2021年12月16日
Discovery Summit Japan 2021
＠WEB

欠測値がある直交表の解析における線形モデルの活用

ー現行水準と最適水準の差の95%信頼区間の推定を例にしてー

高橋 行雄

BioStat 研究所（株）

目　次

1．はじめに

伝統的な実験計画法

- 伝統的な実験計画法は，直交表に限らず因子が互いに直交していることを前提にし，分散分析というシグマを用いた平方和の分解を主体にした解析法である．
- ただし，欠測値が発生した場合にこの方法が適用できないので，分散分析に対する理論的な枠組みを与えてきた「線形推定論」による解析を行う必要がある．
- ただし，「線形推定論」を解説している書物を見いだすことは困難である．

欠測値がある直交表

- JMPの「モデルのあてはめ」は，欠測値がある直交表の解析にも対応している.
- 直交表の解析の定番である「現行水準と最適水準の**差の95%信頼区間**の推定」を，どのようにして求めたら良いのであろうか.
- それらの誤差分散の推定のための「**田口の公式**」あるいは「**伊奈の公式**」は，JMPに組み込まれているのであろうか.

「モデルのあてはめ」の計算原理

- どのような解析手順によって計算されているかを解説した書物なしに，JMPの結果をそのまま信頼して使うことは勇気がいる．
- そこで，L_8 直交表で欠測値が1つある場合について，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．
- その結果として，「モデルのあてはめ」の計算原理について理解を深め，更なる応用ができるようになることを期待したい．

楠ら(1995)の解析事例

- 「線形推定論」は，実験データの解析法としての分散分析に対する理論的な枠組みを与えてきた.
- ただし，朝香ら(1988)，「新版　品質管理便覧　第2版」には，「線形推定論」を見いだすことができない.
- 幸いなことに，楠ら(1995)，「応用　実験計画法」の 第6章「線形モデル」，第2節に「線形推定論」が含まれている.
  - ［例題6.6］ 直交表で欠測値が生じた場合(p206-8)のデータを使を用い，事例紹介をする. .

直交表L_8への要因の割り付け

No.	A (1)	B (2)	(3)	C (4)	(5)	(6)	D (7)	データ y
1	1	1	1	1	1	1	1	32
2	1	1	1	2	2	2	2	23
3	1	2	2	1	1	2	2	19
4	1	2	2	2	2	1	1	17
5	2	1	2	1	2	1	2	19
6	2	1	2	2	1	2	1	20
7	2	2	1	1	2	2	1	欠測
8	2	2	1	2	1	1	2	8
成分	a	b	ab	c	ac	bc	abc	
交互作用			A×B		A×C	A×D		
			C×D		B×D	B×C		

２．JMPの「モデルのあてはめ」による解析

JMPのL_8データ

	No.	A	B	C	D	y
1	1	1	1	1	1	32
2	2	1	1	2	2	23
3	3	1	2	1	2	19
4	4	1	2	2	1	17
5	5	2	1	1	2	19
6	6	2	1	2	1	20
7	7	2	2	1	1	•
8	8	2	2	2	2	8

因子 A, B, C, D は，数値だが名義尺度

「モデルのあてはめ」の結果

分散分析

要因	自由度	平方和	平均平方	F値
モデル	4	303.0952	75.7738	34.9725
誤差	2	4.3333	2.1667	p値(Prob>F)
全体(修正済み)	6	307.4286		0.0280*

効果の検定

要因	自由度	平方和	平均平方	F値	p値(Prob>F)
A	1	77.0417	77.0417	35.5577	0.0270*
B	1	112.6667	112.6667	52.0000	0.0187*
C	1	28.1667	28.1667	13.0000	0.0691
D	1	22.0417	22.0417	10.1731	0.0858

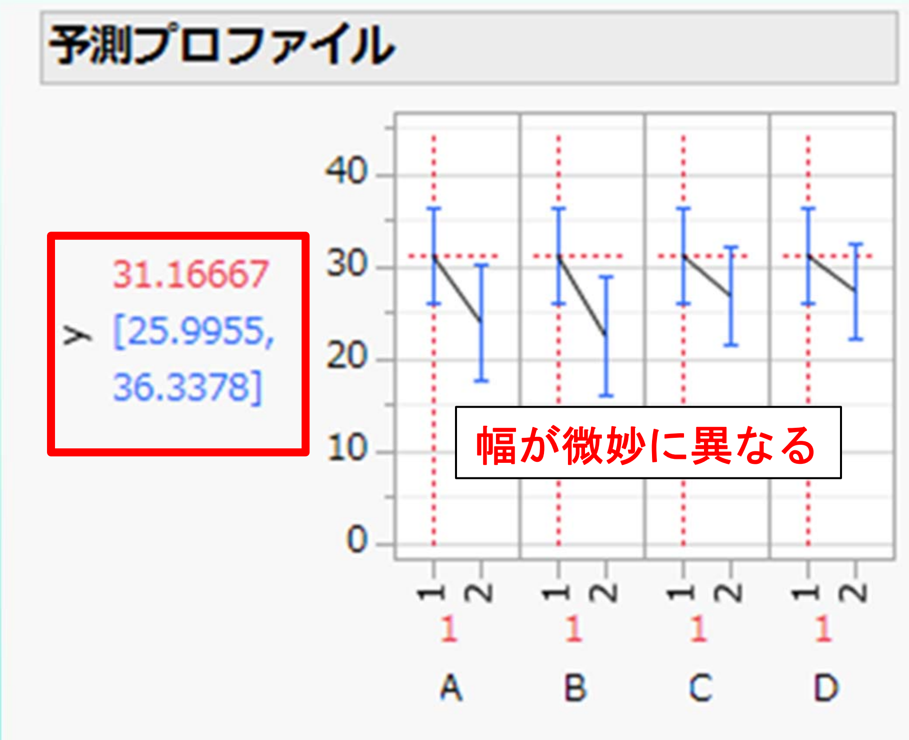
4因子の平方和の合計 77.0 + 112.7 + 28.2 + 22.0 = 239.9

