第 11 回 続・高橋セミナー

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 ―

2023 年 3 月

高橋 行雄　BioStat 研究所（株）

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と必然的に「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い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とすると，「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．

目 次

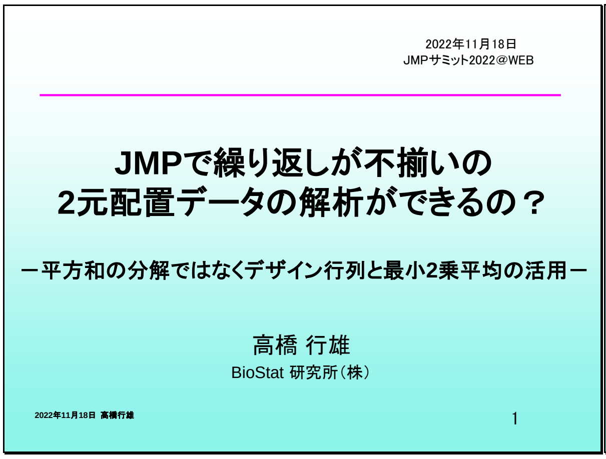
次ページに続く

連絡先:高橋 行雄，〒105-0014 東京都 港区 芝 1-12－3 の 1005
takahashi.stat@nifty.com， FAX：03-342-8035

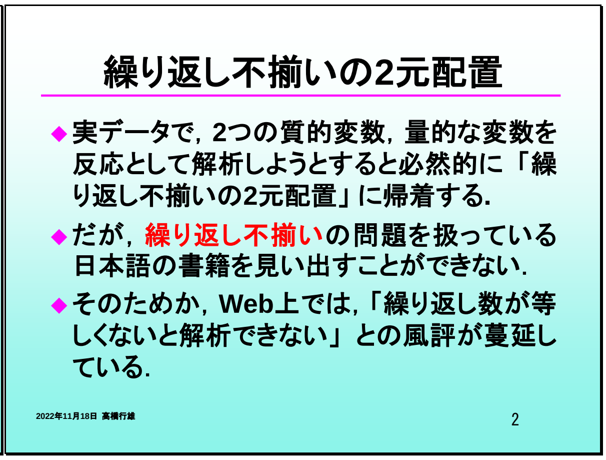
1. はじめに

2022 年 11 月 18 日にオンラインで開催された JMP の Discovery Summit Japan 2022 にて，「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？ ー平方和の分解ではなくデザイン行列と最小 2 乗平均の活用ー」と題して講演した．この講演は，2020 年から取り組んでいる「層別因子を含む探索的な回帰分析入門」の第 3 章の原稿を推敲するために行った．JMP の Discovery Summit Japan 2022 での講演は，予想以上の反響があったので，この講演で用いたスライドを主体にし，続・高橋セミナー第 11 回目とする．

スライド 1

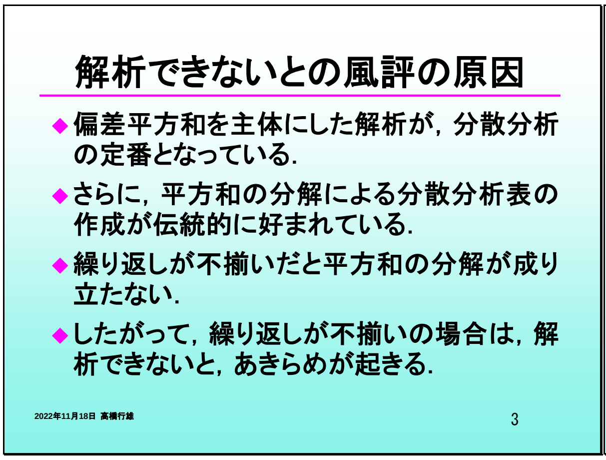


スライド 2

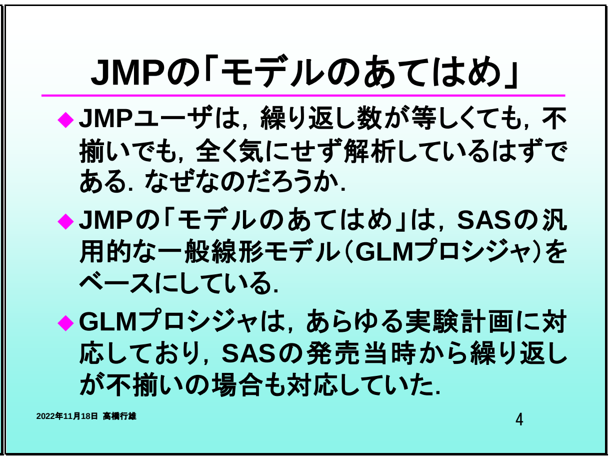


実データで，2 つの質的変数，量的な変数を反応として解析しようとすると必然的に「繰り返しが不揃いの 2 元配置」に帰着する．だが，繰り返しが不揃いの問題を扱っている日本語の書籍を見出すことができない．そのためか，Web 上では，「繰り返し数が等しくないと解析できない」との風評が蔓延している．

スライド 3



スライド 4



解析できないとの風評の原因は，偏差平方和を主体にした解析が，分散分析の定番となっているためである．さらに，平方和の分解による分散分析表の作成が伝統的に好まれており，繰り返しが不揃いだと平方和の分解が成り立たないことも知られている．したがって，「繰り返しが不揃いの場合は，解析できない」とあきらめが起きてしまう．ふがいない事態としか言いようがない．

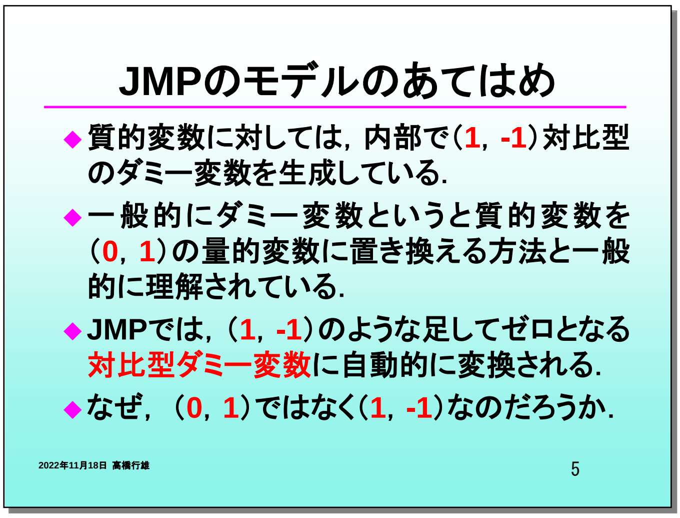
JMP ユーザは，繰り返しが等しくても，不揃いでも，全く気にせず解析しているはずである．なぜなのだろうか．JMP の「モデルのあてはめ」は，無償で継続的に使用できる OnDemand SAS の汎用的な一般線形モデル（GLM プロシジャ）をベースにしている．GLM プロシジャは，あらゆる実験計画モデルに対応しており，SAS の発売当時から繰り返しが不揃いの場合も対応していた．

自由にダウンロードできる最新版の GLM プロシジャの解説書，SAS Institute（2020），「SAS/STAT 15.2 User's Guide, The GLM Procedure」の使い方の事例に「Example 52.3: Unbalanced ANOVA for Two-Way Design with Interaction」が含まれており，SAS が普及していれば，「繰り返し数が等しくないと解析できない」との風評は立たなかったと思われる．このような嘆かわしい状況を改善するために，何らかの啓蒙活動が必要との認識を持つに至った．なお，SAS の英語のマニュアルは，Web 上で自動翻訳された結果が表示される．

2. 繰り返しが不揃いの 2 元配置

各種のダミー変数

スライド 5



繰り返しが不揃いの 2 元配置データの解析には，分散分析表の作成に偏差平方和に基づく平方和の分解による計算手順が使えない．そのために，質的因子をダミー変数に変換し量的変数とし，線形モデルを適用する必要がある．

JMP の「モデルのあてはめ」では，質的変数に対して，内部で（1，–1）対比型ダミー変数を生成している．一般的にダミー変数というと質的変数を（0，1）の量的変数に置き換える方法と一般的に理解されている．だが，JMP では，（1，–1）のように足してゼロとなる対比型ダミー変数に自動的に変換される．なぜ，（0，1）ではなく（1，–1）なのだろうか．

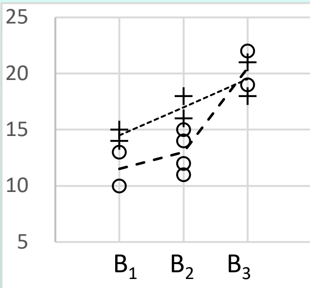
繰り返しが不揃いの事例

スライド 6 に示すのは，高橋行雄，大橋靖雄，芳賀敏郎（1989），「SAS による実験データの解析，307-333」で取り上げられている「繰り返しが不揃いの 2 元配置データ」の事例である．今から 30 年以上前に解析事例として用いていたのであるが，再度，この課題を取り上げて深堀する．

スライド 6

繰り返しが不揃いの事例

	B_1		B_2				B_3		平均 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 y	13.0		14.33				20.0		15.57



高橋行雄，大橋靖雄，芳賀敏郎（1989），
SAS による実験データの解析，307-333，
東京大学出版会．

30年以上前に解析事例として
取り上げていた．

再度，この課題を取り上げ深堀する．

2022年11月18日 高橋行雄

6

このデータは，因子 A が 2 水準，因子 B が 3 水準，繰り返しが 2 のデータに対し，元の A_1B_2 のセルの平均 $(14+12)/2=13$ が変化しないように（15，11）を人工的に加えたものである．このデータは，SAS の GLM プロシジャが提供している 4 種の平方和の特徴を浮き彫りにするために用意した 3 種のデータの一つである．

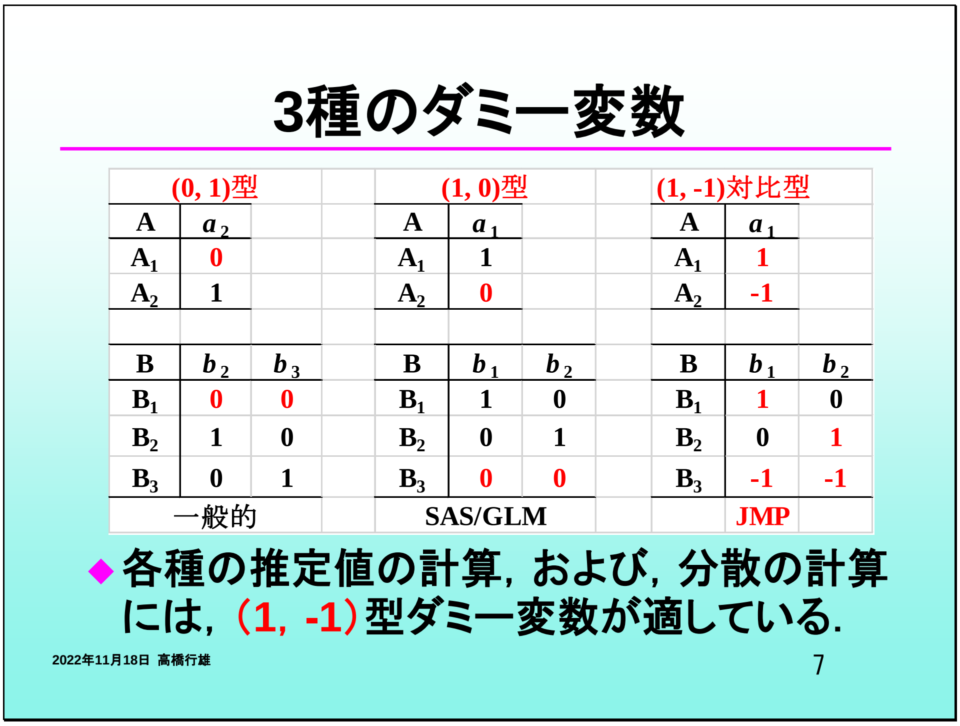
3 種のダミー変数

スライド 7 に示すようにダミー変数は，質的因子を線形モデルで扱うために量的変数に置き換える便宜的な方法である．JMP の「モデルのあてはめ」では，（1，−1）対比型ダミー変数が使われ，SAS の GLM プロシジャでは，（1，0）型ダミー変数がディフォルト使われ，R の lm () 関数では，一般的な（0，1）型ダミー変数が使われている．このように，統計ソフトの開発者の考え方により，各種のダミー変数が使われている．

どのようなダミー変数を使うかは自由であり，質的因子の水準が区別できればいいのである．性別の（男，女）を（1，0）でも，（1，2）でも，（0，1）でも，（1，−1）でも，どのよ

うに与えるかは，解析者の自己責任で選ぶべき課題である．ただし，統計ソフトを使う場合には，質的因子に対してどのようなダミー変数に変換されているかに注意を払う必要がある．JMP は，なぜ（1，-1）型ダミー変数を用いているのだろうか．

スライド 7



(0, 1)型			(1, 0)型			(1, -1)対比型		
A	a_2		A	a_1		A	a_1	
A_1	0		A_1	1		A_1	1	
A_2	1		A_2	0		A_2	-1	
B	b_2	b_3	B	b_1	b_2	B	b_1	b_2
B_1	0	0	B_1	1	0	B_1	1	0
B_2	1	0	B_2	0	1	B_2	0	1
B_3	0	1	B_3	0	0	B_3	-1	-1
一般的			SAS/GLM			JMP		

3. 偏差平方和の計算

修正項 *CT* を用いる方法

繰り返し数が揃っている場合に，スライド 8 に示すような偏差平方和の計算公式が，多くの書籍で示されている．Excel のような表計算ソフトが普及する以前に，できる限り演算回数を減らす工夫がされた優れものである．ただし，繰り返しが等しい場合にのみに適用でき，繰り返しが不揃いの場合は使えない．計算式の中で修正項 *CT* は，Correction Term の略であり，得られたデータの総合計 $T_{..}$ の平方をデータの総数 N で割って得られる．スライド 6 に示した繰り返しが不揃いのデータから，A_1B_2 セルの（15，11）を除き，繰り返しが 2 の 2 元配置データとし，スライド 8 に示す伝統的な計算式によって平方和の計算例を示す．

スライド 8

CT を用いた偏差平方和の計算

◆繰返し数が等しい場合の計算方式

●不揃いの場合は使えない

$CT = T_{\cdots}^2 / N$，$T_{\cdots}$：データの総合計

$$S_\mathrm{T} = \left(\sum_i \sum_j \sum_k y_{ijk}^2\right) - CT$$

$$S_\mathrm{A} = \sum_i \frac{\left(\sum_j \sum_k y_{ijk}\right)^2}{n_\mathrm{B} n_\mathrm{K}} - CT$$

$$S_\mathrm{B} = \sum_j \frac{\left(\sum_i \sum_k y_{ijk}\right)^2}{n_\mathrm{A} n_\mathrm{K}} - CT$$

$$S_\mathrm{AB} = \sum_i \sum_j \frac{\left(\sum_k y_{ijk}\right)^2}{n_k} - CT$$

$$S_e = S_\mathrm{T} - S_\mathrm{AB}$$

$$S_\mathrm{AB} = S_\mathrm{A} + S_\mathrm{B} + S_{\mathrm{A}\times\mathrm{B}}$$

$$S_{\mathrm{A}\times\mathrm{B}} = S_\mathrm{AB} - S_\mathrm{A} - S_\mathrm{B}$$

2022年11月18日 高橋行雄　　8

表 1 に繰り返し 2 の 2 元配置のデータ，表 2 に AB 表として表 1 の周辺和を計算した結果を示す．スライド 8 に示した計算式に忠実に Excel の SumSq () 関数を用い，手計算的な計算手順を示す．

表 1 繰り返し 2 の 2 元配置

	B_1		B_2		B_3		平均
A_1	10	13	14	12	22	19	15.0
A_2	15	14	16	18	21	18	17.0
平均	13.0		15.0		20.0		16.0
N=	12	n_A=	2	n_B=	3	n_K=	2

表 2 AB 表

	B_1	B_2	B_3	計
A_1	23	26	41	90
A_2	29	34	39	102
計	52	60	80	192

$$CT = \frac{T_{\cdots}^2}{N} = \frac{192^2}{12} = 3072.00$$

$$S_\mathrm{T} = \left(\sum_i \sum_j \sum_k y_{ijk}^2\right) - CT = (10^2 + 13^2 + \cdots + 18^2) - 3072 = 148.00$$

$$S_\mathrm{A} = \sum_i \frac{\left(\sum_j \sum_k y_{ijk}\right)^2}{n_\mathrm{B} n_\mathrm{K}} - CT = \frac{90^2 + 102^2}{3 \times 2} - 3072 = 12.00$$

$$S_\mathrm{B} = \sum_j \frac{\left(\sum_i \sum_k y_{ijk}\right)^2}{n_\mathrm{A} n_\mathrm{K}} - CT = \frac{52^2 + 60^2 + 80^2}{2 \times 2} - 3072 = 104.00$$

$$S_\mathrm{AB} = \sum_i \sum_j \frac{\left(\sum_k y_{ijk}\right)^2}{n_k} - CT = \frac{23^2 + 26^2 + \cdots + 39^2}{2} - 3072 = 130.00$$

$$S_{\mathrm{T}} = S_{\mathrm{AB}} + S_e$$

$$S_e = S_{\mathrm{T}} - S_{\mathrm{AB}} = 148.00 - 130.00 = 18.00$$

$$S_{\mathrm{AB}} = S_{\mathrm{A}} + S_{\mathrm{B}} + S_{\mathrm{A\times B}}$$

$$S_{\mathrm{A\times B}} = S_{\mathrm{AB}} - S_{\mathrm{A}} - S_{\mathrm{B}} = 130.00 - 12.00 - 104.00 = 14.00$$

これらの計算公式の簡便性は，残差平方和 S_e および交互作用 $S_{\mathrm{A\times B}}$ を算出に AB 表を活用していることにある．繰り返しが揃っているために総平方和 S_{T} が S_e と S_{AB} の和に分解できること，S_{AB} が S_{A} と S_{B} と $S_{\mathrm{A\times B}}$ の和に分解できることを活用している．ただし，このような平方和の分解を活用して求めた交互作用 $S_{\mathrm{A\times B}}$ に対する理解が，計算公式に依存した理解となり，交互作用の本質的な理解の妨げになりがちである．

繰り返しが不揃いの場合に対する偏差平方和の計算

修正項 CT を用いる偏差平方和の計算は，計算公式で，n_{A}，n_{B}，n_{K} などの繰り返し数が使われており，繰り返しが不揃い場合は，これらを使うことができない．ただし，スライド 9 に示すように偏差平方和の定義式に戻れば，計算することができる．Excel のシート上にデータを行方向に並べ，総平方和 S_{T} は，y_{ijk} と総平均 $\bar{y}_{..} = 15.57$ の偏差平方和，平方和 S_{A} は，

スライド 9

偏差平方和の計算はできるが

			μ^			α^		β^			$(\alpha\beta)$^	ε^
		y_{ijk}	$\bar{y}_{..}$	$y-\bar{y}_{..}$	$\bar{y}_{\mathrm{A}}$	$\bar{y}_{\mathrm{A}}-\bar{y}_{..}$	$\bar{y}_{\mathrm{B}}$	$\bar{y}_{\mathrm{B}}-\bar{y}_{..}$	$\bar{y}_{\mathrm{AB}}$	$\hat{y}_{\mathrm{AB}}$	$\bar{y}_{\mathrm{AB}}-\hat{y}_{\mathrm{AB}}$	$y-\bar{y}_{\mathrm{AB}}$
$\mathrm{A_1}$	$\mathrm{B_1}$	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
	$\mathrm{B_2}$	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
	$\mathrm{B_3}$	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
$\mathrm{A_2}$	$\mathrm{B_1}$	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
	$\mathrm{B_2}$	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
	$\mathrm{B_3}$	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
		平方和		**171.43**		**21.43**		**114.10**			**16.10**	**26.00**
				S_{T}		S_{A}		S_{B}			$S_{\mathrm{A\times B}}$	S_e
							$S_{\mathrm{T}}' = S_{\mathrm{A}}+S_{\mathrm{B}}+S_{\mathrm{A\times B}}+S_{\mathrm{e}} =$			**177.62**		

＋－が逆

2022年11月18日 高橋行雄

9

それぞれの水準平均 $\bar{y}_{i\cdot\cdot}$ と総平均 $\bar{y}_{\cdots}$ の偏差平方和，平方和 S_{B} は，それぞれの水準平均 $\bar{y}_{\cdot j\cdot}$ と総平均 $\bar{y}_{\cdots}$ との偏差平方和で求めることができる．

