

## 第 10 回 続高橋セミナー

# 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―

2022 年 4 月

高橋 行雄 BioStat 研究所(株)

**要約：** 各種の実験デザインに対する平方和の分解をベースとする解析法に対し、その理論的な側面については、線形推定・検定論によって説明されてきた。線形推定・検定論は、各種の実験デザインに対しデザイン行列を設定し、正規方程式を立て、逆行列を使った解析方法を基礎としている。ただし、手計算時代において行列計算を前提にした解析法は、現実の問題に適用することは不可能であり、理論の学習には役立つが実用性に欠けていた。計画した実験データに欠測値が出てしまい伝統的な平方和の分解による解析ができない場合に、線形推定・検定論に基づく解析法があるよとの“守り神”的な存在であった。そこで、Excel の行列計算を用いて「層別因子を含む探索的な回帰分析」の基盤ともいえる伝統的な線形推定・検定論を回顧しつつ Excel の行列関数を用いてリニューアルを計る。

## 目 次

1. はじめに -----1
2. 線形モデルを用いた直交表の解析 -----4  
直交表を用いた実験データの解析、直交表における構造 (DE) モデルと線形モデル、  
直交表のデザイン行列との関連、直交表による実験で欠測値が生じた場合
3. 欠測値がある直交表に対する JMP による解析 -----8  
予測プロファイルを用いた推定、最小 2 乗平均
4. デザイン行列  $X$  の活用 ----- 14  
各種のダミー変数の歴史的背景、デザイン行列を用いたパラメータの推定、  
現行水準と最適水準の組合せ平均、最小 2 乗平均と算術平均の比較、  
JMP の使い勝手
5. SAS/GLM による欠測値がある直交表の解析 ----- 23  
SAS のスクリプト、GLM プロシジャの出力、Estimate ステートメントによる最小 2 乗平均

次ページに続く

|     |  |    |
|-----|--|----|
| 6.  | Excel の回帰分析を用いた線形モデル -----   | 27 |
|     | 正規方程式の Excel シート上での表現, 欠測値がある場合の分散分析表,<br>各種のダミー変数をそのまま用いた線形モデル  |    |
| 7.  | 各種の推定値に対する 95%信頼区間の算出 -----  | 36 |
|     | 従来の分散分析の場合, パラメータの共分散行列,<br>データの相関行列および共分散行列, パラメータの相関行列から共分散行列の算出,<br>パラメータの推定値の標準誤差 $SE$ , 線形和で求めた推定値の 95%信頼区間,<br>SAS の GLM プロシジャでの (1, 0) 型ダミー変数               |    |
| 8.  | 伝統的な実験計画法からの脱却 -----   | 48 |
| 9.  | 構造モデル・回帰モデル・線形モデル -----  | 52 |
|     | 繰り返しの 2 元配置における構造モデル, 線形モデル,<br>正規方程式, デザイン行列を用いた正規方程式,<br>デザイン行列を用いた正規方程式の解, パラメータの共分散行列の活用,<br>組合せ平均の 95%信頼区間, 因子の水準間の差の推定および 95%信頼区間,<br>現行水準と最適水準の差および 95%信頼区間 |    |
| 10. | 伝統的な平方和の分解による解析 -----  | 68 |
|     | 構造モデルから平方和の分解へ, 水準平均の分散,<br>組合せ平均の推定値, 現行水準と最適水準の差, 田口の式(ルール),<br>伊奈の式(ルール), 「平方和の分解」ではなく「線形モデル」による解析の優位性  |    |
| 11. | 平方和の分解から線形モデルへ -----   | 76 |
|     | 絶滅の危機にある平方和の分解, Excel による平方和の分解,<br>有効反復数の功罪, 線形モデルの有用性, Excel の逆行列関数が救いの神,<br>アリ地獄的な状況から抜け出すために   |    |
|     | 文 献 -----  | 79 |
|     | 文 献 索 引 -----  | 80 |
|     | 索 引 -----  | 80 |
|     | 解析用ファイル -----  | 83 |

## 1. はじめに

2021年12月16日にオンラインで開催された JMP の Discovery Summit Japan 2021 にて、「線形モデルによる欠測値がある直交表の解析—現行水準と最適水準の差の 95%信頼区間の推定を例にして—」と題して講演した。この講演は、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 4 章の原稿を推敲するために行った。JMP の Discovery Summit Japan 2021 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 10 回目とする。

スライド 1

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

—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—

高橋 行雄  
BioStat 研究所(株)

2021年12月16日 高橋行雄 1

スライド 2

| 目 次 |                             |
|-----|-----------------------------|
| 1   | はじめに 3                      |
| 2   | JMPの「モデルのあてはめ」による解析 9       |
| 3   | デザイン行列 X の活用 17             |
| 4   | Excel の回帰分析による線形モデル 29      |
| 5   | 各種の推定値に対する<br>95%信頼区間の算出 36 |
| 6   | まとめ 48                      |
|     | 文献 53                       |

2021年12月16日 高橋行雄 2

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平方和の分解に基づく解析方法が適用できないので、分散分析に対する理論的な

スライド 3

---

## 1. はじめに

2021年12月16日 高橋行雄 3

スライド 4

### 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。

2021年12月16日 高橋行雄 4

枠組みを与えてきた「線形推定論」に基づく解析を行う必要がある。ただし、「線形推定論」を丁寧に解説している書物を見いだすことは困難である。そこで、欠測値がある直交表の解析に引き続き、第9章で繰り返しのない $3 \times 2$ の2元配置データを例にし、「構造モデル・回帰モデル・線形モデル」の基礎をExcelの行列関数を用いて詳しく解説する。また、第10章で伝統的な平方和の分解による分散分析対し、Excelの種々の関数を用いた新たな解析法も合わせて示す。

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「現行水準の組合せと最適水準の組合せの差の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「田口の式」あるいは「伊奈の式」は、JMPに組み込まれているのであろうか。

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、結果の正しさについて裏付けが取れない不安が残る。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、JMPの「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ている。

#### スライド5

### 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。

2021年12月16日 高橋行雄 5

#### スライド6

### 「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。

2021年12月16日 高橋行雄 6

注)「現行水準」は、「現行水準の組合せ」を省略して用いている。「最適水準」についても同様に省略して用いている。これらは、工程平均とも言われている。

「線形推定論」は、実験データの解析法としての分散分析に対する理論的な枠組みを与えてきた。ただし、朝香・石川・山口 共同監修 (1988)、「新版 品質管理便覧 第2版」には、「線形推定論」を見いだすことができない。幸いなことに、楠・辻谷・松本・和田 (1995)、「

「応用 実験計画法」の第 6 章「線形モデル」の第 2 節に「線形推定論」が含まれており，  
 [例題 6.6 直交表で欠測値が生じた場合] に丁寧な解説がなされている．ただし，手計算による解析方法なので，他の事例に応用することは困難である．

伝統的な線形推定論では，各種の実験デザインで取り上げる因子をダミー変数とする考え方がきちっと導入され，デザイン行列，正規方程式，逆行列などをベースにした線形モデル入門となっている．各種の実験モデルに対する解析は，線形モデルを拡張することにより，すべて対応できることを高橋・大橋・芳賀(1989)，「SAS による実験データの解析」で示した．ただし，この当時は，表計算ソフトが出始めたころで，線形モデルによる解析法の例示は，SAS/IML という行列計算言語に頼っていたこともあり，その後の発展に結びつかなかった．

あらためて，ダミー変数の数理およびそれらを活用したデザイン行列を活用した解析方法について Excel の行列関数を用い，実用的に対処できるようにリニューアルを図る．スライド 8 に示すのは， $L_8$  直交表に (A, B, C, D) の 4 因子を割り付けて実験をしたが，不幸なことに No.7 の  $A_2B_2C_1D_1$  においてデータが得られなかった場合である．JMP および Excel を用いた解析法について詳しく示す．

スライド 7

### 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．

7

スライド 8

### 直交表 $L_8$ への要因の割り付け

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

8

表 2 で再掲

## 2. 線形モデルを用いた直交表の解析

### 直交表を用いた実験データの解析

直交表を用いた実験データの解析は、手計算による平方和の分解に基づく方法が確立している。平方和の分解による計算の結果は、分散分析表にまとめられ、取り上げた多くの因子および交互作用の効果を一括して統計的に評価できる優れたものである。ただし、2水準あるいは3水準の質的な変数に限定されており、量的な変数をそのまま扱うことができないので、2または3水準を設定して質的な変数として扱うことになる。

扱い易い2水準の直交表に一部の因子を3水準したい場合に、「組合せ法」として、直交表の2列を用いて (1, 1) → 1, (1, 2) → 2, (2, 1) → 3, (2, 2) → 3 のように4水準を3水準に縮小するような割り付けテクニックを用いて平方和を計算するなど、様々な割り付けに対応した平方和の計算なども整備され、実験計画法と言え、直交表と言われるほど普及し、多くの実験計画法に関する書籍によりより解説されている。

直交表に限らず伝統的な実験計画による解析法は、シグマに基づく平方和の分解を主体にしている。直交表を用いた実験データに線形モデルを適用し解析する方法は、要因をすべて量的な変数として取り扱うことになり、まったく普及していない。標準的に用いられている

(1, 2) 型の2水準系の直交表も元々は、(1, -1) 対比型ダミー変数の線形モデルのデザイン行列であったが、手作業での扱いが容易になるように「-1」を「2」に置き換えたものである。スライド20に示すように朝香ら(1988)、「新版 品質管理便覧 第2版」の「第7.1.4.2 節 直交表の性質と利用」に詳しく述べられている。

### 直交表における構造(DE)モデルと線形モデル

2水準系の  $L_8$  直交表のルーツは、すべての因子が2水準の  $2 \times 2 \times 2 = 8$  の3元配置モデルである。因子を A, B, C としたときに、構造 (DE) モデルは、次式で与えられる。

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk} \quad (1)$$

|           |         |              |             |              |                        |                         |                        |                       |
|-----------|---------|--------------|-------------|--------------|------------------------|-------------------------|------------------------|-----------------------|
| $y_{111}$ | $= \mu$ | $+ \alpha_1$ | $+ \beta_1$ | $+ \gamma_1$ | $+ (\alpha\beta)_{11}$ | $+ (\alpha\gamma)_{11}$ | $+ (\beta\gamma)_{11}$ | $+ \varepsilon_{111}$ |
| $y_{112}$ | $= \mu$ | $+ \alpha_1$ | $+ \beta_1$ | $+ \gamma_2$ | $+ (\alpha\beta)_{11}$ | $+ (\alpha\gamma)_{12}$ | $+ (\beta\gamma)_{12}$ | $+ \varepsilon_{112}$ |
| $y_{121}$ | $= \mu$ | $+ \alpha_1$ | $+ \beta_2$ | $+ \gamma_1$ | $+ (\alpha\beta)_{12}$ | $+ (\alpha\gamma)_{11}$ | $+ (\beta\gamma)_{21}$ | $+ \varepsilon_{121}$ |
| $y_{122}$ | $= \mu$ | $+ \alpha_1$ | $+ \beta_2$ | $+ \gamma_2$ | $+ (\alpha\beta)_{12}$ | $+ (\alpha\gamma)_{12}$ | $+ (\beta\gamma)_{22}$ | $+ \varepsilon_{122}$ |
| $y_{211}$ | $= \mu$ | $+ \alpha_2$ | $+ \beta_1$ | $+ \gamma_1$ | $+ (\alpha\beta)_{21}$ | $+ (\alpha\gamma)_{21}$ | $+ (\beta\gamma)_{11}$ | $+ \varepsilon_{211}$ |
| $y_{212}$ | $= \mu$ | $+ \alpha_2$ | $+ \beta_1$ | $+ \gamma_2$ | $+ (\alpha\beta)_{21}$ | $+ (\alpha\gamma)_{22}$ | $+ (\beta\gamma)_{12}$ | $+ \varepsilon_{212}$ |
| $y_{221}$ | $= \mu$ | $+ \alpha_2$ | $+ \beta_2$ | $+ \gamma_1$ | $+ (\alpha\beta)_{22}$ | $+ (\alpha\gamma)_{21}$ | $+ (\beta\gamma)_{21}$ | $+ \varepsilon_{221}$ |
| $y_{222}$ | $= \mu$ | $+ \alpha_2$ | $+ \beta_2$ | $+ \gamma_2$ | $+ (\alpha\beta)_{22}$ | $+ (\alpha\gamma)_{22}$ | $+ (\beta\gamma)_{22}$ | $+ \varepsilon_{222}$ |

$\left. \vphantom{\begin{matrix} y_{111} \\ y_{112} \\ y_{121} \\ y_{122} \\ y_{211} \\ y_{212} \\ y_{221} \\ y_{222} \end{matrix}} \right\} \quad (2)$

式 (1) を線形モデルとして表すと

$$y_{i'} = \theta_0 x_{0,i'} + \theta_1 x_{1,i'} + \theta_2 x_{2,i'} + \cdots + \theta_p x_{p,i'} + \varepsilon_{i'}, \quad i' = 1, 2, \dots, 8 \quad (3)$$

となり，母数（パラメータ） $\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1, \gamma_2, \dots$  を  $\theta_0, \theta_1, \theta_2, \dots$  に対応付け，その変数を  $x_{0,i'}, x_{1,i'}, x_{2,i'}, \dots$  とすることにより線形モデルとなる．変数の数  $p$  は， $p = 3 \times 2 + 3 \times 4 = 18$  となるが，変数の間に

$$x_{1,i'} + x_{2,i'} = 1, \quad x_{3,i'} + x_{4,i'} = 1, \dots \quad (4)$$

などの関係（1次従属）がある．また，構造（DE）モデルにおける母数に関する制約条件

$$\left. \begin{aligned} \alpha_1 + \alpha_2 &= 0, & \beta_1 + \beta_2 &= 0, & \gamma_1 + \gamma_2 &= 0, \\ (\alpha\beta)_{11} + (\alpha\beta)_{12} + (\alpha\beta)_{21} + (\alpha\beta)_{22} &= 0, \\ (\alpha\gamma)_{11} + \cdots \end{aligned} \right\} \quad (5)$$

などを用い

$$\left. \begin{aligned} \alpha_1 &= \alpha, & \alpha_2 &= -\alpha \\ \beta_1 &= \beta, & \beta_2 &= -\beta \\ \gamma_1 &= \gamma, & \gamma_2 &= -\gamma \\ (\alpha\beta)_{11} &= (\alpha\beta)_{22} = \alpha\beta, & (\alpha\beta)_{12} &= (\alpha\beta)_{21} = -\alpha\beta \\ (\alpha\gamma)_{11} &= (\alpha\gamma)_{22} = \alpha\gamma, & \cdots \\ & \cdots \end{aligned} \right\} \quad (6)$$

とする．式 (6) を式 (2) に代入すると

|           |         |            |           |            |                 |                  |                 |                       |
|-----------|---------|------------|-----------|------------|-----------------|------------------|-----------------|-----------------------|
| $y_{111}$ | $= \mu$ | $+ \alpha$ | $+ \beta$ | $+ \gamma$ | $+ \alpha\beta$ | $+ \alpha\gamma$ | $+ \beta\gamma$ | $+ \varepsilon_{111}$ |
| $y_{112}$ | $= \mu$ | $+ \alpha$ | $+ \beta$ | $- \gamma$ | $+ \alpha\beta$ | $- \alpha\gamma$ | $- \beta\gamma$ | $+ \varepsilon_{112}$ |
| $y_{121}$ | $= \mu$ | $+ \alpha$ | $- \beta$ | $+ \gamma$ | $- \alpha\beta$ | $+ \alpha\gamma$ | $- \beta\gamma$ | $+ \varepsilon_{121}$ |
| $y_{122}$ | $= \mu$ | $+ \alpha$ | $- \beta$ | $- \gamma$ | $- \alpha\beta$ | $- \alpha\gamma$ | $+ \beta\gamma$ | $+ \varepsilon_{122}$ |
| $y_{211}$ | $= \mu$ | $- \alpha$ | $+ \beta$ | $+ \gamma$ | $- \alpha\beta$ | $- \alpha\gamma$ | $+ \beta\gamma$ | $+ \varepsilon_{211}$ |
| $y_{212}$ | $= \mu$ | $- \alpha$ | $+ \beta$ | $- \gamma$ | $- \alpha\beta$ | $+ \alpha\gamma$ | $- \beta\gamma$ | $+ \varepsilon_{212}$ |
| $y_{221}$ | $= \mu$ | $- \alpha$ | $- \beta$ | $+ \gamma$ | $+ \alpha\beta$ | $- \alpha\gamma$ | $- \beta\gamma$ | $+ \varepsilon_{221}$ |
| $y_{222}$ | $= \mu$ | $- \alpha$ | $- \beta$ | $- \gamma$ | $+ \alpha\beta$ | $+ \alpha\gamma$ | $+ \beta\gamma$ | $+ \varepsilon_{222}$ |

$$\left. \right\} \quad (7)$$

となり簡単化できる．式 (7) において  $p = 6$  となり，母数  $\mu, \alpha, \beta, \gamma, \alpha\beta, \alpha\gamma, \beta\gamma$  を式 (3) の  $\theta_0, \theta_1, \theta_2, \dots, \theta_6$  に対応付け，母数の (+, -) となる変数を  $x_{0,i'}, x_{1,i'}, x_{2,i'}, \dots, x_{6,i'}$  として定義する．これらの変数の間には，式 (1) のような従属関係は存在せず， $x_{0,i'}, x_{1,i'}, x_{2,i'}, \dots, x_{6,i'}$  は 1 次独立である．

$$\left. \begin{aligned} y_1 &= \theta_0 x_{0,1} + \theta_1 x_{1,1} + \theta_2 x_{2,1} + \cdots + \theta_6 x_{6,1} + \varepsilon_1 \\ y_2 &= \theta_0 x_{0,2} + \theta_1 x_{1,2} + \theta_2 x_{2,2} + \cdots + \theta_6 x_{6,2} + \varepsilon_2 \\ &\vdots \\ y_8 &= \theta_0 x_{0,8} + \theta_1 x_{1,8} + \theta_2 x_{2,8} + \cdots + \theta_6 x_{6,8} + \varepsilon_8 \end{aligned} \right\} \quad (8)$$

## 直交表のデザイン行列との関連

直交表をデザイン行列として表せば、表 1 に示すようになる。一般的に普及している直交表は、手計算の利便性のためにデザイン行列の「-1」を「2」に置き換えている。そのために、列と列との積和が 0 となることが条件の直交性が失われてしまう。スライド 33 に示すように (1, 2) のままで線形モデルとすることも可能であるが、スライド 35 に示すように結果の解釈が煩雑になるので、(1, -1) 対比型ダミー変数とすることが望ましい。ダミー変数といえ (0, 1) ではないかとの疑問も生ずるかもしれない。もちろん (0, 1) とすることは可能であるが、水準平均の推定が煩雑になるので避けた方がよい。

表 1  $L_8$  直交表に対するデザイン行列および (1, 2) 型  $L_8$  直交表

| ——— デザイン行列 $X$ ———                     |       |       |       |       |       |       |       |       |            |                 | $Y$   |     | 列 番 |     |      |     |      |      |       |
|--|-------|-------|-------|-------|-------|-------|-------|-------|------------|-----------------|-------|-----|-----|-----|------|-----|------|------|-------|
| No.                                    | $x_0$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | $\theta$   | $\varepsilon$   | $y$   | No. | (1) | (2) | (3)  | (4) | (5)  | (6)  | (7)   |
| 1                                      | 1     | 1     | 1     | 1     | 1     | 1     | 1     | 1     | $\theta_0$ | $\varepsilon_1$ | $y_1$ | 1   | 1   | 1   | 1    | 1   | 1    | 1    | 1     |
| 2                                      | 1     | 1     | 1     | 1     | -1    | -1    | -1    | -1    | $\theta_1$ | $\varepsilon_2$ | $y_2$ | 2   | 1   | 1   | 1    | 2   | 2    | 2    | 2     |
| 3                                      | 1     | 1     | -1    | -1    | 1     | 1     | -1    | -1    | $\theta_2$ | $\varepsilon_3$ | $y_3$ | 3   | 1   | 2   | 2    | 1   | 1    | 2    | 2     |
| 4                                      | 1     | 1     | -1    | -1    | -1    | -1    | 1     | 1     | $\theta_3$ | $\varepsilon_4$ | $y_4$ | 4   | 1   | 2   | 2    | 2   | 2    | 1    | 1     |
| 5                                      | 1     | -1    | 1     | -1    | 1     | -1    | 1     | -1    | $\theta_4$ | $\varepsilon_5$ | $y_5$ | 5   | 2   | 1   | 2    | 1   | 2    | 1    | 2     |
| 6                                      | 1     | -1    | 1     | -1    | -1    | 1     | -1    | 1     | $\theta_5$ | $\varepsilon_6$ | $y_6$ | 6   | 2   | 1   | 2    | 2   | 1    | 2    | 1     |
| 7                                      | 1     | -1    | -1    | 1     | 1     | -1    | -1    | 1     | $\theta_6$ | $\varepsilon_7$ | $y_7$ | 7   | 2   | 2   | 1    | 1   | 2    | 2    | 1     |
| 8                                      | 1     | -1    | -1    | 1     | -1    | 1     | 1     | -1    | $\theta_7$ | $\varepsilon_8$ | $y_8$ | 8   | 2   | 2   | 1    | 2   | 1    | 1    | 2     |
| $\mu$ $a$ $b$ $ab$ $c$ $ac$ $bc$ $abc$ |       |       |       |       |       |       |       |       |            |                 |       | 成分  | $a$ | $b$ | $ab$ | $c$ | $ac$ | $bc$ | $abc$ |
|  |       |       |       |       |       |       |       |       |            |                 |       | 群   | -1- | -2- | -3-  | -4- | -5-  | -6-  | -7-   |

線形モデルによる解析は、直交の条件が崩れても共通の解析手順により各種の推定ができる汎用的な方法であり、多くの統計ソフトの内部の計算ロジックに使われている。ただし、行列計算を伴うので、実験計画法の入門書では避けられている。また、実験計画法の専門書では、理論が重視されており実際の解析事例は、統計ソフトに丸投げするのが現状であり、行列計算による線形モデルを用いた事例を扱っている書籍は、ごく限られている。

Excel の行列関数を用いて、線形モデルによる解析が手軽にできることを示すが、Excel の分析ツールとして提供されている「回帰分析」により、線形モデルの解析の一部を肩代わりできるので解析手順の簡素化に役立つ。

直交表に対する手計算による解析手順を、Basic などの計算機言語により実現することは、田中・垂水 編 (1986), 「パソコン統計解析ハンドブック III 実験計画法編」に Basic プログラムが示されているように、比較的容易ではあった。ただし、線形モデルによる各種の推定および 95%信頼区間の計算は、行列計算になしには実現不可能に近いので、多くの実験計画法の書物では、第 10 章に示すように水準平均に対する分散から 95%信頼区間を求め、組合せ



平均については、田口の式、あるいは、伊奈の式などによる有効反復数  $n_e$  の計算を計算し分散を計算する方法が定式化されている。

## 直交表による実験で欠測値が生じた場合

最も基本的な  $L_8$  直交表を対象に、何らかのトラブルでデータが得られない状況を設定し、線形モデルによる解析法を示す。取り上げる事例としては、手計算による結果が示されている楠ら（1995）「応用実験計画法」の〔例題 6.6〕を用い、JMP の「モデルのあてはめ」による解析方法を示す。

表 2 に示すのはスライド 8 の  $L_8$  直交表実験で、4 因子（A, B, C, D）を取り上げ、互いに交互作用ないと想定されるが、もしも交互作用があった場合に備えて、因子の割り付けは A を(1)列、B を(2)列、C を(4)列、D を(7)列とした。割り付けがない第(3)列には、成分表示から  $A \times B$  と  $C \times D$ 、第(5)列には、 $A \times C$  と  $B \times D$ 、第(6)列には、 $A \times D$  と  $B \times C$  が出現することがわかる。実際の実験データで、残差の平均平方が大きい場合には、何らかの交互作用が示唆されるので、これらの列を考慮した解析を行うことにより、いくばくかの情報が得られる可能性を期待している。

因子 D を(7)列に割り付けたのは、成分表示が  $abc$  なので成分  $a$  の列、成分  $b$  の列、成分  $c$  の列の 3 因子交互作用が現れる列となり、3 因子（A, B, C）間に 2 因子交互作用がたとえあったとしても、その影響を受けないようにするための配慮である。

表 2 直交表への要因の割り付け [楠ら（1995）、例題 6.6]

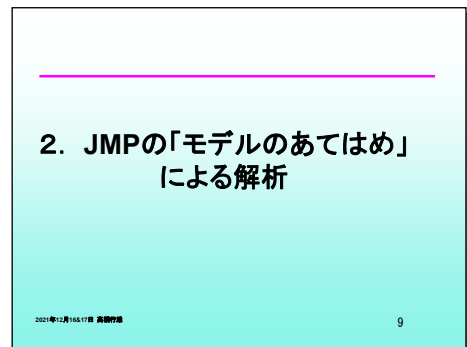
|      | A   | B   |              | C   |              |              | D     | データ |
|------|-----|-----|--------------|-----|--------------|--------------|-------|-----|
| No.  | (1) | (2) | (3)          | (4) | (5)          | (6)          | (7)   | $y$ |
| 1    | 1   | 1   | 1            | 1   | 1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

欠測値があるために伝統的な平方和の分解による方法が適用できないので、デザイン行列を用いた線形モデルを用いた解析が必要となる。

### 3. 欠測値がある直交表に対する JMP による解析

スライド 9

JMP の「モデルのあてはめ」による、欠測値がある直交表の解析を例示する。スライド 10 に欠測値を含む直交表  $L_8$  の JMP データを示す。因子 (A, B, C, D) は、データのタイプは数値としているが、尺度は名義尺度としている。No.7 は、明示的に欠測値であることを示すためにピリオド「.」を与えている。



スライド 10

JMP の「モデルのあてはめ」では、次に示すように「役割変数の選択」で反応変数  $y$  を、「モデル効果の構成」で説明変数 (A, B, C, D) を設定し、手法は「標準最小 2 乗法」とし、強調点としては「最小レポート」を選択する。

## 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**

2021年12月16&17日 高橋行雄 10

**モデルの指定**

列の選択

6列

- ☒ No.
- ☒ A
- ☒ B
- ☒ C
- ☒ D
- ☒ y

役割変数の選択

Y ☒ y

手法: 標準最小2乗

強調点: 最小レポート

ヘルプ 実行

モデル効果の構成

追加 A

交差 B

枝分かれ C

D

スライド 11 に JMP の「モデルのあてはめ」の実行結果を示す。因子が互いに直交することを大前提にする分散分析表では、各因子が先に表示され、その後に「誤差」と「全体 or 合計」が並べるのが一般的である。しかし、因子が互いに直交しない場合に、「各因子の平方和の合計」+「誤差平方和」=「合計」とはならない。そのために、直交を前提にしない JMP での分散分析表は、「モデル」に自由度 4 として全ての因子を込みにした平方和 303.0952 が示され、引き続き「効果の検定」で、各要因の平方和などが示されている。

効果の検定における (A, B, C, D) 4 因子の平方和の合計は、

$$S_A + S_B + S_C + S_D = 77.0417 + 112.6667 + 28.1667 + 22.0417 = 239.9168$$

となり、モデルの平方和 303.0952 とは異なる。このことは、従来の分散分析の解析方法である「平方和の分解」が成り立っていないことが明らかである!! このために、欠測値がある直交表に対し、欠測値に対し「何かしらの値で補完する」などといった便宜的な方法が、まことしやかに語られ続ける元凶となっている。なお、JMP が 1990 年に新たに発売された当時から、「モデルのあてはめ」では、欠測値を含む解析に対応できていたことは、驚嘆すべきことであった。

スライド 11

## 「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      | p値(Prob>F) |
| モデル      | 4   | 303.0952 | 75.7738  | 34.9725 |            |
| 誤差       | 2   | 4.3333   | 2.1667   |         |            |
| 全体(修正済み) | 6   | 307.4286 |          |         | 0.0280*    |
| 効果の検定    |     |          |          |         |            |
| 要因       | 自由度 | 平方和      | 平均平方     | F値      | p値(Prob>F) |
| A        | 1   | 77.0417  | 77.0417  | 35.5577 | 0.0270*    |
| B        | 1   | 112.6667 | 112.6667 | 52.0000 | 0.0187*    |
| C        | 1   | 28.1667  | 28.1667  | 13.0000 | 0.0091     |
| 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。

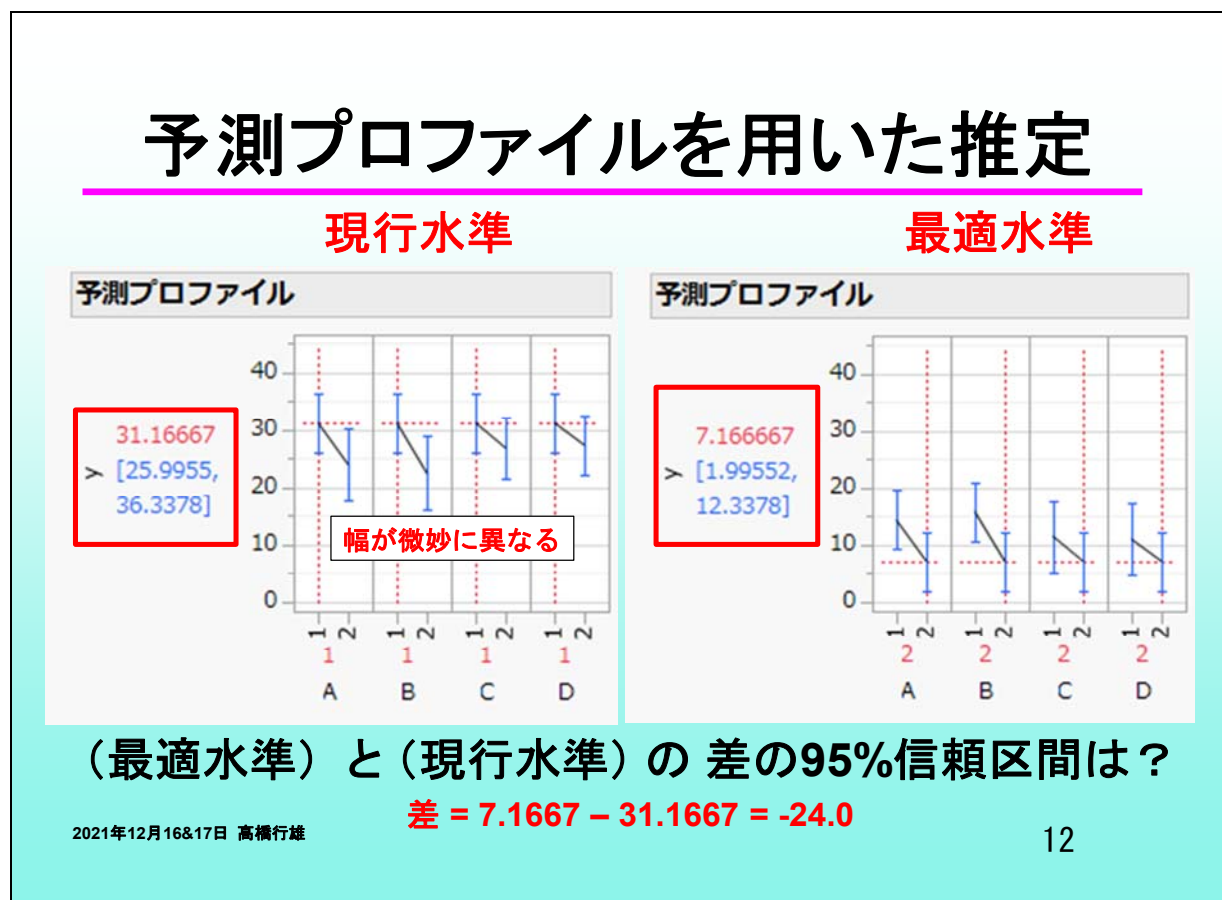
平方和の分解が成り立たない !!