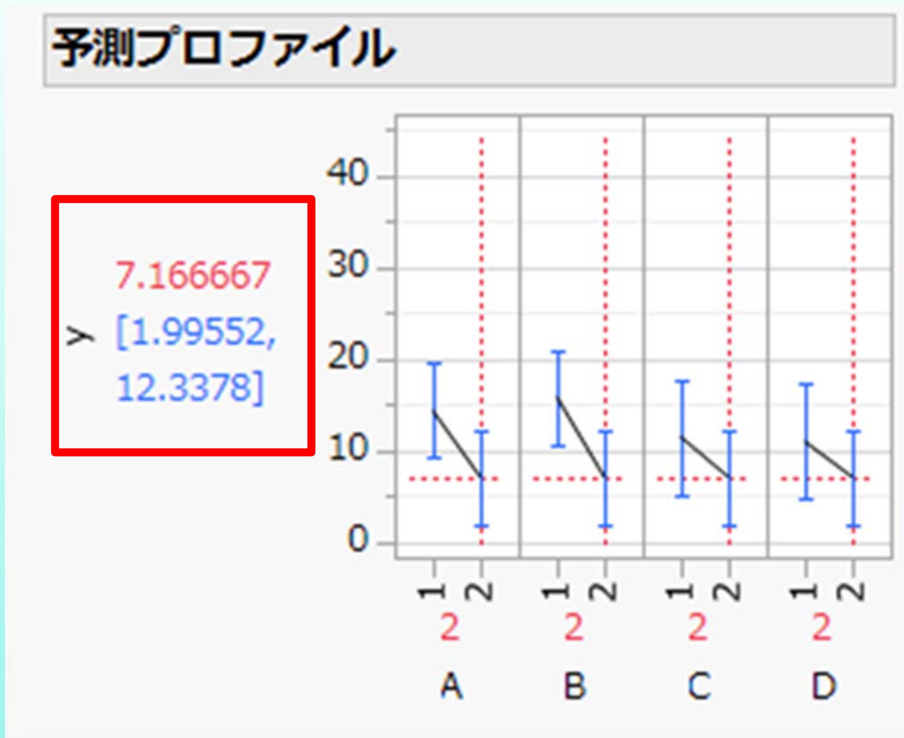
モデルの平方和 303.1 とは異なる.

平方和の分解が成り立たない !!

予測プロファイルを用いた推定

現行水準　　　　最適水準





（最適水準）と（現行水準）の 差の95%信頼区間は？

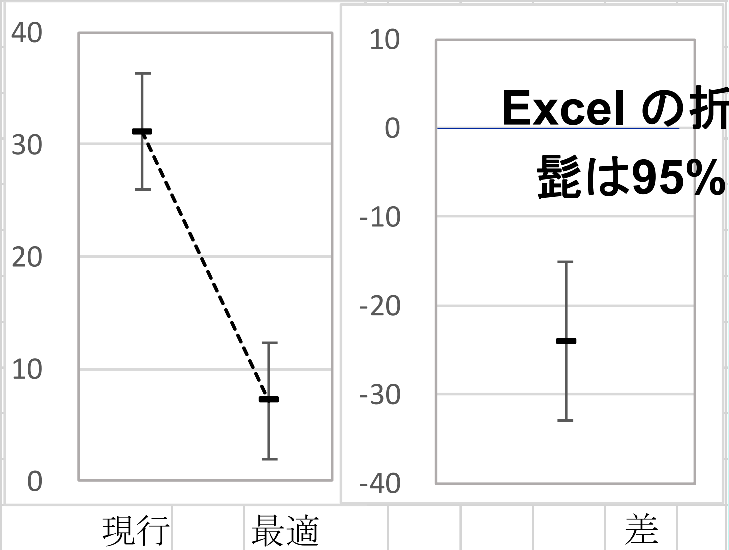
差 = 7.1667 – 31.1667 = -24.0

（最適水準）と（現行水準）の差

カスタム検定

最適－ 現行

パラメータ	
切片	0
A[1]	-2
B[1]	-2
C[1]	-2
D[1]	-2
＝	0
値	-24.0000
標準誤差	2.0817
t値	-11.5292
p値(Prob>|t|)	0.0074
平方和	288.0000



Excel の折れ線グラフ
髭は95%信頼区間

平方和	288.0000
分子自由度	1.0000
F値	132.9231
p値(Prob>F)	0.0074

カスタム検定 の －２ は，何を意味しているのか？
謎めいた数値，本当なのか？

それぞれの因子の水準平均

名義尺度の場合は，因子ごとに作表・作図

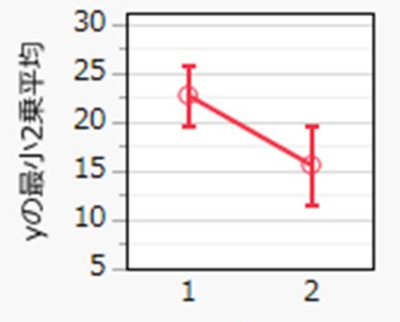
効果の詳細

A

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%	平均
1	22.7500	0.7360	19.5833	25.9167	22.7500
2	15.5833	0.9501	11.4952	19.6715	15.6667

