

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 20 日

第 3 章 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。

第 3 章 目 次

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

次ページに続く

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 \mathbf{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, 主効果も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	
文献索引, 索引, 解析ファイル -----	(121)

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 2 群比較 -----	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

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

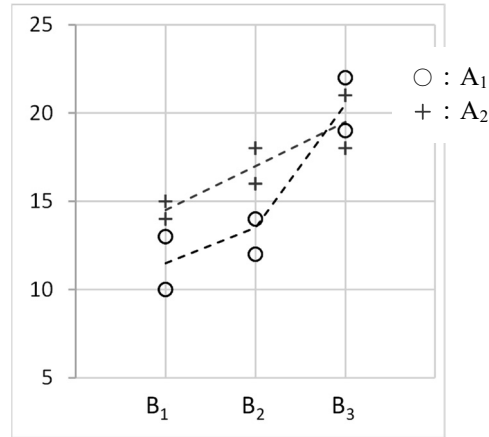


図 3.1 散布図に2因子実験の平均値の重ね書き

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、

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (2.1)$$

であり、主効果については、

$$\sum_i \alpha_i = 0, \quad \sum_j \beta_j = 0 \quad (2.2)$$

交互作用については、

$$\left. \begin{aligned} \sum_i (\alpha\beta)_{ij} &= 0, \quad i=1, 2 \\ \sum_j (\alpha\beta)_{ij} &= 0, \quad j=1, 2, 3 \end{aligned} \right\} \quad (2.3)$$

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、

$$\bar{y}_{...} = \left(\sum_i \sum_j \sum_k y_{ijk} \right) / N = 16.0, \quad N=12 : \text{データ総数} \quad (2.4)$$

$$S_T = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{...})^2 = 148.0 \quad (2.5)$$

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、

$$\bar{y}_{A_i} : \bar{y}_{i..} = \left(\sum_j \sum_k y_{ijk} \right) / (n_B \cdot n_K), \quad n_B = 3, \quad n_K = 2 : \text{繰り返し数} \quad (2.6)$$

$$S_A = \sum_i \sum_j \sum_k (\bar{y}_{i..} - \bar{y}_{...})^2 = 12.0 \quad (2.7)$$

によって求められる。計算の冗長性をなくすために

$$S_A = n_B n_K \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 \quad (2.8)$$

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		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}	$\bar{y}_{...}$	$y_{ijk} - \bar{y}_{...}$	\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

としたいが、Excel の SumSq () 関数を用いる場合は、式 (2.7) 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

$$\bar{y}_{B_j} : \bar{y}_{.j} = \left(\sum_i \sum_k y_{ijk} \right) / (n_A \cdot n_k), \quad n_A = 2, n_k = 2 : \text{繰り返し数} \quad (2.9)$$

$$S_B = \sum_i \sum_j \sum_k (\bar{y}_{.j} - \bar{y}_{...})^2 = 104.0 \quad (2.10)$$

で求める。

交互作用 $A \times B$ 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、

$$\bar{y}_{A_i B_j} : \bar{y}_{ij.} = \left(\sum_k y_{ijk} \right) / n_k \quad (2.11)$$

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 $(\hat{\alpha}_i + \hat{\beta}_j)$ に対応し、

$$\left. \begin{aligned} (\hat{y}'_{ij.} - \bar{y}_{...}) &= (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) \\ \hat{y}'_{ij.} &= \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...} \end{aligned} \right\} \quad (2.12)$$

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

$$S_{A \times B} = \sum_i \sum_j \sum_k (\bar{y}_{ij.} - \hat{y}'_{ij.})^2 = 14.0 \quad (2.13)$$

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

$$S_e = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij.})^2 = 18.0 \quad (2.14)$$

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

$$\left. \begin{aligned} S_T &= S_A + S_B + S_{A \times B} + S_e \\ &= 12.0 + 104.0 + 14.0 + 18.0 = 148.0 \end{aligned} \right\} \quad (2.15)$$

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 (2.1) 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，

$$\left. \begin{aligned} y_{111} &= \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 + (\widehat{\alpha\beta})_{11} + \hat{\varepsilon}_{111} \\ &= 16.0 + (-1.0) + (-3.0) + (-0.5) + (-1.5) = 10.0 \end{aligned} \right\} \quad (2.16)$$

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

$$\left. \begin{aligned} S_T &= \left(\sum_i \sum_j \sum_k y_{ijk}^2 \right) - CT \\ &= 3220 - \frac{192^2}{12} \\ &= 3220 - 3072 = 148.0 \end{aligned} \right\} \quad (2.17)$$

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，

$$\left. \begin{aligned} S_A &= \frac{\sum_i \left(\sum_j \sum_k y_{ijk} \right)^2}{n_B n_K} - CT \\ &= \frac{8100 + 10404}{3 \times 2} - 3072 = 12.0 \end{aligned} \right\} \quad (2.18)$$

$$\left. \begin{aligned} S_B &= \frac{\sum_j \left(\sum_i \sum_k y_{ijk} \right)^2}{n_A n_K} - CT \\ &= \frac{2704 + 3600 + 6400}{2 \times 2} - 3072 = 104.0 \end{aligned} \right\} \quad (2.19)$$

で求める．交互作用の計算のために S_{AB}

$$\left. \begin{aligned} S_{AB} &= \frac{\sum_j \sum_i \left(\sum_k y_{ijk} \right)^2}{n_K} - CT \\ &= \frac{529 + 676 + 1681 + 841 + 1156 + 1521}{2} - 3072 = 130.0 \end{aligned} \right\} \quad (2.20)$$

