

最適計画は、要因効果の推定精度や予測精度を最もよくする実験計画のことを指します。ここでは実験計画法でよく使う直交表実験を例に、最適計画を説明します。

実験計画法と言えば、直交表実験を思い浮かべるのではないのでしょうか？直交表で実験を行うと、なるべく少ない実験回数で、精度の高い結果が得られます。設計や生産において、典型的な SQC（統計的品質管理）改善手法です。つまり、直交表実験は“良い”実験として、世に出回っているのです。

では、なぜ直交表実験は良い実験なのでしょう？ここで繰返しのない 3 元配置を考えます。因子をそれぞれ A, B, C とし、全て 2 水準とします。総実験回数は $2^3 = 8$ 回です。ここで、主効果だけを調べたいとしましょう。そうすると、直交表実験のように実験回数を減らすことが可能です。直交表実験は、主効果と一部の交互作用以外は無視できるのであれば、総当たり（全水準組合せ）を行わなくても、解析対象である主効果と一部の交互作用を解析できるということを教えています。実験回数を減らすだけなら、8 通りある全水準組合せのうち、適当に 4 つだけを行えばよいわけです。例えば、次のように 4 つの組合せを取って実験を行い、データを取りましょう。

3 元配置										
実験 No.	A	B	C							
1	1	1	1							
2	1	1	2							
3	1	2	1							
4	1	2	2							
5	2	1	1							
6	2	1	2							
7	2	2	1							
8	2	2	2							

→

計画 1						
元の実験 No.	実験 No.	A	B	C	y	
1	1	1	1	1	y ₁	
2	2	1	1	2	y ₂	
5	3	2	1	1	y ₃	
8	4	2	2	2	y ₄	

データの構造式は、

$$\begin{aligned}
 y_1 &= \mu + \alpha_1 + \beta_1 + \gamma_1 + \varepsilon_1 \\
 y_2 &= \mu + \alpha_1 + \beta_1 + \gamma_2 + \varepsilon_2 \\
 y_3 &= \mu + \alpha_2 + \beta_1 + \gamma_1 + \varepsilon_3 \\
 y_4 &= \mu + \alpha_2 + \beta_2 + \gamma_2 + \varepsilon_4
 \end{aligned}$$

です。この実験は直交表実験とは異なり、因子 B において第 1 水準と第 2 水準の回数が異なります。したがって、2 水準直交表の解析のように、効果を“(第 1 水準の和 - 第 2 水準の和) / 4”で求めることができません。しかしながら、回帰分析で求めることができます。そこで、2 水準直交表実験の通常の制約条件を考えると、

$$\alpha_1 + \alpha_2 = 0, \beta_1 + \beta_2 = 0, \gamma_1 + \gamma_2 = 0$$

より、

$$\alpha_1 = \alpha, \alpha_2 = -\alpha, \beta_1 = \beta, \beta_2 = -\beta, \gamma_1 = \gamma, \gamma_2 = -\gamma$$

ですので、 α, β, γ を用いてデータの構造式を書き換えると、

$$\begin{aligned}
 y_1 &= \mu + \alpha + \beta + \gamma + \varepsilon_1 \\
 y_2 &= \mu + \alpha + \beta - \gamma + \varepsilon_2 \\
 y_3 &= \mu - \alpha + \beta + \gamma + \varepsilon_3 \\
 y_4 &= \mu - \alpha - \beta - \gamma + \varepsilon_4
 \end{aligned}$$

となります。これを行列の形式にすると、

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha \\ \beta \\ \gamma \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

になります。行列の第1列目は平均値を表すベクトルで、第2列目から第4列目は、偏回帰係数に対応する主効果の説明変数のベクトルになります。この行列を計画行列と言います。データのベクトル (y_1, y_2, y_3, y_4) を $\vec{y}$ 、全平均と主効果のベクトル $(\mu, \alpha, \beta, \gamma)$ を $\vec{b}$ 、計画行列を X で表すと、最小二乗法によって主効果（偏回帰係数）が

$$\hat{\vec{b}} = (X^T X)^{-1} X^T \vec{y}$$

のように求まります。ただし、 X^T は X の転置を表しています。実際に求めてみると、

$$\hat{\mu} = \frac{y_1 + y_4}{2}, \hat{\alpha} = \frac{y_1 - y_3}{2}, \hat{\beta} = \frac{-y_1 + y_2 + y_3 - y_4}{2}, \hat{\gamma} = \frac{y_1 - y_2}{2}$$