最小2乗平均プロット

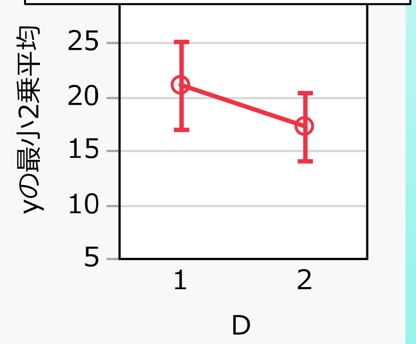


幅が微妙に異なる

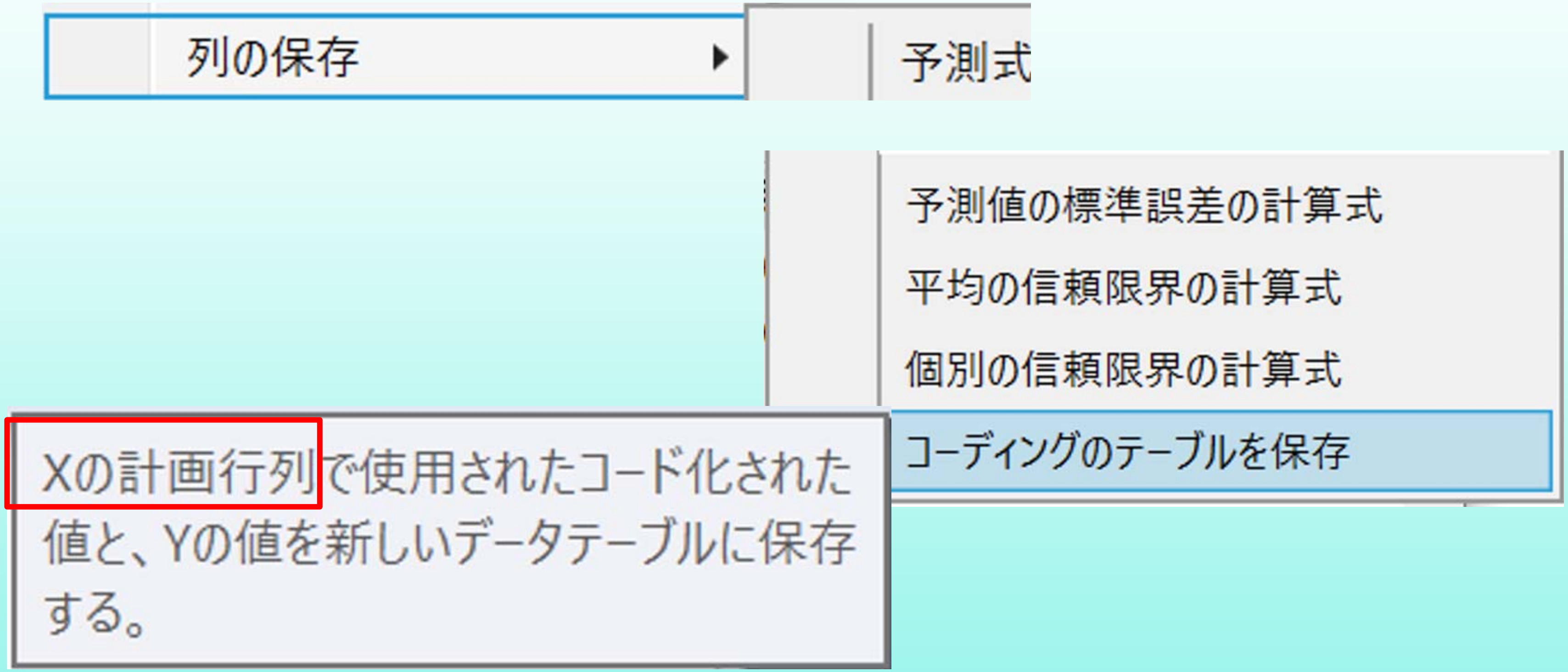
D

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%	平均
1	21.0833	0.9501	16.9952	25.1715	23.0000
2	17.2500	0.7360	14.0833	20.4167	17.2500



JMP：デザイン行列の出力



計画行列*X*, デザイン行列*X* は, 同義語

JMP:コーディングのテーブル

	切片	A[1]	B[1]	C[1]	D[1]	y
1	1	1	1	1	1	32
2	1	1	1	-1	-1	23
	1	1	-1	1	-1	19
	1	1	-1	-1	1	17
5	1	-1	1	1	-1	19
	1	-1	1	-1	1	20
	•	•	•	•	•	•
	1	-1	-1	-1	-1	8

この範囲が計画行列 X

「コーディングのテーブルの保存」により

JMPの内部の計算で用いられる デザイン行列 ***X*** が出力される.

３．デザイン行列 *X* の活用

名義尺度 を 連続尺度 に 変換

- 「モデルのあてはめ」では，名義尺度の変数を扱えるようになっている.
- ただし，JMPの内部の計算では，すべて連続変数に変換されている.
- 計算結果の表示に際しては，元の名義尺度にしユーザに気が付かないように配慮されている.
- 一般的に変換された連続変数は，ダミー変数と言われている.

デザイン行列(計画行列)の出力

元のJMPデータ

	No.	A	B	C	D	y
1	1	1	1	1	1	32
2	2	1	1	2	2	23
3	3	1	2	1	2	19
4	4	1	2	2	1	17
5	5	2	1	1	2	19
6	6	2	1	2	1	20
7	7	2	2	1	1	•
8	8	2	2	2	2	8

A, B, C, D は名義尺度

コーディングのテーブル

デザイン行列 X

	切片	A[1]	B[1]	C[1]	D[1]	y
1	1	1	1	1	1	32
2	1	1	1	-1	-1	23
3	1	1	-1	1	-1	19
4	1	1	-1	-1	1	17
5	1	-1	1	1	-1	19
6	1	-1	1	-1	1	20
7	•	•	•	•	•	•
8	1	-1	-1	-1	-1	8

切片, A[1], . . . ,D[1] は, 連続変数

3種類の L_4 直交表

（a） 基本型

No.	列番 (0)	(1)	(2)	(3)
1	+1	+1	+1	+1
2	+1	+1	−1	−1
3	+1	−1	+1	−1
4	+1	−1	−1	+1

（b） $L_4(2^3)$

No.	列番 (1)	(2)	(3)
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

（c） $H_{2\cdot 4}$

No.	列番 (1)	(2)	(3)
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0

- 朝香ら(1988), 新版 品質管理便覧 第2版, p446-7.
  - (a) 基本形
  - (b) + 1を1, -1 を2 におきかえた．実験計画に使いやすいように，田口[14)]・小西氏[15)]によって提唱された．
  - (c）+1を0, -1 を1 におきかえたもの．嶋田氏[16)]によって提唱された．

14) 田口玄一(1977,1962)：改訂新版実験計画法(上)，実験計画法(下)，丸善
15) 小西省三(1965)：例題演習実験計画法，日刊工業新聞社
16) 島田正三(1958)：やさしい直交配列の話，日本規格協会

各種のダミー変数

- 名義尺度の変数を含むデータに対し，回帰分析を行なう場合は，名義尺度を連続変数に変換する必要がある．
- **一般的**なのは，名義尺度が2水準の場合に，最初の水準を 0，次の水準を 1 とするような**（0，1）**型ダミー変数である．
- **JMP**では，（A_1 ➔ **1**， A_2 ➔ **-1**）のように足して 0 となるような **（1，-1）**対比型ダミー変数が使われている．これは，線形推定論での正統派である．
- ダミー変数の与え方は，**統計ソフトごとに異なる**．

デザイン行列 *X* を用いた解析

- 出力されたコーディングのテーブル（デザイン行列　X）を用いて「モデルのあてはめ」を行なう.
- 「切片」を除いて解析を行う.
  - 切片が出力されていることは，デザイン行列に対する理解を深めるために有益である.
- ダミー変数 A[1], B[1], C[1], D[1] は，連続変数のまま「モデルの効果」に設定する.