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

$$S_e = S_T - S_{AB} = 148.0 - 130.0 = 18.0 \quad (2.21)$$

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

$$S_{A \times B} = S_{AB} - S_A - S_B = 130.0 - 12.0 - 104.0 = 14.0 \quad (2.22)$$

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		
				$n_{\cdot ik}=6$		$n_{i \cdot k}=4$		$n_{\cdot \cdot k}=2$			
	$N=$	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 (2.22) 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 $\bar{y}_{A_i B_j}$ から期待平均 $\hat{y}'_{A_i B_j}$ 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 (2.12) 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 $\bar{y}_{A_i B_j}$ と期待平均 $\hat{y}'_{A_i B_j}$

			セル平均	期待平均	差
			\bar{y}_{AB}	\hat{y}'_{AB}	A×B
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
$S_{A \times B} = 2 \times \text{平方和} =$					14.0

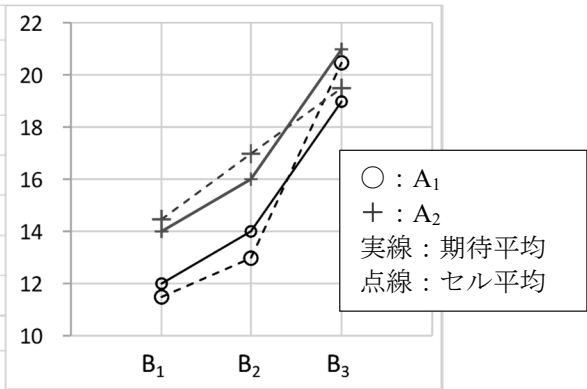


図 3.2 交互作用の図的表現

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×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}

$$\bar{y}_{1..} = \frac{y_{111} + y_{112} + y_{121} + y_{122} + y_{131} + y_{132}}{6} \quad (2.23)$$

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

$$\left. \begin{aligned} \text{Var}(\bar{y}_{1..}) &= \text{Var}\left(\frac{y_{111}}{6} + \frac{y_{112}}{6} + \frac{y_{121}}{6} + \frac{y_{122}}{6} + \frac{y_{131}}{6} + \frac{y_{132}}{6}\right) \\ &= \frac{6\hat{\sigma}^2}{6^2} = \frac{\hat{\sigma}^2}{6} \end{aligned} \right\} \quad (2.24)$$

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
		$\sigma^2 =$	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。

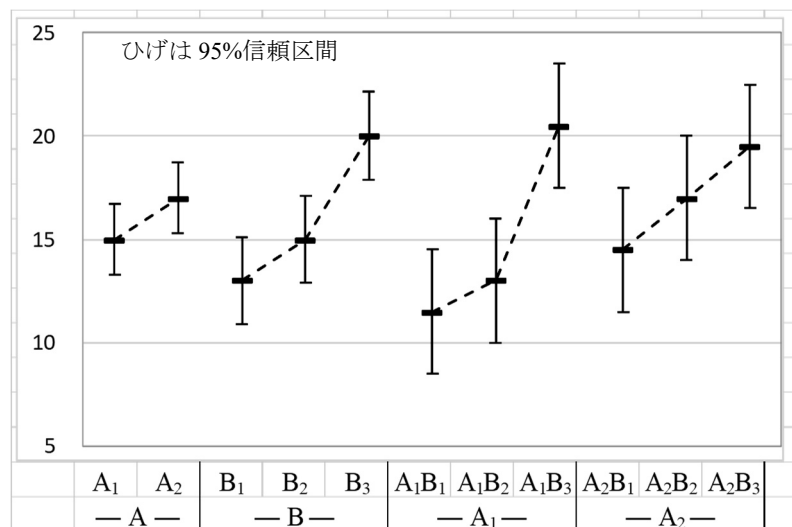


図 3.3 因子 A、因子 B、交互作用 A×B の水準平均と 95% 信頼区間

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、画期的なことである。

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	$t_{0.05} \times SE$	$L_{95\%}$	$U_{96\%}$
$A_1 - A_1$	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
$A_1 B_1 - A_1 B_1$	0.0	0	0.000	0.000	0.000	0.000
$A_1 B_2 - A_1 B_1$	1.5	1	3.000	4.238	-2.738	5.738
$A_1 B_3 - A_1 B_1$	9.0	1	3.000	4.238	4.762	13.238
$A_2 B_1 - A_2 B_1$	0.0	0	0.000	0.000	0.000	0.000
$A_2 B_2 - A_2 B_1$	2.5	1	3.000	4.238	-1.738	6.738
$A_2 B_3 - A_2 B_1$	5.0	1	3.000	4.238	0.762	9.238
$\sigma^2 =$	3.00		$t_{0.05}(6) =$	2.447		

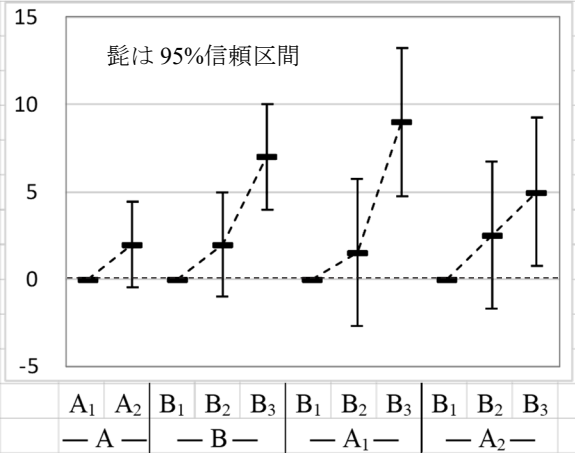


図 3.4 第1水準との差の 95%信頼区間

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

$$\left. \begin{aligned} Var(\bar{y}_{2..} - \bar{y}_{1..}) &= Var(\bar{y}_{2..}) + Var(\bar{y}_{1..}) \\ &= \frac{\sigma^2}{6} + \frac{\sigma^2}{6} = \frac{\sigma^2}{3} \end{aligned} \right\} \quad (2.25)$$