交互作用 $S_{\mathrm{A\times B}}$ は，セル平均 $\bar{y}_{ij\cdot}$ と期待平均 $\hat{y}_{ij\cdot}$ との偏差平方和で求めることができる．ここで，期待平均 $\hat{y}_{ij\cdot}$ は，総平均 $\bar{y}_{\cdots}$ に因子 A の各水準の効果（$\alpha_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$）と因子 B の各水準の効果（$\beta_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$）を加え

$$
\begin{aligned}
\hat{y}_{ij\cdot} &= \bar{y}_{\cdots} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\
&= \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}
\end{aligned}
$$

として計算されている．残差平方和 S_e は，y_{ijk} とセル平均 $\bar{y}_{ij\cdot}$ との偏差平方和として計算している．

偏差平方和の定義式よる計算

スライド 10 に示すように，修正項 CT を使った式ではなく，組み合せ平均または期待平均からの偏差平方和による計算式と計算結果は，

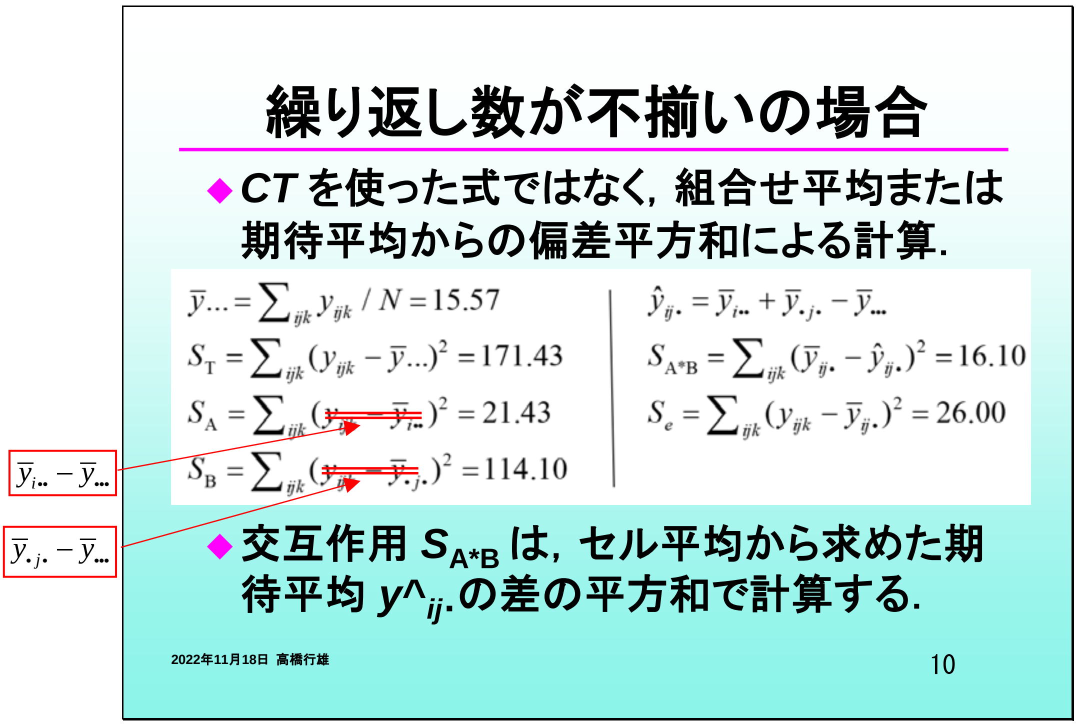
$$\bar{y}_{\cdots} = \frac{\sum_{ijk} y_{ijk}}{N} = \frac{10+13+\cdots+18}{14} = \frac{218}{14} = 15.57$$

$$S_{\mathrm{T}} = \sum_{ijk}(y_{ijk} - \bar{y}_{\cdots})^2 = (10-15.57)^2 + (13-15.57)^2 + \cdots + (18-15.57)^2 = 171.43$$

$$S_{\mathrm{A}} = \sum_{ijk}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2 = (14.50-15.57)^2 + (14.50-15.57)^2 + \cdots + (17.00-15.57)^2 = 21.43$$

$$S_{\mathrm{B}} = \sum_{ijk}(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2 = (13.00-15.57)^2 + (13.00-15.57)^2 + \cdots + (20.00-15.57)^2 = 114.10$$

スライド 10



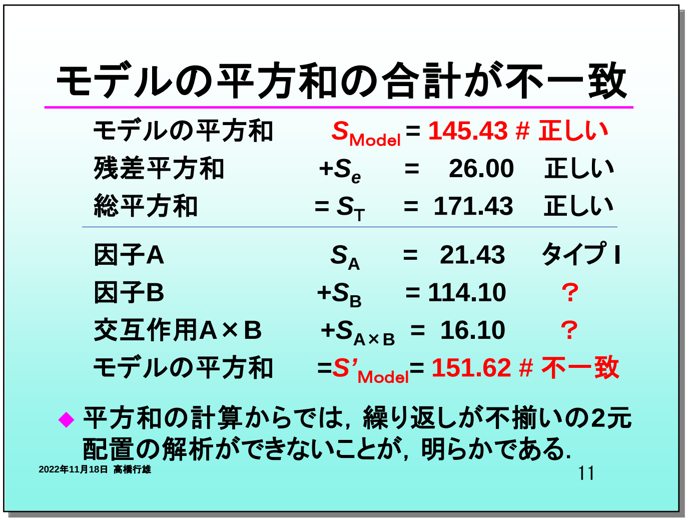
$$\hat{y}_{ij\cdot} = \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}$$

$$S_{\mathrm{A*B}} = \sum_{ijk} (\bar{y}_{ij\cdot} - \hat{y}_{ij\cdot})^2 = (11.50 - 11.93)^2 + (11.50 - 11.93)^2 + \cdots + (19.50 - 21.43)^2 = 16.10$$

$$S_e = \sum_{ijk} (y_{ijk} - \bar{y}_{ij\cdot})^2 = (10 - 11.50)^2 + (13 - 11.50)^2 + \cdots + (18 - 19.50)^2 = 26.00$$

として求められる．

スライド 11



平方和の分解が成り立たない

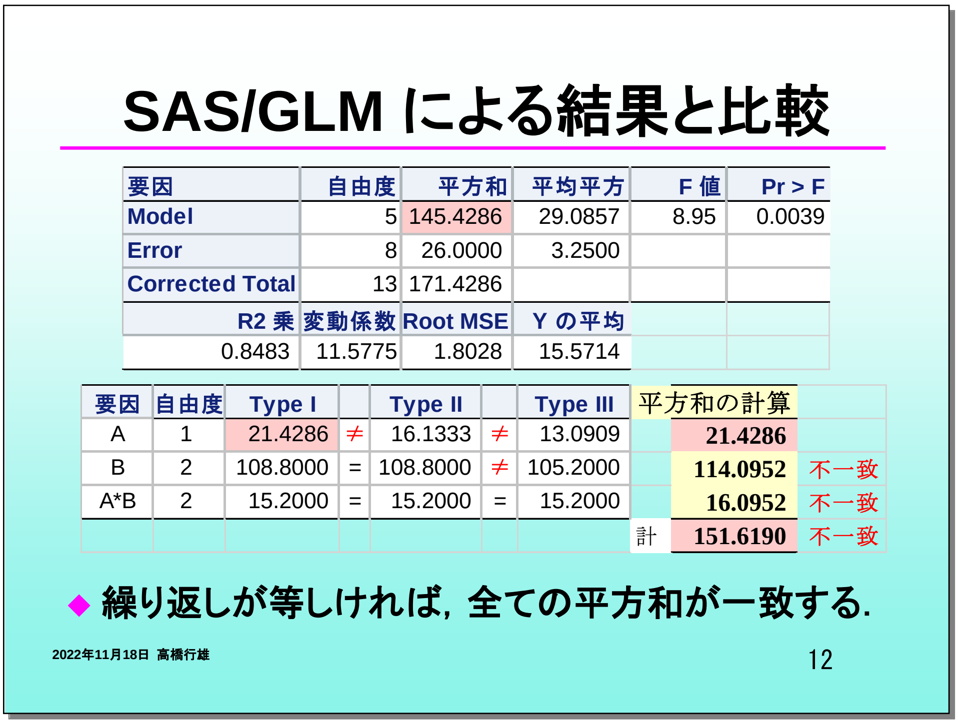
偏差平方和の計算は，定義式による計算をすることができたのであるが，スライド 11 に示すように総平方和 S_{T} から S_e を差し引いた S_{Model} が

$$S_{\mathrm{Model}} = S_{\mathrm{T}} - S_e$$
$$= 171.43 - 26.00 = 145.43$$

として計算される．ところが，S_{A} と S_{B} と $S_{\mathrm{A\times B}}$ の合計が，151.62 となり一致しない．そのために，平方和の分解が成り立たないので，分散分析表にまとめることができない．

SAS の GLM プロシジャによる結果と比較

スライド 12



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			
	R2 乗	変動係数	Root MSE	Y の平均	
	0.8483	11.5775	1.8028	15.5714	

要因	自由度	Type I		Type II		Type III	平方和の計算		
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	151.6190	不一致

無償で継続的に使える OnDemand SAS の GLM プロシジャによる解析は，一般線形モデルによる解析が行われている．繰り返しが不揃いの 2 元配置のデータの解析を行うとスライド 12 に示す結果が得られる．モデルの平方和 S_{Model}，残差平方

和 S_e，および，総平方和 S_T は，スライド 11 上 に示した Excel による計算結果に一致する．しかし， S_A を除いて， S_B と $S_{A\times B}$ の平方和に一致する平方和が見いだせない．偏差平方和の定義式よる計算は，先進的な統計ソフトの結果と残念ながら一致しない．

スライド 12 補足資料　SAS の GLM プロシジャによる分散分析表と 4 種の平方和

```
Title1 'TwoWay_unB_ANOVA.sas 2022-5-
22 Y.Takahashi' ;

data  D1 ;
    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 print data=D1 ;  run ;
proc glm data=D1 ;
    class A B ;
    model  y =  A B A*B
        / solution  ss1 ss2 ss3 ss4 ;
    lsmeans     A B A*B / cl ;
    run ;
```

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			

R2 乗	変動係数	Root MSE	Y の平均
0.8483	11.5775	1.8028	15.5714

要因	自由度	Type I	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

要因	自由度	Type II	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

要因	自由度	Type III	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

要因	自由度	Type IV	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

JMP の「モデルのあてはめ」による解析結果

スライド 13

JMPの「モデルのあてはめ」

JMP 分散分析

要因	自由度	平方和	平均平方	F値
モデル	5	145.4286	29.0857	8.9495
誤差	8	26.0000	3.2500	p値(Prob>F)
全体(修正済み)	13	171.4286		0.0039

JMP		タイプ3 効果の検定		タイプ1		Excel
要因	自由度	平方和		逐次平方和		平方和の計算
A	1	13.0909	≠	21.4286	=	21.4286
B	2	105.2000	≠	108.8000	≠	114.0952
A*B	2	15.2000	=	15.2000	≠	16.0952
モデル	5	133.4909	≠	145.4286	≠	151.6190

◆ JMPでは，SAS/GLMのタイプ 3 およびタイプ 1 の平方和が出力される．平方和の計算は，不一致．

2022年11月18日 高橋行雄　13

JMP の「モデルのあてはめ」は，SAS の GLM プロシジャと同様に，一般線形モデルによる解析法を用いているので，繰り返しが不揃いの2元配置データの解析が行える．スライド 13 に示すように，JMP の「モデルのあてはめ」の出力結果は，スライド 12 に示した GLM プロシジャの出力

と同様の形式となっている．

ディフォルトで出力される「効果の検定の平方和」は，GLM プロシジャの Type III の平方和に一致する．オプションで出力される「逐次平方和」は，Type I の平方和に一致するが，Excel で計算した偏差平方和に一致するものは見当たらない．

Type III の平方和の合計は， $S^{(3)}_{\mathrm{Model}}=133.4909$ となり分散分析表のモデルの平方和 145.4268 に一致しない．Type I の平方和の合計は， $S^{(1)}_{\mathrm{Model}}=145.4268$ と分散分析表のモデルの平方和に一致する．Excel で計算した偏差平方和を合計すると $S^{(\mathrm{Excel})}_{\mathrm{Model}}=151.6190$ となり，これまでに知られている平方和と異なる．

GLM プロシジャ，あるいは，JMP の「モデルのあてはめ」による平方和は，どのようにして求められているのであろうか．Excel の力を借りて JMP「モデルのあてはめ」の計算方法を再現することにより，謎解きをする．

欠測値に対する伝統的な対応

朝香鐵一，石川馨，山口襄 共同監修（1988），「新版 品質管理便覧 第 2 版」に「欠測値のある場合」について述べられている．「7 章 実験計画，7.1.3.3 二因子要因計画，（1）反復のない場合，欠測値のある場合：p735-5」と「第 12 章 数理，12.9 分散分析の数理，12.9.8 欠測値の有る場合の処理：p914-5」が該当する．どちらも，2 元配置の場合について述べられている．共通するのは，繰り返しが揃っている実験計画で，何らかの理由で欠測値が生じた場合を取り上げている．なお，「第 11 章 品質管理とコンピュータ」には，1980 年代における品質管理のためのコンピュータの利用とし，Basic 言語による L_8 直交用の解析プログラムが例示されている時代でもあった．

引用

第 12 章　数理：欠測値の有る場合の処理

操返しのある実験のわりつけの場合には，データ中に欠測値があっても，繰返し数が異なるとみて処理を行えばよい．したがってここでは，繰返しのない二元配置の場合を例にとって説明しよう．まず，欠測値が 1 個の場合について考える．

中略

つぎに欠測値が二つ以上ある場合について考える．この場合， 計算はずっと複雑になるが，もし欠測値の数が少ないときにはつぎのような反復法を用いることができる．欠測値が X，Y の二つの場合，まず Y に適当な値（全体の平均値など）を与え，式(12.296 ）により X の値を求める．つぎにその X の値を用いて，式（12.296）により Y

の値を計算する． この手順を X，Y が収束するまで繰り返せばよい． 三つ以上の欠測値がある場合も同様に行うことができる．

注）式（12.296）は，以下の式（7.82）と同じなので省略する．

第 7 章 実験計画法：欠測値のある場合

この計画法では因子 A，B の各水準の組合せについて実験をすることが要求される．つまり， $\mathrm{A}_i\mathrm{B}_i$ $(i=1,2,\ldots,a;\ \ j=1,2,,\ldots,b)$ の ab 個の組合せについてのデータが必要とされる．ところが，実験の過程において何らかの理由で —例えば電子管の実験で，電子管を実験の途中で割ってしまったなど— そのうち一つのデータが得られなかったとしたら，次の手順によってその欠測値を補って，分散分析を行うことができる．データ x_{rs} が欠測した場合には，

(i) 次式の $\hat{x}_{rs}$ を x_{rs} の推定値とする．

$$\hat{x}_{rs}=\frac{aT'_{r\bullet}+bT'_{\bullet s}-T'_{\bullet\bullet}}{(a-1)(b-1)} \qquad (7.82)$$

ただし， $T'_{r\bullet}$ は r 行の $(b-1)$ 個のデータの和

$T'_{\bullet s}$ は s 行の $(a-1)$ 個のデータの和

$T'_{\bullet\bullet}$ は $(ab-1)$ 個のデータの和

を表している．

(ii) あとは欠測値がなかったと同様に分散分析を行う．ただし，誤差の自由度は欠測値のない場合から1 だけ差し引く．つまり

$$\nu_e=(a-1)(b-1)-1$$

とする．

(iii) A 間平方和 S_{A} から，次項を差し引いて補正する．

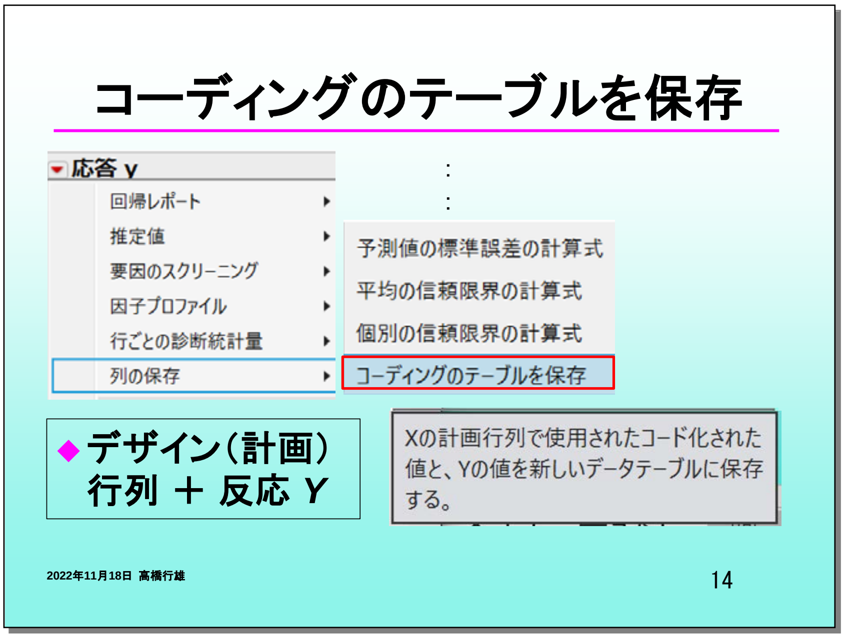
$$\frac{[T'_{\bullet s}-(a-1)\hat{x}_{rs}]^2}{a(a-1)} \qquad (7.83)$$

―――――――――――――― 引用 終り ――――――――――――――

このような手順が，その後にも綿々と語り継がれているのが現状であり，デザイン行列 $\boldsymbol{X}$ を用いた線形モデルによる分散分析について語られている日本語の書籍を見いだすことができないのが現状である．

4. デザイン行列 *X* が救いの神

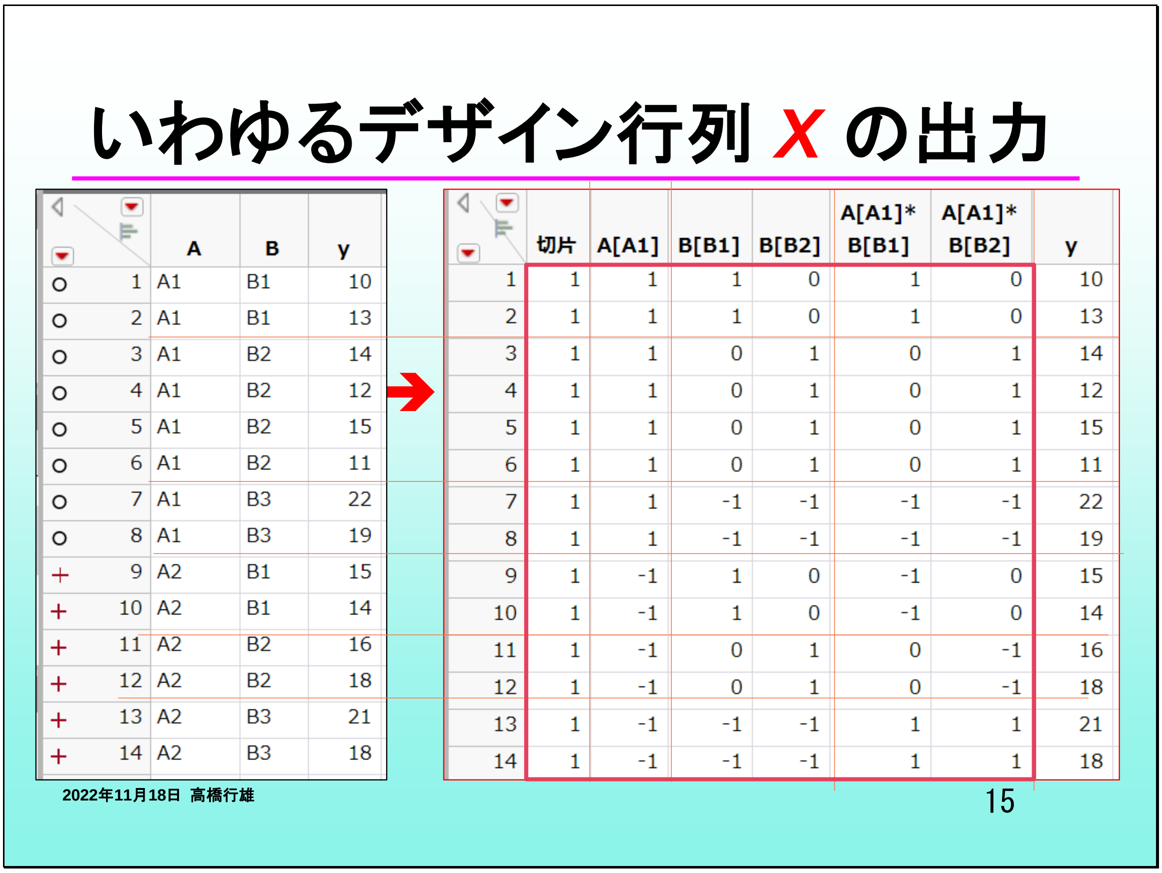
スライド 14



コーディングのテーブル

JMP の「モデルのあてはめ」のオプションで「コーディングのテーブルを保存」する機能がある．意味がまったく分からない造語であるが，「Xの計画行列で使用されたコード化された値と，Yの値を新しいデータテーブルに保存する.」との説明書きがあるので，スライド 15 に示すようにデザイン行列（計画行列）***X*** の出力と理解される．