JMP：モデルのあてはめ

コーディングのテーブル（デザイン行列 *X*）での解析

分散分析

要因	自由度	平方和	平均平方	F値
モデル	4	303.0952	75.7738	34.9725
誤差	2	4.3333	2.1667	p値(Prob>F)
全体(修正済み)	6	307.4286		0.0280*

効果の検定

要因	自由度	平方和	平均平方	F値	Prob>F
A[1]	1	77.0417	77.0417	35.5577	0.0270*
B[1]	1	112.6667	112.6667	52.0000	0.0187*
C[1]	1	28.1667	28.1667	13.0000	0.0691
D[1]	1	22.0417	22.0417	10.1731	0.0858

因子名は，A が A[1] と異なるが，結果は(1, 2)型名義尺度の場合と完全一致

4変数に対する回帰パラメータの出力

パラメータ推定値

項	推定値	標準誤差	t値	Prob>|t|
切片	19.1667	0.6009	31.8953	0.0010*
A[1]	3.5833	0.6009	5.9630	0.0270*
B[1]	4.3333	0.6009	7.2111	0.0187*
C[1]	2.1667	0.6009	3.6056	0.0691
D[1]	1.9167	0.6009	3.1895	0.0858

推定値の相関

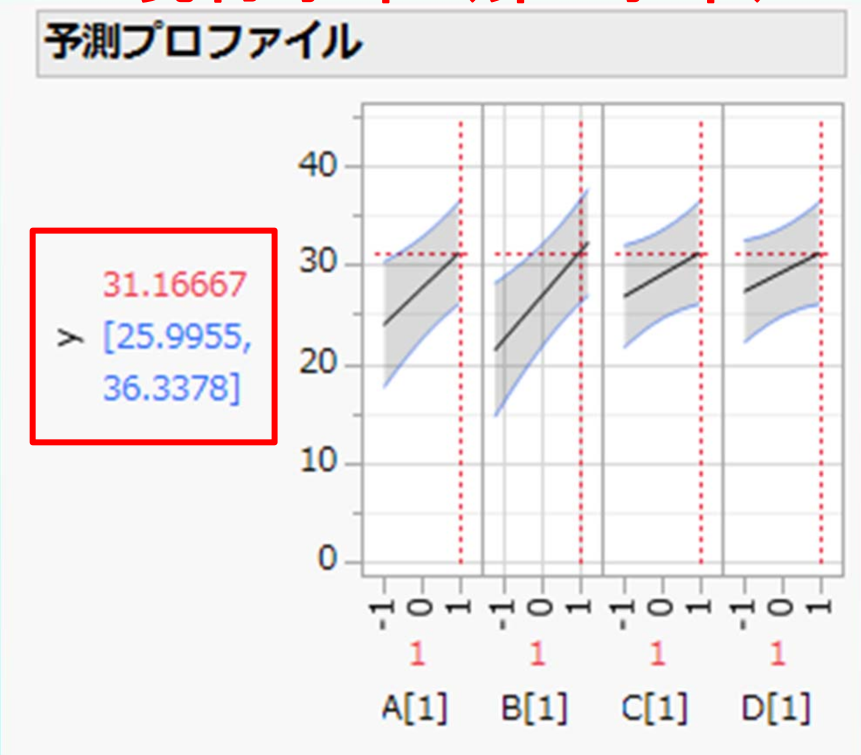
相関

	切片	A[1]	B[1]	C[1]	D[1]
切片	1.0000	-0.2500	-0.2500	0.2500	0.2500
A[1]	-0.2500	1.0000	0.2500	-0.2500	-0.2500
B[1]	-0.2500	0.2500	1.0000	-0.2500	-0.2500
C[1]	0.2500	-0.2500	-0.2500	1.0000	0.2500
D[1]	0.2500	-0.2500	-0.2500	0.2500	1.0000

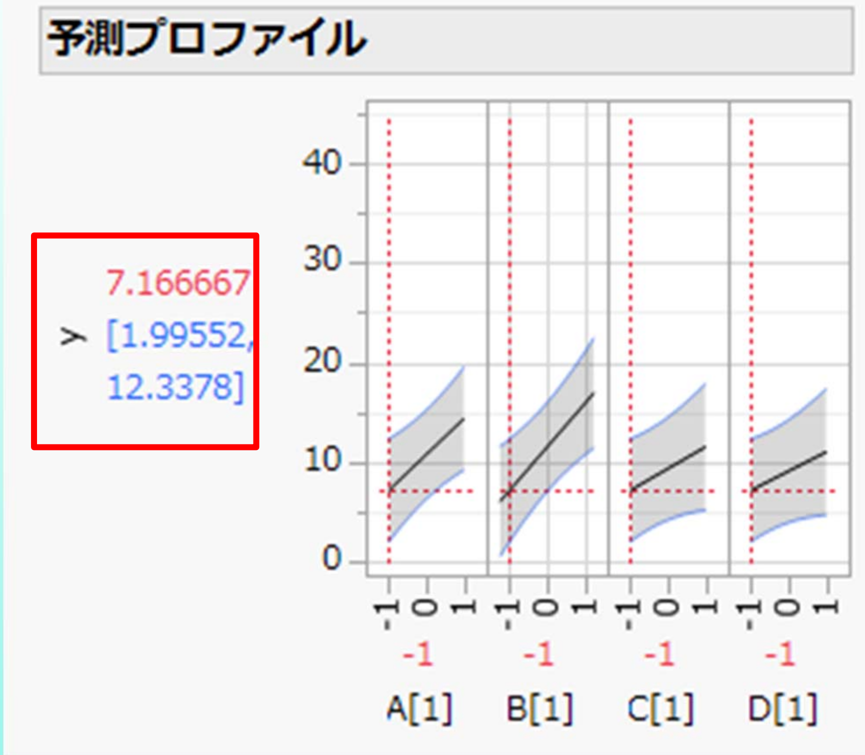
- 回帰パラメータの推定値を出力する.
- 推定値の相関行列も追加で出力する.

予測プロファイル

現行水準（第1水準）



最適水準（第2水準）



水準(1, 2)が(-1, 1)と反転する

(最適水準)と(現行水準)の差

カスタム検定

最適 - 現行	
パラメータ	
切片	0
A[1]	-2
B[1]	-2
C[1]	-2
D[1]	-2
=	0
値	-24.0000
標準誤差	2.0817
t値	-11.5292
p値(Prob>|t|)	0.0074
平方和	288.0000

最適 A_2= -1, ... , D_2= -1

―) 現行 A_1= 1, ... , D_1= 1

-2, ... , -2

名義尺度のままの解析も内部では，(1，-1)対比型ダミー変数に変換されているため，カスタム検定は，同じ設定となる．

平方和	288.0000
分子自由度	1.0000
F値	132.9231
p値(Prob>F)	0.0074

それぞれの因子の水準平均の算出

（1, -1）対比型ダミー変数の場合は，別途計算が必要

パラメータ推定値

項	推定値	標準誤差	t値	Prob>|t|
切片	19.1667	0.6009	31.8953	0.0010*
A[1]	3.5833	0.6009	5.9630	0.0270*
B[1]	4.3333	0.6009	7.2111	0.0187*
C[1]	2.1667	0.6009	3.6056	0.0691
D[1]	1.9167	0.6009	3.1895	0.0858

