

【統計一口メモ 第43話】

＜例数不揃いの2因子分散分析＞

名古屋市立大学大学院医学研究科 非常勤講師 薬学博士 松本一彦

分散分析で1因子の場合は、例数が異なるデータでも解析できるが、2因子になると例数が等しいことが条件となります。実験計画法のテキスト、芳賀敏郎著「医薬品開発のための統計解析2」においても、例数が異なる場合に対応できる手法は載っていません。それは、平方和の分解ができないことによるものです。一方、線形モデルを活用した手法では、例数が異なったデータでも解析できるのですが、それが一般化されていないことが汎用されない理由です。そこで、今回、質的変数をダミー変数に展開して、線形モデルを適用する手法を高橋セミナー¹⁾を参考に解説します。

例題：雄ラット(A₁)と雌ラット(A₂)で3薬剤(B₁,B₂,B₃)による観測値の差の有無を確認したい。試験の結果[A₂ x B₂群]の2例が欠測値となり例数が不揃いとなった。正しい2因子分散分析を実施したい

ラット/薬剤	B1		B2				B3	
雄A1	10	13	14	12	15	11	22	19
雌A2	15	14	16	18			21	18

例数が不揃いのデータでは「平方和の分解による解析」が行えないため、観測値 y を回帰モデルで求めた予測値 \hat{y} により解析します。予測値は、デザイン行列を用いた回帰分析で求めた「係数 β 」と「対比型ダミー変数」を用いて求めます。「係数 β 」は回帰式 ($y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$) に表示される係数です。「対比型ダミー変数」は(1, 0)のダミー変数に対して(1, -1)を使うダミー変数のことです。デザイン行列については、本メモの第39話「最小二乗平均」を参照してください。

§1. エクセルによる予測値 \hat{y} の算出<a列だけのモデル演習>

手順1. 対比型ダミー変数(1-1)を用いたデザイン行列Xの作成。

雌雄	薬剤	y_{ijk}
A ₁	B ₁	10
A ₁	B ₁	13
A ₁	B ₂	14
A ₁	B ₂	12
A ₁	B ₂	15
A ₁	B ₂	11
A ₁	B ₃	22
A ₁	B ₃	19
A ₂	B ₁	15
A ₂	B ₁	14
A ₂	B ₂	16
A ₂	B ₂	18
A ₂	B ₃	21
A ₂	B ₃	18

		— デザイン行列 X —						
		y	x	a	b ₁	b ₂	a b ₁	a b ₂
A1	B1	10	1	1	1	0	1	0
		13	1	1	1	0	1	0
	B2	14	1	1	0	1	0	1
		12	1	1	0	1	0	1
		15	1	1	0	1	0	1
		11	1	1	0	1	0	1
	B3	22	1	1	-1	-1	-1	-1
		19	1	1	-1	-1	-1	-1
	A2	B1	15	1	-1	1	0	-1
14			1	-1	1	0	-1	0
B2		16	1	-1	0	1	0	-1
		18	1	-1	0	1	0	-1
B3		21	1	-1	-1	-1	1	1
		18	1	-1	-1	-1	1	1

デザイン行列 X で a は雌雄 A 群、b は薬剤群 (b_1, b_2) とする。 $ab_1 = a \times b_1$, $ab_2 = a \times b_2$
 ここでは、文章を簡単にするために、デザイン行列 X 中の a のみでの解析とする。

手順2. 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 β)を求める。

回帰モデル $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$

係数 β は共分散行列式 $(X^T X)^{-1} X^T Y$ から求める。

- ① $X^T X$ を X の転置行列と X 行列からエクセルの行列関数 MMULT と TRANSPOSE 関数で求める。
 =MMULT(TRANSPOSE(X の範囲), X の範囲) ← 着色範囲

X	
x	a
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	1
1	-1
1	-1
1	-1
1	-1
1	-1
1	-1

$X^T X$ →

14	2
2	14

- ② $X^T X$ の逆行列 $(X^T X)^{-1}$ を MINVERSE 関数で求める
 =MINVERSE($X^T X$ の範囲)

$(X^T X)^{-1}$

0.073	-0.010
-0.010	0.073

- ③ $X^T Y$ を X の転置行列と Y 行列から MMULT 関数で求める。
 =MMULT(TRANSPOSE(X の範囲), Y の範囲)

y
10
13
14
12
15
11
22
19
15
14
16
18
21
18

$X^T Y$

218
14

- ④ 係数 β を②で求めた $(X^T X)^{-1}$ の逆行列 $X^T X^{-1}$ と③の $X^T Y$ 行列から
 MMULT($(X^T X)^{-1}$ の範囲), $X^T Y$ の範囲) で回帰直線のパラメータ β_0
 (切片) と β_1 (傾き) として求める。
 a 列だけだと回帰式は $y = 15.75 - 1.25x$ となります。