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 (2.1) 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 (2.2), (2.3) 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

$$y_{ijk} = \theta_0 x_0 + \theta_1 a_{1,i} + \theta_2 b_{1,j} + \theta_3 b_{2,j} + \theta_4 (a_1 b_1)_{ij} + \theta_5 (a_1 b_2)_{ij} + \varepsilon_{ijk} \quad (2.26)$$

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, 「定数を 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			係数	標準誤差	t	P -値
		19	1	1	-1	-1	-1	-1	θ^{\wedge}_0 切片 x_0	16.00	0.5000	32.0000	0.0000	
A ₂	B ₁	15	1	-1	1	0	-1	0	θ^{\wedge}_1 a_1	-1.00	0.5000	-2.0000	0.0924	
		14	1	-1	1	0	-1	0	θ^{\wedge}_2 b_1	-3.00	0.7071	-4.2426	0.0054	
	B ₂	16	1	-1	0	1	0	-1	θ^{\wedge}_3 b_2	-1.00	0.7071	-1.4142	0.2070	
		18	1	-1	0	1	0	-1	θ^{\wedge}_4 a_1b_1	-0.50	0.7071	-0.7071	0.5060	
	B ₃	21	1	-1	-1	-1	1	1	θ^{\wedge}_5 a_1b_2	-1.00	0.7071	-1.4142	0.2070	
		18	1	-1	-1	-1	1	1						

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 $\Sigma(\hat{\theta})$

$$\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2, \quad \hat{\sigma}^2 : \text{誤差分散の推定値} \quad (2.27)$$

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 (2.26) に対応したパラメータ θ_i の推定値が“係数”として示されていて、それぞれについて標準誤差 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,

$$\text{推定値 } \hat{\beta}_3 = -\hat{\theta}_2 - \hat{\theta}_3 = 3.0 + 1.0 = 4.0 \quad (2.28)$$

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 (2.1) のデータの構造式の効果 ($\hat{\alpha}_i$, $\hat{\beta}_j$, $(\hat{\alpha}\hat{\beta})_{ij}$) 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 $\bar{y}_{...}$ 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 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

$$L = \sum_i l_i \hat{\theta}_i = \mathbf{l} \hat{\theta} \quad (2.29)$$

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

$$\mathbf{l}^{(5)} = [0 \quad 0 \quad -1 \quad -1 \quad 0 \quad 0] \quad (2.30)$$

と設定することにより,

$$\left. \begin{aligned} \hat{\beta}_3 &= \mathbf{l}^{(5)} \hat{\theta} \\ &= 0 \times 16.0 + 0 \times (-1.0) - 1 \times (-3.0) - 1 \times (-1.0) + 0 \times (-0.5) + 0 \times (-1.0) \\ &= 3.0 + 1.0 = 4.0 \end{aligned} \right\} \quad (2.31)$$

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, $\text{Var}(\hat{\beta}_1) = \text{Var}(\hat{\beta}_2) = 0.7071^2 = 0.50$ であることが示されている. 分散の加法性を仮定して $\text{Var}(\hat{\beta}_3) = \text{Var}(-\hat{\beta}_1 - \hat{\beta}_2)$ を $\text{Var}(\hat{\beta}_3) = 2 \times 0.50 = 1.00$ としてよいのだろうか.

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

$$\text{Var}(\mathbf{l}\hat{\theta}) = \mathbf{l}\Sigma(\hat{\theta})\mathbf{l}^T \quad (2.32)$$

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 $\text{Cov}(\hat{\theta}_2, \hat{\theta}_3) = -0.25$ であり、合成分散の一般式を用いて

$$\begin{aligned} \text{Var}(\hat{\beta}_3) &= \text{Var}(-\hat{\beta}_2 - \hat{\beta}_3) \\ &= \text{Var}(\hat{\beta}_2) + 2\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) + \text{Var}(\hat{\beta}_3) \\ &= 0.50 + 2 \times (-0.25) + 0.50 = 0.50 \end{aligned} \quad (2.33)$$

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	θ^{\wedge}	推定値 $\mathbf{l}\theta^{\wedge}$	分散 $\text{Var}(\mathbf{l}\theta^{\wedge})$	幅 $t_{0.05} \times SE$
			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