第1水準	第2水準
22.7500	15.5833
23.5000	14.8333
21.3333	17.0000
21.0833	17.2500

A_1 水準　切片+ A[1]，　A_2 水準　切片- A[1]

D_1 水準　切片+ D[1]，　D_2 水準　切片- D[1]

JMPの使勝手

- 因子を(1，2)型の名義尺度として解析することにより，欠測値のある直交表の解析が容易にできる.
- 最適水準，現行水準の推定と95%信頼区間の計算も予測プロファイルで GUI 的にできる.
- カスタム検定を用いて(最適水準と現行水準)の差の推定は，JMPの内部計算の仕組みの理解すればできるが，一般的には困難である.
- 要因ごとの水準平均と95%信頼区間のグラフは，個別の表示となり，まとめて表示できない.

４．Excel の回帰分析による線形モデル

Excelで欠測値のある直交表

Excel の「分析ツール」の「回帰分析」を用いて，基本のデザイン行列（1, -1）型に対する解析結果を示す．

	デザイン行列 ***X***									***Y***
	x_0	x_1	x_2	x_4	x_7					
No.	μ	A	B	C	D	$\boldsymbol{\theta}$		$\boldsymbol{\varepsilon}$		***y***
1	1	1	1	1	1	θ_0	+	ε_1	=	32
2	1	1	1	-1	-1	θ_1		ε_2		23
3	1	1	-1	1	-1	θ_2		ε_3		19
4	1	1	-1	-1	1	θ_3		ε_4		17
5	1	-1	1	1	-1	θ_4		ε_5		19
6	1	-1	1	-1	1					20
8	1	-1	-1	-1	-1					8
	μ	a	b	c	abc					

分散分析表				
	自由度	変動	分散	分散比
回帰	4	303.095	75.7738	34.9725
残差	2	4.333	**2.1667**	
合計	6	307.429		
	係数	標準誤差	t	P-値
切片	**19.1667**	0.6009	31.8953	0.0010
A	**3.5833**	0.6009	5.9630	**0.0270**
B	**4.3333**	0.6009	7.2111	**0.0187**
C	**2.1667**	0.6009	3.6056	0.0691
D	**1.9167**	0.6009	3.1895	0.0858

JMPのデザイン行列を用いた結果と完全に一致する．

分散分析表の計算

- Excel の回帰分析では，因子に対する分散分析表の出力がない．
- 自由度1の回帰パラメータの推定値のt値と分散分析表のF値には，$t^2=F$ の関係がある．
- F 値は，因子Aの平方和S_A，誤差分散を$\hat{\sigma}^2$としたときに，$F_A= (S_A/1)/\hat{\sigma}^2$ の関係からS_Aを

$$S_A = t_A^2 \hat{\sigma}^2$$

によって計算し，分散分析表を作成する．

欠測値のある場合の分散分析表

◆ **平方和の分解**で対応できない分散分析表は，Excel の回帰分析の結果に**若干の計算**を付け加えるだけで作成できる.

欠測値がある直交表に対する分散分析表								平方和の計算	
	自由度		変動	分散	分散比				t^2σ^2
要因	*df*		平方和 *S*	平均平法 *V*	*F*	*p* 値	*t* 値	σ^2	平方和 *S*
モデル	4		303.0952	75.7738	34.9725				
A		1	77.0417	77.0417	35.5577	**0.0270**	5.9630	2.1667	77.0417
B		1	112.6667	112.6667	52.0000	**0.0187**	7.2111	2.1667	112.6667
C		1	28.1667	28.1667	13.0000	0.0691	3.6056	2.1667	28.1667
D		1	22.0417	22.0417	10.1731	0.0858	3.1895	2.1667	22.0417
		小計	239.9167						
残差	2		4.3333	**2.1667**					
合計	6		307.4286						

（1, 2）型直交表

◆ （1，2）型のまま，連続変数とした場合のExcelで回帰分析を行なったらどうなるのか．

	デザイン行列 $\boldsymbol{X}$							$\boldsymbol{Y}$
	x_0	x_1	x_2	x_4	x_7			
No.	μ	A	B	C	D	$\boldsymbol{\theta}$	$\boldsymbol{\varepsilon}$	$\boldsymbol{y}$
1	1	1	1	1	1	θ_0	ε_1	32
2	1	1	1	2	2	θ_1	ε_2	23
3	1	1	2	1	2	θ_2	ε_3	19
4	1	1	2	2	1	θ_3	ε_4	17
5	1	2	1	1	2	θ_4	ε_5	19
6	1	2	1	2	1			20
8	1	2	2	2	2			8
	μ	a	b	c	abc			

（行列式：$\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \boldsymbol{y}$）

分散分析表

	自由度	変動	分散	分散比
回帰	4	303.095	75.7738	34.9725
残差	2	4.333	**2.1667**	
合計	6	307.429		

	係数	標準誤差	t	P-値
切片	**55.1667**	3.1798	17.3491	0.0033
A	**-7.1667**	1.2019	-5.9630	**0.0270**
B	**-8.6667**	1.2019	-7.2111	**0.0187**
C	**-4.3333**	1.2019	-3.6056	0.0691
D	**-3.8333**	1.2019	-3.1895	0.0858

◆ 分散分析表の結果は同じ，係数は，まったく異なるが，切片を除いてp値は同じ．

(1, -1)型, (1, 2)型, (0, 1)型

- どのようなダミー変数を与えても，分散分析表は同一となる.
- ただし，パラメータの推定値は，異なる.
- 特に切片の推定値は，大きく異なり，*p*値も異なる.
  - (1, -1)型の場合にのみ切片が$\hat{\mu}$の推定値となる.
- 各因子の推定値は，ダミー変数間の差に比例して変化する.
  - (1, 2)型，(0, 1)型のようにダミー変数の差が1の場合は，水準間の差の推定値になる.