β_0	15.75
β_1	-1.25

この β 係数は a 列についてのみ求めたものなので、これを b_1 ,
 b_2 , ab_1 , ab_2 列についても求めます。

§ 2. デザイン行列 X における係数 β を求める $\langle a, b_1, b_2, ab_1, ab_2 \rangle$

			D	E	F	G	H	I	J
			—— デザイン行列 X ——						
			y	x	a	b ₁	b ₂	a b ₁	a b ₂
26	A1	B1	10	1	1	1	0	1	0
27			13	1	1	1	0	1	0
28			14	1	1	0	1	0	1
29		B2	12	1	1	0	1	0	1
30			15	1	1	0	1	0	1
31			11	1	1	0	1	0	1
32		B3	22	1	1	-1	-1	-1	-1
33			19	1	1	-1	-1	-1	-1
34			15	1	-1	1	0	-1	0
35	A2	B1	14	1	-1	1	0	-1	0
36			16	1	-1	0	1	0	-1
37		B2	18	1	-1	0	1	0	-1
38			21	1	-1	-1	-1	1	1
39		B3	18	1	-1	-1	-1	1	1

デザイン行列から係数 β を次のエクセル関数で求める。

=MMULT(MINVERSE(MMULT(TRANSPOSE(E26:J39),E26:J39)),MMULT(TRANSPOSE
(E26:J39),D26:D39))

			AJ
47	β_0	切片 x_0	16.0
48	β_1	a_1	-1.0
49	β_2	b_1	-3.0
50	β_3	b_2	-1.0
51	β_4	$a_1 b_1$	-0.5
52	β_5	$a_1 b_2$	-1.0

§ 3 予測値 \hat{y} を求める

予測値 \hat{y} はデザイン行列 X (x, a, b₁, b₂, ab₁, ab₂ の全範囲) と係数 β (全範囲) を用いる。

=MMULT(デザイン行列 X 範囲, β 係数範囲) =MMULT(E26:J39,AJ:47:AJ52)

y	\hat{y}
10	11.5
13	11.5
14	13.0
12	13.0
15	13.0
11	13.0
22	20.5
19	20.5
15	14.5
14	14.5
16	17.0
18	17.0
21	19.5
18	19.5

§ 4. 逐次平方和は何のために求めるのだろうか？

逐次平方和とは「回帰の平方和を $S_A, S_{A+B}, S_{A+B+AxB}$ で求め、それらの差分をそれぞれの要因の平方和とする」と定義され、 $S_B = S_{A+B} - S_A$ $S_{AxB} = S_{A+B+AxB} - S_{A+B}$ で求める。

このようにして求めた平方和を全て加えて S_T が求められる。

$$S_T = S_A + S_B + S_{AxB}$$

ここでは、なぜ逐次平方和などを使わなくてはならないのかを紐解いていこう。

答えは最後の手順15の後に書かれている。

(1) 従来の平方和は、次の手順で求める。

手順1. 観測値 y の総平均 15.57 y^{bar} を求め y の横に書き込む。

雌雄	薬剤	y	y^{bar}
A1	B1	10	15.57
		13	15.57
	B2	14	15.57
		12	15.57
		15	15.57
	B3	11	15.57
22		15.57	
A2	B1	19	15.57
		15	15.57
	B2	14	15.57
		16	15.57
	B3	18	15.57
		21	15.57
		18	15.57

手順2. 行ごとに「観測値 - 総平均」($y_i - y^{\text{bar}}$) を求める。

	Z
	$y_i - y^{\text{bar}}$
7	-5.57
8	-2.57
9	-1.57
10	-3.57
11	-0.57
12	-4.57
13	6.43
14	3.43
15	-0.57
16	-1.57
17	0.43
18	2.43
19	5.43
20	2.43

総平方和 $S_T = 171.43$

手順3. 総平方和 $S_T = 171.43$ を得る。

$$\sum (y_i - y^{\text{bar}})^2 = \text{SUMSQ}(Z7:Z20) = 171.43$$

手順4. 因子 A_1, A_2 それぞれの平均値 y_A^{bar} を求める。

$$y_{A1}^{\text{bar}} = 14.50 \quad y_{A2}^{\text{bar}} = 17.00$$

手順5. 因子 A_1, A_2 ごとに 主効果 ($y_A^{\text{bar}} - \hat{y}$) を求める。 A_1 は「-1.07」 A_2 は「1.43」を得る。