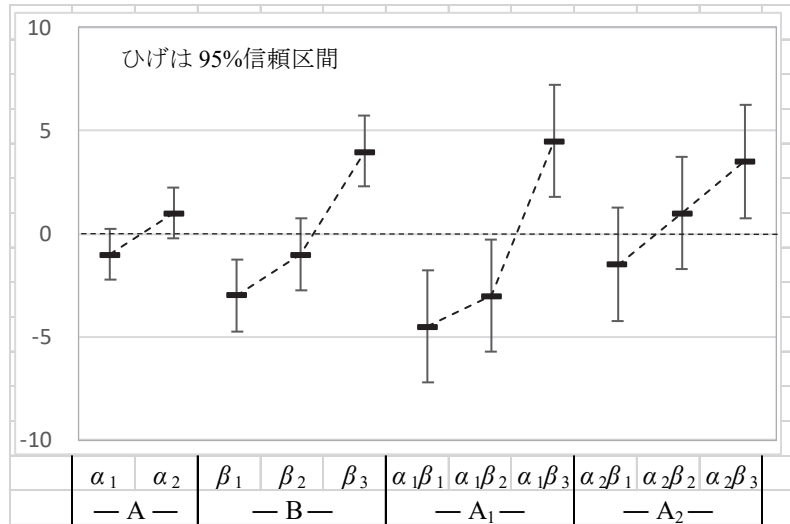


図 3.5 要因効果についての線グラフによる予測プロファイル

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、

$$\bar{y}_{1..} = \bar{y}_{...} + \alpha_1 \quad (2.34)$$

なので、 α_1 について解くと

$$\alpha_1 = \bar{y}_{1..} - \bar{y}_{...} \quad (2.35)$$

が得られる。効果 α_1 の分散 $Var(\alpha_1)$ は、

$$Var(\alpha_1) = Var(\bar{y}_{1..} - \bar{y}_{...}) \quad (2.36)$$

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

$$\begin{aligned} Var(\bar{y}_{1..} - \bar{y}_{...}) &= Var\left(\frac{y_{111} + y_{112} + \cdots + y_{132}}{6} - \frac{y_{111} + y_{112} + \cdots + y_{211} + y_{212} + \cdots + y_{232}}{12}\right) \\ &= Var\left(\frac{y_{111} + y_{112} + \cdots + y_{132}}{12} - \frac{y_{211} + y_{212} + \cdots + y_{232}}{12}\right) \\ &= \frac{6\hat{\sigma}^2}{12^2} + \frac{6\hat{\sigma}^2}{12^2} \\ &= \frac{\hat{\sigma}^2}{12} = \frac{3.00}{12} = 0.250 \end{aligned} \quad (2.37)$$

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 \mathbf{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 \mathbf{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 \mathbf{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 \mathbf{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 $(\mathbf{X}^T \mathbf{X})$ 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 \mathbf{X} の直交性の確認

	$\mathbf{X}^T \mathbf{X}$					
	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(\mathbf{X} の範囲), \mathbf{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」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 ₁	B ₁	1	11.5
		1	13		B ₂	2	13.0
		2	14		B ₃	3	20.5
	B ₂	2	12	A ₂	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 $\bar{y}_{...} =$				15.57			

○ : A₁
+ : A₂

B	A	y
B ₁	A ₁	10
		13
		14
	A ₂	14.5
		17.0
		20.5
B ₂	A ₁	12
		15
		11
	A ₂	14.5
		17.0
		19.5
B ₃	A ₁	22
		19
	A ₂	14.5
		17.0
		19.5
		16.0

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_{1j}) 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 (2.1) で示した場合と同様で

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (2.38)$$

あるが, 制約条件は, 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, $n_{i\cdot}$, $n_{\cdot j}$ を周辺のデータ数として主効果は,

$$\sum_i n_{i\cdot} \alpha_i = 0, \quad \sum_j n_{\cdot j} \beta_j = 0 \quad (2.39)$$

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

$$\left. \begin{aligned} \sum_j n_{ij}(\alpha\beta)_{ij} &= 0, \quad i=1, 2 \\ \sum_i n_{ij}(\alpha\beta)_{ij} &= 0, \quad j=1, 2, 3 \end{aligned} \right\} \quad (2.40)$$

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(2.12)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		
A ₂	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		
	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、

$$\begin{aligned} S'_T &= S_A + S_B + S_{A \times B} + S_e \\ &= 21.43 + 114.10 + 16.10 + 26.00 = 177.62 \end{aligned} \quad (2.41)$$

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

$$p = \text{F.dist.RT}(8.9495, 5, 8) = 0.0039 \quad (2.42)$$

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、

$$\left. \begin{aligned} \hat{y}_{ijk} &= \hat{\theta}_0 x_0 + \hat{\theta}_1 a_1 + \hat{\theta}_2 b_1 + \hat{\theta}_3 b_2 + \hat{\theta}_4 a_1 b_1 + \hat{\theta}_5 a_1 b_2 \\ &= 16.0x_0 + (-1.0)a_1 + (-3.0)b_1 + (-1.0)b_2 + (-0.5)a_1 b_1 + (-1.0)a_1 b_2 \end{aligned} \right\} \quad (2.43)$$

推定されたパラメータは、繰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	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					
		15	1	1	0	1	0	1		係数	標準誤差	t	P -値
		11	1	1	0	1	0	1	θ^{\wedge}_0 切片 x_0	16.00	0.4983	32.1117	0.0000
	B_3	22	1	1	-1	-1	-1	-1	θ^{\wedge}_1 a_1	-1.00	0.4983	-2.0070	0.0797
		19	1	1	-1	-1	-1	-1	θ^{\wedge}_2 b_1	-3.00	0.7205	-4.1639	0.0031
A_2	B_1	15	1	-1	1	0	-1	0	θ^{\wedge}_3 b_2	-1.00	0.6719	-1.4884	0.1750
		14	1	-1	1	0	-1	0	θ^{\wedge}_4 $a_1 b_1$	-0.50	0.7205	-0.6940	0.5073
	B_2	16	1	-1	0	1	0	-1	θ^{\wedge}_5 $a_1 b_2$	-1.00	0.6719	-1.4884	0.1750
		18	1	-1	0	1	0	-1					
	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 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 $S_A^{(I)}$ 、 $S_{A+B}^{(I)}$ 、 $S_{A+B+A \times B}^{(I)}$ を求め、それらの差分

$$S_B^{(I)} = S_{A+B}^{(I)} - S_A^{(I)} \quad (2.44)$$

$$S_{A \times B}^{(I)} = S_{A+B+A \times B}^{(I)} - S_{A+B}^{(I)} \quad (2.45)$$

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

$$S_T = S_A^{(I)} + S_B^{(I)} + S_{A \times B}^{(I)} \quad (2.46)$$

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

	デザイン行列 X						
	切片	A	—B—		—A×B—		回帰の
	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2	平方和
因子Aのみ	\triangle	\bigcirc					$S_A^{(I)}$
因子A+因子B	\triangle	\bigcirc	\bigcirc	\bigcirc			$S_{A+B}^{(I)}$
因子A+因子B+交互作用A×B	\triangle	\bigcirc	\bigcirc	\bigcirc	\bigcirc	\bigcirc	$S_{A+B+A \times B}^{(I)}$