## 予測プロファイルを用いた推定

直交表を用いた実験で、各因子の「現行水準」の組合せによる推定値に対し、各因子の「最適水準」を組合せた推定値を統計的に比較することが、従来の解析法の中で定式化されてきた。そのために、「現行水準」の組合せ、および、「最適水準」の組合せそれぞれについての推定値と 95%信頼区間、さらに、それらの差の推定値と 95%信頼区間の推定も標準的な解析方法として定着してきた。統計的には、かなり込み入った解析方法であるが、現場の技術者が実践できることを目的にした解析方法が整備されてきたことは、画期的なことであった。

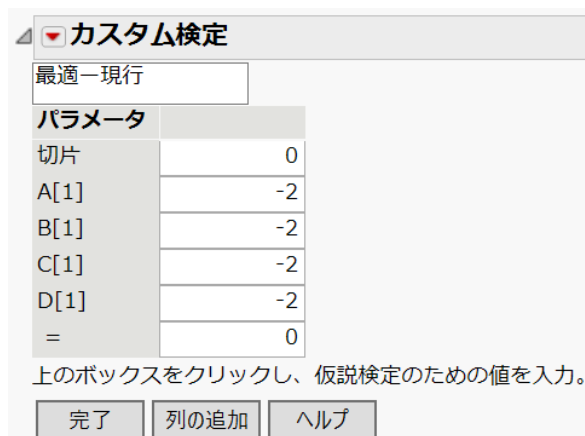
欠測値が存在しない直交表の解析は、電卓レベルの計算で対応できるのであるが、不幸にも欠測値が出てしまったら最後、完全にお手上げの状況となる。なお、JMP に先立つこと 15 年前から SAS の GLM プロシジャでは、欠測値ある直交表の解析が行えるような線形モデルを前提にした解析方法であった。スライド 12 の「予測プロファイルを用いた推定」に「現行水準」および「最適水準」についての推定値および 95%信頼区間の計算結果が示されている。さて、それらの差の推定と 95%信頼区間はどのようにして求めるのであろうか。

スライド 12



スライド 25 に対応

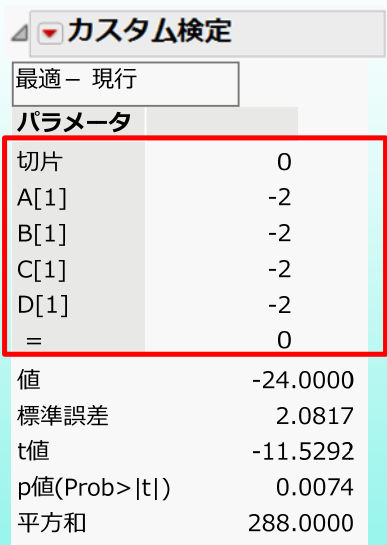
スライド 13 には、（最適水準）と（現行水準）の差を求めるために「推定値」のプルダウン・メニューにある「カスタム検定」の機能を用いた結果が示されている。「カスタム検定」の設定画面で、「最適 - 現行」と入力し、パラメータに謎めいた (-2, -2, -2, -2)

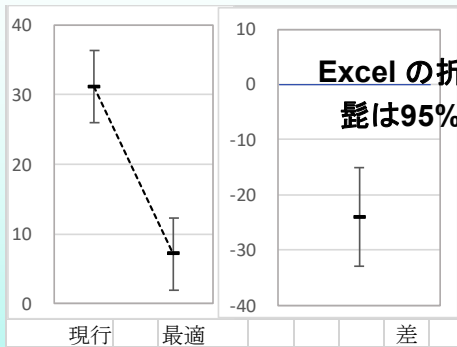


を入力することにより、差の推定値=24.0000 と標準誤差 2.0817 が出力されている。ただし、95%信頼区間は出力されないので、別途 Excel で計算し、折れ線グラフの機能を用いてグラフ化した結果が示されている。ここで、伊奈の式は使われているのであろうか。スライド 47 に Excel による計算法を詳しく示すが、伊奈は用いられていない。

スライド 13

## （最適水準）と（現行水準）の差





Excel の折れ線グラフ  
髭は95%信頼区間

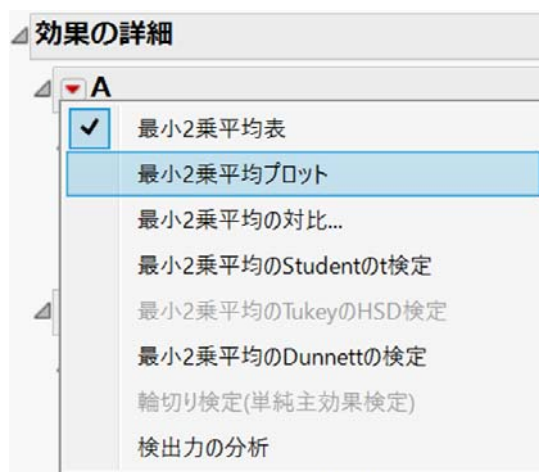
|            |          |
|------------|----------|
| 平方和        | 288.0000 |
| 分子自由度      | 1.0000   |
| F値         | 132.9231 |
| p値(Prob>F) | 0.0074   |

**カスタム検定の -2 は、何を意味しているのか？**

2021年12月16&17日 高橋行雄
謎めいた数値、本当なのか？
13

## 最小2乗平均

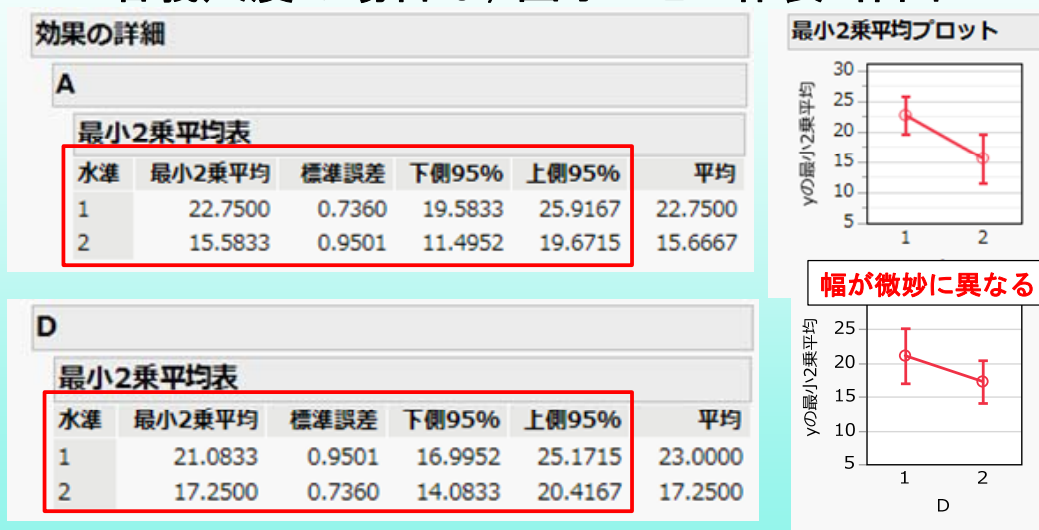
JMP の「モデルのあてはめ」は、汎用的であるがために、出力結果が冗長でまとまりに欠けるきらいがある。その代表例が、それぞれの因子の水準平均の出力とグラフ化である。スライド 14 に示す結果は、「効果の詳細」の因子の ▾ を選択し、「最小2乗平均プロット」を選択した結果である。因子ごとに選択を繰り返す必要があり、一括して表示することができない。なお、スライド 46 に Excel の折れ線グラフを使用した一括表示を例示する。



スライド 14

## 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



スライド 14 で示した「最小 2 乗平均」とは、何なのか？ どのように計算されたものなのか？ 「最小 2 乗平均 (Lsmeans)」は、JMP および SAS で使用されている方言のようなもので、統計の専門用語として認知されていない。欠測値あるは欠測セルがある場合の水準平均をどのように定義するのかに関連しているが、きちっと解説している日本語の書物を見いだすのは困難である。理解を深めるために、Excel の基本機能だけを用いて JMP および SAS から出力される「最小 2 乗平均」を再現し、その結果を元に妥当性を吟味する必要がある。なお、スライド 46 に Excel を用いた計算法を示してある。

線形モデルによる解析では、デザイン行列  $X$  (計画行列ともいう) を用いるのであるが、事例に基づいて解説している日本語の書物は皆無に等しい。そこで、欠測値のある直交表の解析を題材に、丁寧な解説を行うことにする。その第一歩として、スライド 15 に示す「列の保存」のプルダウン・メニューにある「コーディングのテーブルの保存」を実行してみよう。謎めいたプルダウン・メニューであるが、ポップアップされた説明に

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。

とあり、JMP の内部の計算で用いられているザイン行列  $X$  (計画行列) が出力されると理解される。スライド 16 に出力された JMP ファイルを示す。スライド 10 に示した解析に用いた JMP ファイルでは、各因子の水準に (1, 2) を与えていたのであるが、JMP 内部では、(1, -1) に置き換えられていることが推測できる。スライド 10 にはない「切片」が付加されている。これら全体が、線形モデルとした解析するための「デザイン行列  $X$ 」であり、JMP の内部で用いられていることが分かる。

スライド 15

### JMP: デザイン行列の出力

列の保存 ▶ 予測式

予測値の標準誤差の計算式  
平均の信頼限界の計算式  
個別の信頼限界の計算式

**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。**

コーディングのテーブルを保存

計画行列 $X$ , デザイン行列 $X$  は、同義語

2021年12月16日 17時 高橋行雄 15

スライド 16

### 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。

2021年12月16日 17時 高橋行雄 16



## 4. デザイン行列 $X$ の活用

スライド 17

各種の実験デザインに対する解析法は、偏差平方和をベースとする「平方和の分解」により分散分析表の作成に特化されている。そのために「平方和の分解」が不可能な欠測値を含む直交表データの解析が行えない。解析するためには、デザイン行列をベースにする線形モデルの解析を行なう必要がある。

### 3. デザイン行列 $X$ の活用

スライド 18

線形モデルを適用する際は、全ての 변수を連続尺度とすることが必須である。名義尺度の 변수は、連続尺度のダミー 변수として設定し直す必要がある。ダミー 변수に対して、断定的に「(なし, あり) を (0, 1) にすることでしょう」, などの説明が跋扈している。スライド 16 に示した (1, -1) への変換した結果をダミー 변수と言って良いのだろうか。

### 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 변수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 변수**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 변수は、**ダミー 변수**とされている。

スライド 19

「モデルのあてはめ」では、名義尺度の 변수が扱えるが、スライド 19 に示すように内部では名義尺度を全て (1, -1) 対比型の連続尺度のダミー 변수に変換している。ただし、計算結果の表示に際しては、元の名義尺度に戻し、ユーザに気が付かないように配慮されている。

### 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 | 32 |
| 2   | 2 | 1 | 1 | 2 | 23 |
| 3   | 3 | 1 | 2 | 1 | 19 |
| 4   | 4 | 1 | 2 | 1 | 17 |
| 5   | 5 | 2 | 1 | 1 | 19 |
| 6   | 6 | 2 | 1 | 2 | 20 |
| 7   | 7 | 2 | 2 | 1 | 8  |
| 8   | 8 | 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    | 32 |
| 2  | 1    | 1    | 1    | -1   | 23 |
| 3  | 1    | 1    | -1   | 1    | 19 |
| 4  | 1    | 1    | -1   | 1    | 17 |
| 5  | 1    | -1   | 1    | 1    | 19 |
| 6  | 1    | -1   | 1    | -1   | 20 |
| 7  | *    | *    | *    | *    | *  |
| 8  | 1    | -1   | -1   | -1   | 8  |

切片, A[1], ..., D[1] は、連続 변수



## 各種のダミー変数の歴史的背景

平方和の分解を前提にした場合に、直交表として使いやすいのは (1, 2) 型である。これは、各因子の第 1 水準、第 2 水準に対応していて「第 1 水準の合計、第 2 水準の合計を求める」など手順の説明として分かりやすいからであろう。スライド 20 に示すように、朝香ら (1988)、「新版 品質管理便覧 第 2 版, p446-7.」,「7.1.4.2 節 直交表の性質とその利用」に直交表の歴史的背景が詳しく述べられている。

- (a) (1, -1) 基本型
- (b) (1, 2) 型, +1 を 1, -1 を 2 におきかえた。実験計画に使いやすいように、田口<sup>14)</sup>・小西氏<sup>15)</sup>によって提唱された。
- (c) (0, 1) 型, +1 を 0, -1 を 1 におきかえたもの。嶋田氏<sup>16)</sup>によって提唱された。

従って、JMP の内部で使われているデザイン行列でのダミー変数は、(0, 1) 型ではなく直交表の基本型である (1, -1) 型であることが理解される。

スライド 20

## 3種類の $L_4$ 直交表

| (a) 基本型 |     |          |          |     |
|---------|-----|----------|----------|-----|
| No.     | (0) | 列<br>(1) | 番<br>(2) | (3) |
| 1       | +1  | +1       | +1       | +1  |
| 2       | +1  | +1       | -1       | -1  |
| 3       | +1  | -1       | +1       | -1  |
| 4       | +1  | -1       | -1       | +1  |

| (b) $L_4(2^3)$ |          |          |     | (c) $H_{2,4}$ |          |          |     |
|----------------|----------|----------|-----|---------------|----------|----------|-----|
| No.            | 列<br>(1) | 番<br>(2) | (3) | No.           | 列<br>(1) | 番<br>(2) | (3) |
| 1              | 1        | 1        | 1   | 1             | 0        | 0        | 0   |
| 2              | 1        | 2        | 2   | 2             | 0        | 1        | 1   |
| 3              | 2        | 1        | 2   | 3             | 1        | 0        | 1   |
| 4              | 2        | 2        | 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<sup>14)</sup>・小西氏<sup>15)</sup>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<sup>16)</sup>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

スライド 21 で、各種のダミー変数に対する補足説明をする。名義尺度の変数を含むデータに対し、線形モデルを適用する解析は、一般的に回帰分析の手順を準用するので、名義尺度の変数を連続尺度の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、最初的水準を 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、( $A_1 \rightarrow 1$ ,  $A_2 \rightarrow -1$ ) 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、スライド 20 で示したように直交表の (基本型) である。ただし、ダミー変数の与え方は、統計ソフトごとに異なる。SAS の GLM プロシジャでは、最初的水準を 1、次の水準を 0 とするような (1, 0) 型ダミー変数である。R の lm 関数では、(0, 1) 型ダミー変数が使われている。

JMP で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。ここでは、「切片」を除いて解析を行う。JMP のコーディングのテーブルで「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教育的である。コーディングのテーブルのダミー変数 A[1], B[1], C[1], D[1] は、連続変数のまま「モデル効果の構成」に設定する。ただし、「切片」を除いた解析をしているつもりでも、内部では「切片」が自動的に付与される。

原点を通る直線を引きたい、原点を通る 2 本の直線の同時あてはめを行いたい、などの場合は、いわゆる「切片」が自動付与されると解析が行えない。そのために、JMP では「☐ 切片なし」のチェックボックスがあり、Excel の回帰分析にも「☐ 定数に 0 を使用」チェックボックスがある。SAS の GLM プロシジャでは、model ステートメントの「nointercept」オプションを指定するように、対策が講じられている。

#### スライド 21

### 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、最初的水準を 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、( $A_1 \rightarrow 1$ ,  $A_2 \rightarrow -1$ ) 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。

2021年12月16日 17時 高橋行雄

21

#### スライド 22

### 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。

2021年12月16日 17時 高橋行雄

22

スライド 23 に示したのは、コーディングのテーブル（デザイン行列 **X**）を用いた解析結果である。因子 A に対し変数 A[1] と異なるが、表 3 に再掲するように（1, 2）型名義尺度の場合の解析結果と完全に一致する。

表 3 因子を名義尺度として解析した結果（スライド 11 再掲）

| 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     |

スライド 23

## 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

## デザイン行列を用いたパラメータの推定

回帰分析の基本出力は、スライド 24 に示した回帰パラメータの出力であり、スライド 23 に示した分散分析表に引き続き出力されている「効果の検定」は、補足的な出力である。しかし、平方和の分解による分散分析表の作成を主体にした解析手順に、もちろん「回帰パラメータ」の計算過程は含まれていない。

「パラメータの推定値」に4変数に対する推定値、標準誤差が出力され、 $t$ 検定の結果が出力されている。2水準の直交表の解析なので、この結果は、スライド 23 で示した「効果の検定」における  $F$  検定の結果に一致する。したがって、2水準の直交表の解析においては、「効果の検定」よりも「回帰パラメータの推定値」の方が、水準間の差の大きさが示されるので、情報量が多い。

「推定値の相関」は、「応答  $y$ 」の  プルダウン・メニューの「推定値」の中から「推定値の相関」を選択することにより得られる。切片を含めて  $5 \times 5$  の推定値の相関行列が出力されている。これは、スライド 40 に示すデータの相関行列とは、まったく別物であり、各種の推定の基礎となるパラメータの分散共分散行列（スライド 41 参照）から導出される。

スライド 24

## 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  |

◆ 回帰パラメータの推定値を出力する。

推定値の相関

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

◆ 推定値の相関行列も追加で出力する。

## 現行水準および最適水準の組合せ平均

スライド 12 で示した名義尺度 (1, 2) による現行水準および最適水準の組合せ平均に関する予測プロファイルと対比するために、コーディングのテーブル (デザイン行列 **X**) を用いた予測プロファイルをスライド 25 に示す。

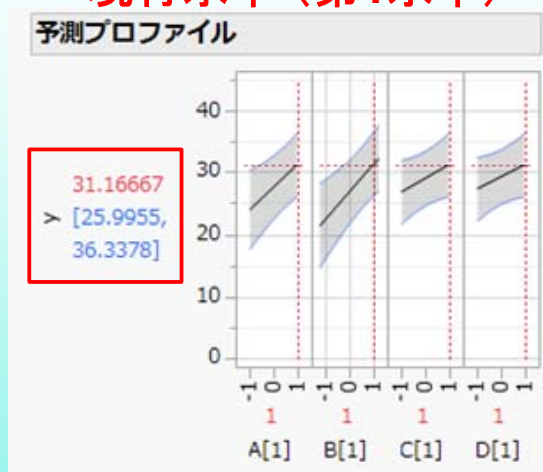
名義尺度 (1, 2) に対して (1, -1) としたので、現行水準 (第 1 水準) は、各変数の X 軸で「1」と設定する。最適水準は、(第 2 水準) の組合せなので、各因子の X 軸で「-1」と設定する。推定された現行水準が 31.16667、最適水準が 7.166667 とスライド 12 に示した結果と一致している。もちろん 95%信頼区間も一致している。

JMP の内部では、このような計算がされているのであるが、変数を名義尺度として解析した場合の出力結果には、内部の事情を感じさせないような出力結果となっている。

スライド 25

## 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

スライド 13 で、（最適水準）と（現行水準）の差に関する統計量の算出の際し、「カスタム検定」の機能で「最適 - 現行」と入力し、パラメータに謎めいた（-2, -2, -2, -2）を入力することにより、差の推定値=24.0000 と 標準誤差=2.0817 が出力されていることを示した。なぜ、（-2, -2, -2, -2）なのかは、「コーディングのテーブル（デザイン行列  $X$ ）」による解析と連動している。

スライド 25 で示したように、現行水準と最適水準の差は、

|         |  |
|---------|--|
| 最適水準    | 切片= 1, $A_2 = -1$ , $B_2 = -1$ , $C_2 = -1$ , $D_2 = -1$ |
| 一) 現行水準 | 切片= 1, $A_1 = 1$ , $B_1 = 1$ , $C_1 = 1$ , $D_1 = 1$     |
|         | 0      -2,      -2,      -2,      -2                     |

であり、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、この内部変数の水準間の差を設定することになる。

スライド 26

## （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**  $A_2 = -1, \dots, D_2 = -1$

一) **現行**  $A_1 = 1, \dots, D_1 = 1$

**-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   |

## 最小 2 乗平均と算術平均の比較

さて、スライド 14 で示した各因子の水準に対する「最小 2 乗平均」は、スライド 27 で示すように、「パラメータの推定値」から計算された線形和に一致する。すなわち、因子 A の第 1 水準と第 2 水準は、

$$\text{切片} + A[1] = 19.1667 + 3.5833 = 22.7500, \quad (n=4)$$

$$\text{切片} - A[1] = 19.1667 - 3.5833 = 15.5833, \quad (n=3)$$

として計算され、これが最小 2 乗平均として表示されている。欠測値があるままの算術平均値を計算すると、表 4 に示すように第 1 水準の最小 2 乗平均は、算術平均 22.7500 に一致するが、第 2 水準の場合は、最小 2 乗平均が 15.5833 であるのに対し、算術平均は 15.6677 と一致しない。因子 D の場合は、水準 1 の場合が一致しない。

表 4 JMP の「表の作成」による算術平均

表の作成

|   | y |   | 1       | 2       |
|---|---|---|---------|---------|
|   | 1 | 2 | y       | y       |
|   | N | N | 平均      | 平均      |
| A | 4 | 3 | 22.7500 | 15.6667 |
| B | 4 | 3 | 23.5000 | 14.6667 |
| C | 3 | 4 | 23.3333 | 17.0000 |
| D | 3 | 4 | 23.0000 | 17.2500 |

このように、JMP および SAS では、欠測値があることを前提にした解析法のために、回帰分析の推定値から算出した線形和による推定値を「最小 2 乗平均」と独自に命名している。

スライド 27

## それぞれの因子の最小 2 乗平均

(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<sub>1</sub> 水準 切片 + A[1],    A<sub>2</sub> 水準 切片 - A[1]

D<sub>1</sub> 水準 切片 + D[1],    D<sub>2</sub> 水準 切片 - D[1]



## JMP の使い勝手

さて、JMP の使い勝手であるが、因子を (1, 2) 型の名義尺度として解析することにより、欠測値のある直交表の解析が容易にできることは、画期的である。最適水準の組合せ、現行水準の組合せの推定と 95%信頼区間の計算も予測プロファイルで GUI 的にできることは、素晴らしい。カスタム検定を用いて（最適水準と現行水準）の差の推定は、JMP の内部計算の仕組みを理解なしには不可能である。因子ごとの水準平均と 95%信頼区間のグラフは、個別の表示となり、まとめて表示できないもどかしさが残る。

スライド 28

### 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.

2021年12月16&17日 高橋行雄

28

JMP の「モデルのあてはめ」を用いて欠測値がある直交表の解析方法を示してきたのであるが、JMP はライセンス契約が必要な有償ソフトであり、誰でも手軽に使える統計ソフトではない。無償の統計ソフトとして R が脚光を浴びているが、欠測値がある直交表の解析は、一筋縄では行かない。無償版の SAS/GLM プロシジャを用いれば、R よりも豊富な結果が得られるので、第 5 章でスクリプトと結果の要約を示す。ただし、第 6 章に示す Excel による解析法を習得してからチャレンジされることを勧める。



## 5. SAS/GLM による欠測値がある直交表の解析

無償版の SAS の GLM プロシジャを用い、欠測値がある直交表の解析方法を示す。GLM プロシジャを用いた解析方法の詳細は、高橋・大橋・芳賀（1989）、「SASによる実験データの解析」に記されているので参照されたい。ここでは、JMPによる結果をSASで実現する方法について示す。用いるデータは、スライド10に示した JMP の  $L_8$  データと同じである。

GLM プロシジャでは、変数を名義尺度として扱う場合は、class ステートメントで定義し、各因子の最小2乗平均の算出は、lsmeans ステートメントを用いる。現行水準と最適水準およびその差は、estimate ステートメントで設定するが、その際に GLM プロシジャの内部でデザイン行列への展開方法に留意しなければならない。基本の考え方は、(1, 0) 型のダミー変数であるが、estimate ステートメントでの設定では、式 (2) で示した母数に無駄の有る1次従属型のデザイン行列に即した設定が求められる。

### SAS のスクリプト

```
Title1 'L8missing.sas 2022-3-15 Y.Takahashi ' ;

data D1 ;
    input No A B C D y ;
/*No. A B C D y */
datalines ;
    1 1 1 1 1 32
    2 1 1 2 2 23
    3 1 2 1 2 19
    4 1 2 2 1 17
    5 2 1 1 2 19
    6 2 1 2 1 20
    7 2 2 1 1 .
    8 2 2 2 2 8
;
proc print data=D1 ; run ;

proc glm data=D1 ;
    class A B C D ;
    model y = A B C D / solution ;
    lsmeans A B C D / stderr cl ;
    estimate '現行' intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
    estimate '最適' intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
    estimate '最適-現行' intercept 0 A -1 1 B -1 1 C -1 1 D -1 1 ;
    estimate '最小2乗平均 A1f' intercept 1 A 1 0 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;
    estimate '最小2乗平均 A1s' intercept 1 A 1 0 ;
    estimate '最小2乗平均 A2f' intercept 1 A 0 1 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;
    estimate '最小2乗平均 A2s' intercept 1 A 0 1 ;
run ;
```

## GLM プロシジャの出力

class ステートメントで名義尺度とした変数のカテゴリの表示

| GLM プロシジャ  |    |     |
|------------|----|-----|
| 分類変数の水準の情報 |    |     |
| 分類         | 水準 | 値   |
| A          | 2  | 1 2 |
| B          | 2  | 1 2 |
| C          | 2  | 1 2 |
| D          | 2  | 1 2 |

反応 y の欠測値を自動的に削除

|                |   |
|----------------|---|
| 読み込んだオブザベーション数 | 8 |
| 使用されたオブザベーション数 | 7 |

スライド 11 の分散分析表に対応

| 要因              | 自由度 | 平方和         | 平均平方       | F 値   | Pr > F |
|-----------------|-----|-------------|------------|-------|--------|
| Model           | 4   | 303.0952381 | 75.7738095 | 34.97 | 0.0280 |
| Error           | 2   | 4.3333333   | 2.1666667  |       |        |
| Corrected Total | 6   | 307.4285714 |            |       |        |

JMP にも同等の出力があるが省略した

| R2 乗     | 変動係数     | Root MSE | y の平均    |
|----------|----------|----------|----------|
| 0.985905 | 7.466465 | 1.471960 | 19.71429 |

JMP の逐次 (タイプ 1) の検定に対応

| 要因 | 自由度 | Type I 平方和  | 平均平方        | F 値   | Pr > F |
|----|-----|-------------|-------------|-------|--------|
| A  | 1   | 86.0119048  | 86.0119048  | 39.70 | 0.0243 |
| B  | 1   | 176.8166667 | 176.8166667 | 81.61 | 0.0120 |
| C  | 1   | 18.2250000  | 18.2250000  | 8.41  | 0.1012 |
| D  | 1   | 22.0416667  | 22.0416667  | 10.17 | 0.0858 |

スライド 11 に対応

| 要因 | 自由度 | Type III 平方和 | 平均平方        | F 値   | Pr > F |
|----|-----|--------------|-------------|-------|--------|
| A  | 1   | 77.0416667   | 77.0416667  | 35.56 | 0.0270 |
| B  | 1   | 112.6666667  | 112.6666667 | 52.00 | 0.0187 |
| C  | 1   | 28.1666667   | 28.1666667  | 13.00 | 0.0691 |
| D  | 1   | 22.0416667   | 22.0416667  | 10.17 | 0.0858 |

最後の水準を基準とする (1, 0) 型のダミー変数での解析, 表 12 に対応

| パラメータ     | 推定値         |   | 標準誤差       | t 値  | Pr >  t |
|-----------|-------------|---|------------|------|---------|
| Intercept | 7.166666667 | B | 1.20185043 | 5.96 | 0.0270  |
| A 1       | 7.166666667 | B | 1.20185043 | 5.96 | 0.0270  |
| A 2       | 0.000000000 | B | .          | .    | .       |
| B 1       | 8.666666667 | B | 1.20185043 | 7.21 | 0.0187  |
| B 2       | 0.000000000 | B | .          | .    | .       |
| C 1       | 4.333333333 | B | 1.20185043 | 3.61 | 0.0691  |
| C 2       | 0.000000000 | B | .          | .    | .       |
| D 1       | 3.833333333 | B | 1.20185043 | 3.19 | 0.0858  |
| D 2       | 0.000000000 | B | .          | .    | .       |

lsmeans ステートメントの出力    lsmeans A B C D / stderr cl

スライド 14 に  
対応

| A | y の最小 2 乗平均 | 標準誤差      | Pr >  t | 95% 信頼限界  |           |
|---|-------------|-----------|---------|-----------|-----------|
| 1 | 22.7500000  | 0.7359801 | 0.0010  | 19.583333 | 25.916667 |
| 2 | 15.5833333  | 0.9501462 | 0.0037  | 11.495184 | 19.671482 |

| B | y の最小 2 乗平均 | 標準誤差      | Pr >  t | 95% 信頼限界  |           |
|---|-------------|-----------|---------|-----------|-----------|
| 1 | 23.5000000  | 0.7359801 | 0.0010  | 20.333333 | 26.666667 |
| 2 | 14.8333333  | 0.9501462 | 0.0041  | 10.745184 | 18.921482 |

| C | y の最小 2 乗平均 | 標準誤差      | Pr >  t | 95% 信頼限界  |           |
|---|-------------|-----------|---------|-----------|-----------|
| 1 | 21.3333333  | 0.9501462 | 0.0020  | 17.245184 | 25.421482 |
| 2 | 17.0000000  | 0.7359801 | 0.0019  | 13.833333 | 20.166667 |

スライド 14 に  
対応

| D | y の最小 2 乗平均 | 標準誤差      | Pr >  t | 95% 信頼限界  |           |
|---|-------------|-----------|---------|-----------|-----------|
| 1 | 21.0833333  | 0.9501462 | 0.0020  | 16.995184 | 25.171482 |
| 2 | 17.2500000  | 0.7359801 | 0.0018  | 14.083333 | 20.416667 |

estimate ステートメントの出力

スライド 25・  
26, 表 14 に  
対応

| パラメータ | 推定値         | 標準誤差       | t 値    | Pr >  t |
|-------|-------------|------------|--------|---------|
| 現行    | 31.1666667  | 1.20185043 | 25.93  | 0.0015  |
| 最適    | 7.1666667   | 1.20185043 | 5.96   | 0.0270  |
| 最適-現行 | -24.0000000 | 2.08166600 | -11.53 | 0.0074  |

スライド 12 に示した JMP での現行水準と最適水準の推定は、GUI 操作により推定できるが、SAS の GLM プロシジャでは、GLM の内部で生成される式 (2) で示した母数に無駄の有るデザイン行列に即して estimate ステートメントで次のように設定する必要がある。

```
estimate '現行' intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
estimate '最適' intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
```

最適水準から現行水準を引く場合は、「最適」から「現行」を引けばよい。JMP では、スライド 13 に示したようにカスタム検定を使うのであるが、唐突に内部で生成されるデザイン行列の知識が求められる。JMP に比べて、GLM プロシジャでは首尾一貫している。ただし、「最適」と「現行」水準の設定ができる事が前提である。

```

estimate '最適'      intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
一) estimate '現行'  intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
estimate '最適-現行' intercept 0 A -1 1 B -1 1 C -1 1 D -1 1 ;

```

## Estimate ステートメントによる最小 2 乗平均

各因子の水準に対する最小 2 乗平均計算は、lsmeans ステートメントで求めるのが一般的であるが、estimate ステートメントを用いてもできる。因子 A の第 1 水準の場合であれば、他の因子のそれぞれの水準の平均値 0.5 を与える必要があるが、設定しなくても 0.5 が自動的に補われる。表 13 の Excel による例示を参照のこと。

```

estimate '最小 2 乗平均 A1f' intercept 1 A 1 0 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;
estimate '最小 2 乗平均 A1s' intercept 1 A 1 0 ;

