

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 1 月 21 日

第 4 章 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紹介する。

第 4 章 目 次

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

次ページに続く

4.2. 直交表における線形モデル-----	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おけ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	
文献索引, 索引, 解析ファイル一覧-----	(155)

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

目 次 (全章)

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-----	55
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 1 因子実験-----	243
9. 前後差データの群間比較に潜む前値の影-----	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

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紹介する。

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， $(\theta_1, \theta_2, \dots, \theta_p)$ 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 を各 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，

$$\left. \begin{aligned} a_{k,k'} &= \sum_{i=1}^n x_{k,i} x_{k',i}, \quad k=0, 1, 2, \dots, p, \quad k'=0, 1, 2, \dots, p \\ B_k &= \sum_{i=1}^n x_{k,i} y_i, \quad k=0, 1, 2, \dots, p \end{aligned} \right\} \quad (4.15)$$

とおくと，式 (4.14) は，

$$\left. \begin{aligned} \hat{\theta}_0 a_{0,0} + \hat{\theta}_1 a_{0,1} + \dots + \hat{\theta}_p a_{0,p} &= B_0 \\ &\dots\dots\dots \\ \hat{\theta}_0 a_{k,0} + \hat{\theta}_1 a_{k,1} + \dots + \hat{\theta}_p a_{k,p} &= B_k \\ &\dots\dots\dots \\ \hat{\theta}_0 a_{p,0} + \hat{\theta}_1 a_{p,1} + \dots + \hat{\theta}_p a_{p,p} &= B_p \end{aligned} \right\} \quad (4.16)$$

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 \mathbf{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 \mathbf{X} の各行に列ベクトル $\boldsymbol{\theta}$ 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 \mathbf{X} を用いた線形モデル

			— デザイン行列 \mathbf{X} —				パラメータ	誤差	反応	
A	B	i	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	$\boldsymbol{\theta}$	$\boldsymbol{\varepsilon}$	\mathbf{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 Y の積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	θ^{\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	θ^{\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 $\text{Minverse}(X^T X \text{ の範囲})$ によって直接

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	0	0	0	0	0	0	6.0	0	0	0	0	0	0	θ^\wedge_0	=	0.167	0	0	0	0	0	0	0	72.20	B_0		
0	0.333	-0.167	0	0	0	0	0	0	4.0	2.0	0	0	0	0	θ^\wedge_1		0	0.333	-0.167	0	0	0	0	-2.10	B_1			
0	-0.167	0.333	0	0	0	0	0	0	2.0	4.0	0	0	0	0	θ^\wedge_2		0	-0.167	0.333	0	0	0	0	0.80	B_2			
0	0	0	0.167	0	0	0	0	0	0	0	6.0	0	0	0	θ^\wedge_3		0	0	0	0.167	0	0	0	-1.20	B_3			

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，画期的なことである．

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 $\hat{\boldsymbol{\theta}}$ との積は，元の $\hat{\boldsymbol{\theta}}$ 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， $\boldsymbol{\theta}$ の推定値 $\hat{\boldsymbol{\theta}}$ となる．回帰モデルでは， $\hat{\boldsymbol{\theta}}$ に代わりに $\hat{\boldsymbol{\beta}}$ を用いて式 (4.19) を次式

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (4.20)$$

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．

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 ε 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、 $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\theta}}$ により予測値 $\hat{\mathbf{y}}$ 、残差 $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$ 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 $\hat{\mathbf{y}}$ 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 \mathbf{X}				予測値 \mathbf{y}^\wedge の推定		誤差分散 $\sigma^{\wedge 2}$ の推定			
			$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	θ^\wedge	\mathbf{y}^\wedge	\mathbf{y}	\mathbf{y}^\wedge	$\varepsilon^\wedge = \mathbf{y} - \mathbf{y}^\wedge$	
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		$\sigma^{\wedge 2} =$ 0.0350	
								$df = n - p - 1 =$ 2			$\sigma^\wedge =$ 0.1871	

デザイン行列 \mathbf{X} とパラメータの推定値 $\hat{\boldsymbol{\theta}}$ の積の計算は、表 4.6 に示すように式 (4.9) に示した構造 (DE) モデルによる推定に対応する。

表 4.6 予測値の構造 (DE) モデルによる推定

A	B	i	μ^\wedge	α^\wedge_1	α^\wedge_2	β^\wedge_1	\mathbf{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 $\hat{\boldsymbol{\theta}}$ 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 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$

	$(\mathbf{X}^T \mathbf{X})^{-1}$	$(c_{k,k'})$	$\sigma^{\wedge 2}$	パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$				
θ^\wedge_0	0.1667	0	0	0.0058	0	0	0	θ^\wedge_0
θ^\wedge_1	0	0.3333	-0.1667	0	0.0117	-0.0058	0	θ^\wedge_1
θ^\wedge_2	0	-0.1667	0.3333	0	-0.0058	0.0117	0	θ^\wedge_2
θ^\wedge_3	0	0	0	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

$$Var(\hat{\theta}_1) = c_{1,1} \hat{\sigma}^2 = 0.3333 \times 0.0350 = 0.0117$$

となり， $\hat{\theta}_1$ と $\hat{\theta}_2$ の共分散は，

$$Cov(\hat{\theta}_1, \hat{\theta}_2) = c_{1,2} \hat{\sigma}^2 = -0.1667 \times 0.0350 = -0.0058$$