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 $S_A^{(I)} = 21.4286$ 、さらに因子 B を加えた $S_{A+B}^{(I)} = 130.2286$ 、フルモデル $S_{A+B+A \times B}^{(I)} = 145.4286$ 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						
		15	1	1	0	1	0	1						
A ₂	B ₃	11	1	1	0	1	0	1	2. 分散分析表 因子A+因子B					
		22	1	1	-1	-1	-1	-1	自由度	変動	分散	分散比		
		19	1	1	-1	-1	-1	-1	回帰	3	130.2286	43.4095	10.5363	
	B ₁	15	1	-1	1	0	-1	0	残差	10	41.2000	4.1200		
		14	1	-1	1	0	-1	0	合計	13	171.4286			
A ₂	B ₂	16	1	-1	0	1	0	-1	3. 分散分析表 因子A+因子B+交互作用A×B					
		18	1	-1	0	1	0	-1	自由度	変動	分散	分散比		
		21	1	-1	-1	-1	1	1	回帰	5	145.4286	29.0857	8.9495	
	B ₃	18	1	-1	-1	-1	1	1	残差	8	26.0000	3.2500		
									合計	13	171.4286			
				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	- $S_{(A)}$	21.43	$= S_{(B)}^{(1)}$	2	108.80	54.40	16.74	0.0014
A×B	$S_{(A+B+A \times B)}$	145.43	- $S_{(A+B)}$	130.23	$= S_{(A \times B)}^{(1)}$	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。同様に、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 $A \times B$ の p 値は、 $p = 0.1586$ であるが、交互作用 $A \times B$ があることを前提にした推定を行なう。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 $A \times B$ 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

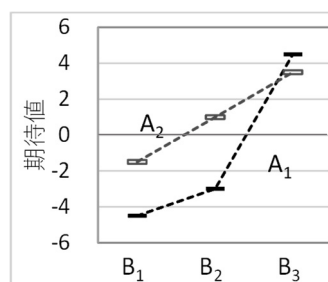


図 3.7 交互作用を考慮した効果の推定

実験は、 A_1B_1 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。 A_1B_1 の効果 $(\hat{\alpha}\beta)_{11} = -4.50$ に対し、最も大きな反応（最適水準） A_1B_3 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、 A_1B_3 と同程度の A_2B_3 についても検討したい。これは、最適でなくとも現行の A_1 水準から A_2 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、 A_2B_2 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、 B_3 水準に変更するよりも B_2 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、

$$l^{(7)} = l^{(3)} - l^{(1)} = [0 \quad 0 \quad -2 \quad -1 \quad -2 \quad -1] \quad (2.47)$$

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

	A	B	L	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
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 (2.32) 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 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$

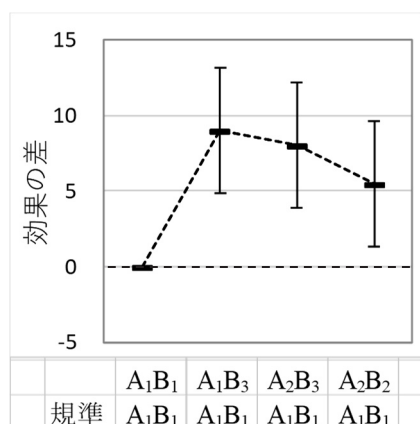
		パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\sigma}^2$						$Var(\hat{\boldsymbol{\theta}})$	SE	標準誤差
θ^{\wedge}_0	x_0	0.2483	-0.0226	0.0226	-0.0451	0.0226	-0.0451	0.2483	0.4983	0.4983
θ^{\wedge}_1	a_1	-0.0226	0.2483	0.0226	-0.0451	0.0226	-0.0451	0.2483	0.4983	0.4983
θ^{\wedge}_2	b_1	0.0226	0.0226	0.5191	-0.2257	-0.0226	0.0451	0.5191	0.7205	0.7205
θ^{\wedge}_3	b_2	-0.0451	-0.0451	-0.2257	0.4514	0.0451	-0.0903	0.4514	0.6719	0.6719
θ^{\wedge}_4	$a_1 b_1$	0.0226	0.0226	-0.0226	0.0451	0.5191	-0.2257	0.5191	0.7205	0.7205
θ^{\wedge}_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、**「繰り返しが不揃い」**、あるいは、**「繰り返しが揃い」**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 ユーザー総会では、**「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か？」**、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 トップ セミナー記録	続・高橋セミナー 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か？－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－ 2023年03月 要約 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のが必然的に「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、ダ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。

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 に一致するといった、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ことを第 3.2 節で示した。

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、2 つの因子に対し、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。

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

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

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

が、因子 A のタイプ II の平方和 $S^{(2)}_A$ となる。分かりづらい表現なので、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

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

が、因子 B に関するタイプ II の平方和 $S^{(2)}_B$ 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

