実験計画法入門

第３章　１因子実験（乱塊法）

乱塊法 … 環境条件に違いが認められる場合、環境条件の同じような場をブロックとしていくつか作り、この各ブロック内で比較したい水準ひとそろいの実験をブロック内での実験順序はランダムに行う実験法。

例

順番	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
水準	A_2	A_3	A_4	A_4	A_4	A_3	A_3	A_2	A_1	A_1	A_4	A_3	A_3	A_2	A_1	A_1	A_1	A_4	A_2	A_2
	第１日				第２日				第３日				第４日				第５日			

上の表は例題 2.1 の完全無作為化法による実験の表である。この表を見ると、第１日目には A_4 水準は２回実験されているが A_1 水準は実験されていない。第２日目では、A_3 が２回実験され、A_1 は実験されていない。３日目以降も複数回実験されている水準と一回も実験されていない水準がある。

ここで、“ 日 ”が実験結果に影響を与えるものとする。

例として日によって実験装置の状態が変化するとする。第１日目と第２日目の装置の状態は良かったが、第３日目、第４日目、第５日目は状態が悪かったとする。すると、A_3 と A_4 水準は状態の良いときに３回も実験されているが、A_1 水準は状態が良いときには１回も実験されていないことになり不公平である。この問題を解決するために今の場合、水準数は４つで、１日に実験できる回数は４回なので、１日のうちに４つの水準すべての実験をすればよい。

いま、“ 日 ” をブロックに選び乱塊法で実験をすれば表 3.1 のような実験をすることになる。

日	1				2				3				4				5			
順序	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
水準	A_3	A_2	A_1	A_4	A_2	A_3	A_4	A_1	A_1	A_2	A_3	A_4	A_1	A_3	A_2	A_4	A_3	A_1	A_4	A_2

表 3.1　乱塊法による実験

表 3.1 では、日を、5 水準を持つひとつの因子と見ることができる。このことからブロックのことをブロック因子と呼ぶ。

例題 3.1

ある化学工場において、製品の収率（%）を高めるための工場実験をする。因子として反応温度 (A) だけをとりあげ、これの最適条件を見つけたい。水準として 150 ℃ (A_1)、175 ℃ (A_2)、200 ℃ (A_3)、225 ℃ (A_4) の 4 つを選んだ。過去の経験によると、日によって収率に変動が認められるので、日をブロック因子 (B) としてとりあげ、乱塊法で実験をすることにした。この工場では 1 日に 6 バッチ生産することができるので、この乱塊法の実験は可能である。

各水準での繰り返し数、すなわちブロック数を 5 とし、各日での実験順序はランダムに決めた結果、表 3.2 のような順序で実験をすることになった。

日＼順序	1	2	3	4
第 1 日	A_2	A_1	A_3	A_4
第 2 日	A_1	A_3	A_2	A_4
第 3 日	A_1	A_2	A_3	A_4
第 4 日	A_3	A_4	A_2	A_1
第 5 日	A_3	A_2	A_4	A_1

表 3.2　実験順序

得られた実験データ（収率）を、因子 A の水準とブロック別に整理すると表 3.3 のようであった。

A \ B	B_1	B_2	B_3	B_4	B_5
A_1	77.7	77.1	77.4	78.1	77.7
A_2	78.3	78.2	78.2	78.4	79.3
A_3	79.3	78.2	80.1	79.7	78.7
A_4	77.0	78.0	78.1	78.4	77.1

表 3.3 収率（%）

実験データを一般に x_{ij} で表すと、表 3.4 のようにまとめられる。

$A \backslash B$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_b	和	平均
A_1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1b}	$x_{1\cdot}$	$\bar{x}_{1\cdot}$
A_2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2b}	$x_{2\cdot}$	$\bar{x}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
A_i	x_{i1}	x_{i1}	$\cdots$	x_{ij}	$\cdots$	x_{ib}	$x_{i\cdot}$	$\bar{x}_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
A_a	x_{a1}	x_{a2}	$\cdots$	x_{aj}	$\cdots$	x_{ab}	$x_{a\cdot}$	$\bar{x}_{a\cdot}$
和	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot j}$	$\cdots$	$x_{\cdot b}$	$x_{\cdot\cdot}$	
平均	$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$	$\cdots$	$\bar{x}_{\cdot j}$	$\cdots$	$\bar{x}_{\cdot b}$		$\bar{x}_{\cdot\cdot}$