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

$$L = \sum_{k=0}^q l_k \theta_k \quad (4.25)$$

の推定量は，

$$\hat{L} = \sum_{k=0}^q l_k \hat{\theta}_k \quad (4.26)$$

となる．その分散は，

$$Var(\hat{L}) = \sum_k \sum_{k'} l_k l_{k'} c_{k,k'} \hat{\sigma}^2 \quad (4.27)$$

で与えられる．ゆえに，線形和 L の $100(1-\alpha)\%$ 信頼限界は，

$$\sum_k l_k \hat{\theta}_k \pm t_{\alpha}(n-p-1) \sqrt{\sum_k \sum_{k'} l_k l_{k'} c_{k,k'} \hat{\sigma}^2} \quad (4.28)$$

となる．

線形和 L に対する分散 $Var(\hat{L})$ 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

$(l_k l_{k'} c_{k,k'} \hat{\sigma}^2)$ の和を計算することが、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 $\boldsymbol{l}^{(1)}$

$$\begin{aligned}\boldsymbol{l}^{(1)} &= [l_0^{(1)} \quad l_1^{(1)} \quad l_2^{(1)} \quad l_3^{(1)}] \\ &= [1 \quad 1 \quad 0 \quad 1]\end{aligned}\quad (4.29)$$

としたときに、式 (4.26) は、

$$\hat{L}^{(1)} = \sum_{k=0}^3 l_k^{(1)} \hat{\theta}_k = \boldsymbol{l}^{(1)} \hat{\boldsymbol{\theta}} = 11.0000 \quad (4.30)$$

	$l_0^{(1)}$	$l_1^{(1)}$	$l_2^{(1)}$	$l_3^{(1)}$			$y^{(1)}$
	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	$\hat{\boldsymbol{\theta}}$		$L^{(1)}$
$\boldsymbol{l}^{(1)}$	1	1	0	1	12.0333	=	11.0000
					-0.8333		
					0.6167		
					-0.2000		

に示すように、Excel の $\text{Mmult}(\boldsymbol{l}^{(1)}$ の範囲, $\hat{\boldsymbol{\theta}}$ 範囲) 関数で計算できる．式 (4.27) の分散の推定量は、

$$\hat{Var}(\hat{L}^{(1)}) = \sum_k \sum_{k'} l_k^{(1)} l_{k'}^{(1)} c_{k,k'} \hat{\sigma}^2 = \boldsymbol{l}^{(1)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) (\boldsymbol{l}^{(1)})^T = 0.0233 \quad (4.31)$$

	$l^{(1)}_0$	$l^{(1)}_1$	$l^{(1)}_2$	$l^{(1)}_3$	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 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ を用いた分散の計算

			$l_0^{(i)}$	$l_1^{(i)}$	$l_2^{(i)}$	$l_3^{(i)}$		y^{\wedge}	分散	幅	95%信頼区間	
$L^{(i)}$	A	B	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	$\hat{\boldsymbol{\theta}}$	L^{\wedge}	$Var(L^{\wedge})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
$L^{(1)}$	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
							$t_{0.05}(6-3-1)=T.\text{inv.}2T(0.05, 2)=$			4.3027		

$\hat{L}^{(1)}$ に対する 95%信頼区間は、式 (4.28) によって、

$$\left. \begin{aligned} & \sum_k l_k^{(1)} \hat{\theta}_k \pm t_{0.05}(n-p-1) \sqrt{\sum_k \sum_{k'} l_k^{(1)} l_{k'}^{(1)} c_{k,k'} \hat{\sigma}^2} \\ &= \hat{L}^{(1)} \pm t_{0.05}(6-3-1) \sqrt{\text{Var}(\hat{L}^{(1)})} \\ &= 11.000 \pm 4.3027 \sqrt{0.0233} \\ &= (10.3428, 11.6572) \end{aligned} \right\} \quad (4.32)$$

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%信頼区間が一括計算できることは、Excel の基本機能であり、四則演算のごとく行列計算ができることは画期的である。

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、マーカを○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