```

| パラメータ        | 推定値        | 標準誤差       | t 値   | Pr >  t |
|--------------|------------|------------|-------|---------|
| 最小 2 乗平均 A1f | 22.7500000 | 0.73598007 | 30.91 | 0.0010  |
| 最小 2 乗平均 A1s | 22.7500000 | 0.73598007 | 30.91 | 0.0010  |
| 最小 2 乗平均 A2f | 15.5833333 | 0.95014619 | 16.40 | 0.0037  |
| 最小 2 乗平均 A2s | 15.5833333 | 0.95014619 | 16.40 | 0.0037  |

最小 2 乗平均は、推定された回帰パラメータを用いて、各因子の水準平均を算出しているのであるが、(1, 0) 型ダミー変数の場合でも、(0, 1) 型ダミー変数の場合でも、ここに示したように推定されたパラメータから簡単には計算することができない。(1, -1) 対比型ダミー変数の場合であれば、スライド 27 に示すように、切片（総平均  $\mu$  の推定値）に A[1] の推定値をプラス・マイナスするといった直観的な計算で求められるし、他者に説明することも容易である。

## 6. Excel の回帰分析を用いた線形モデル

スライド 29

JMP などの多くの統計ソフトで、回帰分析が提供されているので、「今更 Excel の回帰分析なんて」と長年にわたり私自身も見下していた。ただし、線形モデルを用いた様々な事例の解析方法を丁寧に説明しようとしたときに、特定の統計ソフトを前提とすると、統計ソフトの使い方が主体となりがちで、線形モデルの理論に即した解析方法の説明がおろそかになってしまう。

### 4. Excel の回帰分析による線形モデル

2021年12月16&17日 高橋行雄

29

スライド 30 には、Excel で作成したデザイン行列を用いた回帰式の右側に、Excel の「分析ツール」の「回帰分析」の結果が Excel のシート上に並べて表示している。このようなことが簡単にできることが、直観的な理解の助けになると思われる。

スライド 30

## 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 $X$ |       |       |       |       |       | $Y$        |                 |     | 分散分析表 |         |         |         |         |
|------------|-------|-------|-------|-------|-------|------------|-----------------|-----|-------|---------|---------|---------|---------|
|            | $x_0$ | $x_1$ | $x_2$ | $x_4$ | $x_7$ |            |                 |     |       | 自由度     | 変動      | 分散      | 分散比     |
| No.        | $\mu$ | A     | B     | C     | D     | $\theta$   | $\varepsilon$   | $y$ |       |         |         |         |         |
| 1          | 1     | 1     | 1     | 1     | 1     | $\theta_0$ | $\varepsilon_1$ | 32  | 回帰    | 4       | 303.095 | 75.7738 | 34.9725 |
| 2          | 1     | 1     | 1     | -1    | -1    | $\theta_1$ | $\varepsilon_2$ | 23  | 残差    | 2       | 4.333   | 2.1667  |         |
| 3          | 1     | 1     | -1    | 1     | -1    | $\theta_2$ | $\varepsilon_3$ | 19  | 合計    | 6       | 307.429 |         |         |
| 4          | 1     | 1     | -1    | -1    | 1     | $\theta_3$ | $\varepsilon_4$ | 17  |       |         |         |         |         |
| 5          | 1     | -1    | 1     | 1     | -1    | $\theta_4$ | $\varepsilon_5$ | 19  |       | 係数      | 標準誤差    | $t$     | $P$ -値  |
| 6          | 1     | -1    | 1     | -1    | 1     |            |                 | 20  | 切片    | 19.1667 | 0.6009  | 31.8953 | 0.0010  |
| 8          | 1     | -1    | -1    | -1    | -1    |            |                 | 8   | A     | 3.5833  | 0.6009  | 5.9630  | 0.0270  |
|            | $\mu$ | $a$   | $b$   | $c$   | $abc$ |            |                 |     | 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。

2021年12月16&17日 高橋行雄

30

スライド 30 で示した Excel の回帰分析に対し、補足説明をする。表 5 に示すように Excel の回帰分析では、通常「切片  $\mu$ 」列を含めない「デザイン行列」を説明変数として設定する。右下に示したパラメータの共分散行列  $\Sigma(\hat{\theta})$  は、Excel の行列関数および分散分析表に示されている残差の分散（誤差分散の推定値）  $\hat{\sigma}^2 = 2.1667$  を用いて、

$$\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2 = \text{Minverse}(\text{Mmult}(\text{Transpose}(X\text{の範囲}), X\text{の範囲})) * \hat{\sigma}^2 \quad (9)$$

として一括計算した結果である。この計算に用いるデザイン行列  $X$  は、もちろん「切片  $\mu$ 」を含んだデザイン行列  $X$  である。パラメータの共分散行列  $\Sigma(\hat{\theta})$  の計算のため、Excel の回帰分析で必要のない「切片  $\mu$ 」の列も明示したのである。なお、Excel の行列関数の使い方については、表 20 の例示を参照のこと。

回帰分析の標準誤差は、どのようにして計算された結果なのであろうか。パラメータの共分散行列  $\Sigma(\hat{\theta})$  の対角要素が、各パラメータの分散となっているので、それらの平方根が、標準誤差  $SE$  に対応する。

表 5 Excel の回帰分析によるパラメータの推定

| デザイン行列 $X$ |       |       |       |       |       |            |                 | $Y$ | 分散分析表 |   |         |         |         |
|------------|-------|-------|-------|-------|-------|------------|-----------------|-----|-------|---|---------|---------|---------|
|            | $x_0$ | $x_1$ | $x_2$ | $x_4$ | $x_7$ |            |                 |     |       | 自由度   | 変動      | 分散      | 分散比     |
| No.        | $\mu$ | A     | B     | C     | D     | $\theta$   | $\varepsilon$   | $y$ | 回帰    | 4   | 303.095 | 75.7738 | 34.9725 |
| 1          | 1     | 1     | 1     | 1     | 1     | $\theta_0$ | $\varepsilon_1$ | 32  | 残差    | 2   | 4.333   | 2.1667  |         |
| 2          | 1     | 1     | 1     | -1    | -1    | $\theta_1$ | $\varepsilon_2$ | 23  | 合計    | 6   | 307.429 |         |         |
| 3          | 1     | 1     | -1    | 1     | -1    | $\theta_2$ | $\varepsilon_3$ | 19  |       |   |         |         |         |
| 4          | 1     | 1     | -1    | -1    | 1     | $\theta_3$ | $\varepsilon_4$ | 17  |       | 係数  | 標準誤差    | $t$     | $P$ -値  |
| 5          | 1     | -1    | 1     | 1     | -1    | $\theta_4$ | $\varepsilon_5$ | 19  | 切片    | 19.1667   | 0.6009  | 31.8953 | 0.0010  |
| 6          | 1     | -1    | 1     | -1    | 1     |            |                 | 20  | A     | 3.5833  | 0.6009  | 5.9630  | 0.0270  |
| 8          | 1     | -1    | -1    | -1    | -1    |            |                 | 8   | B     | 4.3333  | 0.6009  | 7.2111  | 0.0187  |
|            | $\mu$ | $a$   | $b$   | $c$   | $abc$ |            |                 |     | 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 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \hat{\sigma}^2$  |         |         |         |
|            |       |       |       |       |       |            |                 |     |       | 0.3611  | -0.0903 | -0.0903 | 0.0903  |
|            |       |       |       |       |       |            |                 |     |       | -0.0903   | 0.3611  | 0.0903  | -0.0903 |
|            |       |       |       |       |       |            |                 |     |       | -0.0903   | 0.0903  | 0.3611  | -0.0903 |
|            |       |       |       |       |       |            |                 |     |       | 0.0903  | -0.0903 | -0.0903 | 0.3611  |
|            |       |       |       |       |       |            |                 |     |       | 0.0903  | -0.0903 | -0.0903 | 0.3611  |
|            |       |       |       |       |       |            |                 |     |       | =Minverse(Mmult(Transpose( $X$ の範囲), $X$ の範囲)) * $\hat{\sigma}^2$ |         |         |         |

通常の直交表の解析では、因子ごとの平方和、平均平方を求め、平均平方を誤差分散  $\hat{\sigma}^2 = 2.1667$  で割った  $F$  値から  $p$  値を求めている。Excel の回帰分析では、分散分析表にすべての因子を込みにした変動（平方和）、分散（平均平方）しか出力されない。ただし、回帰係数の  $p$  値が、分散分析表の  $F$  値から求められた  $p$  値に一致するので、この関係を用てスライド 32 に示すように分散分析表にまとめることができる。

## 正規方程式の Excel シート上での表現

Excel の回帰分析によらず，Excel の行列関数を用いて，表 5 を再現してみよう．表 6 に示すのは，表 5 のデザイン行列を用いた回帰式を  $\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \mathbf{Y}$  とし，その両辺にデザイン行列  $\mathbf{X}$  の転置行列  $\mathbf{X}^T$  を掛け，正規方程式  $(\mathbf{X}^T \mathbf{X})\hat{\boldsymbol{\theta}} = \mathbf{X}^T \mathbf{Y}$  とする過程を Excel の行列表記を用いて示したものである．行列表記の  $\mathbf{X}^T \mathbf{X}(a_{kk'})$  の  $a_{kk'}$  は，積和行列  $\mathbf{X}^T \mathbf{X}$  の要素を表す場合に使用する．また， $\mathbf{X}^T \mathbf{Y}$  の  $(B_0, B_1, \dots, B_4)$  は，ベクトル  $\mathbf{X}^T \mathbf{Y}$  の各要素を表す場合に使用する．なお，正規方程式の導出については，第 9 章を参照のこと．

表 6 デザイン行列  $\mathbf{X}$  を用いた正規方程式

|       | 転置行列 $\mathbf{X}^T$ |    |    |    |    |    |    |  | デザイン行列 $\mathbf{X}$ |       |       |       |       |  |                   |   | 転置行列 $\mathbf{X}^T$ |    |    |    |    |    |    |  | $\mathbf{Y}$ |
|-------|---------------------|----|----|----|----|----|----|--|---------------------|-------|-------|-------|-------|--|-------------------|---|---------------------|----|----|----|----|----|----|--|--------------|
|       | 1                   | 2  | 3  | 4  | 5  | 6  | 8  |  | $x_0$               | $x_1$ | $x_2$ | $x_4$ | $x_7$ |  | $\theta^\wedge$   |   | 1                   | 2  | 3  | 4  | 5  | 6  | 8  |  | $y$          |
| $\mu$ | 1                   | 1  | 1  | 1  | 1  | 1  | 1  |  | 1                   | 1     | 1     | 1     | 1     |  | $\theta^\wedge_0$ | = | 1                   | 1  | 1  | 1  | 1  | 1  | 1  |  | 32           |
| A     | 1                   | 1  | 1  | 1  | -1 | -1 | -1 |  | 1                   | 1     | 1     | -1    | -1    |  | $\theta^\wedge_1$ |   | 1                   | 1  | 1  | 1  | -1 | -1 | -1 |  | 23           |
| B     | 1                   | 1  | -1 | -1 | 1  | 1  | -1 |  | 1                   | 1     | -1    | 1     | -1    |  | $\theta^\wedge_2$ |   | 1                   | 1  | -1 | -1 | 1  | 1  | -1 |  | 19           |
| C     | 1                   | -1 | 1  | -1 | 1  | -1 | -1 |  | 1                   | 1     | -1    | -1    | 1     |  | $\theta^\wedge_3$ |   | 1                   | -1 | 1  | -1 | 1  | -1 | -1 |  | 17           |
| D     | 1                   | -1 | -1 | 1  | -1 | 1  | -1 |  | 1                   | -1    | 1     | 1     | -1    |  | $\theta^\wedge_4$ |   | 1                   | -1 | -1 | 1  | -1 | 1  | -1 |  | 19           |
|       |                     |    |    |    |    |    |    |  | 1                   | -1    | 1     | -1    | 1     |  |                   |   |                     |    |    |    |    |    |    |  | 20           |
|       |                     |    |    |    |    |    |    |  | 1                   | -1    | -1    | -1    | -1    |  |                   |   |                     |    |    |    |    |    |    |  | 8            |

| $\mathbf{X}^T \mathbf{X}(a_{kk'})$ |    |    |    |    | $\theta^\wedge$   |   | $\mathbf{X}^T \mathbf{Y}$ |       |
|------------------------------------|----|----|----|----|-------------------|---|---------------------------|-------|
| 7                                  | 1  | 1  | -1 | -1 | $\theta^\wedge_0$ | = | 138.0                     | $B_0$ |
| 1                                  | 7  | -1 | 1  | 1  | $\theta^\wedge_1$ |   | 44.0                      | $B_1$ |
| 1                                  | -1 | 7  | 1  | 1  | $\theta^\wedge_2$ |   | 50.0                      | $B_2$ |
| -1                                 | 1  | 1  | 7  | -1 | $\theta^\wedge_3$ |   | 2.0                       | $B_3$ |
| -1                                 | 1  | 1  | -1 | 7  | $\theta^\wedge_4$ |   | 0.0                       | $B_4$ |

正規方程式の両辺に逆行列  $(\mathbf{X}^T \mathbf{X})^{-1}$  を掛け，正規方程式を解くと表 7 の結果を得る．得られた線形モデル式は，

$$\left. \begin{aligned} \hat{y}_i &= \hat{\theta}_0 x_{0,i} + \hat{\theta}_1 x_{1,i} + \hat{\theta}_2 x_{2,i} + \hat{\theta}_3 x_{4,i} + \hat{\theta}_4 x_{7,i} \\ &= 19.17x_{0,i} + 3.583x_{1,i} + 4.333x_{2,i} + 2.167x_{4,i} + 1.917x_{7,i} \end{aligned} \right\} \quad (10)$$

となる．なお，行列表記中の  $(\mathbf{X}^T \mathbf{X})^{-1}(c_{kk'})$  の  $c_{kk'}$  は， $(\mathbf{X}^T \mathbf{X})^{-1}$  の各要素を表す場合に使用する．

表 7 欠測値がある直交表に対する正規方程式の解

| $(X^T X)^{-1} (c_{kk'})$ |        |        |        |        | $X^T X (a_{kk'})$ |    |    |    |    | $\theta^\wedge$   |   |                      |                 |        | $(X^T X)^{-1} (c_{kk'})$ |        |       |  |  | $X^T Y$ |  |  |  |  |
|--------------------------|--------|--------|--------|--------|-------------------|----|----|----|----|-------------------|---|----------------------|-----------------|--------|--------------------------|--------|-------|--|--|---------|--|--|--|--|
| 0.167                    | -0.042 | -0.042 | 0.042  | 0.042  | 7                 | 1  | 1  | -1 | -1 | $\theta^\wedge_0$ | = | 0.167                | -0.042          | -0.042 | 0.042                    | 0.042  | 138.0 |  |  |         |  |  |  |  |
| -0.042                   | 0.167  | 0.042  | -0.042 | -0.042 | 1                 | 7  | -1 | 1  | 1  | $\theta^\wedge_1$ |   | -0.042               | 0.167           | 0.042  | -0.042                   | -0.042 | 44.0  |  |  |         |  |  |  |  |
| -0.042                   | 0.042  | 0.167  | -0.042 | -0.042 | 1                 | -1 | 7  | 1  | 1  | $\theta^\wedge_2$ |   | -0.042               | 0.042           | 0.167  | -0.042                   | -0.042 | 50.0  |  |  |         |  |  |  |  |
| 0.042                    | -0.042 | -0.042 | 0.167  | 0.042  | -1                | 1  | 1  | 7  | -1 | $\theta^\wedge_3$ |   | 0.042                | -0.042          | -0.042 | 0.167                    | 0.042  | 2.0   |  |  |         |  |  |  |  |
| 0.042                    | -0.042 | -0.042 | 0.042  | 0.167  | -1                | 1  | 1  | -1 | 7  | $\theta^\wedge_4$ |   | 0.042                | -0.042          | -0.042 | 0.042                    | 0.167  | 0.0   |  |  |         |  |  |  |  |
|                          |        |        |        |        |                   |    |    |    |    | $\theta^\wedge$   |   | $(X^T X)^{-1} X^T Y$ |                 |        |                          |        |       |  |  |         |  |  |  |  |
|                          |        |        |        |        |                   |    |    |    |    | $\theta^\wedge_0$ | = | 19.17                | $\mu^\wedge$    |        |                          |        |       |  |  |         |  |  |  |  |
|                          |        |        |        |        |                   |    |    |    |    | $\theta^\wedge_1$ |   | 3.583                | $\alpha^\wedge$ |        |                          |        |       |  |  |         |  |  |  |  |
|                          |        |        |        |        |                   |    |    |    |    | $\theta^\wedge_2$ |   | 4.333                | $\beta^\wedge$  |        |                          |        |       |  |  |         |  |  |  |  |
|                          |        |        |        |        |                   |    |    |    |    | $\theta^\wedge_3$ |   | 2.167                | $\gamma^\wedge$ |        |                          |        |       |  |  |         |  |  |  |  |
|                          |        |        |        |        |                   |    |    |    |    | $\theta^\wedge_4$ |   | 1.917                | $\delta^\wedge$ |        |                          |        |       |  |  |         |  |  |  |  |

表 8 に示すよう正規方程式の解  $\hat{\theta}$  をデザイン行列  $\mathbf{X}$  に掛け、予測値  $\hat{y}_i$  の計算する。元のデータ  $y_i$  と  $\hat{y}_i$  の差  $(y_i - \hat{y}_i)$  の平方和  $S_e = 4.3333$  が計算されている。さらに、自由度  $df = 7 - 4 - 1 = 2$  で割った誤差分散  $\hat{\sigma}^2 = S_e / df = 2.1667$  が求められている。

表 8 予測値  $\hat{\mathbf{y}}$  の推定および誤差分散  $\hat{\sigma}^2$  の推定

|     | $x_0$ | $x_1$ | $x_2$ | $x_4$ | $x_7$ | 予測値 $\hat{\mathbf{y}}$ の計算 |                    |                    | 誤差分散 $\hat{\sigma}^2$ の推定 |                    |   |
|-----|-------|-------|-------|-------|-------|----------------------------|--------------------|--------------------|---------------------------|--------------------|---|
| No. | $\mu$ | A     | B     | C     | D     | $\hat{\theta}$             | $\hat{\mathbf{y}}$ | $\mathbf{y}$       |                           | $\hat{\mathbf{y}}$ | $\varepsilon = \mathbf{y} - \hat{\mathbf{y}}$ |
| 1   | 1     | 1     | 1     | 1     | 1     | 19.167                     | = 31.167           | 32                 | -                         | 31.167             | = 0.8333                                      |
| 2   | 1     | 1     | 1     | -1    | -1    | 3.583                      | 23.000             | 23                 |                           | 23.000             | 0.0000  |
| 3   | 1     | 1     | -1    | 1     | -1    | 4.333                      | 18.667             | 19                 |                           | 18.667             | 0.3333  |
| 4   | 1     | 1     | -1    | -1    | 1     | 2.167                      | 18.167             | 17                 |                           | 18.167             | -1.1667                                       |
| 5   | 1     | -1    | 1     | 1     | -1    | 1.917                      | 20.167             | 19                 |                           | 20.167             | -1.1667                                       |
| 6   | 1     | -1    | 1     | -1    | 1     |                            | 19.667             | 20                 |                           | 19.667             | 0.3333  |
| 8   | 1     | -1    | -1    | -1    | -1    |                            | 7.167              | 8                  |                           | 7.167              | 0.8333  |
|     |       |       |       |       |       |                            |                    | $n =$              | 7                         | $S_e =$            | 4.3333  |
|     |       |       |       |       |       |                            |                    | $p =$              | 4                         | $\hat{\sigma}^2 =$ | 2.1667  |
|     |       |       |       |       |       |                            |                    | $df = n - p - 1 =$ | 2                         | $\hat{\sigma} =$   | 1.4720  |

ここまでの行列計算により、表 5 に示した Excel の回帰分析による分散分析表の残差の分散  $\hat{\sigma}^2 = 2.1667$  が得られている。各種の推定のためのパラメータの共分散行列  $\Sigma(\hat{\theta})$  は、表 9 に示すように積和行列  $\mathbf{X}^T \mathbf{X}$  の逆行列  $(\mathbf{X}^T \mathbf{X})^{-1}$  に誤差分散  $\hat{\sigma}^2 = 2.1667$  を掛けて求めることができる。パラメータの共分散行列  $\Sigma(\hat{\theta})$  の対角要素の平方根が、表 9 に示す回帰パラメータ（係数）の標準誤差  $SE$  となることが確認される。

表 9 パラメータの共分散行列  $\Sigma(\hat{\theta})$

|                  | $(\mathbf{X}^T \mathbf{X})^{-1} (c_{kk'})$ |                  |                  |                  |                  | $\hat{\sigma}^2$ | パラメータの共分散行列 $\Sigma(\hat{\theta})$ |        |        |        |        | $SE$   |
|------------------|--|------------------|------------------|------------------|------------------|------------------|------------------------------------|--------|--------|--------|--------|--------|
| $\hat{\theta}_0$ | 0.167                                      | -0.042           | -0.042           | 0.042            | 0.042            | 2.1667           | 0.361                              | -0.090 | -0.090 | 0.090  | 0.090  | 0.6009 |
| $\hat{\theta}_1$ | -0.042                                     | 0.167            | 0.042            | -0.042           | -0.042           |                  | -0.090                             | 0.361  | 0.090  | -0.090 | -0.090 | 0.6009 |
| $\hat{\theta}_2$ | -0.042                                     | 0.042            | 0.167            | -0.042           | -0.042           |                  | -0.090                             | 0.090  | 0.361  | -0.090 | -0.090 | 0.6009 |
| $\hat{\theta}_3$ | 0.042                                      | -0.042           | -0.042           | 0.167            | 0.042            |                  | 0.090                              | -0.090 | -0.090 | 0.361  | 0.090  | 0.6009 |
| $\hat{\theta}_4$ | 0.042                                      | -0.042           | -0.042           | 0.042            | 0.167            |                  | 0.090                              | -0.090 | -0.090 | 0.090  | 0.361  | 0.6009 |
|                  | $\hat{\theta}_0$                           | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $\hat{\theta}_4$ |                  |                                    |        |        |        |        |        |

ここに示したのは、データ数が 7、変数の数が 4 と小さなデータであり、行列計算の過程を Excel の表として示したのであるが、データ数が多くても、また変数の数が多くとも、Excel の行列関数でデザイン行列の範囲の設定を変えるだけで、計算手順は同じである。

変数の数が 2 以上の場合の回帰分析は、「重回帰分析」と称されるが、ここで示した解析方法と同じである。身近な事例で、切片を加えたデザイン行列を Excel シート上に設定し、自ら Excel の行列関数を用いた重回帰分析を行ない、各種の推定にチャレンジすることにより、線形推定論の理解が深まるであろう。



さて、Excel の回帰分析には、各因子に対する分散分析表の出力がない。欠測値があるので平方和の分解による解析方法が使えないが、2水準系の直交表ならば、Excel の回帰パラメータの出力結果から、各因子の平方和を求めることができる。

表 5 の Excel の回帰分析による結果は、表 7 に示した正規

方程式の解と一致する。したがって、欠測値がある直交表の解析には、ごく一般的な回帰分析を適用すればよいことが分かる。分散分析表を必要とする場合には、パラメータの  $SE$  と誤差分散  $\hat{\sigma}^2 = 2.116$  を用いて、各因子の平方和を計算することができる。自由度 1 の分散分析表の  $F$  値は、パラメータの  $t$  値との関係が  $F = t^2$  となるので、次式が導かれ

$$t_A^2 = F_A = \frac{S_A / 1}{\hat{\sigma}^2} \quad (11)$$

因子 A について次のように平方和  $S_A$  を求めることができる。

$$\left. \begin{aligned} S_A &= t_A^2 \hat{\sigma}^2 \\ &= 5.9630^2 \times 2.1667 \\ &= 77.0417 \end{aligned} \right\} \quad (12)$$

参考のために、JMP の「モデルのあてはめ」を用いた結果を表 10 に示す。因子 A についての  $F$  検定の結果から、式 (12) で正しく計算できていることが確かめられる。

表 10 JMP の「モデルのあてはめ」による線形モデル

| パラメータ推定値 |         |        |         |         | 効果の検定 |     |          |          |         |         |
|----------|---------|--------|---------|---------|-------|-----|----------|----------|---------|---------|
| 項        | 推定値     | 標準誤差   | t値      | Prob> t | 要因    | 自由度 | 平方和      | 平均平方     | F値      | Prob>F  |
| 切片       | 19.1667 | 0.6009 | 31.8953 | 0.0010* | A     | 1   | 77.0417  | 77.0417  | 35.5577 | 0.0270* |
| A[1]     | 3.5833  | 0.6009 | 5.9630  | 0.0270* | B     | 1   | 112.6667 | 112.6667 | 52.0000 | 0.0187* |
| B[1]     | 4.3333  | 0.6009 | 7.2111  | 0.0187* | C     | 1   | 28.1667  | 28.1667  | 13.0000 | 0.0691  |
| C[1]     | 2.1667  | 0.6009 | 3.6056  | 0.0691  | D     | 1   | 22.0417  | 22.0417  | 10.1731 | 0.0858  |
| D[1]     | 1.9167  | 0.6009 | 3.1895  | 0.0858  |       |     |          |          |         |         |

スライド 24 再掲

スライド 11 再掲

## 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  $\sigma^2$  としたときに、 $F_A = (S_A / 1) / \sigma^2$  の関係から  $S_A$  を  

$$S_A = t_A^2 \sigma^2$$
 によって計算し、分散分析表を作成する。

2021年12月16&17日 高橋行雄

31

スライド 32 に示すように、他の因子についても Excel で各パラメータの  $t$  値から平方和を計算し、分散分析表にまとめる。ただし、それぞれの因子の平方和を加え、さらに残差の平方和を加えると

$$\left. \begin{aligned} S'_T &= (S_A + S_B + S_C + S_D) + S_e \\ &= (77.0417 + 112.6667 + 28.1667 + 22.0417) + 4.3333 \\ &= 239.9167 + 4.3333 = 244.2500 \end{aligned} \right\} \quad (13)$$

$S'_T = 244.2500$  となり、スライド 30 および表 5 に示した変動（平方和）の合計  $S_T = 307.4286$  に一致しないので、一般的な分散分析表の形式で表わすことができない。そこで、各因子の平方和の列をずらして表示している。

このようにして作成された分散分析表から読み取れることは、各因子が統計的に有意か有意ではないかだけであり、どのような差なのかを知るためには、各因子の水準平均を計算し、さらにそれらの水準平均の 95%信頼区間を別途計算し、何らかのグラフ表示をする必要がある。したがって、分散分析表の作成は、解析の第 1 歩目にしか過ぎない。

## スライド 32

# 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 |            |                           |
|--------------------|----------|---------------|----------------|------------|--------|--------|------------|---------------------------|
| 要因                 | 自由度 $df$ | 変動<br>平方和 $S$ | 分散<br>平均平方 $V$ | 分散比<br>$F$ | $p$ 値  | $t$ 値  | $\sigma^2$ | $t^2 \sigma^2$<br>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      |                |            |        |        |            |                           |

## 各種のダミー変数をそのまま用いた線形モデル

(1, -1) 対比型のダミー変数を用いた線形モデルについて回帰分析を適用したのであるが、(1, 2) 型の直交表のまま、連続変数として Excel で回帰分析を適用したらどのような結果になるだろうか。

スライド 33 に示すように分散分析表の結果は、スライド 30 と全く同じ結果となる。ただし、パラメータの推定値は、まったく異なるが、切片を除いて  $p$  値は同じとなる。どうしてなのだろうか？ 因子 A の推定値  $\hat{\theta}_1 = -7.1667$  は傾きであり、第 1 水準を数値の 1、第 2 水準を数値の 2 としたので、第 1 水準に対する第 2 水準の差の推定値となっている。

スライド 30 に示した (1, -1) 対比型の場合は、第 2 水準の数値が -1 であり、因子 A の第 1 水準の 1 に対する傾きとして 3.5833 となっている。第 2 水準に対する第 1 水準の差は、 $2 \times 3.5833 = 7.1666$  となり、符号を変えると (1, 2) 型に一致する。それに伴い標準誤差も 0.6009 の 2 倍の 1.2019 となり、結局  $t$  値としては同じとなる。では、切片の 55.1667 は、どのように解釈したら良いのであろうか。全ての変数（因子）が 0 の場合の推定値ではあるが、推定には役に立たない。スライド 30 の (1, -1) 対比型の場合の切片は、構造 (DE) モデルの平均  $\mu$  の推定値  $\hat{\mu} = 19.1667$  となるのとは異なる。

スライド 33

## (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 X |                |                |                |                |                | Y              |                |    | 分散分析表 |     |         |         |         |
|----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----|-------|-----|---------|---------|---------|
|          | x <sub>0</sub> | x <sub>1</sub> | x <sub>2</sub> | x <sub>4</sub> | x <sub>7</sub> |                |                |    |       | 自由度 | 変動      | 分散      | 分散比     |
| No.      | μ              | A              | B              | C              | D              | θ              | ε              | y  |       |     |         |         |         |
| 1        | 1              | 1              | 1              | 1              | 1              | θ <sub>0</sub> | ε <sub>1</sub> | 32 | 回帰    | 4   | 303.095 | 75.7738 | 34.9725 |
| 2        | 1              | 1              | 1              | 2              | 2              | θ <sub>1</sub> | ε <sub>2</sub> | 23 | 残差    | 2   | 4.333   | 2.1667  |         |
| 3        | 1              | 1              | 2              | 1              | 2              | θ <sub>2</sub> | ε <sub>3</sub> | 19 | 合計    | 6   | 307.429 |         |         |
| 4        | 1              | 1              | 2              | 2              | 1              | θ <sub>3</sub> | ε <sub>4</sub> | 17 |       |     |         |         |         |
| 5        | 1              | 2              | 1              | 1              | 2              | θ <sub>4</sub> | ε <sub>5</sub> | 19 |       |     |         |         |         |
| 6        | 1              | 2              | 1              | 2              | 1              |                |                | 20 |       |     |         |         |         |
| 8        | 1              | 2              | 2              | 2              | 2              |                |                | 8  |       |     |         |         |         |
|          | μ              | a              | b              | c              | abc            |                |                |    |       |     |         |         |         |

|    | 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。

結局、(1, -1) 型, (1, 2) 型, (0, 1) 型のどの直交表の数値をダミー変数として用い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。スライドには示さなかったが、SAS/GLM プロシジャでは、表 12 に示すように最後の水準を 0 とするような (1, 0) 型のダミー変数を用いている。

### (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  $\mu$  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。

2021年12月16&amp;17日 高橋行雄

34

どのようなダミー変数を用いても、分散分析表はどれでも同じで、異なるのは切片の推定値であり、 $p$  値も異なる。(1, -1) 型の場合にのみ切片が、一般平均としての  $\hat{\mu}$  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し、(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り、扱い易いと思われるかもしれないが、各因子の水準平均（最小 2 乗平均）をどのようにして求めたら良いのであろうか。各因子の水準に対して算術平均で良しとするならば、簡単に求めることができるのだが。

(0, 1) 型のダミー変数とした場合の結果を表 11 に示す。切片 31.1667 は、全ての因子の第 1 水準が 0 の場合の推定値であり、スライド 12 に示した現行水準の推定値に一致するのであるが、 $t$  検定の結果は無意味である。

