切る点が阻害定数 $(\widehat{K}i)_i$ として推定される。図 12.14 右には、Excel の近似直線で推定された回帰直線 $(\widehat{K}m) = 0.3123 + 0.1520 \cdot [I]$ が示されており、切片の 2 倍となる参照線を回帰直線が横切る点を阻害定数 $(\widehat{K}i)^{\text{回帰}}$ とすると、

$$\left. \begin{aligned} 2 \times 0.3123 &= 0.3123 + 0.1520 \cdot (\widehat{K}i)^{\text{回帰}} \\ (\widehat{K}i)^{\text{回帰}} &= \frac{0.3123}{0.1520} = 2.0546 \end{aligned} \right\} \quad (12.37)$$

それぞれの阻害薬の濃度 $[I]_i$ ごとの阻害定数 $(\widehat{K}i)_i$ の最小値を下回ってしまう。図 12.14 左から、 $(\widehat{K}m)_0 = 0.8767$ を通る阻害薬の濃度 $[I]_i$ ごとの直線をあてはめ、X 切片と交差する点を推定し、表 12.30 に示したように、それらの平均の阻害定数を $(\widehat{K}i)^{\text{平均}} = 10.39$ と推定値する方が、妥当と思われる。

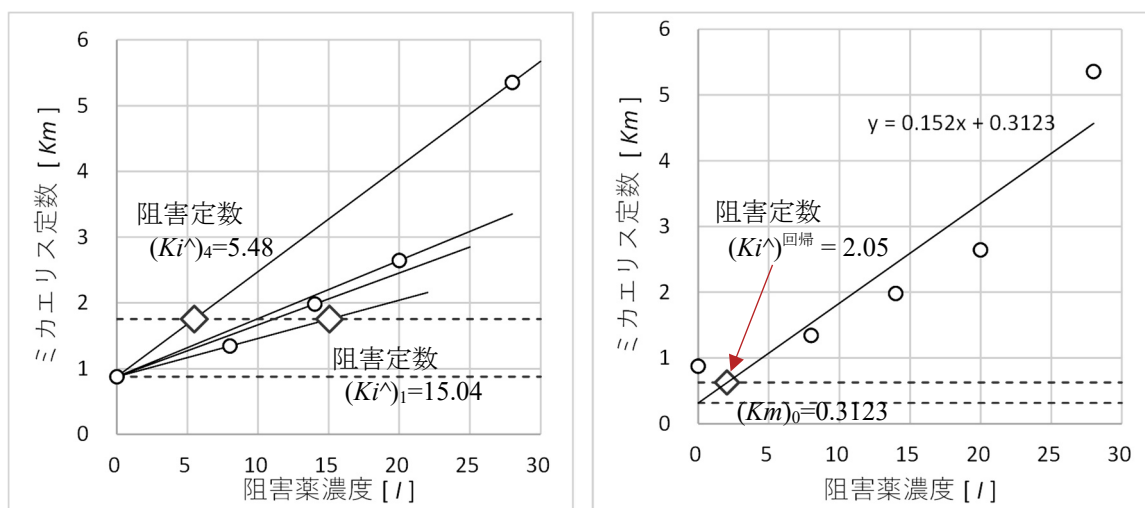


図 12.14 推定されたミカエリス定数 $(\widehat{K}m)_i$ を用いた阻害定数 $(\widehat{K}i)$ の推定

傾き $(K_m)_i / (V_{max})_i$ を用いた阻害定数 K_i の推定

両逆数・Lineweaver-Burk 変換による $(V_{max})_i$ と $(K_m)_i$ の推定に引き続き、阻害定数 (K_i) を推定する場合に、図 12.14 右に示した Y 軸にミカエリス定数 $(K_m)_i$ を用いるのではなく、推定された回帰直線の傾き $(\beta_1)_i = (K_m)_i / (V_{max})_i$ に対して回帰直線をあてはめて阻害定数 (K_i) を推定する方法が定式化されている。

図 12.13 に示すように、阻害薬の濃度 $[I]_i$ ごとの $(V_{max})_i$ と $(K_m)_i$ を推定した場合に、 $(V_{max})_i$ が大きめに推定されると $(V_{max})_i / 2$ となる基質濃度である $(K_m)_i$ も大きめに推定され、逆に $(V_{max})_i$ が小さめに推定されると $(K_m)_i$ も小さめに推定される。そのために $(K_m)_i$ の推定値が不安定になる。 $(K_m)_i$ を $(V_{max})_i$ で割ることにより $(V_{max})_i$ の影響を補正するこ

とになる。幸い、式 (12.2) に示したように両逆数・Lineweaver-Burk 変換の回帰式の傾きが $(\beta_1)_i = (Km)_i / (Vmax)_i$ であり、 $(\beta_1)_i$ を反応とみなすことにより、阻害薬の濃度 $[I]_i$ に関する回帰直線から、図 12.14 右に示したと同様の方法で、阻害定数 $(Ki)^{回帰}$ の推定が可能となる。

表 12.32 に示すのは、表 12.26 および表 12.28 に示した阻害薬濃度 $[I]_0 = 0$ および $[I]_2 = 14$ に示した反応速度 V に対する基質濃度 $[S]$ による回帰直線の切片 $(\hat{\beta}_0)_i = 1/(\widehat{Vmax})_i$ と傾き $(\hat{\beta}_1)_i = (\widehat{Km})_i / (\widehat{Vmax})_i$ に加え、他の阻害薬濃度についての計算結果をまとめ直した結果である。さらに、表 12.30 に示した、阻害薬濃度ごとの阻害定数 $(\widehat{Ki})_i$ も求めている。

表 12.32 両逆数・Lineweaver-Burk 変換の傾きによる阻害薬濃度別の阻害定数 $(Ki)_i$ の推定

No. <i>i</i>	A薬 [I]	切片 $(1/Vmax)_i$		傾き $(Km)_i / (Vmax)_i$				阻害定数 $(Ki)_i$
		$\hat{\beta}_0$	$SE(\hat{\beta}_0)$	$\hat{\beta}_1$	$SE(\hat{\beta}_1)$	L95%	U95%	
0	0	0.0202	0.0069	0.0177	0.0053	0.0052	0.0302	—
1	8	0.0205	0.0090	0.0276	0.0070	0.0111	0.0441	14.40
2	14	0.0193	0.0106	0.0382	0.0082	0.0189	0.0575	12.14
3	20	0.0188	0.0129	0.0497	0.0100	0.0261	0.0732	11.11
4	28	0.0137	0.0172	0.0735	0.0133	0.0421	0.1050	8.90
	平均	0.0185		$t_{0.05(9-2)} =$	2.3646			11.64
	SD	0.0028						2.28

$(\widehat{Ki})_i = \{(\hat{\beta}_1)_0 / [(\hat{\beta}_1)_i - (\hat{\beta}_1)_0]\} * [I]_i$

図 12.15 に示すのは、表 12.32 の傾き $(\hat{\beta}_1)_i = (\widehat{Km})_i / (\widehat{Vmax})_i$ についての阻害薬濃度 $[I]$ に関する回帰直線である。散布図上のそれぞれの点 $(\hat{\beta}_1)_i$ には、標準誤差が付随しているので、計算された 95%信頼区間を重ね書きしている。推定された回帰直線は、

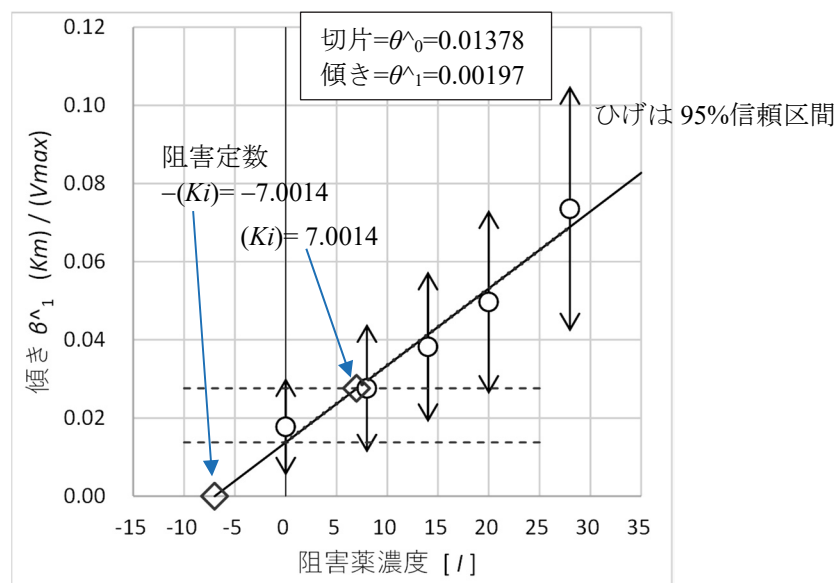


図 12.15 両逆数・Lineweaver-Burk 変換の回帰式の傾きに対する阻害定数 (Ki) の推定

$$\left[\frac{(\widehat{Km})_i}{(\widehat{Vmax})_i} \right] = 0.01378 + 0.00197 \cdot [I] \quad (12.38)$$

であり、X切片は、

$$\left. \begin{aligned} 0 &= 0.01378 + 0.00197 \cdot [I] \\ [I] &= \frac{-0.01378}{0.00197} = -7.0014 \end{aligned} \right\} \quad (12.39)$$

となるので、負の阻害定数 $-(\widehat{Ki})^{\text{回帰}} = -7.0014$ が得られる。推定された阻害定数は、 $(\widehat{Ki})^{\text{回帰}} = 7.0014$ であるが、阻害薬の濃度ごとに求めた最小の $(\widehat{Ki})_4 = 8.90$ よりも小さく、どのように解釈したら良いのであろうか。

推定された阻害定数 $(\widehat{Ki})^{\text{回帰}} = 7.0014$ の 95%信頼区間を求めたいが、どうしたら求められるのであろうか。図 12.15 に示した回帰直線で推定されたパラメータの比から $(\widehat{Ki})^{\text{回帰}}$ が推定されているので、式 (12.18) および式 (12.19) に示したようにパラメータの共分散行列 $\Sigma(\hat{\theta})$ に対する 2 次形式による $Var(\widehat{Ki})$ を適用すれば良いと思われる。しかしながら、元のデータは、推定された傾き $(\hat{\beta}_1)_i$ には、図 12.15 に示したような“揺らぎ”があり、これを考慮した重み付き回帰を適用することも考えられるが、95%信頼区間が広がり推定値精度が悪いことを示すことにしかならない。

このように、両逆数・Lineweaver-Burk 変換のみならず線形変換に基づく阻害定数 (Ki) の推定は、いずれにしても「推定に推定を重ねる」ことになり、推定精度の劣化が免れない。実験で得られたデータの平均値に対し、その 95%信頼区間を示すことは当然のことと思うのであが、線形変換で求められたミカエリス定数の推定値 (\widehat{Km}) に対し、95%信頼区間を示すことすら一般化していない。

阻害定数の推定値 (\widehat{Ki}) についても 95%信頼区間の計算方法を例示したいと思うのであるが、「推定に推定を重ねる」ことになり、推定誤差を考慮した重み付き回帰を適用する必要があり、複雑な式となり実用性に欠けるために、深入りしないことにする。ただし、阻害定数 (Ki) を直接推定するための計算式は知られており、非線形回帰を適用することにより、95%信頼区間を得ることができるので、Excel を用いた計算方法を第 12.5 節で詳しく示す。

12.4. 最大反応 V_{max} を共通とした場合の阻害定数 K_i の推定

(1, 1) 標示型ダミー変数の設定

表 12.30 に示したように、それぞれの阻害薬の濃度 $[I]_i$ ごとの推定された $(\widehat{Km})_i$ の標準誤差 SE が極めて大きくなっているのは、基質濃度 $[S]$ の設定が、わずか 3 点しかないにもかかわらず、 $(V_{max})_i$ と $(Km)_i$ の 2 つのパラメータの推定を試みているためである。もちろん、 $(V_{max})_i$ が共通と仮定できるか否かの検討のためには必要ではある。 $(V_{max})_i$ が阻害薬の濃度 $[I]_i$ に関わらず一定であると仮定できる場合に、阻害薬の濃度 $[I]_i$ に共通の最大反応 $(V_{max})^{\text{共通}}$ とした場合の $(Km)'_i$ を推定したい。阻害薬の濃度ごとの $(Km)'_i$ を推定するために、表 12.33 に示すように (1, 1) 標示型ダミー変数を設定する。

表 12.33 阻害薬 A の濃度別 (1, 1) 標示型ダミー変数

i	阻害薬 A			(1, 1) 標示型ダミー変数 a_i				
	濃度 $[I]$	(V_{max})	$(Km)'$	a_0	a_1	a_2	a_3	a_4
0	0	$(V_{max})^{\text{共通}}$	$(Km)'_0$	1	0	0	0	0
1	8	$(V_{max})^{\text{共通}}$	$(Km)'_1$	0	1	0	0	0
2	14	$(V_{max})^{\text{共通}}$	$(Km)'_2$	0	0	1	0	0
3	20	$(V_{max})^{\text{共通}}$	$(Km)'_3$	0	0	0	1	0
4	28	$(V_{max})^{\text{共通}}$	$(Km)'_4$	0	0	0	0	1

共通の最大反応 $(V_{max})^{\text{共通}}$ に対し、5 つのミカエリス定数 $(Km)'_i$ をもつミカエリス・メンテン式は、5 つのダミー変数 a_i を用いて、

$$\left. \begin{aligned} (Km)' &= (Km)'_0 a_0 + (Km)'_1 a_1 + (Km)'_2 a_2 + (Km)'_3 a_3 + (Km)'_4 a_4 \\ V &= \frac{(V_{max})^{\text{共通}} \cdot [S]}{[S] + (Km)'_i} \end{aligned} \right\} \quad (12.40)$$

として設定する。阻害薬濃度が $[I]_0$ の場合にダミー変数 $a_0=1$ で、他の濃度に対しては、すべて 0、阻害薬濃度が $[I]_1$ の場合にダミー変数 $a_1=1$ で、他の濃度に対しては、すべて 0 となり、結果として最大反応 $(V_{max})^{\text{共通}}$ に対するミカエリス・メンテン式となる。

Excel のソルバーを用いた非線形回帰

表 12.34 に示すのは、表 12.24 に示した阻害薬 A の濃度 $[I]_i$ 別に酵素の濃度 $[S]_j$ の反応速度 V のデータを行方向に展開した全 45 件の Excel のデータで、濃度 $[I]_i$ に対応する (1, 1) 標示型ダミー変数 a_i が設定されている。さらに、式 (12.40) の $(\widehat{Km})'_i$ を行列の積の関数 $\text{Mmult}(a \text{ の範囲}, (\widehat{Km})' \text{ の範囲})$ で求め、最大反応を $(V_{max})^{\text{共通}}$ とする反応速度の推定値 \hat{V} を計算している。

表 12.34 最大反応 V_{max} を共通とする阻害薬 A の濃度別のミカリス定数 $(K_m)_i'$ の推定

		初期値					ソルバー解					
		$(V_{max}^{\wedge})' =$	50.0							$(V_{max}^{\wedge})' =$	47.3925	
		$(K_m^{\wedge})'_0 =$	1.0							$(K_m^{\wedge})'_0 =$	0.7846	
		$(K_m^{\wedge})'_1 =$	2.0							$(K_m^{\wedge})'_1 =$	1.2596	
		$(K_m^{\wedge})'_2 =$	3.0							$(K_m^{\wedge})'_2 =$	1.7071	
		$(K_m^{\wedge})'_3 =$	4.0							$(K_m^{\wedge})'_3 =$	2.2228	
		$(K_m^{\wedge})'_4 =$	5.0							$(K_m^{\wedge})'_4 =$	3.1212	
										残差平方和 $S_e =$	162.1089	
No	i	阻害薬 $[I]$	酵素 $[S]$	(1, 1)型ダミー変数 a_0, a_1, a_2, a_3, a_4					反応 V	予測値 $(K_m^{\wedge})'_i$	残差 $\varepsilon^{\wedge} = V - \hat{V}$	
1	0	0	0.6	1	0	0	0	0	19.488	0.7846	20.537	-1.049
2			0.6	1	0	0	0	0	17.910	0.7846	20.537	-2.627
:												
9			1.2	1	0	0	0	0	32.158	0.7846	28.656	3.502
10	1	8	0.6	0	1	0	0	0	15.039	1.2596	15.292	-0.253
:												
18			1.2	0	1	0	0	0	25.253	1.2596	23.122	2.131
19	2	14	0.6	0	0	1	0	0	12.338	1.7071	12.325	0.013
:												
27			1.2	0	0	1	0	0	21.287	1.7071	19.563	1.724
28	3	20	0.6	0	0	0	1	0	10.179	2.2228	10.074	0.105
:												
36			1.2	0	0	0	1	0	18.145	2.2228	16.615	1.530
37	4	28	0.6	0	0	0	0	1	7.635	3.1212	7.641	-0.006
:												
44			1.2	0	0	0	0	1	12.232	3.1212	13.161	-0.929
45			1.2	0	0	0	0	1	14.020	3.1212	13.161	0.859
$(K_m^{\wedge})'_i = \text{Mmult} (a \text{ の範囲}, (K_m^{\wedge})' \text{ の範囲})$												

元の反応速度 V と推定値 \hat{V} の残差を $\varepsilon_{ij} = V_{ij} - \hat{V}_{ij}$ とし、残差に関する平方和 S_e を SumSq (ε の範囲) によって求めている。ここに示した結果は、表 12.34 左上段の「初期値」を「ソルバー解」に数値のみをペーストし、Excel のソルバーで残差平方和 S_e を最小にするように $(V_{max})^{\text{共通}}$, $(K_m)_0'$, $(K_m)_1'$, ..., $(K_m)_4'$ を変更させて求めた結果である。このように、Excel のソルバーを使うことにより、阻害薬濃度 $[I]_i$ ごとの最大反応を $(V_{max})^{\text{共通}}$ とする非線形回帰の推定値が一気に得られるようになったことは、画期的なことである。両逆数・Lineweaver-Burk 変換式に対し、ダミー変数を用い、共通の切片 $[1/(V_{max})^{\text{共通}}]$ を持つ線形変換式によるパラメータ推定を行なうこともできなくはないが、非線形回帰で直接パラメータ推定を簡単にできるのに、わざわざ面倒なことをする必然性がない。

表 12.35 に示すのは、表 12.34 で推定された $(\widehat{V_{max}})^{\text{共通}} = 47.3925$, $(\widehat{K_m})'_0 = 0.7846$, $(\widehat{K_m})'_2 = 1.7071$, $(\widehat{K_m})'_4 = 3.1212$ を用い、基質濃度 $[S]$ を変化させて反応速度 V をミカエリス・メンテン式で計算した結果である。図 12.16 には、阻害薬濃度 $[I]_i$ ごとの散布図にミカエリス・メンテン曲線の推定値を重ね書きし、さらに、推定された $(\widehat{K_m})'_i$ を $(\widehat{V_{max}})^{\text{共通}} = 47.3965$ の 2 分の 1 の直線上に \diamond を重ね書きしたものである。

表 12.35 V_{max} を共通とする阻害薬 A の濃度別のミカエリス・メンテン曲線の推定

基質 [S]	反応V		
	[I] ₀ =0	[I] ₂ =14	[I] ₄ =28
0.01	0.60	0.28	0.15
0.03	1.75	0.82	0.45
0.1	5.36	2.62	1.47
0.3	13.11	7.08	4.16
0.6	20.54	12.33	7.64
1	26.56	17.51	11.50
2	34.04	25.57	18.51
3	37.57	30.21	23.23
10	43.94	40.48	36.12
30	46.18	44.84	42.93
100	47.02	46.60	45.96
1000	47.36	47.31	47.25
10000	47.39	47.38	47.38

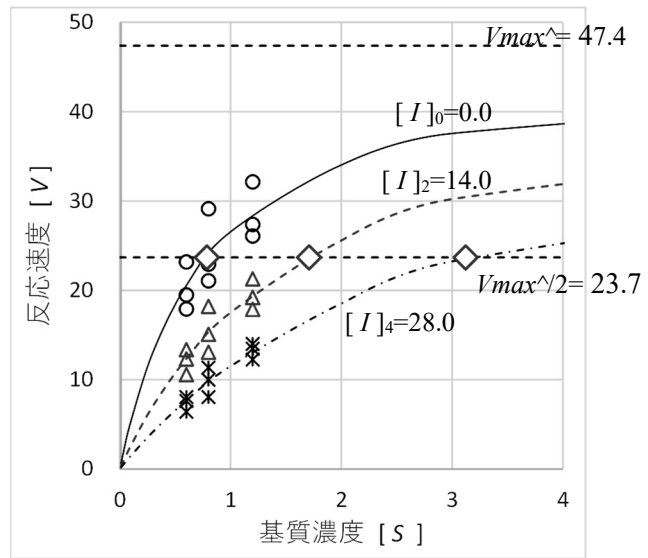


図 12.16 阻害薬濃度 $[I]_i$ ごとのミカエリス・メンテン曲線

表 12.36 に示すのは、表 12.34 で推定された V_{max} を共通としたミカエリス定数 $(K_m)'_i$ を用い、阻害薬の濃度 $[I]_0$ に対するミカエリス定数 $(K_m)'_0$ を基準とした場合の阻害定数 $(K_i)'_i$ の推定結果である。格段に改善しているのが、阻害薬 A の濃度別の阻害定数 $(K_i)'_i$ の推定であり、表 12.32 に示した標準偏差 2.28 から 1.61 と小さくなっている。

表 12.36 非線形回帰による最大反応 V_{max} を共通とした場合の阻害定数 $(K_i)'_i$

No.	A薬 [I]	$(V_{max})_{共通}$	$(K_m)'$	阻害定数 $(K_i)'$
0	0	47.3925	0.7846	—
1	8	47.3925	1.2596	13.22
2	14	47.3925	1.7071	11.91
3	20	47.3925	2.2228	10.91
4	28	47.3925	3.1212	9.40
	平均	47.3925		11.36
	SD	0.0000		1.61

$(K_i)'_i = \{(K_m)'_0 / [(K_m)'_i - (K_m)'_0]\} * [I]_i$

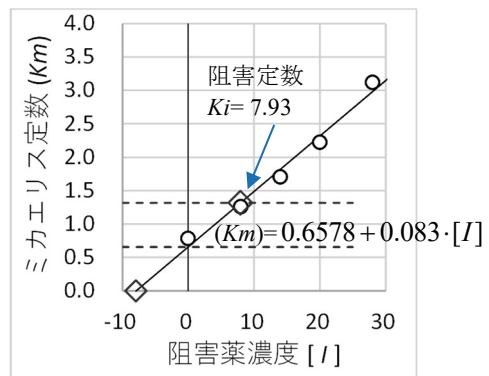


図 12.17 阻害薬 A の濃度と推定されたミカエリス定数との関連

図 12.17 に示すように、阻害薬濃度 $[I]_i$ と推定されたミカエリス定数 $(K_m)'_i$ との関連は、ほぼ直線上と見なせるので、阻害薬濃度 $[I]_i$ 比例して $(K_m)'_i$ が増加していることが確認される。推定される回帰直線は、 $(\widehat{K_m})'_i = 0.6578 + 0.083 \cdot [I]_i$ なので、阻害薬濃度 $[I]$ が 0 のミカエリス定数の推定値は、 $(\widehat{K_m})'_0 = 0.6578$ であり、 $(\widehat{K_m})'_0$ の 2 倍となる阻害定数 $(K_i)^{回帰}$ は、

$$\left. \begin{aligned} 2 \times 0.6578 &= 0.6578 + 0.083 \times (Ki)^{\text{回帰}} \\ (Ki)^{\text{回帰}} &= \frac{0.6578}{0.083} = 7.93 \end{aligned} \right\} \quad (12.41)$$

と推定される。表 12.36 で推定された阻害薬濃度ごとに推定した阻害定数 $(\widehat{Ki})'_i$ よりも小さくなっている。この原因は、 $(\widehat{Km})'_0$ および $(\widehat{Km})'_4$ が若干ではあるが大きめに推定された結果を反映している。

このように、 V_{max} を共通として推定された $(Km)'_i$ を反応とし、阻害薬濃度 $[I]_i$ に対する回帰分析により阻害定数 $(Ki)^{\text{回帰}}$ を推定する方法でも、 $(Km)'_i$ の推定誤差の影響を受けるために阻害定数 $(Ki)^{\text{回帰}}$ の推定が不安定になることが避けられない。

阻害定数 Ki の推定方法による揺らぎ

同一の酵素反応データを用いてもミカエリス・メンテン式による推定方法による最大反応 (V_{max}) およびミカエリス定数 (Km) の推定値が大きく揺らぎ、またそれらの 95%信頼区間も大きく揺らぐことを示したきた。酵素阻害に対する阻害定数 (Ki) の揺らぎについて、これまでの結果を比較する。表 12.30 に示したのは、(1) 阻害薬濃度ごとに両逆数・Lineweaver-Burk 変換により推定したミカエリス定数 $(\widehat{Km})'_i$ を用いた方法である。表 12.32 に示したのは、(2) 両逆数・Lineweaver-Burk 変換により推定した傾き $(\hat{\beta}_1)'_i$ を用いた方法である。表 12.35 に示したのは、(3) V_{max} を共通とする非線形回帰によるミカエリス定数 $(Km)'_i$ を用いた方法であった。これらの推定値を用いた回帰分析から推定される阻害定数 (Ki) は、表 12.37 に示すように、2.05, 7.00, 7.93 と一定しない。また、阻害薬の濃度 $[I]$ ごとに濃度が 0 の基準となる $(\widehat{Km})'_0$ との比較で算出した阻害定数 $(Ki)_i$ の平均は、10.39, 11.64, 11.36 と比較的安定しているが、それらの標準偏差 SD は、3.94, 2.28, 1.64 であり、推定に推定を重ねることによる誤差の伝播による推定値の揺らぎが大きくなることの反映である。

表 12.37 各種の推定方法により求められた阻害定数 (\widehat{Ki}) の比較

i	阻害薬A 濃度 $[I]$	(1) 両逆数 Km		(2) 両逆数 傾き		(3) 非線形 V_{max} 共通	
		$(Km^{\wedge})'$	$(Ki^{\wedge})'$	β^{\wedge}_1	$(Ki^{\wedge})'$	$(Km^{\wedge})'$	$(Ki^{\wedge})'$
0	0	0.8767	—	0.0177	—	0.7846	—
1	8	1.3429	15.04	0.0276	14.40	1.2596	13.22
2	14	1.9810	11.11	0.0382	12.14	1.7071	11.91
3	20	2.6430	9.93	0.0497	11.11	2.2228	10.91
4	28	5.3564	5.48	0.0735	8.90	3.1212	9.40
	平均		10.39		11.64		11.36
	SD		3.94		2.28		1.61
	回帰分析 (Ki) 推定値		2.05		7.00		7.93

12.5. 非線形回帰による阻害定数 K_i の直接推定

表 12.24 に示したように基質濃度 $[S]$ を (0.6, 0.8, 1.2) と 3 水準に設定し, 阻害薬 A の濃度 $[I]_i$ を (0, 8, 14, 20, 28) と 5 水準とした繰り返し 3 の 2 因子型の実験データについて, ミカエリス・メンテン式を用い, 最大反応 $(Vmax)_i$ と最大反応の 1/2 となる基質の濃度・ミカエリス定数 $(Km)_i$ を推定し, ミカエリス定数 $(Km)_0$ を 2 倍にする阻害薬 A の濃度としての阻害定数 $(Ki)_i$ を推定してきた. また, 阻害薬濃度 $[I]_i$ ごとのミカエリス定数 $(Km)_i$ に対する回帰分析により阻害定数 $(Ki)^{回帰}$ も推定したのであるが, 安定した阻害定数 (Ki) を得ることができないだけでなく, 95%信頼区間の推定も行えなかった.

阻害定数 K_i の直接推定するための式の導出

式 (12.32) に示したように, 阻害薬が含まれない $[I]_0 = 0$ の場合のミカエリス定数の推定値を $(Km)_0$, 阻害薬の濃度 $[I]_i$ に対するミカエリス定数の推定値を $(Km)_i$ とした場合に, $(Km)_0$ を 2 倍とする阻害薬の濃度の推定値を $(Ki)_i$ とした.

$$\left. \begin{aligned} (Km)_i &= (Km)_0 + (Km)_0 \frac{[I]_i}{(Ki)_i} \\ &= (Km)_0 \left\{ 1 + \frac{[I]_i}{(Ki)_i} \right\} \end{aligned} \right\} \text{式 (12.32) 再掲}$$

この $(Km)_i$ に対する個別の $(Ki)_i$ ではなく, 共通の $(Ki)^{共通}$ と置きなおし, ミカエリス・メンテン式の (Km) に代入すると,

$$V = \frac{(Vmax) \cdot [S]}{[S] + (Km)_0 \left\{ 1 + \frac{[I]_i}{(Ki)^{共通}} \right\}} \quad (12.42)$$

が得られる. 阻害薬の濃度が, $[I]_0 = 0$ ならば, 通常のみカエリス・メンテン式となる. この式は, それぞれの阻害薬の濃度 $[I]_i$ ごとのミカエリス定数 $(Km)_i$ を推定するのではなく, 阻害定数 $(Ki)^{共通}$ を用いて, 推定値としての $(\widehat{Km})_i$ も含めて同時推定を行なうことにより, 安定した阻害定数 $(Ki)^{共通}$ を推定することが期待できる.

式 (12.42) を用いたミカエリス・メンテン式のあてはめは,

$$\begin{aligned} [I]_0 = 0 \text{ に対しては, } & (\widehat{Km})_0 \text{ を推定し,} \\ [I]_1 = 8 \text{ nM に対しては, } & [(\widehat{Km})_0 + (\widehat{Km})_0 \times 8 / (\widehat{Ki})^{共通}] \text{ を推定し, 順次} \\ & \vdots \\ [I]_4 = 28 \text{ nM について} & [(\widehat{Km})_0 + (\widehat{Km})_0 \times 28 / (\widehat{Ki})^{共通}] \end{aligned}$$

を推定する. このように, 阻害定数 (Ki) の推定の元になる $(\widehat{Km})_0$ の推定に, 阻害薬の濃度 $[I]_i$ の全てのデータが総動員されることに特徴がある.

NLIN プロシジャによる阻害定数 K_i の直接推定

直接推定のため表 12.24 に示したデータを元に SAS データファイルを作成する。基質濃度 $[S]$ は、DATA ステップの do ステートメントで (0.6, 0.8, 1.2) として与え、反応速度 V を 1 つ読み込みでは、SAS データファイルに書き込む。1 行分の 3 つのデータの読み込みと書き込みが終わると、次の行のデータの読み込みと書き込みに移り、最後の行まで継続する。作成された 45 行の「d01」 SAS データセットを、PRINT プロシジャで書き出している。

SAS プログラム 12.7 SAS データファイルの作成

```
TITLE1 '阻害薬 A. SAS Y. Takahashi 2023/07/28' ;
DATA d01 ;
  input A I @ ;
  do S = 0.6, 0.8, 1.2 ;
    input V @ ; ; output ;
  end ;
/* A I S:0.6 S:0.8 S:1.2 */
datalines ;
0 0 19.488 22.922 27.364
0 0 17.910 21.071 26.094
0 0 23.168 29.148 32.158
1 8 15.039 17.896 22.449
1 8 13.204 15.997 20.957
1 8 17.061 22.603 25.253
2 14 12.338 15.076 19.199
2 14 10.568 13.065 17.875
2 14 13.364 18.218 21.287
3 20 10.179 12.718 16.559
3 20 8.640 10.653 15.030
3 20 10.841 14.871 18.145
4 28 7.635 9.979 13.461
4 28 6.415 8.070 12.232
4 28 8.044 11.364 14.020
;
proc print data=d01 ; run;
```

NLIN プロシジャによる阻害定数 (K_i) の直接推定プログラムを次に示す。

SAS プログラム 12.8 /* ミカエリス・メンテン式 阻害定数 K_i の直接推定 */

```
proc nlin data=d01 list ;
  parms Vmax = 50 Km0 = 1 Ki = 10 ;
  Km = Km0 * (1 + I / Ki) ;
  model V = Vmax * S / (S + Km) ;
run ;
```

表 12.38 に NLIN プロシジャによる阻害定数 (K_i) の直接推定の結果を示す。推定された阻害定数は $(\hat{K}_i)^{\text{共通}} = 10.5569$ で、これまで阻害薬の濃度ごとのミカエリス定数から推定した阻害定数の平均的な推定値となっている。NLIN プロシジャによる直接推定では、阻害定数

$(\widehat{K_i})^{\text{共通}}$ の標準誤差が $SE((\widehat{K_i})^{\text{共通}}) = 2.0210$ として推定され、95%信頼区間も (6.4783, 14.6354) として推定することができる。

表 12.38 NLIN プロシジャによる阻害定数 K_i の直接推定

要因	自由度	平方和	平均平方	F 値	近似 Pr > F
Model	3	13623.5	4541.2	1079.44	<.0001
Error	42	176.7	4.207		
Uncorrected	45	13800.2			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		
Vmax	47.6932	6.9798	33.6074	61.7790	
Km0	0.7717	0.2428	0.2818	1.2616	
Ki	10.5569	2.0210	6.4783	14.6354	

推定結果のまとめとしては、表 12.39 に示すように、阻害薬 A の濃度 $[I]_i$ ごとにミカエリス定数 $(Km)_i$ の推定を式 (12.32) 再掲 を用いて次に示すように

$$\left. \begin{aligned} (\widehat{Km})_1 &= (\widehat{Km})_0 \left\{ 1 + \frac{[I]_1}{(\widehat{K_i})^{\text{共通}}} \right\} = 0.7717 \times \left\{ 1 + \frac{8}{10.5569} \right\} = 1.3565 \\ &\vdots \\ (\widehat{Km})_4 &= (\widehat{Km})_0 \left\{ 1 + \frac{[I]_4}{(\widehat{K_i})^{\text{共通}}} \right\} = 0.7717 \times \left\{ 1 + \frac{28}{10.5569} \right\} = 2.8185 \end{aligned} \right\} \quad (12.43)$$

求めることができ、阻害定数 $(\widehat{K_i})_i$ の推定が適切に行なわれているか検証することができる。推定された $(\widehat{Km})_0 = 0.7717$ についての標準誤差 $SE[(\widehat{Km})_0] = 0.2428$ の表示も推定精度を明示するために役に立つ。

表 12.39 NLIN プロシジャにより推定されたパラメータの一覧

i	A薬 $[I]$	最大値		ミカエリス定数		阻害定数	
		$Vmax^{\wedge}$	$SE[(Vmax^{\wedge})]$	(Km^{\wedge})	$SE[(Km^{\wedge})]$	$(K_i^{\wedge})^{\text{共通}}$	$SE[(K_i^{\wedge})]$
0	0	47.6932	6.9798	0.7717	0.2428	—	
1	8	47.6932	6.9798	1.3565	—	10.5569	2.02
2	14	47.6932	6.9798	1.7951	—	10.5569	2.02
3	20	47.6932	6.9798	2.2337	—	10.5569	2.02
4	28	47.6932	6.9798	2.8185	—	10.5569	2.02
$(Km^{\wedge})_i = (Km^{\wedge})_0 [1 + [I]_i / (K_i^{\wedge})^{\text{共通}}]$							

さらに、図 12.18 に示すように、阻害定数 (K_i) を用いたミカエリス定数の推定値を散布図上にプロットし、 $(\widehat{Km})_0 = 0.7717$ の水平線などを重ね書きすることにより、実験結果の理解と解釈に役に立つ。

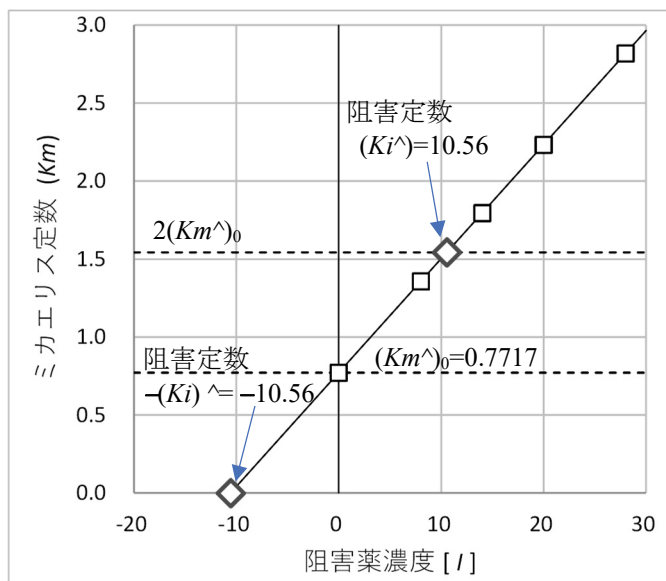


図 12.18 阻害定数 $(\hat{K}i)$ を用いたミカエリス定数 $(\hat{K}m)_i$ の推定値

両逆数・Lineweaver-Burk 変換による阻害定数 Ki の推定結果との比較

両逆数・Lineweaver-Burk 変換により，阻害薬濃度 $[I]_i$ ごとにミカエリス定数 $(Km)_i$ を推定し，表 12.30 に示すように阻害定数 $(Ki)_i$ を推定した．さらに，図 12.14 右に示したように阻害薬 A の濃度 $[I]_i$ に対する回帰直線から，阻害定数 $(Ki)_i$ を推定したが，ミカエリス定数 $(Km)_i$ の推定値の変動により，安定した阻害定数 $(Ki)^{\text{回帰}}$ の推定ができなかった．回帰係数の傾き $(\beta_1)_i = (Km)_i / (Vmax)_i$ を用いることにより，表 12.32 に示すように阻害定数 $(Ki)_i$ の推定は，ミカエリス定数 $(Km)_i$ を用いた場合に比べ安定したとは言え，図 12.15 に示した回帰直線をあてはめて推定した $(Ki)^{\text{回帰}}$ も，回帰係数の傾き $(\beta_1)_i = (Km)_i / (Vmax)_i$ の変動に影響を受け，安定した推定値が得られているとは言えない．

阻害薬濃度 $[I]_i$ に共通の $(Vmax)$ を設定した非線形回帰により，阻害薬濃度 $[I]_i$ ごとにミカエリス定数 $(Km)_i$ を推定し，表 12.36 および図 12.17 に示すように阻害定数 (Ki) の推定を行ってみても，安定した阻害定数 (Ki) の推定ができているとは言いがたい．不安定な阻害定数 (Ki) の推定となる原因は，推定に推定を重ねることによる誤差の伝播による影響であり，さらに，阻害定数 (Ki) の標準誤差の推定，95%信頼区間の推定も思うようにできないと欠点を内在している．

式 (12.42) を用い，阻害薬濃度 $[I]_i$ ごとにミカエリス定数 $(Km)_i$ を推定することなく阻害定数 (Ki) を非線形回帰により直接推定できることは，山崎ら (2008) にも明示されているが，実データを用いた推定結果が示されていない．

ミカエリス定数 $(Km)_i$ が阻害薬の濃度 $[I]_i$ に比例するとした場合の阻害定数 (K_i) の直接推定に際し、NLINプロシジャを用いた。これにより、阻害定数 (K_i) パラメータの標準誤差 SE を推定することが可能となった。

Excel の行列関数を用いた阻害定数 K_i の直接推定

阻害定数 $(K_i)^{\text{共通}}$ の推定を含むミカエリス・メンテン式のあてはめをExcelによって行い、NLINプロシジャの解析方法について理解を深めたい。表 12.2 に示した、ミカエリス・メンテン式を非線形回帰で解いた手順により、阻害薬の濃度 $[I]_i$ を含み、阻害定数 $(K_i)^{\text{共通}}$ の推定を加える。Excelでの解析には、式(12.42)をパラメータに関して偏微分することから始める。

式(12.42)を偏微分しやすいように

$$V = \frac{(Vmax) \cdot [S]}{[S] + (Km)_0 \left\{ 1 + \frac{[I]_i}{(K_i)^{\text{共通}}} \right\}} = \frac{(Vmax) \cdot [S]}{[S] + (Km)} \quad \left. \vphantom{\frac{(Vmax) \cdot [S]}{[S] + (Km)_0 \left\{ 1 + \frac{[I]_i}{(K_i)^{\text{共通}}} \right\}}} \right\} \quad (12.44)$$