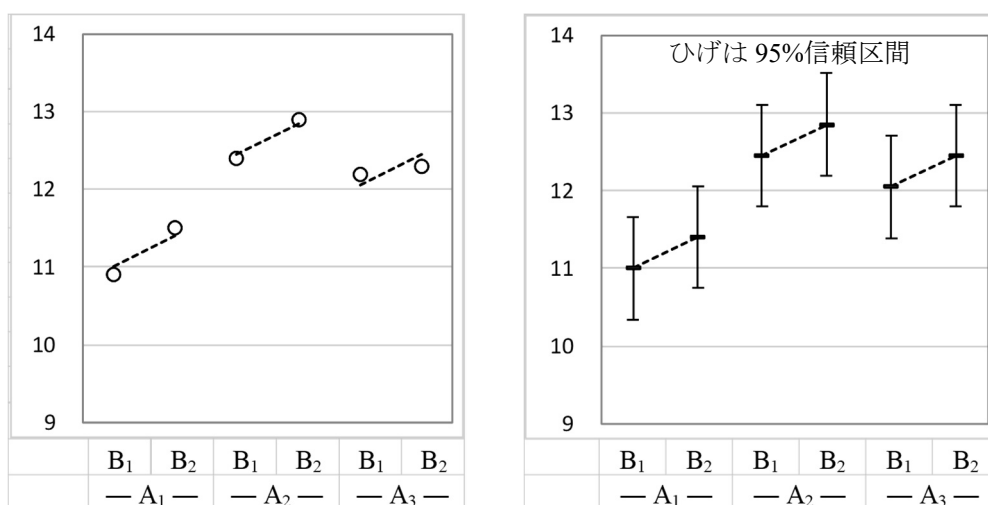


図 4.1 因子 A および B の組み合わせによる推定値と 95%信頼区間

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

$$\left. \begin{aligned} Var(\hat{y}_{1,1}) &= Var(\hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3) \\ &= Var(\hat{\theta}_0) + Var(\hat{\theta}_1) + Var(\hat{\theta}_3) \\ &\quad + 2Cov(\hat{\theta}_0, \hat{\theta}_1) + 2Cov(\hat{\theta}_0, \hat{\theta}_3) + 2Cov(\hat{\theta}_1, \hat{\theta}_3) \\ &= 0.0058 + 0.0117 + 0.0058 + 0.0 + 0.0 + 0.0 \\ &= 0.0233 \end{aligned} \right\} \quad (4.34)$$

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、誤差分散 $\hat{\sigma}^2 = 0.0350$ を用いて、次式で計算している。

$$\left. \begin{aligned} \frac{1}{n_e} &= \frac{1 + \text{推定に無視しない要因の自由度の和}}{\text{実験数 } N} = \frac{1+3}{6} \\ Var(y_{11}) &= \hat{\sigma}^2 \times \frac{1}{n_e} = 0.0350 \times \frac{4}{6} = 0.0233 \end{aligned} \right\} \quad (4.35)$$

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、平均値そのものの自体はピュアな推定値とは言いがたい。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

$$l_3^{(7)} = \frac{[B_1 : x_{3,i} = 1] + [B_2 : x_{3,i} = -1]}{2} = \frac{1 + (-1)}{2} = 0 \quad (4.36)$$

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

			$l_0^{(i)}$	$l_1^{(i)}$	$l_2^{(i)}$	$l_3^{(i)}$		線形和	分散	幅	95%信頼区間	
$L^{(i)}$	A	B	$x_{0,i}$	$x_{1,i}$	$x_{2,i}$	$x_{3,i}$	θ^\wedge	L^\wedge	$Var(L^\wedge)$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
$L^{(7)}$	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、

$$(A_2 - A_1) : \begin{cases} \mathbf{l}^{(11)} = \mathbf{l}^{(8)} - \mathbf{l}^{(7)} \\ = [1 \quad 0 \quad 1 \quad 0] - [1 \quad 1 \quad 0 \quad 0] \\ = [0 \quad -1 \quad 1 \quad 0] \end{cases} \quad (4.37)$$

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，

$$\begin{cases} B_1 : I^{(13)} = [1 \ 0 \ 0 \ 1] \\ B_2 : I^{(14)} = [1 \ 0 \ 0 \ -1] \end{cases} \quad (4.39)$$

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．

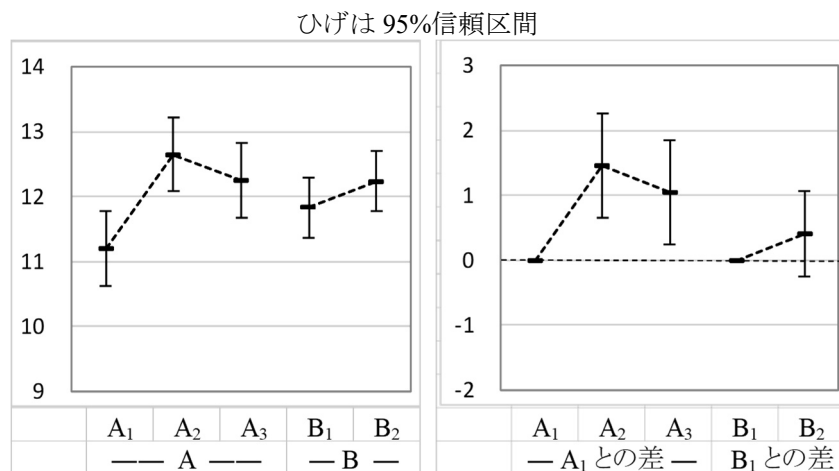


図 4.2 因子 A および B の水準平均および第 1 水準と差の予測プロファイル

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，

$$\begin{aligned} \boldsymbol{l}^{(17)} &= \boldsymbol{l}^{(4)} - \boldsymbol{l}^{(1)} \\ &= [1 \ 0 \ 1 - 1] - [1 \ 1 \ 0 \ 1] \\ &= [0 - 1 \ 1 - 2] \end{aligned} \quad (4.41)$$