表 11 (0, 1) 型のダミー変数を与えた場合の解析結果

| デザイン行列 $X$ |       |       |       |       |       | $Y$        |                 |     | 分散分析表 |         |         |         |         |
|------------|-------|-------|-------|-------|-------|------------|-----------------|-----|-------|---------|---------|---------|---------|
|            | $x_0$ | $x_1$ | $x_2$ | $x_4$ | $x_7$ |            |                 |     | 自由度   | 変動      | 分散      | 分散比     |         |
| No.        | $\mu$ | A     | B     | C     | D     | $\theta$   | $\varepsilon$   | $y$ | 回帰    | 4       | 303.095 | 75.7738 | 34.9725 |
| 1          | 1     | 0     | 0     | 0     | 0     | $\theta_0$ | $\varepsilon_1$ | 32  | 残差    | 2       | 4.333   | 2.1667  |         |
| 2          | 1     | 0     | 0     | 1     | 1     | $\theta_1$ | $\varepsilon_2$ | 23  | 合計    | 6       | 307.429 |         |         |
| 3          | 1     | 0     | 1     | 0     | 1     | $\theta_2$ | $\varepsilon_3$ | 19  |       |         |         |         |         |
| 4          | 1     | 0     | 1     | 1     | 0     | $\theta_3$ | $\varepsilon_4$ | 17  |       | 係数      | 標準誤差    | $t$     | $P$ -値  |
| 5          | 1     | 1     | 0     | 0     | 1     | $\theta_4$ | $\varepsilon_5$ | 19  | 切片    | 31.1667 | 1.2019  | 25.9322 | 0.0015  |
| 6          | 1     | 1     | 0     | 1     | 0     |            |                 | 20  | A     | -7.1667 | 1.2019  | -5.9630 | 0.0270  |
| 8          | 1     | 1     | 1     | 1     | 1     |            |                 | 8   | B     | -8.6667 | 1.2019  | -7.2111 | 0.0187  |
|            | $\mu$ | $a$   | $b$   | $c$   | $abc$ |            |                 |     | C     | -4.3333 | 1.2019  | -3.6056 | 0.0691  |
|            |       |       |       |       |       |            |                 |     | D     | -3.8333 | 1.2019  | -3.1895 | 0.0858  |

(1, 2) 型の場合の各因子の水準平均（最小 2 乗平均）は、どのようにして求めたら良いのであろうか。スライド 35 に計算方法を示す。全ての因子のダミー変数の平均値 1.50 を用いて推定値  $\hat{\mu}$  を次式

$$\left. \begin{aligned} \hat{\mu} &= \hat{\theta}_0 \times 1 + \hat{\theta}_1 \times 1.5 + \hat{\theta}_2 \times 1.5 + \hat{\theta}_3 \times 1.5 + \hat{\theta}_4 \times 1.5 \\ &= 55.1667 + (-7.1667 - 8.6667 - 4.3333 - 3.8333) \times 1.5 \\ &= 19.1667 \end{aligned} \right\} \quad (14)$$

で求めると  $\hat{\mu}=19.1667$  が得られ、スライド 30 に示した切片に一致する。その  $\hat{\mu}$  対し、各因子のパラメータの推定値の 1/2 をプラス・マイナスすれば、水準平均の推定値（最小 2 乗平均）を求めることができる。実際に計算すると、因子 A の第 1 水準の最小 2 乗平均は、

$$\left. \begin{aligned} \hat{y}_{A_1} &= \hat{\mu} - \frac{\hat{\theta}_1}{2} \\ &= 19.1667 - \frac{-7.1667}{2} = 22.7500 \end{aligned} \right\}, \quad \left. \begin{aligned} \hat{y}_{A_2} &= \hat{\mu} + \frac{\hat{\theta}_1}{2} \\ &= 19.1667 + \frac{-7.1667}{2} = 15.5833 \end{aligned} \right\} \quad (15)$$

となり、スライド 27 の結果に一致する。どのようなダミー変数を与えても、このように目的とする推定値は得られるのであるが、めんどろで説明しにくい計算が必要となるので、直交表の解析に際しては (1, -1) 対比型が扱い易い。

スライド 35

## (1, 2) 型の場合の水準平均

| (1, 2) 型 |         | $\mu^\wedge$ 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 |
|          |         | $\mu^\wedge$     | 19.1667  |      |         |      |         |

- ◆  $\mu^\wedge$  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、 $\mu^\wedge$  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。

## 7. 各種の推定値に対する 95%信頼区間の算出

スライド 36

### 従来の分散分析の場合

従来の分散分析の場合は、平方和の分解により誤差分散  $\sigma^2$  を求め、 $F$  検定の分母として用いて分散分析表を作成する。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  $n_e$  から求められたかを勘案し、 $\sigma^2 / n_e$  によって分散を計算し、95%信頼区間を計算する手順が定式化されている。

### 5. 各種の推定値に対する 95%信頼区間の算出

2021年12月16&17日 高橋行雄

36

最適水準の組合せ、現行水準の組合せに対しては、推定に考慮した因子の水準数から、いわゆる「田口の式」または、「伊奈の式」により有効反復数  $n_e$  を求め分散を計算し、95%信頼区間を計算する。

スライド 37

「田口の式」および「伊奈の式」により、因子が互いに直交している直交表などの実験に対し、複雑なパラメータに関する線形和の分散が手軽に求められるようになったことは、画期的であった。ただし、欠測値がある場合、繰り返しが不揃いの場合などに適用できないとの制約があるので、線形モデルによる解析には、適用できないとの制約が付きまとっている。

### 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  $\sigma^2$  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  $\sigma^2$  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。

2021年12月16&17日 高橋行雄

37

スライド 12 に示したように、JMP の「予測プロファイル」を用いて「現行水準」および「最適水準」の推定と 95%信頼区間が求められることを示した。また、スライド 13 では、「最適水準」と「現行水準」の差について「カスタム検定」を用いて標準誤差が算出できることを示した。JMP での「モデルのあてはめ」における各種の推定値の 95%信頼区間は、どのような方法で求められているのだろうか。Excel を用いて確かめてみる。

JMP の「モデルのあてはめ」は、線形モデルの解法によっている。スライド 38 に示すように線形モデルでは、デザイン行列  $X$ 、反応変数の列ベクトル  $Y$  を用いて推定値  $\hat{\theta}$  (列ベクトル) を求め、誤差分散  $\sigma^2$  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、 $c_{kk} \sigma^2$  となる。推定値  $\hat{\theta}_k$  と  $\hat{\theta}_{k'}$  の共分散  $Cov(\hat{\theta}_k, \hat{\theta}_{k'})$  は、 $c_{kk'} \sigma^2$  であり、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  $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$  とする。

## 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  $\sigma^2$  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、 $c_{kk} \sigma^2$  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、 $c_{kk'} \sigma^2$  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  $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$  とする。

2021年12月16&17日 高橋行雄

38

スライド 39 に、Excel の行列関数を用いたパラメータの共分散行列  $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$  の計算手順を示す。ここでは、誤差分散が  $\sigma^2 = 2.1667$  がすでに求められているとしている。

## パラメータの共分散行列 $\Sigma(\hat{\theta})$

|       | 転置行列 $X^T$ |    |    |    |    |    |    | デザイン行列 $X$ |    |    |    |    | デザイン行列の積 $X^T X (a_{kk'})$ |    |    |    |    |
|-------|------------|----|----|----|----|----|----|------------|----|----|----|----|----------------------------|----|----|----|----|
|       | 1          | 2  | 3  | 4  | 5  | 6  | 8  | $\mu$      | A  | B  | C  | D  |                            |    |    |    |    |
| $\mu$ | 1          | 1  | 1  | 1  | 1  | 1  | 1  | 1          | 1  | 1  | 1  | 1  | 7                          | 1  | 1  | -1 | -1 |
| A     | 1          | 1  | 1  | 1  | -1 | -1 | -1 | 1          | 1  | 1  | -1 | -1 | 1                          | 7  | -1 | 1  | 1  |
| B     | 1          | 1  | -1 | -1 | 1  | 1  | -1 | 1          | 1  | -1 | 1  | -1 | 1                          | -1 | 7  | 1  | 1  |
| C     | 1          | -1 | 1  | -1 | 1  | -1 | -1 | 1          | 1  | -1 | -1 | 1  | -1                         | 1  | 1  | 7  | -1 |
| D     | 1          | -1 | -1 | 1  | -1 | 1  | -1 | 1          | -1 | 1  | 1  | -1 | -1                         | 1  | 1  | -1 | 7  |

デザイン行列の積の逆行列

$\sigma^2 = 2.1667$

|                  | $(X^T X)^{-1} (c_{kk'})$ |        |        |        |        | パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$ |         |         |         |         |
|------------------|--------------------------|--------|--------|--------|--------|--|---------|---------|---------|---------|
| $\hat{\theta}_0$ | 0.167                    | -0.042 | -0.042 | 0.042  | 0.042  | 0.3611   | -0.0903 | -0.0903 | 0.0903  | 0.0903  |
| $\hat{\theta}_1$ | -0.042                   | 0.167  | 0.042  | -0.042 | -0.042 | -0.0903  | 0.3611  | 0.0903  | -0.0903 | -0.0903 |
| $\hat{\theta}_2$ | -0.042                   | 0.042  | 0.167  | -0.042 | -0.042 | -0.0903  | 0.0903  | 0.3611  | -0.0903 | -0.0903 |
| $\hat{\theta}_3$ | 0.042                    | -0.042 | -0.042 | 0.167  | 0.042  | 0.0903   | -0.0903 | -0.0903 | 0.3611  | 0.0903  |
| $\hat{\theta}_4$ | 0.042                    | -0.042 | -0.042 | 0.042  | 0.167  | 0.0903   | -0.0903 | -0.0903 | 0.0903  | 0.3611  |

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

2021年12月16&17日 高橋行雄

39



パラメータの共分散行列  $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$  を使うことにより、あらゆる線形和（最小 2 乗平均）に対する分散の計算を可能にする。JMP でもこのパラメータの共分散行列  $\Sigma(\hat{\theta})$  を用い、各種の 95%信頼区間の計算を内部で行っている。

単回帰分析は、線形モデルの最も基本的な事例であるが、偏差平方和  $S_{xx}$ ,  $S_{xy}$ ,  $S_{yy}$  を巧妙に組み立て、行列計算によらず、様々な推定ができよう工夫されていて、極度に標準化されている。そのために、変数が 2 以上となる場合への拡張性が閉ざされてしまい、ガラスの天井のごとく、さらなる応用がきかないような妨げになっている。

身近にある表計算ソフト Excel で行列計算が手軽にできるようになり、線形モデルの理論に基づいた解析ができるようになったことは喜ばしい。しかし、参考になるテキスト無しで、自らスライド 38 に示した手順による Excel での行列計算は、敷居が高すぎて手につかないと思われる。スライド 39 に示したパラメータの共分散行列  $\Sigma(\hat{\theta})$  の計算について Excel の行列関数を用いた計算式

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

を示したのであるが、行列計算の経験がない方は、

高橋行雄 (2021), 最尤法によるポアソン回帰分析入門, 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>

をダウンロードして一読され、添付されている Excel の計算シートを活用されたい。

また、重回帰分析に関しては、新村 (1983a,b) の「行列表現による重回帰分析 (1) および (2)」を参照されたい。この論文は、応答変数  $y$  と説明変数  $x_1$  から  $x_4$  までの 4 個の説明変数からなる 7 個の観測データを用いて行列表現による重回帰分析について詳しく解説されている。また、偏差平方和ベースの重回帰分析について、「規準化データによる重回帰」として例示と文献の引用もあり、時代的な背景の理解に役に立つ。この時代の重回帰分析は、逆行列を求めるために掃き出し計算が用いられており、奥野・久米・芳賀・吉澤 (1981) 「多変量解析法」にも丁寧な解説がある。本書では、Excel の `Minverse()` 関数を使うことを前提にし、逆行列の計算はブラック・ボックスのままにしてきた。詳しくは、新村 (1983c) , 「重回帰分析における掃き出し演算子」を参照のこと。なお、私にとっても掃き出し計算による逆行列の計算は、Fortran を使っていたころ慣れ親しんできたので、なつかしく思うのであるが、Excel の基本の計算機能だけでは逆行列の計算は実現できなかったもので、深入りはしない。



## データの相関行列および共分散行列

パラメータの共分散行列  $\Sigma(\hat{\theta})$  は、データの共分散行列  $\Sigma(\mathbf{x})$  もあり、紛らわしい。一般的に分散共分散行列、または、共分散行列と言ったときには、データの共分散行列  $\Sigma(\mathbf{x})$  が念頭におかれる。2 つの変数  $x$  と  $y$  の共分散  $Cov(x, y)$  は、回帰分析のパラメータ推定で  $S_{xy}$  を次式で計算し、自由度  $n-1$  で割って求められる。

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad Cov(x, y) = \frac{S_{xy}}{n-1} \quad (16)$$

スライド 40 は、 $7 \times 5$  デザイン行列  $\mathbf{X}$  からデータの共分散行列の作成を行列計算で行った結果である。最初にデザイン行列  $\mathbf{X}$  の行方向に Average() 関数を用いて平均値  $\bar{\mathbf{x}}$  を計算し、

$$\Sigma(\mathbf{x}) = (\mathbf{X} - \bar{\mathbf{x}})^T (\mathbf{X} - \bar{\mathbf{x}}) / (7-1) \quad (17)$$

によって  $5 \times 5$  の共分散行列  $\Sigma(\mathbf{x})$  を計算している。Excel の計算式では、

$$\Sigma(\mathbf{x}) = \text{Mmult}(\text{Transpose}((C7:G13)-(J5:N5)), (C7:G13)-(J5:N5)) / 6$$

(  $\mathbf{X}$  の範囲 -  $\bar{\mathbf{x}}$  の範囲 )

として計算した結果である。データの相関係数行列は、個別の分散  $\sigma^2$  を Var.s() 関数で計算

スライド 40

## データの相関 $R(\mathbf{x})$ ・共分散行列 $\Sigma(\mathbf{x})$

|                     |    |    |    |    |   | $\mu$   | A       | B       | C       | D       |
|---------------------|----|----|----|----|---|---------|---------|---------|---------|---------|
| デザイン行列 $\mathbf{X}$ |    |    |    |    | 平均 $\bar{\mathbf{x}}$   | 1.0000  | 0.1429  | 0.1429  | -0.1429 | -0.1429 |
| $\mu$               | A  | B  | C  | D  |   |         |         |         |         |         |
| 1                   | 1  | 1  | 1  | 1  | 共分散行列 $\Sigma(\mathbf{x}) =$<br>(データの)<br>$(\mathbf{X} - \bar{\mathbf{x}})^T (\mathbf{X} - \bar{\mathbf{x}})$ | 0       | 0       | 0       | 0       | 0       |
| 1                   | 1  | 1  | -1 | -1 |   | 0       | 1.14286 | -0.1905 | 0.19048 | 0.19048 |
| 1                   | 1  | -1 | 1  | -1 |   | 0       | -0.1905 | 1.14286 | 0.19048 | 0.19048 |
| 1                   | 1  | -1 | -1 | 1  |   | 0       | 0.19048 | 0.19048 | 1.14286 | -0.1905 |
| 1                   | -1 | 1  | 1  | -1 |   | 0       | 0.19048 | 0.19048 | -0.1905 | 1.14286 |
| 1                   | -1 | 1  | -1 | 1  |   |         |         |         |         |         |
| 1                   | -1 | -1 | -1 | -1 | 個別の分散 $\sigma^2 =$  | 0       | 1.14286 | 1.14286 | 1.14286 | 1.14286 |
|                     |    |    |    |    |   | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! | #DIV/0! |
|                     |    |    |    |    | 相関行列 $R(\mathbf{x}) =$  | #DIV/0! | 1       | -0.1667 | 0.16667 | 0.16667 |
|                     |    |    |    |    | $\Sigma(\mathbf{x}) / \sigma / \sigma^T$  | #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       |

しておき、 $R(x) = \Sigma(x) / \sqrt{\hat{\sigma}^2} / (\sqrt{\hat{\sigma}^2})^T$  で計算した結果である。逆に相関行列  $R(x)$  から共分散行列  $\Sigma(\hat{\theta})$  を求めるためには、 $\Sigma(\hat{\theta}) = R(x) * \sqrt{\hat{\sigma}^2} * (\sqrt{\hat{\sigma}^2})^T$  とすればよい。

## パラメータの相関行列から共分散行列の算出

パラメータの共分散行列  $\Sigma(\hat{\theta})$  については、すでにスライド 39 で示した。ただし、JMP の「モデルのあてはめ」では、内部の計算で使われているが、出力することができない。ただし、スライド 24 に示したようにパラメータの相関行列  $R(\hat{\theta})$  は、得ることができるので、スライド 41 に示したようにパラメータの推定値の標準誤差をベクトル  $SE$  として Excel に取り出し、パラメータの共分散行列  $\Sigma(\hat{\theta})$  は、

$$\Sigma(\hat{\theta}) = R(\hat{\theta}) * SE * SE^T \quad (18)$$

として計算できる。スライド 41 には、Excel の行列計算で求めたパラメータの共分散行列  $\Sigma(\hat{\theta})$  から、パラメータの相関行列  $R(\hat{\theta})$  を

$$R(\hat{\theta}) = \Sigma(\hat{\theta}) / SE / SE^T \quad (19)$$

で計算し、JMP で出力された「推定値の相関」に一致することを確認した結果である。

## スライド 41

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

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

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

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 $\Sigma(\hat{\theta})$ |         |         |         |         |
|------------|------------------------------------|---------|---------|---------|---------|
| $\theta^0$ | 0.3611                             | -0.0903 | -0.0903 | 0.0903  | 0.0903  |
| $\theta^1$ | -0.0903                            | 0.3611  | 0.0903  | -0.0903 | -0.0903 |
| $\theta^2$ | -0.0903                            | 0.0903  | 0.3611  | -0.0903 | -0.0903 |
| $\theta^3$ | 0.0903                             | -0.0903 | -0.0903 | 0.3611  | 0.0903  |
| $\theta^4$ | 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 $R(\hat{\theta})$       |         |         |         |         |
| $\theta^0$ | 1                                  | -0.2500 | -0.2500 | 0.2500  | 0.2500  |
| $\theta^1$ | -0.2500                            | 1       | 0.2500  | -0.2500 | -0.2500 |
| $\theta^2$ | -0.2500                            | 0.2500  | 1       | -0.2500 | -0.2500 |
| $\theta^3$ | 0.2500                             | -0.2500 | -0.2500 | 1       | 0.2500  |
| $\theta^4$ | 0.2500                             | -0.2500 | -0.2500 | 0.2500  | 1       |

## パラメータの推定値の標準誤差 SE

回帰分析では、パラメータの推定値を標準誤差  $SE$  で割った  $t$  値から  $p$  値を求め、統計的判断を行なうのを常としている。平方和の分解に基づいた伝統的な直交表の解析では、水準平均あるいは現行水準の組合せ平均などに基づいて推定を行なうので、パラメータの推定値を用いることはない。

単回帰分析で推定されたパラメータの推定値に対する標準誤差  $SE$  の計算式は、多くの書物で示されているが、2 変数以上の回帰分析に対しては、統計ソフト任せでブラック・ボックス的である。スライド 38 で、「デザイン行列の積  $(X^T X)$  の逆行列  $(X^T X)^{-1}$  の要素を  $c_{kk'}$  とし、推定値  $\hat{\theta}_k$  の分散  $Var(\hat{\theta}_k)$  は、 $c_{kk} \hat{\sigma}^2$  である。」ことを示した。JMP の出力にあるパラメータの推定値の  $SE$  が、どのようにして求められたのかを実際に計算して確かめてみる。スライド 41 に示すように、切片の  $SE=0.6009$  は、パラメータの共分散行列  $\Sigma(\hat{\theta})$  から、

$$SE_{\text{切片}} = \sqrt{Var(c_{00} \hat{\sigma}^2)} = \sqrt{0.3611} = 0.6009 \quad (20)$$

として求められている。他の因子の  $SE$  も同様に  $\Sigma(\hat{\theta})$  対角要素の分散の平方根で計算されている。

スライド 42

## 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 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$ |         |         |         |         |
| 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))\*σ<sup>2</sup>

- ◆ 推定値の標準誤差  $SE$  は、パラメータの共分散行列  $\Sigma(\hat{\theta})$  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。

## 線形和で求めた推定値の 95%信頼区間

欠測値がない場合には、各因子の水準平均と、推定された回帰パラメータを用いた推定値は一致するが、欠測値がある場合には一致しない。因子 A の場合であれば、

$$A_1 \text{の平均} = (32 + 23 + 19 + 17) / 4 = 22.750$$

$$A_2 \text{の平均} = (19 + 20 + 8) / 3 = 15.667$$

であり、回帰パラメータを用いた場合は、スライド 43 に示すように

$$A_1 \text{の推定値} = \hat{\theta}_0 + \hat{\theta}_1 = \hat{\mu} + \hat{\alpha} = 19.1667 + 3.5833 = 22.7500$$

$$A_2 \text{の推定値} = \hat{\theta}_0 - \hat{\theta}_1 = \hat{\mu} - \hat{\alpha} = 19.1667 - 3.5833 = 15.5833$$

であり、A<sub>2</sub>の水準の場合には、一致しないことに注意が必要である。なお、統計ソフト SAS および JMP では、これらの線形和の推定値を最小 2 乗平均 (lsmeans) と称している。

|                     |           | $l_0$ | $l_1$ | $l_2$ | $l_3$ | $l_4$ | $\theta^{\wedge}$ |   | $L^{\wedge(1)}$ |
|---------------------|-----------|-------|-------|-------|-------|-------|-------------------|---|-----------------|
| A <sub>1</sub> の推定値 | $l^{(1)}$ | 1     | 1     | 0     | 0     | 0     | 19.1667           | = | 22.7500         |
|                     |           |       |       |       |       |       | 3.5833            |   |                 |
|                     |           |       |       |       |       |       | 4.3333            |   |                 |
|                     |           |       |       |       |       |       | 2.1667            |   |                 |
|                     |           |       |       |       |       |       | 1.9167            |   |                 |

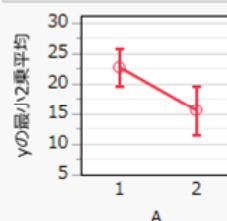
スライド 43

## 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 |

最小2乗平均プロット



|     |                     |                   |         |        | 線形和の例             |                   |
|-----|---------------------|-------------------|---------|--------|-------------------|-------------------|
| 因子  |                     |                   | 係数      | 標準誤差   | A <sub>1</sub> 水準 | A <sub>2</sub> 水準 |
| 切片  | $\theta^{\wedge}_0$ | $\mu^{\wedge}$    | 19.1667 | 0.6009 | 1                 | 1                 |
| A   | $\theta^{\wedge}_1$ | $\alpha^{\wedge}$ | 3.5833  | 0.6009 | 1                 | -1                |
| B   | $\theta^{\wedge}_2$ | $\beta^{\wedge}$  | 4.3333  | 0.6009 | 0                 | 0                 |
| C   | $\theta^{\wedge}_3$ | $\gamma^{\wedge}$ | 2.1667  | 0.6009 | 0                 | 0                 |
| D   | $\theta^{\wedge}_4$ | $\delta^{\wedge}$ | 1.9167  | 0.6009 | 0                 | 0                 |
| 積和= |                     |                   |         |        | 22.7500           | 15.5833           |

2021年12月16&17日 高橋行雄

43

推定値  $\hat{\theta}_k$  と  $\hat{\theta}_{k'}$  の共分散  $Cov(\hat{\theta}_k, \hat{\theta}_{k'})$  は,  $c_{kk'}\hat{\sigma}^2$  であることは, すでに示した. 推定されたパラメータ  $\hat{\theta}_k$  に, ある数値  $l_k$  を掛けた積和を線形和

$L = \sum_k l_k \hat{\theta}_k$  とする. 線形和  $L$  の分散  $Var(L)$  は,

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

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. ただし, Excel の行列関数を使えば線形和  $L$  の分散は, 容易に計算できる.  $A_1$  の分散は, スライド 45 に示すように

$$\left. \begin{aligned} Var(L_{A_1}) &= l_1 l_1 c_{1,1} \hat{\sigma}^2 + l_1 l_2 c_{1,2} \hat{\sigma}^2 + l_2 l_1 c_{2,1} \hat{\sigma}^2 + l_2 l_2 c_{2,2} \hat{\sigma}^2 \\ &= 0.361 + (-0.090) + (-0.090) + 0.361 = 0.5417 \end{aligned} \right\} \quad (21)$$

と計算される. 行列計算では, 任意のベクトル  $l$  に対し  $Var(L) = l \Sigma(\hat{\theta}) l^T$  で計算できる.

## Excel による線形和 $L$ の分散の計算

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

|                     | $l_0$ | $l_1$ | $l_2$ | $l_3$ | $l_4$ | パラメータの共分散行列 $\Sigma(\hat{\theta})$ |        |        |        |        | $l^{(1)T}$              |        |
|---------------------|-------|-------|-------|-------|-------|------------------------------------|--------|--------|--------|--------|-------------------------|--------|
| $A_1$ の分散 $l^{(1)}$ | 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                       |        |
|                     |       |       |       |       |       | =                                  |        |        |        |        | $Var^{\wedge}(L^{(1)})$ |        |
|                     |       |       |       |       |       | 0.2708                             | 0.2708 | 0.0000 | 0.0000 | 0.0000 | 1                       | 0.5417 |
|                     |       |       |       |       |       |                                    |        |        |        |        | 1                       |        |
|                     |       |       |       |       |       |                                    |        |        |        |        | 0                       |        |
|                     |       |       |       |       |       |                                    |        |        |        |        | 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.

スライド 46 に、線形和  $L$  および線形和  $L$  の分散  $Var(L)$  による各因子の水準平均（最小 2 乗平均）と 95%信頼区間の計算結果を示す。因子 A の第 1 水準の線形和は、 $L^{(1)}$  の行ベクトル  $\mathbf{l}^{(1)}$  とパラメータの推定値  $\hat{\boldsymbol{\theta}}$  の列ベクトルの積  $L^{(1)} = \mathbf{l}^{(1)} \hat{\boldsymbol{\theta}} = 22.750$  で計算した結果であり、分散は、スライド 45 に示したパラメータの共分散行列  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$  を使い  $Var(L^{(1)}) = \mathbf{l}^{(1)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{l}^{(1)T} = 0.5417$  で計算した結果である。分散の平方根  $SE$  に  $t_{0.05}(7-5)=4.303$  を掛けたのは、Excel の折れ線グラフで 95%信頼区間の髭を重ね書きするとき用いるためである。

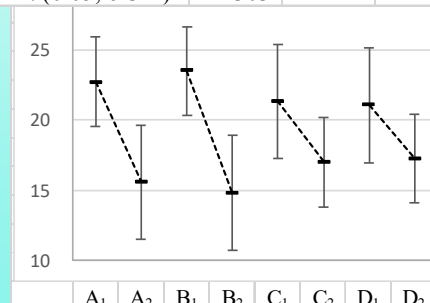
各因子の線形和  $L^{(i)}$  に対応する  $\mathbf{l}^{(i)}$  を設定し、フィルハンドルで  $L^{(i)}$  の計算式をコピーすることにより、一括して計算することができる。作図は、Excel の折れ線グラフで線形和  $L^{(i)}$ （最小 2 乗平均）を一括して描き、因子間の線を消し、「グラフ要素→誤差範囲→その他のオプション→ユーザー設定→値の設定→正および負の誤差の範囲→ $(t \times SE)$  の範囲の選択」により、最小 2 乗平均の 95%信頼区間の髭を付ける。折れ線グラフの X 軸の目盛りは、そのままでは使い物にならないので、Excel のセルに「A<sub>1</sub>, A<sub>2</sub>, ..., D<sub>2</sub>」と入力した結果で代用する。

スライド 46

## 各因子の水準平均と95%信頼区間

|           |                |                |                |                | $l_0$ | $l_1$    | $l_2$   | $l_3$    | $l_4$    |                   | $y^{\wedge}$       |                   |               |            |            |
|-----------|----------------|----------------|----------------|----------------|-------|----------|---------|----------|----------|-------------------|--------------------|-------------------|---------------|------------|------------|
| $L$       | A              | B              | C              | D              | $\mu$ | $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $\theta^{\wedge}$ | $L^{\wedge}$       | $Var(L^{\wedge})$ | $t \times SE$ | $L_{95\%}$ | $U_{95\%}$ |
| $L^{(1)}$ | A <sub>1</sub> |                |                |                | 1     | 1        | 0       | 0        | 0        | 19.167            | 22.750             | 0.5417            | 3.167         | 19.583     | 25.917     |
| $L^{(2)}$ | A <sub>2</sub> |                |                |                | 1     | -1       | 0       | 0        | 0        | 3.583             | 15.583             | 0.9028            | 4.088         | 11.495     | 19.671     |
| $L^{(3)}$ |                | B <sub>1</sub> |                |                | 1     | 0        | 1       | 0        | 0        | 4.333             | 23.500             | 0.5417            | 3.167         | 20.333     | 26.667     |
| $L^{(4)}$ |                | B <sub>2</sub> |                |                | 1     | 0        | -1      | 0        | 0        | 2.167             | 14.833             | 0.9028            | 4.088         | 10.745     | 18.921     |
| $L^{(5)}$ |                |                | C <sub>1</sub> |                | 1     | 0        | 0       | 1        | 0        | 1.917             | 21.333             | 0.9028            | 4.088         | 17.245     | 25.421     |
| $L^{(6)}$ |                |                | C <sub>2</sub> |                | 1     | 0        | 0       | -1       | 0        |                   | 17.000             | 0.5417            | 3.167         | 13.833     | 20.167     |
| $L^{(7)}$ |                |                |                | D <sub>1</sub> | 1     | 0        | 0       | 0        | 1        |                   | 21.083             | 0.9028            | 4.088         | 16.995     | 25.171     |
| $L^{(8)}$ |                |                |                | D <sub>2</sub> | 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  $t \times SE$  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。



2021年12月16&17日 高橋行雄

46

直交表実験では、現行の操業状態( $A_1, B_1, C_1, D_1$ )での推定値に対し、水準を変更したときに良い結果（小さい方がよいとする）が得られる水準に変更したときの推定値を求め、その差の 95%信頼区間により、実験結果の総合評価をするのが常である。スライド 47 に示すように  $L^{(9)}$  が現行水準の組合せで、 $L^{(10)}$  が最適水準の組合せ( $A_2, B_2, C_2, D_2$ )である。その差は  $L^{(11)}$  で計算されていて、差の推定値は  $L^{(12)} = -24.000$ ，分散が 4.333，95%信頼区間は (-32.957, -15.043) と計算されている。

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。

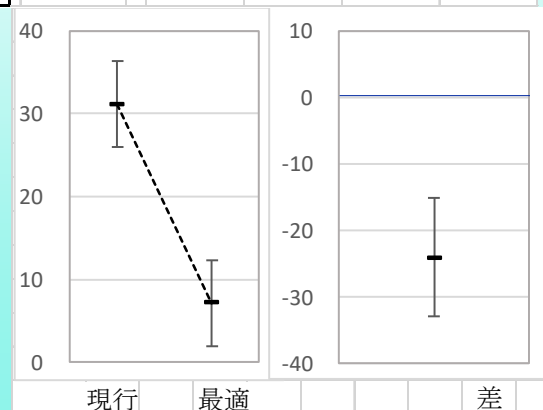
スライド 47

## 現行・最適水準・差の推定

|            |       |       |       |       | $l_0$ | $l_1$    | $l_2$   | $l_3$    | $l_4$    |                   | $y^{\wedge}$      |                   |               |            |            |
|------------|-------|-------|-------|-------|-------|----------|---------|----------|----------|-------------------|-------------------|-------------------|---------------|------------|------------|
| $L$        | A     | B     | C     | D     | $\mu$ | $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $\theta^{\wedge}$ | $L^{\wedge}$      | $Var(L^{\wedge})$ | $t \times SE$ | $L_{95\%}$ | $U_{95\%}$ |
| $L^{(9)}$  | $A_1$ | $B_1$ | $C_1$ | $D_1$ | 1     | 1        | 1       | 1        | 1        | 19.167            | 31.167            | 1.4444            | 5.171         | 25.996     | 36.338     |
| $L^{(10)}$ | $A_2$ | $B_2$ | $C_2$ | $D_2$ | 1     | -1       | -1      | -1       | -1       | 3.583             | 7.167             | 1.4444            | 5.171         | 1.996      | 12.338     |
| $L^{(11)}$ | 現行-最適 |       |       |       | 0     | -2       | -2      | -2       | -2       | 4.333             | -24.000           | 4.3333            | 8.957         | -32.957    | -15.043    |
|            |       |       |       |       |       |          |         |          |          | 2.167             | $t(0.05, 6-3-1)=$ |                   | 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。

