第6章　分散分析

6.1 一元配置法

例）洗剤ごとの洗浄力を確かめる実験をした。数字は、汚れの落ち具合（大きいほどよく落ちているとする）を表す。

表 1

		繰り返し				
洗剤	A_1	58	54	56		
	A_2	64	60	61	58	
	A_3	66	62	68	63	61
	A_4	60	64	62		

1）平方和を求める

a：水準の数
r_i：A_i水準の繰り返し回数
x_{ij}：A_i水準のj番目のデータ
T_i：A_i水準のデータの合計
T：データの総合計
$\bar{x}_i$：A_i水準のデータの平均
$\bar{\bar{x}}$：全データの平均

表 2

		繰り返し					T_i	T_i^2	$\bar{x}_i$
洗剤	A_1	58	54	56			168	28224	56
	A_2	64	60	61	58		243	59049	60.75
	A_3	66	62	68	63	61	320	102400	64
	A_4	60	64	62			186	34596	62

総合計：T=917　総データ数：n=15　全体平均（一般平均）：$\bar{\bar{x}} = 61.13$

修正項：$\mathrm{CT} = \dfrac{T^2}{n} = \dfrac{917^2}{15} = 56059.27$

表 3

							合計
二乗値	A1	3364	2916	3136			9416
	A2	4096	3600	3721	3364		14781
	A3	4356	3844	4624	3969	3721	20514
	A4	3600	4096	3844			11540

総平方和：$S_T = \sum_{i=1}^{a}\sum_{j=1}^{r_i}\left(x_{ij} - \bar{\bar{x}}\right)^2 = \sum\sum {x_{ij}}^2 - CT$

$= (9416 + 14781 + 20514 + 11540) - 56059.27 = 191.73$

A 間平方和：$S_A = \sum_{i=1}^{a}\sum_{j=1}^{r_i}\left(x_{i\cdot} - \bar{\bar{x}}\right)^2 = \sum_{i=1}^{a}\frac{{T_{i\cdot}}^2}{r_i} - CT = \sum_{i=1}^{a}\frac{(A_i\text{水準でのデータの和})^2}{A_i\text{水準でのデータ数}} - CT$

$= \frac{28224^2}{3} + \frac{14781^2}{4} + \frac{20514^2}{5} + \frac{11540^2}{3} - 56059.27 = 122.98$

誤差平方和：$S_E = \sum_{i=1}^{a}\sum_{j=1}^{r_i}\left(x_{\mathrm{ij}} - \bar{x}_{i\cdot}\right)^2 = S_T - S_A = 191.73 - 122.98 = 68.75$

2）自由度を求める

$\phi_T = n - 1 = 15 - 1 = 14\,, \phi_A = a - 1 = 4 - 1 = 3\,, \phi_E = \phi_T - \phi_A = 14 - 3 = 11$

3）分散分析表にまとめる

F_0=V/E

表 4

要因	平方和 S	自由度 ϕ	平均平方 V	F_0
A	122.98	3	40.99	6.56
E	68.75	11	6.25	
計	191.73	14		

α=0.05：F(3 , 11 ; 0.05)=3.59

α=0.01：F(3 , 11 ; 0.01)=6.22

4）検定

有意水準 1%で有意となるので、「洗剤の違いによる汚れの落ち具合に差がある」と言える。

5）推定

最適水準：表 2 の $\bar{x}_i$ の値を見比べて定める。

ここでは、A_3水準の平均が一番大きいので、A_3を最適水準とする。

点推定：$\hat{\mu}_i = \widehat{\mu + \alpha_i} = \bar{x}_{i\cdot}$

$\hat{\mu}_3 = \widehat{\mu + \alpha_3} = \bar{x}_{3\cdot} = 64.0$

区間推定：信頼率 $100(1 - \alpha)\%$ の $\mu_i = \mu + \alpha_i$ の信頼区間

$$\left(\bar{x}_{i\cdot} - t(\phi_E\,, \alpha)\sqrt{\frac{V_E}{r_i}}\,, \bar{x}_{i\cdot} + t(\phi_E\,, \alpha)\sqrt{\frac{V_E}{r_i}}\right)$$

$$\left(64 - t(11\,,0.05)\sqrt{\frac{6.25}{5}}\,,64 + t(11\,,0.05)\sqrt{\frac{6.25}{5}}\right) = \left(64 - 2.201\sqrt{\frac{6.25}{5}}\,,64 + 2.201\sqrt{\frac{6.25}{5}}\right)$$