ただし、 $(Km) = (Km)_0 \left\{ 1 + \frac{[I]_i}{(K_i)^{\text{共通}}} \right\}$

とする。パラメータ $(Vmax)$, $(Km)_0$, $(K_i)^{\text{共通}}$ について偏微分すると

$$\left. \begin{aligned} z_1 &= \frac{\partial V}{\partial (Vmax)} = \frac{[S]}{[S] + (Km)} \\ z_2 &= \frac{\partial V}{\partial (Km)_0} = - \left\{ 1 + \frac{[I]_i}{(K_i)^{\text{共通}}} \right\} \frac{(Vmax) \cdot [S]}{\{[S] + (Km)\}^2} \\ z_3 &= \frac{\partial V}{\partial (K_i)^{\text{共通}}} = \left\{ (Km)_0 \frac{[I]_i}{\{(K_i)^{\text{共通}}\}^2} \right\} \frac{(Vmax) \cdot [S]}{\{[S] + (Km)\}^2} \end{aligned} \right\} \quad (12.45)$$

が得られる。

表 12.24 に示した阻害薬 A のデータを、行方向に展開した結果を表 12.40 に示す。阻害薬の 5 段階の濃度 $[I]_i$ ごとに基質の 3 段階の濃度 $[S]_j$ に対し、繰り返しが 3 のデータなので、全体で $5 \times 3 \times 3 = 45$ 行のデータに対し、反応速度 V が示されている。反応速度の推定値 \hat{V} は、式(12.44)によって計算され、第 1 行目は、

$$\hat{V}_1 = \frac{(\widehat{Vmax}) \cdot [S]}{[S] + (\widehat{Km})_0 \left\{ 1 + \frac{[I]_0}{(\widehat{K_i})^{\text{共通}}} \right\}} = \frac{50 \times [0.6]}{[0.6] + 1 \times \left\{ 1 + \frac{[0]}{10} \right\}} = 18.750$$

$$\text{残差} : \varepsilon_1 = V_1 - \hat{V}_1 = 19.488 - 18.750 = 0.738$$

表 12.40 非線形回帰による阻害定数 K_i の推定のための Excel の計算シート

初期値				$\theta^{(m-1)}$	δ^{\wedge}	$\theta^{(m)}$	$(Z^T Z)^{-1}$			
$(Vmax^{\wedge})=$	50.00	$(Vmax^{\wedge})=$	50.0000	-3.7672	46.2328	21.0589	0.7911	3.5483		
$(Km^{\wedge})_0=$	1.00	$(Km^{\wedge})_0=$	1.0000	-0.2991	0.7009	0.7911	0.0304	0.1449		
$(Ki^{\wedge})_{共通}=$	10.00	$(Ki^{\wedge})_{共通}=$	10.0000	0.6634	10.6634	3.5483	0.1449	0.8856		
		$S_e=$	374.69							
阻害薬 A	酵素	反応	推定値	残差			∂_{Vmax}	∂_{Km_0}	∂_{Ki}	
No	i	I	S	V	V^{\wedge}	$\varepsilon=V-V^{\wedge}$	Km	z_1	z_2	z_3
1	0	0	0.6	19.488	18.7500	0.7380	1.0000	0.3750	-11.7188	0.0000
2		0	0.6	17.910	18.7500	-0.8400	1.0000	0.3750	-11.7188	0.0000
9		0	1.2	32.158	27.2727	4.8853	1.0000	0.5455	-12.3967	0.0000
10	1	8	0.6	15.039	12.5000	2.5390	1.8000	0.2500	-9.3750	0.4167
18		8	1.2	25.253	20.0000	5.2530	1.8000	0.4000	-12.0000	0.5333
19	2	14	0.6	12.338	10.0000	2.3380	2.4000	0.2000	-8.0000	0.4667
27		14	1.2	21.287	16.6667	4.6203	2.4000	0.3333	-11.1111	0.6481
28	3	20	0.6	10.179	8.3333	1.8457	3.0000	0.1667	-6.9444	0.4630
36		20	1.2	18.145	14.2857	3.8593	3.0000	0.2857	-10.2041	0.6803
37	4	28	0.6	7.635	6.8182	0.8168	3.8000	0.1364	-5.8884	0.4339
44		28	1.2	12.232	12.0000	0.2320	3.8000	0.2400	-9.1200	0.6720
45		28	1.2	14.020	12.0000	2.0200	3.8000	0.2400	-9.1200	0.6720

$$(\widehat{Km}) = (\widehat{Km})_0 \left\{ 1 + \frac{[I]_i}{(\widehat{Ki})_{共通}} \right\} = 1 \times \left\{ 1 + \frac{[0]}{10} \right\} = 1.0000$$

$$z_{1,1} = \frac{[S]}{[S] + (\widehat{Km})} = \frac{[0.6]}{[0.6] + 1.0} = 0.3750$$

$$z_{2,1} = - \left\{ 1 + \frac{[I]_i}{(\widehat{Ki})_{共通}} \right\} \frac{(\widehat{Vmax}) \cdot [S]}{\{[S] + (\widehat{Km})\}^2} = - \left\{ 1 + \frac{[0]}{10} \right\} \frac{50 \times [0.6]}{\{[0.6] + 1\}^2} = -11.7188$$

$$z_{3,1} = \left\{ (\widehat{Km})_0 \frac{[I]_i}{\{(\widehat{Ki})_{共通}\}^2} \right\} \frac{(\widehat{Vmax}) \cdot [S]}{\{[S] + (\widehat{Km})\}^2} = \left\{ 1 \times \frac{[0]}{10^2} \right\} \frac{50 \times [0.6]}{\{[0.6] + 1\}^2} = 0.0000$$

として計算されており，計算式を設定した 1 行目をフィルハンドルで，45 行目までコピーして計算させている。

表 12.40 上段に示しているのが，Excel での非線形モデルを解くための設定である．推定したいパラメータは， (\widehat{Vmax}) ， $(\widehat{Km})_0$ ， $(\widehat{Ki})_{共通}$ であり，これらを総称して $\hat{\theta}$ とする．偏微分した z_1 ， z_2 ， z_3 について（45 行×3 列）の微係数行列を Z とし，積和行列 $(Z^T Z)$ の逆行

$$(Z^T Z)^{-1} = \begin{array}{|c|c|c|} \hline \mathbf{21.0589} & 0.7911 & 3.5483 \\ \hline 0.7911 & \mathbf{0.0304} & 0.1449 \\ \hline 3.5483 & 0.1449 & \mathbf{0.8856} \\ \hline \end{array}$$

列が、として計算されている。残差のベクトル $\hat{\boldsymbol{\varepsilon}} = (\mathbf{V} - \hat{\mathbf{V}})$ を反応とし、微係数行列 \mathbf{Z} をデザイン行列と見なした線形回帰（残差線形化法）により、 $\hat{\boldsymbol{\delta}}$ が

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\boldsymbol{\varepsilon}} = \begin{bmatrix} -3.7672 \\ -0.2991 \\ 0.6634 \end{bmatrix}$$

推定されている。

初期値を設定した $\hat{\boldsymbol{\theta}}^{(m-1)}$ に $\hat{\boldsymbol{\delta}}$ を加え、 $\hat{\boldsymbol{\theta}}^{(m)}$ を推定する。

$$\hat{\boldsymbol{\theta}}^{(m)} = \hat{\boldsymbol{\theta}}^{(m-1)} + \hat{\boldsymbol{\delta}} = \begin{bmatrix} 50.0000 \\ 1.0000 \\ 10.0000 \end{bmatrix} + \begin{bmatrix} -3.7672 \\ -0.2991 \\ 0.6634 \end{bmatrix} = \begin{bmatrix} 46.2328 \\ 0.7009 \\ 10.6634 \end{bmatrix}$$

推定された $\hat{\boldsymbol{\theta}}^{(m)}$ をコピーし、 $\hat{\boldsymbol{\theta}}^{(m-1)}$ に値のみをペーストすると、

第 1 回目の反復：

	$\hat{\boldsymbol{\theta}}^{(m-1)}$	$\hat{\boldsymbol{\delta}}$	$\hat{\boldsymbol{\theta}}^{(m)}$	$(\mathbf{Z}^T \mathbf{Z})^{-1}$		
$(Vmax^\wedge) =$	46.2328	1.3779	47.6107	9.4470	0.3178	2.5372
$(Km^\wedge)_0 =$	0.7009	0.0662	0.7671	0.3178	0.0111	0.0953
$(Ki^\wedge)_{共通} =$	10.6634	-0.1278	10.5356	2.5372	0.0953	1.0290
$S_e =$	188.50					

が得られる。同様に第 2 回目の反復、第 3 回目の反復、第 4 回目の反復で、

第 4 回目の反復：

	$\hat{\boldsymbol{\theta}}^{(m-1)}$	$\hat{\boldsymbol{\delta}}$	$\hat{\boldsymbol{\theta}}^{(m)}$	$(\mathbf{Z}^T \mathbf{Z})^{-1}$		
$(Vmax^\wedge) =$	47.6932	0.0000	47.6932	11.5802	0.3964	2.7332
$(Km^\wedge)_0 =$	0.7717	0.0000	0.7717	0.3964	0.0140	0.1037
$(Ki^\wedge)_{共通} =$	10.5569	0.0000	10.5569	2.7332	0.1037	0.9709
$S_e =$	176.69					

$\hat{\boldsymbol{\delta}}$ が全て 0.0000 となり、残差線形化法によるパラメータの推定値が得られたことになる。この推定結果は、表 12.38 下段に示した SAS の NLIN プロシジャの推定結果に一致する。

表 12.38 下段（再掲）SAS の NLIN プロシジャの推定結果

パラメータ	推定値	近似標準誤差	近似 95% 信頼限界	
Vmax	47.6932	6.9798	33.6074	61.7790
Km0	0.7717	0.2428	0.2818	1.2616
Ki	10.5569	2.0210	6.4783	14.6354

さて、推定されたパラメータの標準誤差は、残差線形化法で使った積和行列 $(\mathbf{Z}^T \mathbf{Z})$ の逆行列 $(\mathbf{Z}^T \mathbf{Z})^{-1}$ に誤差分散の推定値

$$\hat{\sigma}^2 = \frac{S_e}{n - \text{パラメータ数}} = \frac{176.69}{45 - 3} = 4.2070$$

を掛けることにより、パラメータ共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ を求めることができる。

$$\Sigma(\hat{\theta}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2 = \begin{bmatrix} 11.5802 & 0.3964 & 2.7332 \\ 0.3964 & 0.0140 & 0.1037 \\ 2.7332 & 0.1037 & 0.9709 \end{bmatrix} \times 4.2070 = \begin{bmatrix} 48.7174 & 1.6676 & 11.4985 \\ 1.6676 & 0.0589 & 0.4362 \\ 11.4985 & 0.4362 & 4.0844 \end{bmatrix} \quad (12.46)$$

その対角要素がパラメータの分散なので、その平方根がパラメータの標準誤差となるので、 $SE(\widehat{Vmax}) = \sqrt{48.718} = 6.9798$ 、 $SE(\widehat{Km_0}) = \sqrt{0.0589} = 0.2428$ 、 $SE(\widehat{Ki}^{共通}) = \sqrt{4.0844} = 2.0210$ となり、この結果は、表 12.38 下段の結果に一致する。

Excel のソルバーを用いた簡便な阻害定数 Ki の直接推定

非線形回帰による阻害定数 $(Ki)^{共通}$ の推定の第 1 歩は、Excel のソルバーによる推定から始めることを勧める。表 12.41 に示すように、推定値 \hat{V} を式 (12.44) により設定し、残差から残差平方和 S_e を設定するだけで、阻害定数 $(Ki)^{共通}$ を推定することができる。パラメータの初期値を設定し、Excel のソルバーで、残差平方和 S_e を最小にするように、 (\widehat{Vmax}) 、 $(\widehat{Km_0})$ 、 $(\widehat{Ki})^{共通}$ を変更すると瞬時に「ソルバー解」とし、阻害定数 $(\widehat{Ki})^{共通} = 10.5563$ が即座に得られる。

表 12.41 Excel ソルバーを用いた阻害定数 Ki の推定

			初期値		初期値		ソルバー 解	
			$(Vmax^{\wedge})=$	50.00	$(Vmax^{\wedge})=$	50.00	$(Vmax)^{\wedge}=$	47.6901
			$(Km^{\wedge})_0=$	1.00	$(Km^{\wedge})_0=$	1.00	$(Km_0)^{\wedge}=$	0.7716
			$(Ki^{\wedge})^{共通}=$	10.00	$(Ki^{\wedge})^{共通}=$	10.00	$(Ki)^{\wedge}=$	10.5563
			残差平方和 $S_e=$		374.69	$S_e=$		176.69
No	阻害薬 A	酵素	反応	推定値	残差	推定値	残差	
	i	I	S	V	V^{\wedge}	V^{\wedge}	$\varepsilon = V - V^{\wedge}$	
1	0	0	0.6	19.4880	18.7500	20.8614	-1.3734	
2		0	0.6	17.9100	18.7500	20.8614	-2.9514	
9		0	1.2	32.1580	27.2727	29.0258	3.1322	
10	1	8	0.6	15.0390	12.5000	14.6259	0.4131	
18		8	1.2	25.2530	20.0000	22.3862	2.8668	
19	2	14	0.6	12.3380	10.0000	11.9475	0.3905	
27		14	1.2	21.2870	16.6667	19.1080	2.1790	
28	3	20	0.6	10.1790	8.3333	10.0983	0.0807	
36		20	1.2	18.1450	14.2857	16.6673	1.4777	
37	4	28	0.6	7.6350	6.8182	8.3708	-0.7358	
44		28	1.2	12.2320	12.0000	14.2417	-2.0097	
45		28	1.2	14.0200	12.0000	14.2417	-0.2217	

この結果は、NLIN プロシジャで計算した表 12.38 下段の結果とは、少数点以下 3 桁目、4 桁目で、収束の判定精度の差異のために一致しない。

12.6. 阻害様式の確認のための手順

阻害薬を加えた場合の阻害定数 (K_i) の推定法について、阻害薬濃度 $[I]$ ごとに両逆数・Lineweaver-Burk 変換を適用した場合の問題点を指摘し、それに代わる非線形回帰を用いることにより、安定した阻害定数 (K_i) の推定法について示してきた。一般的に阻害薬の酵素反応に対する作用は、(1) V_{max} は変化しないが K_m を増加させる、(2) K_m は変化しないが V_{max} を減少させる、(3) V_{max} と K_m を同じ割合で減少させる、(4) K_m の増加と共に V_{max} が減少する、に分類できることが知られている。どのような阻害様式なのかを判定するためには、それぞれの阻害薬の濃度 $[I]$ ごとに、 $(V_{max})_i$ と $(K_m)_i$ を別々に推定し、(1) から (4) で示したどの阻害様式となるかの見当が付ける必要がある。

Excel ソルバーを用いた個別パラメータの一括推定

表 12.33 に示した阻害薬 A の濃度別 (1, 1) 標示型ダミー変数を使い、 $(V_{max})_i$ と $(K_m)_i$ を一括して推定する方法を示す。表 12.42 に示すように、

表 12.42 Excel ソルバーを用いた個別パラメータの一括推定

		$(V_{max})'_i$	$(K_m)'_i$											ソルバー解				
初期値		50.0	1.0												$(V_{max})'_0$	$(K_m)'_0$	47.8120	0.7992
		50.0	2.0												$(V_{max})'_1$	$(K_m)'_1$	45.6536	1.1807
		50.0	3.0												$(V_{max})'_2$	$(K_m)'_2$	47.1047	1.6912
		50.0	4.0												$(V_{max})'_3$	$(K_m)'_3$	48.2438	2.2794
		50.0	5.0												$(V_{max})'_4$	$(K_m)'_4$	56.8061	3.9298
													残差平方和 S_e		161.9270			
No	群	阻害薬	酵素	(1, 1)型ダミー変数					反応			予測値	残差					
	i	$[I]$	$[S]$	a_0	a_1	a_2	a_3	a_4	V	$(V_{max})'_i$	$(K_m)'_i$	V^{\wedge}	$\varepsilon^{\wedge}=V-V^{\wedge}$					
1	0	0	0.6	1	0	0	0	0	19.488	47.8120	0.7992	20.503	-1.015					
2			0.6	1	0	0	0	0	17.910	47.8120	0.7992	20.503	-2.593					
:																		
9			1.2	1	0	0	0	0	32.158	47.8120	0.7992	28.699	3.459					
10	1	8	0.6	0	1	0	0	0	15.039	45.6536	1.1807	15.383	-0.344					
:																		
18			1.2	0	1	0	0	0	25.253	45.6536	1.1807	23.012	2.241					
19	2	14	0.6	0	0	1	0	0	12.338	47.1047	1.6912	12.336	0.002					
:																		
27			1.2	0	0	1	0	0	21.287	47.1047	1.6912	19.551	1.736					
28	3	20	0.6	0	0	0	1	0	10.179	48.2438	2.2794	10.053	0.126					
:																		
36			1.2	0	0	0	1	0	18.145	48.2438	2.2794	16.639	1.506					
37	4	28	0.6	0	0	0	0	1	7.635	56.8061	3.9298	7.524	0.111					
:																		
44			1.2	0	0	0	0	1	12.232	56.8061	3.9298	13.289	-1.057					
45			1.2	0	0	0	0	1	14.020	56.8061	3.9298	13.289	0.731					

$$(Vmax)'_i = Mmult(\mathbf{a} \text{ の範囲}, Vmax' \text{ の範囲}) \quad (12.47)$$

$$(Km)'_i = Mmult(\mathbf{a} \text{ の範囲}, Km' \text{ の範囲}) \quad (12.48)$$

と設定し、

$$\hat{V} = \frac{(Vmax)'_i[S]}{[S] + (Km)'_i} \quad (12.49)$$

により、阻害薬の濃度 $[I]$ ごとの推定値を初期値に対し一括して計算する。残差 $\hat{\varepsilon} = V - \hat{V}$ についての平方和 S_e を最小にするように、 $(Vmax)'_i$ と $(Km)'_i$ を一括して変更することにより、表 12.30 に示した両逆数・Lineweaver-Burk 変換による場合よりも簡便にかつ安定した推定値が得られている。

NLIN プロシジャによるパラメータの一括推定

NLIN プロシジャで $(Vmax)_i$ を共通とし、個別の $(Km)_i$ を推定する際に、ダミー変数 (a_0, a_1, a_2, a_3, a_4) を用いて推定した。このダミー変数を個別の $(Vmax)_i$ の推定に拡張することにより、一括して結果を得ることができる。

SAS プログラム 12.9 /* ミカエリス・メンテン式 ダミー変数 */

```
TITLE1 '阻害薬 A2. SAS Y. Takahashi 2023/03/04 ' ;
DATA d01 ;
  input A a0 a1 a2 a3 a4 I @ ;
  do S = 0.6, 0.8, 1.2 ;
    input V @ ; output ;
  end ;
/* A a0 a1 a2 a3 a4 I S:0.6 S:0.8 S:1.2 */
datalines ;
0 1 0 0 0 0 0 19.488 22.922 27.364
0 1 0 0 0 0 0 17.910 21.071 26.094
0 1 0 0 0 0 0 23.168 29.148 32.158
1 0 1 0 0 0 8 15.039 17.896 22.449
1 0 1 0 0 0 8 13.204 15.997 20.957
1 0 1 0 0 0 8 17.061 22.603 25.253
2 0 0 1 0 0 14 12.338 15.076 19.199
2 0 0 1 0 0 14 10.568 13.065 17.875
2 0 0 1 0 0 14 13.364 18.218 21.287
3 0 0 0 1 0 20 10.179 12.718 16.559
3 0 0 0 1 0 20 8.640 10.653 15.030
3 0 0 0 1 0 20 10.841 14.871 18.145
4 0 0 0 0 1 28 7.635 9.979 13.461
4 0 0 0 0 1 28 6.415 8.070 12.232
4 0 0 0 0 1 28 8.044 11.364 14.020
;
proc print data=d01 ; run;
```

NLIN プロシジャでは、 $Vmax$ を $(Vmax0*a0 + Vmax1*a1 + \dots + Vmax4*a4)$ に置き換え、`parm` ステートメントで初期値を全て 50 に設定する。

SAS プログラム 12.10 /* ミカエリス・メンテン式 個別 Km Vmax */

```

proc nlin data=d01 list ;
  parms Vmax0=50 Vmax1=50 Vmax2=50 Vmax3=50 Vmax4=50 ;
  parms Km0=1 Km1=2 Km2=3 Km3=4 Km4=5 ;
  Vmax_i = (Vmax0*a0+Vmax1*a1+Vmax2*a2+Vmax3*a3+Vmax4*a4) ;
  Km_i = (Km0*a0 +Km1*a1 +Km2*a2 +Km3*a3+ Km4*a4) ;
  model V = Vmax_i * S / (S + Km_i) ;
run ;

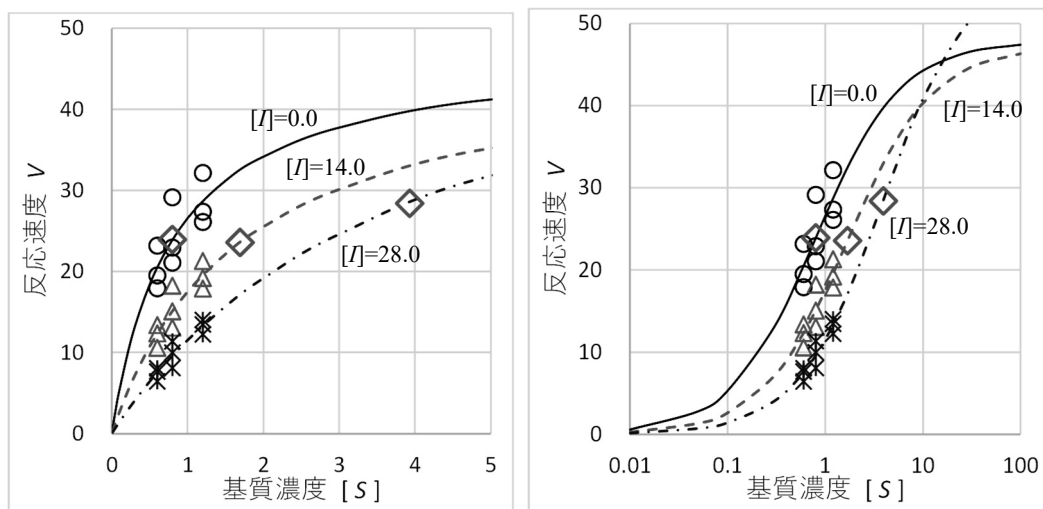
```

表 12.43 に示すように、非線形モデルの適用により $(Vmax)_i$, $(Km)_i$ が一括して出力される。表 12.30 に示した両逆数・Lineweaver-Burk 変換による推定値に比べ、 $(Vmax)_4$ の推定値が、72.8533 から 56.815 へと共通の推定値に近づき、 $(Km)_4$ も 7.6631 から 3.930 へと阻害薬濃度を反映した推定値が得られている。また、それらの標準誤差も全平均で 25%程度減少している。非線形モデルでの標準誤差を基準とした場合には、両逆数・Lineweaver-Burk 変換の標準誤差 SE は、34%増しとなる。

表 12.43 非線形モデルにより推定された個別のパラメーター一覧

阻害薬 $[I]$	パラメータ	推定値	近似標準誤差	パラメータ	推定値	近似標準誤差
$[I] = 0$ nM	Vmax0	47.812	9.490	Km0	0.799	0.333
$[I] = 8$ nM	Vmax1	45.654	14.193	Km1	1.181	0.648
$[I] = 14$ nM	Vmax2	47.111	21.994	Km2	1.692	1.221
$[I] = 20$ nM	Vmax3	48.243	33.107	Km3	2.279	2.209
$[I] = 28$ nM	Vmax4	56.815	76.552	Km4	3.930	6.594

図 12.19 に示すのは、表 12.42 および表 12.43 に示した阻害薬 A の濃度 $[I]_i$ 別に推定された最大反応 $(Vmax)_i$ およびミカエリス定数 $(Km)_i$ を用いて推定したミカエリス・メンテン曲線のプロット図であり、最大反応 $(Vmax)_i$ の 2 分の 1 となる位置にミカエリス定数 $(Km)_i$ を重ね書きした結果である。

図 12.19 阻害薬濃度 $[I]_i$ の増加に伴うミカエリス定数 $(Km)_i$ の変化

阻害様式の判別

図 12.20 に示すのは、阻害薬の濃度ごとの $(Vmax)_i$ と $(Km)_i$ を算出し、 Km を X 軸とし、 $Vmax$ を Y 軸とした散布図である。この図は、阻害様式の定義に従った判別をするために、両逆数・Lineweaver-Burk プロット図を用いる方法よりも簡潔で優れている。

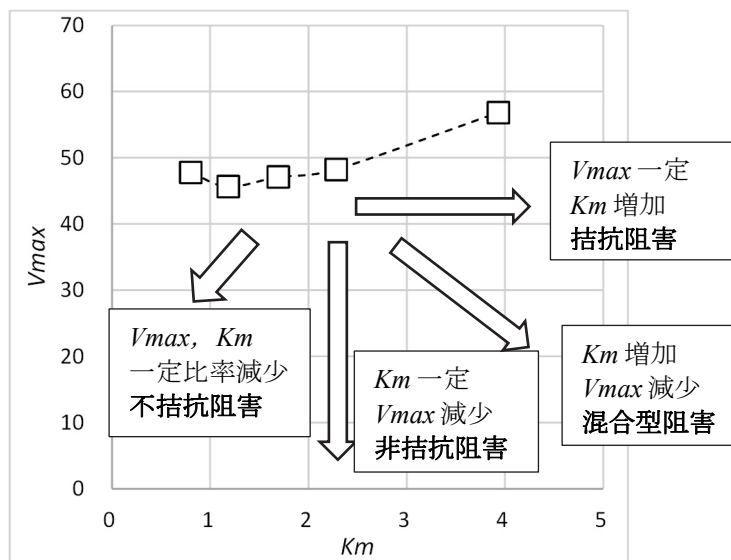


図 12.20 阻害様式の判別のためのミカエリス定数と最大反応の散布図

線形変換で悩むのは終わりにしませんか

酵素反応データの解析の基礎であるミカエリス・メンテン式によるパラメータの推定に、各種の線形変換による解析方法が綿々と受け継がれている。本質的に非線形であるミカエリス・メンテン式のパラメータ推定に多大な貢献を果たしてきたことは間違いない。ただし、線形化の方法により、推定値が不安定となる問題が内在している。もちろん、非線形回帰を使うことの必要性は説かれているものの、しかるべき統計ソフトを使えと言うだけで、その適用事例も計算原理の説明もないがしろにされて続けている。

救いの神が、Excel のソルバーを活用した非線形回帰を用いたカエリス・メンテン式によるパラメータの推定であり、計算の手間は、線形変換による方法よりはるかに簡単であり、安定した推定値も得られることを示した。また、阻害定数の推定にも、従来の線形化の方法に比べ、楽々のできることを例示した。各種の線形変換にこだわり続けるのは、そろそろ終わりにしませんか。

13. ロジスティック曲線のさらなる活用

前章のミカエリス・メンテン式は、酵素反応速度論の代表的な式であり、基質濃度 $[S]$ の対数を取ると 2 パラメータ・ロジスティック曲線となることを示した。多くの生物を用いた実験における化学物質の濃度あるいは用量の設定は、等比で与えることが一般的であり、対数変換した濃度に対する反応についての解析が主体となっている。これまでも示してきたように、Excel のソルバーによりロジスティック曲線のパラメータの推定が、容易にできることを示してきた。ただし、対数変換した濃度（用量）を用いる場合に、厄介な問題が発生する。濃度がゼロの対照群に対しては、対数変換ができないので、低めの濃度を便宜的に与え、最大反応となるような陽性対照群がある場合に、高めの濃度を便宜的に与えることで良いのだろうか。反応がロジスティック曲線となる場合であっても、最大反応の 20% から 80% 程度の範囲の濃度を設定することにより、直線のあてはめによる解析が便宜的に行なわれてきた。実際に、平行な直線をあてはめた事例ではあるが、シグモイド曲線らしき痕跡が残る場合に、平行とみなすことに無理が生ずる。このような場合に、曲線の傾きは同じだが、曲線の位置が異なる「平行なロジスティック曲線」のあてはめにチャレンジする。

13.1. 陰性対照と陽性対照を含むロジスティック曲線のあてはめ

第 11.2 節で取り上げた環境ホルモン投与後の子宮重量データにゴンペルツ・最大極値曲線をあてはめた際に、Vehicle 群について用量段階を 2 段階落とした $dose=0.001$ として扱った。Vehicle 群は、環境ホルモンが含まれない媒体のみが投与された群であり、 $dose=0.0$ であり対数を取るとマイナス無限大となり、対数変換が不可能になってしまうためである。

内臓痛試験データ

本節で取り上げるのは、陰性対照群としての Vehicle 群、さらに陽性対照群としての Sham 群を含む実験データに対するロジスティック曲線のあてはめの事例である。嵯山・大橋・高橋 (2008)、「薬理試験における統計解析の Q&A 用量依存性データへの非線形モデルの活用」で取り上げられた内臓痛試験データを取り上げる。内臓痛を緩和することが既知の薬剤 Alosetron の 3 用量に対し、Vehicle 群と Sham 群を含む 5 群で、1 群たり 8 匹のラット、全体で 40 匹の実験データである。

表 13.1 に内臓痛試験で得られたデータを示す．図 13.1 は，用量ごとの平均値に対する折れ線に対し Excel のスムージングの機能による平滑線を重ね書きした結果である．Vehicle 群と Sham 群は，マイナス無限大とプラスの無限大とすべきであるが，便宜的に

$$x_{\text{Vehicle}} = \log_{10}(0.003) = -2.52, \quad x_{\text{Sham}} = \log_{10}(300000) = 5.48$$

と設定している．

表 13.1 内臓痛試験データ

No	Vehicle	Alosetron ($\mu\text{g/kg}$)			Sham
		3	30	300	
1	30	30	40	35	55
2	20	20	45	45	40
3	25	20	30	35	45
4	20	35	40	40	50
5	25	30	35	50	60
6	30	25	25	45	45
7	25	30	40	50	45
8	20	25	45	40	50
平均	24.38	26.88	37.50	42.50	48.75

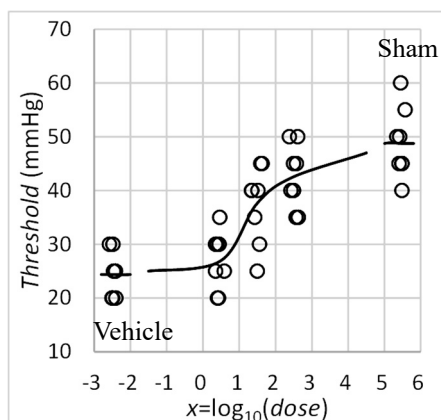


図 13.1 陰性・陽性対照を含むシグモイド状の反応

図 13.1 の散布図で，反応 *Threshold* が同値となる場合に左右に○印を左右にずらしている．これは， $x = \log_{10}(\text{dose})$ に対し，Excel の `Rand ()` 関数による一様乱数 (0~1) から 0.5 を差し引き，幅を調整するような係数 0.20 を掛けて変換

$$x' = x + (\text{Rand} () - 0.5) * 0.20 \tag{13.1}$$

した x' を用いて実現している．また，平均を通る平滑線は，Excel の「データ系列の書式設定」で「スムージング」を設定し，Vehicle の後の空白と Sham の前の空白は，該当する点のみを選択し「線なし」としている．

非線形モデルを用いたパラメータの同時推定

嵯山ら (2008) では，「3) 非線形モデルを用いた同時推定」で次のように述べている．

本実験の反応は，以下のようなモデル式で記述することができる．

$$y = \beta_1 \quad \dots (1)$$

$$y = \beta_1 + \frac{\beta_2 - \beta_1}{1 + \exp\{-\beta_3(x - \beta_4)\}} \quad \dots (2)$$

$$y = \beta_2 \quad \dots (3)$$

すなわち，(1) 式の y は反応の基点と考えられる陰性対照群での反応値で， $x = -\infty$ に相当し，(2) 式は用量反応関係を示すロジスティック曲線をあてはめる

モデル式 (x : 対数用量, y : 対数用量 x における反応値, β_1 : 最小反応値, β_2 : 最大反応値, β_3 : 曲線の勾配, β_4 : 最小と最大反応の和の $1/2$ となる反応を示す対数用量) であり, (3) 式の y は, 反応の終点と考えられる Sham 群での反応値で, $x=+\infty$ に相当する. 現実の実験データではこれに誤差が加わる. 本試験で知りたい情報は, これらのモデル式について同時推定を行うことにより求めた用量反応のパラメータ ($\beta_1 \sim \beta_4$) である.

この当時, Excel による残差線形化法 (ガウス・ニュートン法) によるパラメータの共分散行列を求める方法は, 模索中であり, 近似的な方法によりパラメータの標準誤差 SE を計算したために, JMP の非線形モデルで出力される標準誤差 SE を完全に再現することができなかった. 現在, Excel による残差線形化法によるパラメータの推定法も確立でき, JMP の非線形モデルで推定されたロジスティック曲線の 95%信頼区間の再現もできるようになった.

ダミー変数を用いた非線形の解析モデル

この当時は, β_3 を曲線の勾配を表すパラメータとして扱っていたのであるが, 得られた推定値の統計的な意味付けがしにくい. そのために, $\beta_3 = 1/\sigma_{LGS}$ のように勾配パラメータの逆数とした形状パラメータ σ_{LGS} で置き換え,

$$\beta_3(x - \beta_4) = \frac{1}{\sigma_{LGS}}(x - \beta_4) = \frac{x - \beta_4}{\sigma_{LGS}} \quad (13.2)$$

が得られる. さらに, β_4 を位置パラメータ μ_{LGS} で置き換え規準化

$$\beta_3(x - \beta_4) = \frac{x - \beta_4}{\sigma_{LGS}} = \frac{x - \mu_{LGS}}{\sigma_{LGS}} \quad (13.3)$$

する. 図 13.1 から位置パラメータ μ_{LGS} は, 50%反応量となり $\hat{\mu}_{LGS} = 1.5$, 形状パラメータ σ_{LGS} は, おおよそ 0.5 から 3.5 の範囲の $1/6$ の $\hat{\sigma}_{LGS} = 0.5$ として推定される. あらためて, この内臓痛試験データについて, JMP での解析を行うと共に Excel を用いて追試することにより, ブラック・ボックス的である JMP の「非線形回帰」の解析方法について理解が深められることを期待している.

Vehicle を陰性対照群, Shame を陽性対照群として, 表 13.2 に示すように (1, 1) 標示型ダミー変数を用い,

表 13.2 (1, 1) 標示型ダミー変数の設定

A	群	a_1	a_2	a_3
A ₁	Vehicle; 陰性対照群	1	0	0
A ₂	投与群	0	1	0
A ₃	Sham: 陽性対照群	0	0	1

β_1 を θ_{min} , β_2 を θ_{max} と置き換え, (1, 1) 標示型ダミー変数を用いた式

$$y_i = \theta_{min} a_1 + \left\{ \frac{\theta_{max} - \theta_{min}}{1 + \exp\left[-\left(\frac{x_i - \mu_{LGS}}{\sigma_{LGS}}\right)\right]} \right\} a_2 + \theta_{max} a_3 + \varepsilon_i, \quad \varepsilon_i \sim \text{正規分布} \quad (13.4)$$

を用いる.

JMP の「非線形回帰」

JMP の「非線形回帰」による推定結果を, 表 13.3 に示す. 崑山ら (2008) に示されている推定結果, および, 標準誤差 SE は, $\hat{\beta}_3 = 1/\hat{\sigma}_{LGS}$ を除いて一致する. 推定された $\hat{\theta}_{min} = 23.8358$ は, Vehicle 群の平均値 24.38 より小さい. これは, 投与群のデータに対しロジスティック曲線をあてはめようとする際に θ_{min} が, Vehicle 群の平均値 24.38 より小さ目となり, 双方を満足させるための推定値となっている. 同様に, 推定された $\hat{\theta}_{max} = 48.2225$ は, Shame 群の平均値 48.75 より小さいが, これも Vehicle 群の場合と同様の理由である.

位置パラメータの推定値 $\hat{\mu}_{LGS} = 1.4843$ は, $\hat{\theta}_{min}$ と $\hat{\theta}_{max}$ の平均値 36.0292 ($\hat{\theta}_{min}$ と $\hat{\theta}_{max}$ 間の 50% 点) に対応する逆推定値であり, いわゆる 50% 反応となる用量 ED_{50} である. $\hat{\sigma}_{LGS} = 0.6595$ は, ロジスティック分布の形状パラメータで, 正規分布の標準偏差 σ_{NOR} と同様の性質を持つが, 小さ目の推定値であり, $2\sigma_{NOR}$ と $3\hat{\sigma}_{LGS}$ が同程度と換算できる.

表 13.3 JMP の非線形回帰によるパラメータ推定

解			
	SSE	DFE	MSE
	1256.8502	36	34.9125
パラメータ	推定値	近似標準誤差	
θ_{min}	23.8358	2.0405	
θ_{max}	48.2225	2.0426	
μ_{LGS}	1.4843	0.2616	
σ_{LGS}	0.6595	0.2119	
解法: 解析 Gauss-Newton			

JMP のパラメータの 95% 信頼区間は, プロファイル尤度に基づくために割愛している.

表 13.4 に示すのは, 原著に図示されているロジスティック曲線の 95% 信頼区間を再現するために, 解析用の JMP データファイルの末尾に用量 (0.001, 0.003, ..., 1,000,000) を付与し, その常用対数 (-3.0, -2.52, ..., 6.0) に対する予測値および 95% 信頼区間の出力を Excel に取り込んで整えた結果である. 図 13.2 に示すのは, 元の散布図に推定されたロジスティック曲線の予測値と 95% 信頼区間を重ね書きしている. 各種の参照線は, Excel シート

表 13.4 JMP の非線形回帰で求めた 95%信頼区間

dose	x	y [^]	L 95%	U 95%
Vehicle	-3.00	23.8	19.7	28.0
0.003	-2.52	23.8	19.7	28.0
	-2.00	23.8	19.7	28.0
	-1.52	24.1	20.3	27.8
	-1.00	24.4	20.9	27.9
	-0.52	24.9	21.6	28.2
	0.00	26.2	22.8	29.5
3	0.48	28.2	24.6	31.8
	1.00	31.7	28.1	35.3
30	1.48	36.0	32.6	39.3
	2.00	40.6	36.9	44.2
300	2.48	43.8	40.2	47.4
	3.00	46.0	42.7	49.3
	3.48	47.1	43.8	50.4
	4.00	47.7	44.2	51.2
	4.48	48.0	44.2	51.7
Sham	5.00	48.2	44.1	52.4
300000	5.48	48.2	44.1	52.4
	6.00	48.2	44.1	52.4

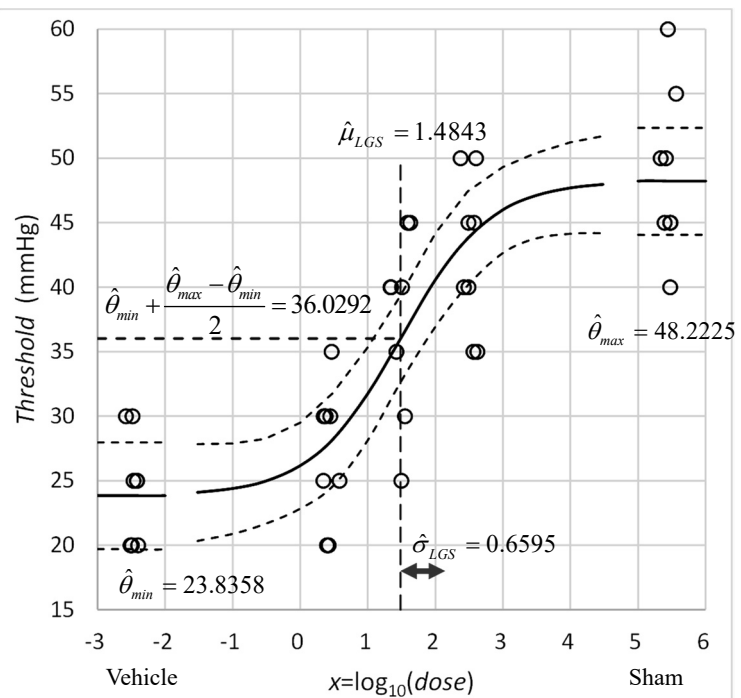


図 13.2 推定されたロジスティック曲線と 95%信頼区間

上に (始点, 終点) の座標を設定して重ね書きし, 散布図中の $\hat{\mu}_{LGS}$, $\hat{\sigma}_{LGS}$, $\hat{\theta}_{min}$ および $\hat{\theta}_{max}$ などの説明書きは, ワードの「テキストボックス」を挿入して書き込んでいる。

Vehicle 群は, 本質的にはマイナス無限大なのであるが, 便宜的に $x = -2.52$ に, 陽性対象は $x = 5.48$ 上に示してある. そのために, ロジスティック曲線を不連続として表示してある. このように, 連続した曲線を途中で不連続とするためには, Excel の散布図で不連続にした曲線の特定の位置のみを選択し, 「データ系列の書式設定」により「線なし」にすると左側の点の位置までが「線なし」となる. 各種の参照点なども Excel シート上に初めの位置と終わりの位置をセットし, 「データの選択」で「追加」して作図している. なお, 散布図上の点が重ならないように左右にずらしてあるが, 常用対数用量 x に対し, 式 (13.1) を適用し左右に振らした結果である.