2021年12月16&17日 高橋行雄



47

欠測値がある場合の  $L_9$  直交表についてデザイン行列を用いた線形モデルを適用したのであるが，スライド 46 に示したグラフが Excel の線グラフによって 95%信頼区間も含めて短時間できれいに仕上げられる。これまでに使用経験のある統計ソフトで結果のグラフ表示に優れているのは，JMP であるが，現行水準と最適水準についての 95%信頼区間付きのグラフは，JMP で作成するすべが思い当たらないので，スライド 13 に示したように，Excel の線グラフの結果で補足する必要があった。



## SAS の GLM プロシジャでの(1, 0)型ダミー変数

第 5 章で、SAS の GLM プロシジャで欠測値がある直交表について各種の推定結果を示した。GLM プロシジャで用いている (1, 0) 型のダミー変数を用いた場合について、表 12 に示すように Excel で再現した結果を示す。パラメータの推定値（係数）が第 5 章で示した GLM プロシジャのパラメータの推定値に一致することが確認できる。もちろん (1, -1) 型のスライド 30, (1, 2) 型のスライド 35 の推定値とは異なるが、切片を除いて  $p$  値は一致している。

表 12 (1, 0) 型のダミー変数を与えた場合の解析結果

| デザイン行列 $X$ |       |       |       |       |       |            |                 |   | $Y$ | 分散分析表  |               |               |               |               |
|------------|-------|-------|-------|-------|-------|------------|-----------------|---|-----|--|---------------|---------------|---------------|---------------|
|            | $x_0$ | $x_1$ | $x_2$ | $x_4$ | $x_7$ |            |                 |   |     |  | 自由度           | 変動            | 分散            | 分散比           |
| No.        | $\mu$ | A     | B     | C     | D     | $\theta$   | $\varepsilon$   |   | $y$ | 回帰   | 4             | 303.095       | 75.7738       | 34.9725       |
| 1          | 1     | 1     | 1     | 1     | 1     | $\theta_0$ | $\varepsilon_1$ | = | 32  | 残差   | 2             | 4.333         | <b>2.1667</b> |               |
| 2          | 1     | 1     | 1     | 0     | 0     | $\theta_1$ | $\varepsilon_2$ |   | 23  | 合計   | 6             | 307.429       |               |               |
| 3          | 1     | 1     | 0     | 1     | 0     | $\theta_2$ | $\varepsilon_3$ |   | 19  |  |               |               |               |               |
| 4          | 1     | 1     | 0     | 0     | 1     | $\theta_3$ | $\varepsilon_4$ |   | 17  |  | 係数            | 標準誤差          | $t$           | $P$ -値        |
| 5          | 1     | 0     | 1     | 1     | 0     | $\theta_4$ | $\varepsilon_5$ |   | 19  | 切片   | <b>7.1667</b> | 1.2019        | 5.9630        | 0.0270        |
| 6          | 1     | 0     | 1     | 0     | 1     |            |                 |   | 20  | A  | <b>7.1667</b> | 1.2019        | 5.9630        | <b>0.0270</b> |
| 8          | 1     | 0     | 0     | 0     | 0     |            |                 |   | 8   | B  | <b>8.6667</b> | 1.2019        | 7.2111        | <b>0.0187</b> |
|            | $\mu$ | $a$   | $b$   | $c$   | $abc$ |            |                 |   |     | C  | <b>4.3333</b> | 1.2019        | 3.6056        | 0.0691        |
|            |       |       |       |       |       |            |                 |   |     | D  | <b>3.8333</b> | 1.2019        | 3.1895        | 0.0858        |
|            |       |       |       |       |       |            |                 |   |     | パラメータの共分散行列 $\Sigma(\hat{\theta}) = (X^T X)^{-1} \sigma^2$ |               |               |               |               |
|            |       |       |       |       |       |            |                 |   |     | <b>1.4444</b>  | -0.7222       | -0.7222       | -0.3611       | -0.3611       |
|            |       |       |       |       |       |            |                 |   |     | -0.7222  | <b>1.4444</b> | 0.3611        | -0.3611       | -0.3611       |
|            |       |       |       |       |       |            |                 |   |     | -0.7222  | 0.3611        | <b>1.4444</b> | -0.3611       | -0.3611       |
|            |       |       |       |       |       |            |                 |   |     | -0.3611  | -0.3611       | -0.3611       | <b>1.4444</b> | 0.3611        |
|            |       |       |       |       |       |            |                 |   |     | -0.3611  | -0.3611       | -0.3611       | 0.3611        | <b>1.4444</b> |
|            |       |       |       |       |       |            |                 |   |     | =Minverse(Mmult(Transpose( $X$ の範囲), $X$ の範囲))* $\sigma^2$ |               |               |               |               |

各因子の水準に対する最小 2 乗平均は、GLM プロシジャの estimate ステートメントで次のように推定できることを示した。

estimate '最小 2 乗平均 A1' intercept 1 A 1 0 B 0.5 0.5 C 0.5 0.5 D 0.5 0.5 ;

ここで、0.5 は、それぞれのダミー変数の平均値であり、Excel でのデザイン行列では、1 次独立になるように第 2 水準を落とした場合なので、次のような設定を行なうことになる。

'最小 2 乗平均 A1' intercept 1 A 1 B 0.5 C 0.5 D 0.5 ;

表 13 に (1, 0) 型のダミー変数における各因子の最小 2 乗平均と 95%信頼区間の計算結果を示す。これは、スライド 46 に示した (1, -1) 対比型ダミー変数における線形和のためのベクトル  $I^{(i)}$  と異なるが、線形和の推定値  $\hat{L}^{(i)}$  も分散  $Var(\hat{L}^{(i)})$  も完全に一致し、もちろん 95% 信頼区間も一致する。



表 13 (1, 0) 型のダミー変数における各因子の最小 2 乗平均と 95%信頼区間

|           |                |                |                |                | $l_0$ | $l_1$    | $l_2$   | $l_3$    | $l_4$    |                 | $y^\wedge$        |                 |       |         |         |
|-----------|----------------|----------------|----------------|----------------|-------|----------|---------|----------|----------|-----------------|-------------------|-----------------|-------|---------|---------|
| $L$       | A              | B              | C              | D              | $\mu$ | $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $\theta^\wedge$ | $L^\wedge$        | $Var(L^\wedge)$ | $SE$  | $L95\%$ | $U95\%$ |
| $L^{(1)}$ | A <sub>1</sub> |                |                |                | 1     | 1        | 0.5     | 0.5      | 0.5      | 7.167           | 22.750            | 0.5417          | 0.736 | 19.583  | 25.917  |
| $L^{(2)}$ | A <sub>2</sub> |                |                |                | 1     | 0        | 0.5     | 0.5      | 0.5      | 7.167           | 15.583            | 0.9028          | 0.950 | 11.495  | 19.671  |
| $L^{(3)}$ |                | B <sub>1</sub> |                |                | 1     | 0.5      | 1       | 0.5      | 0.5      | 8.667           | 23.500            | 0.5417          | 0.736 | 20.333  | 26.667  |
| $L^{(4)}$ |                | B <sub>2</sub> |                |                | 1     | 0.5      | 0       | 0.5      | 0.5      | 4.333           | 14.833            | 0.9028          | 0.950 | 10.745  | 18.921  |
| $L^{(5)}$ |                |                | C <sub>1</sub> |                | 1     | 0.5      | 0.5     | 1        | 0.5      | 3.833           | 21.333            | 0.9028          | 0.950 | 17.245  | 25.421  |
| $L^{(6)}$ |                |                | C <sub>2</sub> |                | 1     | 0.5      | 0.5     | 0        | 0.5      |                 | 17.000            | 0.5417          | 0.736 | 13.833  | 20.167  |
| $L^{(7)}$ |                |                |                | D <sub>1</sub> | 1     | 0.5      | 0.5     | 0.5      | 1        |                 | 21.083            | 0.9028          | 0.950 | 16.995  | 25.171  |
| $L^{(8)}$ |                |                |                | D <sub>2</sub> | 1     | 0.5      | 0.5     | 0.5      | 0        |                 | 17.250            | 0.5417          | 0.736 | 14.083  | 20.417  |
|           |                |                |                |                |       |          |         |          |          |                 | $t(0.05, 6-3-1)=$ |                 | 4.303 |         |         |

現行水準，最適水準，それらの差について GLM プロシジャで設定した estimate ステートメントでは，次のように設定した．

```

estimate '最適'      intercept 1 A 0 1 B 0 1 C 0 1 D 0 1 ;
- ) estimate '現行'   intercept 1 A 1 0 B 1 0 C 1 0 D 1 0 ;
-----
estimate '最適-現行' intercept 0 A -1 1 B -1 1 C -1 1 D -1 1 ;

```

Excel の場合は，表 14 に示すように estimate ステートメントで設定した第 2 水準を落とした第 1 水準による，線形和のためのベクトル  $\mathbf{l}^{(i)}$  を設定する．(1, -1) 型のスライド 47 とは，線形和のためのベクトル  $\mathbf{l}^{(i)}$  は異なるが，線形和の推定値  $\hat{L}^{(i)}$  も分散  $Var(\hat{L}^{(i)})$  も完全に一致し，もちろん  $SE$  も一致する．

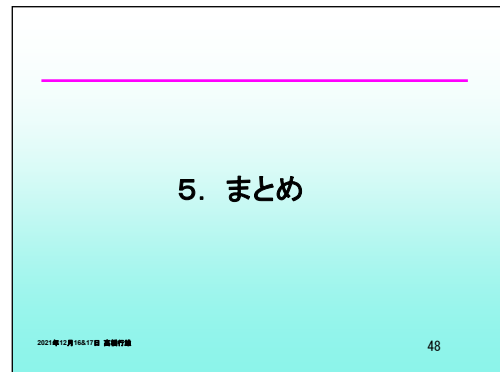
表 14 (1, 0) 型のダミー変数における現行・最適・(最適－現行) の最小 2 乗平均

|            |                |                |                |                | $l_0$ | $l_1$    | $l_2$   | $l_3$    | $l_4$    |                 | $y^\wedge$        |                 |       |         |         |
|------------|----------------|----------------|----------------|----------------|-------|----------|---------|----------|----------|-----------------|-------------------|-----------------|-------|---------|---------|
| $L$        | A              | B              | C              | D              | $\mu$ | $\alpha$ | $\beta$ | $\gamma$ | $\delta$ | $\theta^\wedge$ | $L^\wedge$        | $Var(L^\wedge)$ | $SE$  | $L95\%$ | $U95\%$ |
| $L^{(9)}$  | A <sub>1</sub> | B <sub>1</sub> | C <sub>1</sub> | D <sub>1</sub> | 1     | 1        | 1       | 1        | 1        | 7.167           | 31.167            | 1.4444          | 1.202 | 25.996  | 36.338  |
| $L^{(10)}$ | A <sub>2</sub> | B <sub>2</sub> | C <sub>2</sub> | D <sub>2</sub> | 1     | 0        | 0       | 0        | 0        | 7.167           | 7.167             | 1.4444          | 1.202 | 1.996   | 12.338  |
| $L^{(11)}$ | 最適-現行          |                |                |                | 0     | -1       | -1      | -1       | -1       | 8.667           | -24.000           | 4.3333          | 2.082 | -32.957 | -15.043 |
|            |                |                |                |                |       |          |         |          |          | 4.333           | $t(0.05, 6-3-1)=$ |                 | 4.303 |         |         |
|            |                |                |                |                |       |          |         |          |          | 3.833           |                   |                 |       |         |         |

## 8. 伝統的な実験計画法からの脱却

スライド 48

JMP の「モデルのあてはめ」は、JMP が最初に提供された時代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、手計算による平方和の分解を主体にしており、アリ地獄的な分散分析の世界から抜け出すことができない。



スライド 49

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。

「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。

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優れたものである。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。ただし、「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でようやく線形モデルが、にじみでてきたことにほくそえんでいる。

### 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。

2021年12月16&17日 高橋行雄 49

スライド 50

### 「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。

2021年12月16&17日 高橋行雄 50

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  $\Sigma(\hat{\theta})$  が、Excel の行列関数によって容易にできることは、 $+\alpha$  の機能として重要である。この  $\Sigma(\hat{\theta})$  こそが、各種の推定値に対する 95%信頼区間の計算に無くてはならないものである。

スライド 51

## Excel の回帰分析 $+\alpha +\beta$

- ◆ 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、 **$+\alpha$**  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  **$+\beta$**  の機能として多大な寄与が期待できることも例示した。

2021年12月16&17日 高橋行雄

51

Excel の散布図、あるいは、折れ線グラフによる作図機能も  $+\beta$  の機能として多大な寄与が期待できることも例示した。ただし、Excel の折れ線グラフで 95%信頼区間の髭を付ける際に注意しなければならないのは、髭の長さに対して「固定値」などの選択肢もあり、注意深く操作しないといい加減な誤差範囲が自動的に付けられちゃう恐ろしさがあることを肝に銘ずる必要がある。さらに、「標準偏差」および「標準誤差」のように尤もらしい目くらまし的な設定にも要注意である。

表 15 Excel のいい加減な折れ線グラフの誤差範囲のトラップに注意

| 誤差範囲  |         |
|---|---------|
| <input type="radio"/> 固定値(F)                | 0.1     |
| <input checked="" type="radio"/> パーセンテージ(P) | 5.0 %   |
| <input type="radio"/> 標準偏差(S)               | 1.0     |
| <input type="radio"/> 標準誤差(E)               |         |
| <input type="radio"/> ユーザー設定(C)             | 値の指定(V) |

| ユーザー設定の誤差範囲  |       |
|--|-------|
| 正の誤差の値(P)  | = {1} |
| 負の誤差の値(N)  | = {1} |
| <input type="button" value="OK"/> <input type="button" value="キャンセル"/> |       |

実験データの解析法として、電卓一丁でも分散分析表の作成により、種々の解析ができる解析手順が整備され活用されてきたことは、素晴らしいことであった。ただし、実験の型が異なると別の計算手順が必要となる。本章で示したデザイン行列を用いた線形モデルによる解法は、どのような実験の型にも適用できる汎用的な方法である。Excel の行列関数を用いることにより、自ら各種の推定値に対する 95%信頼区間を計算することができ、統計ソフト JMP を凌駕する各種のグラフ表示も可能であることも示した。

高橋・大橋・芳賀（1989）の第 9 章および第 10 章で取り上げた直交表の解析は、SAS/GLM プロシジャを用いて行った。ただし、現行水準と最適水準の推定に関しては、GLM プロシジャが採用している（1，0）型のデザイン行列を十分に理解した上で estimate ステートメントによる設定が必要であることを示した。JMP では、スライド 12 に示したように GUI を用いた各因子の水準を自由に選択することによって、最適水準および現行水準の推定値と 95%信頼区間が自動的に計算されることは、画期的であり JMP の「モデルのあてはめ」は、SAS の GLM プロシジャよりも進歩している。

JMP は、年間 10 万円を必要とする有償の統計ソフトであり、誰でも手軽に使える状況にはない。この講演は、JMP ユーザを対象にしたのであるが、JMP の「モデルのあてはめ」の計算を Excel の標準機能のみ（マクロなし）で再現できることを示すことにより、JMP の計算原理を深く理解する切っ掛けとなり、更なる応用力を付けてもらいたいと願っている。また、JMP が使えないユーザに対しては、最先端の統計ソフトの計算機能のすごさを実感しつつ、Excel による再現可能であることを認識し、自ら手を動かしてチャレンジされることを期待している。

スライド 52

## 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。

|      |                             |     |
|------|-----------------------------|-----|
| 4.   | 線形モデルによる欠測値がある直交表の解析        | 141 |
| 4.1. | 構造モデル・回帰モデル・線形モデル           | 141 |
| 4.2. | 線形推定論による定式化と EXCEL による行列計算  | 143 |
| 4.3. | 直交表における線形モデル                | 156 |
| 4.4. | 欠測値がある直交表における線形モデルの活用       | 159 |
| 4.5. | JMP の「モデルのあてはめ」を用いた線形モデルの活用 | 166 |

今回の講演は、「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 章の推敲の一環であり、スライド 52 に示す第 4.5 節、第 4.4 節の逆順でスライドを作成した。第 4.3 節については、スライドには含まれていないので第 2 章で取り上げた。第 4.1 節および第 4.2 節については、第 9 章で示す。また、伝統的な分散分析については、第 10 章で簡単に例示する。

スライド 53 に示した文献リストは、講演に用いたスライドで引用したものであり、追加の文献を含めて末尾に示す。

### スライド 53

## 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, 大橋靖雄, 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行列の活用  
<https://www.yukms.com/biostat/takahasi2/rec/009-12.htm>

## 9. 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 16 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  $\mu$  とし、要因効果を示す母数  $\alpha_i$ 、 $\beta_j$  および誤差  $\varepsilon_{ij}$  を用いた 1 次式

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2), \quad i=1,2,\dots,n_A, \quad j=1,2,\dots,n_B \quad (22)$$

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 16 繰り返しのない 2 元配置データ [楠ら (1995), 表 6.5, p191]

|                | B <sub>1</sub> | B <sub>2</sub> |
|----------------|----------------|----------------|
| A <sub>1</sub> | 10.9           | 11.5           |
| A <sub>2</sub> | 12.4           | 12.9           |
| A <sub>3</sub> | 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

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i=1,2,\dots,n \quad (23)$$

となる．これは、回帰モデルと呼ばれている．式 (22)、式 (23) の共通点は、

- 1) データ構造が、式 (22) の母数  $\mu$ 、 $\alpha_i$ 、 $\beta_j$  あるいは式 (23) の回帰係数  $\beta_0$ 、 $\beta_1$  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．

### 線形モデル

式 (22)、式 (23) は、一般的に

$$y_i = \theta_0 x_{0,i} + \theta_1 x_{1,i} + \dots + \theta_p x_{p,i} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i=1,2,\dots,n \quad (24)$$

と書ける．この式を線形モデルと呼ぶ．ただし、式 (24) で  $x_{0,i}$  をすべて 1 とした場合に  $\theta_0$  は、一般平均  $\mu$  となり、 $\theta_1, \theta_2, \dots, \theta_p$  は、式 (22) における各因子の主効果や（交互作用）のパラメータであり、式 (23) で回帰係数（パラメータ） $\beta_0, \beta_1, \beta_2, \dots$  に対応する．式 (22) における、 $3 \times 2$  の繰り返しのない 2 元配置データの構造モデルは、

$$\left. \begin{aligned} y_{11} &= \mu + \alpha_1 && + \beta_1 && + \varepsilon_{11} \\ y_{12} &= \mu + \alpha_1 && + \beta_2 && + \varepsilon_{12} \\ y_{21} &= \mu &+ \alpha_2 && + \beta_1 && + \varepsilon_{21} \\ y_{22} &= \mu &+ \alpha_2 && + \beta_2 && + \varepsilon_{21} \\ y_{31} &= \mu && + \alpha_3 + \beta_1 && + \varepsilon_{31} \\ y_{32} &= \mu && + \alpha_3 && + \beta_2 + \varepsilon_{32} \end{aligned} \right\} \quad (25)$$

として記述される．これは，式 (24) において  $p=5$  とし，母数  $\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2$  を  $\theta_0, \theta_1, \theta_2, \dots, \theta_5$  に対応付け，それらの変数を  $x_{0,i}, x_{1,i}, x_{2,i}, \dots, x_{5,i}$  とした線形モデル

$$\left. \begin{aligned} y_1 &= \theta_0 x_{0,1} + \theta_1 x_{1,1} + \theta_2 x_{2,1} + \dots + \theta_5 x_{5,1} + \varepsilon_1 \\ y_2 &= \theta_0 x_{0,2} + \theta_1 x_{1,2} + \theta_2 x_{2,2} + \dots + \theta_5 x_{5,2} + \varepsilon_2 \\ &\vdots \\ y_6 &= \theta_0 x_{0,6} + \theta_1 x_{1,6} + \theta_2 x_{2,6} + \dots + \theta_5 x_{5,6} + \varepsilon_6 \end{aligned} \right\} \quad (26)$$

となる．さて，式 (25) と式 (26) を比べると，式 (26) の変数  $x_{1,i}, x_{2,i}, \dots, x_{5,i}$  の間には，

$$\left. \begin{aligned} x_{1,i} + x_{2,i} + x_{3,i} &= 1, & i &= 1, 2, \dots, 6 \\ x_{4,i} + x_{5,i} &= 1, & i &= 1, 2, \dots, 6 \end{aligned} \right\} \quad (27)$$

の関係（1次従属）がある．構造（DE）モデルにおける母数に関する制約

$$\alpha_1 + \alpha_2 + \alpha_3 = 0, \quad \beta_1 + \beta_2 = 0 \quad (28)$$

を付け，最後の水準について整理すると

$$\alpha_3 = -\alpha_1 - \alpha_2, \quad \beta_1 = -\beta_2 \quad (29)$$

となり，これらを式 (25) に代入すると

$$\left. \begin{aligned} y_{11} &= \mu + \alpha_1 && + \beta_1 + \varepsilon_{11} \\ y_{12} &= \mu + \alpha_1 && - \beta_1 + \varepsilon_{12} \\ y_{21} &= \mu &+ \alpha_2 + \beta_1 + \varepsilon_{21} \\ y_{22} &= \mu &+ \alpha_2 - \beta_1 + \varepsilon_{22} \\ y_{31} &= \mu - \alpha_1 - \alpha_2 + \beta_1 + \varepsilon_{31} \\ y_{32} &= \mu - \alpha_1 - \alpha_2 - \beta_1 + \varepsilon_{32} \end{aligned} \right\} \quad (30)$$

と書き換えられる．式 (30) は，式 (24) において変数の数を  $p=3$  とし，母数  $\mu, \alpha_1, \alpha_2, \beta_1$  を  $\theta_0, \theta_1, \theta_2, \theta_3$  に対応付け，それらの変数を  $x_{0,i}, x_{1,i}, x_{2,i}, x_{3,i}$  としたものに対応し，これらの変数の間に式 (27) のような従属関係は存在せず， $x_{0,i}, x_{1,i}, x_{2,i}, x_{3,i}$  は1次独立である．

一般に，式 (24) の線形モデルにおいて，変数  $x_{0,i}, x_{2,i}, \dots, x_{p,i}$  が1次独立な場合を，母数  $\theta_0, \theta_1, \dots, \theta_p$  にムダがないという．

線形モデル式 (24) を，データ数が  $n$  の線形モデル式に展開する．この線形モデルに基づく母数（パラメータ）の推定方法について示す．

$$\left. \begin{array}{l} y_1 = \theta_0 x_{0,1} + \theta_1 x_{1,1} + \cdots + \theta_p x_{p,1} + \varepsilon_1 \\ \vdots \\ y_i = \theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i} + \varepsilon_i \\ \vdots \\ y_n = \theta_0 x_{0,n} + \theta_1 x_{1,n} + \cdots + \theta_p x_{p,n} + \varepsilon_n \end{array} \right\} \quad (31)$$

## 正規方程式

式 (31) を残差  $\varepsilon_i$  について

$$\varepsilon_i = y_i - (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i}) \quad (32)$$

として解いたとき， $i=1, 2, \dots, n$  に関する残差平方和  $Q$

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [y_i - (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})]^2 \quad (33)$$

を最小にするパラメータ  $\theta_0, \theta_1, \dots, \theta_p$  の推定値を求めたい．このような方法を最小 2 乗法と呼び，そのとき得られる解を  $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$  とする．

具体的には，式 (33) の残差平方和  $Q$  を各  $\theta_k$  で偏微分して，0 とおいた式

$$\frac{\partial Q}{\partial \theta_k} = 0, \quad k = 0, 1, 2, \dots, p \quad (34)$$

すなわち，

$$\left. \begin{array}{l} \sum_{i=1}^n [x_{0,i} y_i - x_{0,i} (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})] = 0 \\ \dots\dots\dots \\ \sum_{i=1}^n [x_{k,i} y_i - x_{k,i} (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})] = 0 \\ \dots\dots\dots \\ \sum_{i=1}^n [x_{p,i} y_i - x_{p,i} (\theta_0 x_{0,i} + \theta_1 x_{1,i} + \cdots + \theta_p x_{p,i})] = 0 \end{array} \right\} \quad (35)$$

を解き，その解を  $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$  とする．

ここで，簡単のため，

$$\left. \begin{array}{l} a_{k,k'} = \sum_{i=1}^n x_{k,i} x_{k',i}, \quad k = 0, 1, 2, \dots, p, \quad k' = 0, 1, 2, \dots, p \\ B_k = \sum_{i=1}^n x_{k,i} y_i, \quad k = 0, 1, 2, \dots, p \end{array} \right\} \quad (36)$$

とおくと，式 (35) は，



$$\left. \begin{array}{l} \theta_0 a_{0,0} + \theta_1 a_{0,1} + \cdots + \theta_p a_{0,p} = B_0 \\ \cdots \cdots \cdots \\ \theta_0 a_{k,0} + \theta_1 a_{k,1} + \cdots + \theta_p a_{k,p} = B_k \\ \cdots \cdots \cdots \\ \theta_0 a_{p,0} + \theta_1 a_{p,1} + \cdots + \theta_p a_{p,p} = B_p \end{array} \right\} \quad (37)$$

となり，これを正規方程式と呼ぶ．この解を，最小 2 乗推定量といい，線形不偏推定量の中で分散最小の最良線形不偏推定量（BLUE: Best Linear Unbiased Estimator）になる．

### デザイン行列を用いた正規方程式

母数に無駄のない構造（DE）モデルの式 (30) を，線形モデル式で表すと

$$\left. \begin{array}{l} y_1 = \theta_0 x_{0,1} + \theta_1 x_{1,1} + \theta_2 x_{2,1} + \theta_3 x_{3,1} + \varepsilon_1 \\ y_2 = \theta_0 x_{0,2} + \theta_1 x_{1,2} + \theta_2 x_{2,2} + \theta_3 x_{3,2} + \varepsilon_2 \\ \vdots \\ y_6 = \theta_0 x_{0,6} + \theta_1 x_{1,6} + \theta_2 x_{2,6} + \theta_3 x_{3,6} + \varepsilon_6 \end{array} \right\} \quad (38)$$

となる．ここで，線形モデルの変数  $x_{k,i}$  全体をデザイン行列  $\mathbf{X}$  とし，パラメータ  $\theta_k$  を列ベクトル  $\boldsymbol{\theta}$  とすれば，表 17 で示すように行列を用いて線形モデルで表わすことができる．

表 17 デザイン行列  $\mathbf{X}$  を用いた線形モデル

|                |                |     | デザイン行列 $\mathbf{X}$ |           |           |           |                       |   |                            |   | $\mathbf{Y}$ |
|----------------|----------------|-----|---------------------|-----------|-----------|-----------|-----------------------|---|----------------------------|---|--------------|
| A              | B              | $i$ | $x_{0,i}$           | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\boldsymbol{\theta}$ | + | $\boldsymbol{\varepsilon}$ | = | $\mathbf{y}$ |
| A <sub>1</sub> | B <sub>1</sub> | 1   | 1                   | 1         | 0         | 1         | $\theta_0$            |   | $\varepsilon_1$            |   | $y_1$        |
|                | B <sub>2</sub> | 2   | 1                   | 1         | 0         | -1        | $\theta_1$            |   | $\varepsilon_2$            |   | $y_2$        |
| A <sub>2</sub> | B <sub>1</sub> | 3   | 1                   | 0         | 1         | 1         | $\theta_2$            |   | $\varepsilon_3$            |   | $y_3$        |
|                | B <sub>2</sub> | 4   | 1                   | 0         | 1         | -1        | $\theta_3$            |   | $\varepsilon_4$            |   | $y_4$        |
| A <sub>3</sub> | B <sub>1</sub> | 5   | 1                   | -1        | -1        | 1         |                       |   | $\varepsilon_5$            |   | $y_5$        |
|                | B <sub>2</sub> | 6   | 1                   | -1        | -1        | -1        |                       |   | $\varepsilon_6$            |   | $y_6$        |

表 17 は，デザイン行列  $\mathbf{X}$  の各行に列ベクトル  $\boldsymbol{\theta}$  を掛けて加える．すなわち，デザイン行列  $\mathbf{X}$  の第  $i$  行と列ベクトル  $\boldsymbol{\theta}$  の列との積和は，


$$\theta_0 x_{0,i} + \theta_1 x_{1,i} + \theta_2 x_{2,i} + \theta_3 x_{3,i} + \varepsilon_i = y_i \quad (39)$$

と表すことができる．

ここで示したように行列による式を Excel シート上に展開して「表」として示すことにより，行列計算を身近なものにすることができる．行列の表記は，大きな丸括弧（あるいは角括弧）で括った形が一般的であり，文書ファイル上で表記することは多くの時間を要するので避けたい．また，Excel シート上で表 18 に示すように角括弧および大きな丸括弧で括って実現することも可能ではあるが，面倒な操作が発生する．そこで，表 17 に示したように矩

表 18 伝統的な行列の表記法

|                |                |     | デザイン行列 $X$ |           |           |           |  |   |  |   | $Y$  |
|----------------|----------------|-----|------------|-----------|-----------|-----------|--|---|--|---|--|
| A              | B              | $i$ | $x_{0,i}$  | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\theta$   |   | $\varepsilon$  |   | $y$  |
| A <sub>1</sub> | B <sub>1</sub> | 1   | 1          | 1         | 0         | 1         | $\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix}$ | + | $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix}$ | = | $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$ |
|                | B <sub>2</sub> | 2   | 1          | 1         | 0         | -1        |  |   |  |   |  |
| A <sub>2</sub> | B <sub>1</sub> | 3   | 1          | 0         | 1         | 1         |  |   |  |   |  |
|                | B <sub>2</sub> | 4   | 1          | 0         | 1         | -1        |  |   |  |   |  |
| A <sub>3</sub> | B <sub>1</sub> | 5   | 1          | -1        | -1        | 1         |  |   | $\varepsilon_5$  |   | $y_5$  |
|                | B <sub>2</sub> | 6   | 1          | -1        | -1        | -1        |  |   | $\varepsilon_6$  |   | $y_6$  |

形状に並べた要素を「太い外枠」で囲むことにより新たな表記法とした．この表記法は，Excel シートの矩形状の要素（セル）をマウスで選択し，「セルの書式の設定」で  太い外枠(I) を選択することでストレスなく実現できる．

式 (37) の正規方程式は，表 19 に示すように行列の表記を用いて表わすことができる．なお，転置行列  $X^T$  は，Excel の行列関数  $X^T = \text{Transpose}(X \text{ の範囲})$  で求めことができ，デザイン行列  $X$  の行方向を列方向に，または，列方向を行方向に並び変えた行列である．行列の積  $X^T X$  は， $X^T$  の行方向と  $X$  の列方向の積和であり，その要素は，式 (36) の  $a_{k,k'}$  となっている．行列の積は，Excel の行列関数で， $X^T X = \text{Mmult}(X^T \text{ の範囲}, X \text{ の範囲})$  で求めることができる． $X^T Y$  は，転置行列  $X^T$  と列ベクトル  $Y$  の積であり，その要素は，式 (36) の  $B_k$  となっている．