となり，

$$\hat{Var}[\hat{L}^{(17)}] = \sum_k \sum_{k'} l_k^{(17)} l_{k'}^{(17)} c_{k,k'} \hat{\sigma}^2 = \boldsymbol{l}^{(17)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \boldsymbol{l}^{(17)T} = 0.0583 \quad (4.42)$$

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	$l_0^{(i)}$	$l_1^{(i)}$	$l_2^{(i)}$	$l_3^{(i)}$	θ^{\wedge}	線形和 L^{\wedge}	分散 $Var(L^{\wedge})$	幅 $t_{0.05} \times SE$	95%信頼区間	
$L^{(1)}$	A ₁	B ₁	x_{0i}	x_{1i}	x_{2i}	x_{3i}					L95%	U95%
$L^{(1)}$	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	

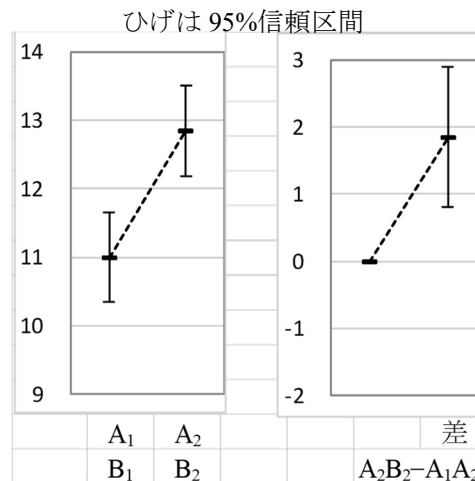


図4.3 現行水準と最適水準との差の予測プロファイル

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、シグマ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

$$\left. \begin{aligned} \alpha_1 + \alpha_2 &= 0, & \beta_1 + \beta_2 &= 0, & \gamma_1 + \gamma_2 &= 0, \\ (\alpha\beta)_{11} + (\alpha\beta)_{12} + (\alpha\beta)_{21} + (\alpha\beta)_{22} &= 0, \\ (\alpha\gamma)_{11} + \dots \end{aligned} \right\} \quad (4.47)$$

を用い

$$\left. \begin{aligned} \alpha_1 &= \alpha, & \alpha_2 &= -\alpha \\ \beta_1 &= \beta, & \beta_2 &= -\beta \\ \gamma_1 &= \gamma, & \gamma_2 &= -\gamma \\ (\alpha\beta)_{11} &= (\alpha\beta)_{22} = \alpha\beta, & (\alpha\beta)_{12} &= (\alpha\beta)_{21} = -\alpha\beta \\ (\alpha\gamma)_{11} &= (\alpha\gamma)_{22} = \alpha\gamma, & \dots \\ & \dots \end{aligned} \right\} \quad (4.48)$$

とする. 式 (4.48) を式 (4.44) に代入すると

y_{111}	$= \mu$	$+ \alpha$	$+ \beta$	$+ \alpha\beta$	$+ \gamma$	$+ \alpha\gamma$	$+ \beta\gamma$	$+ \alpha\beta\gamma$	$+ \varepsilon_{111}$
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}	$= \mu$	$- \alpha$	$- \beta$	$+ \alpha\beta$	$- \gamma$	$+ \alpha\gamma$	$+ \beta\gamma$	$- \alpha\beta\gamma$	$+ \varepsilon_{222}$

$$\left. \right\} \quad (4.49)$$

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ので, (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

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-		

ることが望ましい。ダミー変数といえ（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書籍は、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 \mathbf{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 \mathbf{X} を用いた正規方程式を示す。積和行列 $(\mathbf{X}^T \mathbf{X})$ 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 \mathbf{X}^T						デザイン行列 \mathbf{X}						転置行列 \mathbf{X}^T						\mathbf{Y}			
		1	2	3	4	5	6	8	x_0	x_1	x_2	x_3	x_4	θ^\wedge	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	θ^\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	θ^\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	θ^\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	θ^\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										

$X^T X (a_{k,k'})$					θ^\wedge	$X^T Y$	
7	1	1	-1	-1	θ^\wedge_0	= 138.0	B_0
1	7	-1	1	1	θ^\wedge_1	44.0	B_1
1	-1	7	1	1	θ^\wedge_2	50.0	B_2
-1	1	1	7	-1	θ^\wedge_3	2.0	B_3
-1	1	1	-1	7	θ^\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，

$$\left. \begin{aligned} \hat{y}_i &= \hat{\theta}_0 x_{0,i} + \hat{\theta}_1 x_{1,i} + \cdots + \hat{\theta}_4 x_{4,i} \\ &= 19.17 x_{0,i} + 3.583 x_{1,i} + 4.333 x_{2,i} + 2.167 x_{3,i} + 1.917 x_{4,i} \end{aligned} \right\} \quad (4.51)$$

$\mu \qquad A \qquad B \qquad C \qquad D$

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

$(X^T X)^{-1} (c_{k,k'})$					$X^T X (a_{k,k'})$					θ^\wedge	$(X^T X)^{-1} (c_{k,k'})$					$X^T Y$
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	θ^{\wedge}	y^{\wedge}	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		$\hat{\sigma}^{\wedge}=$	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
θ^{\wedge}_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
θ^{\wedge}_1	-0.042	0.167	0.042	-0.042	-0.042		-0.090	0.361	0.090	-0.090	-0.090	0.6009
θ^{\wedge}_2	-0.042	0.042	0.167	-0.042	-0.042		-0.090	0.090	0.361	-0.090	-0.090	0.6009
θ^{\wedge}_3	0.042	-0.042	-0.042	0.167	0.042		0.090	-0.090	-0.090	0.361	0.090	0.6009
θ^{\wedge}_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