手順6. 主効果 A の平方和 $S_A = 21.43$ 得る。

$$21.43 = \text{SUMSQ}(y_A^{\text{bar}} - \hat{y} \text{ の範囲})$$

y_A^{bar}	$y_A^{\text{bar}} - \hat{y}$
14.50	-1.07
14.50	-1.07
14.50	-1.07
14.50	-1.07
14.50	-1.07
14.50	-1.07
14.50	-1.07
14.50	-1.07
14.50	-1.07
17.00	1.43
17.00	1.43
17.00	1.43
17.00	1.43
17.00	1.43
17.00	1.43

平方和 $S_A = 21.43$

手順7. 因子 B ごとに平均値 y_B^{bar} を求める

$$B_1 = 13.00 \quad B_2 = 14.33 \quad B_3 = 20.00$$

手順8. 因子 B ごとに主効果 ($y_B^{\text{bar}} - y^{\text{bar}}$) を求める。

$$B_1 = 13.00 - 15.57 = -2.57 \quad B_2 = -1.24 \quad B_3 = 4.43$$

手順7-9

手順9. 主効果 B の平方和 $S_B=114.10$ 得る。

$$114.10 = \text{SUMSQ}(y_B^{\text{bar}} - \bar{y}^{\text{bar}} \text{ の範囲})$$

手順10. 因子 A と因子 B の組み合わせ平均 y_{AB}^{bar} を求める。

$$G \text{ 列 } 1 \& 2 \text{ 行目 } y_{AB}^{\text{bar}} = \text{AVERAGE}(C1:C2) = (10+13)/2 = 11.50$$

手順11. 予測平均 (y_{AB}^{\wedge}) を求める。

$$H \text{ 列 } 1 \& 2 \text{ 行目 } y_{AB}^{\wedge} = E1 + F1 - D1 = 11.93$$

手順12. 交互作用の主効果 ($y_{AB}^{\text{bar}} - y_{AB}^{\wedge}$) を求め、その平方和

$$S_{A \times B} = 16.10 \text{ を得る。}$$

$$I \text{ 列 } 1 \& 2 \text{ 行目 } y_{AB}^{\text{bar}} - y_{AB}^{\wedge} = G1 - H1 = -0.43$$

$$\text{交互作用平方和} = \text{SUMSQ}(I1:I14) = 16.10$$

	y_B^{bar}	$y_B^{\text{bar}} - \bar{y}^{\text{bar}}$
B1	13.00	2.57
	13.00	2.57
B2	14.33	1.24
	14.33	1.24
	14.33	1.24
	14.33	1.24
B3	20.00	-4.43
	20.00	-4.43
B1	13.00	2.57
	13.00	2.57
B2	14.33	1.24
	14.33	1.24
B3	20.00	-4.43
	20.00	-4.43

平方和 $S_B = 114.10$

手順 10-12

	A	B	C	D	E	F	G	H	I	
	雌雄	薬剤	y_j	y^{bar}	y_A^{bar}	y_B^{bar}	y_{AB}^{bar}	y_{AB}^{\wedge}	$y_{AB}^{\text{bar}} - y_{AB}^{\wedge}$	
1	A ₁	B1	10	15.57	14.50	13.00	11.50	11.93	-0.43	
2			13	15.57	14.50	13.00	11.50	11.93	-0.43	
3		B2	14	15.57	14.50	14.33	13.00	13.26	-0.26	
4			12	15.57	14.50	14.33	13.00	13.26	-0.26	
5			15	15.57	14.50	14.33	13.00	13.26	-0.26	
6			11	15.57	14.50	14.33	13.00	13.26	-0.26	
7	A ₂	B3	22	15.57	14.50	20.00	20.50	18.93	1.57	
8			19	15.57	14.50	20.00	20.50	18.93	1.57	
9		B1	15	15.57	17.00	13.00	14.50	14.43	0.07	
10			14	15.57	17.00	13.00	14.50	14.43	0.07	
11		B2	16	15.57	17.00	14.33	17.00	15.76	1.24	
12			18	15.57	17.00	14.33	17.00	15.76	1.24	
13			B3	21	15.57	17.00	20.00	19.50	21.43	-1.93
14				18	15.57	17.00	20.00	19.50	21.43	-1.93