表 19 デザイン行列  $X$  を用いた正規方程式

| 転置行列 $X^T$ |   |    |   |    |    |    | デザイン行列 $X$ |           |           |           | 転置行列 $X^T$          |   |   |    |   |    |    | $Y$ |      |       |
|------------|---|----|---|----|----|----|------------|-----------|-----------|-----------|---------------------|---|---|----|---|----|----|-----|------|-------|
|            | 1 | 2  | 3 | 4  | 5  | 6  | $x_{0,i}$  | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\theta^{\wedge}$   |   | 1 | 2  | 3 | 4  | 5  | 6   | $y$  |       |
| $x_{0,i}$  | 1 | 1  | 1 | 1  | 1  | 1  | 1          | 1         | 0         | 1         | $\theta^{\wedge}_0$ | = | 1 | 1  | 1 | 1  | 1  | 1   | 10.9 | $y_1$ |
| $x_{1,i}$  | 1 | 1  | 0 | 0  | -1 | -1 | 1          | 1         | 0         | -1        | $\theta^{\wedge}_1$ |   | 1 | 1  | 0 | 0  | -1 | -1  | 11.5 | $y_2$ |
| $x_{2,i}$  | 0 | 0  | 1 | 1  | -1 | -1 | 1          | 0         | 1         | 1         | $\theta^{\wedge}_2$ |   | 0 | 0  | 1 | 1  | -1 | -1  | 12.4 | $y_3$ |
| $x_{3,i}$  | 1 | -1 | 1 | -1 | 1  | -1 | 1          | 0         | 1         | -1        | $\theta^{\wedge}_3$ |   | 1 | -1 | 1 | -1 | 1  | -1  | 12.9 | $y_4$ |
|            |   |    |   |    |    |    | 1          | -1        | -1        | 1         |                     |   |   |    |   |    |    |     | 12.2 | $y_5$ |
|            |   |    |   |    |    |    | 1          | -1        | -1        | -1        |                     |   |   |    |   |    |    |     | 12.3 | $y_6$ |

| $X^T X (a_{k,k'})$ |     |     |     | $\theta^{\wedge}$   | = | $X^T Y$ |       |
|--------------------|-----|-----|-----|---------------------|---|---------|-------|
| 6.0                | 0.0 | 0.0 | 0.0 | $\theta^{\wedge}_0$ |   | 72.20   | $B_0$ |
| 0.0                | 4.0 | 2.0 | 0.0 | $\theta^{\wedge}_1$ |   | -2.10   | $B_1$ |
| 0.0                | 2.0 | 4.0 | 0.0 | $\theta^{\wedge}_2$ |   | 0.80    | $B_2$ |
| 0.0                | 0.0 | 0.0 | 6.0 | $\theta^{\wedge}_3$ |   | -1.20   | $B_3$ |

式 (36) の  $B_k$  は，積和行列  $(X^T X)$  と列ベクトル  $\hat{\theta}$  の積に等しく， $(X^T X)$  の第  $k$  行について，

$$\hat{\theta}_0 a_{k,0} + \hat{\theta}_1 a_{k,1} + \hat{\theta}_2 a_{k,2} + \hat{\theta}_3 a_{k,3} = B_k \quad (40)$$

このように、実際のデータを用いた行列計算を Excel のシート上で展開することが、行列計算を四則演算のごとく身近なものにする現代の計算手段である．実際に  $\mathbf{X}^T \mathbf{X}$  は、

によって数値計算された結果である。表 20 に示すように「C7:H10」は、 $\mathbf{X}^T$  の左上のセル位置：右下のセル位置であるが、実際の Excel の操作では、マウスで  $\mathbf{X}^T$  の範囲を選択するだけで自動的に設定してくれるので、セル位置の設定の煩わしさはない。ただし、関数の引数としてのこのようなセル位置の表示から内容を類推することはできないので、表記上は

とすることが望ましい。

Figure 1 shows an Excel spreadsheet with the formula bar displaying  $=\text{MMULT}(\text{C7:H10}, \text{J7:M12})$ . The spreadsheet contains the following data:

|    | A | B | C          | D | E  | F | G          | H  | I  | J         | K         | L         | M         | N          | O          | P | Q | R  | S |  |
|----|---|---|------------|---|----|---|------------|----|----|-----------|-----------|-----------|-----------|------------|------------|---|---|----|---|--|
| 4  |   |   |            |   |    |   |            |    |    |           |           |           |           |            |            |   |   |    |   |  |
| 5  |   |   | 転置行列 $X^T$ |   |    |   | デザイン行列 $X$ |    |    |           |           | 転置行列      |           |            |            |   |   |    |   |  |
| 6  |   |   | 1          | 2 | 3  | 4 | 5          | 6  |    | $x_{0,i}$ | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\theta^i$ |            | 1 | 2 | 3  |   |  |
| 7  |   |   | $x_{0,i}$  | 1 | 1  | 1 | 1          | 1  | 1  |           | 1         | 1         | 0         | 1          | $\theta^0$ | = | 1 | 1  | 1 |  |
| 8  |   |   | $x_{1,i}$  | 1 | 1  | 0 | 0          | -1 | -1 |           | 1         | 1         | 0         | -1         | $\theta^1$ |   | 1 | 1  | 0 |  |
| 9  |   |   | $x_{2,i}$  | 0 | 0  | 1 | 1          | -1 | -1 |           | 1         | 0         | 1         | 1          | $\theta^2$ |   | 0 | 0  | 1 |  |
| 10 |   |   | $x_{3,i}$  | 1 | -1 | 1 | -1         | 1  | -1 |           | 1         | 0         | 1         | -1         | $\theta^3$ |   | 1 | -1 | 1 |  |
| 11 |   |   |            |   |    |   |            |    |    |           | 1         | -1        | -1        | 1          |            |   |   |    |   |  |
| 12 |   |   |            |   |    |   |            |    |    |           | 1         | -1        | -1        | -1         |            |   |   |    |   |  |

The bottom right corner of the spreadsheet shows the result of the matrix multiplication  $X^T X$  (labeled  $a_{k,k}$ ):

|      | AA  | AB  | AC  | AD  |
|------|-----|-----|-----|-----|
| (12) | 0.0 | 0.0 | 0.0 | 0.0 |
| 0.0  | 4.0 | 2.0 | 0.0 | 0.0 |
| 0.0  | 2.0 | 4.0 | 0.0 | 0.0 |
| 0.0  | 0.0 | 0.0 | 6.0 | 0.0 |

正規方程式の解は、表 19 の下段の正規方程式に  $(\mathbf{X}^T \mathbf{X})$  の逆行列  $(\mathbf{X}^T \mathbf{X})^{-1}$  を掛けることにより得られる。逆行列の計算は、伝統的に煩雑な「掃き出し法」が用いられていたのであるが、Excel の行列関数 `Minverse`( $\mathbf{X}^T \mathbf{X}$  の範囲) によって直接計算できるようになった。シグマ表記の煩雑な計算法を行列表記で簡潔に示せるだけでなく、実際の計算にも四則演算のごとくそのまま使えることができることは、画期的なことである。

$$\hat{\theta}_k = c_{k,0}B_0 + c_{k,1}B_1 + c_{k,2}B_2 + c_{k,3}B_3 \quad (41)$$

57

表 21 デザイン行列  $X$  を用いた正規方程式の解

| $(X^T X)^{-1} (c_{k,k'})$ |        |        |       | $X^T X (a_{k,k'})$         |     |     |     | $\theta^\wedge$     | $(X^T X)^{-1} (c_{k,k'})$ |                   |        |       | $X^T Y$ |
|---------------------------|--------|--------|-------|----------------------------|-----|-----|-----|---------------------|---------------------------|-------------------|--------|-------|---------|
| 0.167                     | 0      | 0      | 0     | 6.0                        | 0.0 | 0.0 | 0.0 | $\theta^\wedge_0 =$ | 0.167                     | 0                 | 0      | 0     | 72.20   |
| 0                         | 0.333  | -0.167 | 0     | 0.0                        | 4.0 | 2.0 | 0.0 | $\theta^\wedge_1$   | 0                         | 0.333             | -0.167 | 0     | -2.10   |
| 0                         | -0.167 | 0.333  | 0     | 0.0                        | 2.0 | 4.0 | 0.0 | $\theta^\wedge_2$   | 0                         | -0.167            | 0.333  | 0     | 0.80    |
| 0                         | 0      | 0      | 0.167 | 0.0                        | 0.0 | 0.0 | 6.0 | $\theta^\wedge_3$   | 0                         | 0                 | 0      | 0.167 | -1.20   |
|                           |        |        |       |                            |     |     |     |                     |                           |                   |        |       |         |
|                           |        |        |       | $I = (X^T X)^{-1} (X^T X)$ |     |     |     | $\theta^\wedge$     | $(X^T X)^{-1} (c_{k,k'})$ |                   |        |       | $X^T Y$ |
|                           |        |        |       | 1.0                        | 0.0 | 0.0 | 0.0 | $\theta^\wedge_0 =$ | 0.167                     | 0                 | 0      | 0     | 72.20   |
|                           |        |        |       | 0.0                        | 1.0 | 0.0 | 0.0 | $\theta^\wedge_1$   | 0                         | 0.333             | -0.167 | 0     | -2.10   |
|                           |        |        |       | 0.0                        | 0.0 | 1.0 | 0.0 | $\theta^\wedge_2$   | 0                         | -0.167            | 0.333  | 0     | 0.80    |
|                           |        |        |       | 0.0                        | 0.0 | 0.0 | 1.0 | $\theta^\wedge_3$   | 0                         | 0                 | 0      | 0.167 | -1.20   |
|                           |        |        |       |                            |     |     |     |                     |                           |                   |        |       |         |
|                           |        |        |       |                            |     |     |     | $\theta^\wedge$     | $(X^T X)^{-1} X^T Y$      |                   |        |       |         |
|                           |        |        |       |                            |     |     |     | $\theta^\wedge_0 =$ | 12.033                    | $\mu^\wedge$      |        |       |         |
|                           |        |        |       |                            |     |     |     | $\theta^\wedge_1$   | -0.833                    | $\alpha^\wedge_1$ |        |       |         |
|                           |        |        |       |                            |     |     |     | $\theta^\wedge_2$   | 0.617                     | $\alpha^\wedge_2$ |        |       |         |
|                           |        |        |       |                            |     |     |     | $\theta^\wedge_3$   | -0.200                    | $\beta^\wedge_1$  |        |       |         |

したがって、右辺の逆行列  $(X^T X)^{-1}$  と  $X^T Y$  の積が、 $\theta$  の推定値  $\hat{\theta}$  となる．回帰モデルでは、 $\hat{\theta}$  に代わりに  $\hat{\beta}$  を用いて式 (41) を次式で

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (42)$$

表わすのが一般的である．式 (36) の  $a_{k,k'}$  および  $B_k$ ，式 (41) の  $c_{k,k'}$  は、手計算時代における正規方程式の解を求めるために工夫された表記法である．

手計算時代であっても、繰り返しが揃っている場合の正規方程式の解は、手計算レベルでも  $(X^T X)$  の逆行列  $(X^T X)^{-1}$  を求めることができる．ただし、欠測値などがあり繰り返しが不揃いになった場合は、コンピュータを用いた逆行列の計算が不可避である．楠ら (1995) の「第 7.3 節 一般の重回帰」に「Gauss-Doolittle 法による逆行列の計算」の詳細が表として示されている．

このようにして得られた最小 2 乗推定量  $\hat{\theta}_k$  の期待値、分散および共分散は、誤差分散の推定値を  $\hat{\sigma}^2$  としたときに

$$\left. \begin{aligned} E(\hat{\theta}_k) &= \theta_k \\ \text{Var}(\hat{\theta}_k) &= c_{k,k} \hat{\sigma}^2 \\ \text{Cov}(\hat{\theta}_k, \hat{\theta}_{k'}) &= c_{k,k'} \hat{\sigma}^2 \end{aligned} \right\} \quad (43)$$

として求められる．式 (33) の  $\theta_k$  に  $\hat{\theta}_k$  を代入して得られる残差平方和  $S_e$

$$\left. \begin{aligned} S_e &= \sum_{i=1}^n [y_i - (\hat{\theta}_0 x_{0,i} + \hat{\theta}_1 x_{1,i} + \cdots + \hat{\theta}_p x_{p,i})]^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned} \right\} \quad (44)$$

について、(自由度=データ数-独立な母数の個数)により、

$$E[S_e] = (n-p-1)\sigma^2 \quad (45)$$

となる。ゆえに、誤差分散  $\sigma^2$  の不偏推定量は、

$$\hat{\sigma}^2 = \frac{S_e}{n-p-1} \quad (46)$$

で与えられる。なお、残差平方和を  $S_e$  としているが、本来は  $S_{\hat{\varepsilon}}$  とすべきところを、表記の簡単化のための  $\hat{\varepsilon}$  を  $e$  と置き換えている。このような置き換えは、 $\beta$  の推定値  $\hat{\beta}$  とすべきところを、簡単化のために  $b$  と置き換えるような表記法と同様である。

誤差分散  $\sigma^2$  の推定のために表 22 に示すように、 $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\theta}}$  により予測値  $\hat{\mathbf{y}}$  を求める。次に、残差  $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}}$  の平方和から  $S_e = 0.0700$  を求め、誤差の自由度  $df = n - p - 1 = 6 - 3 - 1 = 2$  で割って、誤差分散の推定値  $\hat{\sigma}^2 = 0.0350$  が得られる。

表 22 予測値  $\hat{\mathbf{y}}$  の推定および誤差分散  $\hat{\sigma}^2$  の推定

|                |                |   | デザイン行列 $\mathbf{X}$ |           |           |           | 予測値 $\mathbf{y}^{\wedge}$ の推定  |                       | 誤差分散 $\sigma^{\wedge 2}$ の推定 |                       |  |       |
|----------------|----------------|---|---------------------|-----------|-----------|-----------|--------------------------------|-----------------------|------------------------------|-----------------------|--|-------|
| A              | B              | i | $x_{0,i}$           | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\boldsymbol{\theta}^{\wedge}$ | $\mathbf{y}^{\wedge}$ | $\mathbf{y}$                 | $\mathbf{y}^{\wedge}$ | $\boldsymbol{\varepsilon}^{\wedge} = \mathbf{y} - \mathbf{y}^{\wedge}$ |       |
| A <sub>1</sub> | B <sub>1</sub> | 1 | 1                   | 1         | 0         | 1         | 12.033                         | = 11.000              | 10.9                         | - 11.000              | = -0.100   |       |
|                | B <sub>2</sub> | 2 | 1                   | 1         | 0         | -1        | -0.833                         | 11.400                | 11.5                         | 11.400                | 0.100  |       |
| A <sub>2</sub> | B <sub>1</sub> | 3 | 1                   | 0         | 1         | 1         | 0.617                          | 12.450                | 12.4                         | 12.450                | -0.050   |       |
|                | B <sub>2</sub> | 4 | 1                   | 0         | 1         | -1        | -0.200                         | 12.850                | 12.9                         | 12.850                | 0.050  |       |
| A <sub>3</sub> | B <sub>1</sub> | 5 | 1                   | -1        | -1        | 1         |                                | 12.050                | 12.2                         | 12.050                | 0.150  |       |
|                | B <sub>2</sub> | 6 | 1                   | -1        | -1        | -1        |                                | 12.450                | 12.3                         | 12.450                | -0.150   |       |
|                |                |   |                     |           |           |           |                                |                       | $n =$                        | 6                     | $S_e =$  | 0.070 |
|                |                |   |                     |           |           |           |                                |                       | $p =$                        | 3                     | $\sigma^{\wedge 2} =$  | 0.035 |
|                |                |   |                     |           |           |           |                                |                       | $df = n - p - 1 =$           | 2                     | $\sigma^{\wedge} =$  | 0.187 |

デザイン行列  $\mathbf{X}$  とパラメータの推定値  $\hat{\boldsymbol{\theta}}$  の積の計算は、表 23 に示すように式 (30) の構造 (DE) モデルによる推定に対応する。

表 23 予測値の構造 (DE) モデルによる推定

| A              | B              | i | $\mu^{\wedge}$ |   | $\alpha^{\wedge}_1$ |   | $\alpha^{\wedge}_2$ |   | $\beta^{\wedge}_1$ |   | $\mathbf{y}^{\wedge}$ |
|----------------|----------------|---|----------------|---|---------------------|---|---------------------|---|--------------------|---|-----------------------|
| A <sub>1</sub> | B <sub>1</sub> | 1 | 12.033         | + | -0.833              | + | 0                   | + | -0.200             | = | 11.000                |
|                | B <sub>2</sub> | 2 | 12.033         |   | -0.833              |   | 0                   |   | 0.200              |   | 11.400                |
| A <sub>2</sub> | B <sub>1</sub> | 3 | 12.033         |   | 0                   |   | 0.617               |   | -0.200             |   | 12.450                |
|                | B <sub>2</sub> | 4 | 12.033         |   | 0                   |   | 0.617               |   | 0.200              |   | 12.850                |
| A <sub>3</sub> | B <sub>1</sub> | 5 | 12.033         |   | 0.833               |   | -0.617              |   | -0.200             |   | 12.050                |
|                | B <sub>2</sub> | 6 | 12.033         |   | 0.833               |   | -0.617              |   | 0.200              |   | 12.450                |

## パラメータの共分散行列の活用

推定されたパラメータ  $\hat{\theta}$  の分散および共分散は、式 (43) で示されているが、行列として扱いたい。誤差分散は、表 22 で  $\hat{\sigma}^2 = 0.0350$  と推定されているので、 $c_{k,k'}$  の行列  $(X^T X)^{-1}$  に  $\hat{\sigma}^2$  を掛け、表 24 に示すようにパラメータの共分散行列  $\Sigma(\hat{\theta})$  を求めることができる。共分散行列には、データに関する共分散行列  $\Sigma(\mathbf{x})$  もあり、パラメータの共分散行列を  $\Sigma(\hat{\theta})$  と表記し、区別する必要がある。なお、共分散行列は、分散共分散行列と同義語として用いている。なお、データの相関行列  $R(\mathbf{x})$  と共分散行列  $\Sigma(\mathbf{x})$  の事例については、スライド 40 を参照のこと。

表 24 パラメータの共分散行列  $\Sigma(\hat{\theta})$

|                  | $(X^T X)^{-1} (c_{k,k'})$ |                     |                     |                    | $\hat{\sigma}^2$ | パラメータの共分散行列 $\Sigma(\hat{\theta})$ |         |         |        |
|------------------|---------------------------|---------------------|---------------------|--------------------|------------------|------------------------------------|---------|---------|--------|
| $\hat{\theta}_0$ | 0.1667                    | 0                   | 0                   | 0                  | 0.0350           | 0.0058                             | 0       | 0       | 0      |
| $\hat{\theta}_1$ | 0                         | 0.3333              | -0.1667             | 0                  |                  | 0                                  | 0.0117  | -0.0058 | 0      |
| $\hat{\theta}_2$ | 0                         | -0.1667             | 0.3333              | 0                  |                  | 0                                  | -0.0058 | 0.0117  | 0      |
| $\hat{\theta}_3$ | 0                         | 0                   | 0                   | 0.1667             |                  | 0                                  | 0       | 0       | 0.0058 |
|                  | $\mu^{\wedge}$            | $\alpha^{\wedge}_1$ | $\alpha^{\wedge}_2$ | $\beta^{\wedge}_1$ |                  |                                    |         |         |        |

計算されたパラメータの共分散行列  $\Sigma(\hat{\theta})$  から、 $\hat{\theta}_1$  の分散は

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

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

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

となる。

実験計画法に基づいた実験データの解析上の最大のメリットは、取り上げた多くの因子の組合せデータから、ある注目する因子に対し、他の因子の影響を統計的に取り除いて、あたかも単独で繰り返した実験データのごとく水準の平均と標準誤差を推定することができ、さらに 95%信頼区間も精度よく推定することができることである。

線形推定論によってある線形和  $L$  とその 95%信頼区間の推定方法は、定式化されてはいるが、式 (49) で示す線形和の分散  $Var(L)$  の計算式は、手計算で対応できる範囲を超えている。そのために、第 10 章で示すように簡便な算術平均を用いる計算法が模索され、実用的に用いられてきた。ここでは、線形推定論の原点に戻り、Excel の行列関数を用いた計算方法を示す。

線形推定論による定式化は、次のようである。母数（パラメータ） $\theta_0, \theta_1, \dots, \theta_q (q \leq p)$  に対する任意の 1 次式（線形和）

$$L = \sum_k l_k \theta_k \quad (47)$$

の推定量は,

$$\hat{L} = \sum_k l_k \hat{\theta}_k \quad (48)$$

となる. その分散は,

$$Var(\hat{L}) = \sum_k \sum_{k'} l_k l_{k'} c_{k,k'} \hat{\sigma}^2 \quad (49)$$

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

$$\sum_k l_k \hat{\theta}_k \pm t_{\alpha}(n-p-1) \sqrt{\sum_k \sum_{k'} l_k l_{k'} c_{k,k'} \hat{\sigma}^2} \quad (50)$$

となる.

線形和  $\hat{L}$  に対する分散  $Var(\hat{L})$  が, 式 (49) として 2 次形式で示されているのであるが, 手計算の範囲を超えた式であり, 何らかのコンピュータ言語を用いて計算するしかなく, 実用性に欠けていた. スライド 46 および 47 に示したように救いの神が, Excel の行列関数であり, 線形和  $l_k$  のすべての組合せに対し  $l_k l_{k'} c_{k,k'}$  の和を計算することが, Excel の行列関数によって, Excel シート上に示されている矩形データを参照することにより, 何のストレスもなく四則演算のごとく計算できる.

### 組合せ平均の 95%信頼区間

表 25 に示した  $A_1 B_1$  に対するデザイン行列  $X$  の第 1 行目のデザイン変数ベクトルを

$$\begin{aligned} l^{(1)} &= [l_0^{(1)} \quad l_1^{(1)} \quad l_2^{(1)} \quad l_3^{(1)}] \\ &= [1 \quad 1 \quad 0 \quad 1] \end{aligned} \quad (51)$$

としたときに, 式 (48) は,

$$\hat{L}^{(1)} = \sum_{k=0}^3 l_k \hat{\theta}_k = l^{(1)} \hat{\theta} \quad (52)$$

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

表 25 組合せ平均に対するパラメータの共分散行列  $\Sigma(\hat{\theta})$  を用いた分散の計算

|           |       |       | $l_1$     | $l_2$     | $l_3$     | $l_4$     |                                     | $y^{\wedge}$ |                   |               |         |         |
|-----------|-------|-------|-----------|-----------|-----------|-----------|-------------------------------------|--------------|-------------------|---------------|---------|---------|
| $L$       | A     | B     | $x_{0,i}$ | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\theta^{\wedge}$                   | $L^{\wedge}$ | $Var(L^{\wedge})$ | $t \times SE$ | $L95\%$ | $U95\%$ |
| $L^{(1)}$ | $A_1$ | $B_1$ | 1         | 1         | 0         | 1         | 12.033                              | = 11.000     | 0.023             | 0.657         | 10.343  | 11.657  |
| $L^{(2)}$ |       | $B_2$ | 1         | 1         | 0         | -1        | -0.833                              | 11.400       | 0.023             | 0.657         | 10.743  | 12.057  |
| $L^{(3)}$ | $A_2$ | $B_1$ | 1         | 0         | 1         | 1         | 0.617                               | 12.450       | 0.023             | 0.657         | 11.793  | 13.107  |
| $L^{(4)}$ |       | $B_2$ | 1         | 0         | 1         | -1        | -0.200                              | 12.850       | 0.023             | 0.657         | 12.193  | 13.507  |
| $L^{(5)}$ | $A_3$ | $B_1$ | 1         | -1        | -1        | 1         |                                     | 12.050       | 0.023             | 0.657         | 11.393  | 12.707  |
| $L^{(6)}$ |       | $B_2$ | 1         | -1        | -1        | -1        |                                     | 12.450       | 0.023             | 0.657         | 11.793  | 13.107  |
|           |       |       |           |           |           |           | $t_{0.05}(6-3-1)=T.inv.2T(0.05,2)=$ |              | 4.3027            |               |         |         |

のように，Excel の 行列関数 Mmult( $\mathbf{l}^{(1)}$  の範囲，  $\hat{\boldsymbol{\theta}}$  範囲) で計算できる．式 (49) の分散の推定量は，

$$\hat{Var}(\hat{L}^{(1)}) = \sum_k \sum_{k'} l_k^{(1)} l_{k'}^{(1)} c_{k,k'} \hat{\sigma}^2 = \mathbf{l}^{(1)} \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}) \mathbf{l}^{(1)T} \quad (53)$$

|                    |             |             |             |             |  |         |         |         |                     |                      |
|--------------------|-------------|-------------|-------------|-------------|--|---------|---------|---------|---------------------|----------------------|
|                    | $l_0^{(1)}$ | $l_1^{(1)}$ | $l_2^{(1)}$ | $l_3^{(1)}$ | パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$ |         |         |         | $\mathbf{l}^{(1)T}$ |                      |
| $\mathbf{l}^{(1)}$ | 1           | 1           | 0           | 1           | 0.0058   | 0       | 0       | 0       | 1                   |                      |
|                    |             |             |             |             | 0  | 0.0117  | -0.0058 | 0       | 1                   |                      |
|                    |             |             |             |             | 0  | -0.0058 | 0.0117  | 0       | 0                   |                      |
|                    |             |             |             |             | 0  | 0       | 0       | 0.0058  | 1                   |                      |
|                    |             |             |             |             | =  | 0.0058  | 0.0117  | -0.0058 | 0.0058              | $Var(\hat{L}^{(1)})$ |
|                    |             |             |             |             |  |         |         |         | 1                   | = 0.0233             |
|                    |             |             |             |             |  |         |         |         | 1                   |                      |
|                    |             |             |             |             |  |         |         |         | 0                   |                      |
|                    |             |             |             |             |  |         |         |         | 1                   |                      |

によって計算されている．式 (50) の 95%信頼区間は，

$$\left. \begin{aligned} 95\%CL &= \sum_k l_k^{(1)} \hat{\theta}_k \pm t_{0.05}(n-p-1) \sqrt{\sum_k \sum_{k'} l_k^{(1)} l_{k'}^{(1)} c_{k,k'} \hat{\sigma}^2} \\ &= L^{(1)} \pm t(6-3-1) \sqrt{Var(L^{(1)})} \\ &= 11.000 \pm 4.3027 \sqrt{0.0233} \\ &= (10.3428, 11.6572) \end{aligned} \right\} \quad (54)$$

となる．行列計算は，ややこしいように思われるかも知れないが，行列のサイズが異なっても常に一定形式で計算できることに利便性がある．また，表 25 に示すように Excel のフィルハンドルによる計算式のコピペ（コピー&ペースト）機能により複数の線形和とその分散などの計算が一括してできる．このように複数の線形和（推定値）の分散および 95%がすばやくできることは，画期的なことである．

表 25 で推定した各因子の組合せの推定値に 95%信頼区間を付けた線グラフを図 1 に示す．因子 A と因子 B の各組合せによる実験データは 1 回しかないのに，その 95%信頼区間を表示できるのことは，2 元配置型の実験計画の賜物である．推定結果をきれいな図で示すことは，実験データの結果を解釈する際に必須であり，Excel 行列計算の結果に連結してすぐに簡便に作図できることは，これまでの統計ソフトでは実現が困難であった．

図 1 は，Excel の線グラフで，6 個の推定値に対して折れ線を自動的にあてはめ，マーカーを○印，折れ線を --- 点線に変更し， $A_1B_2$  と  $A_2B_1$ ， $A_2B_2$  と  $A_3B_1$  の間の点線を消した結果である．さらに推定値の 95%信頼区間の設定に際し「ユーザ設定」の機能を用いて髭を付け加えた結果である．なお，髭の長さは「 $t \times SE$ 」の計算結果を用いている．X 軸の設定は，Excel のセル内で記述し，折れ線グラフのサイズに合わせて幅を調整し，折れ線グラフごとコピーして，ワードに「拡張メタファイル」形式で貼り付けたものである．



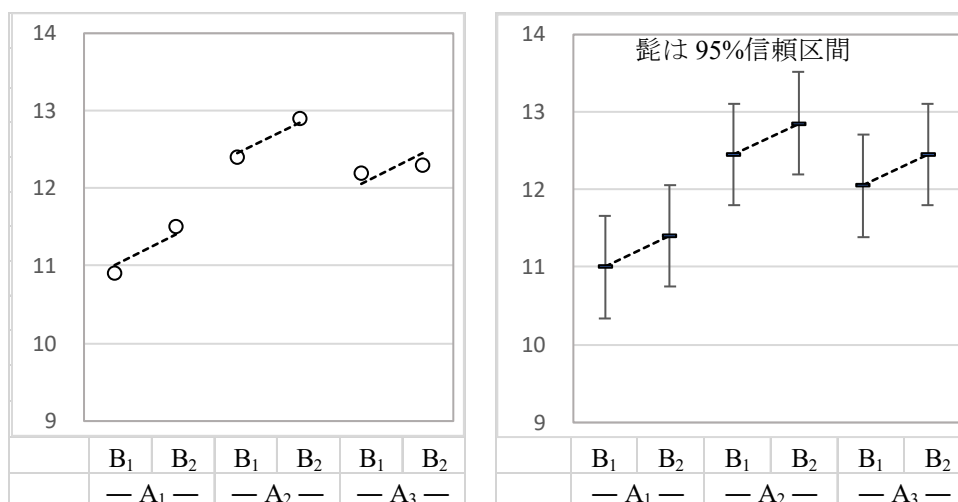


図 1 因子 A および B の組合せによる推定値と 95%信頼区間

$A_1B_1$  の水準平均は、表 23 から

$$\begin{aligned}
 \hat{y}_{A_1B_1} &= \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 \\
 &= \hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3 \\
 &= 12.0333 - 0.8333 - 0.2000 \\
 &= 11.0000
 \end{aligned}
 \quad \left. \vphantom{\begin{aligned} \hat{y}_{A_1B_1} &= \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1 \\ &= \hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3 \\ &= 12.0333 - 0.8333 - 0.2000 \\ &= 11.0000 \end{aligned}} \right\} \quad (55)$$

であることは、自明である。この分散は一般的な合成分散の公式から、表 24 の共分散行列を用いて

$$\begin{aligned}
 Var(\hat{y}_{A_1B_1}) &= Var(\hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3) \\
 &= Var(\hat{\theta}_0) + Var(\hat{\theta}_1) + Var(\hat{\theta}_3) \\
 &\quad + 2Cov(\hat{\theta}_0, \hat{\theta}_1) + 2Cov(\hat{\theta}_0, \hat{\theta}_3) + 2Cov(\hat{\theta}_1, \hat{\theta}_3) \\
 &= 0.0058 + 0.0117 + 0.0058 + 0.0 + 0.0 + 0.0 \\
 &= 0.0233
 \end{aligned}
 \quad \left. \vphantom{\begin{aligned} Var(\hat{y}_{A_1B_1}) &= Var(\hat{\theta}_0 + \hat{\theta}_1 + \hat{\theta}_3) \\ &= Var(\hat{\theta}_0) + Var(\hat{\theta}_1) + Var(\hat{\theta}_3) \\ &\quad + 2Cov(\hat{\theta}_0, \hat{\theta}_1) + 2Cov(\hat{\theta}_0, \hat{\theta}_3) + 2Cov(\hat{\theta}_1, \hat{\theta}_3) \\ &= 0.0058 + 0.0117 + 0.0058 + 0.0 + 0.0 + 0.0 \\ &= 0.0233 \end{aligned}} \right\} \quad (56)$$

計算できるが、式 (53) と本質的に同じ計算方法である。

## 因子の水準間の差の推定および 95%信頼区間

表 16 に示した繰り返しのない 2 元配置データに戻って各種の推定にチャレンジする．表 26 に各水準の平均と基準となる水準との差の平均値を計算した結果を示す．

表 26 水準平均と基準水準からの差

|                | B <sub>1</sub> | B <sub>2</sub> | 平均      | 差      |
|----------------|----------------|----------------|---------|--------|
| A <sub>1</sub> | 10.9           | 11.5           | 11.2000 | 規準     |
| A <sub>2</sub> | 12.4           | 12.9           | 12.6500 | 1.4500 |
| A <sub>3</sub> | 12.2           | 12.3           | 12.2500 | 1.0500 |
| 平均             | 11.8333        | 12.2333        | 12.0333 |        |
| 差              | 規準             | 0.4000         |         |        |
|                |                |                |         |        |