デザイン行列 X									Y	分散分析表, 「定数に 0 を使用」を off				
No.	x_0	x_1	x_2	x_3	x_4	θ	ε	y		自由度	変動	分散	分散比	
1	1	1	1	1	1	θ_0	ε_1	32			4	303.0952	75.7738	34.9725
2	1	1	1	-1	-1	θ_1	ε_2	23			2	4.3333	2.1667	
3	1	1	-1	1	-1	θ_2	ε_3	19			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			係数	標準誤差	t	P -値
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

=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。

$$t_A^2 = F_A = \frac{S_A / 1}{\hat{\sigma}^2} \quad (4.53)$$

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。

$$\begin{aligned} S_A &= t_A^2 \hat{\sigma}^2 \\ &= 5.9630^2 \times 2.1667 \\ &= 77.0417 \end{aligned} \quad (4.54)$$

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	$\sigma^{\wedge 2}$	$t^2 \sigma^{\wedge 2}$
	df	平方和 S	平均平方	F				平方和 S
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 A_1 は一致するが、 A_2 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、このような推定されたパラメータ $\hat{\boldsymbol{\theta}}$ 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、 A_1 および A_2 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

					$l^{(i)}_0$	$l^{(i)}_1$	$l^{(i)}_2$	$l^{(i)}_3$	$l^{(i)}_4$		線形和	分散	幅	95%信頼区間	
$L^{(i)}$	A	B	C	D	x_0	x_1	x_2	x_3	x_4	θ^\wedge	L^\wedge	$Var(L^\wedge)$	$t_{0.05 \times SE}$	$L_{95\%}$	$U_{95\%}$
$L^{(1)}$	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

					$l^{(i)}_0$	$l^{(i)}_1$	$l^{(i)}_2$	$l^{(i)}_3$	$l^{(i)}_4$		線形和	分散	幅	95%信頼区間	
$L^{(i)}$	A	B	C	D	x_0	x_1	x_2	x_3	x_4	θ^{\wedge}	L^{\wedge}	$Var(L^{\wedge})$	$t_{0.05} \times SE$	$L_{95\%}$	$U_{95\%}$
$L^{(9)}$	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					

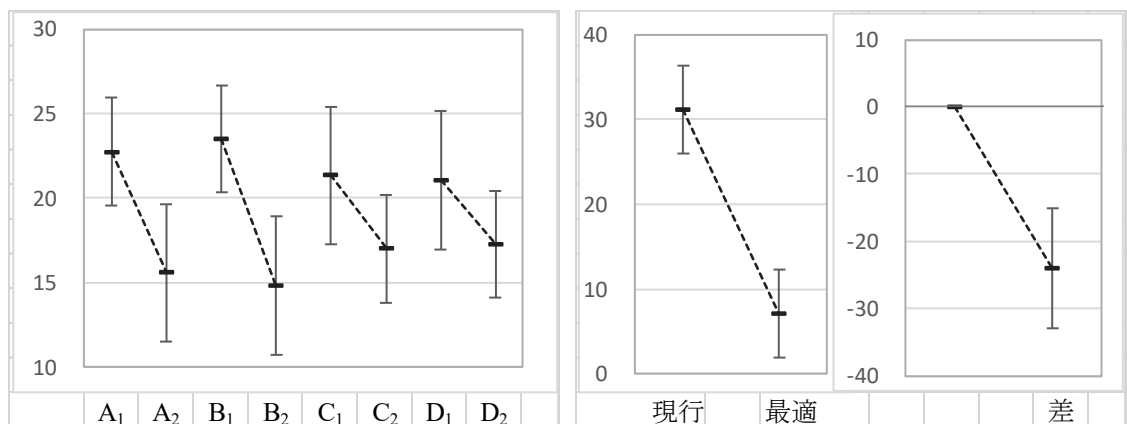


図 4.4 要因ごとの水準平均および現行水準と最適水準についての推定

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	1	32
2	2	1	1	2	2	23	2	1	1	1	-1	-1	23
3	3	1	2	1	2	19	3	1	1	-1	1	-1	19
4	4	1	2	2	1	17	4	1	1	-1	-1	1	17
5	5	2	1	1	2	19	5	1	-1	1	1	-1	19
6	6	2	1	2	1	20	6	1	-1	1	-1	1	20
7	7	2	2	1	1	.	7
8	8	2	2	2	2	8	8	1	-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 乗法」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 での分散分析表は、「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 年に新た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できていたことは、驚嘆すべきことである．