となります。データを全て使っていない推定量があるので、データを全て使う直交表より、精度があまり良くないことがわかります。主効果の推定量 $\hat{\vec{b}}$ の分散と共分散は、 $(X^T X)^{-1} \sigma^2$ と簡単に求まります。ただし、回帰分析における通常の誤差の仮定、不偏性、独立、等分散性 ($= \sigma^2$) を仮定しています。実際に計算してみると、

$$\begin{pmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \begin{pmatrix} 1/2 & 1/4 & -1/2 & 1/4 \\ 1/4 & 1/2 & -1/2 & -1/4 \\ -1/2 & -1/2 & 1 & -1/2 \\ 1/4 & 1/4 & -1/2 & 1/2 \end{pmatrix}$$

のようになります。対角要素は推定量の分散を表し、非対角要素は推定量間の共分散を示しています。例えば、2行2列目は、 $V(\hat{\alpha}) = \sigma^2/2$ であることを示しています。また、2行3列目は、 $Cov(\hat{\alpha}, \hat{\beta}) = -\sigma^2/2$ であることを示しています。

では、 $L_4(2^3)$ ではどうでしょうか？ $L_4(2^3)$ は、

表1 $L_4(2^3)$

元の実験 No.	実験 No.	A	B	C	y
1	1	1	1	1	y_1
4	2	1	2	2	y_2
6	3	2	1	2	y_3
7	4	2	2	1	y_4

です。計画行列は、通常の制約式を入れて、

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

となります。最小二乗法で主効果の推定量を求めると、

$$\hat{\mu} = \frac{y_1 + y_2 + y_3 + y_4}{4}, \hat{\alpha} = \frac{(y_1 + y_2) - (y_3 + y_4)}{4}, \hat{\beta} = \frac{(y_1 + y_3) - (y_2 + y_4)}{4}, \hat{\gamma} = \frac{(y_1 + y_4) - (y_2 + y_3)}{4}$$

です。これは、2水準直交表における平方和の公式でも登場する“第1水準の和と第2水準の和の差”です。このとき、推定量の分散と共分散は、

$$\begin{pmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \begin{pmatrix} 1/4 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{pmatrix}$$

より、 $V(\hat{\mu}) = V(\hat{\alpha}) = V(\hat{\beta}) = V(\hat{\gamma}) = \sigma^2/4$ であり、無相関です。よって、適当に選んで作った計画1よりも、 $L_4(2^3)$ の方が、どの主効果の推定量を見ても分散が小さいですし、無相関なので、“良い（ベターな）計画”と言えます。

さて、ここで、3元配置における実験 No.7 の水準組合せでは、実験対象が変性しまい、実験が無効になる（データが取れない）と仮定しましょう。すなわち、 $L_4(2^3)$ ができない状況です。計画1でも良いのですが、もう少し、分散が小さく、相関係数も小さい計画は作れないでしょうか？計画1において、元の実験 No.5 の代わりに No.4 としてみます。これを計画2とします。計画2において、主効果の分散と共分散は

$$\begin{pmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} \begin{pmatrix} 1/2 & -1/4 & 0 & 1/4 \\ -1/4 & 1/2 & -1/4 & 0 \\ 0 & -1/4 & 1/2 & -1/4 \\ 1/4 & 0 & -1/4 & 1/2 \end{pmatrix}$$