スライド 15



	A	B	y
1	A1	B1	10
2	A1	B1	13
3	A1	B2	14
4	A1	B2	12
5	A1	B2	15
6	A1	B2	11
7	A1	B3	22
8	A1	B3	19
9	A2	B1	15
10	A2	B1	14
11	A2	B2	16
12	A2	B2	18
13	A2	B3	21
14	A2	B3	18

	切片	A[A1]	B[B1]	B[B2]	A[A1]*B[B1]	A[A1]*B[B2]	y
1	1	1	1	0	1	0	10
2	1	1	1	0	1	0	13
3	1	1	0	1	0	1	14
4	1	1	0	1	0	1	12
5	1	1	0	1	0	1	15
6	1	1	0	1	0	1	11
7	1	1	-1	-1	-1	-1	22
8	1	1	-1	-1	-1	-1	19
9	1	-1	1	0	-1	0	15
10	1	-1	1	0	-1	0	14
11	1	-1	0	1	0	-1	16
12	1	-1	0	1	0	-1	18
13	1	-1	-1	-1	1	1	21
14	1	-1	-1	-1	1	1	18

交互作用はダミー変数の積

スライド 7「3 種のダミー変数」で，JMP の「モデルのあてはめ」は，（1，–1）対比型ダミー変数が使われていることを示唆したのであるが，JMP のマニュアルにきちっとした説明がされてはいない．これは，どのようなダミー変数が使われているか，ユーザが意識しなくとも解析がスムーズに行えるような創意工夫がされているためである．したがって，JMP の「モデルのあてはめ」を使う際に，デザイン行列 $\boldsymbol{X}$ およびダミー変数に関する知識は必要とされない．JMP のユーザならば，繰り返しが不揃いの 2 元配置データの解析ができない，などの認識なしに解析を行っているに違いない．ただし，分散分析表に結果をまとめた際に，スライド 13 に示したモデルの平方和 145.4286 と（$S_A + S_B + S_{A\times B}$）の平方和の合計 133.4909 が合わないこと違和感を持つかもしれない．

スライド 16 に，JMP の内部で生成される因子 A および因子 B に対する（1，–1）対比型ダミー変数と交互作用の解析のために必要となるダミー変数間の積が表示されている．スライド 15 に示した JMP の「コーディングのテーブル」から繰り返しの重複を除いた場合に相当することが確かめられるであろう．

スライド 16

交互作用はダミー変数の積

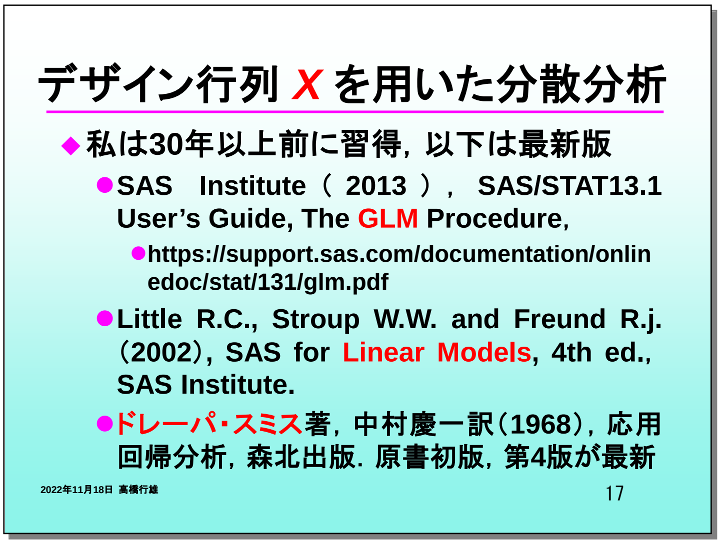
元データ				デザイン行列 X (コーディングのテーブル)						
A	B	*y*		切片	A[A1]	B[B1]	B[B2]	A[A1] *B[B1]	A[A1] *B[B2]	*y*
A_1	B_1			1	1	1	0	1	0	
A_1	B_2		➡	1	1	0	1	0	1	
A_1	B_3			1	1	-1	-1	-1	-1	
A_2	B_1			1	-1	1	0	-1	0	
A_2	B_2			1	-1	0	1	0	-1	
A_2	B_3			1	-1	-1	-1	1	1	

◆ JMPでは，デザイン（計画）行列 X を用いて分散分析を行なっている．

2022年11月18日 高橋行雄 16

デザイン行列 *X*を用いた分析分析

スライド 17



「デザイン行列 ***X*** を用いた分散分析」とは，何のことなのだろうか．SAS の GLM プロシジャも JMP の「モデルのあてはめ」も，「デザイン行列 ***X*** を用いた分散分析」を行っているが，ユーザが認識せずに使えるような工夫がなされている．

分散分析といえば，「平方和の分解」がセットで普及しており，「デザイン行列 ***X***」そのものがマイナーな存在であり，分散分析とは結びつかない．その結果として，「繰り返しが不揃いの 2 元配置」は，解析ができないとの風評が立つ原因でもあり，また，それを打ち消すような論評にも遭遇できない．

普及活動への取り組み

スライド 18

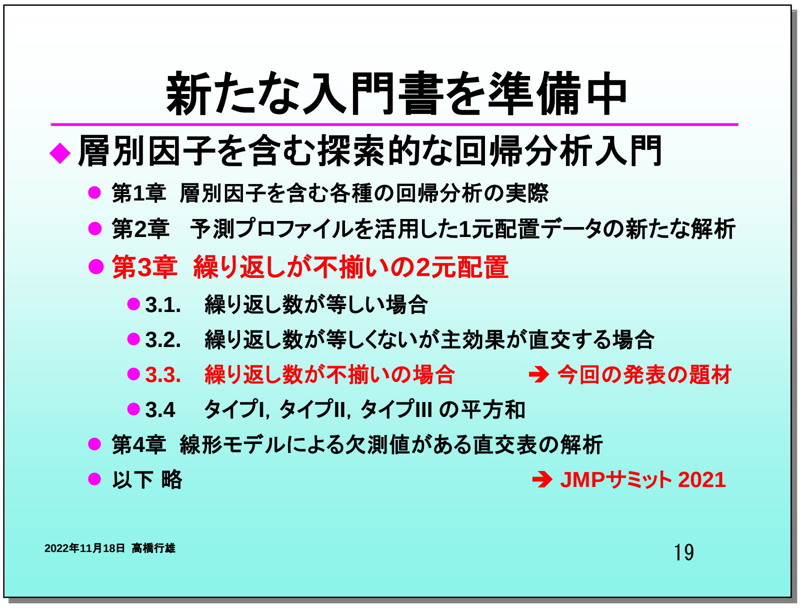
私にとって嘆かわしい状況であり，自ら普及活動に取り組むことにした．JMP サミット 2020 で，最小 2 乗平均の謎を予測プロファイルで解く，－交互作用を考慮した共分散分析を用いて－」と題し講演をした．多くの JMP ユーザが，いつも目にするけれども説明が困難な「最小 2 乗平均」について JMP のユーザにとって身近な「予測プロファイル」を活用して解説を試みた．さらに，Excel を使って，JMP と同様な解析ができることも示した．詳しくは．高橋（2020），「続・高橋セミナー 第 9 回 13 章 最小 2 乗平均の謎を予測プロファイルで解く」参照されたい．

JMP サミット 2021 では，私にとって愛着がある「直交表」を取り上げた．直交表を用いた実験データの解析は，手計算でも容易に分散分析表が作成できるのであるが，不幸なことに欠測値が発生すると，解析不能となってしまう．そのために欠測値を何らかの方法で「補完」するといった便宜的な方法が，まことしやかに語られ続けている．JMP の「モデルのあてはめ」は，そのようなことをせずに「欠測値」としたままでの解析が行えることを示し，

Excel を使って，JMP の「モデルのあてはめ」と同様の解析ができることを示した．詳しくは，高橋（2022），「続・高橋セミナー 第 10 回 線形モデルによる欠測値がある直交表の解析ー謎めいた最小 2 乗平均と 95%信頼区間の活用ー」を参照されたい．

新たな入門書を準備中

スライド 19

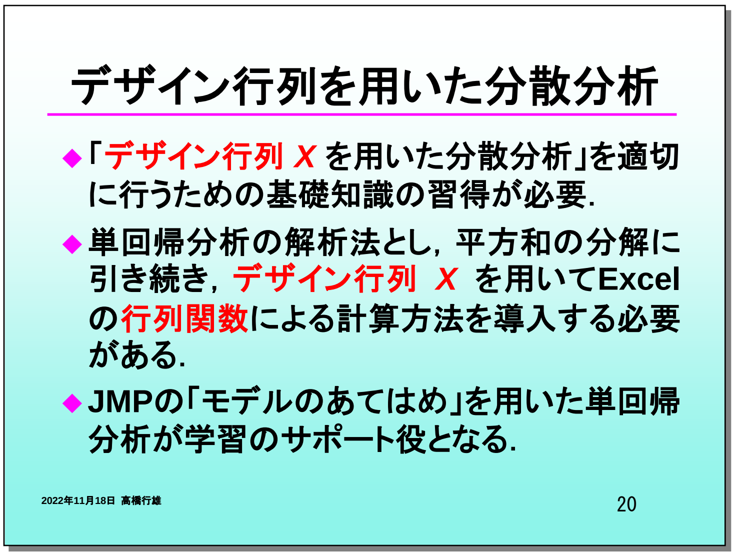


JMP サミット 2022 で取り上げた「繰り返しが不揃いの 2 元配置」は，出版準備中の「層別因子を含む探索的な回帰分析入門」の第 3 章を推敲することも兼ねたものである．

第 4 章は，JMP サミット 2021 で取り上げたテーマであるが，Excel を主体にし，JMP は，Excel の結果を検証するために用いている．ただし，JMP サミットでは，もちろん JMP を主体にし，理解を深めるために Excel による解析方法を示している．

Excel の行列関数の活用

スライド 20



「デザイン行列 $\boldsymbol{X}$ を用いた分散分析」を適切に行うための基礎知識の習得をどのようにして行ったらよいのだろうか．偏差平方和を用いた単回帰分析の解析法に引き続き，デザイン行列 $\boldsymbol{X}$ を用いた Excel の行列関数による計算方法を習得することが望ましい．学習のサポート役には，JMP の「モデルのあてはめ」を用いた単回帰分析が最適である．

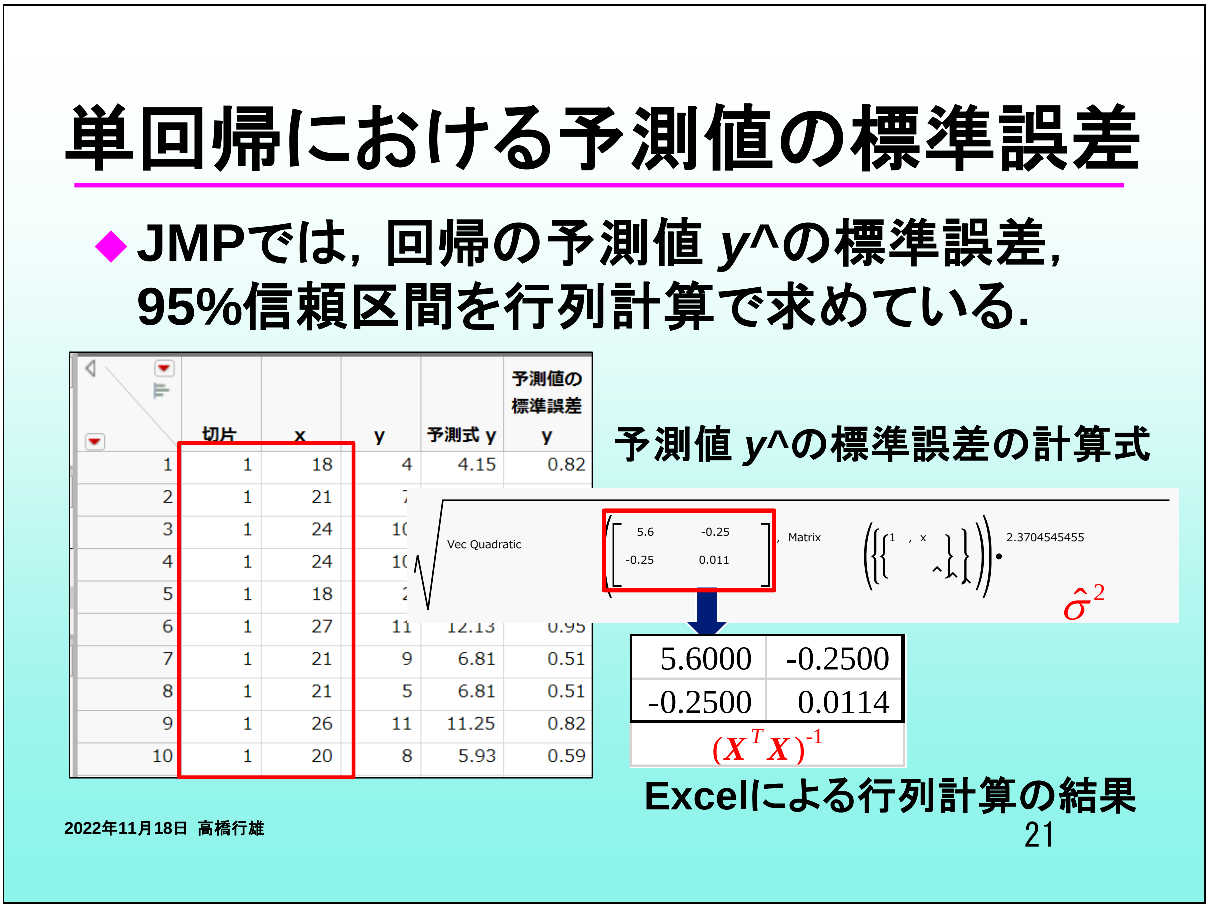
「デザイン行列を用いた回帰分析」は，すでに出版した高橋（2021），「最尤法によるポアソン回帰分析入門」の第 4 章で取り上げている．また，Web でも高橋（2020），「続・高橋セミナー 第 9 回 第 4 章 デザイン行列を用いた回帰分析入門」として公開しているので，参考にしてもらいたい．

単回帰における予測値の標準誤差

スライド 21 に示すのは，JMP の「モデルのあてはめ」を使って単回帰分析を行い，「予測式 y」および「予測値の標準誤差 y」を「列の保存」により JMP ファイルに書き出した結果である．JMP の優れているのは，「予測値の標準誤差 y」の列には，計算式が埋め込まれていて，その計算式を表示することができる．そのために，当該の列に埋め込まれている「計算式」を選択し表示させることにより JMP の内部での計算方法を知ることができる．

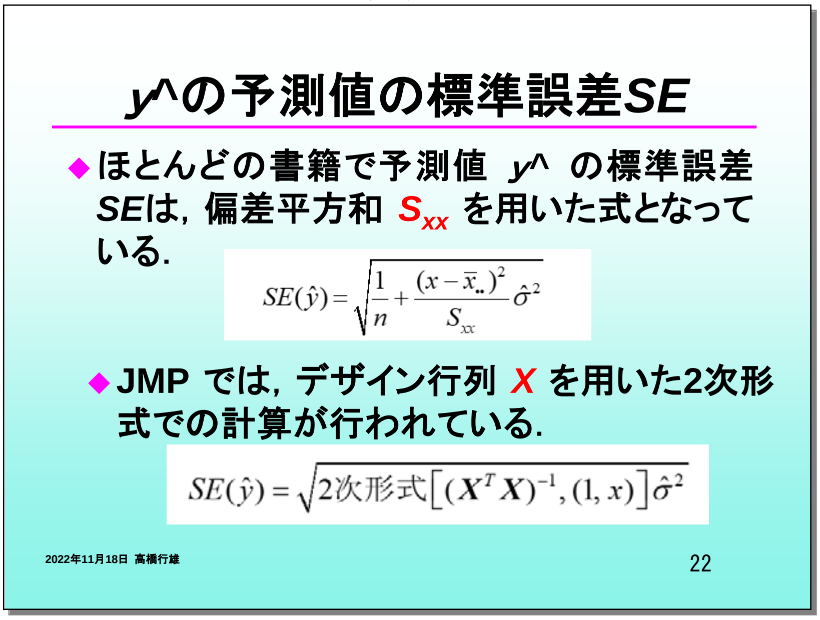
予測値の標準誤差 *SE* は，回帰直線の 95%信頼区間の計算のために用いられるのであるが，ほとんどすべての統計の書籍で示されている計算式とは異なっている．JMP の「モデルのあてはめ」では，デザイン行列 $\boldsymbol{X}$ の積和行列 $(\boldsymbol{X}^T\boldsymbol{X})$ の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ に対する 2 次形式による計算式が用いられている．このことは，別途 Excel で計算した $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ と一致することにより確かめられる．回帰分析の理論を主体にした書籍では，この 2 次形式の計算式が示されているが，計算事例を見出すのは困難である．

スライド 21



予測値 $\hat{y}$ の標準誤差 *SE*

スライド 22



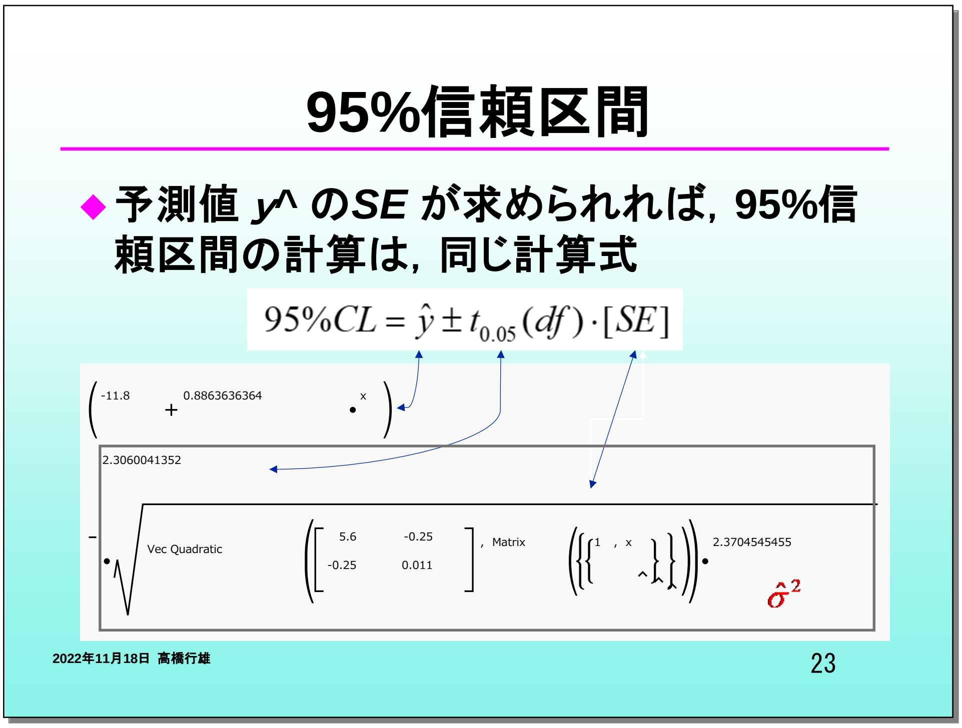
スライド 22 に示すように，単回帰分析について記述のあるほとんどすべての書籍では，予測値 $\hat{y}$ の標準誤差 *SE* の計算式は，偏差平方和 S_{xx} を用いている．この式は，単回帰分析に特化した式であり，2 変数以上のへの拡張性が欠如している．