(1, 2)型の場合の水準平均

(1, 2)型		μ^の推定		第1水準		第2水準	
	係数			数値	推定値	数値	推定値
切片	**55.1667**	1.0	55.1667				
A	**-7.1667**	1.5	-10.7500	-0.5	22.7500	0.5	15.5833
B	**-8.6667**	1.5	-13.0000	-0.5	23.5000	0.5	14.8333
C	**-4.3333**	1.5	-6.5000	-0.5	21.3333	0.5	17.0000
D	**-3.8333**	1.5	-5.7500	-0.5	21.0833	0.5	17.2500
		μ^	19.1667				

- ***μ*^は，(1+2)/2=1.5 の場合の推定値を計算する．**
- **各水準の推定値は， *μ*^に推定値の半分をマイナス・プラスする．**
- **ダミー変数は，推定したい目的に合わせて選択することが望ましい．**

５. 各種の推定値に対する 95%信頼区間の算出

従来の分散分析の場合

- 平方和の分解により，分散分析表を完成させ，誤差分散 **σ^2** を求める.
- 各水準の推定値に対しては，その推定値が幾つの**データ数**から求められたかを**勘案**してσ^2から分散を手計算で求め，95%信頼区間を計算する.
- 最適水準，現行水準に対しては，推定に考慮した因子の水準数から**田口の式**により**有効反復数**を求め分散を計算し，95%信頼区間を計算する.
- 最適水準と現行水準の**差**については，**伊奈の式**により，有効反復数を計算する.

線形モデルの場合

- デザイン行列 X，反応変数 Y を用いて推定値 $\hat{\theta}$ を求め，誤差分散 $\hat{\sigma}^2$ を計算する．
- デザイン行列の積　$(X^T X)$ の逆行列 $(X^T X)^{-1}$ の要素を $c_{kk'}$ とする．
- 推定値 $\hat{\theta}_k$ の分散 $Var(\hat{\theta}_k)$ は，$c_{kk}\hat{\sigma}^2$ である．
- 推定値　$\hat{\theta}_k$ と $\hat{\theta}_{k'}$ の共分散　$Cov(\hat{\theta}_k, \hat{\theta}_{k'})$ は，$c_{kk'}\hat{\sigma}^2$ である．
- これらの分散共分散を行列としてまとめ，パラメータの共分散行列を $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$ とする．

パラメータの共分散行列 $\Sigma(\theta\hat{})$

転置行列 X^T

	1	2	3	4	5	6	8
μ	1	1	1	1	1	1	1
A	1	1	1	1	-1	-1	-1
B	1	1	-1	-1	1	1	-1
C	1	-1	1	-1	1	-1	-1
D	1	-1	-1	1	-1	1	-1

デザイン行列 X

μ	A	B	C	D
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	-1	1
1	-1	-1	-1	-1

デザイン行列の積

= $X^T X$ $(a_{kk'})$

7	1	1	-1	-1
1	7	-1	1	1
1	-1	7	1	1
-1	1	1	7	-1
-1	1	1	-1	7

デザイン行列の積の逆行列

$(X^T X)^{-1}$ $(c_{kk'})$

$\theta\hat{}_0$	0.167	-0.042	-0.042	0.042	0.042
$\theta\hat{}_1$	-0.042	0.167	0.042	-0.042	-0.042
$\theta\hat{}_2$	-0.042	0.042	0.167	-0.042	-0.042
$\theta\hat{}_3$	0.042	-0.042	-0.042	0.167	0.042
$\theta\hat{}_4$	0.042	-0.042	-0.042	0.042	0.167

$\sigma\hat{}\hat{}2$=2.1667

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

0.3611	-0.0903	-0.0903	0.0903	0.0903
-0.0903	**0.3611**	0.0903	-0.0903	-0.0903
-0.0903	0.0903	**0.3611**	-0.0903	-0.0903
0.0903	-0.0903	-0.0903	**0.3611**	0.0903
0.0903	-0.0903	-0.0903	0.0903	**0.3611**

=Minverse(Mmult(Transpose(Xの範囲),Xの範囲))*$\sigma\hat{}^2$

データの相関$R(x)$・共分散行列$\Sigma(x)$

デザイン行列 X

μ	A	B	C	D
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	-1	1
1	-1	-1	-1	-1

	μ	A	B	C	D
平均 $\bar{\boldsymbol{x}}$=	1.0000	0.1429	0.1429	-0.1429	-0.1429
共分散行列$\boldsymbol{\Sigma}(\boldsymbol{x})$=	0	0	0	0	0
(データの)	0	**1.14286**	-0.1905	0.19048	0.19048
$(\boldsymbol{X}-\bar{\boldsymbol{x}})^T(\boldsymbol{X}-\bar{\boldsymbol{x}})$	0	-0.1905	**1.14286**	0.19048	0.19048
	0	0.19048	0.19048	**1.14286**	-0.1905
	0	0.19048	0.19048	-0.1905	**1.14286**
個別の分散$\boldsymbol{\sigma}^2$=	0	1.14286	1.14286	1.14286	1.14286
	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
相関行列$\boldsymbol{R}(\boldsymbol{x})$=	#DIV/0!	**1**	-0.1667	0.16667	0.16667
$\boldsymbol{\Sigma}(\boldsymbol{x})/\boldsymbol{\sigma}/\boldsymbol{\sigma}^T$	#DIV/0!	-0.1667	**1**	0.16667	0.16667
	#DIV/0!	0.16667	0.16667	**1**	-0.1667
	#DIV/0!	0.16667	0.16667	-0.1667	**1**

パラメータの相関$R(\theta\hat{})$・共分散行列$\Sigma(\theta\hat{})$

JMPの「モデルのあてはめ」では，パラメータ（推定値）の共分散行列は出力されないので推定値の相関から

$\Sigma(\theta\hat{}) = R(\theta\hat{}) * SE * SE^T$

で計算する．

推定値の相関

相関

	切片	A[1]	B[1]	C[1]	D[1]
切片	1.0000	-0.250	-0.250	0.2500	0.2500
A[1]	-0.250	1.0000	0.2500	-0.250	-0.250
B[1]	-0.250	0.2500	1.0000	-0.250	-0.250
C[1]	0.2500	-0.250	-0.250	1.0000	0.2500
D[1]	0.2500	-0.250	-0.250	0.2500	1.0000

パラメータの共分散行列 $\Sigma(\theta\hat{})$

$\theta\hat{}_0$	**0.3611**	-0.0903	-0.0903	0.0903	0.0903
$\theta\hat{}_1$	-0.0903	**0.3611**	0.0903	-0.0903	-0.0903
$\theta\hat{}_2$	-0.0903	0.0903	**0.3611**	-0.0903	-0.0903
$\theta\hat{}_3$	0.0903	-0.0903	-0.0903	**0.3611**	0.0903
$\theta\hat{}_4$	0.0903	-0.0903	-0.0903	0.0903	**0.3611**