Excel のソルバーを用いたパラメータ推定

Excel のソルバーを用いることにより, ロジスティック曲線のパラメータ推定が簡単にできることを示す. 表 13.1 の内臓痛のデータを表 13.5 に示すように, 行方向に展開し, 表 13.2 に示した (1, 1) 標示型ダミー変数 (a_1, a_2, a_3) を Vehicle 群, 投与群, Sham 群に与え, 常用対数用量 $x = \log_{10}(dose)$ の計算式を入力する. 推定式を \hat{y} とし, Excel の計算式で

$$\left. \begin{aligned}
 \hat{y} &= \hat{\theta}_{min} * a_1 \\
 &+ (\hat{\theta}_{min} + (\hat{\theta}_{max} - \hat{\theta}_{min}) / (1 + \exp(-(x - \hat{\mu}_{LGS}) / \hat{\sigma}_{LGS}))) * a_2 \\
 &+ \hat{\theta}_{max} * a_3
 \end{aligned} \right\} \quad (13.5)$$

を与える。残差平方和 S_e を

$$S_e = \text{SumSq}(\mathbf{y} \text{ の範囲} - \hat{\mathbf{y}} \text{ の範囲}), \quad S_e = \sum_i (y_i - \hat{y}_i)^2 \quad (13.6)$$

のようにベクトル「 \mathbf{y} の範囲」で与えることにより、残差の計算の列を Excel シート上に設定することなしに行列計算で直接求められる。

パラメータの欄に初期値を設定し、Excel のソルバーで「目的セルの設定」に残差平方和 S_e を、「目標値」を「最小値」にする。「変数セルの変更」でパラメータの範囲を設定し、「解決」により、ロジスティック曲線のパラメータ推定が瞬時に行なわれる。解は、初期値を設定したセルに得られるのであるが、表 13.5 では、得られた解を、右横にコピーし、あらためて初期値を設定し直した結果が表示されている。解の右側に JMP の結果を示すが、 $\hat{\theta}_{max}$ が少数点 4 桁目で異なるものの他は一致していることが確認できる。

表 13.5 Excel のソルバーによるパラメータ推定

		初期値	25			$\hat{\theta}_{min} =$	25.0000	解	23.8358
			50			$\hat{\theta}_{max} =$	50.0000		48.2224
			1.5			$\hat{\mu}_{LGS} =$	1.5000		1.4843
			0.5			$\hat{\sigma}_{LGS} =$	0.5000		0.6595
					残差平方和	$S_e =$	1385.0421		1256.8502
No	dose	a_1	a_2	a_3	x	y	\hat{y}		\hat{y}
1	Vehicle	1	0	0	0	30	25.0000		23.8358
2	Vehicle	1	0	0	0	20	25.0000		23.8358
:									
8	Vehicle	1	0	0	0	20	25.0000		23.8358
9	3	0	1	0	0.4771	30	27.8620		28.1863
:									
16	3	0	1	0	0.4771	25	27.8620		28.1863
17	30	0	1	0	1.4771	40	37.2141		35.9623
:									
24	30	0	1	0	1.4771	45	37.2141		35.9623
25	300	0	1	0	2.4771	35	46.8977		43.7930
:									
32	300	0	1	0	2.4771	40	46.8977		43.7930
33	Sham	0	0	1	0	55	50.0000		48.2224
:					0				
39	Sham	0	0	1		45	50.0000		48.2224
40	Sham	0	0	1	0	50	50.0000		48.2224
					$x = \text{if}((\text{IsNumber} = \text{dose}), \log_{10}(\text{dose}), 0)$				

推定値
23.8358
48.2225
1.4843
0.6595

表 13.3 の JMP の出力結果

Vehicle 群の推定値 $\hat{y}_{\text{Vehicle}} = 23.8358$ は、推定された最小値 $\hat{\theta}_{min} = 23.8358$ に一致し、Shame 群の推定値は $\hat{y}_{\text{Shame}} = 48.2224$ は、推定された最大値 $\hat{\theta}_{max} = 48.2224$ に一致している。これらの推定値は、表 13.1 で示したそれぞれの群平均 (24.38, 48.75) とは異なる。最小値 $\hat{\theta}_{min}$

= 23.8358 と最大値 $\hat{\theta}_{max} = 48.2224$ の推定値は、投与群のデータから推定される最小値および最大値も同時に考慮し、全体の残差平方和が最小になるように推定された結果となっている。位置パラメータ $\hat{\mu}_{LGS} = 1.4843$ は、図 13.1 に示したロジスティック曲線の最大値 $\hat{\theta}_{max}$ と最小値 $\hat{\theta}_{min}$ の平均値 36.0292 に対する逆推定値となっている。

Excel のソルバーは、表 13.5 に示したように、どのような推定式に対しても、残差平方和 S_e を最小化するようなパラメータの推定を瞬時に行ってくれる優れたものである。パラメータの標準誤差を求めるためには、非線形の式 (13.4) をパラメータによって偏微分した微係数行列を用いる必要がある。幸い SAS の NLIN プロシジャの「list」オプションにより表 13.9 に示すように偏微分式を出力することができるので、Excel による偏微分式の設定に際し助けになる。

SAS の NLIN プロシジャによる解析

最近、統計分析フリーソフト「R」が多くの人達によって使われるようになり、Web 上でも R によるロジスティック（シグモイド）曲線をあてはめてパラメータの推定値および標準誤差を求めている事例を数多く見いだすことができるようになったことは、喜ばしいことである。ただし、図 13.2 に示したように、推定されたロジスティック曲線に 95%信頼区間を重ね書きするような事例を見いだすことができない。私もチャレンジしてはいるが、いまだに満足する結果が得られない。

そこで、無償で継続的に使える OnDemand SAS の NLIN プロシジャを用いて、ロジスティック曲線の 95%信頼区間を得る方法を示す。表 13.4 の JMP で求めた 95%信頼区間の推定方法も、同様の方法である。SAS データファイルを作成する DATA ステップを次に示す。デー

SAS プログラム 13.1 /* SAS データファイルの作成 */

```
TITLE1 '内臓痛.SAS << ラットの内臓痛 >>' ;

DATA d01 ;
  input dose a1 a2 a3 @ ;
  x = log10(dose) ;
  do i=1 to 8 ;
    input y @ ; output ;
  end;
/*dose a1 a2 a3 1 2 3 4 5 6 7 8 */
datalines ;
0.003 1 0 0 30 20 25 20 25 30 25 20
3 0 1 0 30 20 20 35 30 25 30 25
30 0 1 0 40 45 30 40 35 25 40 45
300 0 1 0 35 45 35 40 50 45 50 40
300000 0 0 1 55 40 45 50 60 45 45 50
;
proc print data=d01 ; run;
```

タの並びは、ダミー変数をデータとして与えるために、表 13.1 に示した元のデータを Excel の「行と列の入れ替え」によって転置した並びにし、「タブ」を「スペース」に置き換え、整形している。表 13.6 に SAS の DATA ステップで作成されたデータファイルを PRINT プロシジャで出力した結果をコピーし、Excel シート上にペーストした結果を示す。全体で $5 \times 8 = 40$ 行となっている。

表 13.6 作成された SAS データファイル

OBS	dose	a1	a2	a3	x	i	y
1	0.003	1	0	0	-2.5229	1	30
2	0.003	1	0	0	-2.5229	2	20
:							
8	0.003	1	0	0	-2.5229	8	20
9	3	0	1	0	0.4771	1	30
:							
16	3	0	1	0	0.4771	8	25
17	30	0	1	0	1.4771	1	40
:							
24	30	0	1	0	1.4771	8	45
25	300	0	1	0	2.4771	1	35
:							
32	300	0	1	0	2.4771	8	40
33	300000	0	0	1	5.4771	1	55
:							
39	300000	0	0	1	5.4771	7	45
40	300000	0	0	1	5.4771	8	50

SAS データファイル data=d01 の作成に引き続き、滑らかなロジスティック曲線を描くための *dose* を読み込み、SAS データファイル data=d02 を作成し、DATA ステップの set ステートメントそれらを行方向に併合したファイルを data=d03 として作成する。

SAS プログラム 13.2 /* dose のみの SAS データファイルの作成 */

```

data d02 ;
  input a1 a2 a3 @ ;
  do i=1 to 7 ;
    input dose @ ;
    if dose = . then delete ;
    x = log10(dose) ;
    output ;
  end ;
datalines ;
1 0 0 0.001 0.003 0.01 . . .
0 1 0 0.03 0.1 0.3 1 3 10 .
0 1 0 30 100 300 1000 3000 10000 30000
0 0 1 100000 300000 1000000 . . .
;
data d03 ;
  set d01 d02 ;

```

```
proc print data=d03 ; run;
```

表 13.7 に行方向に併合された SAS データファイル d03 を示す。OBS=1~40 が d01 ファイルであり、OBS=41~59 が d02 ファイルのデータであり、行方向に連結されている。

表 13.7 行方向に併合された SAS データファイル d03

OBS	dose	a1	a2	a3	x	i	y
1	0.003	1	0	0	-2.5229	1	30
:							
40	300000	0	0	1	5.4771	8	50
41	0.001	1	0	0	-3	1	.
42	0.003	1	0	0	-2.5229	2	.
43	0.01	1	0	0	-2	3	.
44	0.03	0	1	0	-1.5229	1	.
:							
56	30000	0	1	0	4.4771	7	.
57	100000	0	0	1	5	1	.
58	300000	0	0	1	5.4771	2	.
59	1000000	0	0	1	6	3	.

このように反応 y が含まれない、解析に用いられない無駄なデータを結合するのは、最初に SAS の NLIN プロシジャで OBS=1~40 の反応 y を含むデータを対象にしたロジスティック曲線のパラメータ推定を行なわせる。次に、OBS=1~59 全体に対し、推定されたパラメータを用い、ロジスティック曲線の推定値、その標準誤差と 95%信頼区間を計算させ、別のデータファイルに出力させる。さらに、その出力ファイルを用いて SAS で作図することができる。ただし、手間の割には、きれいな図の作成ができないので、Excel に取り込み、表 13.4、および、図 13.2 で例示した Excel による作図を参考にしてもらいたい。

SAS プログラム 13.3 /* ロジスティック曲線 */

```
proc nlin data=d03 ;
  parms      theta_min=25  theta_max=50  mu_LGS=1.5  sigma_LGS=0.5;
  model y =  theta_min*a1
             + (theta_min + (theta_max - theta_min)/(1 + exp(- (x - mu_LGS)/sigma_LGS) ))*a2
             + theta_max*a3 ;
  output out=out1  predicted=y_hat  stdp=SE  L95M=L95  U95M=U95 ;
run ;
proc print data=out1 ; run ;
```

NLIN プロシジャに限らず SAS のプロシジャに共通する方法であるが、予測値および 95% 信頼区間などの計算結果を計算させるためには、output ステートメントの中で出力したい統計量に関するキーワードに引き続き出力ファイルでの変数名を

```
predicted = y_hat,  stdp = SE,  L95M = L95,  M95M = M95
```

のように設定する。計算結果は、out=out1 で指定された「out1」SAS ファイルに出力されるので、PRINT プロシジャで書き出し、出力結果をコピーして Excel シート上にペーストする。

若干の書式の設定した結果を、表 13.8 に示す。表 13.4 に示した JMP の非線形回帰によって得られた結果に一致する。

表 13.8 SAS の NLIN プロシジャによる予測値の 96%信頼区間の Excel への取り込み

OBS	dose	a1	a2	a3	x	i	y	y_hat	SE	L95	U95
1	0.003	1	0	0	-2.52	1	30	23.8	2.04	19.70	27.97
:											
40	300000	0	0	1	5.48	8	50	48.2	2.04	44.08	52.36
41	0.001	1	0	0	-3.00	1	.	23.8	2.04	19.70	27.97
42	0.003	1	0	0	-2.52	2	.	23.8	2.04	19.70	27.97
43	0.01	1	0	0	-2.00	3	.	23.8	2.04	19.70	27.97
44	0.03	0	1	0	-1.52	1	.	24.1	1.85	20.33	27.85
45	0.1	0	1	0	-1.00	2	.	24.4	1.73	20.88	27.90
46	0.3	0	1	0	-0.52	3	.	24.9	1.63	21.65	28.24
47	1	0	1	0	0.00	4	.	26.2	1.64	22.83	29.49
48	3	0	1	0	0.48	5	.	28.2	1.78	24.57	31.80
49	10	0	1	0	1.00	6	.	31.7	1.78	28.14	35.35
50	30	0	1	0	1.48	1	.	36.0	1.65	32.61	39.32
51	100	0	1	0	2.00	2	.	40.6	1.79	36.94	44.19
52	300	0	1	0	2.48	3	.	43.8	1.79	40.16	47.43
53	1000	0	1	0	3.00	4	.	46.0	1.64	42.67	49.32
54	3000	0	1	0	3.48	5	.	47.1	1.63	43.78	50.39
55	10000	0	1	0	4.00	6	.	47.7	1.74	44.16	51.23
56	30000	0	1	0	4.48	7	.	48.0	1.85	44.20	51.72
57	100000	0	0	1	5.00	1	.	48.2	2.04	44.08	52.36
58	300000	0	0	1	5.48	2	.	48.2	2.04	44.08	52.36
59	1000000	0	0	1	6.00	3	.	48.2	2.04	44.08	52.36

NLIN プロシジャによるパラメータに関する偏微分

Excel のソルバーを用いることにより、ロジスティック曲線のパラメータの推定値を手軽に推定できることを示してきた。ただし、推定されたパラメータの標準誤差を求めるために、ロジスティック曲線式をパラメータに関する偏微分式を必要とする。なかなか難儀ではあるが、偏微分式が得られれば、残差線形化の方法（ガウス・ニュートン法）により統一した手順で Excel によりパラメータの共分散行列 $\Sigma(\hat{\theta})$ を求め、対角要素の平方根により標準誤差を求め、一般的な方法で 95%信頼区間を求めることができる。

地道に微分公式を適用することにより、各種のロジスティック曲線のパラメータに関する偏微分式を得ることができるのだが、緻密で細かな配慮が必要であることから得られた偏微分式が正しいのか疑心暗鬼となる。正しいことの証明は、実データを用いた Excel による残差線形化法により、残差平方和が収束することで検証できる。得られた偏微分式に少しでも不具合があれば、残差平方和が発散してしまう。ただし、どこに不具合があるのか発見することは難儀であり、何らかの思い込みに阻まれるので、日を改めてプラス・マイナスなどの確認をするなどすると細かなミスを見い出すことができる。

パラメータに関する正しい偏微分式を得るために、JMP の自動偏微分の結果を参照したり、SAS の NLIN プロシジャによる偏微分式の出力を参考にしたりしてきた。ここでは、NLIN プロシジャの list オプションによる偏微分式の出力結果の見方を示す。その偏微分式を用いた Excel による残差線形化法によるパラメータの推定を行ない、微係数行列 \mathbf{Z} を用いたパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の計算方法を示す。さらに、 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の対角要素の平方根が、パラメータの標準誤差 SE となることを示す。

これまで、SAS の NLIN プロシジャの model ステートメントでのモデル式を一つの式で設定すると、得られる偏微分式が冗長となり活用しづらい。そこで、NLIN プロシジャで提供されているプログラミング機能を使い、モデル式を階層的に与えることにより偏微分式を簡潔化する方法を示す。式 (13.4) を次のように段階的に与え、

$$\left. \begin{aligned} g &= \exp\left(-\frac{x - \mu_{LGS}}{\sigma_{LGS}}\right) \\ y &= \theta_{min} a_1 + \left(\theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + g}\right) a_2 + \theta_{max} a_3 \end{aligned} \right\} \quad (13.7)$$

NLIN プロシジャの list オプションで表 13.9 に示すように偏微分式を出力させる。

```
SAS プログラム 13.4          /* ロジスティック曲線 */
proc nlin data=d03 list ;
  parms      theta_min=25  theta_max=50  mu_LGS=1.5  sigma_LGS=0.5;
            g = exp(-(x-mu_LGS)/sigma_LGS) ;
  model  y =  theta_min*a1
            + (theta_min + (theta_max - theta_min)/(1 + g)) * a2
            + theta_max*a3 ;
run;
```

表 13.9 NLIN プロシジャによる偏微分式の出力

Listing of Compiled Program Code		
Stmt	Line:Col	Statement as Parsed
	1 71:11:00	g = EXP(- (x - muLGS) / sigmaLGS);
	1 71:11:00	@g/@muLGS = g / sigmaLGS;
	1 71:11:00	@g/@sigmaLGS = - (- (x - muLGS) / sigmaLGS) / sigmaLGS * g;
	2 72:05:00	MODEL.y = theta_min * a1 + (theta_min + (theta_max - theta_min) / (1 + g)) * a2 + theta_max * a3;
	2 72:05:00	@MODEL.y/@theta_min = a1 + (1 + -1 / (1 + g)) * a2;
	2 72:05:00	@MODEL.y/@theta_max = a2 / (1 + g) + a3;
	2 72:05:00	@MODEL.y/@muLGS = - (@g/@muLGS * (theta_max - theta_min) / (1 + g)) / (1 + g) * a2;
	2 72:05:00	@MODEL.y/@sigmaLGS = - (@g/@sigmaLGS * (theta_max - theta_min) / (1 + g)) / (1 + g) * a2;

表 13.9 に示す偏微分式を Excel シート上で再現するためには、きちっとした数式で示すことが望ましいので、偏微分式の出力の順に数式化する。

$$\left. \begin{aligned} g &= \exp\left(-\frac{x - \mu_{LGS}}{\sigma_{LGS}}\right) \\ \frac{\partial g}{\partial \mu_{LGS}} &= \frac{1}{\sigma_{LGS}} \cdot g \\ \frac{\partial g}{\partial \sigma_{LGS}} &= \frac{x - \mu_{LGS}}{(\sigma_{LGS})^2} \cdot g \end{aligned} \right\} \quad (13.8)$$

$$\left. \begin{aligned} y &= \theta_{min} a_1 + \left(\theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + g}\right) a_2 + \theta_{max} a_3 \\ z_1 = \frac{\partial y}{\partial \theta_{min}} &= a_1 + \left(1 - \frac{1}{1 + g}\right) a_2 \\ z_2 = \frac{\partial y}{\partial \theta_{max}} &= \left(\frac{1}{1 + g}\right) a_2 + a_3 \end{aligned} \right\} \quad (13.9)$$

$$\left. \begin{aligned} z_3 = \frac{\partial y}{\partial \mu_{LGS}} &= -\frac{\partial g}{\partial \mu_{LGS}} \cdot \frac{\theta_{max} - \theta_{min}}{(1 + g)^2} \cdot a_2 = \frac{g}{\sigma_{LGS}} \cdot \frac{\theta_{max} - \theta_{min}}{(1 + g)^2} \cdot a_2 \\ z_4 = \frac{\partial y}{\partial \sigma_{LGS}} &= -\frac{\partial g}{\partial \sigma_{LGS}} \cdot \frac{\theta_{max} - \theta_{min}}{(1 + g)^2} \cdot a_2 = \frac{(x - \mu_{LGS})g}{(\sigma_{LGS})^2} \cdot \frac{\theta_{max} - \theta_{min}}{(1 + g)^2} \cdot a_2 \end{aligned} \right\} \quad (13.10)$$

残差線形化法によるロジスティック曲線のあてはめ

表 13.10 に示す Excel シートの \hat{y} 列までは、表 13.5 に示したソルバーを用いたパラメータの推定と同じ内容で、残差平方和 $S_e = 1256.85 = \text{SumSq}(y \text{ の範囲} - \hat{y} \text{ の範囲})$ が計算されている。次の残差 $\hat{\varepsilon} = y - \hat{y}$ 列を明示的に与えたのは、 $\hat{\delta}^{(m-1)}$ の計算で必要となるためであり、 \hat{g} 列は、式 (13.8) による計算結果であり、パラメータによる偏微分式による計算結果が順次示されている。表 13.10 右上段に (40行×4列) の微係数行列 \mathbf{Z} の積和行列 ($\mathbf{Z}^T \mathbf{Z}$) の逆行列 ($\mathbf{Z}^T \mathbf{Z}$)⁻¹ が示され、残差 $\hat{\varepsilon}_i$ を反応とする回帰分析の推定値 $\hat{\delta}^{(m-1)}$ ベクトル、 $\hat{\delta}^{(m-1)}$ を増分とした $\hat{\theta}^{(m)} = \hat{\theta}^{(m-1)} + \hat{\delta}^{(m-1)}$ も計算されている。 $\hat{\theta}^{(m-1)}$ に対する増分 $\hat{\delta}^{(m-1)}$ は、

$$\hat{\delta}^{(m-1)} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\varepsilon} \quad (13.11)$$

のように、残差のベクトル $\hat{\varepsilon}$ を反応とし、微係数行列 \mathbf{Z} をデザイン行列とみなした回帰分析による推定結果である。推定値は、 $\hat{\delta}^{(m-1)} = [0 \ 0 \ 0 \ 0]^T$ となっていて、 $\hat{\theta}^{(m-1)}$ と $\hat{\theta}^{(m)}$ が等しく、解が求まった状態が示されている。

すっきりと解を求めることができれば嬉しいのであるが、小さくなるはずの残差線形和 S_e が逆に大きくなることも、しばしば経験する。収束しない原因のほとんどは、Excel シート上に偏微分式を設定する際に式のマイナスを付け忘れていたりすることもあるし、入力した偏微分式にミスがある場合もあり、細心の注意が必要である。

表 13.10 残差線形化法による 4P ロジスティック曲線のあてはめ

				$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$	$(Z^T Z)^{-1}$						
初期値		25	$\theta^{min} =$	23.8358	0.0000	23.8358	0.1193	-0.0056	0.0079	-0.0049			
		50	$\theta^{max} =$	48.2225	0.0000	48.2225	-0.0056	0.1195	0.0079	0.0048			
		1.5	$\mu^{LGS} =$	1.4843	0.0000	1.4843	0.0079	0.0079	0.0020	0.0000			
		0.5	$\sigma^{LGS} =$	0.6595	0.0000	0.6595	-0.0049	0.0048	0.0000	0.0013			
残差平方和 $S_e =$				1256.85	----- 微係数行列 Z -----								
ダミー変数							残差						
No	dose	a_1	a_2	a_3	x	y	y^{\wedge}	$\varepsilon^{\wedge} = y - y^{\wedge}$	g^{\wedge}	z_1 $\partial \theta^{min}$	z_2 $\partial \theta^{max}$	z_3 $\partial \mu^{LGS}$	z_4 $\partial \sigma^{LGS}$
1	Vehicle	1	0	0		30	23.8358	6.1642	9.4944	1	0	0	0
2	Vehicle	1	0	0		20	23.8358	-3.8358	9.4944	1	0	0	0
:												0	
8	Vehicle	1	0	0		20	23.8358	-3.8358	9.4944	1	0	0	0
9	3	0	1	0	0.477	30	28.1863	1.8137	4.6055	0.8216	0.1784	-5.4198	8.2774
:													
16	3	0	1	0	0.477	25	28.1863	-3.1863	4.6055	0.8216	0.1784	-5.4198	8.2774
17	30	0	1	0	1.477	40	35.9623	4.0377	1.0110	0.5027	0.4973	-9.2440	0.1013
:													
24	30	0	1	0	1.477	45	35.9623	9.0377	1.0110	0.5027	0.4973	-9.2440	0.1013
25	300	0	1	0	2.477	35	43.7931	-8.7931	0.2219	0.1816	0.8184	-5.4964	-8.2738
:													
32	300	0	1	0	2.477	40	43.7931	-3.7931	0.2219	0.1816	0.8184	-5.4964	-8.2738
33	Sham	0	0	1		55	48.2225	6.7775	9.4944	0	1	0	0
:													
39	Sham	0	0	1		45	48.2225	-3.2225	9.4944	0	1	0	0
40	Sham	0	0	1		50	48.2225	1.7775	9.4944	0	1	0	0
		$x = \log_{10}(dose)$											

初期値=[25 50 1.5 0.5]^T を $\hat{\theta}^{(m-1)}$ にセットすると、表 13.11 に示す ($m=1$) の結果が得られる。増分 $\hat{\delta}^{(m-1)}$ は、式 (13.11) で計算された結果で、 $\hat{\delta}^{(m-1)}$ 加えた $\hat{\theta}^{(m)}$ が計算されている。これを $\hat{\theta}^{(m)}$ コピーし、 $\hat{\theta}^{(m-1)}$ に値のみをペーストすると自動的に ($m=2$) の計算が瞬時に得られる。これを繰り返し、増分 $\hat{\delta}^{(m-1)}$ が全て 0.0000 となるまで繰り返す。表 13.11 に示すように ($m=9$) 回目で、増分 $\hat{\delta}^{(m-1)}$ が全て 0.0000 となる。ただし、表示型式を指数とすると少数点以下 5 桁以下に何がしかの推定誤差が含まれていることが確認できる。

表 13.11 残差線形化法による繰り返し計算の過程

	$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$		$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$
$m=1$	25.0000	-1.4205	23.5795	$m=2$	23.5795	0.2802	23.8597
	50.0000	-1.9760	48.0240		48.0240	0.1419	48.1660
	1.5000	-0.0634	1.4366		1.4366	0.0434	1.4800
	0.5000	0.1478	0.6478		0.6478	0.0018	0.6496
$m=3, 4, 5, 6, 7,$							
	$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$		$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$
$m=8$	23.8357	0.0001	23.8358	$m=9$	23.8358	0.0000	23.8358
	48.2224	0.0001	48.2225		48.2225	0.0000	48.2225
	1.4843	0.0000	1.4843		1.4843	0.0000	1.4843
	0.6595	0.0000	0.6595		0.6595	0.0000	0.6595

初期値をセットし、残差平方和 S_e が最小になるようにソルバーで $\hat{\theta}^{(m-1)}$ を変化させると表 13.5 に示した結果が得られるが、増分 $\hat{\delta}^{(m-1)}$ が、全て 0.0000 とならないので、 $\hat{\theta}^{(m)}$ を $\hat{\theta}^{(m-1)}$ に値のみをペーストすると $\hat{\delta}^{(m-1)}$ が、全て 0.0000 となる。

残差 $\hat{\epsilon}$ を反応とした回帰分析を繰り返すことにより、 $m=9$ 回目でパラメータの推定が行なえたので、パラメータの推定値に対する標準偏差 SE を計算し、さらに、推定値 \hat{y} の分散の計算を行い 95%信頼区間の計算を行なう。

表 13.12 に示すようにパラメータの分散は、パラメータの共分散行列 $\Sigma(\hat{\theta}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$ の対角要素であり、標準誤差 SE は、分散の平方根で計算されていて、表 13.3 に示した JMP の結果に一致する。推定値 \hat{y} は、式 (13.4) によって求められ、 \hat{y} の分散は、パラメータの共分散行列 $\Sigma(\hat{\theta})$ に関する推定値 \hat{y} に対応する微係数ベクトルを \mathbf{z} としたときの 2 次形式

$$Var(\hat{y}) = \mathbf{z} \Sigma(\hat{\theta}) \mathbf{z}^T \tag{13.12}$$

によって計算される。分散が計算されれば、一般的な計算により、(L95%, U95%) が計算できる。この結果は、表 13.8 に示した SAS の NLIN プロシジャによる結果に一致する。もちろん表 13.4 に示した JMP の結果にも一致する。

表 13.12 残差線形化法による 4P ロジスティック曲線の 95%信頼区間の計算

		$\hat{\theta}^{(m-1)}$		$\Sigma(\hat{\theta}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$						SE	$L95\%$	$U95\%$
		$\hat{\theta}_{min}$	$\hat{\theta}_{max}$	μ_{LGS}	σ^2							
		23.8358	48.2225	1.4843	0.6595	4.163	-0.196	0.276	-0.169	2.040	19.70	27.97
						-0.196	4.172	0.276	0.168	2.043	44.08	52.37
						0.276	0.276	0.068	0.000	0.262	0.954	2.015
						-0.169	0.168	0.000	0.045	0.212	0.230	1.089
誤差平方和 $S_e =$		1256.85	36.966	—— 微係数ベクトル \mathbf{z} ——				$t_{0.05}(40-4) =$		2.028		
ダミー変数				z_1	z_2	z_3	z_4	分散	95%信頼区間			
$dose$	a_1 a_2 a_3	x	\hat{y}	\hat{g}	$\partial \hat{\theta}_{min}$	$\partial \hat{\theta}_{max}$	$\partial \mu_{LGS}$	$\partial \sigma_{LGS}$	$Var(\hat{y})$	SE	$L95\%$	$U95\%$
0.003	1 0 0	-2.523	23.84	435.3	1	0	0	0	4.163	2.040	19.70	27.97
	0 1 0	-1.000	24.39	43.25	0.977	0.023	-0.817	3.077	2.995	1.731	20.88	27.90
	0 1 0	0.000	26.16	9.494	0.905	0.095	-3.188	7.175	2.692	1.641	22.83	29.49
3	0 1 0	0.477	28.19	4.605	0.822	0.178	-5.420	8.277	3.178	1.783	24.57	31.80
	0 1 0	1.000	31.74	2.084	0.676	0.324	-8.102	5.950	3.155	1.776	28.14	35.35
30	0 1 0	1.477	35.96	1.011	0.503	0.497	-9.244	0.101	2.733	1.653	32.61	39.31
	0 1 0	2.000	40.57	0.458	0.314	0.686	-7.964	-6.227	3.198	1.788	36.94	44.19
300	0 1 0	2.477	43.79	0.222	0.182	0.818	-5.496	-8.274	3.210	1.792	40.16	47.43
	0 1 0	3.000	46.00	0.100	0.091	0.909	-3.067	-7.049	2.692	1.641	42.67	49.32
	0 1 0	4.000	47.70	0.022	0.022	0.978	-0.781	-2.977	3.039	1.743	44.16	51.23
300000	0 0 1	5.477	48.22	0.002	0	1	0	0	4.172	2.043	44.08	52.37

13.2. 平行でない直線のあてはめによる効力比の推定

これまでは、観察されたデータに対し、層別因子を考慮した回帰直線をあてはめ、各種の推定の問題を扱ってきた。本節で取り上げるのは、血圧を下げる薬物 S (標準薬) と新たな薬物 T (試験薬) の効果の比較するために動物を用いた実験データである。実験の目的は、ある一定量の血圧を下げるために必要となる薬物の投与量の比較である。動物に薬剤を少しずつ投与して、ある一定の血圧が得られるまで投与量を増やして行く直接法もあるが、ここでは、幾つかの投与量をあらかじめ設定する間接法による実験データを取り上げる。

血圧の上昇作用の抑制

表 13.13 は、ピース編, 中上・森川監訳 (1992), 「医薬統計学, 第 3 章 薬理学的活性の評価」に示されている「アンジオテンシン I 注入後の血圧上昇データ」である。アンジオテンシン I をラットの大腿静脈に注入すると、血圧上昇が起きる。降圧薬を前もって投与しておくことにより血圧の上昇を抑制する。したがって、血圧上昇が少ないほど降圧効果があると判断される。

表 13.13 アンジオテンシン I 注入後の血圧上昇 (mmHg)

	S 薬 dose (mg/kg)				T 薬 dose (mg/kg)				
	10	30	100	300	1	3	10	30	100
1	48	50	26	20	44	35	23	10	6
2	49	37	20	14	48	39	32	19	5
3	52	36	25	12	48	42	33	19	20
4	53	39	26	16	56	52	48	27	17
5	34	34	27	15	47	41	33	21	15
6	50	36	24	11	56	44	28	16	9
7	58	41	28	18					
8	48	40	25	16					
9	46	30	22	14					
10	56	40	23	13					
平均	49.4	38.3	24.6	14.9	49.8	42.2	32.8	18.7	12.0
(総平均, SD)	=(31.8, 14.0)								

用量別の比較

このような実験データの解析方法として思いつくのは、標準薬 S と試験薬 T の同一用量に対し、対応のない t 検定を行なうことであろう。その前に、用量を一致させた平均値とその 95% 信頼区間の線グラフを描くことも必要となる。図 13.3 に示すように、用量を一致させた標準薬 S と試験薬 T の平均値の 95% 信頼区間が重なり合うことがないので、 t 検定をするまでもなく明らかな差がある。

表 13.14 薬剤別の血圧上昇の基本統計量

薬剤	dose	n	平均	分散	$t_{0.05} \times SE$	L95%	U95%
S薬	10	10	49.40	43.38	4.97	44.43	54.37
	30	10	38.30	27.79	3.97	34.33	42.27
	100	10	24.60	5.82	1.82	22.78	26.42
	300	10	14.90	7.43	2.06	12.84	16.96
T薬	1	6	49.83	24.97	5.74	44.09	55.58
	3	6	42.17	32.57	6.56	35.61	48.73
	10	6	32.83	70.17	9.63	23.20	42.46
	30	6	18.67	38.40	7.12	11.54	25.79
	100	6	12.00	38.40	7.12	4.88	19.12
				$t_{0.05}(9)=$	2.262		
				$t_{0.05}(5)=$	2.571		

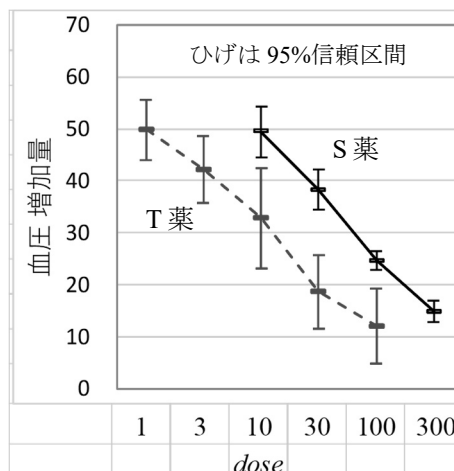


図 13.3 用量群間の比較

この実験の目的は、標準薬 S の用量反応関係に対する試験薬 T の用量反応関係の相対的な比較検討である。そのために、標準薬 S に対する 4 用量に対する血圧上昇が、対数用量に関して直線的であることを確認しつつ試験薬 T の用量反応関係を総合的に評価したい。図 13.3 から、試験薬 T の用量反応関係も対数用量に関して直線的であることが示されている。試験薬 T は少ない用量で標準薬 S と同様の血圧上昇を抑える作用を持つことが確認され、実験の目的は達成できたと解される。

同一の反応となる用量の比較

薬物の作用の評価として、既存の標準薬と同等の効果を発現するための用量の比較が一般的に行なわれている。これは、薬物の持つ多面的な薬理作用を比較検討するための方法であり、図 13.3 に血圧の増加量の抑制を 30 mmHg とするための標準薬 S の用量と試験薬 T の用量の比較を行いたい。そのためには、回帰直線をあてはめて、30 mmHg となる用量を逆推定する必要がある。

図 13.4 に示すのは、用量 *dose* を常用対数変換した変数 *x* に対して Excel による散布図に「近似曲線」を用いた回帰直線を重ね書きした結果である。標準薬 S および試験薬 T の回帰直線は、

$$\left. \begin{aligned} y^{(S)} &= 72.957 - 23.673x \\ y^{(T)} &= 50.735 - 19.816x \end{aligned} \right\} \quad (13.13)$$

が示されている。血圧の増加が 30 mmHg となる標準薬 S と試験薬 T の対数用量 $x_{y=30}^{(S)}$ と $x_{y=30}^{(T)}$ は、式 (13.13) の $y^{(S)}$ および $y^{(T)}$ に 30 を代入して *x* について解くと

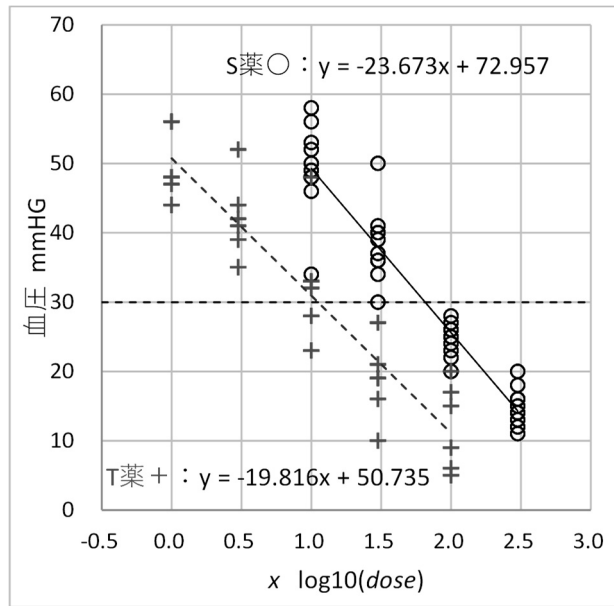


図 13.4 標準薬 S と試験薬 T の対数用量 x に対する用量反応直線

$$\left. \begin{aligned} \hat{x}_{y=30}^{(S)} &= \frac{30 - 72.957}{-23.673} = 1.815 \\ \hat{x}_{y=30}^{(T)} &= \frac{30 - 50.735}{-19.816} = 1.046 \end{aligned} \right\} \quad (13.14)$$

と逆推定され、元の用量 $dose$ に戻すと

$$\left. \begin{aligned} \widehat{dose}_{y=30}^{(S)} &= 10^{1.815} = 65.25 \\ \widehat{dose}_{y=30}^{(T)} &= 10^{1.046} = 11.13 \end{aligned} \right\} \quad (13.15)$$

と推定される。効力比 ρ は、標準薬 S の用量と等価となる ρ 倍の試験薬 T の用量

$$dose_{y=30}^{(S)} = \rho_{y=30} \cdot dose_{y=30}^{(T)} \quad (13.16)$$

として定義されているので、効力比は、

$$\hat{\rho}_{y=30} = \frac{\widehat{dose}_{y=30}^{(S)}}{\widehat{dose}_{y=30}^{(T)}} = \frac{65.25}{11.13} = 5.86 \text{ (倍)} \quad (13.17)$$

5.86 倍となる。同様に、血圧の増加が 20 mmHg の場合は、表 13.15 に示ように 4.85 倍、40 mmHg で 7.09 倍となる。

表 13.15 標準薬 S に対する試験薬 T の効力比

	S 薬			T 薬		
回帰係数	$\beta^{\wedge}_0 =$	72.9570		50.7345		
	$\beta^{\wedge}_1 =$	-23.6730		-19.8158		
		対数用量	$dose$	対数用量	$dose$	効力比
	血圧 y_0	x^{\wedge}_0	$10^{x^{\wedge}_0}$	x^{\wedge}_0	$10^{x^{\wedge}_0}$	ρ^{\wedge}
逆推定	20	2.2370	172.59	1.5510	35.56	4.85
	30	1.8146	65.25	1.0464	11.13	5.86
	40	1.3922	24.67	0.5417	3.48	7.09
		$x^{\wedge}_0 = (y_0 - \beta^{\wedge}_0) / \beta^{\wedge}_1$				

効力比 ρ の 95%信頼区間

推定された効力比 $\hat{\rho}$ の 95%信頼区間を求めたい. 効力比 ρ の常用対数は, 式 (13.17) から

$$\left. \begin{aligned} \log_{10} \hat{\rho}_{y_0} &= \log_{10} \widehat{dose}_{y_0}^{(S)} - \log_{10} \widehat{dose}_{y_0}^{(T)} \\ \text{ただし, } \log_{10} \widehat{dose}_{y_0}^{(S)} &= \hat{x}_{y_0}^{(S)} = \frac{y_0 - \hat{\beta}_0^{(S)}}{\hat{\beta}_1^{(S)}} \\ \log_{10} \widehat{dose}_{y_0}^{(T)} &= \hat{x}_{y_0}^{(T)} = \frac{y_0 - \hat{\beta}_0^{(T)}}{\hat{\beta}_1^{(T)}} \end{aligned} \right\} \quad (13.18)$$

であり, $\log_{10} \widehat{dose}_{y_0}^{(S)}$ と $\log_{10} \widehat{dose}_{y_0}^{(T)}$ は互いに独立である. したがって, $\log_{10} \hat{\rho}_{y_0}$ の分散は, $Var(\log_{10} \widehat{dose}_{y_0}^{(S)})$ および $Var(\log_{10} \widehat{dose}_{y_0}^{(T)})$ の和として求めることができる. それぞれの分散は, 逆推定値 $\hat{x}_{y_0}^{(S)}$ および $\hat{x}_{y_0}^{(T)}$ なので, 回帰パラメータで偏微分し, パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式 (デルタ法) で求めることができる.

$$\left. \begin{aligned} \hat{x}_{y_0}^{(S)} &= \frac{y_0 - \hat{\beta}_0^{(S)}}{\hat{\beta}_1^{(S)}}, & \hat{x}_{y_0}^{(T)} &= \frac{y_0 - \hat{\beta}_0^{(T)}}{\hat{\beta}_1^{(T)}} \\ d_0^{(S)} &= \frac{\partial \hat{x}_{y_0}^{(S)}}{\partial \hat{\beta}_0^{(S)}} = -\frac{1}{\hat{\beta}_1^{(S)}}, & d_0^{(T)} &= \frac{\partial \hat{x}_{y_0}^{(T)}}{\partial \hat{\beta}_0^{(T)}} = -\frac{1}{\hat{\beta}_1^{(T)}} \\ d_1^{(S)} &= \frac{\partial \hat{x}_{y_0}^{(S)}}{\partial \hat{\beta}_1^{(S)}} = -\frac{y_0 - \hat{\beta}_0^{(S)}}{(\hat{\beta}_1^{(S)})^2}, & d_1^{(T)} &= \frac{\partial \hat{x}_{y_0}^{(T)}}{\partial \hat{\beta}_1^{(T)}} = -\frac{y_0 - \hat{\beta}_0^{(T)}}{(\hat{\beta}_1^{(T)})^2} \end{aligned} \right\} \quad (13.19)$$

効力比の 95%信頼区間を求めるためには, パラメータの共分散行列 $\Sigma(\hat{\beta})$ が必要となる. 標準薬 S および試験薬 T について別々に求めることも可能であるが, 2 つの直線のパラメータの共分散が 0 であることを示すために (1, 1) 標示型ダミー変数 (a_1, a_2) を用いて一括して推定することにする. 表 13.16 左に示すように変数名を (a_1, a_1x, a_2, a_2x) とし, 標準薬 S の場合に a_1 および $a_1x = a_1 \log_{10}(dose)$, 試験薬 T の場合に a_2 および $a_2x = a_2 \log_{10}(dose)$ を与える. ただし, 試験薬 T の場合に (a_1, a_1x) は共に 0, 標準薬 S の場合に (a_2, a_2x) は共に 0 となる (70 行 \times 4 列) のデザイン行列 X をセットする.