JMP での標準誤差 *SE* の計算式は，行列計算が含まれていることもあり，敷居が高すぎると思われるかもしれない．ただし，2 変数以上の場合でも計算式中の(1 *x*)をベクトルに拡張するだけで，計算式そのものに変わりはなく汎用的に使える優れものである．

直線のあてはめが悪いので，2 次式あるいは 3 次式をあてはめて 95%信頼区間も示したい．その場合に，標準誤差 *SE* の計算式はどうなるのだろうか．2 次式の予測値 $\hat{y}$ の標準誤差 *SE* の計算式は，4 次式になり，偏差平方和 S_{xx} を用いた式で表すことは絶望的ですらある．

予測値 $\hat{y}$ の 95%信頼区間

スライド 23



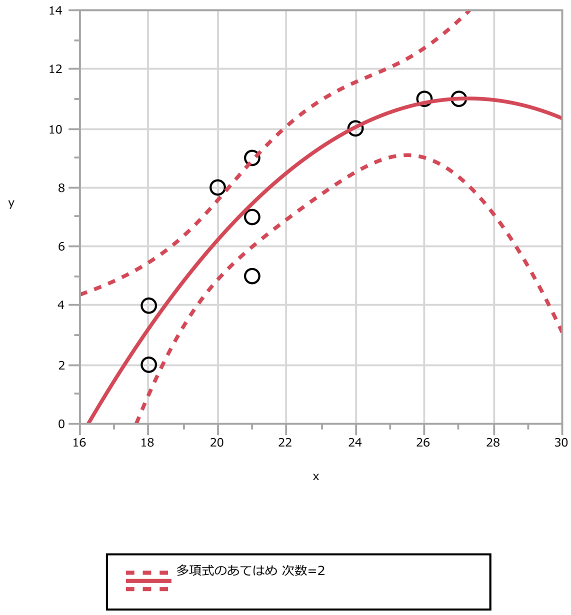
JMP の「モデルのあてはめ」で，予測値 $\hat{y}$ の 95%信頼区間を「列の保存」で JMP ファイルに書き出し，標準誤差 *SE* の場合と同様に「計算式」で下限の場合を表示させることができる．

これにより，予測値の計算式，*t* 値の計算結果も示されており，行列計算を用いた単回帰分析の学習に役に立つ．

JMP による 2 次式のあてはめと 95%信頼区間の表示

2 次式をあてはめた際の 95%信頼区間について，JMP の「二変量の関係」を用いて例示し，その計算式も合わせて示す．データは，スライド 21 に示した JMP ファイルを用い，多項式のあてはめ」から「2」次 を選択し，散布図の下のプルダウンメニューから「回帰の信頼区間」，「予測値の保存」，「平均の信頼区間の計算式」を選択する．以下に結果を示す．

2 次式の 95%信頼区間の表示

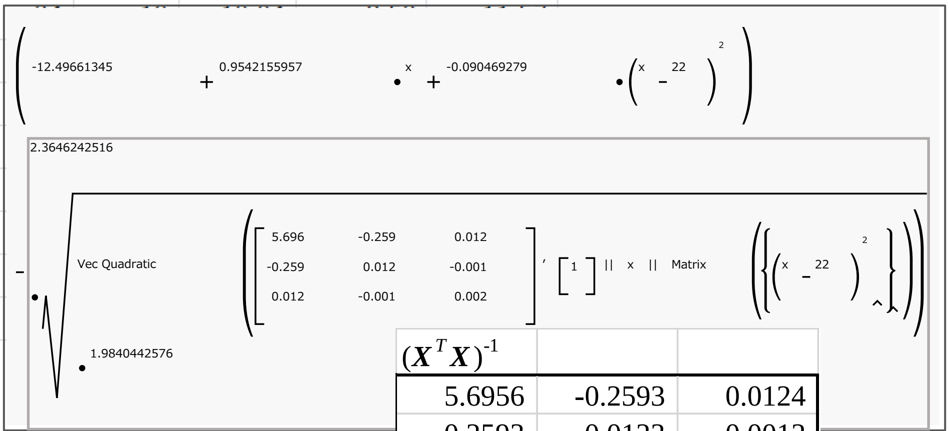


プルダウンメニュー

2 次式の予測値と 95%信頼区間の計算式の表示

	切片	x	y	予測値 y	平均 yの下側95%	平均 yの上側95%
1	1	18	4	3.23	1.00	5.47
2	1	21	7	7.45	5.99	8.91
7	1					
8	1					
9	1					
10	1					

2 次式の予測値の 95%信頼区間の下限の計算式



$(X^TX)^{-1}$		
5.6956	-0.2593	0.0124
-0.2593	0.0123	-0.0012
0.0124	-0.0012	0.0016

Excel による 2 次式の 95%信頼区間の推定の実際

表 3　2 次式の 95%信頼区間の計算表

デザイン行列 $\boldsymbol{X}$								
切片	x	$(x\text{-}22)^2$	y	y^	$Var(y$^$)$	SE	L95%	U95%
1	16	36		-0.4861	4.2253	2.0555	-5.3466	4.3745
1	18	16	4	3.2318	0.8920	0.9445	0.9984	5.4651
1	18	16	2	3.2318	0.8920	0.9445	0.9984	5.4651
1	20	4	8	6.2258	0.3234	0.5687	4.8810	7.5706
1	21	1	7	7.4514	0.3800	0.6164	5.9938	8.9091
1	21	1	9	7.4514	0.3800	0.6164	5.9938	8.9091
1	21	1	5	7.4514	0.3800	0.6164	5.9938	8.9091
1	24	4	10	10.0427	0.4156	0.6447	8.5183	11.5670
1	24	4	10	10.0427	0.4156	0.6447	8.5183	11.5670
1	26	16	11	10.8655	0.6156	0.7846	9.0102	12.7207
1	27	25	11	11.0055	1.2580	1.1216	8.3533	13.6576
1	28	36		10.9645	2.6588	1.6306	7.1088	14.8202
1	30	64		10.3398	9.3863	3.0637	3.0953	17.5843
x でソートしている						$t_{0.05}(10\text{-}3)=$	2.3646	

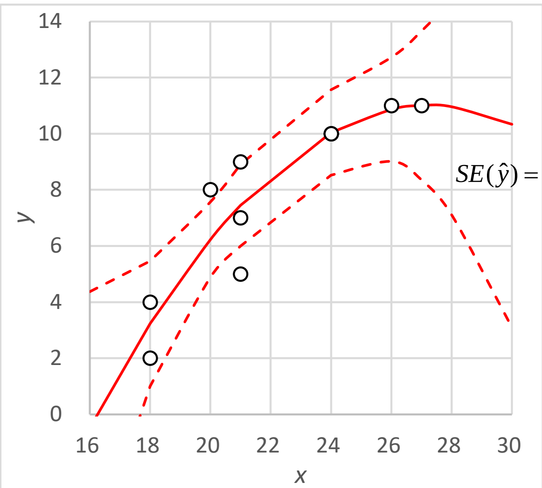
表 4　Excel の回帰分析によるパラメータの推定

分散分析表				
	自由度	変動	分散	分散比
回帰	2	74.2117	37.1058	18.7021
残差	7	13.8883	**1.9840**	$=\sigma$^2
合計	9	88.1000		

	係数	標準誤差	t	P-値
切片	**-12.4966**	3.3616	-3.7175	0.0075
x	**0.9542**	0.1560	6.1155	0.0005
$(x\text{-}22)^2$	**-0.0905**	0.0566	-1.5994	0.1538

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		
5.6956	-0.2593	0.0124
-0.2593	**0.0123**	-0.0012
0.0124	-0.0012	**0.0016**

$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^$2$		
11.3003	-0.5145	0.0246
-0.5145	0.0243	-0.0024
0.0246	-0.0024	0.0032



2 次曲線に対する 95%信頼区間の重ね書き

$$\hat{y} = -12.4966 + 0.9542x - 0.0905\cdot(x-22)^2$$

$$SE(\hat{y}) = \sqrt{\begin{bmatrix}1 & x & (x-22)^2\end{bmatrix}\begin{bmatrix}5.6956 & -0.2593 & 0.0124 \\ -0.2593 & 0.0123 & -0.0012 \\ 0.0124 & -0.0012 & 0.0016\end{bmatrix}\begin{bmatrix}1 \\ x \\ (x-22)^2\end{bmatrix}\hat{\sigma}^2}$$

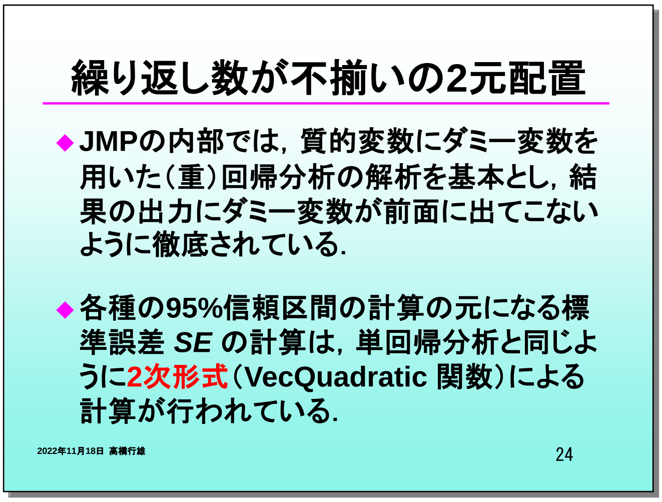
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		
5.6956	-0.2593	0.0124
-0.2593	0.0123	-0.0012
0.0124	-0.0012	0.0016

$$\hat{\sigma}^2 = 1.9840$$

5. 繰り返しが不揃いの 2 元配置

質的変数をダミー変数に

スライド 24

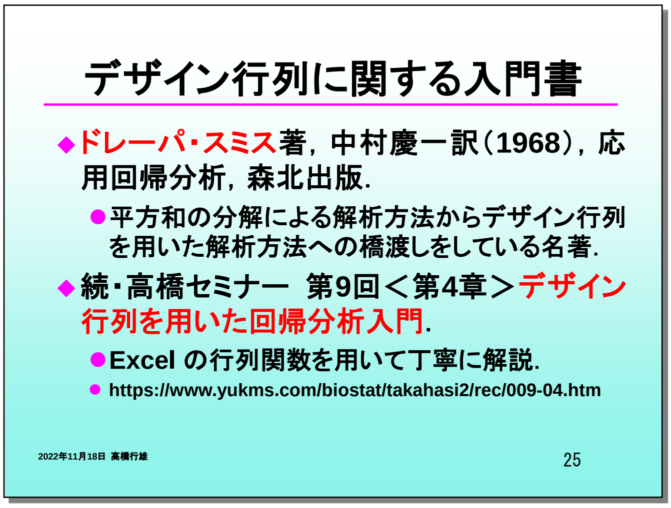


繰り返しが不揃いの 2 元配置データの解析に際し，偏差平方和の計算および平方和の分解による解析がなぜ行えないかを数値例のよって具体的に示してきた．このような状況を解決するためには，デザイン行列を用いた一般線形モデルを適用するしかないが，行列計算を伴うことから忌避されているのが現状であり，このことが，啓蒙活動を始めたきっかけである．

JMP の内部では，質的変数をダミー変数に置き換えた（重）回帰分析の解析を基本とし，結果の出力にダミー変数が前面に出てこないように徹底されている．各種の 95%信頼区間の計算の元になる標準誤差 *SE* の計算は，単回帰分析と同じように 2 次形式（VecQuadratic 関数）による計算が行われている．

デザイン行列に関する入門書

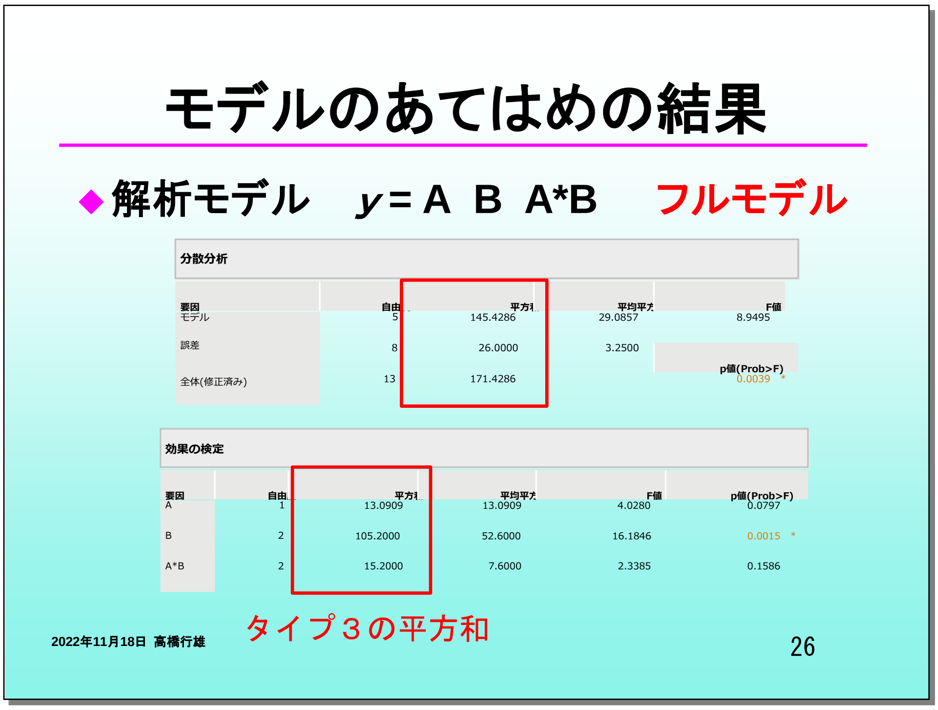
スライド 25



デザイン行列を用いた回帰分析に関する書籍は，ドレーパ，スミス著，中村慶一訳（1968），「応用回帰分析」しか見当たらない．平方和の分解による解析方法からデザイン行列を用いた解析方法への橋渡しをしている名著である．なお，原著は，第 3 版（1998）となっている．

この「応用回帰分析」でのデザイン行列を用いた回帰分析の導入法を参考に，Excel の行列関数を用いて高橋（2020），「続・高橋セミナー 第 9 回，第 4 章 デザイン行列を用いた回帰分析入門」を公開しているので参考にしてもらいたい．

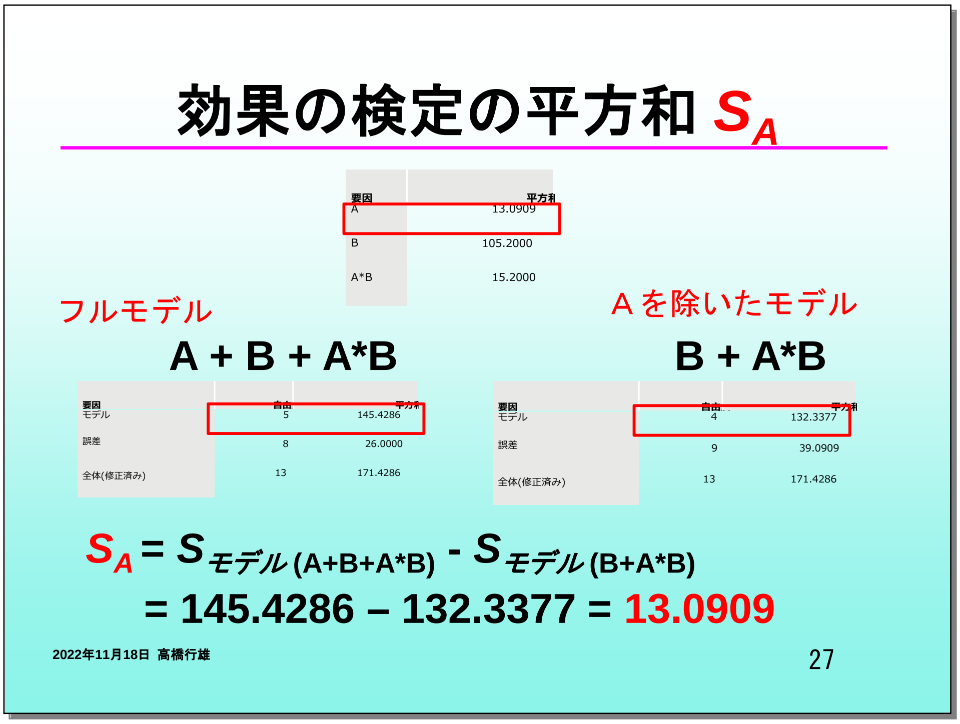
繰り返しが不揃いの 2 元配置における平方和　　　　　　スライド 26



スライド 26 に JMP の「モデルのあてはめ」による分散分析表を示す．この平方和は，スライド 13 で示したように，SAS の GLM プロシジャで出力されるタイプ III の平方和に一致する．ただし，タイプ III の平方和とは，どのようなものなのか，謎につつまれている．

Excel を用いた偏差平方和の計算では，全体の平方和 $S_{\mathrm{T}}=171.4286$， 残差平方和 $S_e=26.0000$，それらの差から求められたモデル全体の平方和 $S_{\mathrm{Model}}=145.4286$ が一致するが，S_{A}， S_{B}， $S_{\mathrm{A\times B}}$ は不一致であることを示した．

効果の検定の平方和 S_{A}　　　　　　スライド 27

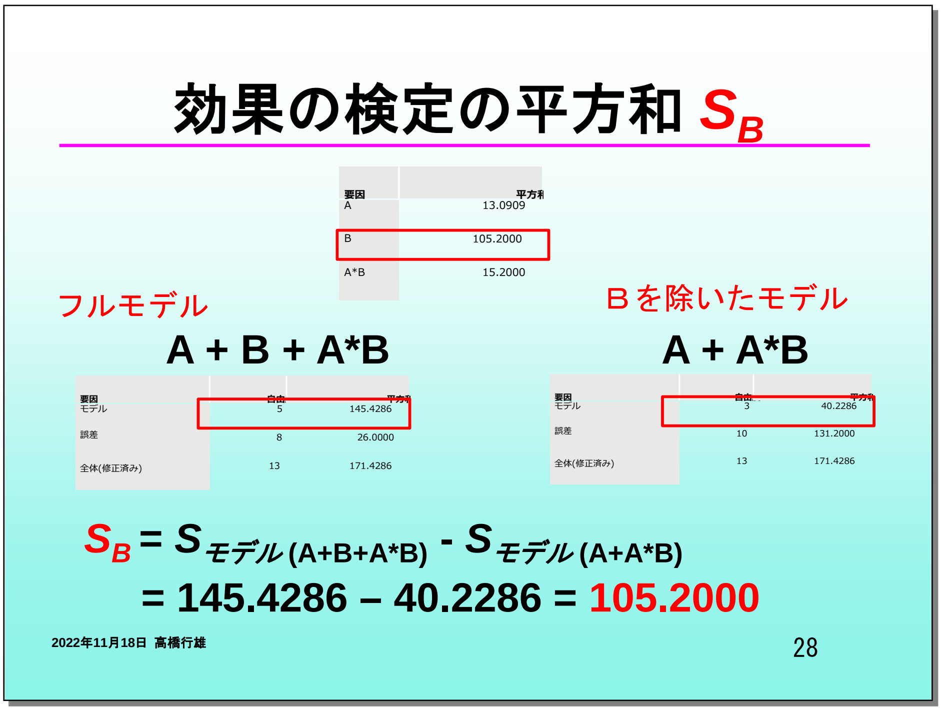


「効果の検定」で出力される平方和 $S_{\mathrm{A}}=13.0909$ は，スライド 27 に示すように，フルモデル A+B +A*B の平方和 145.4286 から，因子 A を除いたモデル B+A*B の平方和 132.3377 を差し引いた 13.0909 に一致する．