交互作用 = 16.10

手順 13

手順13. 観測値 y_i 組み合わせ平均 y_{AB}^{bar} の差 =

残差 ($y_i - y_{AB}^{\text{bar}}$) を求める。

$$K \text{ 列 } 1 \text{ 行目} = C - H = -1.50$$

手順14. 残差平方和 $Se=26.00$ を求める。

$$\text{SUMSQ}(K1:K14) = 26.00$$

手順15. 総平方和 S'_T を求める

$$S'_T = S_A + S_B + S_{A \times B} + Se =$$

$$21.43 + 114.10 + 16.10 + 26.00 = 177.62$$

得られた平方和の合計177.62は手順3で得られた総平方和171.43と一致しない。そこで、登場するのが逐次平方和である。

	A	B	C	H	K	
	雌雄	薬剤	y_j	y_{AB}^{bar}	$y_i - y_{AB}^{\text{bar}}$	
1	A ₁	B1	10	11.50	-1.50	
2			13	11.50	1.50	
3		B2	14	13.00	1.00	
4			12	13.00	-1.00	
5			15	13.00	2.00	
6			11	13.00	-2.00	
7	A ₂	B3	22	20.50	1.50	
8			19	20.50	-1.50	
9		B1	15	14.50	0.50	
10			14	14.50	-0.50	
11		B2	16	17.00	-1.00	
12			18	17.00	1.00	
13			B3	21	19.50	1.50
14				18	19.50	-1.50

残差平方和 = 26.00

要因A、B、A x Bを求めるために回帰平方和の差分を算出する					
要因	回帰平方和		差分平方和		因子平方和
A=	S(A)	14.86	S(A)	14.86	14.86
B=	S(A+B)	123.14	S(B)=S(A+B)- S(A)	123.14-14.86	108.29
AxB=	S(A+B+AxB)	145.43	S(AxB) = S(A+B+AxB)-S(A+B)	145.43-123.14	22.29

残差の回帰平方和は通常平方和で求めた値と同値の 26.00 を得る。

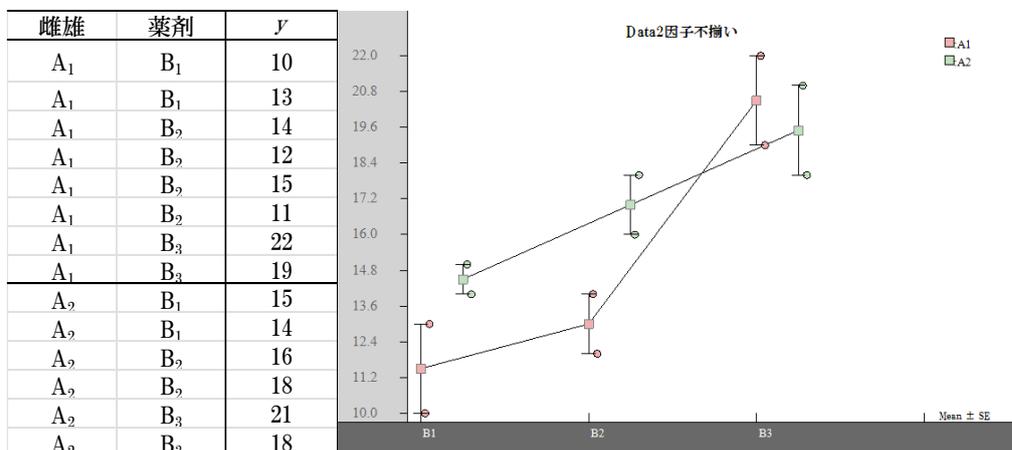
逐次平方和による分散分析表

要因	平方和	自由度	平均平方	F比	p値	
雌雄	14.86	1	14.86	4.57	0.065	*
薬剤	108.29	2	54.14	16.66	0.001	**
交互作用	22.29	2	11.14	3.43	0.084	
誤差	26.00	8	3.25			
全体	171.43	13				

§ 5. Pharmaco ソフト (Pharmaco ANOVA Ver.2)による解析²⁾



Pharmaco 入力形式



要因	平方和	自由度	平均平方	F比	p値	
雌雄	14.86	1	14.86	4.57	0.065	
薬剤	108.29	2	54.14	16.66	0.001	**
交互作用	22.29	2	11.14	3.43	0.084	
誤差	26.00	8	3.25			
全体	171.429	13				

§ 6. JMP ソフトによる解析

JMP Ver.10 による入力形式は § 5 の Pharmaco 入力形式と同じで、「分析」では「モデルのあてはめ」を用いて次のような結果が得られる。

応答 y						
モデル全体						
効果の検定						
要因	パラメータ数	自由度	平方和	F値	p値(Prob>F)	
雌雄	1	1	13.09091	4.0280	0.0797	
薬剤	2	2	105.20000	16.1846	0.0015*	
雌雄*薬剤	2	2	15.20000	2.3385	0.1586	

JMP も解析にデザイン行列を使用しており、結果は、エクセル、Pharmaco と同じ傾向を示している。本報告は高橋セミナー主幹の高橋行雄氏に校閲していただいた。心から御礼申し上げます。

- 1) 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 繰り返し不揃いの 2 因子実験データの解析
高橋行雄 2024.1.20
Biostat 研究所(株) takahasi stat@nifty.com
- 2) Pharmaco 工房:<https://pharmaco.jp>