表 3.4 乱塊法による実験データ

補足

$$x_{i\cdot} = \sum_{j}^{b} x_{ij} = A_i \text{ 水準でのデータの和}$$

$$x_{\cdot j} = \sum_{i}^{a} x_{ij} = \text{ブロック } B_j \text{ でのデータの和}$$

$$x_{\cdot\cdot} = \sum_{i}^{a} \sum_{j}^{b} x_{ij} = \text{全データの和}$$

$$\bar{x}_{i\cdot} = \frac{1}{b} x_{i\cdot} = A_i \text{ 水準でのデータの平均値}$$

$$\bar{x}_{\cdot j} = \frac{1}{a} x_{i\cdot} = \text{ブロック } B_j \text{ でのデータの平均値}$$

$$\bar{x}_{\cdot\cdot} = \frac{1}{ab} x_{\cdot\cdot} = \text{全データの平均値}$$

データの構造模型

x_{ij} の構造模型を次のように仮定する。

$$x_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \qquad (3.1)$$

μ : 一般平均

α_i : A の主効果 $\quad \sum_{i}^{a} \alpha_i = 0$

β_j : ブロック因子 B の主効果 $\quad \sum_{j}^{b} \beta_j = 0$

e_{ij} : 誤差、正規分布 $N(0,\ \sigma^2)$ に従う

完全無作為化法のときと同様に、まず因子 A の水準 A_1, A_2, $\cdots$, A_a の間に差があるかどうかを調べる。すなわち、仮説

$$H_0 \ : \ \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

の検定をする。

ブロック間に差があるかどうか調べるときも同様に検定すればよい。

次に最適水準の決定には因子 A の主効果 α_i を推定すればよい。

因子 A の水準間に差がないという仮説を、簡単に"因子 A（について）の検定"と呼ぶ。同様にブロック間に差がないという仮説の検定は"ブロック（について）の検定"と呼ぶ。

分散分析

因子 A についての検定とブロックについての検定をするために分散分析をする。総平方和 S_T は

$$S_T = \sum_{i}^{a} \sum_{j}^{b} (x_{ij} - \bar{x}_{..})^2 \tag{3.2}$$

で表される。これは全データのばらつきを表すので、これをばらつきを与える原因ごとに分解すると次のようになる。

$$S_T = \underbrace{b \sum_{i}^{a} (\bar{x}_{i.} - \bar{x}_{..})^2}_{A\text{ 間平方和 }(S_A)} + \underbrace{a \sum_{j}^{b} (\bar{x}_{.j} - \bar{x}_{..})^2}_{\text{ブロック間平方和 }(S_B)} + \underbrace{\sum_{i}^{a} \sum_{j}^{b} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2}_{\text{誤差平方和 }(S_e)} \tag{3.3}$$

(3.3) 式の証明

まず $x_{ij} - \bar{x}_{..}$ を次のように分解する。

$$x_{ij} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

次に両辺を 2 乗し、両辺の i、j についての和をとる。すると、左辺は総平方和 S_T となり、右辺は次のようになる。

$$\sum_i^a \sum_j^b (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_i^a \sum_j^b (\bar{x}_{.j} - \bar{x}_{..})^2 + \sum_i^a \sum_j^b (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})^2$$

$$+2 \sum_i^a \sum_j^b (\bar{x}_{i.} - \bar{x}_{..})(\bar{x}_{.j} - \bar{x}_{..})$$

$$+2 \sum_i^a \sum_j^b (\bar{x}_{i.} - \bar{x}_{..})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

$$+2 \sum_i^a \sum_j^b (\bar{x}_{.j} - \bar{x}_{..})(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