$SE=$	0.6009	0.6009	0.6009	0.6009	0.6009

パラメータの相関行列 $R(\theta\hat{})$

$\theta\hat{}_0$	**1**	-0.2500	-0.2500	0.2500	0.2500
$\theta\hat{}_1$	-0.2500	**1**	0.2500	-0.2500	-0.2500
$\theta\hat{}_2$	-0.2500	0.2500	**1**	-0.2500	-0.2500
$\theta\hat{}_3$	0.2500	-0.2500	-0.2500	**1**	0.2500
$\theta\hat{}_4$	0.2500	-0.2500	-0.2500	0.2500	**1**

推定値の *SE* の計算

	係数	標準誤差	t	P-値
切片	**19.1667**	0.6009	31.8953	0.0010
A	**3.5833**	0.6009	5.9630	**0.0270**
B	**4.3333**	0.6009	7.2111	**0.0187**
C	**2.1667**	0.6009	3.6056	0.0691
D	**1.9167**	0.6009	3.1895	0.0858

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

0.3611	-0.0903	-0.0903	0.0903	0.0903
-0.0903	**0.3611**	0.0903	-0.0903	-0.0903
-0.0903	0.0903	**0.3611**	-0.0903	-0.0903
0.0903	-0.0903	-0.0903	**0.3611**	0.0903
0.0903	-0.0903	-0.0903	0.0903	**0.3611**

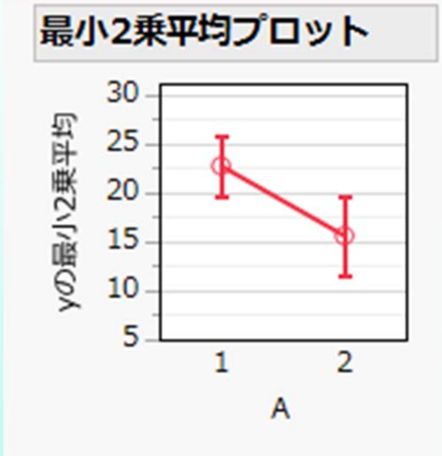
=Minverse(Mmult(Transpose(***X***の範囲),***X***の範囲))*$\sigma\hat{}^{2}$

- 推定値の**標準誤差*SE***は，パラメータの共分散行列**$\Sigma(\theta\hat{})$**対角要素**の平方根**で計算されている.
- 伝統的な分散分析法では，回帰パラメータの推定は行なわれないが，線形モデルでは，**基本中の基本である.**

線形和の例，因子Aの水準平均

最小2乗平均表

水準	最小2乗平均	標準誤差	下側95%	上側95%	平均
1	22.7500	0.7360	19.5833	25.9167	22.7500
2	15.5833	0.9501	11.4952	19.6715	15.6667



						線形和の例	
因子			係数	標準誤差		A_1水準	A_2水準
切片	$\theta\hat{}_0$	$\mu\hat{}$	**19.1667**	0.6009		1	1
A	$\theta\hat{}_1$	$\alpha\hat{}$	**3.5833**	0.6009		1	-1
B	$\theta\hat{}_2$	$\beta\hat{}$	**4.3333**	0.6009		0	0
C	$\theta\hat{}_3$	$\gamma\hat{}$	**2.1667**	0.6009		0	0
D	$\theta\hat{}_4$	$\delta\hat{}$	**1.9167**	0.6009		0	0
					積和=	22.7500	15.5833

線形和*L*の 分散・*SE*

- 推定値 $\theta_k^と\theta_{k'}$^の共分散 $Cov(\theta_k\text{^}, \theta_{k'}\text{^})$は，$c_{kk'}\sigma$^2 であることは，すでに示した．
- 推定されたパラメータθ_k^に，ある数値 l_k を掛けた積和を 線形和 $L=\Sigma_k l_k\theta_k$^とする．
- 線形和 *L* の分散は，

  $Var(L)= \Sigma_k\Sigma_{k'}\, l_k\, l_{k'}\, c_{kk'}\sigma$^2 である．
- 手計算が困難なことから統計関連の書物では，実例が極めて乏しく，忘れ去られたがごとくである．
- Excel の行列関数を使って容易に計算できる．

Excel による線形和Lの分散の計算

◆ A_1の分散 $Var(L) = \boldsymbol{l}\,\Sigma(\hat{\theta})\,\boldsymbol{l}^T = \Sigma_k \Sigma_{k'} l_k l_{k'} c_{kk'} \sigma$^2

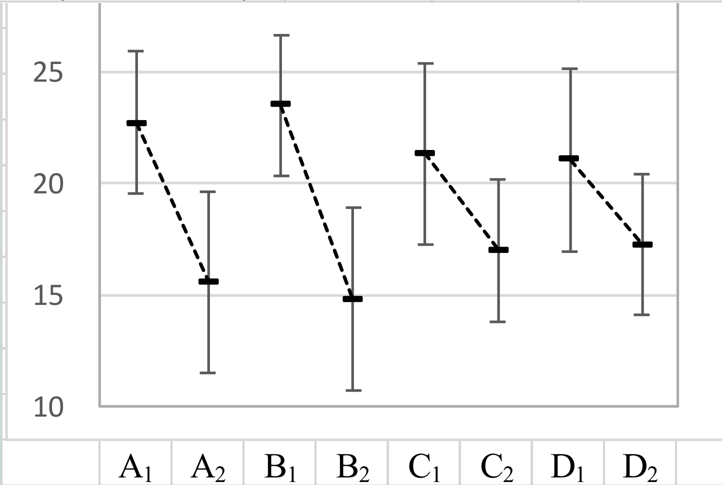
		l_0	l_1	l_2	l_3	l_4	パラメータの共分散行列 $\Sigma(\theta$^)					$\boldsymbol{l}^{(1)T}$	
A_1の分散	$\boldsymbol{l}^{(1)}$	1	1	0	0	0	**0.361**	-0.090	-0.090	0.090	0.090	1	
							-0.090	**0.361**	0.090	-0.090	-0.090	1	
							-0.090	0.090	**0.361**	-0.090	-0.090	0	
							0.090	-0.090	-0.090	**0.361**	0.090	0	
							0.090	-0.090	-0.090	0.090	**0.361**	0	
													Var^$(L$^$^{(1)})$
						=	0.2708	0.2708	0.0000	0.0000	0.0000	1	= 0.5417
												1	
												0	
Var^$(L$^$^{(1)})$ =Mmult(Mmult($\boldsymbol{l}$ の範囲, $\Sigma(\theta$^) の範囲), Transpose($\boldsymbol{l}^T$ の範囲))												0	
												0	

◆ Excel の行列関数を使えば簡単である.

◆ 任意の線形和の分散も，目の前シート上で計算.