言い換えると，モデル B+A*B に因子 A をモデルに追加したときの平方和の増分ともいえる．

効果の検定の平方和 S_B　　　　　　　　スライド 28

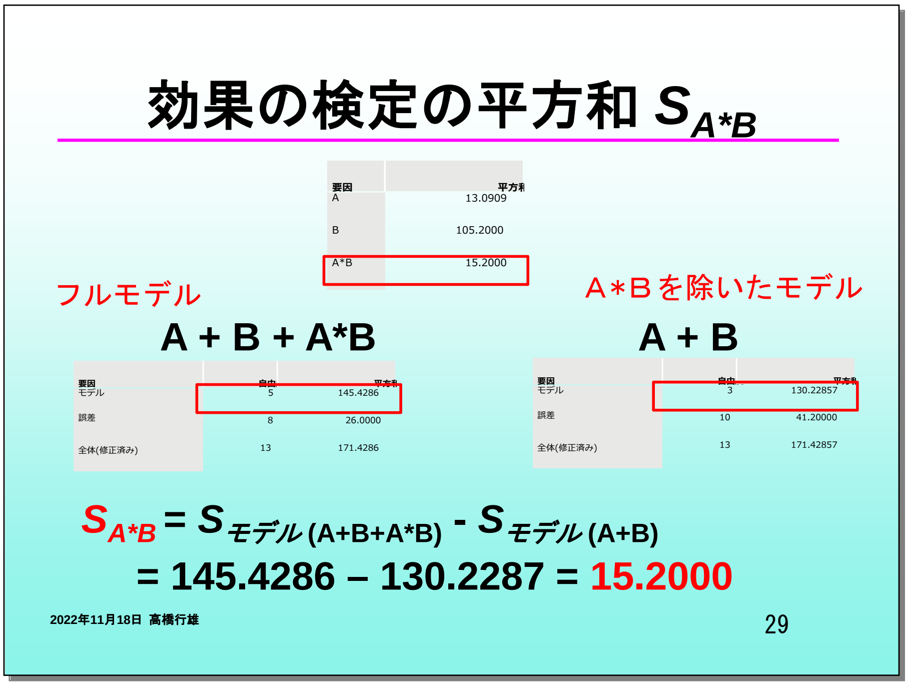
　「効果の検定」で出力される平方和 S_B =105.2000 は，スライド 28 に示すように，フルモデル A+B+A*B の平方和 145.4286 から，因子 B を除いたモデル A+A*B の平方和 40.2286 を差し引いた 105.2000 に一致する．



　言い換えると，モデル A+A*B に因子 B をモデルに追加したときの平方和の増分ともいえる．

効果の検定の平方和 S_{A*B}　　　　　　　　スライド 29

　「効果の検定」で出力される平方和 S_{A*B} =15.2000 は，スライド 29 に示すように，フルモデル A+B+A*B の平方和 145.4286 から，交互作用 A*B を除いたモデル A+ B の平方和 130.2286 を差し引いた 15.2000 に一致する．



　言い換えると，モデル A+ B に交互作用 A*B をモデルに追加したときの平方和の増分ともいえる．

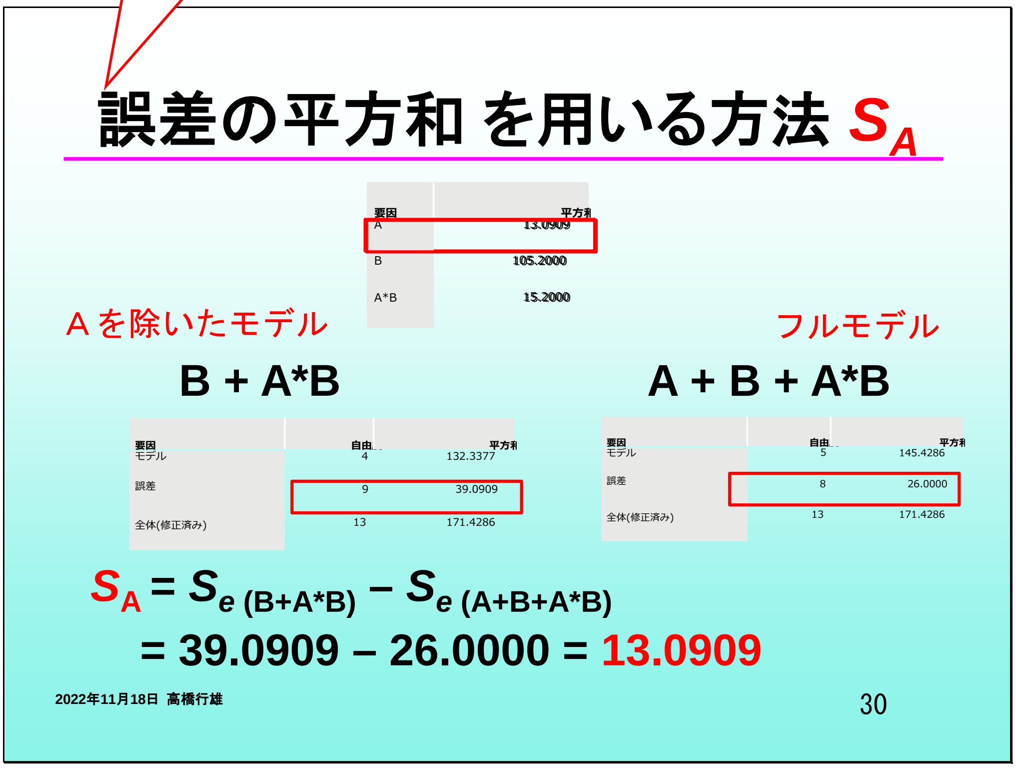
残差（誤差）の平方和を用いる方法 S_A

効果の検定で出力される平方和 S_A =13.0909 は，スライド 30 に示すように，因子 A を除いたモデル B+A*B の残差（誤差）平方和 39.0909 からフルモデル A+B+A*B の残差平方和 26.0000 を差し引いた平方和 13.0909 に一致する．

言い換えると，モデル B+A*B に因子 A をモデルに追加したときの平方和の減少分ともいえる．

残差の平方和が正しい．
分散分析表の“誤差”
に惑わされやすい

スライド 30



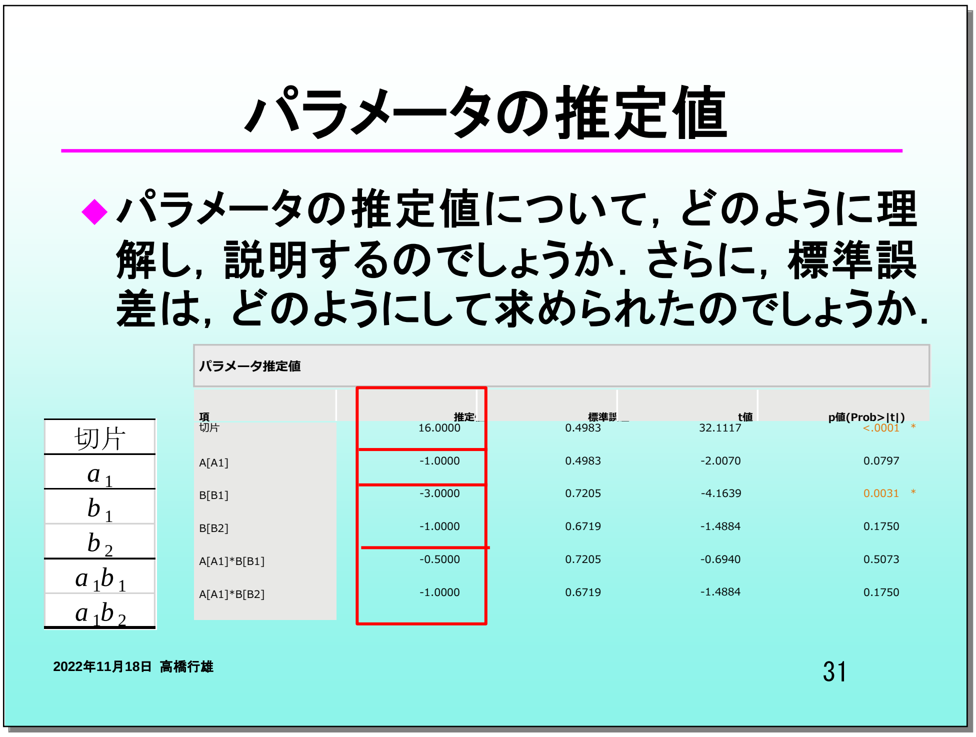
6. パラメータの推定値と標準誤差 *SE*

パラメータの推定値

分散分析と言えば，各種の偏差平方和，分散の加法性を用いて分散分析表を作成し，残差の平均平方を誤差分散の推定値$\hat{\sigma}^2$とし，各要因の平均平方を誤差分散$\hat{\sigma}^2$で除したF値による検定を行うことと理解されている．F検定の結果が有意となった場合に，各種の水準平均に対する標準誤差*SE*の推定は，有効反復数n_eを別途計算し，$(\hat{\sigma}^2/n_e)$で求めることが定式化されている．ただし，繰り返しが不揃いの2元配置の場合には，誤差分散の推定値$\hat{\sigma}^2$が得られることは示したが，各要因の平方和を求めることができなかった．

そのために，JMP の内部では，スライド 15 に示したデザイン行列 $\boldsymbol{X}$ （別名：コーディングのテーブル）に展開され，（重)回帰分析が JMP の内部で適用されている．その結果として，スライド 31 に示すように，デザイン行列 $\boldsymbol{X}$ の変数に対するパラメータの推定値が出力される．

スライド 31



因子 A に対する変数名は，（1，–1）対比型ダミー変数の「1」に対応する第 1 水準の意味の A[A1] となっている．

因子 B は，3 水準なのでスライド 7 に示すように第 1 水準として B[B1]，第 2 水準として B[B2] が変数名となっている．交互作用 A×B は，因子 A と因子 B のダミー変数の積 A[A1]*B[B1]， A[A1]*B[B2] が変数名となっている．

さて，パラメータの推定値について，どのように理解し，説明するのだろうか．さらに，標準誤差は，どのようにして求められたのだろうか．Excel の行列計算を用いてパラメータの推定値を再現する．

パラメータ *θ*^の推定値　手順1）(X^TX)

変数名は，（切片， A[A1]，. . . ， A[A1]*B[B2]）であるが，それらの推定値を $\hat{\boldsymbol{\theta}}=[\hat{\theta}_0,\ \hat{\theta}_1, . . . ,\ \hat{\theta}_5]^T$ とする．デザイン行列を $\boldsymbol{X}$ としたときに推定値 $\hat{\boldsymbol{\theta}}$ は，線形モデルの公式

$$\hat{\boldsymbol{\theta}}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

によって推定される．この式は，単回帰分析の切片を $\hat{\beta}_0$，傾きを $\hat{\beta}_1$ とし，それらの列ベクトルを $\hat{\boldsymbol{\beta}}$ としたときに， $\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ と全く同じであるが，ダミー変数を用いていることを強調するために $\hat{\boldsymbol{\theta}}$ としている．

スライド 32 に示したのは，スライド 15 に示したデザイン行列 $\boldsymbol{X}$ を，Excel シート上に再現した結果である． $\boldsymbol{X}^T$ はデザイン行列の転置（Transpose）行列で，デザイン行列 $\boldsymbol{X}$ の行と列を入れ替えた行列である．実際の操作は， $\boldsymbol{X}^T$ の範囲（6 行×14 列）の範囲を選択し，その中で Excel の関数，「=Transpose($\boldsymbol{X}$ の範囲の選択)」を入力し，コントロールキーとシフトキーを同時に選択し，エンターすると計算結果が得られる. .

実際には転置行列を Excel シート上に作成しなくとも積和行列 ($\boldsymbol{X}^T\boldsymbol{X}$) の計算結果を Excel シート上に作成できることをスライド 32 の下に計算式を示している．

スライド 32

パラメータ *θ*^ の推定値 1

$$\theta\text{\^{}} = (X^TX)^{-1}X^Ty$$

$\boldsymbol{X}^T$ デザイン行列

切片	1	1	1	1	1	1	1	1	1	1	1	1	1	1
A[A1]	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
B[B1]	1	1	0	0	0	0	-1	-1	1	1	0	0	-1	-1
B[B2]	0	0	1	1	1	1	-1	-1	0	0	1	1	-1	-1
A[A1] *B[B1]	1	1	0	0	0	0	-1	-1	-1	-1	0	0	1	1
A[A1] *B[B2]	0	0	1	1	1	1	-1	-1	0	0	-1	-1	1	1

$\boldsymbol{X}$

1	1	1	0	1	0
1	1	1	0	1	0
1	1	0	1	0	1
1	1	0	1	0	1
1	1	0	1	0	1
1	1	0	1	0	1
1	1	-1	-1	-1	-1
1	1	-1	-1	-1	-1
1	-1	1	0	-1	0
1	-1	1	0	-1	0
1	-1	0	1	0	-1
1	-1	0	1	0	-1
1	-1	-1	-1	1	1
1	-1	-1	-1	1	1

= $\boldsymbol{X}^T\boldsymbol{X}$

14	2	0	2	0	2
2	**14**	0	2	0	2
0	0	**8**	4	0	0
2	2	4	**10**	0	2
0	0	0	0	**8**	4
2	2	0	2	4	**10**

=Mmult(Transpose($\boldsymbol{X}$ の範囲),$\boldsymbol{X}$ の範囲)

2022年11月18日 高橋行雄

32

パラメータ *θ*^の推定値　手順 2）（$X^T y$）

スライド 33 に示すのは，デザイン行列 $\boldsymbol{X}$ の転置行列 $\boldsymbol{X}^T$ と，反応 $\boldsymbol{y}$ の列ベクトルとの積和により $(\boldsymbol{X}^T\boldsymbol{y})$ の列ベクトルを Excel シート上に作成した結果である． $\boldsymbol{X}^T$ が Excel シート上にあれば，Excel の Mmult () 関数を用い

$$(\boldsymbol{X}^T\boldsymbol{y}) = \text{Mmult}\ (\boldsymbol{X}^T \text{の範囲},\ \boldsymbol{y} \text{の範囲})$$

として計算できる． $\boldsymbol{X}^T$ が Excel シート上に無くとも

$$(\boldsymbol{X}^T\boldsymbol{y}) = \text{Mmult}\ (\text{Transpose}(\boldsymbol{X} \text{の範囲}),\ \boldsymbol{y} \text{の範囲})$$

計算式の引数に Transpose () 関数を入れ子状に設定することもできる．

行列の計算は，左側の行列の「行」と右側の行列（ベクトル）の「列」に対し，対応するセル同士の積を計算し，全てのセルの積に対する和で計算されている．

スライド 33

パラメータ *θ*^ の推定値 2

$$\theta\hat{}\ = (X^TX)^{-1}X^Ty$$

	X^T														y		X^Ty
切片	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10	=	218
A[A1]	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	13		14
B[B1]	1	1	0	0	0	0	-1	-1	1	1	0	0	-1	-1	14		-28
B[B2]	0	0	1	1	1	1	-1	-1	0	0	1	1	-1	-1	12		6
A[A1] *B[B1]	1	1	0	0	0	0	-1	-1	-1	-1	0	0	1	1	15		-8
A[A1] *B[B2]	0	0	1	1	1	1	-1	-1	0	0	-1	-1	1	1	11		16
															22		
															19		
															15		
															14		
															16		
															18		
															21		
															18		

2022年11月18日　高橋行雄

33

パラメータ $\boldsymbol{\theta}$^の推定値　手順 3）（$\boldsymbol{X}^T\boldsymbol{X}$）$^{-1}$（$\boldsymbol{X}^T\boldsymbol{y}$）

スライド 34 に示すように，手順 1 で計算された積和行列 ($\boldsymbol{X}^T\boldsymbol{X}$) の逆行列を Minverse () 関数で求めた ($\boldsymbol{X}^T\boldsymbol{X}$)$^{-1}$ と手順 2 で計算された ($\boldsymbol{X}^T\boldsymbol{y}$) との積和によってパラメータの推定値 $\hat{\boldsymbol{\theta}}$ が

$$\hat{\boldsymbol{\theta}} = \text{Mmult (} (\boldsymbol{X}^T\boldsymbol{X})^{-1} \text{ の範囲, } (\boldsymbol{X}^T\boldsymbol{y}) \text{ の範囲)}$$

として求められている．

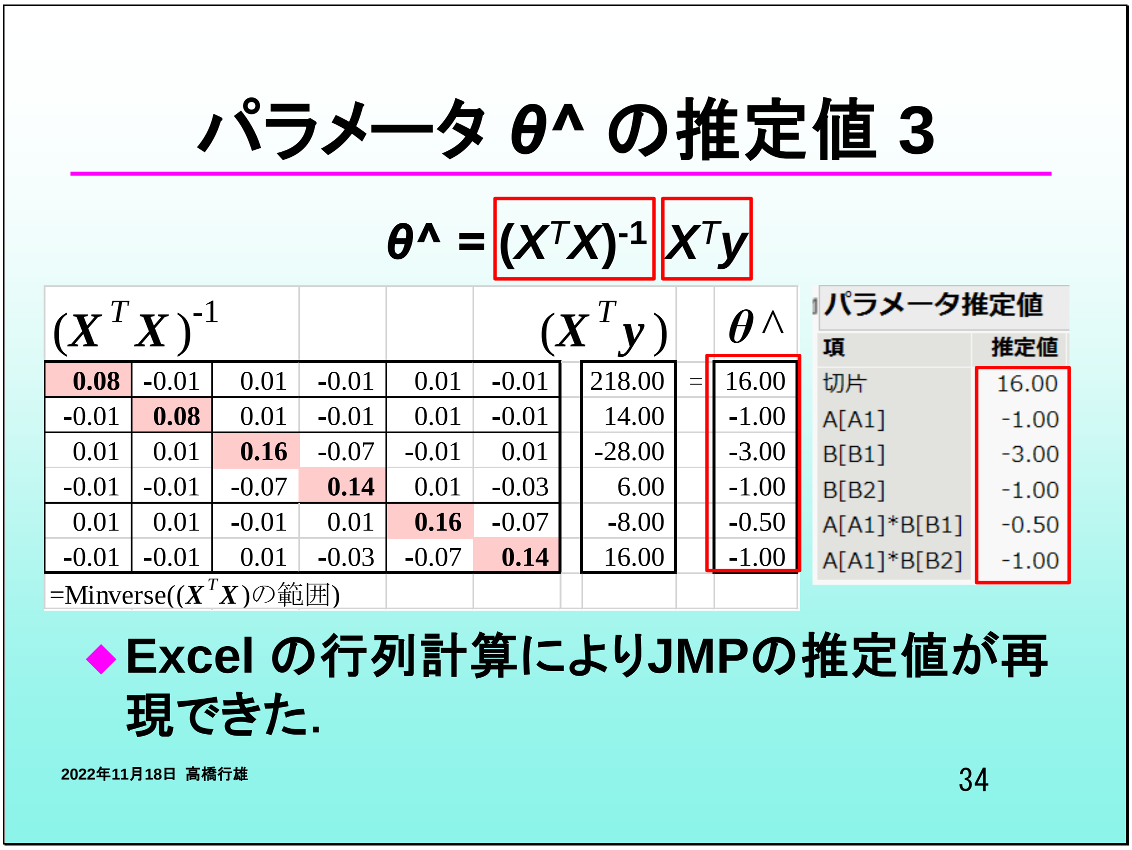
計算の途中結果を示しながらパラメータの推定値 $\hat{\boldsymbol{\theta}}$ を求める手順を示したきたのであるが，元々あるのはデザイン行列 $\boldsymbol{X}$ と反応ベクトル $\boldsymbol{y}$ だけであり，中間結果を Excel シート上に展開するのは煩わしいので，一気呵成にパラメータの推定値 $\hat{\boldsymbol{\theta}}$ を求めるためには，

$$\hat{\boldsymbol{\theta}} = \text{Mmult (Minverse (Mmult (Transpose(}\boldsymbol{X}\text{), } \boldsymbol{X}\text{))), Mmult (Transpose (}\boldsymbol{X}\text{), } \boldsymbol{y}\text{))}$$

のように計算することも可能である．

Excel の行列計算で求めたパラメータの推定値 $\hat{\boldsymbol{\theta}}$ が，JMP のモデルのあてはめで求めた結果に一致することが確かめられる．

スライド 34



パラメータ $\boldsymbol{\theta}$^の Excel の回帰分析を用いた計算

ダミー変数を用いたデザイン行列が Excel シート上にあれば，行列計算をしなくても Excel の分析ツールの回帰分析を使えば，容易にパラメータの推定値 $\hat{\boldsymbol{\theta}}$ を Excel シート上のデザイン行列 $\boldsymbol{X}$ の横に書き込むことができる．

表 5 Excel の回帰分析を用いた分散分析表とパラメータの推定値

			—— デザイン行列 $\boldsymbol{X}$ ——					
		y	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2
A_1	B_1	10	1	1	1	0	1	0
		13	1	1	1	0	1	0
	B_2	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
	B_3	22	1	1	-1	-1	-1	-1
		19	1	1	-1	-1	-1	-1
A_2	B_1	15	1	-1	1	0	-1	0
		14	1	-1	1	0	-1	0
	B_2	16	1	-1	0	1	0	-1
		18	1	-1	0	1	0	-1
	B_3	21	1	-1	-1	-1	1	1
		18	1	-1	-1	-1	1	1

	分散分析表				
		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		
		係数	標準誤差	t	P-値
θ^$_0$	切片	16.00	0.4983	32.1117	0.0000
θ^$_1$	a_1	-1.00	0.4983	-2.0070	0.0797
θ^$_2$	b_1	-3.00	0.7205	-4.1639	0.0031
θ^$_3$	b_2	-1.00	0.6719	-1.4884	0.1750
θ^$_4$	a_1b_1	-0.50	0.7205	-0.6940	0.5073
θ^$_5$	a_1b_2	-1.00	0.6719	-1.4884	0.1750

Excel の統計計算にかぎらず数値計算には，不具合がありなかなか改善されないとの指摘が，McCullough ら(2008) などによって 2010 年ごろまであったが，最近は見かけなくなってきた．2010 年以後，Excel の回帰分析および行列関数を用いた啓蒙活動を継続的に続けているが，これまで不具合に遭遇したことはない．逆行列の計算で多重共線性が原因と思われるおかしな結果が出たので，SAS でも JMP で計算をしたところ，同じようにおかしな結果となったことを経験したことがある．これらのことからも，Excel の数値計算について信頼性が担保されているように思われ，教育目的には安心して使えると判断している．

McCullough B.D., David A. Heiser D.A.(2008), On the accuracy of statistical procedures in Microsoft Excel 2007, Computational Statistics and Data Analysis 52 (2008) 4570–4578.