第４項を計算すると

$$2\sum_{i}^{a}\sum_{j}^{b}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})(\bar{x}_{\cdot j}-\bar{x}_{\cdot\cdot}) = 2\underbrace{\sum_{i}^{a}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})}_{0}\underbrace{\sum_{j}^{b}(\bar{x}_{\cdot j}-\bar{x}_{\cdot\cdot})}_{0} = 0$$

第５項を計算すると

$$2\sum_{i}^{a}\sum_{j}^{b}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})(x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x}_{\cdot\cdot})$$

$$= 2\sum_{i}^{a}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})\sum_{j}^{b}(x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x}_{\cdot\cdot})$$

$$= 2\sum_{i}^{a}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})\underbrace{[x_{\cdot j}-a\bar{x}_{\cdot\cdot}-a\bar{x}_{\cdot j}+a\bar{x}_{\cdot\cdot}]}_{0} = 0$$

第６項は第５項と同様に計算できる。

結局、右辺で残ったのは

$$\sum_{i}^{a}\sum_{j}^{b}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2+\sum_{i}^{a}\sum_{j}^{b}(\bar{x}_{\cdot j}-\bar{x}_{\cdot\cdot})^2+\sum_{i}^{a}\sum_{j}^{b}(x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x}_{\cdot\cdot})^2$$

となり、(3.3) 式が成り立つ。

平方和の分解式 (3.3) から次の分散分析表を得る。

変動因	平方和	自由度	平均平方	F_0	平均平方の期待値
A 間	S_A	$\phi_A = a - 1$	$V_A = S_A/\phi_A$	V_A/V_e	$\sigma^2 + b\sigma_A^2$
ブロック間	S_B	$\phi_B = b - 1$	$V_B = S_B/\phi_B$	V_B/V_e	$\sigma^2 + a\sigma_B^2$
誤差	S_e	$\phi_e = (a-1)(b-1)$	$V_e = S_e/\phi_e$		σ^2
総	S_T	$\phi_T = ab - 1$			

表 3.5　乱塊法の分散分析表

最適水準の決定

検定をした結果、因子 A の水準間に有意な差があると判定されたとする。このとき、水準間の差や最適水準はどれか？といった問題が起こる。いま、完全無作為化法のときと同様に因子 A の主効果を α_i のかわりに $\mu+\alpha_i$ で考える。A の主効果を $\mu+\alpha_i$ とおいているので、$\mu+\alpha_i$ を推定することとなる。

$$\bar{x}_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b} x_{ij} = \frac{1}{b}\sum_{j=1}^{b}(\mu+\alpha_i+e_{ij}) = \mu+\alpha_i+\bar{e}_{i\cdot}$$

から

$$E\{\bar{x}_{i\cdot}\} = E\{\mu+\alpha_i+\bar{e}_{i\cdot}\} = \mu+\alpha_i$$

$$\mathrm{Var}\{\bar{x}_{i\cdot}\} = \mathrm{Var}\{\bar{e}_{i\cdot}\} = \frac{1}{b}\sigma^2$$

を得る。

次に $\mu+\alpha_i$ の区間推定を考える。$\mu+\alpha_i$ の 95 %信頼区間は、分散分析表から V_e は σ^2 の普遍推定量であることと、V_e の自由度が $(a-1)(b-1)$ であることがわかるので

$$\bar{x}_{i\cdot} \pm t((a-1)(b-1),\ 0.05)\sqrt{\frac{V_e}{b}}$$

で与えられる。

まとめると

点推定：A_i 水準でのデータの平均値 $= \bar{x}_i$

信頼区間（９５％）：$\bar{x}_{i\cdot} \pm t((a-1)(b-1),\ 0.05)\sqrt{\frac{V_e}{b}}$