Excel の回帰分析を起動し, 「定数に 0 を使用」をオンとし, 「入力 X の範囲」に (70 行 \times 4 列) のデザイン行列 X 全体を設定する. 表 13.16 右上段に回帰分析の分散分析表とパラメータの推定値を示す. 推定された標準薬 S と試験薬 T のパラメータは, 図 13.4 に示されている回帰式に一致する. 表 13.16 右下段は, Excel の行列関数を用いて計算したパラメータの共分散行列 $\Sigma(\hat{\beta})$ であり, 標準薬 S の場合は左上の (2 \times 2), 試験薬 T の場合は右下の (2 \times 2) を用いる. $\Sigma(\hat{\beta})$ の対角要素は, パラメータの分散であり, 平方根を取るとパラメータの SE となる. ちなみに 1 行 1 列目は, $\sqrt{7.5846} = 2.7540$ と標準薬 S の切片 $\hat{\beta}_0^{(S)}$ の SE に一致することが確かめられる.

表 13.16 標準薬 S および試験薬 T の回帰直線のパラメータの共分散行列 $\Sigma(\hat{\beta})$ の推定

No.	drug	dose	デザイン行列 X					分散分析表 (定数に 0 を使用) on					
			a_1	a_{1x}	a_2	a_{2x}	y	自由度	変動	分散	分散比		
1	S	10	1	1.00	0	0.00	48	回帰	4	82237.27	20559.32	735.16	
:								残差	66	1845.73	27.9657	$=\sigma^2$	
11	S	30	1	1.48	0	0.00	50	合計	70	84083.00			
:													
21	S	100	1	2.00	0	0.00	26		係数	標準誤差	t	P-値	
:								切片	0.0000	#N/A	#N/A	#N/A	
31	S	300	1	2.48	0	0.00	20	$\hat{\beta}_0^{(S)}$	a_1	72.9570	2.7540	26.4911	0.0000
:								$\hat{\beta}_1^{(S)}$	a_{1x}	-23.6730	1.5093	-15.6847	0.0000
41	T	1	0	0.00	1	0.00	44	$\hat{\beta}_0^{(T)}$	a_2	50.7345	1.6620	30.5267	0.0000
:								$\hat{\beta}_1^{(T)}$	a_{2x}	-19.8158	1.3653	-14.5144	0.0000
47	T	3	0	0.00	1	0.48	35						
:													
53	T	10	0	0.00	1	1.00	23	パラメータの共分散行列 $\Sigma(\hat{\beta}) = (X^T X)^{-1} \sigma^2$					
:								$\hat{\beta}_0^{(S)}$		7.5846	-3.9604	0	0
59	T	30	0	0.00	1	1.48	10	$\hat{\beta}_1^{(S)}$		-3.9604	2.2780	0	0
:								$\hat{\beta}_0^{(T)}$		0	0	2.7621	-1.8469
65	T	100	0	0.00	1	2.00	6	$\hat{\beta}_1^{(T)}$		0	0	-1.8469	1.8639
:													
70	T	100	0	0.00	1	2.00	9						
			$x = \log_{10}(\text{dose})$										

表 13.17 に示すのは、標準薬 S および試験薬 T の逆推定値の分散と SE の計算結果であり、標準薬 S の $y_0^{(S)} = 20\text{mmHg}$ に対する逆推定 $\hat{x}_0^{(S,20)} = 2.2370$ の偏微分値 $d_0^{(S,20)}$ と $d_1^{(S,20)}$ は、

$$\left. \begin{aligned} d_0^{(S,20)} &= -\frac{1}{\hat{\beta}_1^{(S)}} = -\frac{1}{-23.6730} = 0.0422 \\ d_1^{(S,20)} &= -\frac{y_0^{(S,20)} - \hat{\beta}_0^{(S)}}{(\hat{\beta}_1^{(S)})^2} = -\frac{20 - 72.9570}{(-23.6730)^2} = 0.0945 \end{aligned} \right\} \quad (13.20)$$

と計算されている。

表 13.17 標準薬 S および試験薬 T の逆推定値の分散と SE

	y_0	$x_0^{(S)}$	$d_0^{(S)}$	$d_1^{(S)}$	分散	SE
S薬	20	2.2370	0.0422	0.0945	0.0023	0.0475
	30	1.8146	0.0422	0.0767	0.0013	0.0357
	40	1.3922	0.0422	0.0588	0.0017	0.0417
	y_0	$x_0^{(T)}$	$d_0^{(T)}$	$d_1^{(T)}$	分散	SE
T薬	20	1.5510	0.0505	0.0783	0.0039	0.0622
	30	1.0464	0.0505	0.0528	0.0024	0.0489
	40	0.5417	0.0505	0.0273	0.0033	0.0577
分散 $Ver(x^\wedge) = d \Sigma(\hat{\beta}) d^T$						

この偏微分値をベクトル $d^{(S,20)} = [0.0422 \quad 0.0945]$ としたときに、 $\hat{x}_0^{(S,20)}$ の分散は、標準薬 S についてのパラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する $d^{(S,20)} = [d_0^{(S,20)} \quad d_1^{(S,20)}]$ の 2 次形式を用い

$\hat{d}^{(S,20)}$		S薬: $\Sigma(\hat{\beta})$		$\hat{d}^{(S,20)T}$	$Var[x\hat{\alpha}_0^{(S,20)}]$
0.0422	0.0945	7.5846	-3.9604	0.0422	= 0.0023
		-3.9604	2.2780	0.0945	

として計算されている。標準薬 S の 30 mmHg, 40 mmHg については、式 (13.20) の $y_0^{(S,20)}$ に (30, 40) に代入して計算されている。試験薬 T についても同様に Excel シート上で計算した結果である。

表 13.17 で求められた標準薬 S の逆推定値と T 薬の逆推定値の差

$$\left. \begin{aligned} \log_{10} \hat{\rho}_{y_0^{(20)}} &= \hat{x}_{y_0}^{(S,20)} - \hat{x}_{y_0}^{(T,20)} \\ &= 2.2370 - 1.5510 = 0.6860 \end{aligned} \right\} \quad (13.21)$$

を求める。差の分散については、分散の加法性が成り立つので、表 13.17 の分散を用いて

$$\left. \begin{aligned} Var(\hat{x}_{y_0}^{(S,20)} - \hat{x}_{y_0}^{(T,20)}) &= Var(\hat{x}_{y_0}^{(S,20)}) + Var(\hat{x}_{y_0}^{(T,20)}) \\ &= 0.0023 + 0.0039 = 0.0061 \end{aligned} \right\} \quad (13.22)$$

が得られ、95%信頼区間が (0.5299, 0.8422) として計算されている。効力比 ρ は、

$$\hat{\rho}_{y_0^{(20)}} = 10^{0.6860} = 4.85 \quad (13.23)$$

となり、95%信頼区間も同様に 10 の冪乗を取り (3.39, 6.95) となる。同様に 30mmHg の場合は、5.86 (4.44, 7.74)、40mmHg の場合は、7.09 (5.11, 9.83) が計算されている。

表 13.18 効力比の 95%信頼区間の推定

y_0	$x\hat{\alpha}_0^{(S)} - x\hat{\alpha}_0^{(T)}$	分散	SE	L95%	U95%
20	0.6860	0.0061	0.0782	0.5298	0.8422
30	0.7682	0.0037	0.0605	0.6475	0.8890
40	0.8505	0.0051	0.0712	0.7083	0.9926
y_0	ρ^{\wedge}			L95%	U95%
20	4.85	—	—	3.39	6.95
30	5.86	—	—	4.44	7.74
40	7.09	—	—	5.11	9.83
			$t_{0.05(70-4)}=$	1.9966	

効力比を求める反応を設定する際に、図 13.4 から切の良い (20, 30, 40) を選択したのであるが、実験に先立ち設定する必要もある。その際には、標準薬 S の反応 y の全体の平均と標準偏差 SD を用いて (平均 - SD , 平均, 平均 + SD) とあらかじめ設定しておく必要性もある。その際には、表 13.13 に示されている y の総平均 $\hat{\mu}=31.8$, $SD=14.0$ から ($y_0^{\mu-SD} = 17.8$, $y_0^{\mu} = 31.8$, $y_0^{\mu+SD} = 45.8$) における効力比の算出を行なうことになる。

13.3. 平行な直線のあてはめによる効力比の推定

平行でない直線のあてはめでは、用量によって効力比が異なるので、平行な直線をあてはめることにより、どの用量でも均一となる効力比を求めたい。

切片が異なる平行な直線のあてはめ

標準薬 S と試験薬 T で切片が異なるが傾き β_1 は共通とした平行な直線は、標準薬 S を 0、試験薬 T を 1 とする (0, 1) 型ダミー変数 a_2 を用い、 $x = \log_{10}(\text{dose})$ としたときに、

$$y_i = \beta_0^{(S)} + \beta_0^{(T-S)} a_2 + \beta_1 x_i + \varepsilon_i \quad (13.24)$$

により推定できる。パラメータ $\beta_0^{(T-S)}$ は、標準薬 S の切片と試験薬の切片との差分 ($\beta_0^{(T)} - \beta_0^{(S)}$) となる。効力比 ρ は、標準薬 S の用量 $\text{dose}_{y=y_0}^{(S)}$ と同じ反応 y_0 が得られる試験薬 T の用量 $\text{dose}_{y=y_0}^{(T)}$ とした場合に、 $\text{dose}_{y=y_0}^{(S)} = \rho \cdot \text{dose}_{y=y_0}^{(T)}$ のように定義されている。試験薬 T が、標準薬 S に比べて少ない用量で同じ反応となれば、 $\rho > 1.0$ となり、逆に同じ反応を得るために用量を増やさなければならない場合には、 $\rho < 1.0$ となる。

用量 dose の設定は、(1, 3, 10, 30, ...)、(1, 10, 100, 1000, ...) のように等比で与え常用対数 $x = \log_{10}(\text{dose})$ に対し、反応 y が平行な直線となることを期待している。式 (13.24) では、Y 軸方向の差が $\beta_0^{(T-S)}$ となるので、X 軸方向の差を $\rho' = \log_{10} \rho$ としたときに、

$$\left. \begin{aligned} \rho &= \frac{\text{dose}_{y=y_0}^{(S)}}{\text{dose}_{y=y_0}^{(T)}} \\ \log_{10} \rho = \rho' &= x_{y_0}^{(S)} - x_{y_0}^{(T)} = \frac{\beta_0^{(T-S)}}{\beta_1} \end{aligned} \right\} \quad (13.25)$$

との関係が導き出される。したがって、効力比 ρ は、

$$\rho = \frac{\text{dose}_{y=y_0}^{(S)}}{\text{dose}_{y=y_0}^{(T)}} = \frac{10^{x_{y_0}^{(S)}}}{10^{x_{y_0}^{(T)}}} = \frac{10^{x_{y_0}^{(S)}}}{10^{x_{y_0}^{(S)} - \rho'}} = 10^{\rho'} \quad (13.26)$$

として求めることができる。

対数効力比 ρ' の分散は、 $\beta_0^{(T-S)}$ と β_1 の比なので、 ρ' を式 (13.24) のパラメータで偏微分し、

$$\left. \begin{aligned} \rho' &= \frac{\beta_0^{(T-S)}}{\beta_1} \\ d_0 &= \frac{\partial \rho'}{\partial \beta_0^{(S)}} = 0 \\ d_1 &= \frac{\partial \rho'}{\partial \beta_0^{(T-S)}} = \frac{1}{\beta_1} \\ d_2 &= \frac{\partial \rho'}{\partial \beta_1} = -\frac{\beta_0^{(T-S)}}{\beta_1^2} \end{aligned} \right\} \quad (13.27)$$

得られた偏微分値を $\mathbf{d}=[d_0 \ d_1 \ d_2]$ ベクトルにまとめ、パラメータの共分散行列 $\Sigma(\hat{\beta})$ を挟んだ 2 次形式 $Var(\rho')=\mathbf{d}\Sigma(\hat{\beta})\mathbf{d}^T$ で対数効力比 ρ' の分散を求めることができる。

表 13.19 に示すのは、式 (13.24) に対応した (70 行×3 列) のデザイン行列 \mathbf{X} を用いた Excel の回帰分析による結果である。推定されたパラメータは、 $\hat{\beta}_0^{(T-S)}=-16.8144$ 、 $\hat{\beta}_1=-21.5516$ なので、対数効力比 $\rho'=(-16.8144)/(-21.5516)=0.7802$ 、効力比 $\rho=10^{0.7802}=6.03$ が得られる。

表 13.19 切片が異なる平行な直線のあてはめ

No.	drug	dose	デザイン行列 \mathbf{X}				分散分析表 (定数に 0 を使用) off					
			x_0	a_2	x	y		自由度	変動	分散	分散比	
1	S	10	1	0	1.00	48	回帰	2	12679.31	6339.66	218.25	
2	S	10	1	0	1.00	49	残差	67	1946.19	29.0476		
:							合計	69	14625.50			
40	S	300	1	0	2.48	13						
41	T	1	1	1	0.00	44		係数	標準誤差	t 値	p 値	
42	T	1	1	1	0.00	48	$\hat{\beta}_0^{(S)}=$ 切片 x_0	69.2688	1.9861	34.8768	0.0000	
:							$\hat{\beta}_0^{(T-S)}=$ a_2	-16.8144	1.5132	-11.1119	0.0000	
70	T	100	1	1	2.00	9	$\hat{\beta}_1=$ x	-21.5516	1.0319	-20.8857	0.0000	
			$x=\log_{10}(\text{dose})$									

効力比 $\rho=6.03$ の 95%信頼区間を求めるために、対数効力比 $\rho'=0.7802$ の 95%信頼区間を先に求め、それらから $(10^{L_{95\%}(\rho')}, 10^{U_{95\%}(\rho')})$ を計算する。そのためには、表 13.20 に示すように、デザイン行列 \mathbf{X} と誤差分散の推定値 $\hat{\sigma}^2=29.0476$ を用いたパラメータの共分散行列 $\Sigma(\hat{\beta})=(\mathbf{X}^T\mathbf{X})^{-1}\hat{\sigma}^2$ を必要とする。

表 13.20 パラメータの共分散行列 $\Sigma(\hat{\beta})=(\mathbf{X}^T\mathbf{X})^{-1}\hat{\sigma}^2$

$(\mathbf{X}^T\mathbf{X})$				$(\mathbf{X}^T\mathbf{X})^{-1}$				$\hat{\sigma}^2$
70.0	30.0	99.2679		0.1358	-0.0727	-0.0637	*	29.0476
30.0	30.0	29.7255		-0.0727	0.0788	0.0274		
99.2679	29.7255	177.6374		-0.0637	0.0274	0.0367		
			=	3.9446	-2.1103	-1.8512		$\beta_0^{(S)\wedge}$
				-2.1103	2.2897	0.7962		$\beta_0^{(T-S)\wedge}$
				-1.8512	0.7962	1.0648		β_1^\wedge
				パラメータの共分散行列 $\Sigma(\beta^\wedge)$				

次に、式 (13.27) に示したパラメータに関する偏微分式を用い

$$d_0=0, \quad d_1=\frac{1}{\hat{\beta}_1}=-0.0464, \quad d_2=-\frac{\hat{\beta}_0^{(T-S)}}{\hat{\beta}_1^2}=0.0362 \quad (13.28)$$

を求める。これらをベクトル $\mathbf{d}=[d_0 \ d_1 \ d_2]$ としてまとめ、表 13.21 に示すように、パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式により、対数効力比 ρ' の分散 $Var(\rho')=0.0037$ を求め、その平方根で $SE(\rho')=\sqrt{0.0037}=0.0604$ が得られる。

表 13.21 2次形式による対数効力比 ρ' の分散の推定

d_0	d_1	d_2	パラメータの共分散行列 $\Sigma(\beta^{\wedge})$			d^T	$Var(\rho')$	$SE(\rho')$
0.0000	-0.0464	0.0362	3.9446	-2.1103	-1.8512	0.0000	0.0037	0.0604
			-2.1103	2.2897	0.7962	-0.0464		
			-1.8512	0.7962	1.0648	0.0362		

対数効力比 $\rho' = 0.7802$ の 95%信頼区間は、自由度 $(70 - 3)$ の両側 5%の $t_{0.05}(67) = 1.9960$ から、

$$L95\% \rho' = 0.7802 - 1.9960 \times 0.0604 = 0.6596$$

$$U95\% \rho' = 0.7802 + 1.9960 \times 0.0604 = 0.9008$$

が得られる。効力比 $\rho = 10^{0.7802} = 6.03$ の 95%信頼区間は、

$$\left. \begin{aligned} L95\% \rho &= 10^{0.6596} = 4.57 \\ U95\% \rho &= 10^{0.9008} = 7.96 \end{aligned} \right\} \quad (13.29)$$

となる。図 13.5 に、これらの結果を示す。

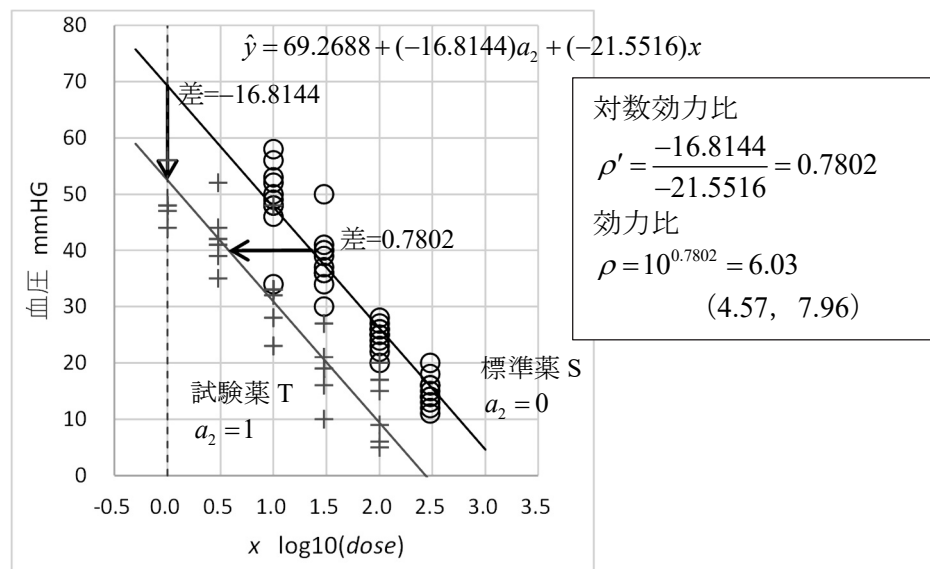


図 13.5 対数効力比 ρ' の算出方法

SAS の NLIN プロシジャによる効力比の直接推定

無償で継続的に提供されている OnDemand SAS の NLIN プロシジャを用いることにより、効力比とその 95%信頼区間を直接推定することができる。同様に JMP の「非線形のあてはめ」、R の nls() 関数によっても同様の結果が得られる。

平行な回帰直線のあてはめは、一般的に式 (13.24) を用いて、対数効力比 $\rho' = 0.7802$ から効力比 $\rho = 10^{0.7802} = 6.03$ は、簡単に求まるのであるが、その 95%信頼区間の算出に際しては、面倒な行列計算を必要とした。式 (13.24) の $\beta_0^{(T-S)}$ に代えて、式 (13.25) から導出される $\beta_0^{(T-S)}$

= $\beta_1 \rho'$ を代入し整理し、対数効力比 ρ' をパラメータとした推定式

$$\left. \begin{aligned} y &= \beta_0^{(S)} + \beta_1 \rho' a_2 + \beta_1 x \\ &= \beta_0^{(S)} + \beta_1 (x + \rho' a_2) \end{aligned} \right\} \quad (13.30)$$

が得られる。標準薬 S の場合は、 $a_2 = 0$ なので $y = \beta_0^{(S)} + \beta_1 x$ となり、試験薬 T の場合は、 $a_2 = 1$ であり、 $y = \beta_0^{(S)} + \beta_1 (x + \rho')$ のように対数用量 x に代えて $(x + \rho')$ のように試験薬 T すべての対数用量 x に ρ' を加えて回帰直線を推定することになる。実際の指定に際し、試験薬 T の回帰直線が、標準薬 S の回帰直線に重なるように X 軸方向にオフセットすることになる。

式 (13.30) は、推定したいパラメータ β_1 と ρ' の積が含まれパラメータに関して線形ではないので、非線形回帰による解析が必要となる。効力比のみの推定であれば、式 (13.24) の回帰式が手軽であるが、煩雑な効力比の分散を行わなければならない。式 (13.30) による非線形回帰を適用することにより、対数効力比 ρ' の標準誤差 SE および 95%信頼区間を直接推定することができる。

SAS で解析を行うためには、前もって解析用の SAS データファイルを作成しておく必要がある。ここで示すのは、表 13.13 のデータを Excel で転置し、スペース区切りのテキストデータとして、SAS の DATA ステップの中で、解析用の SAS データファイルを作成している。

SAS プログラム 13.5 /* SAS データファイルの作成 */

```

TITLE1 'アンジオテンシン.SAS << アンジオテンシンによる血圧上昇 >>' ;
DATA d01 ;
  input D$ a2 dose @ ;
  x = log10(dose) ;
  do i=1 to 10 ;
    input y @ ; if y=. then delete ; output ;
  end;

/* D a2 dose 1 2 3 4 5 6 7 8 9 10 */
datalines ;
S 0 10 48 49 52 53 34 50 58 48 46 56
S 0 30 50 37 36 39 34 36 41 40 30 40
S 0 100 26 20 25 26 27 24 28 25 22 23
S 0 300 20 14 12 16 15 11 18 16 14 13
T 1 1 44 48 48 56 47 56 . . . .
T 1 3 35 39 42 52 41 44 . . . .
T 1 10 23 32 33 48 33 28 . . . .
T 1 30 10 19 19 27 21 16 . . . .
T 1 100 6 5 20 17 15 9 . . . .
;
proc print data=d01 ; run;

```

作成された SAS データファイル “d01” は、表 13.22 に示すように PRINT プロシジャで 70 行が出力される。

表 13.22 解析用 SAS データファイル “d01”

OBS	D	a2	dose	x	i	y
1	S	0	10	1.0000	1	48
:						
11	S	0	30	1.4771	1	50
:						
21	S	0	100	2.0000	1	26
:						
31	S	0	300	2.4771	1	20
:						
40	S	0	300	2.4771	10	13
41	T	1	1	0.0000	1	44
:						
47	T	1	3	0.4771	1	35
:						
53	T	1	10	1.0000	1	23
:						
59	T	1	30	1.4771	1	10
:						
65	T	1	100	2.0000	1	6
:						
70	T	1	100	2.0000	6	9

式 (13.30) を用いて NLIN プロシジャによる解析を行う。式 (13.24) の場合、GLM プロシジャならば、`model y = a2 x ;` のように変数名の指定だけであるが、NLIN プロシジャの場合は、`parms` ステートメントでパラメータの初期値を設定し、パラメータを含めたモデル式にする必要がある。

SAS プログラム 13.6 /* 対数効力比の直接推定 */
 /* X 軸方向の移動 */

```
proc nlin data=d01 ;
  parms      β0=70  β1=-20  log10ρ=1 ;
  model y = β0 + β1 * (x + (log10ρ)*a2) ;
run ;
```

SAS/NLIN プロシジャの結果を表 13.23 に示す。ここで得られた対数効力比 ρ' (SAS の計算式で使えない記号が含まれているので $\log_{10}\rho$ としている) ように効力比は、10 の冪乗で

表 13.23 対数効力比 ($\log_{10}\rho$) の直接推定

要因	自由度	平方和	平均平方	F 値	近似 Pr > F
Model	2	12679.3	6339.7	218.25	<.0001
Error	67	1946.2	29.0476		
Corrected Total	69	14625.5			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		
β0	69.2688	1.9861	65.3045	73.2331	
β1	-21.5516	1.0319	-23.6113	-19.4920	
log10ρ	0.7802	0.0604	0.6596	0.9008	

$\rho = 10^{0.7802} = 6.03$ となり、95%信頼区間は、 $(10^{0.6596}, 10^{0.9008}) = (4.57, 7.96)$ と Excel の表計算で求めている。この推定値は、式 (13.29) の結果に一致する。

表 13.24 Excel による効力比 ρ と 95%信頼区間の計算

パラメータ	推定値	近似標準誤差	近似 95% 信頼限界	
log10 ρ	0.7802	0.0604	0.6596	0.9008
$\rho = 10^{\log_{10}\rho} =$	6.03		4.57	7.96

平行でない回帰直線 vs. 平行な回帰直線

平行な直線をあてはめて効力比を求める解析方法は「共分散分析」と類似しているが、対数用量 x は、実験者が能動的に設定している。したがって、 x は共変量ではなく実験計画法的には、制御因子と呼ぶべき変数である。ただ、どのような場合でも平行な直線をあてはめる際に、交互作用がないことの確認が形式的に求められる。この事例で交互作用は、幸い $p = 0.0624$ と有意な差ではないが、微妙な結果ではある。交互作用が有意な差であった場合に、用量ごとに効力比を出すことは、表 13.15 に示したように容易であるが、それらの使い方・表現方法などについては、今後の課題である。

原田 (2017) は、「平行線検定を利用した薬物の効力比較」で、標準薬 S と試験薬 T 薬の傾きの差が同一の場合であっても、ばらつきが小さい場合に交互作用が有意な差となり、ばらつきが大きい場合には有意な差とならないことを指摘し、「交互作用が有意だから平行線をあてはめることができない」などの短絡的な判断をいさめている。それに代えて、標準薬 S の推定される傾きに対し、T 薬との傾きの差について「同等と見なせる傾きの差の範囲」をあらかじめ設定し、傾きの差の 95% が、同等と見なせる範囲に入っていれば、交互作用があったとしても平行線のあてはめる手順を提言している。この内容は、原田・吉池 (2018) によって「平行線検定 (直線及びシグモイド曲線) による効力比較」によって詳しく解説されている。

なお、平行な回帰直線をあてはめて効力比を算出する方法は、「平行線検定法」として Finney (1978), 「Statistical Method in Biological Assay 3rd ed., Chap. 4 Parallel line assay」で偏差平方和を用いて定式化されているが、煩雑なのでダミー変数を用いた回帰分析を適用する方法が簡潔で見通しが良い。佐久間著, 五所ら編 (2017), 「新版 薬効評価, 第 7 章 計量的反応」に、偏差平方和を用いた平行線検定法が、「7.1 平行線検定法の原理, 7.2 平行線検定法の計算, 7.3 対称計画, 7.4 精度係数と観測例数, 7.5 変則的な平行線検定法」に詳しく示されているので、参考にされたい。

13.4. 平行なロジスティック曲線による効力比の推定

直線のあてはめは便宜的

表 13.13 に示したように、降圧剤を前もって投与した後に、アンジオテンシン I を規定用量投与し血圧を上昇させているので、降圧剤を投与しない場合に血圧の上昇には、上限が存在する。降圧剤の用量を増やすことにより、血圧の上昇は抑えられ、降圧剤の用量をさらに増やした場合に、血圧の上昇は、0 に漸近する。

図 13.3 に示した試験薬 T の平均値の折れ線グラフから、ごくわずかにロジスティック曲線らしき形状が浮かび上がっている。標準薬 S については、試験薬 T ほどではないがロジスティック曲線らしき形状となっている。そのために、共通の最大値および最小値を持ち、標準薬 S の X 軸上の位置パラメータを基準として、試験薬 T で左方向に移動するが、ロジスティック曲線の形状は同じであると仮定する。単調減少のロジスティック曲線なので、生存関数 $S_{LGS}(x) = 1 - F_{LGS}(x)$ を使うことになり、式を整理すると指数の中のマイナスがプラスに代わった式となる。

$$F_{LGS}(x) = \frac{1}{1 + \exp\left[-\frac{(x - \mu_{LGS})}{\sigma_{LGS}}\right]}, \quad S_{LGS}(x) = 1 - F_{LGS}(x) = \frac{1}{1 + \exp\left[\frac{(x - \mu_{LGS})}{\sigma_{LGS}}\right]} \quad (13.31)$$

ロジスティック曲線による効力比の直接推定

最大値を θ_{max} 、最小値を θ_{min} とする生存関数型のロジスティック曲線 $S_{LGS}(x)$ に拡張する。薬剤は (0, 1) 型ダミー変数として a_2 を用い、そのパラメータを ρ' とした場合に、

$$S_{LGS}(x) = \theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + \exp\left[\frac{(x + \rho' a_2 - \mu_{LGS})}{\sigma_{LGS}}\right]} \quad (13.32)$$

により、対数効力比 ρ' を直接推定することができる。

表 13.25 左にロジスティック曲線のパラメータ推定のための Excel シートを示す。推定値の計算式の設定のために、前もっておおよそのパラメータの初期値を設定する必要がある。最大値 θ_{max} は、得られたデータの最大値が増加量の 58 mmHg なので、区切りよく $\hat{\theta}_{max} = 60$ とする。最小値 θ_{min} は、増加量が 0 mmHg と設定したいのであるが、少し下駄をはかせて $\hat{\theta}_{min} = 1$ とする。位置パラメータ μ_{LGS} は、標準薬 S にロジスティック曲線をあてはめた場合の変曲点（中心位置）の x なので、おおよそ 1.75 となるが、少しずらして $\hat{\mu}_{LGS} = 1$ とする。形状パラメータ σ_{LGS} は、正規分布の標準偏差の半分弱であり、対数用量 x の 0~3 の 6 分の 1 として

$\hat{\sigma}_{LGS} = 0.5$ ぐらいと推測されるが、切のよい $\hat{\sigma}_{LGS} = 1$ と設定する. X 軸方向の標準薬 S と試験薬 T の差 ρ' は、標準薬 S を試験薬 T に合わせるためのオフセットなので、おおよそ $\hat{\rho}' = 1$ とする.

表 13.25 左は、初期値に対して式 (13.32) による推定値 \hat{y} の計算, $\text{SumSq}(y - \hat{y})$ による残差平方和 $S_e = 14372.5$ が計算されている. 表 13.25 右は、ロジスティック曲線を描くため、用量の設定範囲を広げにし、式 (13.32) を標準薬 S と試験薬 T に別々に適用して推定した結果である. 図 13.6 左に初期値に対する 2 本のロジスティック曲線を描いた結果を示す.

表 13.25 Excel ソルバーによる初期値に対するロジスティック曲線のパラメータ推定

初期値:		1.00	$\theta_{min}^{\wedge} =$	1.0000	$\hat{y} = \theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + \exp\left[\frac{(x + \rho' a_2 - \mu_{LGS})}{\sigma_{LGS}}\right]}$							
		60.00	$\theta_{max}^{\wedge} =$	60.0000								
		1.00	$\mu_{LGS}^{\wedge} =$	1.0000								
		1.00	$\rho'^{\wedge} =$	1.0000								
		1.00	$\sigma_{LGS}^{\wedge} =$	1.0000								
$(y - \hat{y})^2$ 残差平方和 $S_e =$					14372.51			S 薬		T 薬		
No.	drug	dose	x	a_2	y	\hat{y}	dose	x	a_2	$\hat{y}^{(S)}$	a_2	$\hat{y}^{(T)}$
1	S	10	1.00	0	48	30.500	0.01	-2.00	0	57.202	1	52.967
2	S	10	1.00	0	49	30.500	0.10	-1.00	0	52.967	1	44.132
:							1	0.00	0	44.132	1	30.500
40	S	300	2.48	0	13	11.966	10	1.00	0	30.500	1	16.868
41	T	1	0.00	1	44	30.500	50	1.70	0	20.590	1	10.121
42	T	1	0.00	1	48	30.500	100	2.00	0	16.868	1	8.033
:							1000	3.00	0	8.033	1	3.798
70	T	100	2.00	1	9	8.033	10000	4.00	0	3.798	1	2.061
$S_e = \text{Sumsq}(y \text{ の範囲} - \hat{y} \text{ の範囲})$							100000	5.00	0	2.061	1	1.395

表 13.26 Excel ソルバーによるロジスティック曲線のパラメータ推定

初期値:		1.00	$\theta_{min}^{\wedge} =$	7.3674	$\hat{y} = \theta_{min} + \frac{\theta_{max} - \theta_{min}}{1 + \exp\left[\frac{(x + \rho' a_2 - \mu_{LGS})}{\sigma_{LGS}}\right]}$							
		60.00	$\theta_{max}^{\wedge} =$	55.0173								
		1.00	$\mu_{LGS}^{\wedge} =$	1.7623								
		1.00	$\rho'^{\wedge} =$	0.7675								
		1.00	$\sigma_{LGS}^{\wedge} =$	0.4337								
$(y - \hat{y})^2$ 残差平方和 $S_e =$					1813.86			S 薬		T 薬		
No.	drug	dose	x	a_2	y	\hat{y}	dose	x	a_2	$\hat{y}^{(S)}$	a_2	$\hat{y}^{(T)}$
1	S	10	1.00	0	48	48.009	0.01	-2.00	0	55.009	1	54.970
2	S	10	1.00	0	49	48.009	0.10	-1.00	0	54.936	1	54.543
:							1	0.00	0	54.212	1	50.651
40	S	300	2.48	0	13	15.055	10	1.00	0	48.009	1	31.050
41	T	1	0.00	1	44	50.651	50	1.70	0	32.928	1	15.215
42	T	1	0.00	1	48	50.651	100	2.00	0	24.821	1	11.640
:							1000	3.00	0	9.963	1	7.831
70	T	100	2.00	1	9	11.640	10000	4.00	0	7.640	1	7.414
$S_e = \text{Sumsq}(y \text{ の範囲} - \hat{y} \text{ の範囲})$							100000	5.00	0	7.395	1	7.372

表 13.26 に示すのは、表 13.25 に対し、Excel のソルバーにより残差平方和 S_e を、最小にするようにパラメータ $[\hat{\theta}_{min}, \hat{\theta}_{max}, \hat{\mu}_{LGS}, \hat{\sigma}_{LGS}, \hat{\rho}']$ を変更させた結果である。誤差平方和は、 $S_e = 1813.86$ となっている。

表 13.19 および表 13.23 に示した平行な直線をあてはめた場合の残差平方和 S_e は、1946.2 であるので、パラメータ数が 3 から 5 に増えたとしてもその残差平方和の減少は、 $1946.2 - 1813.86 = 132.34$ と大きく、ロジスティック曲線のあてはめが適していると判断される。図 13.6 左に示すのがパラメータに初期値を与えた場合で、図 13.6 右に示すのが、パラメータをソルバーで変更させ残差平方和 S_e を最小化した平行なロジスティック曲線である。

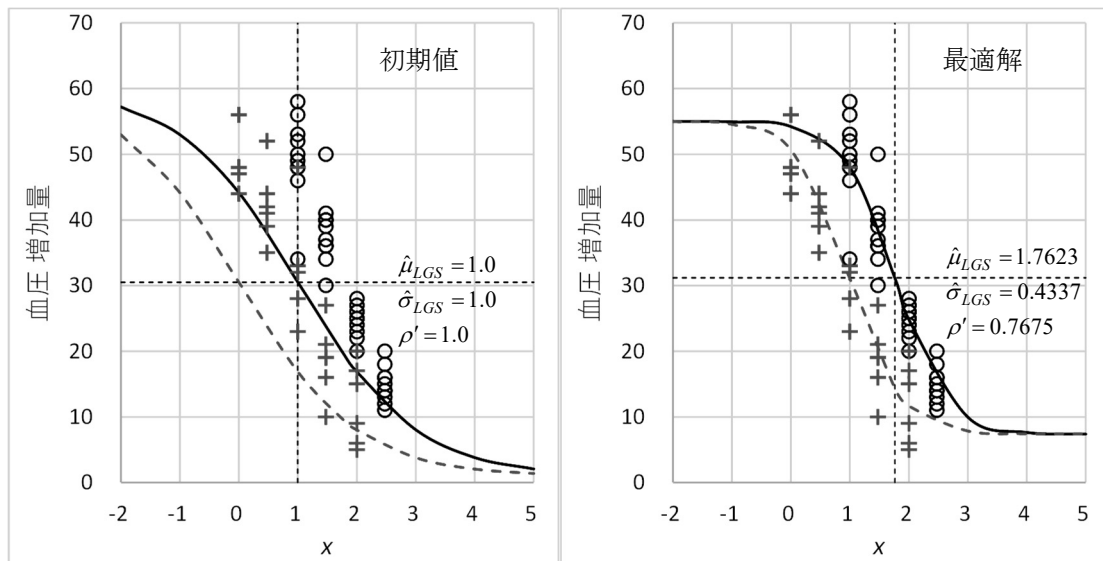


図 13.6 平行なロジスティック曲のあてはめ

図 13.4 に示した 2 本の直線をあてはめた結果を、平行なロジスティック曲線をあてはめた図 13.6 右と対比して見ると、標準薬 S および試験薬 T どちらの場合でも、各用量のデータの平均値に近づく方向に移動し、図 13.3 の平均値を通る折れ線グラフに近い形状が得られている。