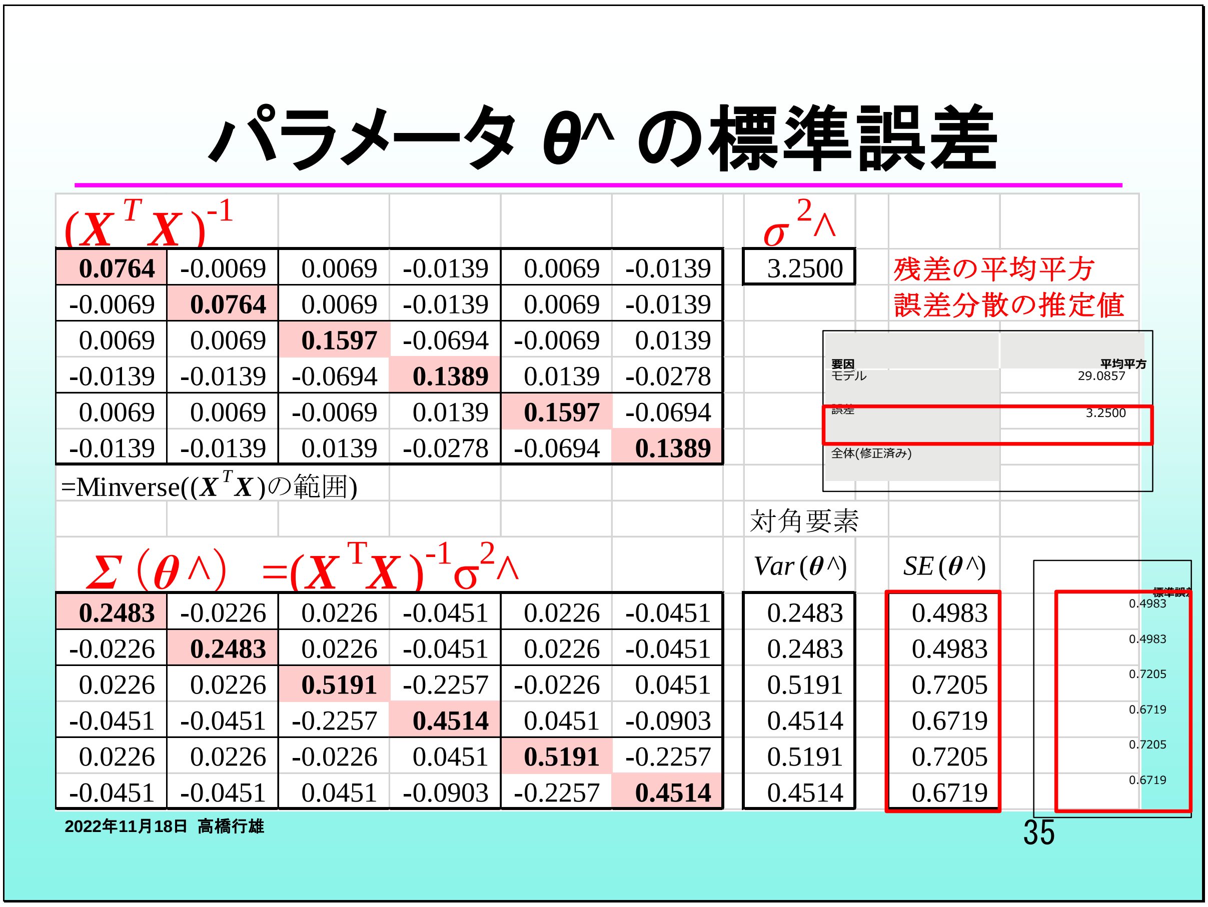
パラメータ $\boldsymbol{\theta}$^の標準誤差

Excel の行列関数を用いることにより，パラメータの推定が JMP の「モデルのあてはめ」と同様の結果が得られることを示した．スライド 21，22，23 で単回帰分析の予測値 $\hat{y}$ の標準誤差，および，95%信頼区間の計算にデザイン行列 $\boldsymbol{X}$ の積和行列 $(\boldsymbol{X}^T\boldsymbol{X})$ の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ が使われていることを示した．

スライド 35 に示すように，繰り返しが不揃いの 2 元配置でダミー変数を使った場合でも，単回帰分析の場合と同様にパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$ が，各種の 95%信頼区間の推定に用いることができる．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の計算は，パラメータの推定で用いた $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ に分散分析表で推定されている誤差分散の推定値 $\hat{\sigma}^2$ を掛けて求めることができる．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の対角要素は，パラメータの分散であり，平方根を取ることにより，パラメータの標準誤差となることが確かめられる．

このように，デザイン行列 $\boldsymbol{X}$ を使った単回帰分析での計算方法は，変数の数が増えても全く同じである．きちっとした学習のためには，偏差平方和による単回帰分析を学習した後に，デザイン行列 $\boldsymbol{X}$ を用いた解析法の学習をすることが必要不可欠である．

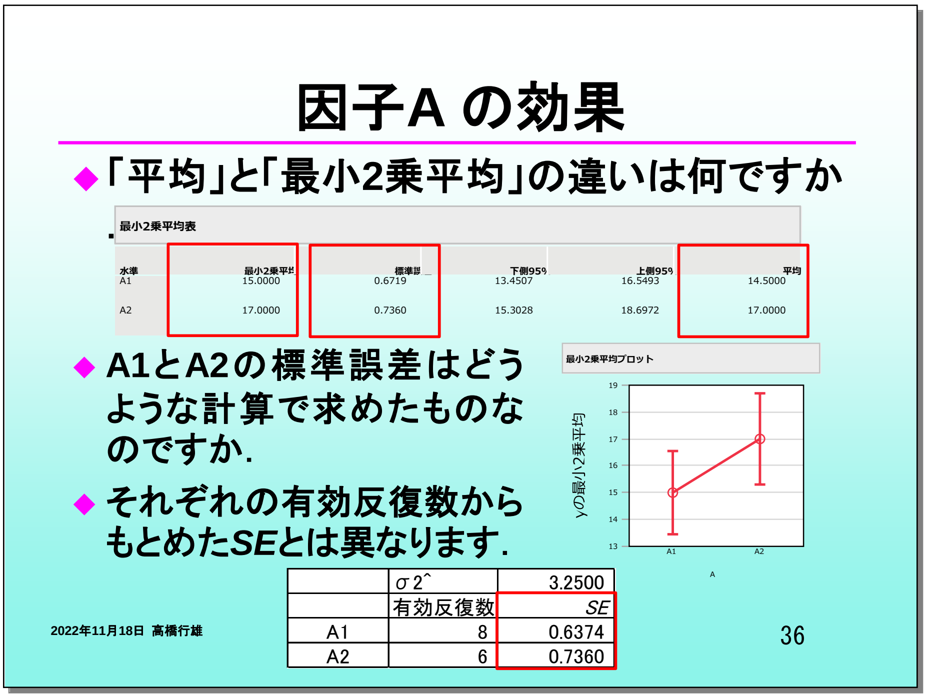
スライド 35



7. 「平均」と「最小 2 乗平均」の違いは何ですか

因子 A の効果

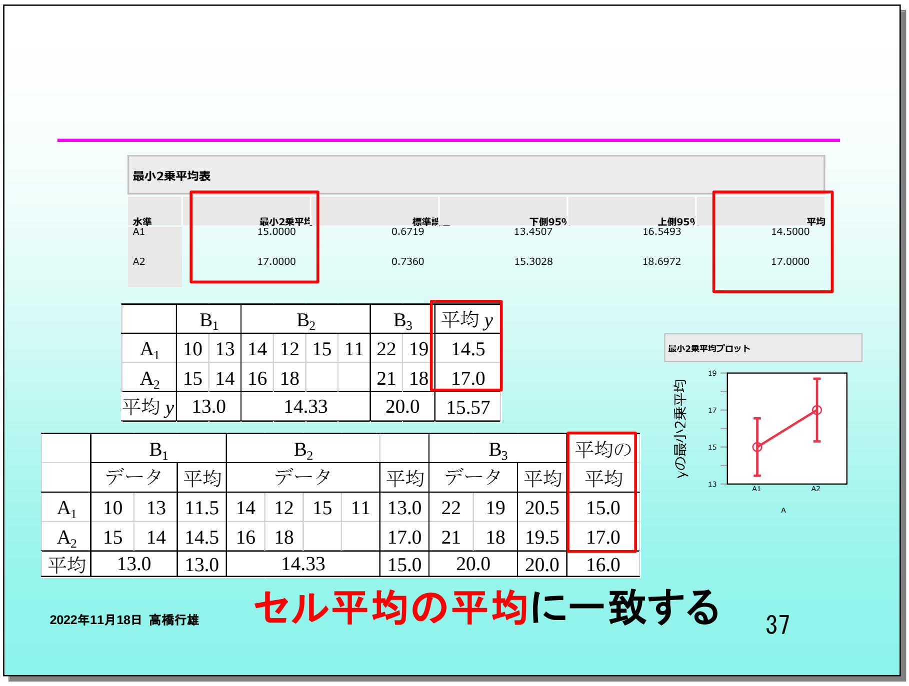
スライド 36



JMP の「モデルのあてはめ」を使い，「効果の詳細」で因子 A について展開すると必ず「最小 2 乗平均」に遭遇する．また，「平均」も合わせて示されているが，繰り返しが不揃いの A_1 水準では，15.0 と 14.5 と食い違いが起きる．

「最小 2 乗平均」は謎につつまれた統計量で，きちっとした説明をこれまで見たことがない．原因は，「最小 2 乗平均」の生みの親である SAS を使っている統計の専門家が，日本ではほとんどいないためでもある．

スライド 37



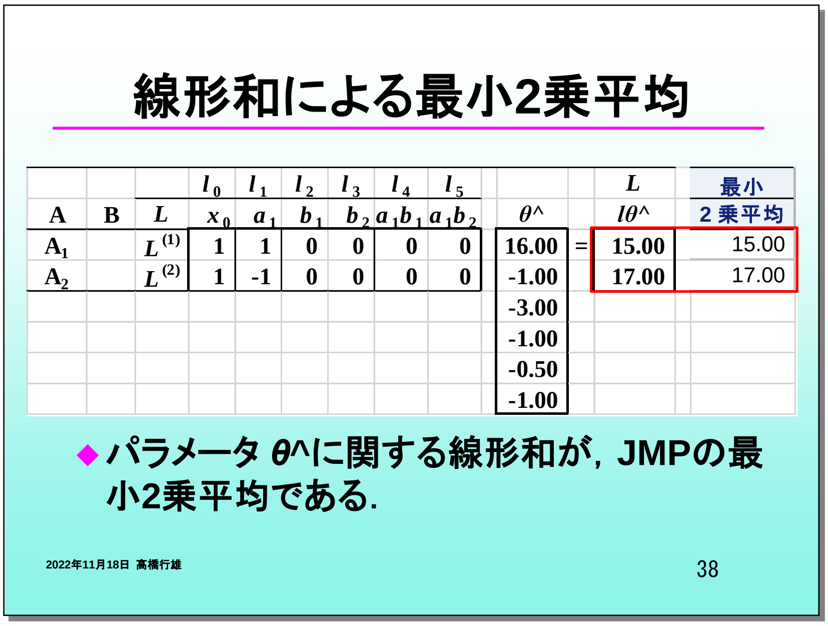
「最小2乗平均」について，「セル平均の平均です」との論評は，2 元配置では適合するが，連続変数が含まれている場合にに，説明に窮する．

また，その標準誤差は，「どのような計算なのか」と問われたときにも答えられる必要性もある．

パラメータ $\boldsymbol{\theta}$^に関する線形和が最小 2 乗均　　　　　　スライド 38

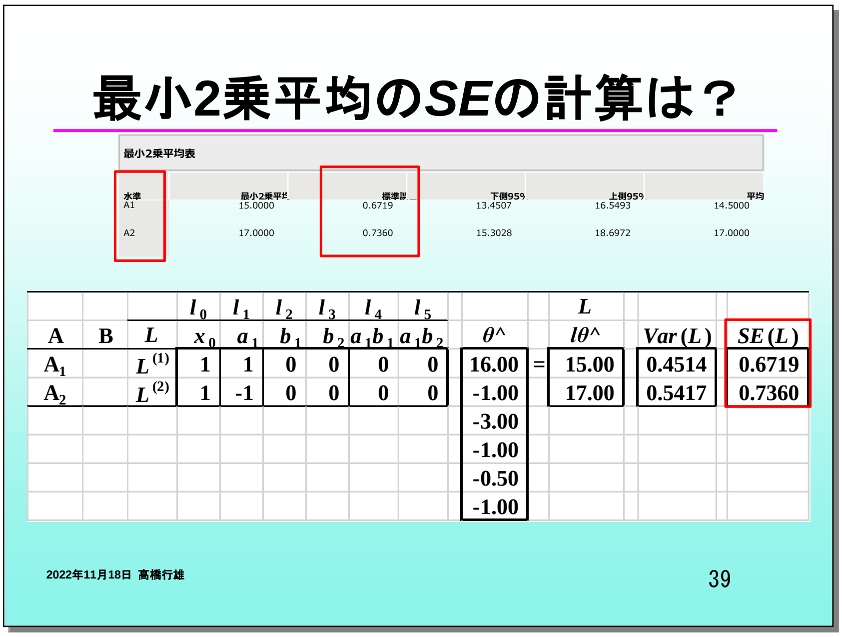
JMP の「モデルのあてはめ」で出力される「コーディングのテーブル」で使用されている変数名のままでは扱いにくいので，スライド 7 に示した（1, –1）対比型ダミー変数名を用いる．

さらに，パラメータの推定値に関する線形和を簡潔に表すため，変数名として l_i を用いる．スライド 38 に示す因子 A の第 1 水準 A_1 の最小 2 乗平均を $L^{(1)}$ としたときに， $L^{(1)} = \boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}} = 15.00$ が計算されている．実際の計算では，$L^{(1)} = 1\times\hat{\theta}_0 + 1\times\hat{\theta}_1 = 15.00$ のように，パラメータの推定値を用いた単純な足し算となっている．因子 A の第 2 水準の最小 2 乗平均 $L^{(2)} = \boldsymbol{l}^{(2)}\hat{\boldsymbol{\theta}} = 17.00$ は， $L^{(2)} = 1\times\hat{\theta}_0 + (-1)\times\hat{\theta}_1 = 17.00$ として計算されている．



最小 2 乗平均の標準誤差 *SE* の計算は？　　　　　　スライド 39

最小 2 乗平均は，パラメータの推定値の線形和として説明できる．その標準誤差 *SE* は，スライド 39 に JMP の「モデルのあてはめ」で出力された最小 2 乗平均の標準誤差 *SE* を貼り付けている．



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

スライド 40

パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ が，各種の推定値の分散を計算するために大活躍する．JMP の「モデルのあてはめ」でも SAS の GLM プロシジャでも内部の計算で使われている．

スライド 35 に示したパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ をスライド 40 に再掲する．これは，Excel シート上で行列関数を用いて計算したシート上の結果をスライドにコピー&ペーストしたものである．

パラメータの共分散行列Σ(θ^)

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

$\theta\hat{}_0$	切片	**0.2483**	-0.0226	0.0226	-0.0451	0.0226	-0.0451
$\theta\hat{}_1$	a_1	-0.0226	**0.2483**	0.0226	-0.0451	0.0226	-0.0451
$\theta\hat{}_2$	b_1	0.0226	0.0226	**0.5191**	-0.2257	-0.0226	0.0451
$\theta\hat{}_3$	b_2	-0.0451	-0.0451	-0.2257	**0.4514**	0.0451	-0.0903
$\theta\hat{}_4$	a_1b_1	0.0226	0.0226	-0.0226	0.0451	**0.5191**	-0.2257
$\theta\hat{}_5$	a_1b_2	-0.0451	-0.0451	0.0451	-0.0903	-0.2257	**0.4514**

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

線形和 $L = \boldsymbol{l}\hat{\boldsymbol{\theta}}$ の分散 $Var(\boldsymbol{l}\hat{\boldsymbol{\theta}})$ は， $\boldsymbol{l}$ に関する 2 次形式

$$Var(\boldsymbol{l}\hat{\boldsymbol{\theta}}) = \boldsymbol{l}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2]\boldsymbol{l}^T = \boldsymbol{l}\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\boldsymbol{l}^T$$

2022年11月18日 高橋行雄 40

A_1 についての標準誤差

スライド 41

A_1 の最小 2 乗平均についての標準誤差 *SE* の計算過程をスライド 41 に示す．スライド 40 下に示した一般式を，数値が入った事例により理解を深めてもらうことを期待していている．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ に対する $\boldsymbol{l}^{(1)}$ の 2 次形式であり， $\boldsymbol{l}^{(1)}$ の 0 の変数を除いた形での行列計算の過程を示してある．因子の A_1 水準の標準誤差は， $SE(\boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}}) = 0.6719$ として計算されている．分散は $Var(\boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}}) = 0.4514$ であり， 2×2 のパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ は，スライド 40 の 6×6 の共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の左上の 2×2 の部分である．共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ の要素を c_{ij} とすると

$$Var(\boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}}) = 1 \times c_{11} + 1 \times c_{12} + 1 \times c_{21} + 1 \times c_{22}$$
$$= 0.2483 - 2 \times 0.0226 + 0.2483 = 0.4514$$

として計算されている．

A_1 についての標準誤差

最小2乗平均表

水準	最小2乗平均	標準誤	下側95%	上側95%	平均
A1	15.0000	0.6719	13.4507	16.5493	14.5000
A2	17.0000	0.7360	15.3028	18.6972	17.0000

		l_0	l_1	$\boldsymbol{\Sigma}(\boldsymbol{\theta}\hat{})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma\hat{}^2$				
A	L	x_0	a_1	x_0	a_1	$\boldsymbol{l}^T$		
A_1	$L^{(1)}$	**1**	**1**	**0.2483**	-0.0226	**1**		
				-0.0226	**0.2483**	**1**		
								$Var(A_1)$
				0.2257	0.2257	**1**	=	0.4514
						1		
								$SE(A_1)$
								0.6719

2022年11月18日 高橋行雄 41

最小 2 乗平均に対する 95%信頼区間

因子 A の A_1 水準に対する最小 2 乗平均とその分散と標準誤差 SE の計算過程を詳しく示してきたのであるが，因子 B，および，交互作用 $A \times B$ の各種準についても推定したい．スライド 42 に示すように，行方向に最小 2 乗平均と 95%信頼区間の計算結果が並べられている．