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

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_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	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					
	B ₃	15	1	1	0	1	0	1	2. 分散分析表 (B)				
		11	1	1	0	1	0	1		自由度	変動	分散	分散比
A ₂	B ₁	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 ₂	15	1	-1	1	0	-1	0	合計	13	171.4286		
		14	1	-1	1	0	-1	0					
	B ₃	16	1	-1	0	1	0	-1	3. 分散分析表 (A)				
		18	1	-1	0	1	0	-1		自由度	変動	分散	分散比
		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	
				A	B				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 $S_{(A+B+A \times B)}$ の平方和の増分が、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×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主効果を 1 つずつ除いたモデル、因子 B を除いたモデルの平方和 $S_{(A+A \times B)}$ とし、フルモデルの平方和から因子 B を除いたモデルの平方和の減少分

$$S_A^{(3)} = S_{(A+B+A \times B)} - S_{(A+A \times B)}$$

が、因子 A に関するタイプ III の平方和 $S_A^{(3)}$ である。同様に、因子 A を除いたモデルの平方和 $S_{(B+A \times B)}$ であり、フルモデルの平方和から因子 B を除いたモデルの平方和の減少分

$$S_B^{(3)} = S_{(A+B+A \times B)} - S_{(B+A \times B)}$$

が、因子 B に関するタイプ III の平方和である。

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、フルモデルから主効果（A+B）を除いたモデルの平方和の減少分

$$S_{A \times B}^{(3)} = S_{(A+B+A \times B)} - S_{(A+B)}$$

が、交互作用 $A \times B$ に関するタイプ 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 $S_{(A+B+A \times B)}$ の増分とし、逐次（タイプ I）の平方和 $S_{A \times B}^{(1)}$ に一致する。

表 3.27 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、 $(A+A \times B)$ 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 $A \times B$ 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 ($A+B+A \times B$)				
		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					
	B_3	15	1	1	0	1	0	1	2. 分散分析表 ($B+A \times B$)				
		11	1	1	0	1	0	1		自由度	変動	分散	分散比
A_2	B_1	22	1	1	-1	-1	-1	-1	回帰	4	132.3377	33.0844	7.6171
		19	1	1	-1	-1	-1	-1	残差	9	39.0909	4.3434	
	B_2	16	1	-1	0	1	0	-1	合計	13	171.4286		
		18	1	-1	0	1	0	-1					
	B_3	21	1	-1	-1	-1	1	1	3. 分散分析表 ($A+A \times B$)				
		18	1	-1	-1	-1	1	1		自由度	変動	分散	分散比
			A		B		$A \times B$		回帰	3	40.2286	13.4095	1.0221
									残差	10	131.2000	13.1200	
									合計	13	171.4286		
									4. 分散分析表 ($A+B$)				
										自由度	変動	分散	分散比
									回帰	3	130.2286	43.4095	10.5363
									残差	10	41.2000	4.1200	
									合計	13	171.4286		

「3. 分散分析表 ($A+A \times B$)」の場合には、
デザイン行列 X の並びを変えている。

		x_0	a_1	a_1b_1	a_1b_2	b_1	b_2
--	--	-------	-------	----------	----------	-------	-------

これは、Excel での変数の設定に際し、
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	$S_{(A+B+A \times B)}$	145.43	-	$S_{(B+A \times B)}$	132.34	$= S^{(3)}_A$	1	13.09	13.09	4.03	0.0797
B	$S_{(A+B+A \times B)}$	145.43	-	$S_{(A+A \times B)}$	40.23	$= S^{(3)}_B$	2	105.20	52.60	16.18	0.0015
A×B	$S_{(A+B+A \times B)}$	145.43	-	$S_{(A+B)}$	130.23	$= S^{(3)}_{A \times B}$	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 3.28 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, 多少の手間により 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

A_1 と A_2 の差のベクトル $\mathbf{l}_{A_2-A_1}$ を表 3.34 に示す. 平方和の計算のために $\mathbf{L}_A = \mathbf{l}_{A_2-A_1}$ 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
	\mathbf{l}_{A1}	1	1	0	0	0	0
	\mathbf{l}_{A2}	1	-1	0	0	0	0
\mathbf{L}_A	差 \mathbf{l}_{A1-A2}	0	2	0	0	0	0

表 3.17 から, フルモデルの推定値は, $\hat{\boldsymbol{\theta}} = [16.0 \ -1.0 \ -3.0 \ -1.0 \ -0.5 \ -1.0]^T$ なので, $\mathbf{L}_A \hat{\boldsymbol{\theta}} = 2 \times (-1.0) = -2$ であり, $\mathbf{L}_A (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{L}_A^T = (-2) \times 0.07638889 \times (-2)$ と共にスカラーになるので, 式 (2.48) 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.

$$\begin{aligned}
 SS(\mathbf{L}_A \boldsymbol{\theta}) &= (\mathbf{L}_A \boldsymbol{\theta})^T [\mathbf{L}_A (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{L}_A^T]^{-1} (\mathbf{L}_A \boldsymbol{\theta}) \\
 &= [2 \times (-1.00)] \times [2 \times 0.07638889 \times 2]^{-1} [2 \times (-1.00)] \\
 &= (-2)^2 / (4 \times 0.07638889) \\
 &= 13.0909
 \end{aligned} \quad (2.48)$$