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。

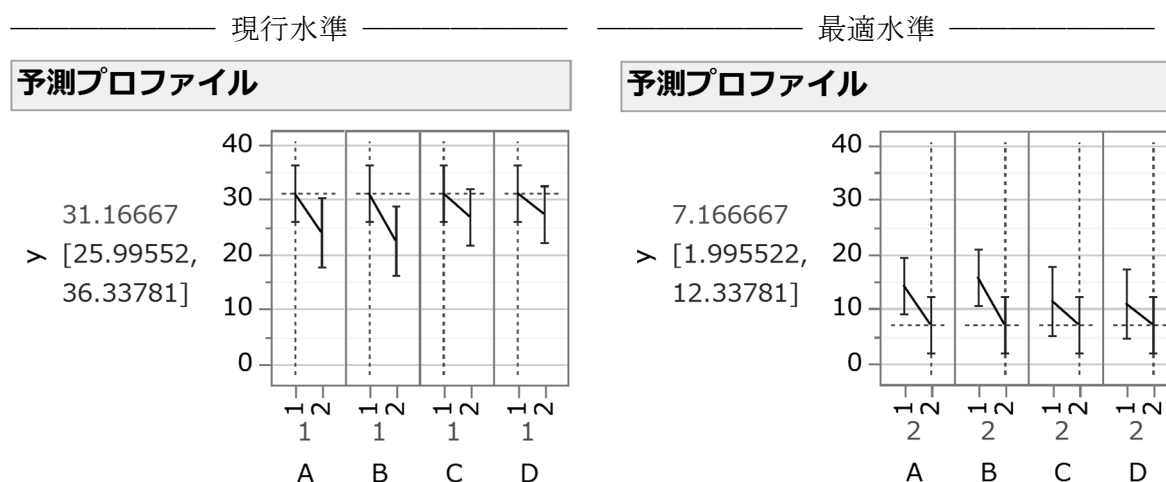


図 4.5 予測プロファイルを用いた現行水準と最適水準の推定
表 4.21 に示した Excel による結果に一致する。

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 の生成方式に精通し，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．

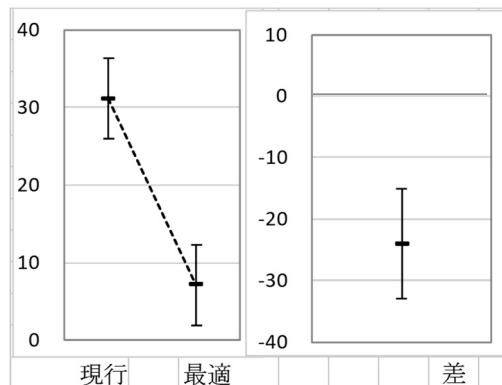
表 4.25 「カスタム検定」による謎めいたパラメータ設定

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

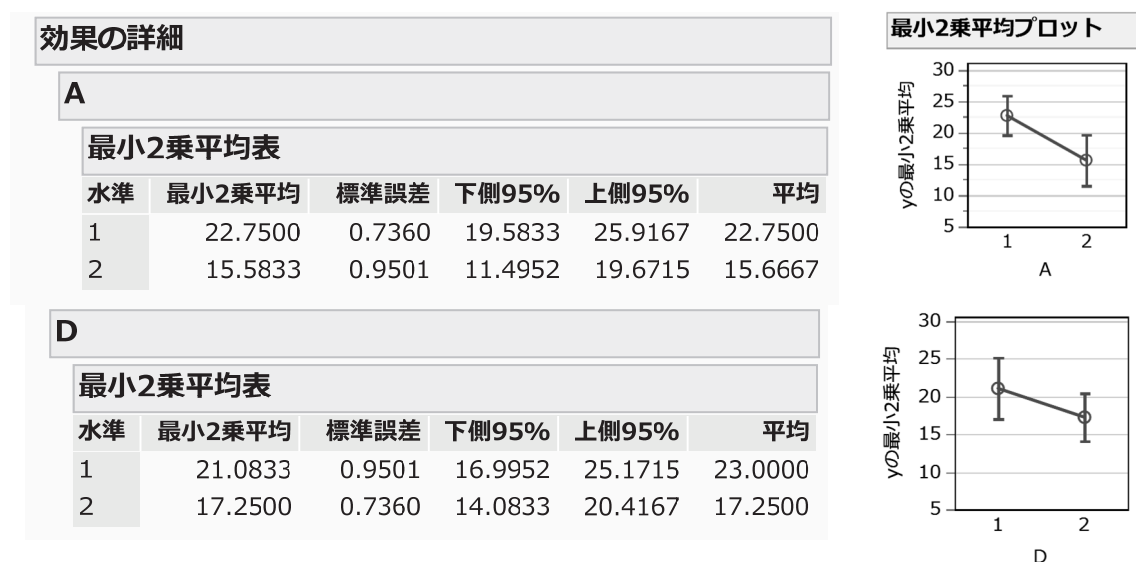
最小 2 乗平均

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」について理解を深められる。

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 水準の合計を求める」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 (1) (2) (3)			No.	列 番 (1) (2) (3)			No.	列 番 (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ことは、画期的である。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、素晴らしい。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。

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、(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) 型ダミー変数を採用している 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

$$\widehat{A_2 B_2 C_2 D_2} = 7.1667 \quad (4.59)$$

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、

$$\hat{\mu} = 7.1667 + \frac{7.1667 + 8.6667 + 4.3333 + 3.8333}{2} = 19.1667 \quad (4.60)$$