実際の計算は，推定したい最小 2 乗平均に対する線形和の係数 $\boldsymbol{l}^{(i)}$ を設定し，スライド 39 に示したパラメータの推定値 $\hat{\boldsymbol{\theta}}$ との積和を A_1 水準の $L^{(1)}$ 行で

$$L^{(1)} = \boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}} \quad := \text{Mmult}\,(\ \boldsymbol{l}^{(1)}\,\text{の範囲},\ \hat{\boldsymbol{\theta}}\,\text{の範囲})$$

設定し，セルのフィルハンドルによって，行方向の計算式を連続的に変化させた結果である．線形和の分散 $Var(\boldsymbol{l}^{(i)}\hat{\boldsymbol{\theta}})$ も同様に

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

$$= \text{Mmult}\,(\ \text{Mmult}\,(\boldsymbol{l}^{(i)}\,\text{の範囲},\ \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})\,\text{の範囲}),\ \text{Transpose}\,(\boldsymbol{l}^{(i)}\,\text{の範囲})\,)$$

として計算している．

95%信頼区間の計算に，

$$t_{0.05} \times SE \quad := \text{T.inv.2T}(0.05,\ 14\text{–}6)$$

を用いている．これは，Excel の折れ線グラフで 95%信頼区間の幅を付けるための準備である．

スライド 42

最小2乗平均に対する95%信頼区間

			l_0	l_1	l_2	l_3	l_4	l_5					幅		
A	B	L	x_0	a_1	b_1	b_2	a_1b_1	a_1b_2	θ^		$l\theta$^	$Var(l\theta$^$)$	$t_{0.05} \times SE$	L95%	U95%
A_1		$L^{(1)}$	1	1	0	0	0	0	16.00	=	15.00	0.4514	1.5493	13.45	16.55
A_2		$L^{(2)}$	1	-1	0	0	0	0	-1.00		17.00	0.5417	1.6972	15.30	18.70
	B_1	$L^{(3)}$	1	0	1	0	0	0	-3.00		13.00	0.8125	2.0786	10.92	15.08
	B_2	$L^{(4)}$	1	0	0	1	0	0	-1.00		15.00	0.6094	1.8001	13.20	16.80
	B_3	$L^{(5)}$	1	0	-1	-1	0	0	-0.50		20.00	0.8125	2.0786	17.92	22.08
A_1	B_1	$L^{(6)}$	1	1	1	0	1	0	-1.00		11.50	1.6250	2.9396	8.56	14.44
	B_2	$L^{(7)}$	1	1	0	1	0	1			13.00	0.8125	2.0786	10.92	15.08
	B_3	$L^{(8)}$	1	1	-1	-1	-1	-1			20.50	1.6250	2.9396	17.56	23.44
A_2	B_1	$L^{(9)}$	1	-1	1	0	-1	0			14.50	1.6250	2.9396	11.56	17.44
	B_2	$L^{(10)}$	1	-1	0	1	0	-1			17.00	1.6250	2.9396	14.06	19.94
	B_3	$L^{(11)}$	1	-1	-1	-1	1	1			19.50	1.6250	2.9396	16.56	22.44

$l\theta$^ =Mmult(l の範囲, θ^ の範囲)　SE =sqrt($Var(l\theta$^$)$)　$t_{0.05}$= T.Inv.2T(0.05, 8)= 2.3060

$Var(l\theta$^$)$=Mmult(Mmult(l の範囲, Σ (θ^) の範囲),Transpose(l の範囲))

◆ GLM の最小2乗平均の95%信頼区間に一致

2022年11月18日 高橋行雄　42

最小 2 乗平均の差に対する 95%信頼区間

スライド 42 では，各因子の水準，および，交互作用の組合せ水準に対する最小 2 乗平均に対する 95%信頼区間の推定について示した．推定された最小 2 乗平均の差 $(\mathrm{A}_2-\mathrm{A}_1)$ などの 95%信頼区間を推定したい．「最小 2 乗平均の差」も「最小 2 乗平均」であり，適切な線形和の係数 $\boldsymbol{l}^{(i)}$ を設定すれば，スライド 42 で示したとまったく同じ計算式が適用できる．$(\mathrm{A}_2-\mathrm{A}_1)$ の最小 2 乗平均は，

$$(\mathrm{A}_2-\mathrm{A}_1)\ :\ \begin{cases} L^{(2)}-L^{(1)}=\boldsymbol{l}^{(2)}\hat{\boldsymbol{\theta}}-\boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}} \\ \qquad\qquad\quad =(\boldsymbol{l}^{(2)}-\boldsymbol{l}^{(1)})\hat{\boldsymbol{\theta}} \end{cases}$$

なので，係数ベクトルの差であり，

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

を与えればよい．他も同様に係数ベクトルを与える．

スライド 43

最小2乗平均の差に対する95%信頼区間

A	B	L	l_0 x_0	l_1 a_1	l_2 b_1	l_3 b_2	l_4 a_1b_1	l_5 a_1b_2	θ^		$l\theta$^	$Var(l\theta$^$)$	幅 $t_{0.05}\times SE$	L95%	U95%
A_1-A_1		$L^{(12)}$	0	0	0	0	0	0	16.00	=	0.00	0.0000	0.0000	0.00	0.00
A_2-A_1		$L^{(13)}$	0	-2	0	0	0	0	-1.00		2.00	0.9931	2.2980	-0.30	4.30
	B_1-B_1	$L^{(14)}$	0	0	0	0	0	0	-3.00		0.00	0.0000	0.0000	0.00	0.00
	B_2-B_1	$L^{(15)}$	0	0	-1	1	0	0	-1.00		2.00	1.4219	2.7497	-0.75	4.75
	B_3-B_1	$L^{(16)}$	0	0	-2	-1	0	0	-0.50		7.00	1.6250	2.9396	4.06	9.94
A_1B_1	$-A_1B_1$	$L^{(17)}$	0	0	0	0	0	0	-1.00		0.00	0.0000	0.0000	0.00	0.00
A_1B_2	$-A_1B_1$	$L^{(18)}$	0	0	-1	1	-1	1			1.50	2.4375	3.6002	-2.10	5.10
A_1B_3	$-A_1B_1$	$L^{(19)}$	0	0	-2	-1	-2	-1			9.00	3.2500	4.1572	4.84	13.16
A_2B_1	$-A_2B_1$	$L^{(20)}$	0	0	0	0	0	0			0.00	0.0000	0.0000	0.00	0.00
A_2B_2	$-A_2B_1$	$L^{(21)}$	0	0	-1	1	1	-1			2.50	3.2500	4.1572	-1.66	6.66
A_2B_3	$-A_2B_1$	$L^{(22)}$	0	0	-2	-1	2	1			5.00	3.2500	4.1572	0.84	9.16

$l\theta$^=Mmult(l の範囲, θ^の範囲)　　$t_{0.05}$= T.Inv.2T(0.05, 8)= 2.3060

$Var(l\theta$^$)$=Mmult(Mmult(l の範囲, Σ (θ^) の範囲),Transpose(l の範囲))

◆GLMのEstimateステートメントの結果に一致

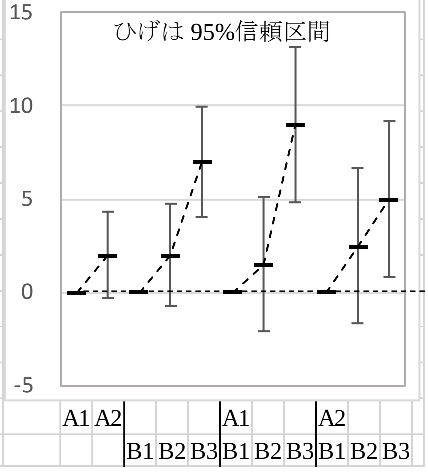
2022年11月18日 高橋行雄　　43

GLM プロシジャによる最小 2 乗平均の差

表 6 に estimate ステートメントを使って，A1，B1 を基準とした水準間の差，（A1 B1），（A2 B1）を基準とした組合せ水準の差と *SE* を求めた結果を示す．水準間の差について lsmeans ステートメントのオプション pdiff，tdiff などで出力することができる．ただし，*p* 値と *t* 値がマトリックス状に出力されるが，95%信頼区間，あるいは，*SE* の出力が無いために，ひげ付き線グラフが書けない．そこで，estimate ステートメントを用いて標準誤差 *SE* を得ることにした．Estimate ステートメントは，lsmeans ステートメントで推定できる最小 2 乗平均も含め，あらゆる推定ができるが，GLM プロシジャが用いているダミー変数についての知識が必要であり，ここでは，estimate ステートメントの使用法に立ち入らない．

表 6 Estimate ステートメントによる最小 2 乗平均の差に関する 95%信頼区間の線グラフ

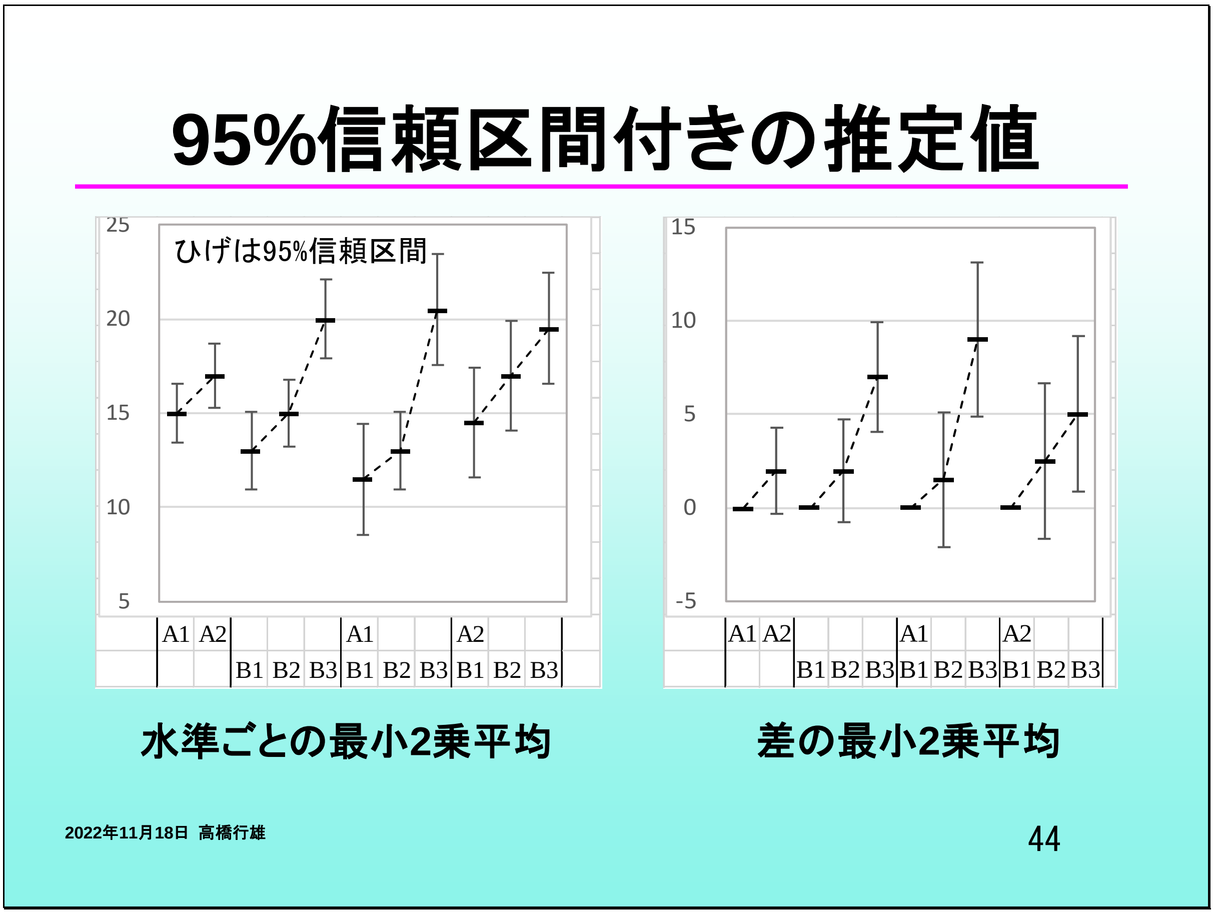
パラメータ	推定値	標準誤差	t 値	Pr > |t|	$t_{0.05}SE$	L95%	U95%
	0.00						
A2-A1	2.00	0.9965	2.01	0.0797	2.2980	-0.30	4.30
	0.00						
B2-B1	2.00	1.1924	1.68	0.1320	2.7497	-0.75	4.75
B3-B1	7.00	1.2748	5.49	0.0006	2.9396	4.06	9.94
	0.00						
A1*B2-A1*B1	1.50	1.5612	0.96	0.3648	3.6002	-2.10	5.10
A1*B3-A1*B1	9.00	1.8028	4.99	0.0011	4.1572	4.84	13.16
	0.00						
A2*B2-A2*B1	2.50	1.8028	1.39	0.2029	4.1572	-1.66	6.66
A2*B3-A2*B1	5.00	1.8028	2.77	0.0242	4.1572	0.84	9.16



95%信頼区間付きの推定値の折れ線グラフ

推定された結果は，必ず何らかのグラフ表示が，結果の理解，他者への説明のために必須である．最小 2 乗平均とその 95%信頼区間のグラフ表示は，スライド 44 に示すような Excel の折れ線グラフで描くのが生産的である．

スライド 44



Excel の折れ線グラフ作成のヒント

1）計算されたスライド 42 と 43 の 最小 2 乗平均 12 行を一括して折れ線グラフとする．

2）折れ線グラフの幅と Excel シートの列の幅を調整し，グラフの外側のセルに水準名を書き込む．

3）線を点線に変更し，マーカを横棒に変更し，切れ目のマーカを選択し，線を消す．

4）「グラフ要素」の「誤差範囲」➔「その他のオプション」➔「ユーザ設定」➔「値の設定」➔「正の誤差の値」･「負の誤差の値」に「$t_{0.05}\times SE$」の 12 行分を設定する．

5）外側の説明も含めてコピーし，ワードに「図（拡張メタファイル)」で貼り付ける．

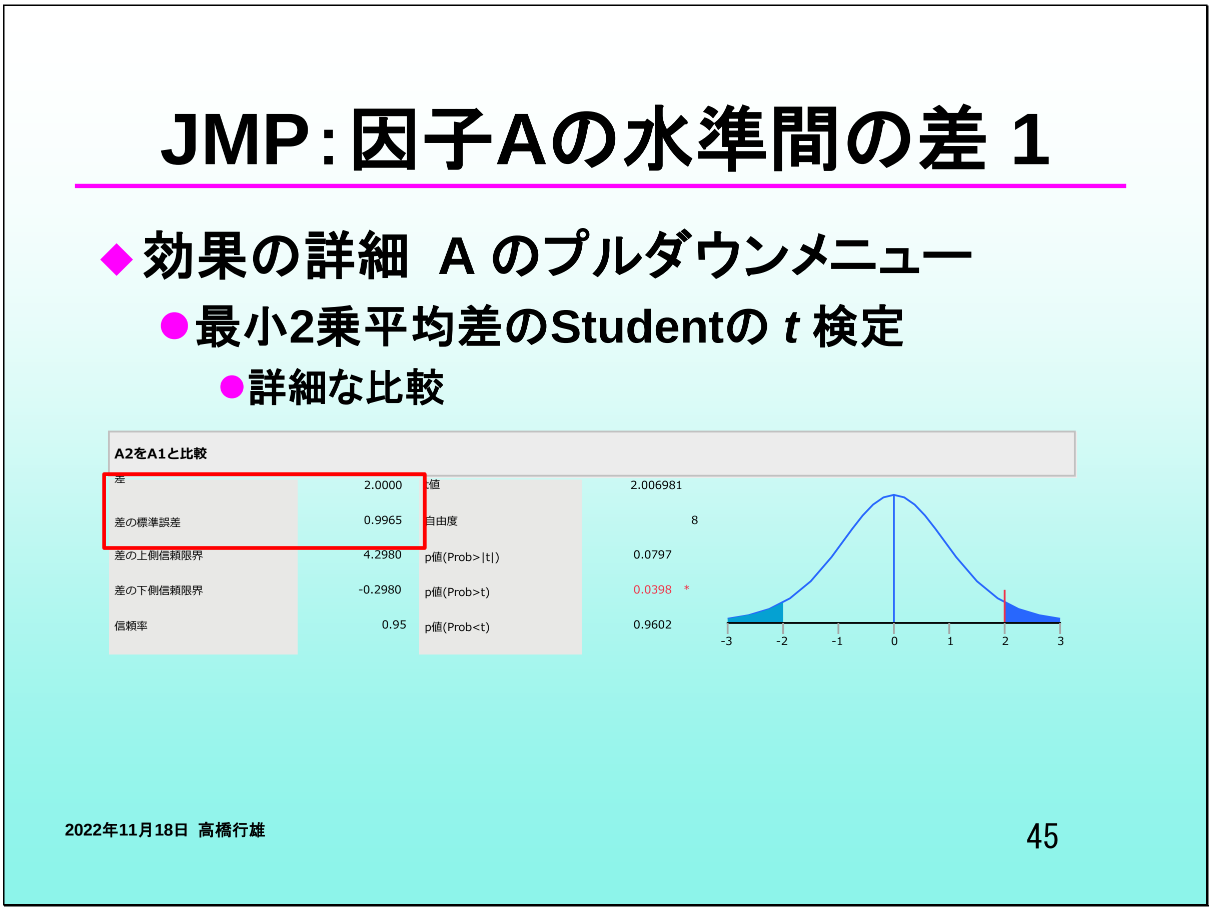
注意）Excel の折れ線グラフで 95%信頼区間の幅を正確に付けるヒントを示したが，Excel には，多様でいい加減な幅の設定が多数用意されており，途中で引っ掛からないように細心の注意が必要である．

因子 A の水準間の差の推定 1（Student の t 検定）

JMP の「モデルのあてはめ」で，スライド 44 に示したように，各水準をまとめて表示する機能はないので，スライド 37 に示したように各因子，交互作用ごとに折れ線グラフを描く必要がある．

厄介なのは，水準間の差である．スライド 45 に示したのは，「効果の詳細 A」のプルダウンメニューから「最小 2 乗平均差の Student の t 検定」を選択し，「詳細な比較」で出力した結果であるが，スライド 44 右 に示したような線グラフの作成は，現在のバージョン 16 ではサポートされていない．

スライド 45



因子 A の水準間の差の推定 2（対比の設定による場合）

スライド 45 と同様に「効果の詳細　A 」のプルダウンメニューの「最小 2 乗平均の対比」を用いても「Student の t 検定」と同じ結果が得られる．対比の設定は，最小 2 乗平均の差 $(\mathrm{A}_2-\mathrm{A}_1)$ に対応した符号となるように $\boxed{+}\,\boxed{-}$ のボックスを選択して設定する．差の推定値は 2.0，標準誤差は 0.9965 とスライド 45 と同じ結果が得られる．この対比の設定は，JMP の内部のデザイン行列 $\boldsymbol{X}$ を反映してる訳ではなく，ユーザが対応しやすいような GUI となっている．