各因子の水準平均と95%信頼区間

					l_0	l_1	l_2	l_3	l_4		y^				
L	A	B	C	D	μ	α	β	γ	δ	θ^	L^	Var (L^)	t×SE	L95%	U95%
$L^{(1)}$	A_1				1	1	0	0	0	19.167	22.750	0.5417	3.167	19.583	25.917
$L^{(2)}$	A_2				1	-1	0	0	0	3.583	15.583	0.9028	4.088	11.495	19.671
$L^{(3)}$		B_1			1	0	1	0	0	4.333	23.500	0.5417	3.167	20.333	26.667
$L^{(4)}$		B_2			1	0	-1	0	0	2.167	14.833	0.9028	4.088	10.745	18.921
$L^{(5)}$			C_1		1	0	0	1	0	1.917	21.333	0.9028	4.088	17.245	25.421
$L^{(6)}$			C_2		1	0	0	-1	0		17.000	0.5417	3.167	13.833	20.167
$L^{(7)}$				D_1	1	0	0	0	1		21.083	0.9028	4.088	16.995	25.171
$L^{(8)}$				D_2	1	0	0	0	-1		17.250	0.5417	3.167	14.083	20.417
												t(0.05, 6-3-1)=	4.303		

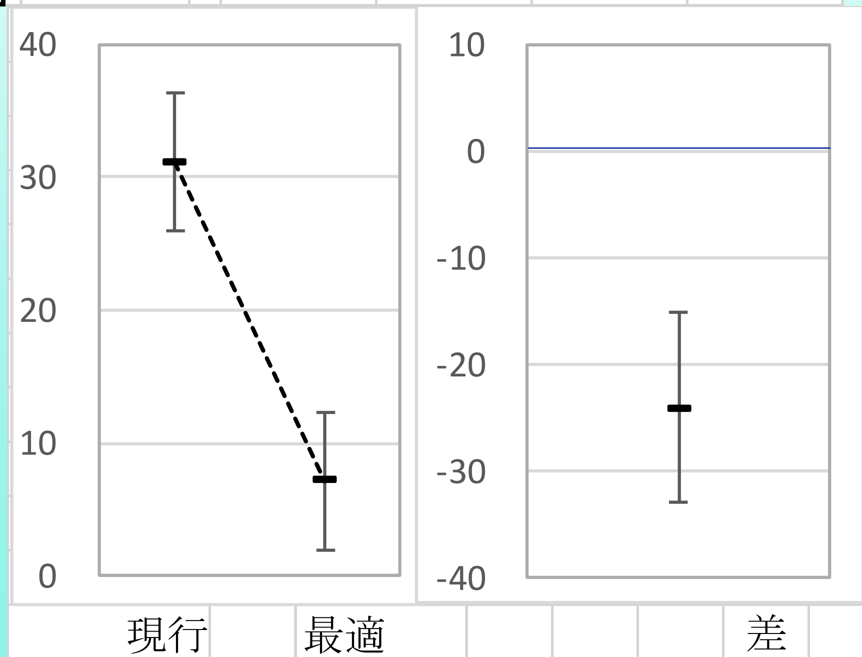


作図は，Excel の折れ線グラフにtxSEの幅を付け，因子間の点線を消去する．見栄えは上々である．

現行・最適水準・差の推定

					l_0	l_1	l_2	l_3	l_4		y^				
L	A	B	C	D	μ	α	β	γ	δ	θ^	L^	Var (L^)	t×SE	L95%	U95%
$L^{(9)}$	A_1	B_1	C_1	D_1	1	1	1	1	1	19.167	31.167	1.4444	5.171	25.996	36.338
$L^{(10)}$	A_2	B_2	C_2	D_2	1	-1	-1	-1	-1	3.583	7.167	1.4444	5.171	1.996	12.338
$L^{(11)}$	現行−最適				0	-2	-2	-2	-2	4.333	-24.000	4.3333	8.957	-32.957	-15.043
										2.167		t(0.05, 6-3-1)=	4.303		
										1.917					

線形和のための係数 l_kを変えれば，現行水準，最適水準，その差 など，自在な推定が行える．その分散から，95%信頼区間の計算は，フィルハンドル操作による計算式のコピペで即座に計算できる．



5．まとめ

伝統的な実験計画法からの脱却

- JMPの「モデルのあてはめ」は，最初から伝統的な実験計画法から脱却している.
- 多くの「実験計画法」に関する書物は，アリ地獄的な平方和の分解に頼る分散分析から抜け出すことができない.
- JMPの「モデルのあてはめ」は，平方和の分解ではなく，線形モデルに基盤をおいている.
- また，「実験計画法」における交互作用は，重要な概念であり，JMP「モデルのあてはめ」で自由に設定できるようになっている.

「モデルのあてはめ」の深い理解

- JMPの「モデルのあてはめ」は，使用者に線形モデルをオブラートにくるんだがごとく感じさせない．
- 伝統的な実験計画法では，対応できない欠測値を含む直交表の解析に対し，JMPの「モデルのあてはめ」を使えば線形モデルの有用性を浮き彫りにできると考えた．
- 「現行水準と最適水準の差の推定と95%信頼区間」におけるカスタム検定においてようやく線形モデルがにじみでてきたことは，喜ばしい．

Excel の回帰分析 +α +β

- Excel の回帰分析は，線形モデルの入門の第一歩として適している.
- 各種の線形和と95%信頼区間の推定を行なうために必要なパラメータの共分散行列がExcel の行列関数によって容易にできることは，+α の機能として重要である.
- Excel の散布図，あるいは，折れ線グラフによる作図機能も +β の機能として多大な寄与が期待できることも例示した.

出版に向けて

今回の発表は,「**層別因子を含む探索的な回帰分析入門**」と題する執筆中の本の第4章の推敲するためであり, 4.5節・4.4節の順でスライドを作成した.

文 献

- 楠正，辻谷将昭，松本哲夫，和田武夫（1995），「応用 実験計画法」，p206-8，日科技連出版社.
- 朝香鐵一，石川馨，山口襄 共同監修（1988） 新版 品質管理便覧 第2版，446-7，日本規格協会.
- 高橋行雄，大橋靖男，芳賀敏郎（1989），SASによる実験データの解析，東大出版会.
  - 第15章 4種の平方和とLSMEAN
  - 第16章 GLMプロシジャの計算方式
- 高橋行雄（2021），最尤法によるポアソン回帰入門，カクワークス社.
  - 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
  - 第12章 パラメータの共分散分析の活用
    https://www.yukms.com/biostat/takahasi2/rec/009-12.htm

ご清聴ありがとうございました