NLIN プロシジャによるパラメータの標準誤差と 95%信頼区間

さて、X 軸方向の差 $\hat{\rho}' = 0.7675$ の 95%信頼区間を Excel で求めたいのであるが、式 (13.32) をパラメータで偏微分した式を用いて計算した微係数行列 \mathbf{Z} から $\mathbf{\Sigma}(\hat{\theta}) = (\mathbf{Z}^T \mathbf{Z})^{-1}$ を計算し、その対角要素から $\hat{\rho}'$ の分散 $\text{Var}(\hat{\rho}')$ を求めればよいのであるが、言うが易しである。ここでも、無償で継続的に使える OnDemand SAS の NLIN プロシジャを使った事例を示しつつ、NLIN プロシジャの list オプションで出力されるパラメータで偏微分した式の見方も合わせて示す。SAS プログラム 13.7 に示すのは、SAS プログラム 13.5 で作成された SAS データファ

イル「data=d01」を用いて、表 13.25 に示したパラメータを初期値として設定した 2 本の平行なロジスティック曲線をあてはめ検出力 ρ' を直接求めるための SAS プログラムである。

SAS プログラム 13.7 /* ロジスティック曲線 効力比 */

```
proc nlin data=d01 list ;
  parms       $\theta_{max}=60$   $\theta_{min}=1$   $\mu_{LGS}=1$   $\sigma_{LGS}=1$   $\log_{10}\rho=10$  ;
  g = exp( (x + ( $\log_{10}\rho$ )*a2 -  $\mu_{LGS}$ ) /  $\sigma_{LGS}$ ) ;
  model y =  $\theta_{min} + (\theta_{max}-\theta_{min})/(1 + g)$  ;
run ;
```

表 13.27 に示すのは、NLIN プロシジャの実行結果である。分散分析表の誤差平方和 $S_e = 1813.9$ は、表 13.26 に示した Excel のソルバーで求めた 1813.86 に一致し、引き続き出力されているパラメータの推定値も一致している。X 軸方向の 2 本のロジスティック曲線の X 軸方向の間隔は、 $\log_{10}\rho = 0.7675$ となっていて、 $SE = 0.0663$ 、95%信頼区間が (0.6350, 0.8999) として求められている。

表 13.27 NLIN プロシジャによるによる平行なロジスティック曲線のあてはめ

要因	自由度	平方和	平均平方	F 値	近似Pr > F
Model	4	12811.6	3202.9	114.78	<.0001
Error	65	1813.9	27.9055		
Corrected Total	69	14625.5			
パラメータ	推定値	近似標準誤差	近似 95% 信頼限界		
θ_{min}	7.3667	5.2326	-3.0836	17.817	
θ_{max}	55.0179	5.2038	44.6252	65.4106	
μ_{LGS}	1.7623	0.1013	1.5599	1.9646	
$\log_{10}\rho$	0.7675	0.0663	0.635	0.8999	
σ_{LGS}	0.4337	0.137	0.16	0.7074	

表 13.27 に示したように、対数効力比 ρ' のままでは、評価しにくいので表 13.28 に示すように Excel に取り込んだ結果に対し、10 の冪乗として効力比 ρ および 95%信頼区間を計算する。試験薬 T は標準薬 S に対し 5.85 倍の効力があり、その 95%信頼区間は (4.32, 7.94) となる。

表 13.28 Excel により計算した効力比と 95%信頼区間

パラメータ	推定値	近似標準誤差	近似 95% 信頼限界	
$\log_{10}\rho$	0.7675	0.0663	0.635	0.8999
$\rho=10^{\log_{10}\rho}$	5.85		4.32	7.94
	効力比		効力比の95%信頼区間	

Excel による効力比の 95%信頼区間の推定

実務的には、表 13.28 に示すようにロジスティック曲線をあてはめた場合の効力比の推定と 95%信頼区間を正確に求めることが、解析の最終的な目的である。ただし、対数効力比 ρ' がどのようにして求められたのか、また、その 95%信頼区間を求めるための標準誤差 SE の算出方法については、統計ソフトまかせでブラック・ボックスのままである。第 10 章では、最小値が異なる 3 本のロジスティック曲線の 95%信頼区間の推定について Excel を主体にした解析方法を示し、JMP と SAS による検証結果を示したので、同様の手順で Excel による効力比 ρ の 95%信頼区間の推定方法を示す。

偏微分式の簡単化のために式 (13.32) の分母の指数項を $g(x)$ とした NLIN プロシジャから出力されるパラメータに関する偏微分式の出力結果を次に示す。

表 13.29 NLIN プロシジャによる偏微分式の出力

Listing of Compiled Program Code		
Stmt	Line:Col	Statement as Parsed
1	72:12:00	$g = \text{EXP}((x + \log_{10}\rho * a_2 - \mu_{LGS}) / \sigma_{LGS});$
1	72:12:00	$@g/@\mu_{LGS} = -1 / \sigma_{LGS} * g;$
1	72:12:00	$@g/@\log_{10}\rho = a_2 / \sigma_{LGS} * g;$
1	72:12:00	$@g/@\sigma_{LGS} = -((x + \log_{10}\rho * a_2 - \mu_{LGS}) / \sigma_{LGS}) / \sigma_{LGS} * g;$
2	73:05:00	$\text{MODEL.y} = \theta_{\min} + (\theta_{\max} - \theta_{\min}) / (1 + g);$
2	73:05:00	$@\text{MODEL.y}/@\theta_{\min} = 1 - 1 / (1 + g);$
2	73:05:00	$@\text{MODEL.y}/@\theta_{\max} = 1 / (1 + g);$
2	73:05:00	$@\text{MODEL.y}/@\mu_{LGS} = -(@g/@\mu_{LGS} * (\theta_{\max} - \theta_{\min}) / (1 + g)) / (1 + g);$
2	73:05:00	$@\text{MODEL.y}/@\log_{10}\rho = -(@g/@\log_{10}\rho * (\theta_{\max} - \theta_{\min}) / (1 + g)) / (1 + g);$
2	73:05:00	$@\text{MODEL.y}/@\sigma_{LGS} = -(@g/@\sigma_{LGS} * (\theta_{\max} - \theta_{\min}) / (1 + g)) / (1 + g);$

$$\left. \begin{aligned}
 g &= \exp\left(\frac{x + \rho' a_2 - \mu_{LGS}}{\sigma_{LGS}}\right) \\
 \frac{\partial g}{\partial \sigma_{LGS}} &= -\frac{1}{\sigma_{LGS}} \cdot g \\
 \frac{\partial g}{\partial \rho'} &= \frac{a_2}{\sigma_{LGS}} \cdot g \\
 \frac{\partial g}{\partial \sigma_{LGS}} &= -\frac{x + \rho' a_2 - \mu_{LGS}}{(\sigma_{LGS})^2} \cdot g
 \end{aligned} \right\} \quad (13.33)$$

$$\left. \begin{aligned}
 y &= \theta_{\min} + \frac{\theta_{\max} - \theta_{\min}}{1 + g} \\
 z_1 &= \frac{\partial y}{\partial \theta_{\min}} = 1 - \frac{1}{1 + g} \\
 z_2 &= \frac{\partial y}{\partial \theta_{\max}} = \frac{1}{1 + g}
 \end{aligned} \right\} \quad (13.34)$$

$$\left. \begin{aligned} z_3 &= \frac{\partial y}{\partial \mu_{LGS}} = -\frac{\partial g}{\partial \sigma_{LGS}} \cdot \frac{\theta_{max} - \theta_{min}}{(1+g)^2} \\ z_4 &= \frac{\partial y}{\partial \rho'} = -\frac{\partial g}{\partial \rho'} \cdot \frac{\theta_{max} - \theta_{min}}{(1+g)^2} \\ z_5 &= \frac{\partial y}{\partial \sigma_{LGS}} = -\frac{\partial g}{\partial \sigma_{LGS}} \cdot \frac{\theta_{max} - \theta_{min}}{(1+g)^2} \end{aligned} \right\} \quad (13.35)$$

表 13.30 に示すように、表 13.26 に示した Excel ソルバーによるパラメータ推定欄の右側に式 (13.33) から式 (13.35) に示したパラメータの偏微分式を順次入力し、微係数行列 \mathbf{Z} を完成させる。この微係数行列 \mathbf{Z} を線形モデルのデザイン行列 \mathbf{X} とみなし、残差ベクトル $\hat{\boldsymbol{\varepsilon}}$ を反応ベクトル \mathbf{y} とみなし、パラメータ $\hat{\boldsymbol{\delta}}$ を残差による回帰分析

$$\hat{\boldsymbol{\delta}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\boldsymbol{\varepsilon}} \quad (13.36)$$

により推定する。表 13.30 に示された残差 $\hat{\varepsilon}_i$ 、微係数行列 \mathbf{Z} を用いて $(\mathbf{Z}^T \mathbf{Z})^{-1}$ が表の上部に、パラメータ $\hat{\boldsymbol{\delta}}^{(m-1)}$ の推定結果が中ほどに示されている。左側の $\hat{\boldsymbol{\theta}}^{(m-1)}$ は、表 13.26 のソルバーで求めた推定値であり、その推定値を用いて微係数行列 \mathbf{Z} が計算され、その \mathbf{Z} から $\hat{\boldsymbol{\delta}}^{(m-1)}$ が推定されている。パラメータ $\hat{\boldsymbol{\delta}}^{(m-1)}$ の値が全て 0 となれば、表示されている小数点 4 桁まで正確に推定されていることになるが、 $\hat{\theta}_{min}$ と $\hat{\theta}_{max}$ は (0.0009, -0.0009) と 0.0000 ではない。

表 13.30 Excel による残差線形化法による効力比 ρ と 95%信頼区間

		θ	$\theta^{(m-1)}$	$\delta^{(m-1)}$	$\theta^{(m)}$	$(\mathbf{Z}^T \mathbf{Z})^{-1}$							
初期値:	1.00	$\theta^{min} =$	7.3674	-0.0009	7.3665	0.9810	-0.6986	-0.0065	0.0003	-0.0231			
	60.00	$\theta^{max} =$	55.0173	0.0009	55.0182	-0.6986	0.9702	-0.0065	0.0004	0.0231			
	1.00	$\mu^{LGS} =$	1.7623	0.0000	1.7623	-0.0065	-0.0065	0.0004	0.0000	0.0000			
	1.00	$\rho^{log10} =$	0.7675	0.0000	0.7675	0.0003	0.0004	0.0000	0.0002	0.0000			
	1.00	$\sigma^{LGS} =$	0.4337	0.0000	0.4337	-0.0231	0.0231	0.0000	0.0000	0.0007			
$(y - \hat{y})^2$ 誤差平方和 $S_e =$			1813.86			-----微係数行列 \mathbf{Z} -----							
No.	drug	dose	x	a_2	y	\hat{y}	$\hat{\varepsilon}$	\hat{g}	$\partial \theta^{min}$	$\partial \theta^{max}$	$\partial \mu^{LGS}$	$\partial \rho^{log10}$	$\partial \sigma^{LGS}$
1	S	10	1.00	0	48	48.009	-0.009	0.1724	0.1471	0.8529	13.7834	0.000	-24.226
2	S	10	1.00	0	49	48.009	0.991	0.1724	0.1471	0.8529	13.7834	0.000	-24.226
:													
40	S	300	2.48	0	13	15.055	-2.055	5.1981	0.8387	0.1613	14.8665	0.000	24.504
41	T	1	0.00	1	44	50.651	-6.651	0.1009	0.0916	0.9084	9.1456	-9.146	-20.978
42	T	1	0.00	1	48	50.651	-2.651	0.1009	0.0916	0.9084	9.1456	-9.146	-20.978
:													
70	T	100	2.00	1	9	11.640	-2.640	10.1532	0.9103	0.0897	8.9679	-8.968	20.786

表 13.26 に示した Excel のソルバーで求めたパラメータの推定値に対し、表 13.27 の NLIN プロシジャの結果は、小数点 4 桁目で異なる結果が示されていた。推定結果の有効数字をどこまで求めるかを変更することができるようになっているが、解析の目的によって使い分けが必要がある。

表 13.30 に示されている $\hat{\theta}^{(m-1)}$ に $\hat{\delta}^{(m-1)}$ を加えた $\hat{\theta}^{(m)}$ の結果を $\hat{\theta}^{(m-1)}$ に値のみをペーストすると、改めて $\hat{\delta}^{(m-1)}$ が計算されるようになっている。表 13.31 に示すように 2 回ほどコピー&ペーストを繰り返すと $\hat{\delta}^{(m-1)}$ が全て 0.0000 となる。もちろん、こ少数点 5 桁以下には、なにがしかの数値が存在している。Excel での結果は、表 13.27 の NLIN プロシジャの結果と比較すると、 θ_{max} で 0.0002 の差異があるが、NLIN プロシジャの収束の判定基準を変更することにより、Excel での計算結果に一致する。

表 13.31 残差線形化法によるパラメータの推定精度の改善

θ	$\hat{\theta}^{(m-1)}$	$\hat{\delta}^{(m-1)}$	$\hat{\theta}^{(m)}$	$\hat{\theta}^{(m-1)}$	$\hat{\delta}^{(m-1)}$	$\hat{\theta}^{(m)}$	推定値
$\theta_{min}^{\wedge} =$	7.3674	-0.0009	7.3665	7.3667	0.0000	7.3666	7.3667
$\theta_{max}^{\wedge} =$	55.0173	0.0009	55.0182	55.0181	0.0000	55.0181	55.0179
$\mu_{LGS}^{\wedge} =$	1.7623	0.0000	1.7623	1.7623	0.0000	1.7623	1.7623
$\rho_{\log 10}^{\wedge} =$	0.7675	0.0000	0.7675	0.7675	0.0000	0.7675	0.7675
$\sigma_{LGS}^{\wedge} =$	0.4337	0.0000	0.4337	0.4337	0.0000	0.4337	0.4337

少数定以下 3 から 4 桁まで一致 NLIN プロシジャ

厄介な微係数行列 \mathbf{Z} を Excel で作成し、残差線形化法により非線形モデルのパラメータ推定を行う目的は、求めたい対数効力比 ρ' の 95%信頼区間を計算するための SE を自ら計算できる力を付けてもらいと願うからである。表 13.32 に示したのは、表 13.30 で示した $(\mathbf{Z}^T \mathbf{Z})^{-1}$ に誤差分散 $\hat{\sigma}^2$ を掛けパラメータの共分散行列 $\Sigma(\hat{\theta})$ を求め、該当する対角要素の平方根により求めた $SE = 0.0663$ である。表 13.27 に示した NLIN プロシジャの結果と一致している。

表 13.32 対数効力比 ρ' の SE の計算

$\hat{\theta}$	$\Sigma(\hat{\theta}) = (\mathbf{Z}^T \mathbf{Z})^{-1} \hat{\sigma}^2$					SE
$\hat{\theta}_{min}$	27.3814	-19.5016	-0.1818	0.0091	-0.6441	5.2327
$\hat{\theta}_{max}$	-19.5016	27.0809	-0.1827	0.0109	0.6452	5.2039
$\hat{\mu}_{LGS}$	-0.1818	-0.1827	0.0103	0.0012	-0.0001	0.1013
$\hat{\rho}_{\log 10}$	0.0091	0.0109	0.0012	0.0044	0.0000	0.0663
$\hat{\sigma}_{LGS}$	-0.6441	0.6452	-0.0001	0.0000	0.0188	0.1370
	$\hat{\sigma}^2 = S_e / df = 1813.86 / (70 - 5) = 24.1848$					

線形モデルと非線形モデルの解法の対比

SAS/NLIN プロシジャでは、線形モデル式であれ、非線形モデル式であれ、モデル式に対し内部でパラメータに関して自動的に偏微分が行われ、微係数行列 \mathbf{Z} が内部で生成される。微係数行列 \mathbf{Z} をデザイン行列 \mathbf{X} に代え、反応 y_i に対する推定値 $\hat{y}_i^{(m)}$ との残差 $\hat{\varepsilon}_i^{(m)}$ の列ベクトルを $\hat{\varepsilon}^{(m)}$ としたときに、線形モデルによるパラメータの推定式に対応し、

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad \Leftrightarrow \quad \hat{\delta}^{(m-1)} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \hat{\varepsilon}^{(m-1)} \tag{13.37}$$

のように、非線形モデルでは、残差 $\hat{\epsilon}^{(m-1)}$ に対して線形モデルと同様なパラメータ推定式を用い、 $\hat{\theta}^{(m-1)}$ が $\hat{\theta}^{(m)}$ とほぼ等しくなるまで繰り返し演算が行なわれている。そのために、非線形モデルの解法は、残差線形モデルとも称されている。また、パラメータの共分散行列は、線形モデルの場合と同様に、

$$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2 \Leftrightarrow \Sigma(\hat{\theta}) = (Z^T Z)^{-1} \hat{\sigma}^2 \quad (13.38)$$

で与えられる。

パラメータの標準誤差 *SE* は、線形モデルの場合と同様にパラメータの共分散行列 $\Sigma(\hat{\theta})$ の対角要素の平方根で与えられる。表 13.33 に示すように SAS/NLIN プロシジャでは、outest オプションでパラメータの共分散行列を別の SAS データセットに出力されるので、PRINT プロシジャにより出力された末尾にパラメータの共分散行列 $\Sigma(\hat{\theta})$ が計算されている。この $\Sigma(\hat{\theta})$ は、表 13.32 で示した Excel による反復計算の結果に対応する。

```
SAS プログラム 13.8 /* ロジスティック曲線 パラメータの共分散行列 */
proc nlin data=d01 outest=o01 converge=0.0000000001 ;
  parms  theta_min=1 theta_max=60 theta_min=1  muLGS=1 log10p=1  sigmaLGS=1 ;
  g = exp( (x + (log10p)*a2 - muLGS) / sigmaLGS ) ;
  model y = theta_min + (theta_max - theta_min) / (1 + g) ;
proc print data=o01 ;
run ;
```

パラメータの共分散行列 $\Sigma(\hat{\theta})$ の計算に際し、誤差分散の推定値 $\hat{\sigma}^2$ も加わるので、Excel の結果と NLIN プロシジャの結果が部妙異なる。そのために、Excel では、表 13.31 に示した再計算をさらに繰り返し表 13.32 の精度を高め、NLIN プロシジャでは、収束の条件設定 $\text{converge}=10^{-5}$ を 10^{-9} に変更して実行した結果を表 13.33 に示す。これにより、小数点以下 4 桁まで一致する。

表 13.33 NLIN プロシジャによるパラメータの共分散行列 $\Sigma(\hat{\theta})$ の出力

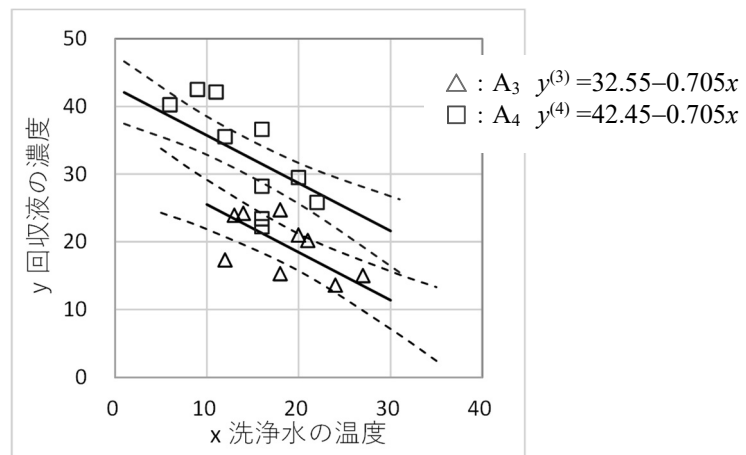
OBS	_TYPE_	_STATUS_	_NAME_	_SSE_	theta_min	theta_max	muLGS	log10p	sigmaLGS
11	FINAL	0 Converged		1813.86	7.3667	55.0179	1.7623	0.7675	0.4337
12	COVB	0 Converged	theta_min	1813.86	27.3814	-19.5016	-0.1818	0.0091	-0.6441
13	COVB	0 Converged	theta_max	1813.86	-19.5016	27.0809	-0.1827	0.0109	0.6452
14	COVB	0 Converged	muLGS	1813.86	-0.1818	-0.1827	0.0103	0.0012	-0.0001
15	COVB	0 Converged	log10p	1813.86	0.0091	0.0109	0.0012	0.0044	0.0001
16	COVB	0 Converged	sigmaLGS	1813.86	-0.6441	0.6452	-0.0001	0.0001	0.0188

層別因子を含む探索的な回帰分析入門

文献・索引 目次

文 献	491
文 献 索 引	495
索 引	497
解析用ファイル一 覧	523

洗浄水の温度 x に対する季節 A_4 および季節 A_3 の予測プロファイ



文 献

Web アクセスは全て 2024 年 6 月

—あ—

- 1) アーミテージ, ベリー著, 椿美智子, 椿広計 共訳 (2001), 医学研究のための統計的方法, 原著第 3 版, 82-84, 267-72, サイエンティスト社.
- 2) Armitage P., Berry G. and Matthews J.N.S. (2002), *Statistical Methods in Medical Reserch*, 4th ed., 408-18, 730-37, Blakwell.
- 3) 朝香鐵一, 石川馨, 山口襄 共同監修 (1988), 新版 品質管理便覧 第 2 版. 446-47, 日本規格協会.
- 4) 岩崎学 (2002), 「処置前—処置後」データの解析と平均への回帰, 日本行動計量学, Vol. 29 : 247-73, https://www.jstage.jst.go.jp/article/jbhmk/29/2/29_2_247/_pdf
- 5) 岩崎学 (2006), 統計的データ解析入門, 単回帰分析, 74-75, 東京図書.
- 6) 大和田章一 (2010), 線形モデルと非線形モデルの基本的な考え方—逆推定の解析, 標準誤差と信頼限界—, 第 2 期 医薬安全性研究会, 第 7 回定例会.
https://biostat.jp/archive_teireikai_2_download.php?id=19
- 7) 奥野忠一, 久米均, 芳賀敏郎, 吉澤正 (1981), 多変量解析法《改訂版》, 112-23, 日科技連出版社.
- 8) 奥野忠一 応用統計ハンドブック編集委員会編 代表者 (1978), 応用統計ハンドブック, 315-17, 養賢堂.

—か—

- 9) 楠正, 辻谷将昭, 松本哲夫, 和田武夫 (1995), 185-208, 応用実験計画法, 日科技連出版社.
- 10) Collett D. (2003), *Modelling Binary Data*, 2nd ed., 106-11, Chapman & Hall.

—さ—

- 11) 寄山陽二郎, 大橋雅津代, 高橋行雄 (2008), 薬理試験における統計解析の Q&A : 用量依存性データへの非線形モデルの活用, 日薬理誌, Vol 132, 199~206.
https://www.jstage.jst.go.jp/article/fpj/132/4/132_4_199/_pdf/-char/ja
- 12) 佐久間著, 五所正彦, 酒井弘憲, 佐藤泰憲, 竹内久朗 編, 新版 薬効評価, 232-55, 東大出版会.
- 13) SAS Institute Inc. (2020), *The GLM Procedure, SAS/STAT 15.2 User's Guide*, 4023-26,
<https://go.documentation.sas.com/api/docsets/statug/15.2/content/glm.pdf>
- 14) SAS Institute Inc. (2020), *The NLIN Procedure, SAS/STAT 15.2 User's Guide*,
<https://documentation.sas.com/api/docsets/statug/15.2/content/nlin.pdf>

- 15) 新村秀一(1983), 重回帰分析における掃き出し演算子, オペレーションズ・リサーチ, vol.28: 565-669. http://orsj.or.jp/~archive/pdf/bul/Vol.28_11_565.pdf
- 16) スネデカー, コクラ ン著, 畑村又好, 奥野忠一, 津村善郎 訳(1972), 統計的方法, 原著第 6 版, 394-402, 岩波書店.
- た—
- 17) 高橋行雄, 大橋靖雄, 芳賀敏郎(1989), SAS による実験データの解析, 111-27, 289-300, 東京大学出版会.
- 18) 高橋行雄(2012), 平均への回帰を考慮した投与前値の調整, 第 2 回 続高橋セミナー.
https://www.yukms.com/biostat/takahasi2/rec/archive/takahashi02_2012_08_17.pdf
- 19) 高橋(2013a), 応用回帰分析 I —各種の重み付き回帰における逆推定—, 第 3 回 続高橋セミナー. <https://www.yukms.com/biostat/takahasi2/rec/003.htm>
- 20) 高橋(2013b), 回帰分析・再入門 —統計ソフトが対応していない生物統計の各種の課題を Excel でサクサク解こう, 第 7 回 じっくり勉強すれば身につく統計入門.
<https://scientist-press.com/wp-content/uploads/2019/07/seminar7.pdf>
- 21) 高橋行雄(2015), 寿命試験データの統計解析, 第 4 回 続高橋セミナー.
[https://www.yukms.com/biostat/takahasi2/rec/archive/takahashi_04_2\(2015_04_10\).pdf](https://www.yukms.com/biostat/takahasi2/rec/archive/takahashi_04_2(2015_04_10).pdf)
- 22) 高橋行雄(2017), 一般化線形モデルを Excel で極め活用する —プロビット法・ロジット法・補 2 重対数法—, 第 6 回 続高橋セミナー
<https://www.yukms.com/biostat/takahasi2/rec/006.htm> :takahashi06(2017-08).zip
- 23) 高橋行雄(2018), 二値データの用量反応関係における新たな解析 —プロビット・ロジット・補 2 重対数の使い分け—, 第 2 期 医薬安全性研究会, 第 22 回定例会.
https://biostat.jp/archive_teireikai_2_download.php?id=150
- 24) 高橋行雄(2021), 最尤法によるポアソン回帰分析入門, 423-39, カクワークス社.
<https://www.yukms.com/biostat/takahasi2/rec/archive/takahashi09all-main.pdf>
- 25) 高橋行雄(2022), 線形モデルによる欠測値がある直交表の解析 —謎めいた最小 2 乗平均と 95%信頼区間の活用—, 第 10 回 続高橋セミナー.
<https://www.yukms.com/biostat/takahasi2/rec/010.htm>
- 26) 高橋行雄(2023), JMP で繰り返し不揃いの 2 元配置データの解析ができるの? —平方和の分解ではなくデザイン行列と最小 2 乗平均の活用—, 第 11 回 続高橋セミナー.
<https://www.yukms.com/biostat/takahasi2/rec/011.htm> :takahashi11(2023-03-16).zip
- 27) 竹内啓(1963), 数理統計学, 310-15, 東洋経済.
- 28) 竹内啓 編集委員代表(1989), 統計学辞典, 621-25, 東洋経済新報社.
- 29) 田中豊, 垂水共之(1986), パソコン統計解析ハンドブック, III 実験計画法編, 共立出版.

- 30) 知名秀泰, 岡田豊(2014), 原典からの酵素反応速度論, 生物工程, Vol. 92 : 20-25.
https://www.sbj.or.jp/wp-content/uploads/file/sbj/9201/9201_yomoyama.pdf
- 31) ドブソン著, 田中豊, 森川敏彦, 山中竹春, 富田誠 訳(2008), 141-50, 共立出版.
- 32) Dobson A.J. and Barnett A.G. (2008), An Introduction to Generalized Linear Model 4th ed., 124-31, CRP press.
- 33) ドレーパ, スミス著, 中村慶一 訳(1968), 応用回帰分析, 8-18, 47-60, 森北出版.
- 34) Draper N.R. and Smith H. (1998), Applied Regression Analysis 3rd ed., 20-40, 115-31, John Wiley & Sons.

— な —

- 35) 中西展大(2016), 非線形回帰を用いた逆推定の基礎, 第 12 回じっくり勉強すれば身につく統計入門. <https://scientist-press.com/wp-content/uploads/2019/07/seminar12.zip>
- 36) ネルソン著, 奥野忠一監訳, 柴田義貞, 藤野和研, 鎌倉稔成 訳(1988), 34-41, 日科技連出版社.

— は —

- 37) 芳賀敏郎(2014), 医薬品開発のための統計解析 第 2 部 実験計画法 改訂版, 15-45, 87-96, 133-51, サイエнтиスト社.
- 38) 芳賀敏郎(2016), 医薬品開発のための統計解析, 第 3 部 非線形モデル 改訂版, 14-16, 42-56, サイエнтиスト社.
- 39) 原田淳(2017), 平行線検定を利用した薬物の効力比較, 日薬理誌, Vol. 150, 16-22.
https://www.jstage.jst.go.jp/article/fpj/150/1/150_16/_pdf
- 40) 原田淳, 吉池通晴(2018), 平行線検定(直線及びシグモイド曲線)による効力比較, 第 2 期 医薬安全性研究会, 第 21 回定例会.
https://biostat.jp/archive_teireikai_2_download.php?id=140
- 41) ピース編, 中上節夫, 森川敏彦 監訳(1992), 医薬統計学 医薬品開発のための統計学, 97-104, サイエнтиスト社.
- 42) 広瀬明彦, 西浦博(2021), ベンチマークドース手法の適用の現状と課題 —動物実験データへの適用を中心に—, 産業医学レビュー Vol. 34, 1-15.
https://www.jstage.jst.go.jp/article/ohpfrev/34/1/34_1/_pdf
- 43) Finney D.J. (1978), Statistical Method in Biological Assay 3rd ed., 69-104. Charles Griffin.
- 44) フライス著, KR(関西臨床データ解析)研究会 訳(2004), 臨床試験のデザインと解析, 189-223. アーム.

- 45) Frison L and, Pocock S.J. (1992), Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and its Implications for Design. Stat Med, Vol. 11 :1685-704.

—ま—

- 46) MaCullough B.D. and Heiser D.H. (2008), On the accuracy of statistical procedures in Microsoft Excel 2007, Computational Statistics and Data Analysis Vol. 52:4570–4578
- 47) 宮川雅巳(2008), 問題発見と解決の科学 SQC の基本, 54-62. 日本規格協会.

—や—

- 48) 山崎亜紀子, 久米英介, 田中澄子, 浜田知久馬, 吉村功(2008), 非線形回帰法によって薬理研究の精度は向上する —酵素阻害薬の阻害定数の推定の事例—, 日薬理誌 Vol. 131; 195~203. https://www.jstage.jst.go.jp/article/fpj/131/3/131_3_195/_pdf
- 49) 山田雅之, 吉田光宏, 高橋行雄(2003), 陰性および陽性対照があるシグモイド曲線—ダミー変数を持つ非線型回帰モデルの応用—, SAS ユーザ総会論文集 2003; 51-60. https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/sugj2003.pdf
- 50) 吉村編著(1987), 毒性・薬効データの統計解析 —事例研究によるアプローチ—, 101-109, 110-114, 226-32, サイエンティスト社.

—ら—

- 51) Littell R.C., Strop W.W. and Freund R.J. (2002), SAS for Linear Models 4th ed., 230-38, SAS Institute.
- 52) Rouaud M. (2017), Probability, Statistics and Estimation, 80-85. <https://www.incertitudes.fr/book.pdf>

文献索引

あ	アーミテイジら著, 椿・椿訳(2001) - 医学研究のための統計的方法, 第3.6節 分散に関するその他の公式	184
	第9.4節 群内の回帰	327
	Armitageら(2002) - Statistical Methods in Medical Research, 4th edition. Sec. 20.5 Michaelis-Menten assay	405
	朝香・石川・山口 監修(1988) - 新版 品質管理便覧 第2版	121
	朝香ら(1988) - 新版 品質管理便覧 第2版 7.1.4.2節 直交表の性質とその利用	148
	岩崎(2002) - 「処置前-処置後」データの解析と平均への回帰	290
	岩崎(2006) - 単回帰分析, 第4-3節 原点を通る回帰式(Excel)	53
	大和田(2010) - 線形モデルと非線形モデルの基本的な考え方-逆推定の解析, 標準誤差と信頼限界-	194
	奥野・久米・芳賀・吉澤(1981) - 多変量解析法《改訂版》, 第II部, 第7章 層別因子を含む回帰分析	19, 195
	奥野ら(1978) - 応用統計ハンドブック, 3.7.3節 乱塊法における共分散分析	47
か	楠・辻谷・松本・和田(1995) - 応用実験計画法, 第6章 線形モデル, 6.2節 線形推定論	121
	Collett(2003) - Modeling Binary Data 2nd. ed. Sec. 4.2 Estimating an effective dose	183
	崑山・大橋・高橋(2008) - 薬理試験における統計解析のQ&A 用量依存性データへの非線形モデルの活用	455
さ	佐久間著, 五所ら編(2017) - 新版 薬効評価, 第7章 計量的反応	480
	SAS Institute Inc. (2020) - The GLM Procedure, SAS/STAT 15.2 User's Guide	102
	- The NLIN Procedure, SAS/STAT 15.2 User's Guide.	358
	新村(1983) - 重回帰分析における掃き出し演算子	126
	スネデガー・コクラン(1972) - 統計的方法 原著 第6版, 第14章 共分散分析	282
た	高橋・大橋・芳賀(1989) - SASによる実験データの解析	121
	- SASによる実験データの解析, 第15章 4種の平方和とLSMEANS	89
	- SASによる実験データの解析, 第7章 共分散分析	260
	高橋(2012) - 平均への回帰を考慮した投与前値の調整	310
	高橋(2013a) - 応用回帰分析 I - 各種の重み付き回帰における逆推定 -	183
	高橋(2013b) - 回帰分析・再入門 - 統計ソフトが対応していない生物統計の各種の課題をExcelでサクサク解こう	183
	高橋(2015) - 第4回 続高橋セミナー: 寿命試験データの統計解析	378
	高橋(2017) - 一般化線形モデルをExcelで極め活用する - プロビット法・ロジット法・補2重対数法 -	336, 369
	高橋(2018) - 二値データの用量反応関係における新たな解析 - プロビット・ロジット・補2重対数の使い分け -	336
	高橋(2021) - 最尤法によるポアソン回帰分析入門, 第13.2節 交互作用を考慮した共分散分析	19
	高橋(2022) - 線形モデルによる欠測値がある直交表の解析	144
	高橋(2023) - 続・高橋セミナー 第11回 JMPで繰り返しが不揃いの2元配置データの解析ができるの?	102
	竹内(1963) - 数理統計学, 第29章 IV 回帰直線自体についての推論	183
	竹内ら(1989) - 統計学辞典 IV分野別応用, 第4.3節 毒性評価	243
	田中・垂水(1986) - パソコン統計解析ハンドブック III 実験計画法編	137
	知名・岡田(2014) - 原典からの酵素反応速度論	405

文献索引

Dobsonら(2008) - An Introduction to Generalized Linear Model 4th Ed.	369
ドブソン著・田中ら訳(2008) - 一般化線形モデル入門, 7.3節 用量反応モデル	369
Draper, Smith(1998) - Applied Regression Analysis 3rd ed.	156
ドレーパ・スミス著, 中村訳(1968) - 応用回帰分析	156
な 中西(2016) - 非線形回帰を用いた逆推定の基礎	194
ネルソン著・奥野監訳(1988) - 寿命データの解析	378
は 芳賀(2014) - 医薬品開発のための統計解析 第2部 実験計画法 改訂版. 第1章 質的因子の1因子実験	55
第2章 量的因子の1因子実験	80
第4章 共分散分析	7
芳賀(2016) - 医薬品開発のための統計解析, 第3部 非線形モデル 改訂版	183, 337, 372
原田(2017) - 平行線検定を利用した薬物の効力比較	480
原田, 吉池(2018) - 平行線検定(直線及びシグモイド曲線)による効力比較,	480
ピース編, 中上・森川監訳(1992) - 医薬統計学, 第3章 薬理学的活性の評価	469
広瀬ら(2021) - ベンチマークドース手法の適用の現状と課題 -動物実験データへの適用を中心に-	382
Finney(1978) - Statistical Method in Biological Assay 3rd ed., Chap. 4 Parallel line assay	480
フライス(2004) - 臨床試験のデザインと解析, 第7章 共分散分析と変化量の検討	30
Frison, Pocock(1992) - Repeated Measures in Clinical Trials: Analysis Using Mean Summary Statistics and its Implications for Design	294
ま MaCulloughら(2008) - On the accuracy of statistical procedures in Microsoft Excel 2007	155
宮川(2008) - 問題発見と解決の科学-SQCの基本-, 第6.3節 計画的に採取した3元配置データの解析例	219
や 山崎ら(2008) - 非線形回帰法によって薬理研究の精度は向上する -酵素阻害薬の阻害定数の推定の事例-	428
山田・吉田・高橋(2003) - 陰性および陽性対照があるシグモイド曲線 ' -ダミー変数を持つ非線型回帰モデルの応用-	379
吉村編著(1987) - 毒性・薬効データの統計解析, 第3.2節 臓器重量の解析における体重データの利用	260
第3.3節 体重は増加量でみるべきか?	243
第5.4節 LD50の推定法	369
ら Littellら(2002) - SAS for Linear Models 4th ed., Chapter 7 Analysis of Covariance	38
Rouaud(2017) - Probability, Statistics and Estimation. https://www.incertitudes.fr/book.pdf	420

索引

あ Armitageら(2002) - 散布図から復元	407	い Eadie-Hofstee 変換 - Scatchard 変換	406
- 推定値に若干の差異	420	Emaxモデル - 芳賀(2016)	372
- Direct linear plot	420	閾値 p 値 - 停止ルール	256
- Non-linear regression	405	(1, 1) 標示型 - 阻害薬Aの濃度別	439
- 標準誤差が示されていない	420	- ダミー変数	277, 336, 439, 457, 472
- Michaelis-Menten assay	405	(1, 1) 標示型ダミー変数 - セル平均モデル	68
- 両逆数 GLM/ gamma errors	420	1 因子実験 - 共分散分析	230
アーミテージら(2001) - カドミウムガス	327	- 繰り返し等しい	55
- 共分散分析	327	- 繰り返し不揃い	69
- 群内の回帰	327	- 質的因子	55
- 積の分散	184	- (0, 1)型ダミー変数	77
- 比の分散	184	- 多項式回帰	80
R が脚光を浴びている - 無償の統計ソフト	149	1 元配置型 - 共分散分析	38
R の lm () 関数 - 原点共通	54	1 次従属 - 制約条件	136
RMSE - 誤差の平均平方	258	一次式 - 回帰モデル	122
outest オプション - NLINプロシジャ	488	(1, 0)型 - GLMプロシジャ	150, 213
- パラメータの共分散行列	488	- 切片の推定値は?	152
OUTEST= - パラメータの共分散行列	360	- (0, 1) 型	153
outest=cov1 オプション - NLINプロシジャ	424	- ダミー変数	10, 39
outputステートメント - NLINプロシジャ	403, 424	(1, 0)型か - (1, 2)型ダミー変数	58
- 95%信頼区間	463	- (0, 1)型か	58
- DATAステップ	400	(1, 0)型ダミー変数 - パラメータ推定値	152
朝香ら(1988) - 直交表の歴史的背景	148	(1, 2)型 - (1, -1)基本型	148
- 品質管理便覧, 新版 第2版	121	- L_8 直交表	137
「値」のみ - ペースト	313	(1, 2)型ダミー変数 - (1, 0)型か	58
あてはまりの悪さ - LOF解析	87	(1, 2)型直交表 - (1, -1)対比型	135
あてはめの基礎 - ロジスティック曲線	335	1.0 に近づく - 時点間の相関	313
後値 Y - 共分散分析	296	位置パラメータ - 規準化	457
- 偶然に有意な差	292, 301	- 形状パラメータ	340
- 群間比較	290	位置パラメータ μ - 形状パラメータ σ	367
- 前後差 D	290	位置パラメータ μ_{GMP} - 形状パラメータ σ_{GMP}	375
- 前値 X を共変量	297	位置パラメータ μ_{MEV} - 95%信頼区間	389
- 名目の α エラーに接近	314	位置パラメータ μ_{SEV} - 形状パラメータ σ_{SEV}	377
- 有意な差	292	(1, -1)基本型 - (1, 2)型	148
後値 Y $p=0.020$ - 前後差 D $p=0.640$	313	- (0, 1)型	148
後値 Y $p=0.404$ - 前後差 D $p=0.023$	313	(1, -1)対比型 - (1, 2)型直交表	135
後値 Y $p=0.615$ - 前後差 D $p=0.651$	314	- 効果 α_i の推定	79
後値 Y および群間差 - 最小2乗平均の図示	300	- コーディングのテーブル	144
後値 Y および前後差 D - 共に有意な差	293	- (0, 1)型	230, 305
- 共に有意な差	305	- (-1, 1)対比型	258
Analysys of Covariance - Littellら(2002)	38	(1, -1)対比型ダミー変数-JMPの解析用ファイル	144
Average () 関数 - 平均	57, 283	(1- β)エラー - 検出力(1- β)	310
- Excel	167	一樣乱数 - Rand () 関数	456
新たな統計解析 - ミカエリス・メンテン式	405	位置を変える - 邪魔された分	398
パラメータ - 共分散行列	127	一括推定 - 個別パラメータ	451
Arithmetic mean - design matrix	163	一定の検出力 - 前値 X を共変量	326
α エラー - 第1種の過誤	310	一般式を拡張 - 合成分散	131
α エラーが変化 - 前値 X の大小関係	316	一般線形モデル - SAS	45
α エラーが保持 - 前値 X を共変量	316	- GLMプロシジャ	45
Alosetronの3用量 - 内臓痛を緩和	455	伊奈の式 - 芳賀(2014)	66
ANCOVA - 投与前値を共変量	294	- 有効反復数	66
- Frison, Pocock (1992)	294	if () 関数 - Excel	224
アンジオテンシン I - 血圧上昇	469	「いぶしろん」 - 「 ϵ 」に変換	157
Unbalanced ANOVA-Getting Started, PROC GLM	102	岩崎(2002) - 処置前-処置後	290
- PROC GLM	102	- 平均への回帰	290
い η 分布 - 基準化した変数	397	岩崎(2006) - 原点を通る回帰式	53
η 分布に関して解く - パーセント点 π 分布	397	- 単回帰分析	53