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、

$$\hat{A}_1 = 7.1667 + 7.1667 + \frac{8.6667 + 4.3333 + 3.8333}{2} = 22.7500 \quad (4.61)$$

であり、A₂ 水準の最小 2 乗平均は

$$\hat{A}_2 = 7.1667 + 0.0 + \frac{8.6667 + 4.3333 + 3.8333}{2} = 15.5833 \quad (4.62)$$

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% 信頼限界	
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% 信頼限界	
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% 信頼限界	
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% 信頼限界	
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

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」としてまとめたのであるが、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。

第4章 文献索引

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
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
新村(1983) - 重回帰分析における掃き出し演算子	126
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・大橋・芳賀(1989) - SASによる実験データの解析	121
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







第4章 索引

あ R が脚光を浴びている - 無償の統計ソフト	149	か 画期的 - JMP の使い勝手	149
朝香ら(1988) - 直交表の歴史的背景	148	逆行列 - Minverse () 関数	125
パラメータ - 共分散行列	127	- 掃き出し法	125
1次従属 - 制約条件	136	- パズル的な代数計算	126
一次式 - 回帰モデル	122	逆行列の計算 - Gauss-Doolittle 法	126
(1, 0)型 - GLMプロシジャ	150	- 楠ら(1995)	126
- 切片の推定値は?	152	95%信頼区間 - 折れ線グラフ	130
- (0, 1)型	153	- 組合せ水準	129
(1, 0)型ダミー変数 - パラメータ推定値	152	- 現行水準と最適水準の差	133
(1, 2)型 - (1, -1)基本型	148	- 水準の差	143
- L_8 直交表	137	- 水準間の差の推定	132
(1, 2)型直交表 - (1, -1)対比型	135	- 線形和の推定	128
(1, -1)基本型 - (0, 1)型	148	- 予測プロファイル	146
- (1, 2)型	148	共分散行列 - パラメータ	127, 140
(1, -1)対比型 - (1, 2)型直交表	135	行列計算 - 線形モデル	137
- コーディングのテーブル	144	楠ら(1995) - 逆行列の計算	126
- JMPの解析用ファイル	144	- 欠測値がある直交表	138
一般式を拡張 - 合成分散	131	- 線形モデル・線形推定論	121
Excel による行列計算 - 線形推定論	123	- 2因子実験のデータ	121
Excel の回帰分析 - 欠測値がある直交表	140	組合せによる推定値 - 現行水準	146
estimate ステートメント - 最適水準の推定	152	組合せ水準 - 95%信頼区間	129
- 水準の差の推定	152	欠測セルがある場合 - 最小2乗平均	148
HTML形式 - GLMプロシジャの出力	151	欠測値 - 直交表	138
Minverse () 関数 - 逆行列	125	欠測値がある場合 - 各種の推定	142
lsmeans - 最小2乗平均	142	欠測値がある直交表 - Excel の回帰分析	140
- 最小2乗平均	147	- 楠ら(1995)	138
lsmeans ステートメン - 最小2乗平均	152	- SAS/GLM	150
- L_8 直交表 - (1, 2)型	137	- 正規方程式の解	139
- 線形モデル	135	- 線形モデル	138
- データの構造式	135	- 分散分析表	141
- デザイン行列 X	137	現行水準 - 組合せによる推定値	146
折れ線グラフ - 95%信頼区間	130	- 最適水準の差	143
OnDemand SAS - GLMプロシジャ	150	現行水準と最適水準の差 - 95%信頼区間	133
- 無償で継続的に	149	交互作用 - 成分表示	138
か 回帰モデル - 一次式	122	合成分散 - 一般式を拡張	131
- 構造モデル	121	構造(DE)モデル - 推定	127
- 線形モデル	121	- 線形モデル	135
Gauss-Doolittle 法 - 逆行列の計算	126	- 母数に関する制約	122
各種のダミー変数 - 歴史的背景	148	構造モデル - 回帰モデル	121
各種の推定 - 欠測値がある場合	142	- 線形モデル	121
カスタム検定 - 水準の差の推定	146	- DE モデル	121
- デザイン行列 X	147	コーディングのテーブル - (1, -1)対比型	144