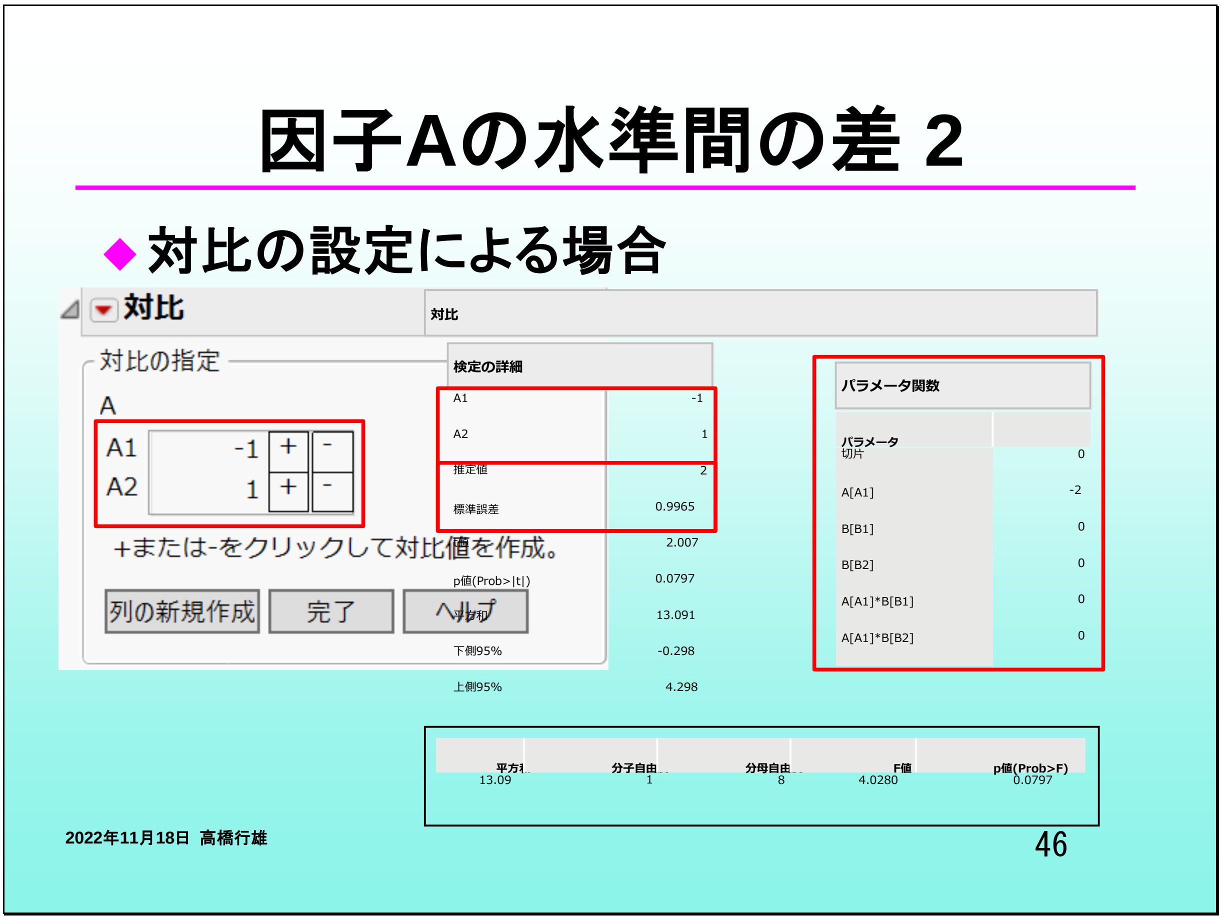
スライド 46 右に示す「パラメータ関数」が，内部で用いられている線形和のためのベクトルであり，スライド 43 に示した

$$(\mathrm{A}_2-\mathrm{A}_1)\ :\ \left\{\begin{aligned} L^{(2)}-L^{(1)} &= \boldsymbol{l}^{(2)}\hat{\boldsymbol{\theta}}-\boldsymbol{l}^{(1)}\hat{\boldsymbol{\theta}} \\ &= (\boldsymbol{l}^{(2)}-\boldsymbol{l}^{(1)})\hat{\boldsymbol{\theta}} \end{aligned}\right.$$

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

に対応している．

スライド 46



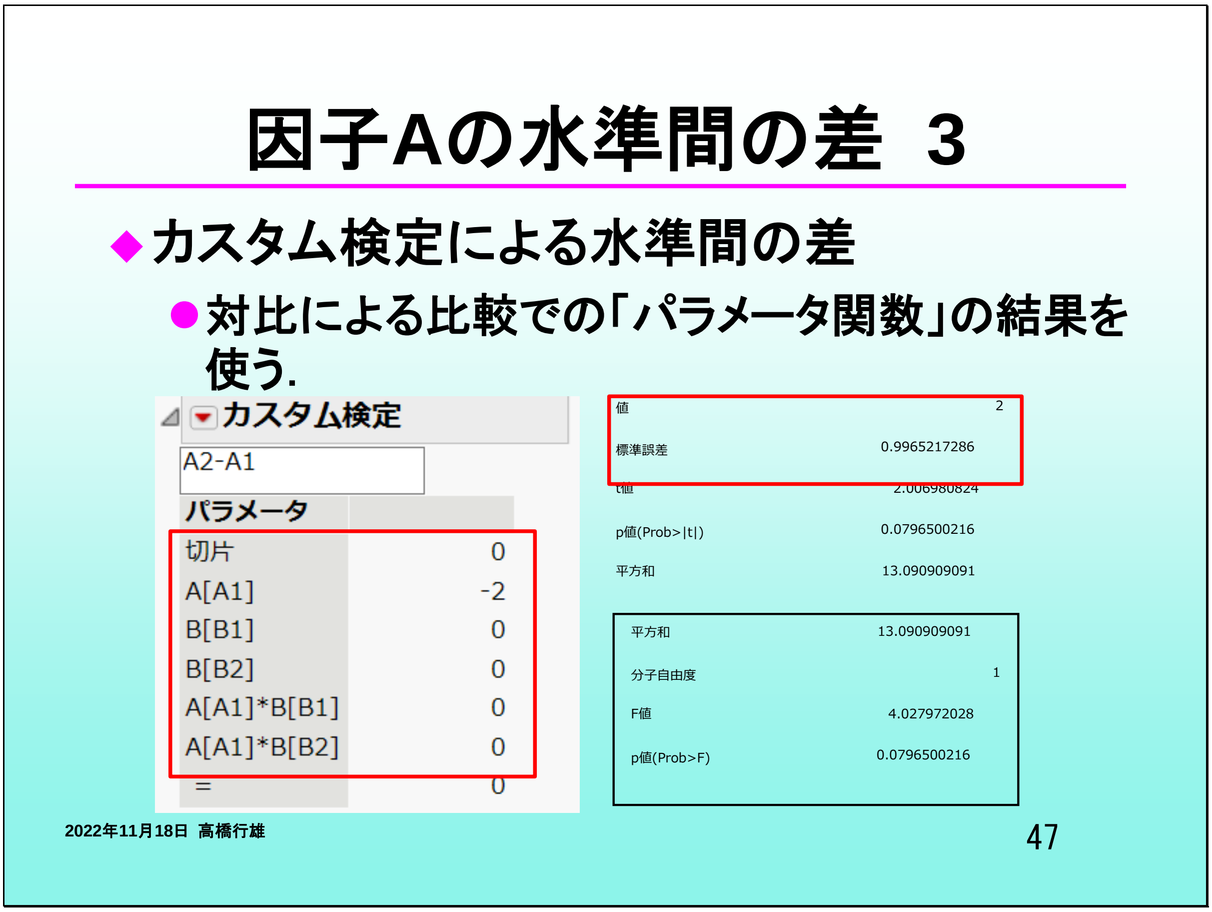
因子 A の水準間の差の推定 3（カスタム検定）

JMP の内部のデザイン行列 $\boldsymbol{X}$ に最も忠実なのは，スライド 47 に示した「カスタム検定」を使う方法である．スライド 43 に示した線形和のための係数ベクトルを直接入力する方法である．$(\mathrm{A}_2 - \mathrm{A}_1)$ を推定するために

$$(\mathrm{A}_2 - \mathrm{A}_1) = [0 \quad -2 \quad 0 \quad 0 \quad 0 \quad 0]$$

と入力するのであるが，どのような推定であろうとも自己責任ではあるが，係数ベクトルを設定すれば計算できる優れものである．

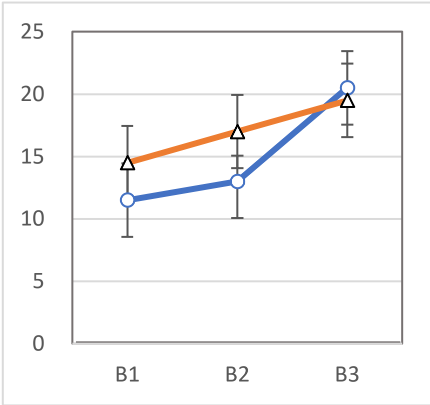
スライド 47



最小 2 乗平均の交互作用プロット

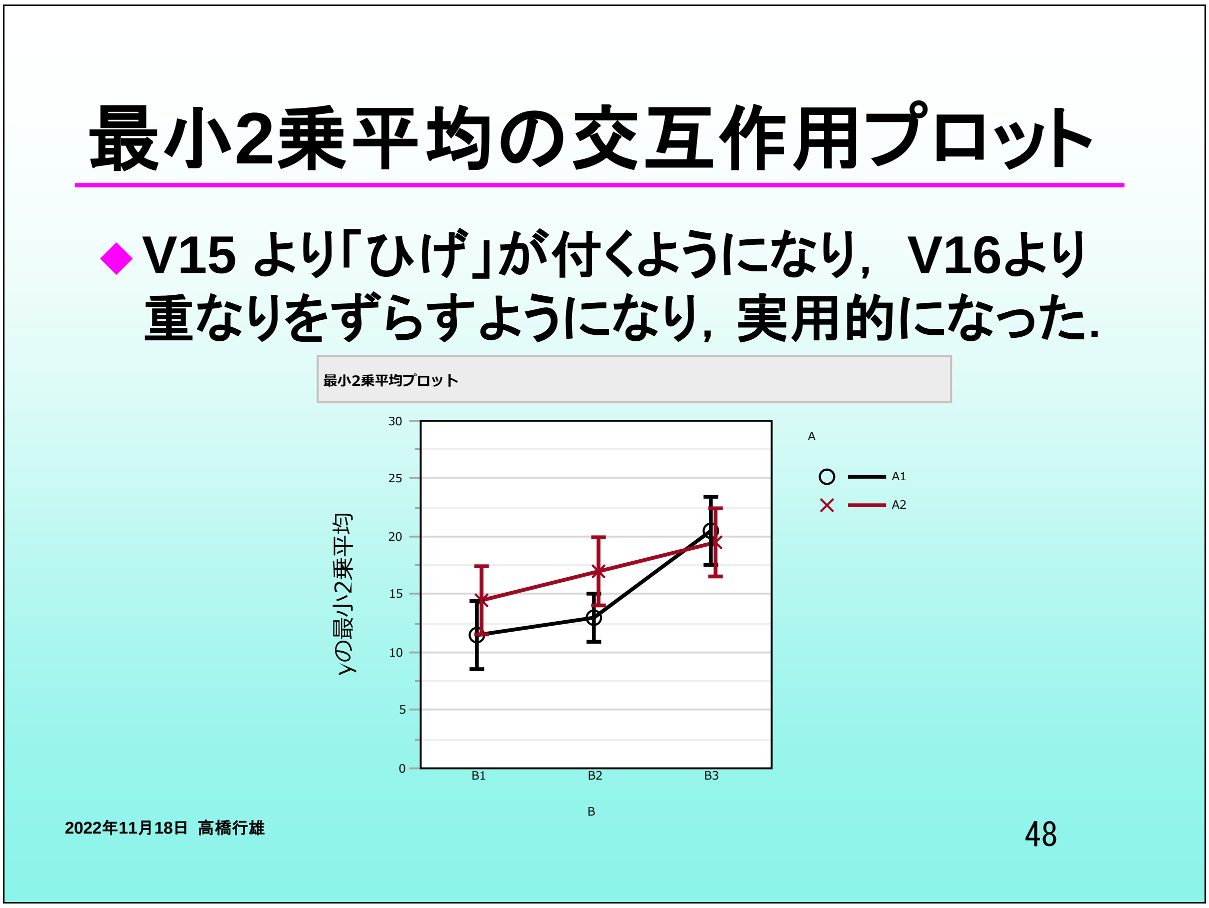
最小 2 乗平均の交互作用プロットは，Excel で簡単に描くことができないが，JMP の「モデルのあてはめ」で簡単に得られる折れ線グラフもある．スライド 48 に示した交互作用のプロットである．JMP バージョン 14 以前は，実用に耐えなかったのであるが，バージョン 15 より「ひげ」が付くようになり，さらにバージョン 16 より重なりをずらすようになり，実用的になった．

Excel の折れ線
グラフによる作図例



なお，Excel の折れ線グラフでは，JMP のバージョン 15 と同様な折れ線グラフを作成することはできるが，95%信頼区間のひげが重なりあい，ずらすことができない．Excel の散布図を使えば，プロット点をずらすことは容易であるが，ひげの作図をプロット点ごとに設定し直す必要があり，生産的でない．

スライド 48

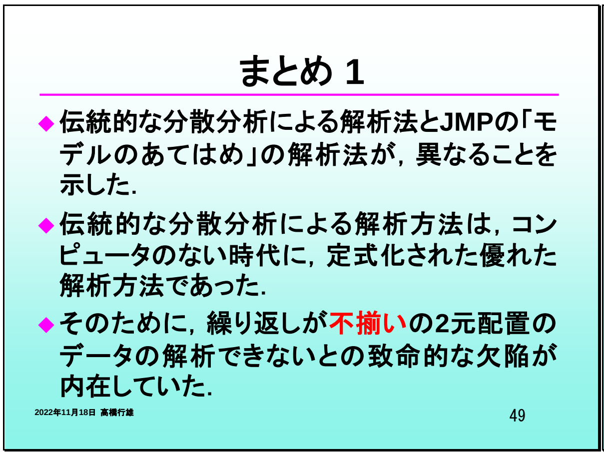


8. まとめ

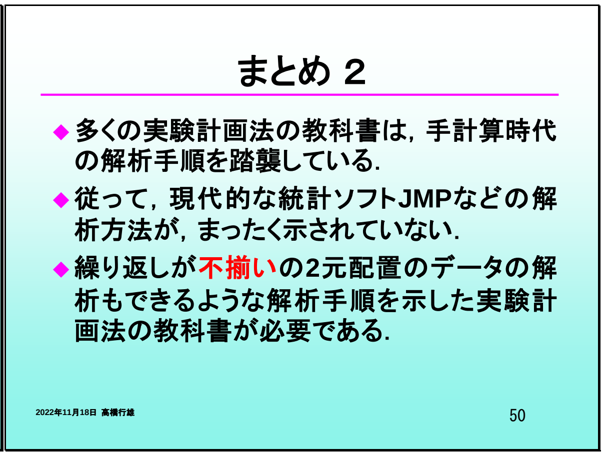
伝統的な分散分析による解析法と JMP の「モデルのあてはめ」の解析法が，異なることを示した．伝統的な分散分析による解析方法は，コンピュータのない時代に，定式化された優れた解析方法であった．そのために，繰り返しが不揃いの 2 元配置のデータの解析できないとの致命的な欠陥が内在していた．

多くの実験計画法の教科書は，手計算時代の解析手順を踏襲している．従って，現代的な統計ソフト JMP などの解析方法が，まったく示されていない．繰り返しが不揃いの 2 元配置のデータの解析もできるような解析手順を示した実験計画法の教科書が必要である．

スライド 49



スライド 50



デザイン行列 ***X*** を用いた分散分析の基礎については，講演時間の制約もあり割愛した．段階的な学習のためには，繰り返しなしの 2 元配置データについて，「構造モデル・回帰モデル・線形モデル」を最初に学習することが望ましい．とはいえ，推薦できる書籍は皆無に近い．唯一，楠正，辻谷将昭，松本哲夫，和田武夫（1995），「応用 実験計画法」の「第 6 章 線形モデル」に繰り返しなしの 2 元配置データを例にした線形モデルによる解析事例が示され，理論的な説明もしっかりしている．高橋（2022）「続・高橋セミナー 第 10 回　線形モデルによる欠測値がある直交表の解析 －謎めいた最小 2 乗平均と 95%信頼区間の活用－」の第 9 章以降に楠らが取り上げている例題について，Excel の行列計算による説明を付けて詳細に示している．

本セミナーでは，取り扱わなかった「構造モデル」，「デザイン行列を用いた正規方程式」，「デザイン行列を用いた正規方程式の解」，「組合せ平均の 95%信頼区間」，「現行水準と最適

水準の差および 95%信頼区間」,「有効反復数の功罪」,「田口の式（ルール）」,「伊奈の式（ルール）」などについて解説して入る．第 9 章の書き出しを示す.

―――――――――――― 引用 ――――――――――――

第 9 章　構造モデル･回帰モデル･線形モデル

繰り返しなしの 2 元配置における構造モデル

表 3 に示すのは，繰り返しのない 2 元配置データである．データの構造は，一般平均を μ とし，要因効果を示す母数 α_i，β_j および誤差 ε_{ij} を用いた 1 次式

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0,\ \sigma^2), \quad i = 1, 2, \ldots, n_{\mathrm{A}}, \quad j = 1, 2, \ldots, n_{\mathrm{B}} \qquad (1)$$

で表すことができる．このモデルを構造モデルまたは DE（Design of Experiment）モデルと言う．

表 3 繰返しのない 2 元配置データ［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

他方，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 が x_i の 1 次式

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0,\ \sigma^2), \quad i = 1, 2, \cdots, n \qquad (2)$$

となる．これは，回帰モデルと呼ばれている．式 (1), 式 (2) の共通点は,

1）データ構造が，式 (1) の母数 μ，α_i，β_j あるいは式 (2) の回帰係数 β_0，β_1 などの未知母数（パラメータ）と，確率変数である誤差の 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.「回帰モデル」を線形回帰モデルとも言うのであるが，取り上げる変数が連続尺度であることを前提にしている．名義尺度を連続尺度としてのダミー変数を主体にした場合に,「線形モデル」として区別して扱うことにする.

――― 以下 略 ―――

―――――――――――― 引用 終り――――――――――――

文 献

1）朝香鐵一, 石川馨, 山口襄 共同監修（1988）新版 品質管理便覧 第 2 版, 434-5, 914-5, 日本規格協会.

2）楠正, 辻谷将昭, 松本哲夫, 和田武夫（1995）, 応用 実験計画法, 185-95, 日科技連出版社.

3）SAS Institute（2013）, SAS/STAT 13.1 User’s Guide, The GLM Procedure, 3569-74, SAS Institute Inc., Cary, NC, USA.

4）SAS Institute（2020）, SAS/STAT 15.2 User’s Guide, The GLM Procedure, 4157-62, SAS Institute Inc., Cary, NC, USA. （自動翻訳付き）
https://documentation.sas.com/doc/en/statug/15.2/statug_glm_toc.htm .

5）高橋行雄, 大橋靖雄, 芳賀敏郎（1989）, SAS による実験データの解析, 289-333, 東大出版会.

6）高橋行雄（2020a）, 最尤法によるポアソン回帰分析入門, 第 4 章 デザイン行列を用いた回帰分析入門, https://www.yukms.com/biostat/takahasi2/rec/009-04.htm .

7）高橋行雄（2020b）, 最尤法によるポアソン回帰分析入門, 第 13 章 最小 2 乗平均の謎を予測プロファイルで解く, https://www.yukms.com/biostat/takahasi2/rec/009-13.htm .

8）高橋行雄（2021）, 最尤法によるポアソン回帰分析入門, 135-74, 383-420, カクワークス社.

9）高橋行雄（2022）, 線形モデルによる欠測値がある直交表の解析 －謎めいた最小 2 乗平均と 95%信頼区間の活用－, https://www.yukms.com/biostat/takahasi2/rec/010.htm .

10）ドレーパ, スミス著, 中村慶一訳（1968）, 応用回帰分析, 森北出版. 原書初版,

11）Draper N.R., Smith H.（1998）, Applide Regression Analisys 3th ed., Wiley.

12）McCullough B.D., David A. Heiser D.A. (2008), On the accuracy of statistical procedures in Microsoft Excel 2007, Computational Statistics and Data Analysis 52 (2008) 4570–4578.

13）Little R.C., Stroup W.W. and Freund R.j.（2002）, SAS for Linear Models, 4th ed., SAS Institute.

文 献 索 引

索 引

解析用ファイル

名前	サイズ
2元不揃い_02_データリスト_ダミー変数	35 KB
2元不揃い_03_2元不揃いSAS	1 KB
2元不揃い_03_平方和の計算	77 KB
2元不揃い_04_0_不揃い2元	2 KB
2元不揃い_04_1_コーディングのテーブル	2 KB
2元不揃い_04_2_交互作用	19 KB
2元不揃い_04_3_単回帰	4 KB
2元不揃い_04_4_2次式	4 KB
2元不揃い_04_5_2次式95%CL	31 KB
2元不揃い_05_タイプIIIの平方和	11 KB
2元不揃い_06_1_パラメータの推定値	8 KB
2元不揃い_06_2_推定値と標準誤差SE	31 KB
2元不揃い_07_1因子Aの効果	7 KB
2元不揃い_07_2_因子Aの水準間の差	12 KB
2元不揃い_07_3_最小2乗平均	84 KB

空白ページ

第 11 回 続高橋セミナー

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 －

BioStat 研究所（株）
〒105-0014 東京都 港区 芝 1‐12－3 の 1005
2023 年 3 月　高橋 行雄
takahashi.stat@nifty.com，FAX：03-342-8035