因子 A の各水準の平均は，因子 B の 2 つの水準の影響が含まれており，平均値そのものの自体はピュアな推定値とはいえない．ただし，A<sub>1</sub> 水準と A<sub>2</sub> 水準との差は，因子 B の影響が除去され，信頼できる推定値になっている．そこで，その差 1.450 の 95 %信頼区間を求めたい．ここで，母数  $\theta_0, \theta_1, \dots, \theta_q (q \leq p)$  に対し任意の線形和  $L^{(j)} = \sum_{k=0}^q l_k^{(j)} \theta_k$  に着目する．因子 A<sub>1</sub> 水準と A<sub>2</sub> 水準の差の推定値と分散を求めて 95%信頼区間を求める．さらに，因子 A<sub>1</sub> 水準と A<sub>3</sub> 水準の差についても同様に求めたい．

表 27 に示すように因子 B の水準をプールした因子 A<sub>1</sub> のデザイン変数ベクトル

$$A_1 : l^{(7)} = [1 \ 1 \ 0 \ 0] \quad (57)$$

は，因子 B に対応するデザイン変数  $x_{3,i}$  について，因子 B の水準のダミー変数の平均

$$l_3^{(7)} = \frac{[B_1 : x_{3,i} = 1] + [B_2 : x_{3,i} = -1]}{2} = \frac{1 + (-1)}{2} = 0 \quad (58)$$

表 27 因子ごとの水準および差に対する推定値と 95%信頼区間

| L          | A                              | B              | $l_1$<br>$x_{0,i}$ | $l_2$<br>$x_{1,i}$ | $l_3$<br>$x_{2,i}$ | $l_4$<br>$x_{3,i}$ | $\theta^{\wedge}$                   | $L^{\wedge}$ | $Var(L^{\wedge})$ | $t \times SE$ | L95%   | U95%   |
|------------|--------------------------------|----------------|--------------------|--------------------|--------------------|--------------------|-------------------------------------|--------------|-------------------|---------------|--------|--------|
| $L^{(7)}$  | A <sub>1</sub>                 |                | 1                  | 1                  | 0                  | 0                  | 12.033                              | 11.200       | 0.018             | 0.569         | 10.631 | 11.769 |
| $L^{(8)}$  | A <sub>2</sub>                 |                | 1                  | 0                  | 1                  | 0                  | -0.833                              | 12.650       | 0.018             | 0.569         | 12.081 | 13.219 |
| $L^{(9)}$  | A <sub>3</sub>                 |                | 1                  | -1                 | -1                 | 0                  | 0.617                               | 12.250       | 0.018             | 0.569         | 11.681 | 12.819 |
| $L^{(10)}$ | A <sub>1</sub> -A <sub>1</sub> |                | 0                  | 0                  | 0                  | 0                  | -0.200                              | 0.000        | 0.000             | 0.000         | 0.000  | 0.000  |
| $L^{(11)}$ | A <sub>2</sub> -A <sub>1</sub> |                | 0                  | -1                 | 1                  | 0                  |                                     | 1.450        | 0.035             | 0.805         | 0.645  | 2.255  |
| $L^{(12)}$ | A <sub>3</sub> -A <sub>1</sub> |                | 0                  | -2                 | -1                 | 0                  |                                     | 1.050        | 0.035             | 0.805         | 0.245  | 1.855  |
| $L^{(13)}$ |                                | B <sub>1</sub> | 1                  | 0                  | 0                  | 1                  |                                     | 11.833       | 0.012             | 0.465         | 11.369 | 12.298 |
| $L^{(14)}$ |                                | B <sub>2</sub> | 1                  | 0                  | 0                  | -1                 |                                     | 12.233       | 0.012             | 0.465         | 11.769 | 12.698 |
| $L^{(15)}$ | B <sub>1</sub> -B <sub>1</sub> |                | 0                  | 0                  | 0                  | 0                  |                                     | 0.000        | 0.000             | 0.000         | 0.000  | 0.000  |
| $L^{(16)}$ | B <sub>2</sub> -B <sub>1</sub> |                | 0                  | 0                  | 0                  | -2                 |                                     | 0.400        | 0.023             | 0.657         | -0.257 | 1.057  |
|            |                                |                |                    |                    |                    |                    | $t_{0.05}(6-3-1)=T.inv.2T(0.05,2)=$ | 4.303        |                   |               |        |        |

が 0 となることにより水準平均および分散が求められる．他の  $A_2$  および  $A_3$  についても表 27 に示すように計算される．各線形和  $L^{(j)}$  の分散は，パラメータの共分散行列  $\Sigma(\hat{\theta})$  を用いて式 (53) を用いて計算している

水準  $A_1$  を基準とした  $(A_2 - A_1)$  のデザイン変数のベクトル  $I^{(11)}$  は，

$$(A_2 - A_1) \begin{cases} I^{(11)} = I^{(8)} - I^{(7)} \\ = [1 \ 0 \ 1 \ 0] - [1 \ 1 \ 0 \ 0] \\ = [0 \ -1 \ 1 \ 0] \end{cases} \quad (59)$$

となり，同様に  $(A_3 - A_1)$  は，

$$(A_3 - A_1) \begin{cases} I^{(12)} = I^{(9)} - I^{(7)} \\ = [1 \ -1 \ -1 \ 0] - [1 \ 1 \ 0 \ 0] \\ = [0 \ -2 \ -1 \ 0] \end{cases} \quad (60)$$

となり，推定値，分散，95%信頼区間が計算されている．

因子 B については，因子 A についてのデザイン変数  $x_{1,i}$  および  $x_{2,i}$  について，因子 A の水準のダミー変数の平均は，それぞれ 0 になることから，

$$B_1 : I^{(13)} = [1 \ 0 \ 0 \ 1] \quad (61)$$

であり，水準  $B_1$  を基準とした  $(B_2 - B_1)$  のデザイン変数のベクトルは

$$(B_2 - B_1) \begin{cases} I^{(16)} = I^{(14)} - I^{(13)} \\ = [1 \ 0 \ 0 \ -1] - [1 \ 0 \ 0 \ 1] \\ = [0 \ 0 \ 0 \ -2] \end{cases} \quad (62)$$

とする．これらの計算は，表 25 の  $A_1 B_1$  に対する  $I^{(1)}$  の計算式をコピーしたものである．このように，求めたい推定値に対し，対応するデザイン変数ベクトル  $I$  を設定するだけで半自動的に計算することができる．

表 27 で推定した各因子の水準平均と 95%信頼区間，基準となる水準との差の推定値と 95%信頼区間を図 2 に示す．図 1 の因子を組合せた場合に比べ 95%信頼区間の幅が小さくなっていることが確認される．

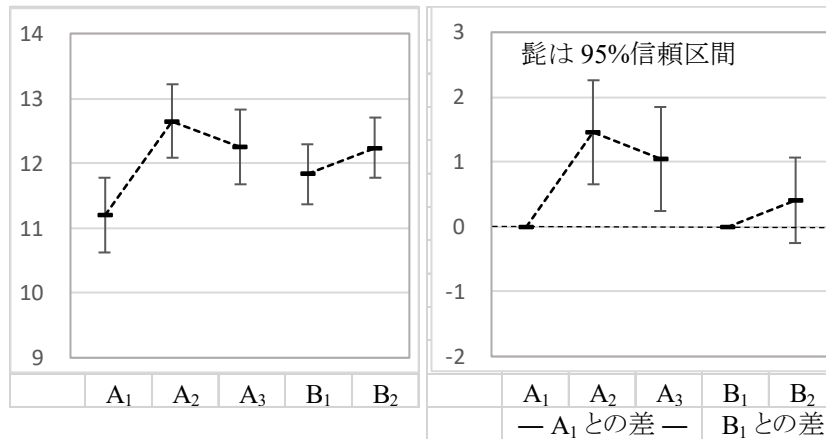


図2 因子 A および B の水準平均および第 1 水準と差の予測プロファイル

### 現行水準と最適水準の差および 95%信頼区間

実験の目的が、現行の作業条件から何らかの改善をしたいとする．現行水準が  $A_1B_1$  とし、あった場合に、最適水準が  $A_2B_2$  であった場合に、その差の 95%信頼区間を算出し、統計的な判断をしたい．2つの組合せ平均の推定値および分散は、表 25 で既に計算している．表 28 に示すように、それぞれの行ベクトル  $l^{(4)}$  と  $l^{(1)}$  の差は、

$$\begin{aligned} l^{(17)} &= l^{(4)} - l^{(1)} \\ &= [1 \ 0 \ 1 - 1] - [1 \ 1 \ 0 \ 1] \\ &= [0 - 1 \ 1 - 2] \end{aligned} \quad (63)$$

となり、

$$\hat{Var}[\hat{L}^{(17)}] = \sum_k \sum_{k'} l_k^{(17)} l_{k'}^{(17)} c_{k,k'} \hat{\sigma}^2 = l^{(17)} \Sigma(\hat{\theta}) l^{(17)T} = 0.058 \quad (64)$$

が計算され、差の 95%信頼区間は、(0.811, 2.889) と 0 を含んでいないので、有意な差であると判断される．

表 28 現行水準と最適水準の差の 95%信頼区間

|            |                |                | $l_1$     | $l_2$     | $l_3$     | $l_4$     |                 | $y^\wedge$        |                 |       |         |         |
|------------|----------------|----------------|-----------|-----------|-----------|-----------|-----------------|-------------------|-----------------|-------|---------|---------|
| $l$        | A              | B              | $x_{0,i}$ | $x_{1,i}$ | $x_{2,i}$ | $x_{3,i}$ | $\theta^\wedge$ | $L^\wedge$        | $Var(L^\wedge)$ | $SE$  | $L95\%$ | $U95\%$ |
| $l^{(1)}$  | A <sub>1</sub> | B <sub>1</sub> | 1         | 1         | 0         | 1         | 12.033          | = 11.000          | 0.023           | 0.153 | 10.343  | 11.657  |
| $l^{(4)}$  | A <sub>2</sub> | B <sub>2</sub> | 1         | 0         | 1         | -1        | -0.833          | 12.850            | 0.023           | 0.153 | 12.193  | 13.507  |
|            | 差 0            |                | 0         | 0         | 0         | 0         | 0.617           | 0.000             | 0.000           | 0.000 | 0.000   | 0.000   |
| $l^{(17)}$ | 差              |                | 0         | -1        | 1         | -2        | -0.200          | 1.850             | 0.058           | 0.242 | 0.811   | 2.889   |
|            |                |                |           |           |           |           |                 | $t(0.05, 6-3-1)=$ |                 | 4.303 |         |         |

図 3 に Excel の折れ線グラフを用いて、現行水準と最適水準との差の予測プロファイルを示す．

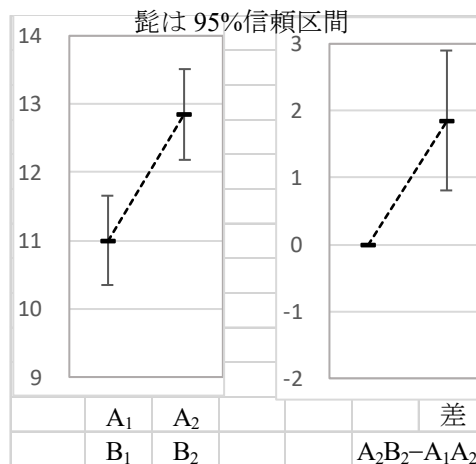


図3 現行水準と最適水準との差の予測プロファイル

これまで示してきた Excel による結果については、JMP の「モデルのあてはめ」によって検証してあるが、JMP の結果を示すことは冗長になるので、現行水準、最適水準、それらの差について、表 28 に示した線形和  $L$  に対するベクトル  $\mathbf{l}$  を用いた「カスタム検定」の結果を表 29 に示す。線形和および  $SE$  が表 28 の結果に一致していることが確認される。

表 29 JMP のカスタム検定を用いた検証

| カスタム検定        |              |                               |                               |   |
|---------------|--------------|-------------------------------|-------------------------------|---|
| 現行 最適 差       |              |                               |                               |   |
| パラメータ         |              | A <sub>1</sub> B <sub>1</sub> | A <sub>2</sub> B <sub>2</sub> | A <sub>2</sub> B <sub>2</sub> - A <sub>1</sub> B <sub>1</sub> |
| 切片            |              | 1                             | 1                             | 0   |
| A[A1]         |              | 1                             | 0                             | -1  |
| A[A2]         |              | 0                             | 1                             | 1   |
| B[B1]         |              | 1                             | -1                            | -2  |
| =             |              | 0                             | 0                             | 0   |
| 線形和 $\hat{L}$ | 値            | 11.0000                       | 12.8500                       | 1.8500  |
| SE            | 標準誤差         | 0.1528                        | 0.1528                        | 0.2415  |
|               | t値           | 72.0119                       | 84.1230                       | 7.6597  |
|               | p値(Prob> t ) | 0.0002                        | 0.0001                        | 0.0166  |
|               | 平方和          | 181.5000                      | 247.6838                      | 2.0535  |

なお、JMP のカスタム検定については、スライド 13 およびスライド 26 に直交表での事例が示されているので参照のこと。

## 10. Excel を用いた新たな平方和の分解による解析

平方和の分解による伝統的な分散分析については、これまで例示してこなかった。表 16 に示した繰り返しのない 2 元配置データを用い、Excel の計算機能を活用した平方和の分解による新たな分散分析を行なう。表 30 に示す 2 元表の周辺に各因子の水準平均が示され、総平均 12.033 からの差（効果）も示されている。

表 30 2 元配置のデータに対する構造モデルの適用

|                               | B <sub>1</sub>                | B <sub>2</sub>                | 平均     | $\bar{y}_{i.}$ | $\alpha_i$ | $\bar{y}_{i.} - \bar{y}_{..}$ |
|-------------------------------|-------------------------------|-------------------------------|--------|----------------|------------|-------------------------------|
| A <sub>1</sub>                | 10.9                          | 11.5                          | 11.200 | $\bar{y}_{1.}$ | -0.833     | $\bar{y}_{1.} - \bar{y}_{..}$ |
| A <sub>2</sub>                | 12.4                          | 12.9                          | 12.650 | $\bar{y}_{2.}$ | 0.617      | $\bar{y}_{2.} - \bar{y}_{..}$ |
| A <sub>3</sub>                | 12.2                          | 12.3                          | 12.250 | $\bar{y}_{3.}$ | 0.217      | $\bar{y}_{3.} - \bar{y}_{..}$ |
| 平均                            | 11.833                        | 12.233                        | 12.033 |                |            |                               |
| $\bar{y}_{.j}$                | $\bar{y}_{.1}$                | $\bar{y}_{.2}$                |        | $\bar{y}_{..}$ |            |                               |
| $\beta_j$                     | -0.200                        | 0.200                         |        |                |            |                               |
| $\bar{y}_{.j} - \bar{y}_{..}$ | $\bar{y}_{.1} - \bar{y}_{..}$ | $\bar{y}_{.2} - \bar{y}_{..}$ |        |                |            |                               |

手計算を前提にした場合には、できるだけ平均値の計算のような割り算を極力排除している。それは、表 30 に示すように「各水準の平均」ではなく「各水準の和」を求めて展開するのが常である。しかし、Excel を用いた計算では、逆に煩雑になり、見通しが悪くなるので平方和の分解を素直に実現するために、各水準の平均値を用いた計算方法が適している。

### 構造モデルから平方和の分解へ

式 (22) に示した構造モデルは、次のようであった。

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{正規分布} \quad (65)$$

$$\text{ただし, } \sum_i \alpha_i = 0, \sum_j \beta_j = 0, \sum_i \varepsilon_{ij} = 0, \sum_j \varepsilon_{ij} = 0, \sum_i \sum_j \varepsilon_{ij} = 0$$

この構造モデルを線形モデル化し、デザイン行列を用い、正規方程式を解き、 $\alpha_i$  と  $\beta_j$  の推定を推定する方法を第 9 章で示した。ここでは、表 30 の周辺に示した各因子の水準平均を用いた解析方法を示す。

因子 A と因子 B は互いにバランスが取れ、互いに直交しているので、 $\hat{\mu}$  が総平均  $\bar{y}_{..}$  と等しくなり、因子 A の効果  $\alpha_i$  は、それぞれの平均  $\bar{y}_{i.}$  から総平均  $\bar{y}_{..}$  を引いて表すことができる。因子 B についても同様である。構造モデル式 (65) は、

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + \varepsilon_{ij} \quad (66)$$

として表することができる．この式から， $\varepsilon_{ij}$  は，

$$\varepsilon_{ij} = y_{ij} - [\bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})] \quad (67)$$

として求めることができる．式 (66) の  $\bar{y}_{..}$  を左辺に移すと

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + \varepsilon_{ij} \quad (68)$$

が得られる．式 (68) を左辺と右辺を 2 乗し全ての  $i$  と  $j$  について加えた場合について整理すると，式の展開は省略するが，次式のように，簡単化することができる．

$$\left. \begin{aligned} \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 &= \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_i \sum_j (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_i \sum_j (\varepsilon_{ij})^2 \\ S_T &= S_A + S_B + S_e \end{aligned} \right\} \quad (69)$$

それぞれの  $y_{ij}$  から総平均  $\bar{y}_{..}$  を引いた偏差平方和  $S_T$  が，因子 A の偏差平方和  $S_A$ ，因子 B の偏差平方和  $S_B$ ，残差平方和  $S_e$  の和として表すことができる．平方和  $S_T$  を ( $S_A$ ， $S_B$ ， $S_e$ ) の平方和の和に分解することを「平方和の分解」と言う．

実際に平方和の分解が成り立つか確かめてみよう．表 31 に示したのは，実際のデータ  $y_{ij}$  を ( $\mu$ ， $\alpha_i$ ， $\beta_j$ ， $\varepsilon_{ij}$ ) に分解し，それらを足し合わせた結果が元の  $y_{ij}$  となることが示されている．

表 31 実際のデータを構造モデル化した場合

|                |                |          | $\mu + \alpha_i$ | $\mu + \beta_j$ | $\mu$  | $\alpha_i$                    | $\beta_j$                     | $\varepsilon_{ij}$ |          |
|----------------|----------------|----------|------------------|-----------------|--|-------------------------------|-------------------------------|--------------------|----------|
| A              | B              | $y_{ij}$ | $\bar{y}_{i.}$   | $\bar{y}_{.j}$  | $\bar{y}_{..}$   | $\bar{y}_{i.} - \bar{y}_{..}$ | $\bar{y}_{.j} - \bar{y}_{..}$ | $\varepsilon_{ij}$ | $y_{ij}$ |
| A <sub>1</sub> | B <sub>1</sub> | 10.9     | 11.200           | 11.833          | 12.033   | -0.833                        | -0.200                        | -0.100             | 10.90    |
| A <sub>1</sub> | B <sub>2</sub> | 11.5     | 11.200           | 12.233          | 12.033   | -0.833                        | 0.200                         | 0.100              | 11.50    |
| A <sub>2</sub> | B <sub>1</sub> | 12.4     | 12.650           | 11.833          | 12.033   | 0.617                         | -0.200                        | -0.050             | 12.40    |
| A <sub>2</sub> | B <sub>2</sub> | 12.9     | 12.650           | 12.233          | 12.033   | 0.617                         | 0.200                         | 0.050              | 12.90    |
| A <sub>3</sub> | B <sub>1</sub> | 12.2     | 12.250           | 11.833          | 12.033   | 0.217                         | -0.200                        | 0.150              | 12.20    |
| A <sub>3</sub> | B <sub>2</sub> | 12.3     | 12.250           | 12.233          | 12.033   | 0.217                         | 0.200                         | -0.150             | 12.30    |
|                |                |          |                  |                 | $y_{ii} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + \varepsilon_{ii}$ |                               |                               |                    | 検算       |

表 32 に示したのは，表 31 で計算された ( $\alpha_i$ ， $\beta_j$ ， $\varepsilon_{ij}$ ) のそれぞれの平方和が計算されている．さらに，それらの平方和の合計 ( $S_A + S_B + S_e$ ) が，データ  $y_{ij}$  から総平方和  $\bar{y}_{i.}$  を引いた偏差平方和  $S_T$  になることを

$$\left. \begin{aligned} S_T &= S_A + S_B + S_e \\ &= 2.243 + 0.240 + 0.070 \\ &= 2.553 \end{aligned} \right\} \quad (70)$$

として示している．

表 32 平方和の分解が成り立っていることの検証

|                |                |          | $y_{ij}-\mu$          | $\alpha_i$                  | $\beta_j$                   | $\varepsilon_{ij}$ |       |
|----------------|----------------|----------|-----------------------|-----------------------------|-----------------------------|--------------------|-------|
| A              | B              | $y_{ij}$ | $y_{ij}-\bar{y}_{..}$ | $\bar{y}_{i.}-\bar{y}_{..}$ | $\bar{y}_{.j}-\bar{y}_{..}$ | $\varepsilon_{ij}$ | 検証    |
| A <sub>1</sub> | B <sub>1</sub> | 10.9     | -1.133                | -0.833                      | -0.200                      | -0.100             |       |
| A <sub>1</sub> | B <sub>2</sub> | 11.5     | -0.533                | -0.833                      | 0.200                       | 0.100              |       |
| A <sub>2</sub> | B <sub>1</sub> | 12.4     | 0.367                 | 0.617                       | -0.200                      | -0.050             |       |
| A <sub>2</sub> | B <sub>2</sub> | 12.9     | 0.867                 | 0.617                       | 0.200                       | 0.050              |       |
| A <sub>3</sub> | B <sub>1</sub> | 12.2     | 0.167                 | 0.217                       | -0.200                      | 0.150              |       |
| A <sub>3</sub> | B <sub>2</sub> | 12.3     | 0.267                 | 0.217                       | 0.200                       | -0.150             |       |
|                |                | 平方和      | 2.553                 | 2.243                       | 0.240                       | 0.070              | 2.553 |
|                |                |          | $S_T$                 | $S_A$                       | $S_B$                       | $S_e$              | 合計    |

分解された平方和を元にして、表 33 に示すように分散分析表を作成する。この結果、因子 A の 3 水準間に有意な差 ( $p=0.030$ ) があり、因子 B の 2 水準間には有意な差がないことが分かる。

表 33 2 元配置に対する分散分析表

| 要因 | 自由度 | 平方和   | 平均平方  | F 値    | p 値   |    |
|----|-----|-------|-------|--------|-------|----|
| A  | 2   | 2.243 | 1.122 | 32.048 | 0.030 | *  |
| B  | 1   | 0.240 | 0.240 | 6.857  | 0.120 | NS |
| 残差 | 2   | 0.070 | 0.035 |        |       |    |
| 合計 | 5   | 2.553 |       |        |       |    |

このようにまとめられた分散分析表により、各因子について統計的な有意性について総合的に判断を行いことができる。ただし、どのような差なのか検討する際には、表 30 に示した 2 元表の周辺の算術平均を主体にし、分散分析表の残差の平均平方（誤差分散  $\hat{\sigma}^2 = 0.035$ ）を用い、各因子の水準平均の分散を推定し、95%信頼区間を図示することが必要となる。

### 水準平均の分散

各因子の水準平均の標準誤差、水準間の差の標準誤差を算出し、これらの 95%を推定したい。さらに、現行水準 A<sub>1</sub>B<sub>1</sub> と最適水準 A<sub>2</sub>B<sub>2</sub> 標準誤差、これらの差の標準誤差を算出したい。表 34 に示すように、因子 A の A<sub>i</sub>水準平均は、式 (65) から、

$$\bar{y}_{i.} = \left( \frac{y_{i1} + y_{i2}}{2} \right) = \hat{\mu} + \hat{\alpha}_i \quad (71)$$

であり、分散  $Var(\hat{y}_{i.})$  は、 $y_{i1}$  と  $y_{i2}$  が互いに独立なので、

$$Var(\bar{y}_{i.}) = Var\left(\frac{y_{i1} + y_{i2}}{2}\right) = \frac{\hat{\sigma}^2 + \hat{\sigma}^2}{2^2} = \frac{\hat{\sigma}^2}{2} \quad (72)$$



となり，誤差分散  $\hat{\sigma}^2$  の  $1/2$  となる．このように，誤差分散  $\hat{\sigma}^2$  に対する  $1/2$  の逆数，すなわち有効反復数  $n_e = 2$  と定義する．この有効反復数  $n_e$  を用いて，表 34 に示すように，各因子の水準に対する 95%信頼区間を計算することができる．

表 34 各因子の水準平均および水準間平均の 95%信頼区間

|                                | 水準<br>平均                | 有効<br>反復数 | 分散     | SE              | L 95%  | U 95%  |
|--------------------------------|-------------------------|-----------|--------|-----------------|--------|--------|
| A <sub>1</sub>                 | 11.200                  | 2         | 0.0175 | 0.1323          | 10.631 | 11.769 |
| A <sub>2</sub>                 | 12.650                  | 2         | 0.0175 | 0.1323          | 12.081 | 13.219 |
| A <sub>3</sub>                 | 12.250                  | 2         | 0.0175 | 0.1323          | 11.681 | 12.819 |
| A <sub>2</sub> -A <sub>1</sub> | 1.450                   | 1         | 0.0350 | 0.1871          | 0.645  | 2.255  |
| B <sub>1</sub>                 | 11.833                  | 3         | 0.0117 | 0.1080          | 11.369 | 12.298 |
| B <sub>2</sub>                 | 12.233                  | 3         | 0.0117 | 0.1080          | 11.769 | 12.698 |
| B <sub>2</sub> -B <sub>1</sub> | 0.400                   | 1.5       | 0.0233 | 0.1528          | -0.257 | 1.057  |
|                                | 誤差分散 $\hat{\sigma}^2 =$ |           | 0.0350 | $t_{0.05}(2) =$ | 4.3027 |        |

因子 A の 2 つの水準平均の差，A<sub>2</sub> 水準と A<sub>1</sub> 水準の差は，

$$\left. \begin{aligned} \bar{y}_{2\cdot} - \bar{y}_{1\cdot} &= \frac{y_{21} + y_{22}}{2} - \frac{y_{11} + y_{12}}{2} \\ &= (\hat{\mu} + \hat{\alpha}_2) - (\hat{\mu} + \hat{\alpha}_1) = \hat{\alpha}_2 - \hat{\alpha}_1 \end{aligned} \right\} \quad (73)$$

であり，これらの  $y_{ij}$  は互いに独立なので，分散は，

$$Var(\bar{y}_{2\cdot} - \bar{y}_{1\cdot}) = \frac{2\hat{\sigma}^2}{2^2} + \frac{2\hat{\sigma}^2}{2^2} = \frac{\hat{\sigma}^2}{1} \quad (74)$$

となり，有効反復数は  $n_e = 1$  となる．

因子 B の水準平均は，式 (65) から，

$$\bar{y}_{\cdot j} = \left( \frac{y_{1j} + y_{2j} + y_{3j}}{3} \right) = \hat{\mu} + \hat{\beta}_j \quad (75)$$

となり，分散  $Var(\bar{y}_{\cdot j})$  は，

$$Var(\bar{y}_{\cdot j}) = Var\left(\frac{y_{1j} + y_{2j} + y_{3j}}{3}\right) = \frac{\hat{\sigma}^2 + \hat{\sigma}^2 + \hat{\sigma}^2}{3^2} = \frac{\hat{\sigma}^2}{3} \quad (76)$$

となり，有効反復数は  $n_e = 3$  となる．第 2 水準と第 1 水準の差の分散は

$$\left. \begin{aligned} Var(\hat{y}_{2\cdot} - \hat{y}_{1\cdot}) &= Var\left(\frac{y_{12} + y_{22} + y_{32}}{3} - \frac{y_{11} + y_{21} + y_{31}}{3}\right) \\ &= \frac{3\hat{\sigma}^2}{3^2} + \frac{3\hat{\sigma}^2}{3^2} = \frac{\hat{\sigma}^2}{1.5} \end{aligned} \right\} \quad (77)$$

となり，有効反復数は  $n_e = 1.5$  となる．

## 組合せ平均の推定値

さて、2つの因子を組合せた場合の推定値として  $A_1B_1$  の測定値  $y_{1,1}=10.9$  として良いのだろうか。平方和の分解の過程で、

$$\left. \begin{aligned} y_{ij} &= \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + \varepsilon_{ij} \\ y_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \varepsilon_{ij} \\ y_{ij} &= \hat{y}_{ij} + \varepsilon_{ij} \end{aligned} \right\} \quad (78)$$

としたので、 $A_1B_1$  の測定値  $y_{1,1}=10.9$  には誤差  $\varepsilon_{1,1}$  が含まれている。したがって、誤差  $\varepsilon_{1,1}$  が含まれない推定値  $\hat{y}_{ij}$  を使いたい。表 35 に推定値  $\hat{y}_{ij}$  の計算結果を示す。 $A_1B_1$  の測定値は、 $y_{1,1}=10.9$  であるが、推定値は  $\hat{y}_{1,1}=11.000$  となる。ところで、分散  $Var(\hat{y}_{1,1})$  の有効反復数  $n_e$  は、幾つになるのだろうか。

表 35 組合せ平均の推定値

|                | B <sub>1</sub> | B <sub>2</sub> | 平均     | 効果 $\alpha$    |
|----------------|----------------|----------------|--------|----------------|
| A <sub>1</sub> | 10.9           | 11.5           | 11.200 | -0.833         |
| 推定値            | 11.000         | 11.400         |        |                |
| A <sub>2</sub> | 12.4           | 12.9           | 12.650 | 0.617          |
| 推定値            | 12.450         | 12.850         |        |                |
| A <sub>3</sub> | 12.2           | 12.3           | 12.250 | 0.217          |
| 推定値            | 12.050         | 12.450         |        |                |
| 平均             | 11.833         | 12.233         | 12.033 |                |
| 効果 $\beta$     | -0.200         | 0.200          |        | $\bar{y}_{..}$ |

式 (78) から、組合せ平均の推定値は  $\hat{y}_{1,1} = \hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1$  なので分散  $Var(\hat{y}_{1,1})$  を求めるためには、個々の  $y_{ij}$  の線形和に戻し、 $Var(y_{ij}) = \hat{\sigma}^2$  に戻って分散の加法性を用いて合成する必要がある。これは、 $\hat{\alpha}_1$  を推定する際の  $(\bar{y}_{1.} - \bar{y}_{..})$  の  $\bar{y}_{1.}$  と  $\bar{y}_{..}$  には共通の  $(y_{11}, y_{12})$  が含まれており、互いに独立ではないので分散の加法性が成り立たないためである。同様に  $\hat{\beta}_1$  を加えたのであるから、個々の  $y_{ij}$  に再編成し分散の加法性を適用する必要がある。式 (79) に示すように、細かな計算ではあるが、最終的に分散の加法性が適用できるようにまとめられる。

$$\left. \begin{aligned} Var(\hat{y}_{1,1}) &= Var(\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1) \\ &= Var[\bar{y}_{..} + (\bar{y}_{1.} - \bar{y}_{..}) + (\bar{y}_{.1} - \bar{y}_{..})] \\ &= Var(\bar{y}_{1.} + \bar{y}_{.2} - \bar{y}_{..}) \\ &= Var\left(\frac{y_{11} + y_{12}}{2} + \frac{y_{11} + y_{21} + y_{31}}{3} - \frac{y_{11} + y_{12} + y_{21} + y_{22} + y_{31} + y_{32}}{6}\right) \\ &= Var\left(\frac{4y_{11}}{6} + \frac{2y_{12}}{6} + \frac{y_{21}}{6} + \frac{y_{31}}{6} - \frac{y_{22}}{6} - \frac{y_{32}}{6}\right) \\ &= \frac{4^2 + 2^2 + 4 \times 1^2}{6^2} \hat{\sigma}^2 = \frac{24}{36} \hat{\sigma}^2 = \frac{\hat{\sigma}^2}{1.50} \end{aligned} \right\} \quad (79)$$