か	誤差分散 - 不偏推定量	126	さ	線形モデル - 正規方程式	123
	- 残差の自由度	127		- デザイン行列 X	124
	- 残差平方和	127		- パラメータ(係数)	122
	誤差分散の推定 - 予測値の推定	140		線形モデル・線形推定論 - 楠ら(1995)	121
さ	最小2乗推定量 - 正規方程式	124		線形モデルを拡張 - 高橋ら(1989)	121
	最小2乗平均 - lsmeans	142, 147		線形推定論 - Excel による行列計算	123
	- lsmeans ステートメン	152		- 理論的な枠組み	121
	- 欠測セルがある場合	148		線形和 L - 最小2乗平均	142
	- 線形和 L	142		線形和 L に対する分散 - 2次形式	128
	- モデルのあてはめ	148		線形和の推定 - 95%信頼区間	128
	最小2乗法 - 残差平方和	123	た	タイプ I の平方和 - デフォルト	151
	最適水準の差 - 現行水準	143		タイプ III の平方和 - デフォルト	151
	最適水準の推定 - estimate ステートメント	152		高橋ら(1989) - GLMプロシジャ	150
	最良線形不偏推定量 - BLUE	124		- GLMプロシジャの計算方式	153
	SAS/GLM - 欠測値がある直交表	150		- 線形モデルを拡張	121
	残差の自由度 - 誤差分散	127		- 4 種の平方和	153
	残差平方和 - 誤差分散	127		田口の式 - 伝統的な実験計画法	131
	- 最小2乗法	123		- 無視しない要因の自由度	131
	残差平方和 Q - 偏微分	123		- 有効反復数 n_e	134
	GLMプロシジャ - (1, 0) 型	150		直交性を保ちたい - 補間	145
	- OnDemand SAS	150		直交表 - 欠測値	138
	- 高橋ら(1989)	150		- 実験計画法	135
	GLMプロシジャの計算方式 - 高橋ら(1989)	153		直交表の歴史的背景 - 朝香ら(1988)	148
	GLMプロシジャの出力 - HTML形式	151		デフォルト - タイプ I の平方和	151
	実験計画法 - 直交表	135		- タイプ III の平方和	151
	質的な変数に限定 - 平方和の分解	135		DE モデル - 構造モデル	121
	JMP - モデルのあてはめ	144		- Design of Experiment	121
	- 有償ソフト	149		データの構造式 - L8 直交表	135
	JMP の「モデルのあてはめ」 - 分散分析表	145		- 2因子実験	122
	JMP の解析用ファイル - (1, -1) 対比型	144		- 母数	121
	JMP の使い勝手 - 画期的	149		データは数値 - 文字変数として設定	144
	新村(1983) - 掃き出し演算子	126		手計算時代 - 正規法式の解	126
	水準の差 - 95%信頼区間	143		Design of Experiment - DE モデル	121
	水準の差の推定 - estimate ステートメント	152		デザイン行列 X - L8 直交表 L8	137
	- カスタム検定	146		- カスタム検定	147
	水準間の差の推定 - 95%信頼区間	132		- 正規方程式	124, 125, 139
	推定 - 構造(DE)モデル	127		- 正規方程式の解	125
	正規方程式 - 最小2乗推定量	124		- 線形モデル	124
	- 線形モデル	123		伝統的な実験計画法 - 田口の式	131
	- デザイン行列 X	124, 125, 139	な	2因子実験 - データの構造式	122
	正規方程式の解 - 欠測値がある直交表	139		2因子実験のデータ - 楠ら(1995)	121
	- 手計算時代	126		2次形式 - 線形和 L に対する分散	128
	- デザイン行列 X	125	は	掃き出し演算子 - 新村(1983)	126
	成分表示 - 交互作用	138		掃き出し法 - 逆行列	125
	制約条件 - 1次従属	136		パズル的な代数計算 - 逆行列	126
	切片の推定値は? - (1, 0) 型	152		パラメータ - 共分散行列	140
	(0, 1) 型 - (1, 0) 型	153		パラメータ(係数) - 線形モデル	122
	- (1, -1) 基本型	148		パラメータ推定値 - (1, 0) 型	152
	線形モデル - L_8 直交表	135		不偏推定量 - 誤差分散	127
	- 回帰モデル	121		BLUE - 最良線形不偏推定量	124
	- 行列計算	137		- best linear unbiased estimator	124
	- 欠測値がある直交表	138		分散分析表 - 欠測値がある直交表	141
	- 構造(DE)モデル	135		- JMP の「モデルのあてはめ」	145
	- 構造モデル	121		- 平方和の分解	135

は	平方和の分解 - 質的な変数に限定	135	ま	モデルのあてはめ - 最小2乗平均	148
	- 分散分析表	135		- JMP	144
	best linear unbiased estimator - BLUE	124	や	有効反復数 n_e - 田口の式	134
	偏微分 - 残差平方和 Q	123		有償ソフト - JMP	149
	補間 - 直交性を保ちたい	145		予測プロファイル - 95%信頼区間	146
	母数 - データの構造式	121		予測値の推定 - 誤差分散	127
	母数に関する制約 - 構造(DE)モデル	122		- 誤差分散の推定	140
ま	無視しない要因の自由度 - 田口の式	131		4種の平方和 - 高橋ら(1989)	153
	無償の統計ソフト - R が脚光を浴びている	149	ら	理論的な枠組み - 線形推定論	121
	無償で継続的に - OnDemand SAS	149		歴史的背景 - 各種のダミー変数	148
	文字変数として設定 - データは数値	144			

第4章 解析用ファイル一覧

サイズ	名前	種類
 64 KB	第04章_01_線形モデル_2因子	Microsoft Excel ワークシート
 23 KB	第04章_02_線形モデル_直交	Microsoft Excel ワークシート
 59 KB	第04章_03_線形モデル_直交欠測	Microsoft Excel ワークシート
 9 KB	第04章_04_L8欠測値	JMP Data Table
 1 KB	第04章_05_欠測値がある直交表_SAS	テキストドキュメント
 21 KB	第04章_05_線形モデル_SAS	Microsoft Excel ワークシート

空白ページ

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

第 4 章 欠測値がある直交表の線形モデルによる解析

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2024 年 1 月 21 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035