となります。計画 2 の方が計画 1 よりも無相関なものもあり、 $V(\hat{\beta})$ も小さく、良さそうな計画です。

このように、同じ実験回数なら、推定量の分散がなるべく小さく、共分散が 0 に近い計画が望ましいと考えられます。ここでは 3 因子で主効果のみを考えていますが、もっと多くの因子を取り上げ、主効果および 2 因子交互作用を取り上げた場合、このように分散と共分散を一つずつ比較することで計画の優劣をうまくつけることができるのでしょうか？一般的には分散と共分散個々の比較では、部分的に勝ったり負けたりで、優劣をつけることができません。そこで、計画の良さの総合指標を作り、それで比較します。よく用いられる総合指標は、分散と共分散の行列（分散共分散行列） $(X^T X)^{-1} \sigma^2$ の行列式です。この行列式は推定量の分散と共分散の全てが詰め込まれていますから、それを最小にする計画を最も良い計画とするわけです。この最も良い計画を D 最適計画と言います。 D は行列式の英語、*Determinant* の頭文字です。

上述の計画、 $L_4(2^3)$ 、計画 1、計画 2 のうち、どれが D 最適計画でしょうか。それぞれの計画について、 $(X^T X)^{-1} (\sigma^2$ は計画に依存しなので省略) の行列式を求めてみると、

$$L_4(2^3) : \frac{1}{256}, \text{計画 1} : \frac{1}{64}, \text{計画 2} : \frac{1}{64}$$

です。したがって、この 3 つの中では $L_4(2^3)$ が D 最適計画であることがわかります。この規準では計画 1 と計画 2 は同等の良さです。3 元配置の 8 通りから 4 通りを取る選び方はこの他にもありますから、厳密には全通りを尽くして、 D 最適計画が求まります。しかし、ホテリングの定理から、 $L_4(2^3)$ は厳密な意味での D 最適計画であることがわかっています。また、他のサイズの直交表においても同様なことが言え、したがって、直交表による実験計画は、このような意味で“良い計画”であるのです。

行列式の他にも、 $(X^T X)^{-1}$ の対角成分の和、すなわち、推定量の分散の和を総合指標とするものもあります。この和を最小にする計画を A 最適計画と言います。 $L_4(2^3)$ 、計画 1、計画 2 で比較してみますと、

$$L_4(2^3) : 1, \text{計画 1} : 2\frac{1}{2}, \text{計画 2} : 2$$

です。 $L_4(2^3)$ が 3 つの中の A 最適計画で、計画 2 の方が計画 1 よりも良いということになります。

ここまで直交表実験をベースに最適計画を説明してきましたが、思い切って、一般化します。ただし、因子は全て量的因子とします。 p 個の因子があり、水準を振れる範囲を全て $[-1, 1]$ とします。また、実験データ（応答）と因子との関数関係を $f(x_1, \dots, x_p)$ とし、実験の総回数を n に制限して、関数関係を精度よく推定したいとします。例えば、 $p = 2$ 、 $f(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$ （データの構造式は $y = f(x_1, x_2) + \varepsilon$ ）、 $n = 10$ としましょう。“精度が良い計画” = D 最適計画と定義した時、 $f(x_1, x_2)$ から導かれる計画行列 X について、 $|(X^T X)^{-1}|$ を最小とする、もしくは、 $|(X^T X)^{-1}| = |X^T X|^{-1}$ より $|X^T X|$ を最大にする $x_1, x_2 \in [-1, 1]$ の $n = 10$ 個の組合せが本設定における D 最適計画です。

簡単な例として、単回帰における D 最適計画を求めてみましょう。データの構造式は $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ 、 $i = 1, \dots, n$ (n 偶数) で、 x_i の取れる範囲は $[-1, 1]$ としましょう。ここで、

$$|X^T X| = \left| \sum_{i=1}^n x_i \quad \sum_{i=1}^n x_i^2 \right| = n \left(\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \right) = n \sum_{i=1}^n (x_i - \bar{x})^2$$

です。 D 最適計画は $|X^T X|$ を最大にする、すなわち、説明変数の分散をなるべく大きくする x_i 、 $i = 1, \dots, n$ の組合せですから、 n (偶数) 個の x を -1 と 1 に半分ずつ振り分けばよいのです。

最適計画は、 n 数、水準を振れる領域、データの構造式（モデル）の様々なケースに対して柔軟に生成できますが、ある種の“良さ”とそのケースの下に限定された最適な実験計画であることを忘れずに使用すべきです。