い	因子が直交 - デザイン行列 X	101	え	- 分析ツール	59
	因子が直交しない場合 - 繰り返しが不揃い	102		- 偏微分式の計算	344
	- 高橋ら(1989)	102		- 本質的な脆弱性	421
	主効果から因子Aを除く - タイプ II の平方和	112		- Mmult () 関数	161, 271
	主効果から因子Bを除く - タイプ II の平方和	112		- 予測プロファイル	26
	陰性および陽性対照 - シグモイド曲線	379		- Rand () 関数	283, 456
	陰性対照群 - Vehicle群	455		Excel 2007 に対する批判 - MaCullough ら(2008)	155
	Intecept () 関数 - Excel	168		Excel 2007 以前 - 計算ミス	178
	Minverse () 関数 - 逆行列	226		- 互換性・メッセージに注意	178
	インフレーション - 前後差 D の分散	287		Excel シート - 行列計算の実際 $X\beta$	159
	inputステートメント - DATAステップ	400		- 検出力の計算手順	319
う	Wikipedia - Gauss-Newton algorithm	420		- 検出力を可視化	318
	- ガウス・ニュートン法	336, 365, 366, 420		- デザイン行列 X を入力	157
	- 計画行列	163		Excel ソルバー - 逆推定の実際	189
	- design matrix	163		- 正確な95%信頼区間	188
	- 非線形回帰	366		Excel に - HTML出力を	45
	- ミカエリス・メンテン式	366		Excel により手軽に - 非線形回帰	419
	ウス不良 - 流れ不良	220		Excel による解析 - 本質的な脆弱性	391
	内側の数が一致 - 積和の計算	160		Excel による行列計算 - 線形推定論	123
	うつ症状 - Hamilton の評価尺度	30		Excel による非線形回帰 - SASによる検証	421
	埋め込まれている計算式 - 予測値 y^{\wedge}	353		- 阻害定数 K_i の直接推定	447
	上側95%信頼区間が「0」 - ソルバー	334		Excel の 1 枚のシート上で - 非線形回帰	420
え	A_1 群の平均値がやや大きい - 前値 X	312		Excel の X 軸の書式設定 - 対数目盛	410
	HTML出力を - Excelに	45		Excel のソルバー - 残差平方和 S_e を最小に	408
	exp () 関数の中の線形式 - 初期値	340		- 阻害定数 K_i の直接推定	450
	Excel - Average () 関数	167		- パラメータ推定	459, 460
	- if () 関数	224		- 非線形回帰	439
	- Intecept () 関数	168		Excel のソルバー/非線形回帰 - 線形変換で悩む	454
	- Minverse () 関数	170		Excel の回帰分析 - 共分散分析	204, 468
	- 折れ線グラフ	18, 28, 95		- 結果が怪しい	53
	- 解析マナーの習得	156		- 欠測値がある直交表	140
	- ガウス・ニュートン法	457		- 現実的な対応	181
	- 行列関数	155		- SASのGLMプロシジャ	115
	- 行列計算	7		- 前値 X を共変量	295
	- 矩形データ	55		- デザイン行列 X	296
	- 計算式の設定ミス	53		- 定数に 0 を使用	9
	- 効力比の95%信頼区間	485		- 変数は隣り合う	231
	- Correl () 関数	264, 286		- 変数減少法	230
	- 再計算/自動的	283		Excel の計算シート - 現代の算盤	410
	- SumSq () 関数	91, 161		- 阻害定数 K_i の推定	448
	- SumProduct () 関数	161, 168		Excel の行列の積 - Mmult () 関数	410
	- 散布図	56		Excel の行列関数 - 回帰パラメータの推定	311
	- Slope () 関数	168		Excel の行列計算 - 回帰パラメータ	174
	- 正規乱数	283		- 回帰分析の実際	174
	- 積和行列の逆数	344		Excel の線グラフ - SAS と Excel の連携法	361
	- 線形近似	8		Excel で再現 - SASの結果は正しいのか?	404
	- 操作ミス	53		Excel の作図機能 - 統計ソフトの出力	425
	- 層別散布図	8, 196		Excel の散布図 - 重ね書き	296
	- ソルバーの活用法	341		行列の転置/Excel - Transpose() 関数	410
	- 単回帰分析	86		S_R - 回帰の平方和	176
	- 探索的な回帰分析	250		S_e - 残差平方和	176
	- Mdetarm () 関数	169		S_e と S_R に分解 - 全体の平方和 S_T	182
	- T.test () 関数	245		SS1 SS2 SS3 - タイプ I, II, III	212
	- DevSq () 関数	168, 245		SSE - 誤差の平方和	258
	- Transpose () 関数	161		SQCの基本 - 宮川(2008)	219
	- Norm.dist () 関数	367		estimateステートメント - 各種の推定値	45
	- 箱ひげ図	56		- 最小2乗平均の差	218
	- 分散分析表の作成	176		- 最適水準の推定	152
	- データ分析ツール	158		- GLMプロシジャ	120

え	- 水準の差の推定	152	え	- 直線関係か否か	80
	- 推定値	46		- 芳賀(2014)	80
S_T	- 全体の偏差平方和	176		- 分散分析表	87
$S_T = S_e + S_R$	- 平方和の分解	177	LD_{50}	- 50パーセント致死量	367
$(X^T X)^{-1} \sigma^2$	- $\Sigma(\theta^{\wedge})$	109	LD_{50}	- 推定法 - 吉村編著(1987)	369
	- パラメータの共分散行列	109	L_8	- 直交表 - (1, 2)型	137
X軸方向	- オフセット	478		- 線形モデル	135
「X」	- 「X」をボールド&イタリック	157		- データの構造式	135
XとY	- 相関があるか	264		- デザイン行列 X	137
X切片	- 負の阻害定数 K_i の推定	431	お	応用実験計画法 - 楠ら(1995)	121
	- $-(Km)_0$	433		大きい場合の検出力 - 前値 X の差	319
HTML形式	- SASの標準出力	359		大きく変化 - 回帰パラメータ	233
	- SASの標準出力	421		扇型線グラフ - 増加率	247
	- GLMプロシジャの出力	151		大和田(2010) - 逆推定の解析	194
NLINの逆推定のSE	- 直接推定	402		- 非線形モデル	194
NLINプロシジャ	- outest オプション	488		起きた結果 - 平均への回帰	313
	- outest=cov1 オプション	424		奥野ら(1978) - 除草剤の効果	47
	- outputステートメント	403, 424		- 乱塊法における共分散分析	47
	- オフセット変数	401		奥野ら(1981) - 共分散分析	203
	- 効力比の直接推定	477		- 層別因子を含む回帰分析	19
	- ゴンペルツ曲線	399		- 多変量解析法	19
	- SAS	397, 461		- 伝統的な共分散分析	195
	- SAS の非線形	358		驚くべき検出力の差 - 2群間の大小関係	325
	- SAS Institute(2020)	399	off	- 定数に 0 を使用	40, 49
	- 自動偏微分	358	オフ	- 定数に0を使用	59
	- 阻害定数 K_i の直接推定	444	オフ(デフォルト)	- 定数に 0 を使用	32
	- parms ステートメント	401	offset from reference group	- One-way ANOVA	163
	- パラメータに関する偏微分	464	オフセット	- X軸方向	478
	- パラメータの一括推定	452		- 逆推定	397
	- 非線形回帰	421, 444		- ゴンペルツ曲線	399
	- プログラミング機能	465		オフセットの推定値 - 各シグモイド曲線	398
	- 平行なロジスティック曲線	484		オフセット式 - 逆推定値	193
	- Hanes-Woolf 変換	426		オフセット値 η 分布 $\pi=0.10$ - 邪魔をする	398
	- 偏微分した式	483		オフセット変数 - NLINプロシジャ	401
	- 偏微分式の出力	465, 485	折れ線グラフ	- Excel	18, 28, 95
	- modelステートメント	401		- 95%信頼区間	95, 95
	- 理解を深めたい	447		- 作図のヒント	64
	- list オプション	465		- ひげの長さの設定	208
F.dist.RT () 関数	- F分布の右側確率	13		- 予測プロファイル	75
	- F分布の上側確率	178		- ロジスティック曲線	481
$f_{NOR}(x)$	- 正規分布の確率密度関数	368	折れ線グラフ/Excel	- 行/列の切り替え	248
	- 正規分布の累積分布関数 $F_{NOR}(x)$	368	終わりにしませんか	- 線形変換で悩む	454
F分布の右側確率	- F.dist.RT () 関数	13	on	- 定数に 0 を使用	24, 199
F分布の上側確率	- F.dist.RT () 関数	178	オン	- 定数に 0 を使用	9, 51
$F_{LGS}(x)$	- ロジスティック分布の累積分布関数	369	OnDemand SAS	- SAS	44
Minverse () 関数	- 逆行列	15, 125		- GLMプロシジャ	150
	- Excel	170		- 無償で継続的に	44, 149
MCP剤	- 共通の切片	49		- 無償で継続的	195, 211, 394
Mdetarm () 関数	- 行列式	169		- 無償で継続的	211, 397, 421, 483
Mmult () 関数	- 行列関数	15	か	回帰の平方和 - 共通の傾き	198
	- Excel の行列の積	410		- S_R	176
lsmeans	- 最小2乗平均	142, 147, 218, 297		- 差分	107
	- 4種の平方和	89		- 水準ごと	198
lsmeansステートメント	- 最小2乗平均	45		回帰の平方和の差 - 交互作用の平方和	33
	- 最小2乗平均	46, 152, 216, 280		回帰パラメータ - Excel の行列計算	174
	- SASのGLM	280		- 大きく変化	233
LOF解析	- あてはまりの悪さ	87		- t検定	176
	- タイプ I の平方和	88		- 分散および共分散の推定	175
	- 多項式回帰	80			

か 回帰パラメータの推定 - 偏差平方和	164	か - 定数に 0 を使用	472
- Excel の行列関数	311	回帰分析・再入門 - 高橋 (2013b)	183
- 行列計算	15	回帰分析の実際 - Excel の行列計算	174
回帰パラメータの比 - 逆推定値	183	回帰分析の適用 - 重層的	197
回帰モデル - 一次式	122	回帰分析の反復 - 残差を反応	363
- 構造モデル	121	回帰分析の併合 - 水準ごと	197
- 主効果モデル	204	回帰分析を繰り返す - 残差 ε^{\wedge} を反応	468
- 線形モデル	121	回収液の濃度 - 洗浄水の温度	19
回帰式 - 交互作用	32	解析のための決定樹 - 臓器重量	280
回帰式が複雑 - 中心化	306	解析の妥当性 - 交互作用を含める	309
回帰直線 - 傾きが異なる	268	解析の目的に合わせた - ダミー変数の選択	79
- 季節ごと	195	解析マナーの習得 - Excel	156
- 逆推定	470	解析手法の選択 - 決定樹	263
- 95%信頼区間	27, 42, 178, 190	解法の対比 - 線形モデルと非線形モデル	487
- 苦し紛れの解釈	338	Gauss-Doolittle 法 - 逆行列の計算	126
- 原点	9	Gauss-Newton algorithm - Wikipedia	420
- 原点を共通	48, 52	ガウス・ニュートン法 - Wikipedia	336
- 原点を通る	199, 266	- Wikipedia	365, 366, 420
- 個別データの95%信頼区間	224	- Excel	457
- 散布図	28	- 残差線形化法	457
- 推定されたパラメータ	234	- 数値例	366
- 切片を共通	48	- ミカエリス・メンテン式	336
- 層別散布図	219	- Rouaud (2017)	420
- 同一直線上か	265, 273	カキの成長 - 設置場所	39
- パラメータの共分散行列 $\Sigma(\beta^{\wedge})$	473	- 発電プラント	38
- 非平行	31	カキの成長に関する実験 - Littellら (2002)	38
- 別々	8	各群の最小値 - 95%信頼区間	345
- 便宜的な方法	335	各種の95%信頼区間 - ゴンペルツ曲線	385
回帰直線の95%信頼区間-Scatchard・Eadie-Hofstee変換	417	各種のシグモイド曲線 - オフセットの推定値	398
- 伝統的な方法	180	各種 - ダミー変数	10
- Hanes-Woolf 変換	415	各種のダミー変数 - 歴史的背景	148
- Lineweaver-Burk 変換	413	各種の群間比較 - 検出力(1- β エラー)	318
回帰直線の傾きの差 - t検定	330	各種の推定 - 回帰分析	155
回帰直線の差 - 95%信頼区間	272	- 共通の傾き	205
- 95%信頼区間	332	- 欠測値がある場合	142
- 非平行	34	各種の推定値 - estimateステートメント	45
回帰直線の推定値 - 分散	225	各種の推定方法 - 阻害定数 K_i の比較	442
回帰分析 - 各種の推定	155	各種の線形変換 - ミカエリス・メンテン式	417
- 計算公式の範囲内	168	各種の線形和 L - 推定値	206
- 交互作用モデル	204	拡張 - 4パラメータモデル	372
- 最小2乗法	164	- 合成分散の一般式	26
- (0, 1)型	78	確認のための手順 - 阻害様式	451
- 前値を共変量	281	確率楕円 - 散布図	264
- 全データ	197	確率密度関数 - 累積分布関数	367
- 体重を交絡変数	268	下限が存在 - 血圧の上昇	481
- 対比型デザイン行列 X	97	下限に漸近 - ロジスティック曲線	339
- タイプ I, タイプ II, タイプ III の平方和	115	化合物 - 毒性	260
- 探索的な解析	230	重ね書き - Excel の散布図	296
- デザイン行列	7	- 層別散布図	103
- デザイン行列 X	71, 105, 225	- 平均 \pm SD	222
- デザイン行列 X の一部	101	可視化 - 交互作用	93
- 2次式	83	- 前値 X の揺らぎ	285
- 複数の共変量	250, 225	- 平均への回帰	283
- 分散分析表	72	数多くの変数 - 多重比較	243
- データ分析ツール	158	カスタム検定 - 水準の差の推定	146
- 分析ツール	59	- デザイン行列 X	147
- モデルのあてはめ	20	重ね書き - 3本の回帰直線	329
- LinEst () 関数	60	傾き β^{\wedge}_1 と切片 β^{\wedge}_0 の比 - 偏微分	412
回帰分析/Excel - ソルバー	341	- ミカエリス定数 K_m^{\wedge}	412

か 傾き $(Km)_i / (Vmax)_i$ - 阻害定数 K_i の推定	436	き	- パズル的な代数計算	126
- 阻害薬濃度 $[I]$ に関する回帰	437	- 両辺に掛ける	169	
傾き β_1 の逆数 - Hanes-Woolf 変換式	427	逆行列の計算 - Gauss-Doolittle 法	126	
傾きが異なる - 回帰直線	268	- 楠ら(1995)	126	
傾きが異なる回帰直線の差 - t 検定	333	逆行列を前掛 - 単位行列	169	
傾きが同じでない - 2本の直線	267	逆推定 - オフセット	397	
傾きに差 - 95%信頼区間	272	- 回帰直線	470	
傾きに統計的に差 - どのように説明	331	- グラフ表示	193	
傾きの差 - 検定統計量	268	- Gompertz 4P	393	
- 交互作用モデル	269	- シグモイド曲線	367	
傾きの差の範囲 - 同等と見なせる	480	- 10パーセント反応	381	
傾きは同じか - 2本の直線	265	- 非線形回帰	395	
隔靴搔痒的な解釈 - 分散分析表の解釈	115	逆推定の解析 - 大和田(2010)	194	
画期的 - JMP の使い勝手	149	逆推定の基礎 - 中西(2016)	194	
画期的な解析方法 - 中心化	309	逆推定の実際 - Excel ソルバー	189	
葛藤 - 平均への回帰	281	逆推定式を偏微分 - JMPの逆推定のSE	402	
活用/さらなる - ロジスティック曲線	455	逆推定値 - オフセット式	193	
過度に標準化 - ガラスの天井	86	- 回帰パラメータの比	183	
- 計算公式	86	- 95%信頼区間	183, 390, 397	
カドミウムガス - アーミテージら(2001)	327	- 95%信頼区間の直接推定	193	
- 加齢と共に約 0.6リットル減少	348	- 個別データ	187	
- 肺活量の減少	327, 328, 337	- 正確な95%信頼区間	185, 187	
ガラスの天井 - 過度に標準化	86	- 分散とSE	473	
- 計算公式の範囲内	168	逆推定値の95%信頼区間 - JMP の回帰分析	190	
- 2次式の95%信頼区間	181	- 竹内(1963)	183	
加齢と共に約 0.6リットル減少 - カドミウムガス	348	- モデルのあてはめ	192	
環境ホルモン - ラットの子宮重量	379	逆推定値の式 - パラメータで偏微分	184	
関数の形式はまちまち - ワイブル分布	376	逆推定値の比較 - 95%信頼区間	187	
ガンスプレー - 吐出量 x	219	逆数あるいは比 - 95%信頼区間	417	
間接的な作用 - 直接作用	245	逆数の標準誤差 - 誤差の伝播の式	427	
簡単な数値を与えて例示 - 阻害定数 K_i	430	逆数変換 $(1/V)$ - 誤差が不均一	407	
管理限界 - 個別データの95%信頼区間	235	95%信頼区間 - outputステートメント	463	
- 上限および下限	221	- 位置パラメータ μ_{MEV}	389	
- ソルバー	228	- 折れ線グラフ	95, 130	
- 中心膜厚	219	- 回帰直線	27, 42, 178, 190	
- 統計的な判断	224	- 回帰直線の差	272, 332	
管理限界を超える - 吐出量 x	235	- 各群の最小値	345	
き 基質濃度 $[S]$ - Substrate concentration	407	- 傾きに差	272	
- 対数目盛	373	- 期待値の差	109	
標準化した変数 - η 分布	397	- 逆推定値	183, 390, 397	
規準化 - 位置パラメータ	457	- 逆推定値の比較	187	
基準 - 要因の取舍選択	232	- 逆数あるいは比	417	
疑心暗鬼 - 偏微分式が正しいのか	464	- 組合せ水準	129	
季節 - 洗浄水の温度	19	- グラフビルダー	356	
季節ごと - 回帰直線	195	- グラフ表示	94	
期待値の差 - 95%信頼区間	109	- 計算結果	458	
期待平均 - 効果の和	91	- 現行水準と最適水準の差	120, 113	
- 交互作用	91	- 原点を通る	279	
- セル平均	93	- 効果	73	
期待平均とセル平均の差 - 交互作用	93	- 合成分散の一般式	298	
拮抗阻害 - 共通の (V_{max}) 共通	430	- 効力比 ρ	472, 474	
- 山崎ら(2008)	428	- 個別データ	180, 224, 227	
逆推定/JMPによる - 芳賀(2016)	183	- ゴンペルツ曲線	387, 402	
逆推定/重み付き回帰 - 高橋(2013a)	183	- 最小2乗平均	298, 300	
逆に大きい場合 - 前値 X の差	322	- 最大反応 V_{max}	418	
逆行列 - Minverse () 関数	15, 125, 226	- 3次曲線	82	
- 対角行列	51	- 10パーセント点	390	
- 掃き出し演算子	126	- 水準の差	143	
- 掃き出し法	125	- 水準間の差の推定	132	

き 95%信頼区間 - 水準間の比較	64	き - 1元配置型	38
- 水準平均	63, 74, 94	- Excelの回帰分析	204, 268
- 水準平均の差	74	- 奥野ら(1981)	203
- 推定値	14	- 計算が難しい厄介な方法	294
- 推定値の差	17, 95	- 交互作用	30
- 切片の推定値	266	- 交互作用を考慮	195
- 線形モデル	342	- 最小2乗平均	297
- 線形和の推定	128	- GLMプロシージャ	195, 211
- 阻害定数 K_i	438	- スネデガーら(1972)	282, 294
- 阻害薬 A 14 nM	434	- 線形モデル	195
- 阻害薬 A なし	433	- 前後差 D	296
- ソルバー	36	- 高橋ら(1989)	260
- 対数効力比 ρ'	476	- ダミー変数	203
- 単回帰分析	86	- 難解なのか?	203
- 直線間の差	35	- 難解な解析法	195
- 同一直線上か	276	- 芳賀(2014)	7
- 2次式	83	- 平行な直線	480
- 2次式の解の公式	36	- 変化量の検討	30
- 2水準間の差	209	- 迷信	268
- パラメータに関する	344	- 妄想	280
- パラメータの差	345	- 乱塊法	47
- ひげの長さ	208	共分散分析と変化量の検討 - フライス(2004)	30
- 非線形回帰/ JMPの	459	共分散分析の適用 - 難点	196
- 平行な回帰直線	210	共変量 - 交絡変数	243
- ミカエリス・メンテン曲線	409, 411, 423, 425	- 交絡変数	268
- ミカエリス定数 K_m	419	- 前値	30, 38
- 要因効果	98	- 前値 X	294
- 予測プロファイル	146, 210, 351	- 前値の影響	248
- ロジスティック曲線	342, 346, 352, 458	- 胎仔数の影響	249
- ロジスティック曲線の差	347	共変量 x - 均一性	222
95%信頼区間の計算 - 残差線形化法	468	共変量/ 癖のある - 癖のある共変量	282
95%信頼区間の出力 - JMPの「二変量の関係」	191	共変量として前値 - 前後差	250
95%信頼区間の推定 - ゴンペルツ曲線	388	共変量の交互作用 - 中心化	255
95%信頼区間の直接推定 - 逆推定値	193	共変量の分布 - 均一	197
95%信頼区間の表示 - 個別データ	229	共変量を含む - 3因子実験データ	219, 230
95%信頼区間の幅 - $t_{0.05}(df) \times SE$	95	“行・列”の順番 - “列・行”ではなく	160
共分散行列/パラメータ - Collett(2003)	183	行/列の切り替え - 折れ線グラフ/Excel	248
共通の (V_{max})共通 - 拮抗阻害	430	行ベクトル d - 偏微分式	184
共通の傾き - 回帰の平方和	198	行方向に併合 - set ステートメント	462
- 各種の推定	205	行列の基本 - 矩形データ	155
- 3本の回帰直線	11	行列の計算 - 嫌悪感	155
共通の最大値 - そこそこの95%信頼区間	345	行列の積 - Mmult () 関数	226, 271
共通の切片 - MCP剤	49	行列の表記 - 太い外枠	157
共分散 - パラメータ	206	行列関数 - Excel	155
共分散行列 - パラメータ	127	- Mmult () 関数	15
- 誤差分散の推定値	40	行列計算 - Excel	7
- 線形和の分散	271	- 回帰パラメータの推定	15
- 対角要素	15, 41	- 拒絶反応	410
- デザイン行列 X	40	- 線形モデル	137
- 2次形式	271	- パラメータの推定	169
- パラメータ	14, 25, 50, 72	行列計算の基礎 - デザイン行列 X	157
- パラメータ	140, 172, 226, 457	行列計算の実際 $X\beta$ - Excelシート	159
共分散行列 $\Sigma(\beta^{\wedge})$ - パラメータ	295	行列計算の不具合 - 多重共線性	251
共分散行列 $\Sigma(\theta^{\wedge})$ - パラメータ(推定値)	351	行列式 - Mdetarm () 関数	169
共分散行列/パラメータ - 相関行列/パラメータ	422	行列表記 - 正規方程式	165
共分散分析 - アーミテージら(2001)	327	- Draperら(1989)	168
- 後値 Y	296	- ドレーパラ(1968)	168
- 1因子実験	230	行列表記に対応付け - 正規方程式	168

き	行/列の入れ替え - 転置/行列	15	群内の回帰 - アーミーテージら(2001)	327
	曲線のあてはめ - JMP	392	け (K_m) ₀ の2倍となる阻害薬Aの濃度 - 阻害定数 K_i	429
	拒絶反応 - 行列計算	410	計画行列 - Wikipedia	163
	「ε」に変換 - 「いぶしろん」	157	- デザイン行列	163
	ギリシャ文字 β - 「ベータ」と入力	157	- デザイン行列とは何か	18
	均一 - 共変量の分布	197	計算が容易でない - 有効反復数	76
	均一性 - 共変量 x	222	計算が難しい厄介な方法 - 共分散分析	294
	近似95%信頼区間 - デルタ法	184	計算ミス - Excel 2007 以前	178
く	偶然に有意な差 - 後値 Y	292	計算結果 - 95%信頼区間	458
	- 後値 Y	301	- 初期値	386
	- 前後差 D	290	計算公式 - 過度に標準化	86
	群分け - 前値	281	- 平方和	86
	矩形データ - Excel	55	計算公式の範囲内 - 回帰分析	168
	- 行列の基本	155	- ガラスの天井	168
	楠ら(1995) - 応用 実験計画法	121	計算式 - 予測値の標準誤差 y^{\wedge}	353
	- 逆行列の計算	126	計算式エディター - 微分した式	357
	- 欠測値がある直交表	138	計算式の設定ミス - Excel	53
	- 線形モデル・線形推定論	121	計算式の設定 - 予測値 y^{\wedge}	349
	- 2 因子実験のデータ	121	計算式の保存 - 予測信頼限界	396
	癖のある共変量 - 共変量/癖のある	282	計算式を含む - 予測値の95%信頼区間	352
	区別 - 要因	232	計算精度 - 倍精度実数	126
	組み合わせ - ルール	256	計算精度を保つ - 64ビット版	126
	組合せによる推定値 - 現行水準	146	計算方法がブラック・ボックス - 非線形回帰	417
	組合せ水準 - 95%信頼区間	129	経時的に測定 - 前値	281
	classステートメント - 質的変数	212	形状 - ワイブル分布	378
	グラフ・ビルダー - ゴンペルツ曲線	396	形状パラメータ - 位置パラメータ	340
	- JMP	19	- 勾配パラメータの逆数	457
	- 層別散布図	20	形状パラメータ σ - 位置パラメータ μ	367
	グラフィビルダー - 95%信頼区間	356	形状パラメータ σ_{GMP} - 位置パラメータ μ_{GMP}	375
	- 格子状に散布図	264	形状パラメータ σ_{SEV} - 位置パラメータ μ_{SEV}	377
	- スプライン曲線	334	啓蒙活動 - 繰り返しが不揃い	110
	- 予測値 y^{\wedge} の重ね合わせ	355	$K_m - V_{max}$ の2分の1となる基質濃度	407
	グラフ表示 - 逆推定	193	- ミカエリス定数	407
	- 95%信頼区間	94	K_m で偏微分 - ミカエリス・メンテン式	407
	繰り返しが2 - 2因子実験	89	血圧が高めの人たち - 臨床試験	281
	繰り返しが等しい - 1 因子実験	55	血圧の上昇 - 下限が存在	481
	- 高橋ら(1989)	89	- 上限が存在	481
	- 2 因子実験	89	血圧上昇 - アンジオテンシン I	469
	- 分散分析表	93	血圧上昇データ - 薬理学的活性の評価	469
	繰り返しが不揃い - 1 因子実験	69	結果が怪しい - Excel の回帰分析	53
	- 因子が直交しない場合	102	結果の表示 - 予測プロファイル	259
	- 啓蒙活動	110	欠測セルがある場合 - 最小2乗平均	148
	- JMP	102	欠測値 - 直交表	138
	- 線形モデル	102	欠測値がある場合 - 各種の推定	142
	- 高橋(2023)	102	欠測値がある直交表 - Excel の回帰分析	140
	- 2 因子実験	89, 102	- 楠ら(1995)	138
	- 芳賀(2014)	69	- SAS/GLM	150
	- 分散分析表	111	- 正規方程式の解	139
	繰り返し計算 - 残差線形化法	467	- 線形モデル	138
	苦し紛れの解釈 - 回帰直線	338	- 分散分析表	141
	Cressie-Keightley プロット - Direct linear plot	420	決定樹 - 解析手法の選択	263
	Cleland-Wilkinson 法 - 知名ら(2014)	420	Getting Started, PROC GLM - Unbalanced ANOVA	102
	黒子的な役割 - 最小極値分布	378	嫌悪感 - 行列の計算	155
	群間の差 - 切片の差	280	限界 - 分散の加法性	75
	群間比較 - 後値 Y	290	現業実習 - 冷蔵庫の塗装不良	219
	- 前後差 D	282	現行水準 - 組合せによる推定値	146
	- 前後差データ	281	- 最適水準	107
	- 第1種の過誤 (α エラー)	310	- 最適水準の差	143
	群間比較を適切 - 交互作用を含めたモデル	301	現行水準と最適水準の差 - 95%信頼区間	120, 133

け 現実的な対応 - Excel の回帰分析	181	こ - 分散分析表	21, 33
- 交絡変数	279	- 平方和	13
検出力 - $\rho=0.50$ の場合	324	- 変数減少法	230
検出力($1-\beta$) - ($1-\beta$) エラー	310	交互作用 $x \times A$ - GLM プロシジャ	212
- シミュレーション	324	交互作用あり非平行線 - 交互作用なし平行線	303
検出力($1-\beta$) の比較 - 時点間の相関を変化	325	交互作用がある場合 - 2本の直線の推定値	302
検出力($1-\beta$ エラー) - 各種の群間比較	318	交互作用がないこと - 前提条件	301
検出力が極端に減弱 - 前後差 D	322	交互作用とはどのようなものなのか? -	93
検出力に反映 - 分散の期待値の違い	325	交互作用なし平行線 - 交互作用あり非平行線	303
検出力の可視化 - $\rho=0.10$ の場合	321, 323	交互作用の解釈 - 光明を与える	308
- $\rho=0.90$ の場合	321, 323	交互作用の検討 - デザイン行列	251
検出力の計算手順 - Excel シート	319	交互作用の分離 - 残差平方和の差	331
検出力を可視化 - Excel シート	318	交互作用の平方和 - 回帰の平方和の差	33
検証/NLIN プロシジャ - 標準誤差	422	- 残差平方和の差	33, 330
減少効果/両群の - マイナス 10.0	311	交互作用プロット - 予測プロファイル	242
現代の算盤 - Excel の計算シート	410	交互作用プロファイル - JMP	242
検定統計量 - 傾きの差	268	交互作用モデル - 回帰分析	204
原典からの酵素反応速度論 - 知名ら(2014)	405	- 傾きの差	269
原点 - 回帰直線	9	- JMP	306
原点を共通 - 回帰直線	48, 52	- 主効果モデル	31, 204, 278, 330
- PCP 剤	51	- デザイン行列 X	32
原点を通る - 回帰直線	199, 266	交互作用を含む - 線形モデル	231
- 95%信頼区間	279	- 分散分析表	200
- 2本の回帰直線	277	- 変数選択	239
原点を通る回帰式 - 岩崎(2006)	53	交互作用を含む場合 - 最小2乗平均	304
原点共通 - Rのlm() 関数	54	交互作用を含めたモデル - 群間比較を適切	301
こ 抗うつ薬 - プラセボ群	157	交互作用を含める - 解析の妥当性	309
抗うつ薬投与前後 - フライス(2004)	30	交互作用を考慮 - 共分散分析	195
効果 - 95%信頼区間	73	- デザイン行列 X	195
- データの構造式	72	交互作用を考慮した共分散分析 - 高橋(2021)	19
- 予測プロファイル	75	交互作用を念頭に - 分野	219
効果 α_i - 予測プロファイル	62	考察は不十分 - 分散分析表だけ	120
効果 α_i の推定 - (1, -1)対比型	79	格子状 - 層別散布図行列	221
効果の分散 - 独立でない	76	格子状に散布図 - グラフビルダー	264
- 有効反復数	66	合成分散 - 一般式を拡張	131
効果の推定 - 線形和	108	- 拡張	26
- データの構造式	61, 62	- 差の分散	34
効果の和 - 期待平均	91	- 差 d^{\wedge} の分散	332
- データの構造式	92	- 推定値の分散	14
交互作用 - 回帰式	32	- 線形式	43
- 可視化	93	- 正しい分散の計算	99
- 期待平均	91	- 2次形式	178
- 期待平均とセル平均の差	93	- パラメータの共分散行列	330
- 共分散分析	30	- パラメータ間の差	345
- 残差の平方和	14	- パラメータの比	184
- 残差平方和ヘプーリング	200	合成分散の一般式 - 95%信頼区間	298
- 質的変数と量的変数	239	合成分散の一般式(デルタ法) - 知る人ぞ知る	427
- 自由度	13	酵素阻害実験データ - 量×量の2因子実験	429
- 図的表現	94	酵素反応速度論 - ヒルの式	372
- 成分表示	138	構造(DE)モデル - 推定	127
- 切片の差の推定値	225	- 線形モデル	135
- 線形モデル	224	- 母数に関する制約	122
- 前後差が拡大	247	構造モデル - 回帰モデル	121
- 層別因子を含む回帰分析	327	- 線形モデル	121
- ダミー変数	24, 204, 225	- DE モデル	121
- ダミー変数の積	96	構造式 - データの構造式	56
- 中心化	240, 250, 306	勾配/曲線の - パラメータ	457
- デザイン行列 X	301	勾配パラメータの逆数 - 形状パラメータ	457
- 複数の共変量	250	光明を与える - 交互作用の解釈	308

こ 交絡変数 - 共変量	243, 268	こ - NLINプロシジャ	399
- 現実的な対応	279	- オフセット	399
- 総摂餌量	244	- 各種の95%信頼区間	385
- 総摂餌量の影響	247	- 95%信頼区間	387, 402
- 多彩な影響	260	- グラフ・ビルダー	396
- 投与終了時の体重	260	- 様々な表記法	374
- 2群比較	243	- 10%点の推定	404
効力比 - 平行でない直線	469	- パラメータで偏微分	385
効力比 ρ - 95%信頼区間	472, 474	- パラメータの分散	389
- 残差線形化法	486	- 予測プロファイル	395
- 対数効力比 ρ'	484	- 累積バグ数の経時変化	374
- 定義	471	- ロジスティック曲線	379
効力比の95%信頼区間 - Excel	485	ゴンペルツ曲線の逆推定 - JMP	392
効力比の推定 - 平行なロジスティック曲線	481	- 非線形回帰	394
- 平行な直線	475	さ 催奇形性 - 生殖発生毒性試験	243
効力比の直接推定 - NLINプロシジャ	477	再計算/シート - 正規乱数	283
- ロジスティック曲線	481	再計算/自動的 - Excel	283
<i>Coefficient of Variation</i> - CV	285	最終モデル - 主効果+交互作用モデル	234
コーディングのテーブル - (1, -1)対比型	144	最小2乗法によるパラメータの推定 - デザイン行列	18
互換性・メッセージに注意 - Excel 2007 以前	178	(最小値から最大値)に拡張 - (0 から 1)の範囲	327
誤差分散 - 不偏推定量	126	最小化 - 残差平方和 S_e	461
誤差が不均一 - 逆数変換 (1/V)	407	最小2乗推定量 - 正規方程式	124
- 推定値が不安定	407	最小2乗平均 - lsmeans	142, 147, 297
誤差の伝播の式 - 逆数の標準誤差	427	- LSMEANS	218
誤差伝播の式 - 面積の標準偏差	189	- lsmeansステートメント	45, 46, 152, 216, 280
誤差伝播の式の一般化 - 分散共分散行列	189	- 95%信頼区間	298, 300
誤差の伝播 - 不安定な阻害定数 K_i の推定	446	- 共分散分析	297
誤差の平均平方 - RMSE	258	- 欠測セルがある場合	148
誤差の平方和 - SSE	258	- 交互作用を含む場合	304
誤差の平方和 Q - 偏微分	164	- 算術平均	259
誤差分散 - 残差の自由度	127	- GLMプロシジャ	217
- 残差平方和	127, 172	- 推定値	217
誤差分散の推定 - 予測値の推定	140	- 推定平均	43
誤差分散の推定値 - 共分散行列	40	- 線形和 L	142
- 残差の分散	226	- 線形和の推定値	271
50%反応量D50 - 2値データ	368	- 前後差 D	298
50パーセント致死量 - LD_{50}	367	- 中心化された場合	307
- プロビット法	367	- 非平行な直線	309
50パーセント反応 - 用量の推定	383	- モデルのあてはめ	148
個体内の揺らぎ - 時点間の相関	285	最小2乗平均の差 - estimateステートメント	218
個別 K_m V_{max} - SASプログラム	453	最小2乗平均の活用 - デザイン行列 X	102
個別データ - 逆推定値	187	最小2乗平均の図示 - 後値 Y および群間差	300
- 95%信頼区間	180, 224, 227	- 前後差 D および群間差	300
- 95%信頼区間の表示	229	最小2乗平均の対比 - JMP	308
個別データの95%信頼区間 - 回帰直線	224	最小2乗平均の例示 - 2本の回帰直線	298
- 管理限界	235	最小2乗法 - 回帰分析	164
個別の信頼限界の計算式 - JMP の回帰分析	190	- 残差平方和	123
個別パラメータ - 一括推定	451	最小に - 残差平方和 S_e	483
込みにした変数選択 - 質的変数と量的変数	241	最小の反応を θ_{min} - 最大の反応を θ_{max}	372
Collett (2003) - 共分散行列/パラメータ	183	最小極値分布 - 黒子的な役割	378
Correl () 関数 - Excel	264, 286	- Smallest Extreme Value Distribution	377
- 相関係数	264, 286	- ワイブル分布	376, 378
Gompertz 4P - 逆推定	393	最大の反応を θ_{max} - 最小の反応を θ_{min}	372
- パラメータの推定	393	最大極値分布 - ゴンペルツ曲線	374
- ロジスティック 4P	392	- Maximam Extreme Value Distribution	375
ゴンペルツ曲線 - 95%信頼区間の推定	388	最大値・最小値 - ロジスティック曲線	336
- 最大極値分布	374	最大反応 V_{max} - 95%信頼区間	418
		- 標準誤差 SE	418