この結果から、 $Var(\hat{y}_{1,1})$ の有効反復数は、 $n_e=1.50$ であることが分かる。最適水準  $A_2B_2$  の推定値は、 $\hat{y}_{2,2}=12.850$ であり、式 (79) と同様に分散  $Var(\hat{y}_{2,2})$  の有効反復数は、 $n_e=1.50$  であり、表 36 に示すように

$$Var(\hat{y}_{2,2}) = \frac{\hat{\sigma}^2}{1.50} = \frac{0.0350}{1.50} = 0.0233 \quad (80)$$

となる。この分散は、表 25 に示されている線形和  $L$  の分散で計算された結果に一致している。

### 現行水準と最適水準の差

さて、現行水準と最適水準の差は、正確には、（現行水準の組合せ）と（最適水準の組合せ）と言うべきであるが、表 35 の推定値を用いて

$$\hat{y}_{2,2} - \hat{y}_{1,1} = 12.850 - 11.000 = 1.850 \quad (81)$$

となる。ただし、分散  $Var(\hat{y}_{2,2} - \hat{y}_{1,1})$  の  $\hat{y}_{2,2}$  と  $\hat{y}_{1,1}$  には、共通の  $y_{2,1}$  と  $y_{1,2}$  が含まれているので、互いに独立ではなく、分散の加法性が成り立たないため、個々の  $y_{ij}$  にもどって再編成する必要がある。

$$\left. \begin{aligned} Var(\hat{y}_{2,2} - \hat{y}_{1,1}) &= Var[(\hat{\mu} + \hat{\alpha}_2 + \hat{\beta}_2) - (\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1)] \\ &= Var[(\bar{y}_{2.} - \bar{y}_{..}) + (\bar{y}_{.2} - \bar{y}_{..}) - (\bar{y}_{1.} - \bar{y}_{..}) - (\bar{y}_{.1} - \bar{y}_{..})] \\ &= Var(\bar{y}_{2.} + \bar{y}_{.2} - \bar{y}_{1.} - \bar{y}_{.1}) \\ &= Var\left(\frac{y_{21} + y_{22}}{2} + \frac{y_{12} + y_{22} + y_{32}}{3} - \frac{y_{11} + y_{12}}{2} - \frac{y_{11} + y_{21} + y_{31}}{3}\right) \\ &= Var\left(-\frac{5y_{11}}{6} - \frac{y_{12}}{6} + \frac{y_{21}}{6} + \frac{5y_{22}}{6} - \frac{2y_{31}}{6} + \frac{2y_{32}}{6}\right) \\ &= \frac{2 \times 5^2 + 2 \times 2^2 + 2 \times 1^2}{6^2} \hat{\sigma}^2 = \frac{60}{36} \hat{\sigma}^2 = \frac{\hat{\sigma}^2}{0.60} \end{aligned} \right\} \quad (82)$$

式 (82) から、 $Var(\hat{y}_{2,2} - \hat{y}_{1,1})$  の有効反復数は、 $n_e=0.60$  となる。これらの結果をまとめ、表 36 に推定値の 95%信頼区間の計算結果を示す。この結果は、表 28 に示した線形モデルでの線形和に対する分散の推定結果に一致する。

表 36 現行水準と最適水準の推定値および差の 95%信頼区間（表 28 に一致）

|                   | $\bar{y}_{..}$ | $\bar{y}_{i.} - \bar{y}_{..}$ | $\bar{y}_{.j} - \bar{y}_{..}$ | 推定値                      | 有効   |        |                |        |        |
|-------------------|----------------|-------------------------------|-------------------------------|--------------------------|------|--------|----------------|--------|--------|
|                   | $\mu^{\wedge}$ | $\alpha^{\wedge}_i$           | $\beta^{\wedge}_j$            | $y^{\wedge}_{ij}$        | 反復数  | 分散     | SE             | L 95%  | U 95%  |
| $A_1B_1$          | 12.033         | -0.833                        | -0.200                        | 11.000                   | 1.50 | 0.0233 | 0.1528         | 10.343 | 11.657 |
| $A_2B_2$          | 12.033         | 0.617                         | 0.200                         | 12.850                   | 1.50 | 0.0233 | 0.1528         | 12.193 | 13.507 |
| $A_2B_2 - A_1B_1$ | 0.000          | 1.450                         | 0.400                         | 1.850                    | 0.60 | 0.0583 | 0.2415         | 0.811  | 2.889  |
|                   |                |                               |                               | 誤差分散 $\sigma^{\wedge 2}$ |      | 0.0350 | $t_{0.05}(2)=$ | 4.3027 |        |

## 田口の式(ルール)

現行水準の組合せの推定値の分散を求めるために、式 (79) で示したように個々の  $y_{ij}$  に戻って有効反復数  $n_e$  を計算したのであるが、簡便な算出式として田口の式 (ルール) 知られている。

$$\frac{1}{n_e} = \frac{1 + \text{推定に無視しない要因の自由度の和}}{\text{実験数}N} \quad (83)$$

式 (79) の  $Var(\hat{y}_{1,1}) = Var(\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1)$  に適用してみよう。この式には、因子 A と因子 B が含まれているので、それらの自由度は、(3-1), (2-1) なので、有効反復数  $n_e$  は、

$$n_e = \frac{6}{1 + (3-1) + (2-1)} = \frac{6}{4} = 1.5 \quad (84)$$

により  $n_e = 1.5$  として求まる。

さて、式 (82) の組合せ平均の差  $Var(\hat{y}_{2,2} - \hat{y}_{1,1})$  の場合は、どうなるのだろうか。

$$Var(\hat{y}_{2,2} - \hat{y}_{1,1}) = Var[(\hat{\mu} + \hat{\alpha}_2 + \hat{\beta}_2) - (\hat{\mu} + \hat{\alpha}_1 + \hat{\beta}_1)] \quad (85)$$

なので、「推定に無視しない要因の自由度の和」は、どのように考えたらよいであろうか。とりあえず、計算してみると

$$n_e = \frac{6}{1 + (3-1) + (2-1) + (3-1) + (2-1)} = \frac{6}{7} = 0.857 \quad (86)$$

と  $n_e = 0.857$  が計算されるのだが、個々の  $y_{ij}$  に戻って計算した  $n_e = 0.60$  と一致しない。したがって、田口の式 (ルール) は、この場合には使えない。

## 伊奈の式(ルール)

田口の式 (ルール) に加えて伊奈の式 (ルール) も知られている。

$$\frac{1}{n_e} = \text{点推定の式でデータの合計にかかっている係数の和} \quad (87)$$

式 (79) の  $Var(\hat{y}_{1,1}) = Var(\bar{y}_{1.} + \bar{y}_{.2} - \bar{y}_{..})$  のそれぞれの係数は、1/2, 1/3, 1/6 なので、有効反復数  $n_e$  の逆数は、

$$\left. \begin{aligned} \frac{1}{n_e} &= \frac{1}{2} + \frac{1}{3} - \frac{1}{6} \\ &= \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{1}{1.5} \end{aligned} \right\} \quad (88)$$

となり、 $n_e = 1.5$  が得られる。

さて、式 (82) の  $Var(\hat{y}_{2,2} - \hat{y}_{1,1})$  の場合は、どうなるのだろうか.

$$Var(\hat{y}_{2,2} - \hat{y}_{1,1}) = Var(\bar{y}_{2\cdot} + \bar{y}_{\cdot 2} - \bar{y}_{1\cdot} - \bar{y}_{\cdot 1}) \quad (89)$$

のそれぞれの計数は、 $1/2$ ,  $1/3$ ,  $1/2$ ,  $1/3$  なので,

$$\frac{1}{n_e} = \frac{1}{2} + \frac{1}{3} - \frac{1}{2} - \frac{1}{3} = 0 \quad (90)$$

となり、 $n_e$  を求めることができない. したがって、式 (82) に示したように個々の  $y_{ij}$  に戻って有効反復数  $n_e$  の計算をする必要がある.

## 11. 平方和の分解から線形モデルへ

### 絶滅の危機にある平方和の分解

平方和の分解により複数の因子の個々の分散を求め、その残りを誤差分散とし分散分析表にまとめる過程は、統計の魔術と言えるほど美しい。完成した分散分析表から、各因子の分散を誤差分散  $\sigma^2$  で除し  $F$  値を求め、 $p$  値を計算し統計的な判定を行う過程の華麗さに多くの人がしびれたに違いない。その結果として多くの実験デザインが考案され、さらに多くの人を虜にしてきたのであろう。しかしながら、PC 上で多くの統計ソフトが自由に使える現代では、電卓もない時代に手計算でもできる平方和の分解に基づいた分散分析の計算方法は、絶滅の危機に瀕しているのではないだろうか。

### Excel による平方和の分解

Excel の計算機能を活用するためには、表 31 に示したようにデータを行方向に展開し、構造モデルに即した計算表を作成することが望ましい。その結果として、表 32 に示したように平方和の計算が即座にでき、平方和の分解の理解にも役に立つ。各因子の水準平均の 95% 信頼区間のために必要な分散の計算も、表 34 に示すように「有効反復数」を与えることにより Excel のフィルハンドルを使用した計算式のコピペにより、ストレスなく計算できる。ここまでの計算ならば、「線形モデル」を使った計算よりも簡便であり、理解しやすいことは間違いなし、教育的でもある。

### 有効反復数の功罪

推定の応用課題である組合せ平均の 95% 信頼区間、2 つの組合せ平均間の差の 95% 信頼区間を求める際に、分散の加法性が使えなくなり、式 (79) および式 (82) に示したような煩雑な計算による有効反復数  $n_e$  を計算する必要となる。組合せ平均の推定の場合であれば、田口の式および伊奈の式により、簡便に有効反復数  $n_e$  を計算することができることを示したが、現行水準（の組合せ）と最適水準（の組合せ）の差についての有効反復数  $n_e$  には、対応できないことを示した。

### 線形モデルの有用性

第 9 章では、 $3 \times 2$  の 2 元配置データに対し、線形モデルによる解析で、表 28 に示したように現行水準（の組合せ）と最適水準（の組合せ）の差について線形和  $L$  に基づいた 95% 信

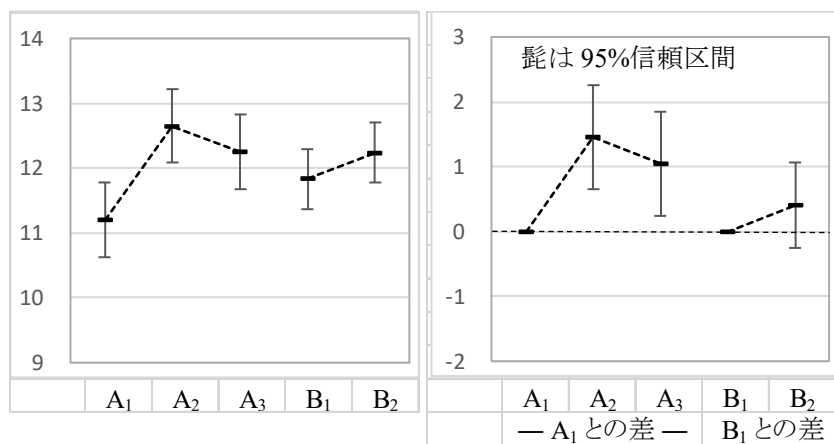
頼区間の計算結果を例示した。線形モデルを活用した解析では、式 (82) のような面倒な計算なし分散が推定できることは当たり前ではあるが、改めて線形モデルによる解析の優位性を実感できる。

### Excel の逆行列関数が救いの神

平方和の分解に基づく計算法の最大の難点は、対象とする因子または変数が、すべて名義尺でなければならないことである。そのために量的な変数の場合に対しても名義尺度にすることが強要される。量的な変数を扱える回帰分析の場合であっても、変数が 1 つの場合には手計算を前提にした偏差平方和をベースにした解析方法が整備されてきた。しかし、2 変数以上への拡張は、逆行列の計算が必須となり、手計算を前提にした定式化は絶望的である。また、共変量を含む共分散分析に対しても、平方和の分解に基づく計算ができない。このような観点から、伝統的な分散分析法にこだわることは、自ら解析能力の発展を阻害することになる。このような閉塞状態に薄明りを指しているのが、Excel の逆行列関を用いる線形モデルによる解析である。

### アリ地獄的な状況から抜け出すために

量的変数をあえて名義尺度化することにより、直交表に割り付けて多くの変数（因子）の解析が手計算でできることは、画期的であった。この成功体験が、伝統的な平方和の分解による分散分析法に絡み取られ、多くの人達がアリ地獄的な状況から抜け出せなくなってるに違いない。このセミナーにより、線形モデルを用いた分散分析法に興味を持って頂ける人たちが現れることを願っている。





## 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, 393-556, 446-7, 日本規格協会.
- 2) 奥野忠一, 久米均, 芳賀敏郎, 吉澤正(1981), 多変量解析法 <<改訂版>>, 71-78, 日科技連出版社.
- 3) 楠正, 辻谷将昭, 松本哲夫, 和田武夫(1995), 応用 実験計画法, 191-5, 206-8, 239-42, 日科技連出版社.
- 4) 新村秀一(1983a), 行列表現による重回帰分析(1), オペレーションズ・リサーチ学会, Vol 28, 439-45, [www.orsj.or.jp/~archive/pdf/bul/Vol.28\\_09\\_439.pdf](http://www.orsj.or.jp/~archive/pdf/bul/Vol.28_09_439.pdf).
- 5) 新村秀一(1983b), 行列表現による重回帰分析(2), オペレーションズ・リサーチ学会, Vol 28, 506-12, [www.orsj.or.jp/~archive/pdf/bul/Vol.28\\_10\\_506.pdf](http://www.orsj.or.jp/~archive/pdf/bul/Vol.28_10_506.pdf).
- 6) 新村秀一(1983c), 重回帰分析における掃き出し演算子, オペレーションズ・リサーチ学会, Vol 28, 585-89, [www.orsj.or.jp/~archive/pdf/bul/Vol.28\\_11\\_565.pdf](http://www.orsj.or.jp/~archive/pdf/bul/Vol.28_11_565.pdf).
- 7) 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, 289-333, 東大出版会.
- 8) 高橋行雄(2020), 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>.
- 9) 高橋行雄(2020), 最尤法によるポアソン回帰分析入門, 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>.
- 10) 高橋行雄(2021a), 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.
- 11) 高橋行雄(2021b), 線形モデルによる欠測値がある直交表の解析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, JMP discovery summit Japan 2021, <https://discoverysummit.jmp/ja/2021/japan/agenda.html>.
- 12) 田中豊, 垂水共之 編(1986), パソコン統計解析ハンドブック III 実験計画法編, 161-220, 共立出版.

| 文 献 索 引 |                                     |           |                              |
|---------|-------------------------------------|-----------|------------------------------|
|         | 朝香ら(1988) - 新版 品質管理便覧 第2版           |           | 2, 4, 15                     |
|         | 奥野ら(1981) - 多変量解析法 改訂版              |           | 38                           |
|         | 楠ら(1995) - 応用実験計画法                  |           | 2, 52, 58                    |
|         | 新村(1983a) - 行列表現による重回帰分析(1)         |           | 38                           |
|         | 新村(1983b) - 行列表現による重回帰分析(2)         |           | 38                           |
|         | 新村(1983c) - 重回帰分析における掃き出し演算子        |           | 38                           |
|         | 高橋ら(1989) - SASによる実験データの解析          |           | 3, 23, 50                    |
|         | 高橋(2020) - デザイン行列を用いた回帰分析入門         |           | 38                           |
|         | - パラメータの共分散分析の活用                    |           | 38                           |
|         | 高橋(2021) - 最尤法によるポアソン回帰分析入門         |           | 38                           |
|         | - 線形モデルによる欠測値がある直交表の解析              |           | 1                            |
|         | 高橋(準備中) - 層別因子を含む探索的な回帰分析入門         |           | 50                           |
|         | 田中ら(1986) - パソコン統計解析ハンドブックIII実験計画法編 |           | 6                            |
| 索 引     |                                     |           |                              |
| あ       | Avarage関数 - Excel                   | 39        | Mmult関数 - Excel 28           |
|         | アリ地獄 - 分散分析法                        | 48, 77    | - Excel 62                   |
| い       | 一次従属 - 一次独立                         | 5, 53     | - 行列の積 28                    |
|         | 一次独立 - 一次従属                         | 5, 53     | lsmeans - 最小2乗平均 13          |
|         | (1,0)型ダミー変数 - SAS/GLMプロシジャ          | 16, 46    | lsmeansステートメント - GLMプロシジャ 23 |
|         | (1,2)型 - ダミー変数                      | 15        | お 大きな丸括弧 - 行列 55             |
|         | (1,2)型ダミー変数 - 最小2乗平均                | 35        | 折れ線グラフ - Excel 44, 63        |
|         | - 線形モデル                             | 33        | か 回帰分析 - Excel 6             |
|         | (1,-1)基本型 - ダミー変数                   | 15        | 回帰モデル - 構造モデル 52             |
|         | (1,-1)対比型 - ダミー変数                   | 6         | - 線形モデル 52                   |
|         | 伊奈の式 - 有効反復数                        | 7, 36, 74 | Gauss-Doolittle法 - 逆行列 58    |
|         | 色分け - 引数の設定                         | 57        | 角括弧 - 行列 55                  |
|         | 因子 - 直交                             | 9         | カスタム検定 - 組合せ平均の差 20          |
| え       | Excel - Avarage関数                   | 39        | - 最適 - 現行の差 11               |
|         | - Minverse関数                        | 28        | - JMP 67                     |
|         | - Mmult関数                           | 28, 62    | ガラスの天井 - 単回帰分析 38            |
|         | - 折れ線グラフ                            | 44, 63    | き 逆行列 - Minverse関数 28        |
|         | - 回帰分析                              | 6         | - Minverse関数 57              |
|         | - 行列関数                              | 28        | - Gauss-Doolittle法 58        |
|         | - Transpose関数                       | 28        | - 積和行列 30                    |
|         | - Var.S関数                           | 39        | - 掃き出し計算 38                  |
|         | - 分析ツール                             | 6         | 共分散行列 - 線形モデル 37             |
|         | - 平方和の分解                            | 68        | - 対角要素 28                    |
|         | Excel 回帰分析 - 定数に0                   | 16        | - データの 39                    |
|         | Excelシート上 - 正規方程式                   | 29        | - パラメータ 28, 37, 40, 60       |
|         | Excelの回帰分析 - 重回帰分析                  | 30        | - 標準誤差SE 41                  |
|         | - 線形モデル                             | 27, 49    | 行列 - 丸括弧 55                  |
|         | - パラメータ推定                           | 28        | - 角括弧 55                     |
|         | - 理解の助け                             | 27        | - 外枠 太い 56                   |
|         | Excelの行列計算 - 引数の設定                  | 57        | - 表記法 55                     |
|         | estimateステートメント - 最小2乗平均            | 26, 47    | 行列関数 - Excel 28              |
|         | - GLMプロシジャ                          | 23, 46    | 行列計算言語 - SAS/IML 3           |
|         | F値 - t値から                           | 31        | 行列の積 - Mmult関数 28, 56        |
|         | Minverse関数 - Excel                  | 28        | 極度の標準化 - 単回帰分析 38            |
|         | - 逆行列                               | 28, 57    | く 組合せ水準 - 信頼区間95% 61         |
|         |                                     |           | 組合せ平均 - 現行水準 19              |
|         |                                     |           | - 最適水準 19                    |

|   |                            |        |                         |            |
|---|----------------------------|--------|-------------------------|------------|
|   | - 分散                       | 72     | JMP - カスタム検定            | 67         |
|   | - 有効反復数                    | 72     | - 切片なし                  | 16         |
|   | - 予測プロファイル                 | 19     | - 直交表                   | 8          |
|   | 組合せ平均の差 - 分散               | 73     | - 使勝手                   | 22         |
|   | - 有効反復数                    | 73     | - 発売1990                | 9          |
|   | 組合せ法 - 直交表                 | 4      | - モデルのあてはめ              | 2          |
|   | 組合せ平均の差 - カスタム検定           | 20     | - 有償                    | 22         |
| け | 計画行列 $X$ - コーディングのテーブル     | 13     | JMPのモデルのあてはめ - GLMプロシジャ | 50         |
|   | - デザイン行列 $X$               | 13     | 重回帰分析 - Excelの回帰分析      | 30         |
|   | 欠測値 - 正規方程式                | 29     | 定数に0 - Excel 回帰分析       | 16         |
|   | - 直交表                      | 2, 7   | 信頼区間95% - 組合せ水準         | 61         |
|   | - ビリオド「.」                  | 8      | - 現行水準                  | 66         |
|   | - 分散分析表                    | 31     | - 最適水準                  | 66         |
|   | - 補完                       | 9      | - 水準間の差                 | 64         |
|   | - 線形モデル                    | 14     | - 推定値                   | 36         |
|   | 検証 - 平方和の分解                | 70     | - 線形和                   | 42, 65     |
|   | 現行・最適水準の差 - 線形和            | 45     | す 水準間の差 - 信頼区間95%       | 64         |
|   | 現行水準 - 組合せ平均               | 19     | 水準平均 - 標準誤差             | 70         |
|   | - 信頼区間95%                  | 66     | 水準平均の差 - 有効反復数          | 71         |
|   | - 予測プロファイル                 | 10     | 推定値 - 信頼区間95%           | 36         |
|   | 原点を通る - 切片なし               | 16     | - 相関行列                  | 18         |
| こ | 構造 (DE) モデル - 線形モデル        | 4      | - 標準誤差SE                | 41         |
|   | 構造モデル - 回帰モデル              | 52     | 推定値の線形和 - 最小2乗平均        | 21         |
|   | - 線形モデル                    | 52     | せ 正規方程式 - Excelシート上     | 29         |
|   | - 2元配置                     | 52, 68 | - 欠測値                   | 29         |
|   | - 平方和の分解                   | 68     | - 残差平方和                 | 54         |
|   | コーディングのテーブル - 計画行列 $X$     | 13     | - デザイン行列 $X$            | 29, 55, 56 |
|   | - 分散分析                     | 17     | 正規方程式の解 - デザイン行列        | 58         |
|   | 誤差分散の推定 - 予測値              | 59     | 成分表示 - 直交表              | 7          |
|   | 誤差分散の推定値 - 残差の分散           | 28     | 制約条件 - 母数               | 5          |
| さ | 最小2乗平均 - (1,2) 型ダミー変数      | 35     | 積和行列 - 逆行列              | 30         |
|   | - estimateステートメント          | 26, 47 | 説明変数 - モデル効果の構成         | 8          |
|   | - lsmeans                  | 13     | 絶滅の危機 - 分散分析            | 76         |
|   | - 算術平均                     | 21     | 切片なし - 原点を通る            | 16         |
|   | - 推定値の線形和                  | 21     | - JMP                   | 16         |
|   | - 線形和の推定値                  | 42     | (0,1) 型 - ダミー変数         | 15         |
|   | - プロット                     | 12     | (0,1) 型ダミー変数 - 線形モデル    | 34         |
|   | - 方言 (JMP, SAS)            | 13     | 線形モデル - (1,2) 型ダミー変数    | 33         |
|   | 最適水準 - 組合せ平均               | 19     | - Excelの回帰分析            | 27         |
|   | - 信頼区間95%                  | 66     | - 回帰モデル                 | 52         |
|   | - 予測プロファイル                 | 10     | - 共分散行列                 | 37         |
|   | 最適-現行の差 - カスタム検定           | 11     | - 欠測値                   | 14         |
|   | SAS/GLM - 直交表の解析           | 23     | - 構造モデル                 | 52         |
|   | SAS/GLMプロシジャ - (1,0) 型     | 16     | - (0,1) 型ダミー変数          | 34         |
|   | - (1,0) 型ダミー変数             | 46     | - 直交表                   | 2, 4       |
|   | - nointersept              | 16     | - デザイン行列 $X$            | 55         |
|   | SAS/IML - 行列計算言語           | 3      | - 平方和の分解                | 76         |
|   | 算術平均 - 最小2乗平均              | 21     | - モデルのあてはめ              | 48         |
|   | 残差の分散 - 誤差分散の推定値           | 28     | 線形モデルの入門 - Excelの回帰分析   | 49         |
|   | 残差平方和 - 正規方程式              | 54     | 線形和 - 現行・最適水準の差         | 45         |
|   | - 偏微分                      | 54     | - 信頼区間95%               | 42, 65     |
| し | GLMプロシジャ - estimateステートメント | 23, 46 | - 分散                    | 60         |
|   | - lsmeansステートメント           | 23     | - 分散の計算                 | 38         |
|   | - 母数にむだ                    | 25     | 線形和 $L$ - 分散            | 43         |
|   | - JMPのモデルのあてはめ             | 50     | 線形和の推定値 - 最小2乗平均        | 42         |
|   | 実験計画法 - 伝統的                | 1, 48  |                         |            |

|   |                       |           |   |                       |           |
|---|-----------------------|-----------|---|-----------------------|-----------|
| そ | 相関行列 - 推定値            | 18        |   | 表記法 - 行列              | 55        |
|   | - データ                 | 18, 39    |   | 標準誤差 - 水準平均           | 70        |
|   | 外枠 太い - 行列            | 56        |   | 標準誤差SE - 共分散行列        | 41        |
| た | 対角要素 - 共分散行列          | 28        |   | - 推定値                 | 41        |
|   | 対角要素の平方根 - 標準誤差SE     | 41        |   | - 対角要素の平方根            | 41        |
|   | 田口の式 - 有効反復数          | 7         |   | ピリオド「.」 - 欠測値         | 8         |
|   | - 有効反復数               | 36, 74    | ふ | プロット - 最小2乗平均         | 12        |
|   | ダミー変数 - (1,2) 型       | 15        |   | 分散 - 組合せ平均            | 72        |
|   | - (1,-1) 基本型          | 15        |   | - 組合せ平均の差             | 73        |
|   | - (1,-1) 対比型          | 6         |   | - 線形和                 | 60        |
|   | - (0,1) 型             | 15        |   | - 線形和L                | 43        |
|   | - 名義尺度                | 14        |   | - 2次形式                | 61        |
|   | - 歴史的背景               | 15        |   | 分散の計算 - 線形和           | 38        |
|   | 単回帰分析 - ガラスの天井        | 38        |   | 分散分析 - アリ地獄的          | 48        |
|   | - 極度の標準化              | 38        |   | - コーディングのテーブル         | 17        |
| ち | 直交 - 因子               | 9         |   | - 絶滅の危機               | 76        |
|   | 直交表 - デザイン行列          | 6         |   | - 平方和の分解              | 9, 76     |
|   | - Basic言語             | 6         |   | 分散分析表 - 欠測値           | 31        |
|   | - $L_8$               | 3         | へ | 分散分析法 - アリ地獄          | 77        |
|   | - 組合せ法                | 4         |   | 平方和 - $t$ 値から         | 32        |
|   | - 欠測値                 | 2, 7      |   | 平方和の分解 - Excel        | 68        |
|   | - JMP                 | 8         |   | - 検証                  | 70        |
|   | - 成分表示                | 7         |   | - 構造モデル               | 68        |
|   | - 線形モデル               | 2, 7      |   | - 線形モデル               | 76        |
|   | 直交表の解析 - SAS/GLM      | 23        |   | - 統計の魔術               | 76        |
| つ | 使勝手 - JMP             | 22        |   | - 分散分析                | 9, 76     |
| て | DEモデル - 構造モデル         | 4         |   | - 偏差平方和               | 14, 69    |
|   | $t$ 値から - $F$ 値       | 31        |   | - 量的変数を名義尺度へ          | 77        |
|   | - 平方和                 | 32        |   | 変換 - 連続尺度へ            | 14        |
|   | データ - 相関行列            | 18        |   | 偏差平方和 - 平方和の分解        | 14, 69    |
|   | データの - 共分散行列          | 39        |   | 偏微分 - 残差平方和           | 54        |
|   | - 相関行列                | 39        | ほ | 方言 (JMP,SAS) - 最小2乗平均 | 13        |
|   | デザイン行列 - 正規方程式        | 56        |   | 補完 - 欠測値              | 9         |
|   | - 正規方程式の解             | 58        |   | 母数 - 制約条件             | 5         |
|   | - 直交表                 | 6         |   | - パラメータ               | 5         |
|   | - パラメータの推定            | 18        |   | 母数にむだ - GLMプロシジャ      | 25        |
|   | デザイン行列 $X$ - 計画行列 $X$ | 13        | む | 無償 - SAS/GLM          | 22        |
|   | - 正規方程式               | 29, 55    | め | 名義尺度 - ダミー変数          | 14        |
|   | - 線形モデル               | 55        | も | モデルのあてはめ - 線形モデル      | 48        |
|   | 転置行列 - Transpose関数    | 28, 56    |   | - 役割変数の選択             | 8         |
|   | 伝統的 - 実験計画法           | 1,,48     |   | モデル効果の構成 - 説明変数       | 8         |
| と | 統計の魔術 - 平方和の分解        | 76        | や | 役割変数の選択 - 反応変数 $y$    | 8         |
|   | Transpose関数 - Excel   | 28        | ゆ | 有効反復数 - 伊奈の式          | 7, 36, 74 |
|   | - 転置行列                | 28, 56    |   | - 組合せ平均               | 72        |
| に | 2元配置 - 構造モデル          | 52, 68    |   | - 組合せ平均の差             | 73        |
|   | - 線形モデル               | 52        |   | - 水準平均の差              | 71        |
|   | 2次形式 - 分散             | 61        |   | - 田口の式                | 7, 36, 74 |
| は | Var.S関数 - Excel       | 39        |   | 有償 - JMP              | 22        |
|   | 掃き出し計算 - 逆行列          | 38        | よ | 予測値 - 誤差分散の推定         | 59        |
|   | パラメータ - 共分散行列         | 28, 37 60 |   | 予測プロファイル - 組合せ平均      | 19        |
|   | - 母数                  | 5         |   | - 現行水準                | 10        |
|   | パラメータ推定 - Excelの回帰分析  | 28        | り | 理解の助け - Excelの回帰分析    | 27        |
|   | パラメータの - 共分散行列        | 40        |   | 量的変数を名義尺度へ - 平方和の分解   | 77        |
|   | パラメータの推定 - デザイン行列     | 18        | れ | 歴史的背景 - ダミー変数         | 15        |
|   | 反応変数 $y$ - 役割変数の選択    | 8         |   | 連続尺度へ - 変換            | 14        |
| ひ | 引数の設定 - 色分け           | 57        |   |                       |           |
|   | - Excelの行列計算          | 57        |   |                       |           |

## 解析用ファイル

|   |        |                      |
|---|--------|----------------------|
|  | 20 KB  | 線形モデル_02_直交表         |
|  | 9 KB   | 線形モデル_03_L8欠測値       |
|  | 82 KB  | 線形モデル_03_L8欠測値       |
|  | 5 KB   | 線形モデル_04_コーディングのテーブル |
|  | 58 KB  | 線形モデル_04_デザイン行列      |
|  | 3 KB   | 線形モデル_04_算術平均        |
|  | 1 KB   | 線形モデル_05_L8直交表_SAS   |
|  | 35 KB  | 線形モデル_06_Excel線形回帰   |
|  | 85 KB  | 線形モデル_07_各種の推定値      |
|  | 160 KB | 線形モデル_08_折れ線グラフ      |
|  | 60 KB  | 線形モデル_09_3x2構造モデル    |
|  | 14 KB  | 線形モデル_09_カスタム検定      |
|  | 158 KB | 線形モデル_10_3x2平方和の分解   |

非売品，無断複製を禁ずる

第 10 回 続高橋セミナー

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 —

BioStat 研究所(株)

〒105-0014 東京都港区 芝 1-12-3 の 1005

2022 年 4 月 高橋 行雄

[takahashi.stat@nifty.com](mailto:takahashi.stat@nifty.com)，FAX : 03-342-8035