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_{B1-B2}, l_{B1-B3}) 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_{B1}	1	0	1	0	0	0
	l_{B2}	1	0	0	1	0	0
	l_{B3}	1	0	-1	-1	0	0
L_B	差 l_{B1-B3}	0	0	2	1	0	0
	差 l_{B2-B3}	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。

$$L_B \theta = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -3.00 \\ -1.00 \end{bmatrix} = \begin{bmatrix} -7.00 \\ 1.00 \end{bmatrix}$$

$$(L_B (X^T X)^{-1} L_B^T)^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0.1597 & -0.0694 \\ -0.0694 & 0.1389 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{bmatrix}^{-1}$$

$$(L_B \theta)^T (L_B (X^T X)^{-1} L_B^T)^{-1} (L_B \theta) = \begin{bmatrix} -7.00 \\ 1.00 \end{bmatrix}^T \begin{bmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{bmatrix}^{-1} \begin{bmatrix} -7.00 \\ 1.00 \end{bmatrix}$$

$$= 105.2000$$

3) 対比による交互作用 $A \times B$ に関するタイプ III の平方和

表 3.37 に交互作用 $A \times B$ 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 $A \times B$ 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

$$L_{A \times B} \theta = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -0.50 \\ -1.00 \end{bmatrix} = \begin{bmatrix} -2.00 \\ -1.50 \end{bmatrix}$$

$$(L_{A \times B} (X^T X)^{-1} L_{A \times B}^T)^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0.1597 & -0.0694 \\ -0.0694 & 0.1389 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{bmatrix}^{-1}$$

$$(L_{A \times B} \theta)^T (L_{A \times B} (X^T X)^{-1} L_{A \times B}^T)^{-1} (L_{A \times B} \theta) = \begin{bmatrix} -2.00 \\ -1.50 \end{bmatrix}^T \begin{bmatrix} 0.5000 & -0.2500 \\ -0.2500 & 0.9931 \end{bmatrix}^{-1} \begin{bmatrix} -2.00 \\ -1.50 \end{bmatrix}$$

$$= 15.2000$$

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	$S_A=(\boldsymbol{L}_A\boldsymbol{\theta}^\wedge)^T(\boldsymbol{L}_A(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{L}_A^T)^{-1}(\boldsymbol{L}_A\boldsymbol{\theta}^\wedge)$				$S_{(A)}^{(3)}$	1	13.0909		
B	2	$S_B=(\boldsymbol{L}_B\boldsymbol{\theta}^\wedge)^T(\boldsymbol{L}_B(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{L}_B^T)^{-1}(\boldsymbol{L}_B\boldsymbol{\theta}^\wedge)$				$S_{(B)}^{(3)}$	2	105.2000		
A×B	2	$S_{A\times B}=(\boldsymbol{L}_{A\times B}\boldsymbol{\theta}^\wedge)^T(\boldsymbol{L}_{A\times B}(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{L}_{A\times B}^T)^{-1}(\boldsymbol{L}_{A\times B}\boldsymbol{\theta}^\wedge)$				$S_{(A\times B)}^{(3)}$	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ことが、極めて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

第3章 文献 索引





SAS Institute Inc. (2020) - The GLM Procedure, Getting Started ,PROC GLM for Unbalanced ANOVA	102
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
高橋・大橋・芳賀(1989) - SASによる実験データの解析, 第15章 4種の平方和とLSMEANS	89

第3章 索引

あ Unbalanced ANOVA - PROC GLM	102	か 啓蒙活動 - 繰り返しが不揃い	110
因子が直交 - デザイン行列 X	101	現行水準 - 最適水準	107
因子が直交しない場合 - 高橋ら(1989)	102	現行水準と最適水準の差 - 95%信頼区間	120
主効果から因子Aを除く - タイプ II の平方和	112	効果の推定 - 線形和	108
主効果から因子Bを除く - タイプ II の平方和	112	効果の和 - 期待平均	91
Excel - 折れ線グラフ	95	- データの構造式	92
- SumSq () 関数	91	交互作用 - 可視化	93
Excel の回帰分析 - SASのGLMプロシジャ	115	- 期待平均	91
estimateステートメント - GLMプロシジャ	120	- 期待平均とセル平均の差	93
$(X^T X)^{-1} \sigma^2$ - $\Sigma(\theta^*)$	109	- 図的表現	94
- パラメータの共分散行列	109	- ダミー変数の積	96
LSMEANS - 4種の平方和	89	交互作用とはどのようなものなのか？	93
折れ線グラフ - Excel	95	考察は不十分 - 分散分析表だけ	120
- 95%信頼区間	95	合成分散の一般式 - 正しい分散の計算	99
か 回帰の平方和 - 差分	107	さ 最小2乗平均の活用 - デザイン行列 X	102
回帰分析 - 対比型デザイン行列 X	97	最適水準 - 現行水準	107
- タイプ I, タイプ II, タイプ III の平方和	115	作成手順 - 層別散布図	103
- デザイン行列 X	105	SASのGLMプロシジャ - Excel の回帰分析	115
- デザイン行列 X の一部	101	- 無償で継続的	120
重ね書き - 層別散布図	103	差分 - 回帰の平方和	107
可視化 - 交互作用	93	SumSq () 関数 - Excel	91
隔靴搔痒的な解釈 - 分散分析表の解釈	115	- 平方和の計算	91
期待値の差 - 95%信頼区間	109	GLMプロシジャ - estimateステートメント	120
期待平均 - 効果の和	91	- タイプ I の平方和	116
- 交互作用	91	- タイプ I, II, III, IV の平方和	106
- セル平均	93	- タイプ II の平方和	116
期待平均とセル平均の差 - 交互作用	93	- タイプ III の平方和	116
95%信頼区間 - 折れ線グラフ	95	- 分散分析表	116
- 期待値の差	109	- 無償で継続的	120
- グラフ表示	94	GLMプロシジャの世界だけ - 理解が上滑り	117
- 現行水準と最適水準の差	120	恣意性を排除 - 逐次平方和	111
- 水準平均	94	$\Sigma(\theta^*)$ - $(X^T X)^{-1} \sigma^2$	109
- 推定値の差	95	- パラメータの共分散行列	109
- 要因効果	98	JMP - 繰り返しが不揃い	102
95%信頼区間の幅 - $t_{0.05}(df) \times SE$	95	修正項 CT - 平方和の計算	92
グラフ表示 - 95%信頼区間	94	- 平方和の計算表	93
繰り返しが2 - 2因子実験	89	主効果・交互作用 - データの構造式	90
繰り返しが等しい - 高橋ら(1989)	89	主効果と交互作用を平等 - タイプIIIの平方和	113, 114
- 2因子実験	89	主効果を主体 - タイプ II の平方和	111, 113
- 分散分析表	93	水準平均 - 95%信頼区間	94
繰り返しが不揃い - JMP	102	推定値の差 - 95%信頼区間	95
- 啓蒙活動	110	図的表現 - 交互作用	94
- 線形モデル	102	制約条件 - データの構造式	90
- 高橋(2023)	102	- データの構造式	103
- 2因子実験	89, 102	セル平均 - 期待平均	93
- 分散分析表	111		