さ 最大反応 V_{max} - 標準誤差 SE	418	さ - 平方和の計算	91
- ミカエリス定数 K_m	405	SumProduct () 関数 - Excel	161, 168
最大反応 V_{max} を共通 - ミカエリス定数	439, 440	さらなる改良 - <i>Scatchard</i> ・ <i>Eadie-Hofstee</i> 変換	416
最適解 - 初期値	483	3 因子モデル - 2 因子モデル	232
最適水準 - 現行水準	107	3 因子実験データ - 共変量を含む	219, 230
最適水準の差 - 現行水準	143	3 区分 - 前値 X での t 検定	315
最適水準の推定 - estimate ステートメント	152	- 前値 X/t 検定	324
最尤法で解く方法 - 反復重み付き回帰	369	残差 ε^{\wedge} を反応 - 回帰分析を繰り返す	468
最良線形不偏推定量 - BLUE	124	残差のベクトル ε^{\wedge} - 残差線形化法 δ^{\wedge}	449
寄山ら(2008) - 推定結果	458	残差の自由度 - 誤差分散	127
- 非線形モデル	455, 456	残差の分散 - 誤差分散の推定値	226
- 薬理試験における統計解析	455	残差の平方和 - 交互作用	14
作図のヒント - 折れ線グラフ	64	残差を反応 - 回帰分析の反復	363
作成手順 - 層別散布図	103	残差線形モデル - 非線形モデルの解法	488
佐久間(2017) - 平行線検定法	480	残差線形化法 - ガウス・ニュートン法	457
SAS - 一般線形モデル	45	- 95%信頼区間の計算	468
- NLIN プロシジャ	397, 461	- 繰り返し計算	467
- OnDemand SAS	44	- 効力比 ρ	486
- シミュレーション	315	- 推定精度の改善	487
- DATA ステップ	211, 421	- パラメータの推定	385
- テキストデータ	211	- パラメータ推定	363
- PROC ステップ	421	- ロジスティック曲線 / 4P	467
SAS と Excel の連携法 - Excel の線グラフ	361	- ロジスティック曲線のあてはめ	466
SAS の非線形 - NLIN プロシジャ	358	残差線形化法 δ^{\wedge} - 残差のベクトル ε^{\wedge}	449
SAS プログラム - 3 本のロジスティック曲線	359	残差平方和 - S_e	176
SAS/ OnDemand - 無償で継続的	461, 358	- 誤差分散	127, 172
SAS データセット化 - DATA ステップ	358	- 最小2乗法	123
SAS による実験データの解析 - 高橋ら(1989)	121	- 微小なパラメータ増減値	336
SAS の結果は正しいのか? - Excel で再現	404	残差平方和 Q - 偏微分	123
- JMP はデルタ法	404	残差平方和 S_e - 最小化	461
SAS Institute (2020) - NLIN プロシジャ	399	- 最小に	483
- 非線形モデル	358	残差平方和 S_e を最小に - Excel のソルバー	408
SAS/ OnDemand - 無償で継続的	477	残差平方和の差 - 交互作用の分離	331
SAS/GLM - 欠測値がある直交表	150	- 交互作用の平方和	33, 330
SAS データセット化 - ラットの子宮重量	400	残差平方和の最小化 - ソルバー	337
SAS データファイル - DATA ステップ	461	残差平方和へプーリング - 交互作用	200
- DATA ステップ	478	3 次曲線 - 95%信頼区間	82
SAS による検証 - Excel による非線形回帰	421	- 予測プロファイル	82
SAS の GLM - lsmeans ステートメント	280	3 次式のあてはめ - デザイン行列 X	81
SAS の GLM プロシジャ - Excel の回帰分析	115	3 種のシングモイド曲線 - 性能評価	380
- 無償で継続的	120	3 種の関数 - 正規分布	368
SAS の標準出力 - HTML 形式	359	算術平均 - デザイン行列	163
- HTML 形式	421	- 最小2乗平均	259
SAS プログラム - 個別 $K_m V_{max}$	453	3 通りに分ける - 有意差検定の p 値	316
- 阻害定数 K_i の直接推定	444	3 倍強の違い - ミカエリス定数の SE	419
- 追加データファイル	423	2×2 格子状の散布図 - 複数の共変量	252
- ミカエリス・メンテン式	422	散布図 - Excel	56
雑草重量 - 被覆度	47	- 回帰直線	28
差の95%信頼区間 - 平行な回帰直線	201	- 確率楕円	264
差の t 検定 - 対象群と処置群	245	- 2×2 区分	254
差 d^{\wedge} の分散 - 合成分散の一般式	332	散布図から復元 - <i>Armitage</i> ら(2002)	407
差の分散 - 合成分散の一般式	34	散布図上に重ね書き - ロジスティック曲線	338, 347
- 分散の加法性	474	層別散布図行列 - 格子状	221
Substrate concentration - 基質濃度 $[S]$	407	3 本のロジスティック曲線 - SAS プログラム	359
差分 - 回帰の平方和	107	- JMP	349
様々な表記法 - ゴンペルツ曲線	374	3 本の回帰直線 - 重ね書き	329
SumSq () 関数 - Excel	91, 161	- 共通の傾き	11
- 平方和	57	- 切片が異なる	12

さ	- デザイン行列 X	10, 328	し	時点間の相関 - 1.0 に近づく	313
し	GLMプロシジャ - (1, 0)型	150, 213		- 個体内の揺らぎ	285
	- 一般線形モデル	45		- 変化させた場合	316
	- estimateステートメント	120		- 0.0 に近づく	314
	- OnDemand SAS	150		時点間の相関が 0.5 - 実質の α エラー	315
	- 共分散分析	195, 211		時点間の相関を変化 - 検出力(1- β) の比較	325
	- 交互作用 $x \times A$	212		時点間相関を変化 - 実質の α エラー	317
	- 最小2乗平均	217		自動的 - 変数選択	237
	- タイプ I の平方和	116		自動的に行う機能 - パラメータで偏微分	399
	- タイプ I, II, III, IVの平方和	106		自動的に偏微分 - 統計ソフト	342
	- タイプ II の平方和	116		自動微分 - 非線形回帰/ 統計ソフト	351
	- タイプ III の平方和	116		自動偏微分 - NLINプロシジャ	358
	- 高橋ら(1989)	150		シミュレーション - 検出力(1- β)	324
	- デザイン行列 X	212		- SAS	315
	- パラメータの推定値	213		- 条件設定	310
	- 分散分析表	116		- 第1種の過誤	315
	- 無償で継続的	120		邪魔された分 - 位置を変える	398
	GLMプロシジャの計算方式 - 高橋ら(1989)	153		邪魔をする - オフセット値 η 分布 $\pi=0.10$	398
	GLMプロシジャの出力 - HTML形式	151		Sham群 - 陽性対照群	455
	GLMプロシジャの世界だけ - 理解が上滑り	117		JMP - 3本のロジスティック曲線	349
	恣意性を排除 - 逐次平方和	111		- 曲線のあてはめ	392
	- 変数の自動選択	237		- グラフ・ビルダー	19
	シート再計算 - 数式タブ	283		- 繰り返しが不揃い	102
	CV - Coefficient of Variation	285		- 交互作用プロファイル	242
	- 変動係数	285		- 交互作用モデル	306
	GUI による偏微分する機能 - JMP	357		- ゴンペルツ曲線の逆推定	392
	Generalized Linear Model - Dobson ら(2018)	369		- 最小2乗平均の対比	308
	$\Sigma(\beta^{\wedge})$ - パラメータの共分散行列	178		- GUI による偏微分する機能	357
	$\Sigma(\theta^{\wedge})$ - $(X^T X)^{-1} \sigma^{\wedge 2}$	109		- 主効果モデル	305
	- パラメータの共分散行列	109, 206, 227		- ステップワイズ法	256
	$\Sigma(\theta^{\wedge})$ の対角要素 - パラメータの分散	408		- 対比の設定	23
	σ_{LGS} - σ_{NOR} の 0.5513倍	370		- 探索的な回帰分析	255
	σ_{NOR} の 0.5513倍 - σ_{LGS}	370		- 逐次(タイプ1)の平方和	88
	シグマ - 読み飛ばしたくなる	155		- 二変量の関係	266
	シグマを用いた - 積和の計算	161		- パラメータに関する偏微分	356
	シグモイド曲線 - 陰性および陽性対照	379		- 非線形回帰	350, 458
	- 逆推定	367		- 分散分析表	37
	- 山田ら(2003)	379		- Vec Quadratic () 関数	191, 354
	- 累積分布関数	367		- 変数選択	256
	実験計画法 - 直交表	135		- モデルのあてはめ	20, 144
	実験計画法 改訂版 - 芳賀(2014)	55		- 有償ソフト	149
	実験計画法 - 伝統的	55		- 予測プロファイル	21, 242
	実際の α エラー - 名目の $\alpha=0.05$	311		JMP の「モデルのあてはめ」 - 分散分析表	145
	実質の α エラー - 時点間の相関が 0.5	315		JMP の「二変量の関係」-95%信頼区間の出力	191
	- 時点間相関を変化	317		JMP の解析用ファイル - (1, -1)対比型	144
	実質の α エラーが増大 - 前後差 D	314		JMP の回帰分析 - 逆推定値の95%信頼区間	190
	質的因子 - 芳賀(2014)	55		- 個別の信頼限界の計算式	190
	質的な変数に限定 - 平方和の分解	135		- 二変量のあてはめ	190
	質的因子 - 1 因子実験	55		- 平均の信頼限界の計算式	190
	質的変数 - classステートメント	212		JMP の使い勝手 - 画期的	149
	- 線形モデル	24		JMPの逆推定のSE - 逆推定式を偏微分	402
	- ダミー変数	55		JMPの自動偏微分 - 正しい偏微分式	465
	質的変数と量的変数 - 交互作用	239		JMPの非線形回帰 - 標準誤差	350
	- 込みにした変数選択	241		JMPはデルタ法 - SASの結果は正しいのか?	404
	質的変数のまま - データリスト	240		収束しない原因 - 元の偏微分式にミス	466
	質的変数を - 量的変数に	55		修正項CT - 平方和の計算	92, 93
	実薬 - プラセボ	281		重回帰分析 - ドレーパら(1968)	156
	時点間 - 相関	311		重層的 - 回帰分析の適用	197
				重層的に適用 - 単回帰分析	195

し 自由度 - 交互作用	13	す 水準平均の分散 - 分散の加法性	65
- 分散分析表	177	推定 - 構造(DE)モデル	127
主効果+交互作用モデル - 最終モデル	234	推定されたパラメータ - 回帰直線	234
- 2 因子モデル	233	推定に推定を重ねる線形変換 - 推定精度の劣化	438
主効果・交互作用 - データの構造式	90	推定曲線を重ね書 - 年齢 x が小さい順	338
主効果と交互作用を平等 - タイプ III の平方和	113	推定結果 - 崚山ら(2008)	458
- タイプ III の平方和	114	推定結果との比較 - 阻害定数 K_i	446
主効果に重き - 分野	219	推定精度の改善 - 残差線形化法	487
主効果モデル - 回帰モデル	204	推定精度の劣化 - 推定に推定を重ねる線形変換	438
- 交互作用モデル	31, 204, 274, 330	推定値 - estimateステートメント	46
- JMP	305	- 各種の線形和 L	206
主効果を主体 - タイプ II の平方和	111, 113	- 95%信頼区間	14
種々の関連 - 臓器重量と体重	260	- 最小2乗平均	216, 217
取水路・放水路 - 水深	38	- 線形和 L	26, 44
主たる共変量 - 補助の共変量	252	- 分散・共分散	61
出力ファイルでの変数名 - 統計量に関するキーワード	463	推定値 V^{\wedge} - 反応速度	410
10%点の推定 - ギンペルツ曲線	404	推定値 V^{\wedge} の分散 - 反応速度	410
10パーセント反応 - 逆推定	381	推定値 y^{\wedge} の分散の分散 - 2次形式	179
- 用量の推定	383	推定値が一定しない - 線形変換	454
10パーセント点 - 95%信頼区間	390	推定値が不安定 - 誤差が不均一	407
- ソルバー	390	- Lineweaver-Burk 変換	413
寿命データの解析 - ネルソン(1988)	378	推定値に若干の差異 - Armitageら(2002)	420
寿命試験データの統計解析 - 高橋(2015)	378	推定値の分散 - 合成分散の一般式	14
上限および下限 - 管理限界	221	推定値の差 - 95%信頼区間	17, 95
上限が存在 - 血圧の上昇	481	推定値の出力要求 - solutionオプション	212
上限と下限 - 平均への回帰	31	推定平均 - 最小2乗平均	43
条件設定 - シミュレーション	310	推定方法 - 阻害定数 K_i	430
初期パラメータ - ロジスティック曲線	341	推定方法による揺らぎ - 阻害定数 K_i	442
初期値 - exp () 関数の中の線形式	340	数式タブ - シート再計算	283
- 計算結果	386	数値に基づく対策 - 統計的分析	229
- 最適解	483	数値例 - ガウス・ニュートン法	366
- 2 本のロジスティック曲線	482	Scatchard・Eadie-Hofstee 変換 - 回帰直線の95%信頼区間	417
初期値を最小化 - ソルバー	341	- さらなる改良	416
初期値を設定 - parms ステートメント	479	- パラメータ推定	416
除草剤 - 成長抑制	47	Scatchard 変換 - Eadie-Hofstee 変換	406
- 用量反応関係	48	- 線形化	406
除草剤の効果 - 奥野ら(1978)	47	Stdev.S () 関数 - 標準偏差	283
処置前 - 処置後 - 岩崎(2002)	290	図的表現 - 交互作用	94
飼料効率 - 体重の増加抑制	246	ステップワイズ法 - JMP	256
知る人ぞ知る - 合成分散の一般式(デルタ法)	427	- (1, -1) 対比型	239
新村(1983) - 掃き出し演算子	126	- 手作業	257
水深 - 取水路・放水路	38	- 手作業による変数減少法	237
す 水準ごと - 回帰の平方和	198	- 変数選択	255
- 回帰分析の併合	197	- 変数増加・減少・増減	255
水準ごとの切片 - 平行線	201	- 名義尺度, 順序尺度	256
水準の差 - 95%信頼区間	143	- モデルのあてはめ	256
水準の差の推定 - estimateステートメント	152	- モデルの作成	257
- カスタム検定	146	スナップショット - 線グラフと散布図	287
水準の平均の差 - 分散の加法性	68	- $p=0.105$	311
水準間の差 - 対比	22	スネデガーら(1972) - 共分散分析	282, 294
- 予測プロファイル	28, 75, 209	- 統計的方法 原著 第6版	282
水準間の差の推定 - 95%信頼区間	132	スプライン曲線 - グラフビルダー	334
水準間の比較 - 95%信頼区間	64	- 直線のあてはめの妥当性	334
水準効果 - 分散の推定	66	すべての組み合わせ - マクロ	240
水準平均 - 95%信頼区間	63, 74, 94	スムージング/ Excel の - 平滑線	456
水準平均の差 - 95%信頼区間	74	Smallest Extreme Value Distribution - 最小極値分布	377
- (0, 1) 型	79	3 パラメータモデル - 4 パラメータモデル	372
- 分散	65	Slope () 関数 - Excel	168

せ 正確な95%信頼区間 - Excel ソルバー	188	せ セル平均モデル - (1, 1) 標示型ダミー変数	68
- 逆推定値	185, 187	- cell means model	68
- 2次式の解の公式	186	cell means model - セル平均モデル	68
正規分布 - 3種の関数	368	- One-way ANOVA	163
- Norm.dist () 関数	367	(0, 1)型 - (1, 0)型	153
- ロジスティック分布	384	- (1, -1)基本型	148
正規分布の σ_{NOR} - ロジスティック分布 σ_{LGS}	370	- (1, -1)対比型	77, 230, 305
正規分布の確率密度関数 - $f_{NOR}(x)$	368	- 回帰分析	78
正規分布の逆関数 - Norm.Inv () 関数	283	- 水準平均の差	79
正規分布の累積分布関数 - $F_{NOR}(x)$	368	- ダミー変数	10, 59, 475
正規方程式 - 行列表記	165	- パラメータ推定	78
- 行列表記に対応付け	168	(0, 1)型か - (1, 0)型か	58
- 最小2乗推定量	124	(0, 1)型ダミー変数 - 1 因子実験	77
- 線形モデル	123	- 線形モデル	224
- デザイン行列 X	124, 125, 139	- デザイン行列 X	204
正規方程式の解 - 欠測値がある直交表	139	(0 から 1) の範囲 - (最小値から最大値) に拡張	327
- 手計算時代	126	線グラフ - 前値と後値	247
- デザイン行列 X	125	- ひげ付き	201
- パラメータの推定	165	線グラフと散布図 - スナップショット	287
正規乱数 - Excel	283	線形ではない - パラメータに関して	478
- 再計算/シート	283	線形モデル - L_8 直交表	135
- Norm.Inv (Rand (), 0, 1) 関数	286	- 回帰モデル	121
- 分析ツールの乱数	283	- 95%信頼区間	342
制御因子 - 対数用量 x	480	- 共分散分析	195
正攻法/設定 - パラメータの初期値	340	- 行列計算	137
生殖発生毒性試験 - 催奇形性	243	- 繰り返しが不揃い	102
成長期 - 実験動物	243	- 欠測値がある直交表	138
成長抑制 - 除草剤	47	- 交互作用	224
性能評価 - 3種のシグモイド曲線	380	- 交互作用を含む	231
成分表示 - 交互作用	138	- 構造 (DE) モデル	135
制約条件 - 1 次従属	136	- 構造モデル	121
- データの構造式	59, 90, 103	- 質的変数	24
積 $X\beta$ - 「積和」の計算	158	- 正規方程式	123
積の分散 - アーミテージら(2001)	184	- (0, 1)型ダミー変数	224
「積和」の計算 - 積 $X\beta$	158	- (1, -1)対比型ダミー変数	105
積和 - デザイン行列 X	162	- ダミー変数	55, 58, 96
積和 $l\theta^{\wedge}$ - 線形和 L	227	- デザイン行列 X	124
積和 XY - デザイン行列 X と Y	162	- パラメータ(係数)	122
積和の計算 - 内側の数が一致	160	- LinEst () 関数	60
- シグマを用いた	161	線形モデル・線形推定論 - 楠ら(1995)	121
積和行列 - 対角行列	51	線形モデルではない - 非線形モデル	336
積和行列 ($Z^T Z$) - 微係数行列 Z	448	線形モデルと非線形モデル - 解法の対比	487
- 微係数行列 Z	466	線形モデルを拡張 - 高橋ら(1989)	121
積和行列の逆行列 ($Z^T Z$) ⁻¹ - 微係数行列 Z	343	線形化 - Scatchard・Eadie-Hofstee 変換	406
積和行列の逆数 - Excel	344	- パラメータ	370
設置場所 - カキの成長	39	- パラメータの推定精度	405, 417
set ステートメント - 行方向に併合	462	- パラメータ推定の脆弱性	405
- DATAステップ	462	- Hanes-Woolf 変換	406
切片の扱い - 定数に 0 を使用	158	- ミカエリス・メンテン式	405
切片 β^{\wedge}_0 の逆数 - β^{\wedge}_0 で偏微分	412	線形化にこだわる - 悪あがき	419
切片が異なる - 3 本の回帰直線	12	線形近似 - Excel	8
- 平行な直線	475, 476	- 層別散布図/Excel	328
切片の差 - 群間の差	280	線形式 - 合成分散の一般式	43
切片の差の推定値 - 交互作用	225	- ロジット	335
切片の推定 - 線形和 L	207	線形推定論 - Excel による行列計算	123
切片の推定値 - 95%信頼区間	266	- 理論的な枠組み	121
切片の推定値は? - (1, 0)型	152	線形変換 - 推定値が一定しない	454
切片を共通 - 回帰直線	48	線形変換で悩む - Excel のソルバー/非線形回帰	454
セル平均 - 期待平均	93	- 終わりにしませんか	454

せ 線形和 - 効果の推定	108	せ 前値 X を共変量 - 後値 Y	297
線形和 L - 最小2乗平均	142	- α エラーが保持	316
- 推定値	26, 44	- 一定の検出力	326
- 積和 10^{\wedge}	227	- Excel の回帰分析	295
- 切片の推定	207	- 前後差 D	297
- 2次形式	299	前値 X/t 検定 - 3 区分	324
- 2水準間の差	29	前値 X での t 検定 - 3 区分	315
- 分散	26, 41, 227	- p 値の比較	315
線形和 L に対する分散 - 2 次形式	128	前値 X の揺らぎ - 可視化	285
線形和 L の分散 - 2 次形式	206	前値 X を共変量 - 名目の α エラー 5%	317
線形和の推定 - 95%信頼区間	128	前値と後値 - 線グラフ	247
線形和の推定値 - 最小2乗平均	271	- 平均への回帰	281
線形和の分散 - 共分散行列	271	前値の影 - 前後差データ	281
前後差 - 共変量として前値	250	前値の影響 - 共変量	248
- 増加率	246	前値を共変量 - 回帰分析	281
- 平行な線グラフ	247	前提条件 - 交互作用がないこと	301
前後差 D - 後値 Y	290	全データ - 回帰分析	197
- 共分散分析	296	そ 総当たり法 - 変数選択	237
- 偶然に有意な差	290	相関 - 時点間	311
- 群間比較	282	相関があるか - X と Y	264
- 検出力が極端に減弱	322	相関の期待値 - $\rho=0.10$	321, 323
- 最小2乗平均	298	- $\rho=0.90$	320
- 実質の α エラーが増大	314	相関係数 - Correl () 関数	264, 286
- 前値 X を共変量	297	相関行列 - パラメータ(推定値)	351
- t 検定	282	相関行列/パラメータ - 共分散行列/パラメータ	422
- 平均への回帰	283	増加率 - 扇型線グラフ	247
- 有意な差	291	- 前後差	246
前後差 D $p=0.023$ - 後値 Y $p=0.404$	313	臆器重量の有意差 - 体重の有意差	262
前後差 D $p=0.640$ - 後値 Y $p=0.020$	313	臓器の絶対重量 - 体重比(相対重量)	260
前後差 D $p=0.651$ - 後値 Y $p=0.615$	314	臓器重量 - 解析のための決定樹	280
前後差 D および群間差 - 最小2乗平均の図示	300	- 体重	260
前後差 D の分散 - インフレーション	287	臓器重量と体重 - 種々の関連	260
前後差が拡大 - 交互作用	247	- 常に探索的	264
前後差か増加率か - 判断基準	247	- 6 通りの典型例	261
前後差データ - 群間比較	281	臓器重量と体重の関係 - 高橋ら(1989)	260
- 前値の影	281	- 吉村編著(1987)	260
前後差の SD - 大小関係	289	臓器重量の比較 - 体重の総平均	270
前後差の相関が 0 - 分散が 2 倍	288	操作ミス - Excel	53
洗浄水の温度 x - 予測プロファイル	210	総摂餌量 - 交絡変数	244
洗浄用水の温度 - 回収液の濃度	19	- 体重の増加抑制	243
- 季節	19	- 妊娠18日の体重	243
選択されたモデル - 変数増減法	239	総摂餌量の影響 - 交絡変数	247
選択された変数 - モデルのあてはめ	241	相対的な比較検討 - 用量反応関係	470
全体の平方和 S_T - S_e と S_R に分解	182	総平均 - 平均の平均	71
全体の偏差平方和 - S_T	176	総平方和 - 平方和の分解	92
前値 - 共変量	30, 38	層別因子 - ロジスティック曲線	327
- 群分け	281	層別因子を含む回帰分析 - 奥野ら(1981)	19
- 経時的に測定	281	- 交互作用	327
- 2元配置	38	層別散布図 - Excel	8, 196
- 妊娠 7 日の体重	244	- 回帰直線	219
前値 X - A_1 群の平均値がやや大きい	312	- 重ね書き	103
- 共変量	294	- グラフ・ビルダー	20
- 小さい場合の検出力	322	- 作成手順	103
- 平均への回帰	30	- 2×2 の格子状	220
- ほぼ等しい	314	- 併合	223
前値 X の差 - 大きい場合の検出力	319	層別散布図/Excel - 線形近似	328
- 逆に大きい場合	322	阻害定数 $(K_i)_i$ - 両逆数・Lineweaver-Burk 変換	446
前値 X の大小関係 - α エラーが変化	316	阻害定数 $(K_i)'_i$ - V_{max} を共通	441

そ 阻害定数 K_i - 簡単な数値を与えて例示	430	た 対角要素 - 共分散行列	15, 41
- 95%信頼区間	438	- 分散	15
- $(Km)_0$ の2倍となる阻害薬Aの濃度	429	対角要素の平方根 - 標準誤差	226
- 推定結果との比較	446	第三者 - ブライド化	282
- 推定方法	430	対照群 - 2元配置	38
- 推定方法による揺らぎ	442	胎仔数の影響 - 共変量	249
- 濃度 $[I]$ を逆推定	430	体重 - 臓器重量	260
- 揺らぎ	442	体重の総平均 - 臓器重量の比較	270
- 揺らぎの実際	435	- 直線上の推定値	275
阻害定数 K_i が不安定 - V_{max} を共通	442	体重の増加抑制 - 飼料効率	246
阻害定数 K_i の推定 - Excel の計算シート	448	- 総摂餌量	243
- 傾き $(Km)_i / (Vmax)_i$	436	- 吉村編著(1987)	246
- 2段階の線形化	444	体重の有意差 - 臓器重量の有意差	262
- 2段階目	432	体重は増加量でみるべきか - 吉村編著(1987)	243
- ミカエリス・メンテン式	428	体重を交絡変数 - 回帰分析	268
阻害定数 K_i の直接推定 - Excel による非線形回帰	447	体重増加データ - 妊娠ラット	244
- Excel のソルバー	450	体重比(相対重量) - 臓器の絶対重量	260
- NLINプロシジャ	444	対象群と処置群 - 差の t 検定	245
- SASプログラム	444	大小関係 - 前後差の SD	289
- 非線形回帰	443	対数効力比 ρ' - 95%信頼区間	476
阻害定数 K_i の比 - 2種類の推定方法	428	- 効力比 ρ	484
阻害定数 K_i の比較 - 各種の推定方法	442	- 直接推定	478
阻害定数 K_i を用いた式 - ミカエリス定数 $(K_m)_i$	431	- パラメータ	478
阻害定数の推定 - 非線形回帰	428	- パラメータで偏微分	475
- 山崎ら(2008)	428	- $\log_{10} \rho$	479
阻害薬 A 14 nM - 95%信頼区間	434	対数効力比 ρ' の分散 - 2次形式	477
- ミカエリス定数の推定	434	対数変換した濃度(用量) - 濃度がゼロ	455
阻害薬 A なし - 95%信頼区間	433	対数目盛 - Excel のX軸の書式設定	410
- ミカエリス定数の推定	432	- 基質濃度 $[S]$	373
阻害薬Aの濃度 $[I]_i$ ごと - ミカエリス定数 $(Km)_i$	445	対数用量 x - 制御因子	480
阻害薬Aの濃度別 - (1, 1)標示型	439	- 用量反応直線	471
阻害薬濃度 $[I]$ に関する回帰 - 傾き $(Km)_i / (Vmax)_i$	437	(1, -1)対比型 - ダミー変数	10
阻害薬の酵素反応 - 4種類の阻害様式	429	- ステップワイズ法	239
阻害薬の濃度別 - $(Vmax)_i$ と $(Km)_i$ を推定	451	- (0, 1)型	77
阻害薬濃度 $[I]_i$ ごと - ミカエリス・メンテン曲線	441	- ダミー変数	23, 59, 71, 96
阻害様式 - 確認のための手順	451	- データの構造式	59
第11回 続・高橋セミナー - 高橋(2023)	110	(1, -1)対比型ダミー変数 - 線形モデル	105
そこそこの95%信頼区間 - 共通の最大値	345	対比 - 水準間の差	22
solutionオプション - 推定値の出力要求	212	- タイプ III の計算	117
ソルバー - 管理限界	228	対比の設定 - JMP	23
- 95%信頼区間	36	- パラメータ関数	23
- 上側95%信頼区間が「0」	334	対比型ダミー変数 - デザイン行列 X	60
- 回帰分析/ Excel	341	- パラメータ推定	25
- 残差平方和の最小化	337	対比型デザイン行列 X - 回帰分析	97
- 10パーセント点	390	タイプ I の平方和 - LOF解析	88
- 初期値を最小化	341	- GLMプロシジャ	116
- ミカエリス・メンテン式	341	- 逐次平方和	88, 106
- 目標値	341	- デフォルト	151
ソルバー/ Excel - パラメータ推定	482	- 分散分析表	107
- 変数セルの変更	460	タイプ I 平方和 - 逐次平方和	37
- ロジスティック曲線	481	タイプ I, II, III - SS1 SS2 SS3	212
ソルバーの活用法 - Excel	341	- 平方和の比較	216
た 第1種の過誤 - α エラー	310	タイプ I, II, III, IVの平方和 - GLMプロシジャ	106
- シミュレーション	315	タイプ I, タイプ II, タイプ III の平方和	111
第1種の過誤(α エラー) - 群間比較	310	- 回帰分析	115
対応のない t 検定 - 同一用量	469	タイプ II の平方和の合計は総平方和 S_T に一致しない	113
対角行列 - 逆行列	51	タイプ II の平方和 - 主効果から因子Aを除く	112
- 積和行列	51	- 主効果から因子Bを除く	112
		- GLMプロシジャ	116

た	- 主効果を主体	111, 113	- (1, 1)標示型	10, 49, 51, 439, 457, 472
	- フルモデルからA×Bを除く	112	- (1, 0)型	10, 39
	- 分散分析表	113	- 各種	10
タイプ III の平方和の合計は 総平方和 S_T に一致しない		114	- 共分散分析	203
タイプ III の計算 - 対比		117	- 交互作用	24, 204, 225
タイプ III の平方和 - GLMプロシジャ		116	- 質的変数	55
- 主効果と交互作用を平等	113, 114		- (0, 1)型	10, 59, 475
- 使い方	119		- 線形モデル	55, 58, 69
- デフォルト	151		- (1, -1)対比型	10, 23, 59, 71, 96, 256
- フルモデルからA×Bを除く	113		- 多項式型	81
- フルモデルからAを除く	113		- 非線形の解析モデル	457
- フルモデルからBを除く	113		ダミー変数の積 - 交互作用	96
タイプ III の平方和算出 - 分散分析表	114		ダミー変数の選択 - 解析の目的に合わせた	79
Times New Roman - 半角英数字のフォント	157		多様な揺らぎ - 長期的には安定	281
Direct linear plot - Armitageら(2002)	420		単位行列 - 逆行列を前掛	169
- Cressie-Keightley プロット	420		単回帰分析 - 岩崎(2006)	53
互いに独立 - 分散の加法性	96		- Excel	86
高橋ら(1989) - 因子が直交しない場合	102		- 95%信頼区間	86
- 共分散分析	260		- 重層的に適用	195
- 繰り返しが等しい	89		- デザイン行列 X	9, 85
- SASによる実験データの解析	121		段階的に適用 - 2 群間の t 検定	245
- GLMプロシジャ	150		探索的な解析 - 回帰分析	230
- GLMプロシジャの計算方式	153		探索的な回帰分析 - JMP	255
- 線形モデルを拡張	121		- Excel	250
- 臓器重量と体重の関係	260		探索的な推定 - パラメータの初期値	340
- 4 種の平方和	153	ち	小さい場合の検出力 - 前値 X	322
- 4 種の平方和とLSMEANS	89		CHANGE vs ANCOVA - Frison, Pocock (1992)	294
高橋(2012) - 投与前値の調整	310		逐次(タイプ1)の平方和 - JMP	88
- 平均への回帰	310		逐次的な回帰分析 - 平方和の増加	205
高橋(2013a) - 逆推定/重み付き回帰	183		逐次平方和 - 恣意性を排除	111
高橋(2013b) - 回帰分析・再入門	183		- タイプ I の平方和	88, 106
高橋(2015) - 寿命試験データの統計解析	378		- タイプ I 平方和	37
高橋(2017) - 反復重み付き回帰	371		- 分散分析表	205
- プロビット法・ロジット法・補2重対数法	369		- 平方和の差分	106
高橋(2021) - 交互作用を考慮した共分散分析	19		知名ら(2014) - Cleland-Wilkinson 法	420
高橋(2023) - 繰り返しが不揃い	102		- 原典からの酵素反応速度論	405
- 第11回 続・高橋セミナー	110		- Tseng-Hsu 法	420
高めの濃度を便宜的に - 陽性対照群	455		- 非線形回帰	405
田口の式 - 伝統的な実験計画法	131		- 歴史的背景	405
- 無視しない要因の自由度	131		中心化 - 回帰式が複雑	306
- 有効反復数	66		- 画期的な解析方法	309
- 有効反復数 n_e	134		- 共変量の交互作用	255
竹内(1963) - 逆推定値の95%信頼区間	183		- 交互作用	240, 250, 306
竹内ら(1989) - 統計学辞典	243		- 多重共線性	240
- 毒性評価	243		中心化された場合 - 最小2乗平均	307
多項式の中心化 - モデルの設定	21		中心膜厚 - 管理限界	219
多項式回帰 - 1 因子実験	80		- 不良の発生	220
- LOF解析	80		長期的には安定 - 多様な揺らぎ	281
- デザイン行列 X	80		直接作用 - 間接的な作用	245
多項式型 - ダミー変数	81		直接推定 - 対数効力比 ρ'	478
多彩な影響 - 交絡変数	260		直線のあてはめ - 便宜的	481
多重共線性 - 行列計算の不具合	251		直線のあてはめの妥当性 - スプライン曲線	334
- 中心化	240		直線間の差 - 95%信頼区間	35
多重比較 - 数多くの変数	243		直線関係か否か - LOF解析	80
正しい分散の計算 - 合成分散の一般式	99		直線上の推定値 - 体重の総平均	275
正しい偏微分式 - JMPの自動偏微分	465		直線的 - 用量反応	470
他の変数を定数 - 偏微分 β_0	164		直接推定 - NLINの逆推定のSE	402
多変量解析法 - 奥野ら(1981)	19		直交性を保ちたい - 補完	145
ダミー変数 - (1, 1)標示型	277, 336			

ち 直交表 - 欠測値	138	て	- デザイン行列とは何か	18
- 実験計画法	135	Design of Experiment - DE モデル		121
直交表の歴史的背景 - 朝香ら(1988)	148	デザイン行列 - 回帰分析		7
つ 追加データファイル - SASプログラム	423	- 計画行列		163
2 パラメータ - ロジスティック曲線	410	- 交互作用の検討		251
使い方 - タイプ III の平方和	119	- 最小2乗法によるパラメータの推定		18
繋げる配慮 - 標準最小2乗法	239	- 算術平均		163
常に探索的 - 臓器重量と体重	264	- 定義		163
て 定義 - 効力比 ρ	471	- 統計ソフト		7
- デザイン行列	163	デザイン行列 X - 因子が直交		101
t 検定 - 回帰パラメータ	176	- Excel の回帰分析		296
- 回帰直線の傾きの差	330	- L_8 直交表		137
- 傾きが異なる回帰直線の差	333	- 回帰分析	71, 105, 225	
- 前後差 D	282	- カスタム検定		147
停止ルール - 閾値 p 値	256	- 共分散行列		40
定数に 0 を使用 - Excel の回帰分析	9	- 行列計算の基礎		157
- off	40, 29	- 交互作用		301
- オフ(デフォルト)	32	- 交互作用モデル		32
- on	24, 199	- 交互作用を考慮		195
- オン	9, 51	- 最小2乗平均の活用		102
- 回帰分析/ Excel	472	- 3 次式のあてはめ		81
- 切片の扱い	158	- 3 本の回帰直線	10, 328	
定数に 0 を使用 - オフ	59	- GLM プロシジャ		212
Mdetarm () 関数 - Excel	169	- 正規方程式	124, 125, 139	
$t_{0.05}(df) \times SE$ - 95%信頼区間の幅	95	- 正規方程式の解		125
T.test () 関数 - Excel	245	- 積和		162
T.dist.2T () 関数 - 両側 t 分布	284	- (0, 1) 型ダミー変数		204
- t 分布の両側確率	175	- 線形モデル		124
t 分布の両側確率 - T.dist.2T ()	175	- 対比型ダミー変数		60
デフォルト - タイプ I の平方和	151	- 多項式回帰		80
- タイプ III の平方和	151	- 単回帰分析	9, 85	
デフォルトで出力 - パラメータの相関行列	360	- データの構造式		68
DE モデル - 構造モデル	121	- 転置と積和		160
- Design of Experiment	121	- ドレーパラ(1968)		156
DATA ステップ - output ステートメント	400	- 2 次式のあてはめ		82
- input ステートメント	400	- パラメータの共分散行列	61, 173	
- SAS	211, 421	- パラメータの共分散行列 $\Sigma (\beta^{\wedge})$		299
- SAS データセット化	358, 461, 478	- 微係数行列 Z	342, 364, 486	
- set ステートメント	462	- 変幻自在		18
- do ステートメント	211, 400, 444	- 偏差平方和		170
データの構造式 - L_8 直交表	135	- 偏微分行列 Z		194
- 効果	72	デザイン行列 X と Y - 積和 XY		162
- 効果の推定	61, 62	デザイン行列 X の一部 - 回帰分析		101
- 効果の和	92	デザイン行列 X を入力 - Excel シート		157
- 構造式	56	デザイン行列とは何か - 計画行列		18
- 主効果・交互作用	90	- Design Matrix		18
- 制約条件	59, 90, 103	デザイン行列の変数 - 別名		209
- (1, -1) 対比型	59	手作業 - ステップワイズ法		257
- デザイン行列 X	68	- 変数選択		256
- 2 因子実験	122	手作業による変数減少法 - ステップワイズ法		237
- 母数	121	手順 - 変数選択		232
データは数値 - 文字変数として設定	144	Tseng-Hsu 法 - 知名ら(2014)		420
データファイル - 併合	424	DevSq () 関数 - 偏差平方和		57
データリスト - 質的変数のまま	240	- Excel	168, 245	
適している - ロジスティック曲線	483	- 偏差平方和		245
手計算時代 - 正規方程式の解	126	デルタ法 - 近似95%信頼区間		184
- 偏差平方和	7	- 2次形式		472
design matrix - Arithmetic mean	163	典型的な事例 - 「平均への回帰」現象		320
- Wikipedia	163	転置 / 行列 - 行/列の入れ替え		15

て	- Transpose () 関数	15	な	内臓痛試験 - 用量依存性データ	455
	転置と積和 - デザイン行列 X	160		内部で計算 - 偏微分式	360
	転置行列 - Transpose () 関数	226		中西(2016) - 逆推定の基礎	194
	伝統的 - 実験計画法	55		- 非線形回帰	194
	伝統的な共分散分析 - 奥野ら(1981)	195		流れ不良 - ウス不良	220
	伝統的な実験計画法 - 田口の式	131		成り立たない - 分散の加法性	100
	伝統的な方法 - 回帰直線の95%信頼区間	180		難解なのか? - 共分散分析	203
と	同一の反応 - 用量の比較	470		難解な解析法 - 共分散分析	195
	同一直線上か - 回帰直線	265, 273		難点 - 共分散分析の適用	196
	- 95%信頼区間	276	に	2 因子モデル - 3 因子モデル	232
	同一用量 - 対応のない t 検定	469		- 主効果 + 交互作用モデル	233
	doステートメント - DATAステップ	211, 400, 444		2 因子実験 - 繰り返しが 2	89
	統計ソフト - 自動的に偏微分	342		- 繰り返しが等しい	89
	- デザイン行列	7		- 繰り返しが不揃い	89, 102
	- 微妙に異なる結果	391		- データの構造式	122
	- 理解の妨げ	195		2 因子実験のデータ - 楠ら(1995)	121
	統計ソフト SAS - ミカエリス・メンテン式	421		量×量の 2 因子実験 - 酵素阻害実験データ	429
	統計ソフトの使い方 - 見方さえ習得	155		2×2 の格子状 - 層別散布図	220
	統計ソフトの使用 - 非線形回帰	417		2×2 区分 - 散布図	254
	統計ソフトの出力 - Excelの作図機能	425		2 区分に分類 - 補助の共変量	252
	統計学辞典 - 竹内ら(1989)	243		2 群の前値 X - 揺らぎ	284
	統計的な判断 - 管理限界	224		2 群間の t 検定 - 段階的に適用	245
	統計的分析 - 数値に基づく対策	229		2 群間の大小関係 - 驚くべき検出力の差	325
	統計的方法 原著 第6版 - スネデガーら(1972)	282		2 群比較 - 交絡変数	243
	統計量に関するキーワード - 出力ファイルでの変数名	463		2 元配置 - 前値	38
	同時あてはめ - ロジスティック曲線/ 3本	337		- 対照群	38
	- ロジスティック曲線/ 複数	335		2 次曲線 - 予測プロファイル	83
	同等と見なせる - 傾きの差の範囲	480		2 次形式 - 合成分散の一般式	178
	等比で - 濃度あるいは用量の設定	455		- 推定値 y^{\wedge} の分散の分散	179
	実験動物 - 成長期	243		- 共分散行列	271
	投与終了時の体重 - 交絡変数	260		- 線形和 L	299
	投与前値の調整 - 高橋(2012)	310		- 線形和 L に対する分散	128
	投与前値を共変量 - ANCOVA	294		- 線形和 L の分散	206
	毒性 - 化合物	260		- 対数効力比 ρ' の分散	477
	毒性試験 - 反復投与	243, 260		- デルタ法	472
	毒性評価 - 竹内ら(1989)	243		- 2 次式の95%信頼区間	181
	独立でない - 効果の分散	76		- パラメータの共分散行列 $\Sigma(\theta^{\wedge})$	412
	吐出货量 - ガンスプレー	219		- 微係数ベクトル z	346
	吐出货量 x - 管理限界を超える	235		- V^{\wedge}_{max} の分散	412
	塗装不良の改善 - 宮川(2008)	219		- 分散の計算	74
	どのように説明 - 傾きに統計的に差	331		- 分散共分散行列	189
	Dobsonら(2018) - Generalized Linear Model	369		- Vec Quadratic () 関数	191, 354
	ドブソン(2008) - 反復重み付き回帰	371		2 次式 - 回帰分析	83
	- 用量反応モデル	369		- 95%信頼区間	83
	共に有意な差 - 後値 Y および前後差 D	293		2 次式の95%信頼区間 - ガラスの天井	181
	- 後値 Y および前後差 D	305		- 2 次形式	181
	Transpose() 関数 - 行列の転置/ Excel	410		2 次式のあてはめ - デザイン行列 X	82
	- 転置 / 行列	15		2 次式の解の公式 - 95%信頼区間	36
	- 転置行列	226		- 正確な95%信頼区間	186
	- Excel	161		2 次式の推定値 - 分散の計算過程	84
	Draperら(1989) - 行列表記	168		2 種類の推定方法 - 阻害定数 K_i の比	428
	ドレーパら(1968) - 行列表記	168		2 水準間の差 - 95%信頼区間	209
	- 重回帰分析	156		- 線形和 L	29
	- デザイン行列 X	156		2 段階による推定値 - 両逆数・Lineweaver-Burk 変換	428
	内臓痛を緩和 - Alosetronの3用量	455		2 段階の線形化 - 阻害定数 K_i の推定	444
				2 段階目 - 阻害定数 K_i の推定	432