さ	線形モデル - 繰り返しが不揃い	102	た	デザイン行列 X - 因子が直交	101
	- (1, -1) 対比型ダミー変数	105		- 回帰分析	105
	- ダミー変数	96		- 最小2乗平均の活用	102
	線形和 - 効果の推定	108		デザイン行列 X の一部 - 回帰分析	101
	総平方和 - 平方和の分解	92	な	成り立たない - 分散の加法性	100
	層別散布図 - 重ね書き	103		2因子実験 - 繰り返しが2	89
	- 作成手順	103		- 繰り返しが等しい	89
た	(1, -1) 対比型 - ダミー変数	96		- 繰り返しが不揃い	89, 102
	(1, -1) 対比型ダミー変数 - 線形モデル	105	は	パラメータ(係数) - β_i or θ_i	97
	対比 - タイプ III の計算	117		パラメータの共分散行列 - $(X^T X)^{-1} \sigma^2$	109
	対比型デザイン行列 X - 回帰分析	97		- $\Sigma(\theta^*)$	109
	タイプ I の平方和 - GLMプロシジャ	116		不一致 - 平方和の計算結果	117
	- 逐次平方和	106		不揃い - 分散の加法性が成り立たない	104
	- 分散分析表	107		- 平方和の計算	104
	タイプ I, II, III, IV の平方和 - GLMプロシジャ	106		フルモデルから $A \times B$ を除く - タイプ II の平方和	112
	タイプ I, タイプ II, タイプ III の平方和	111		- タイプ III の平方和	113
	- 回帰分析	115		フルモデルから A を除く - タイプ III の平方和	113
	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		フルモデルから B を除く - タイプ III の平方和	113
	タイプ II の平方和 - 主効果から因子 A を除く	112		PROC GLM - Unbalanced ANOVA	102
	- 主効果から因子 B を除く	112		分散の加法性 - 互いに独立	96
	- GLMプロシジャ	116		- 成り立たない	100
	- 主効果を主体	111, 113		分散の加法性が成り立たない - 不揃い	104
	- フルモデルから $A \times B$ を除く	112		分散の推定 - 有効反復数 n_e	100
	- 分散分析表	113		分散分析表 - 繰り返しが等しい	93
	タイプ III の計算 - 対比	117		- 繰り返しが不揃い	111
	タイプ III の平方和 - GLMプロシジャ	116		- GLMプロシジャ	116
	- 主効果と交互作用を平等	113, 114		- タイプ I の平方和	107
	- 使い方	119		- タイプ II の平方和	113
	- フルモデルから $A \times B$ を除く	113		- タイプ III の平方和算出	114
	- フルモデルから A を除く	113		- 平方和の分解	89
	- フルモデルから B を除く	113		分散分析表だけ - 考察は不十分	120
	タイプ III の平方和算出 - 分散分析表	114		分散分析表の解釈 - 隔靴搔痒的な解釈	115
	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		平方和の差分 - 逐次平方和	106
	互いに独立 - 分散の加法性	96		平方和の計算 - SumSq () 関数	91
	高橋(2023) - 繰り返しが不揃い	102		- 修正項 CT	92
	- 第11回 続・高橋セミナー	110		- 不揃い	104
	高橋ら(1989) - 因子が直交しない場合	102		平方和の計算結果 - 不一致	117
	- 繰り返しが等しい	89		平方和の計算表 - 修正項 CT	93
	正しい分散の計算 - 合成分散の一般式	99		平方和の分解 - 総平方和	92
	ダミー変数 - 線形モデル	96		- 分散分析表	89
	- (1, -1) 対比型	96		β_i or θ_i - パラメータ(係数)	97
	ダミー変数の積 - 交互作用	96	ま	無償で継続的 - SASのGLMプロシジャ	120
	逐次平方和 - 恣意性を排除	111		- GLMプロシジャ	120
	- タイプ I の平方和	106		モデルに含めるダミー変数の順番 -	111
	- 平方和の差分	106	や	有効反復数 n_e - 分散の推定	100
	使い方 - タイプ III の平方和	119		要因効果 - 95%信頼区間	98
	$t_{0.05}(df) \times SE$ - 95%信頼区間の幅	95		4種の平方和 - LSMEANS	89
	データの構造式 - 効果の和	92	ら	理解が上滑り - GLMプロシジャの世界だけ	117
	- 主効果・交互作用	90			
	- 制約条件	90, 103			

第3章 解析用ファイル一覧

サイズ	名前 ^		種類
 204 KB	第03章_01_2元繰り返しが等しい		Microsoft Excel ワークシート
 125 KB	第03章_02_2元不揃い		Microsoft Excel ワークシート
 56 KB	第03章_03_各種の平方和		Microsoft Excel ワークシート
 1 KB	第03章_03_各種の平方和_SAS		テキストドキュメント

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 3 章 繰り返しが不揃いの 2 因子実験データの解析

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2024 年 1 月 20 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035