に2 値データ - 50%反応量 D_{50}	368	は	- 共分散行列 $\Sigma(\beta^{\wedge})$	295
2 値反応 - ロジスティック回帰	327		- 勾配/ 曲線の	457
- ロジスティック回帰	335		- 線形化	370
二変量のあてはめ - JMP の回帰分析	190		- 対数効力比 ρ'	478
二変量の関係 - JMP	266		- 分散と共分散	298
2 本の直線の推定値 - 交互作用がある場合	302		- 偏微分式の出力結果	485
2 本のロジスティック曲線 - 初期値	482		パラメータ(係数) - β_i or θ_i	97
2 本の回帰直線 - 原点を通る	277		- 線形モデル	122
- 最小2乗平均の例示	298		パラメータ(推定値) - 共分散行列 $\Sigma(\theta^{\wedge})$	351
- 非平行	273		- 相関行列	351
2 本の直線 - 傾きが同じでない	267		パラメータで偏微分 - 逆推定値の式	184
- 傾きは同じか	265		- ギンペルツ曲線	385
妊娠18日の体重 - 総摂餌量	243		- 自動的に行う機能	399
妊娠7日の体重 - 前値	244		- 対数効力比 ρ'	475
妊娠ラット - 体重増加データ	244		- モデル式	342
ね ネルソン(1988) - 寿命データの解析	378		- ロジスティック曲線の式	342
年収の比較データ - 芳賀(2014)	7		パラメータとして直接 - V_{max} および K_m	426
年収 - 比較調査	7		パラメータに関し - 偏微分	447
年齢 x が小さい順 - 推定曲線を重ね書	338		パラメータに関して - 線形ではない	478
の 濃度 $[I]$ を逆推定 - 阻害定数 K_i	430		パラメータに関する - 95%信頼区間	344
濃度あるいは用量の設定 - 等比で	455		- 偏微分式	343
濃度がゼロ - 対数変換した濃度(用量)	455		パラメータに関する偏微分 - NLINプロシジヤ	464
- 低めの濃度を与え	455		- JMP	356
Norm.dist () 関数 - Excel	367		パラメータの95%信頼区間 - 非線形回帰	409
- 正規分布	367		パラメータの一括推定 - NLINプロシジヤ	452
Norm.Inv (Rand (), 0, 1) 関数 - 正規乱数	286		パラメータの共分散行列 - outest オプション	488
Norm.Inv () 関数 - 正規分布の逆関数	283		- OUTEST=	360
Non-linear regression - Armitageら(2002)	405		- $(X^T X)^{-1} \sigma^2$	109
は パーセント点 π 分布 - η 分布に関して解く	397		- 合成分散の一般式	330
バイアス - 平均への回帰	282		- $\Sigma(\beta^{\wedge})$	178
肺活量の減少 - カドミウムガス	327, 328, 337		- $\Sigma(\theta^{\wedge})$	109, 206, 227
倍精度実数 - 計算精度	126		- デザイン行列 X	61, 173
芳賀(2014) - 伊奈の式	66		- パラメータの相関行列	361
- LOF解析	80		パラメータの共分散行列 $\Sigma(\beta^{\wedge})$ - 回帰直線	473
- 共分散分析	7		- デザイン行列 X	299
- 繰り返しが不揃い	69		パラメータの共分散行列 $\Sigma(\theta^{\wedge})$ - 2次形式	412
- 実験計画法 改訂版	55		- 反復の過程	425
- 質的因子	55		- 微係数行列 Z	408, 565
- 年収の比較データ	7		- ロジスティック曲線	344
芳賀(2016) - Emaxモデル	372		パラメータの差 - 95%信頼区間	345
- 逆推定/JMPによる	183		パラメータの初期値 - 正攻法/ 設定	340
- 非線形モデル	194, 337, 349		- 探索的な推定	340
- 非線形回帰/ JMP	339		パラメータの推定 - 行列計算	169
- 非線形回帰に関する必読の書	337		- Gompertz 4P	393
掃き出し演算子 - 逆行列	126		- 残差線形化法	385
- 新村(1983)	126		- 正規方程式の解	165
掃き出し法 - 逆行列	125		- ロジスティック 4P	393
箱ひげ図 - Excel	56		パラメータの推定の実際 - 偏差平方和	167
パズル的な代数計算 - 逆行列	126		パラメータの推定精度 - 線形化	405
発電プラント - カキの成長	38		- 線形化	417
Hanes-Woolf 変換式 - 傾き β_1^{\wedge} の逆数	427		パラメータの推定値 - GLMプロシジヤ	213
Hamilton の評価尺度 - うつ症状	30		パラメータの線形式 - メリット	371
原田(2017) - 平行線検定	480		パラメータの相関行列 - デフォルトで出力	360
- 薬物の効力比較	480		- パラメータの共分散行列	361
parms ステートメント - NLNプロシジヤ	401		パラメータの同時推定 - 非線形モデル	456
- 初期値を設定	479		パラメータの比 - 合成分散	184
パラメータ - 共分散	206		パラメータの分散 - ギンペルツ曲線	389
- 共分散行列	14, 25, 50, 72, 140		- $\Sigma(\theta^{\wedge})$ の対角要素	408
	172, 226, 457		パラメータの分散の推定 - 偏差平方和	171

- はパラメータ間の差 - 合成分散の一般式 345
 パラメータ関数 - 対比の設定 23
 パラメータ推定 - Excel のソルバー 459, 460
 - 残差線形化法 363
 - *Scatchard・Eadie-Hofstee* 変換 416
 - (0, 1) 型 78
 - ソルバー/ Excel 482
 - 対比型ダミー変数 25
 - 非線形回帰 458
 - *Hanes-Woolf* 変換 414
 - ミカエリス・メンテン式 405
 - *Lineweaver-Burk* 変換 412
 - ロジスティック曲線/ Excel 459
 パラメータ推定の頑健性 - 非線形モデル 405
 パラメータ推定の脆弱性 - 線形化 405
 パラメータ推定値 - (1, 0) 型ダミー変数 152
 - モデルのあてはめ 242
 半角英数字のフォント - Times New Roman 157
Hanes-Woolf 変換 - 線形化 406
 判断基準 - 前後差か増加率か 247
 反応が量的変数 - ロジスティック曲線 327
 反応速度 - 推定値 V^{\wedge} 410
 - 推定値 V^{\wedge} の分散 410
 反応速度 V - Velocity 407
 反応速度 V の最大値 - V_{max} 407
 反復の過程 - パラメータの共分散行列 $\Sigma(\theta^{\wedge})$ 425
 反復重み付き回帰 - 最尤法で解く方法 369
 - 高橋(2017) 371
 - ドブソン(2008) 371
 - ロジット 371
 反復投与 - 毒性試験 243, 260
 $\hat{p}=0.105$ - スナップショット 311
 p 値の比較 - 前値 X での t 検定 315
 Vehicle群 - 陰性対照群 455
 PCP剤 - 原点を共通 51
 ピース編(1992) - 薬理学的活性の評価 469
 比較調査 - 年収 7
 引数 - Rand () 関数 283
 低めの濃度を与え - 濃度がゼロ 455
 微係数ベクトル z - 2 次形式 346
 微係数行列 Z - 積和行列 ($Z^T Z$) 448, 466
 - 積和行列の逆行列 ($Z^T Z$)⁻¹ 343
 - デザイン行列 X 342, 364, 486
 - パラメータの共分散行列 $\Sigma(\theta^{\wedge})$ 408, 465
 ひげ付き - 線グラフ 201
 ひげの長さ - 95%信頼区間 208
 ひげの長さの設定 - 折れ線グラフ 208
 微小なパラメータ増減値 - 残差平方和 336
 非線型回帰モデル - 山田ら(2003) 379
 非線形 - ミカエリス・メンテン式 406
 非線形の解析モデル - ダミー変数 457
 非線形の式 - 偏微分/ パラメータ 461
 非線形モデル - 大和田(2010) 194
 - 寄山ら(2008) 455, 456
 - SAS Institute(2014) 358
 - 線形モデルではない 336
 - 芳賀(2016) 194, 337, 349
 - パラメータの同時推定 456
 - パラメータ推定の頑健性 405
- ひ - ミカエリス・メンテン式 374, 405, 408
 - 薬理試験における統計解析 455
 非線形モデルの解法 - 残差線形モデル 488
 非線形回帰 - Wikipedia 366
 - Excel により手軽に 419
 - Excel の 1 枚のシート上で 420
 - Excel のソルバー 439
 - NLINプロシジャ 421, 444
 - 逆推定 395
 - 計算方法がブラック・ボックス 417
 - ギンペルツ曲線の逆推定 394
 - JMP 350, 458
 - 阻害定数の推定 428
 - 阻害定数 K_i の直接推定 443
 - 知名ら(2014) 405
 - 統計ソフトの使用 417
 - 中西(2016) 194
 - パラメータの95%信頼区間 409
 - パラメータ推定 458
 - *Hanes-Woolf* 変換 425, 426
 - ミカエリス・メンテン式 366
 - 予測式の保存 395
 - ロジスティック曲線 349
 非線形回帰/ JMP - 芳賀(2016) 339
 - 95%信頼区間 459
 非線形回帰/ 統計ソフト - 自動微分 351
 非線形回帰に関する必読の書 - 芳賀(2016) 337
 1 つのパラメータ - 1 つの変数 336
 1 つの変数 - 1 つのパラメータ 336
 比の分散 - アーミテージら(2001) 184
 非曝露・曝露群 - ロジスティック曲線 339
 被覆度 - 雑草重量 47
 微分した式 - 計算式エディター 357
 非平行 - 回帰直線 31
 - 回帰直線の差 34
 - 2本の回帰直線 273
 非平行な直線 - 最小2乗平均 309
 微妙に異なる結果 - 統計ソフト 391
 (1, 1) 標示型 - ダミー変数 10, 49, 51
 標準誤差 - 検証/ NLINプロシジャ 422
 - JMPの非線形回帰 350
 - 対角要素の平方根 226
 標準誤差 SE - 最大反応 V_{max} 418
 - ミカエリス定数 K_m 419
 標準誤差が示されていない - Armitageら(2002) 420
 標準最小2乗法 - 繋げる配慮 239
 - 予測プロファイル 258
 標準偏差 - Stdev.S () 関数 283
 ヒル係数が $h=1$ - ミカエリス・メンテン式 374
 ヒルの式 - 酵素反応速度論 372
 - ロジスティック曲線 373
 広瀬ら(2021) - ベンチマークドース 382
 品質管理便覧, 新版 第2版 - 朝香ら(1988) 121
 不安定な阻害定数 K_i の推定 - 誤差の伝播 446
 不一致 - 平方和の計算結果 117
 Finney(1978) - 平行線検定法 480
 (V_{max}) _{i} と (K_m) _{i} を推定 - 阻害薬の濃度別 451
 $1/(V_{max})_0$ - Y切片 433

ふ V_{max} の分散 - 2次形式	412	ふ 分散の加法性が成り立たない - 不揃い	104
V_{max} - 反応速度 V の最大値	407	分散の期待値の違い - 検出力に反映	325
V_{max} および K_m - パラメータとして直接	426	- Frison, Pocock (1992)	325
V_{max} で偏微分 - ミカエリス・メンテン式	407	分散の計算 - 2次形式	74
V_{max} の2分の1となる基質濃度 - K_m	407	分散の計算過程 - 2次式の推定値	84
V_{max} を共通 - 阻害定数 (K_i) _{i}	441	分散の推定 - 水準効果	66
- 阻害定数 K_i が不安定	442	- 有効反復数 n_e	100
- ミカエリス・メンテン曲線	441	分散共分散行列 - 誤差伝播の公式の一般化	189
4 パラメータモデル - 拡張	372	- 2次形式	189
- 3パラメータモデル	372	分散分析表 - LOF解析	87
複数の共変量 - 回帰分析	250, 255	- 回帰分析	72
- 交互作用	250	- 繰り返しが等しい	93
- (2×2) 格子状の散布図	252	- 繰り返しが不揃い	111
不揃い - 分散の加法性が成り立たない	104	- 欠測値がある直交表	141
- 平方和の計算	104	- 交互作用	21, 33
太い外枠 - 行列の表記	157	- 交互作用を含む	200
負の阻害定数 K_i の推定 - X切片	431	- GLMプロシジャ	116
不偏推定量 - 誤差分散	126	- JMP	37
フライス(2004) - 共分散分析と変化量の検討	30	- JMPの「モデルのあてはめ」	145
- 抗うつ薬投与前後	30	- 自由度	177
ブラインド化 - 第三者	282	- タイプ I の平方和	107
プラセボ - 実薬	281	- タイプ II の平方和	113
プラセボ群 - 抗うつ薬	157	- タイプ III の平方和算出	114
Frison, Pocock (1992) - ANCOVA	294	- 逐次平方和	205
- CHANGE vs ANCOVA	294	- 平方和の分解	56, 72, 89, 135
- 分散の期待値の違い	325	分散分析表だけ - 考察は不十分	120
- POST vs ANCOVA	294	分散分析表の解釈 - 隔靴搔痒的な解釈	115
不良の発生 - 中心膜厚	220	分散分析表の作成 - Excel	176
BLUE - 最良線形不偏推定量	124	データ分析ツール - Excel	158
- best linear unbiased estimator	124	- 回帰分析	158
フルモデル - 4次元交互作用	238	分析ツール - Excel	59
フルモデルからA×Bを除く - タイプ II の平方和	112	- 回帰分析	59
- タイプ III の平方和	113	分析ツールの乱数 - 正規乱数	283
フルモデルからAを除く - タイプ III の平方和	113	分野 - 交互作用を念頭に	219
フルモデルからBを除く - タイプ III の平方和	113	- 主効果に重き	219
プログラミング機能 - NLINプロシジャ	465	へ 平滑線 - スムージング/ Excelの	456
PROC GLM - Unbalanced ANOVA	102	「平均への回帰」現象 - 典型的な事例	320
PROCステップ - SAS	421	平均 - Average () 関数	57, 283
プロビット法 - 50パーセント致死量	367	平均±SD - 重ね書き	222
プロビット法・ロジット法・補2重対数法 - 高橋(2017)	369	平均の信頼限界の計算式 - JMPの回帰分析	190
分散 - 回帰直線の推定値	225	平均の平均 - 総平均	71
- 水準平均の差	65	平均への回帰 - 岩崎(2002)	290
- 線形和 L	26, 41, 227	- 起きた結果	313
- 対角要素	15	- 可視化	283
- 面積 $z = xy$	189	- 葛藤	281
分散・共分散 - 推定値	61	- 上限と下限	31
分散および共分散の推定 - 回帰パラメータ	175	- 前後差 D	283
分散が2倍 - 前後差の相関が0	288	- 前値 X	30
分散とSE - 逆推定値	473	- 前値と後値	281
分散と共分散 - パラメータ	298	- 高橋(2012)	310
分散の引き算 - 分散の加法性	67	- バイアス	282
分散の加法性 - 限界	75	- 厄介な現象	281
- 差の分散	474	平行でない回帰直線 - 平行な回帰直線	480
- 水準の平均の差	68	平行でない直線 - 効力比	469
- 水準平均の分散	65	平行な - ロジスティック曲線	455
- 互いに独立	96	平行なロジスティック曲線 - NLINプロシジャ	484
- 成り立たない	100	- 効力比の推定	481
- 分散の引き算	67	平行な回帰直線 - 差の95%信頼区間	201
- 平方和の分解	55	- 95%信頼区間	210

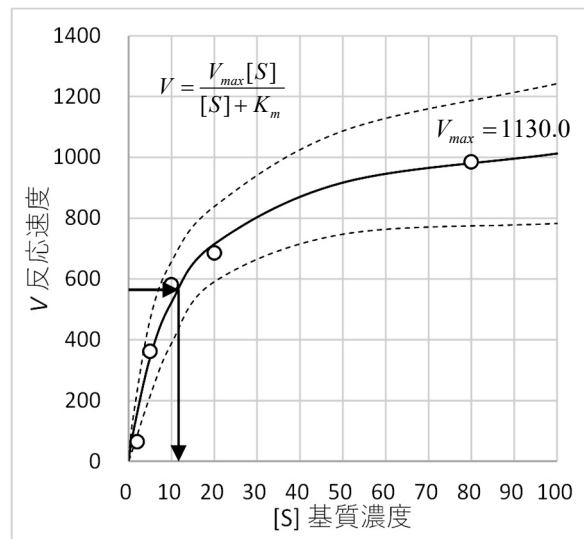
へ	- 平行でない回帰直線	480	へ	変数名 - 別名	227
	平行な線グラフ - 前後差	247		変数セルの変更 - ソルバー/ Excel	460
	平行な直線 - 共分散分析	480		変数の自動選択 - 恣意性を排除	237
	- 効力比の推定	475		変数は隣り合う - Excel の回帰分析	231
	- 切片が異なる	475, 476		変数減少法 - Excel の回帰分析	230
	- ロジスティック曲線らしき痕跡	455		- 交互作用	230
	平行線 - 水準ごとの切片	201		変数選択 - 交互作用を含む	239
	平行線検定 - 原田(2017)	480		- 自動的	237
	平行線検定法 - 佐久間(2017)	480		- JMP	256
	- Finney(1978)	480		- ステップワイズ法	255
	- 偏差平方和	480		- 総当たり法	237
	併合 - 層別散布図	223		- 手作業	256
	- データファイル	424		- 手順	232
	平方和の差分 - 逐次平方和	106		変数増加・減少・増減 - ステップワイズ法	255
	平方和 - 計算公式	86		変数増減 - 方向	256
	- 交互作用	13		変数増減法 - 選択されたモデル	239
	- SumSq () 関数	57		変数名として S_V - 変換式 ($[S]/V$)	426
	平方和の計算 - SumSq () 関数	91		ベンチマークドース - 広瀬ら(2021)	382
	- 修正項 CT	92		変動係数 - CV	285
	- 不揃い	104		偏微分 - 傾き β^{\wedge}_1 と切片 β^{\wedge}_0 の比	412
	平方和の計算結果 - 不一致	117		偏微分式の出力 - list オプション	401
	平方和の計算表 - 修正項 CT	93		偏微分 - 誤差の平方和 Q	164, 123
	平方和の増加 - 逐次的な回帰分析	205		- パラメータに関し	447
	平方和の比較 - タイプ I, II, III	216		偏微分 β_0 - 他の変数を定数	164
	平方和の分解 - $S_T = S_e + S_R$	177		偏微分/パラメータ - 非線形の式	461
	- 質的な変数に限定	135		偏微分した式 - NLINプロシジヤ	483
	- 総平方和	92		- list オプション	483
	- 分散の加法性	55		偏微分行列 Z - デザイン行列 X	194
	- 分散分析表	56, 72, 89, 135		偏微分式 - 行ベクトル d	184
	- 補足	182		- 内部で計算	360
	Hanes-Woolf 変換 - NLINプロシジヤ	426		- パラメータに関する	343
	- 回帰直線の95%信頼区間	415		- list オプション	360
	- パラメータ推定	414		偏微分式が正しいのか - 疑心暗鬼	464
	- 非線形回帰	425, 426		偏微分式の計算 - Excel	344
	ペースト - 「値」のみ	313		偏微分式の出力 - NLINプロシジヤ	465, 485
	「ベータ」と入力 - ギリシャ文字 β	157		- list オプション	424
	β^{\wedge}_{ob} で偏微分 - 切片 β^{\wedge}_0 の逆数	412		偏微分式の出力結果 - パラメータ	485
	β_i or θ_i - パラメータ(係数)	97		偏微分式を出力 - list オプション	461
	Vec Quadratic () 関数 - JMP	191, 354	ほ	方向 - 変数増減	256
	- 2次形式	191, 354		「 X 」をボード&イタリック - 「 X 」	157
	best linear unbiased estimator - BLUE	124		補完 - 直交性を保ちたい	145
	別々 - 回帰直線	8		補助の共変量 - 主たる共変量	252
	別名 - デザイン行列の変数	209		- 2区分に分類	252
	- 変数名	227		母数 - データの構造式	121
	Velocity - 反応速度 V	407		母数に関する制約 - 構造(DE)モデル	122
	変換式 ($[S]/V$) - 変数名として S_V	426		POST vs ANCOVA - Frison, Pocock(1992)	294
	変化させた場合 - 時点間の相関	316		補足 - 平方和の分解	182
	変化量の検討 - 共分散分析	30		補2重対数法 - ワイブル分布	376
	便宜的 - 直線のあてはめ	481		ほぼ等しい - 前値 X	314
	便宜的な方法 - 回帰直線	335		本質的な脆弱性 - Excel	421
	変幻自在 - デザイン行列 X	18		- Excel による解析	391
	偏差平方和 - DevSq () 関数	57	ま	(-1, 1)対比型 - (1, -1)対比型	258
	- パラメータの推定の実際	167		- ダミー変数	256
	- 回帰パラメータの推定	164		$-(Km)_0$ - X 切片	433
	- 手計算時代	7		マイナス 10.0 - 減少効果/両群の	311
	- デザイン行列 X	170		Maximam Extreme Value Distribution - 最大極値分布	375
	- DevSq () 関数	245		マクロ - すべての組み合わせ	240
	- パラメータの分散の推定	171		MaCulloughら(2008) - Excel 2007 に対する批判	155
	- 平行線検定法	480			

ま Mmult () 関数 - Excel	161, 271	も 目標値/最小値 - 目的セルの設定/ソルバー	460
- 行列の積	226, 271	文字変数として設定 - データは数値	144
み Michaelis-Menten - ミカエリス・メンテン	405	modelステートメント - NLNプロシジャ	401
Michaelis-Menten assay - Armitageら(2002)	405	- モデル式	212
ミカエリス・メンテン - Michaelis-Menten	405	モデル式 - modelステートメント	212
ミカエリス・メンテン曲線 - 95%信頼区間	409	- パラメータで偏微分	342
- 95%信頼区間	411, 423, 425	モデルに含めるダミー変数の順番 -	111
- 阻害薬濃度 $[I]_i$ ごと	441	モデルのあてはめ - 回帰分析	20
- V_{max} を共通	441	- 逆推定値の95%信頼区間	192
ミカエリス・メンテン式 - 新たな統計解析	405	- 最小2乗平均	148
- Wikipedia	366	- JMP	20, 144
- ガウス・ニュートン法	336	- ステップワイズ法	256
- 各種の線形変換	417	- 選択された変数	241
- K_m で偏微分	407	- パラメータ推定値	242
- SASプログラム	422	モデルの作成 - ステップワイズ法	257
- 線形化	405	モデルの設定 - 多項式の中心化	21
- 阻害定数 K_i の推定	428	元の偏微分式にミス - 収束しない原因	466
- ソルバー	341	や 薬物の効力比較 - 原田(2017)	480
- 統計ソフト SAS	421	薬理学的活性の評価 - 血圧上昇データ	469
- パラメータ推定	405	- ピース編(1992)	469
- 非線形	406	薬理試験における統計解析 - 嵯山ら(2008)	455
- 非線形モデル	374, 405, 408	- 非線形モデル	455
- 非線形回帰	366	厄介な現象 - 平均への回帰	281
- ヒル係数が $h=1$	374	山崎ら(2008) - 拮抗阻害	428
- V_{max} で偏微分	407	- 阻害定数の推定	428
- ロジスティック曲線式	411	山田ら(2003) - シグモイド曲線	379
ミカエリス定数 - K_m	407	- 非線形回帰モデル	379
- 最大反応 V_{max} を共通	439, 440	- ロジスティック曲線	379
ミカエリス定数 (K_m) _i - 阻害定数 K_i を用いた式	431	ゆ 有意な差 - 後値 Y	292
- 阻害薬Aの濃度 $[I]_i$ ごと	445	- 前後差 D	291
- 両逆数・Lineweaver-Burk 変換	432	有意差検定の p 値 - 3 通りに分ける	316
ミカエリス定数 K_m - 95%信頼区間	419	有効反復数 - 伊奈の式	66
- 最大反応 V_{max}	405	- 計算が容易でない	76
- 標準誤差 SE	419	- 効果の分散	66
ミカエリス定数 K_m の推定 - 阻害薬 A なし	432	- 田口の式	66
ミカエリス定数 K_m ^ - 傾き β^{\wedge}_1 と切片 β^{\wedge}_0 の比	412	有効反復数 n_e - 分散の推定	100
ミカエリス定数の SE - 3 倍強の違い	419	- 田口の式	134
ミカエリス定数の推定 - 阻害薬 A 14 nM	434	有償ソフト - JMP	149
見方さえ習得 - 統計ソフトの使い方	155	揺らぎ - 阻害定数 K_i	442
宮川(2008) - SQCの基本	219	- 2 群の前値 X	284
- 塗装不良の改善	219	揺らぎの実際 - 阻害定数 K_i	435
む 無視しない要因の自由度 - 田口の式	131	よ 要因 - 区別	232
無償で継続的 - OnDemand SAS	44, 149, 195	要因の取舍選択 - 基準	232
211, 397, 421, 483, 461, 477		要因効果 - 95%信頼区間	98
- SASのGLMプロシジャ	120	陽性対照群 - Sham群	455
- SAS/ OnDemand	461, 477	- 高めの濃度を便宜的に	455
- GLMプロシジャ	120	用量の推定 - 50パーセント反応	383
無償の統計ソフト - R が脚光を浴びている	149	- 10パーセント反応	383
め 名義尺度, 順序尺度 - ステップワイズ法	256	用量の比較 - 同一の反応	470
迷信 - 共分散分析	268	用量依存性データ - 内臓痛試験	455
名目の $\alpha=0.05$ - 実際の α エラー	311	用量反応 - 直線的	470
名目の α エラー 5% - 前値 X を共変量	317	用量反応モデル - ドブソン(2008)	369
名目の α エラーに接近 - 後値 Y	314	用量反応関係 - 除草剤	48
メリット - パラメータの線形式	371	- 相対的な比較検討	470
面積 $z = xy$ - 分散	189	用量反応直線 - 対数用量 x	471
面積の標準偏差 - 誤差伝播の公式	189	4 次元交互作用 - フルモデル	238
も 妄想 - 共分散分析	280	吉村編著(1987) - LD_{50} の推定法	369
目的セルの設定/ソルバー - 目標値/最小値	460	- 臓器重量と体重の関係	260
目標値 - ソルバー	341		

よ 吉村編著(1987) - 体重の増加抑制	246	り	- ミカエリス定数 (Km) _i の推定	432
- 体重は増加量でみるべきか?	243	両側 t 分布 - T.dist.2T () 関数		284
予測プロファイル - Excel	26	両辺に掛ける - 逆行列		169
- 折れ線グラフ	75	量的な反応 - 累積ロジスティック分布		372
- 95%信頼区間	146, 210, 351	- ロジスティック曲線		335
- 結果の表示	259	量的変数に - 質的変数を		55
- 効果	75	理論的な枠組み - 線形推定論		121
- 効果 α_i	62	臨床試験 - 血圧が高めの人たち		281
- 交互作用プロット	242	る 累積バグ数の経時変化 - ゴンペルツ曲線		374
- ゴンペルツ曲線	395	累積ロジスティック分布 - 量的な反応		372
- 3次曲線	82	累積分布関数 - 確率密度関数		367
- JMP	21, 242	- シグモイド曲線		367
- 水準間の差	28, 75, 209	ルール - 組み合わせ		256
- 洗浄水の温度 x	210	冷蔵庫の塗装不良 - 現業実習		219
- 2次曲線	83	れ 0.0 に近づく - 時点間の相関		314
- 標準最小2乗法	258	歴史的背景 - 各種のダミー変数		148
予測式の保存 - 非線形回帰	395	- 知名ら(2014)		405
予測信頼限界 - 計算式の保存	396	“列・行”ではなく - “行・列”の順番		160
予測値 y^{\wedge} - 埋め込まれている計算式	353	る Rouaud (2017) - ガウス・ニュートン法		420
- 計算式の設定	349	$\rho=0.10$ - 相関の期待値		321
予測値 y^{\wedge} の重ね合わせ - グラフビルダー	355	- 相関の期待値		323
予測値の95%信頼区間 - 計算式を含む	352	$\rho=0.10$ の場合 - 検出力の可視化		321
予測値の推定 - 誤差分散	127	- 検出力の可視化		323
- 誤差分散の推定	140	$\rho=0.50$ の場合 - 検出力		324
予測値の標準誤差 y^{\wedge} - 計算式	353	$\rho=0.90$ - 相関の期待値		320
読み飛ばしたくなる - シグマ	155	$\rho=0.90$ の場合 - 検出力の可視化		321
4種類の阻害様式 - 阻害薬の酵素反応	429	- 検出力の可視化		323
4種の平方和 - LSMEANS	89	64ビット版 - 計算精度を保つ		126
- 高橋ら(1989)	153	$\log_{10} \rho$ - 対数効力比 ρ'		479
4種の平方和と LSMEANS - 高橋ら(1989)	89	6通りの典型例 - 臓器重量と体重		261
ら Lineweaver-Burk 変換 - 回帰直線の95%信頼区間	413	ロジスティック 4P - Gompertz 4P		392
- 推定値が不安定	413	- パラメータの推定		393
- パラメータ推定	412	ロジスティック回帰 - 2値反応	327, 335	
- 両逆数	406, 411	- ロジットに関する線形式		371
ラットの子宮重量 - 環境ホルモン	379	ロジスティック曲線 - あてはめの基礎		335
- SASデータセット化	400	- 折れ線グラフ		481
乱塊法 - 共分散分析	47	- 下限に漸近		339
乱塊法における共分散分析 - 奥野ら(1978)	47	- 活用/さらなる		455
Rand () 関数 - 一様乱数	456	- 95%信頼区間	342, 346, 352, 485	
- Excel	283, 446	- 効力比の直接推定		481
- 引数	283	- ゴンペルツ曲線		379
り 理解が上滑り - GLMプロシジャの世界だけ	117	- 最大値・最小値		336
理解の妨げ - 統計ソフト	195	- 散布図上に重ね書き	338, 347	
理解を深めたい - NLINプロシジャ	447	- 初期パラメータ		341
list オプション - NLINプロシジャ	465	- 層別因子		327
- 偏微分式の出力	401	- ソルバー/ Excel		481
- 偏微分した式	483	- 2パラメータ		410
- 偏微分式を出力	461	- 適している		483
listオプション - 偏微分式	360	- パラメータの共分散行列 $\Sigma(\theta^{\wedge})$		344
- 偏微分式の出力	424	- 反応が量的変数		327
Littellら(2002) - Analisis of Covariance	38	- 非線形回帰		349
- カキの成長に関する実験	38	- 非曝露・曝露群		339
LinEst () 関数 - 回帰分析	60	- ヒルの式		373
- 線形モデル	60	- 平行な		455
両逆数 - Lineweaver-Burk 変換	406, 411	- 山田ら(2003)		379
両逆数 GLM/ gamma errors - Armitageら(2002)	420	- 量的な反応		335
両逆数・Lineweaver-Burk 変換 - 阻害定数 (Ki) _i	446	- ワイブル曲線		379
- 2段階による推定値	428	ロジスティック曲線/3本 - 同時あてはめ		337
		ロジスティック曲線/4P - 残差線形化法		467

ろ	ロジスティック曲線/ Excel - パラメータ推定	459	わ	Y切片 - $1 / (Vmax)_0$	433
	ロジスティック曲線/ 複数 - 同時あてはめ	335		ワイブル曲線 - ロジスティック曲線	379
	ロジスティック曲線のあてはめ - 残差線形化法	466		ワイブル分布 - 関数の形式はまちまち	376
	ロジスティック曲線の差 - 95%信頼区間	347		- 形状	378
	ロジスティック曲線の式 - パラメータで偏微分	342		- 最小極値分布	376, 378
	ロジスティック曲線らしき痕跡 - 平行な直線	455		- 補2重対数法	376
	ロジスティック曲線式 - ミカエリス・メンテン式	411		悪あがき - 線形化にこだわる	419
	ロジスティック分布 - 正規分布	384		One-way ANOVA - offset from reference group	163
	ロジスティック分布の σ_{LGS} - 正規分布の σ_{NOR}	370		- cell means model	163
	ロジスティック分布の確率密度関数 - $f_{LGS}(x)$	369			
	ロジスティック分布の累積分布関数 - $F_{LGS}(x)$	369			
	ロジット - 線形式	335			
	- 反復重み付き回帰	371			
	ロジットに関する線形式 - ロジスティック回帰	371			

推定されたミカエリス・メンテン曲線に対する 95%信頼区間



解析用ファイル 一覧

ファイル名	ファイル名
第01章_01_会社と年齢と給料.xlsx	第06章_03_1_SASプログラムSAS.txt
第01章_02_1_奥野_層別回帰.xlsx	第06章_03_2_SASの出力編集.xlsx
第01章_02_2_奥野_層別回帰.jmp	
第01章_03_1_プライス_抗うつ薬_交互作用.xlsx	第07章_01_層別散布図.xlsx
第01章_03_2_プライス_抗うつ薬_交互作用.jmp	第07章_02_95%信頼区間.xlsx
第01章_04_1_牡蠣.xlsx	第07章_03_変数選択.xlsx
第01章_04_2_牡蠣sas.txt	第07章_04_1_JMP_ステップ.jmp
第01章_05_除草剤.xlsx	第07章_04_2_JMP_ステップ名義.jmp
第02章_01_1因子_等しい.xlsx	第08章_01_妊娠ラット.xlsx
第02章_02_1因子_不揃い.xlsx	第08章_02_妊娠ラット回帰分析.xlsx
第02章_03_1因子_不揃い01型.xlsx	第08章_03_ステップワイズ.jmp
第02章_04_1_多項式.xlsx	第08章_04_1_臓器重量.xlsx
第02章_04_2_多項式.jmp	第08章_04_2_臓器重量_T_相関.jmp
	第08章_04_3_臓器重量_回帰.jmp
	第08章_05_臓器重量_交絡_回帰.xlsx
第03章_01_2元繰り返しが等しい.xlsx	第09章_02_1_前値X.xlsx
第03章_02_2元不揃い.xlsx	第09章_02_2_時点間の相関.xlsx
第03章_03_1_各種の平方和.xlsx	第09章_03_前値後値_群間比較.xlsx
第03章_03_2_各種の平方和_SAS.txt	第09章_04_1_共変量X.xlsx
	第09章_04_2_EX3.jmp
第04章_01_線形モデル_2因子.xlsx	第09章_04_3_共変量X.xlsx
第04章_02_線形モデル_直交.xlsx	第09章_05_αエラー.xlsx
第04章_03_線形モデル_直交欠測.xlsx	第09章_06_検出力.xlsx
第04章_04_L8欠測値.jmp	
第04章_05_1_欠測値がある直交表_SAS.txt	
第04章_05_2_線形モデル_SAS.xlsx	
第05章_02_行列計算_入門.xlsx	第10章_01_1_3本の直線.xlsx
第05章_03_回帰_偏差平方和.xlsx	第10章_01_2_カドミウムガス_スプライン.jmp
第05章_04_回帰_正規方程式.xlsx	第10章_02_3本のロジスティック曲線.xlsx
第05章_05_回帰_デザイン行列.xlsx	第10章_03_ロジスティック曲線95%.xlsx
第05章_06_回帰_逆推定.xlsx	第10章_04_1_カドミウムガス_ロジスティック曲線.jmp
第05章_07_1_回帰_JMP逆推定.jmp	第10章_04_2_ロジスティック曲線計算式.xlsx
第05章_07_2_回帰_逆推定直接.xlsx	第10章_05_1_非線形回帰_SAS.txt
	第10章_05_2_SAS_out.xlsx
	第10章_06_1_ロジ_残差線形化.xlsx
第06章_01_伝統的共分散分析.xlsx	第10章_06_2_ミカエリス_メンテン_Wikipedia.xlsx
第06章_02_層別共分散.xlsx	

解析用ファイル一覧

ファイル名

第11章_01_1_シグモイド曲線.xlsx
 第11章_01_2_ワイブル_最小極値.xlsx
 第11章_02_3種のシグモイドのあてはめ.xlsx
 第11章_03_ゴンペルツ.xlsx
 第11章_04_1_曲線のあてはめ.jmp
 第11章_04_2_ゴンペルツ非線形D10.jmp
 第11章_05_1_ゴンペルツSAS.txt
 第11章_05_2_ゴンペルツ_SAS.xlsx

 第12章_01_ミカエリス_メンテン_非線形_線形化.xlsx
 第12章_02_1_ミカエリス_SAS.txt
 第12章_02_2_ミカエリスWoolf_SAS.txt
 第12章_02_3_ミカエリス_SAS検証.xlsx
 第12章_03_1_酵素阻害.xlsx
 第12章_03_2_酵素阻害.xlsx

ファイル名

第12章_04_Vmax共通.xlsx
 第12章_05_1_阻害定数 Ki 直接推定_SAS.txt
 第12章_05_2_Ki直接.xlsx
 第12章_06_1_Vmax_Km 個別.xlsx
 第12章_06_2_Vmax_Km個別推定_SAS.txt

 第13章_01_1_内臓痛.xlsx
 第13章_01_2_内臓痛.jmp
 第13章_01_3_内臓痛_SAS.txt
 第13章_02_1_アンジオテンシン.jmp
 第13章_02_2_平行でない直線.xlsx
 第13章_03_1_平行な直線.xlsx
 第13章_03_2_平行な直線_SAS.txt
 第13章_04_1_ロジスティック.xlsx
 第13章_04_2_ロジスティック_SAS.txt

著者紹介

高橋 行雄 (たかはし ゆきお)

1971年 中央大学理工学部管理工学科終了
富士通電算機専門学校研究科終了

同年 日本ロシュ株式会社
研究所 研究統計課
前臨床および臨床試験の統計解析に従事

2002年 中外製薬株式会社 統計解析部

2011年 同社 退社

同年 BioStat 研究所 (株) 設立 現在に至る

著書 毒性・薬効データの統計解析, サイエンティスト社, 1987

SASによる実験データの解析, 東京大学出版会, 1989

毒性試験データの統計解析, 地人書館, 1992

最尤法によるポアソン回帰入門, カクワークス社, 2021

非売品, 無断複製を禁ずる

第 12 回 続高橋セミナー
層別因子を含む探索的な回帰分析入門

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3 の 1005

2024 年 6 月 高橋 行雄

takahashi.stat@nifty.com , FAX : 03-3452-8035