$=(61.5\,,66.5)$

6）A_i 水準でもう一度データをとるときの予測（おまけ）

点予測：$\hat{x}_{i0} = \widehat{\mu + \alpha_i} = \bar{x}_{i\cdot}$

$$\hat{\mu}_3 = \widehat{\mu + \alpha_3} = \bar{x}_{3\cdot} = 64.0$$

予測区間：信頼率 $100(1-\alpha)\%$ の x_{i0} の予測区間

$$\left(\bar{x}_{i\cdot} - t(\phi_E\,,\alpha)\sqrt{\left(1+\frac{1}{r_i}\right)V_E}\,,\bar{x}_{i\cdot} + t(\phi_E\,,\alpha)\sqrt{\left(1+\frac{1}{r_i}\right)V_E}\right)$$

$$\left(64 - 2.201\sqrt{\left(1+\frac{1}{5}\right)6.25}\,,64 + 2.201\sqrt{\left(1+\frac{1}{5}\right)6.25}\right) = (58.0\,,70.0)$$

6.2 繰り返しのある二元配置法

例）ある接着剤メーカーの S 製品は、原料 A、B によって接着力が決まる。今それぞれの候補を 3 つずつ選び、それぞれの組み合わせで 2 回ずつ粘着力テストを行ったところ、以下のような結果が出た。データの解析をせよ。

表 5

	B_1	B_2	B_3
A_1	13.6	14.5	13.9
	13.3	14.8	14.3
A_2	12.6	13.5	13.6
	12.4	13.7	13.5
A_3	13.3	14.0	14.2
	13.2	14.3	13.9

a：A 水準の数
b：B 水準の数
r：繰り返し回数
x_{ijk}：A_i水準, B_j水準のk番目のデータ
$T_{i\cdot\cdot}$：A_i水準のデータの合計
$\bar{x}_{i\cdot\cdot}$：A_i水準のデータの平均
$T_{\cdot j\cdot}$：B_j水準のデータの合計
$\bar{x}_{\cdot j\cdot}$：B_j水準のデータの平均

T：データの総合計
$\bar{\bar{x}}$：全データの平均

1）データの集計表（表 6)、AB 二元表（表 7)、$\bar{x}_{ij}$ 表（表 8）を作成する。

表 6 データの集計表

	B_1	B_2	B_3	計$T_{i\cdot\cdot}$	平均$\bar{x}_{i\cdot\cdot}$
A_1	13.6 13.3	14.5 14.8	13.9 14.3	84.4	14.07
A_2	12.6 12.4	13.5 13.7	13.6 13.5	79.3	13.22
A_3	13.3 13.2	14 14.3	14.2 13.9	82.9	13.82
計$T_{\cdot j\cdot}$	78.4	84.8	83.4	総合計：T=246.6	
平均$\bar{x}_{\cdot j\cdot}$	13.07	14.13	13.90	全平均：x=13.70	

表 7 AB 二元表（$T_{ij\cdot}$表）

	B_1	B_2	B_3	計$T_{i\cdot\cdot}$	平均$\bar{x}_{i\cdot\cdot}$
A_1	26.9	29.3	28.2	84.4	14.07
A_2	25.0	27.2	27.1	79.3	13.22
A_3	26.5	28.3	28.1	82.9	13.82
計$T_{\cdot j\cdot}$	78.4	84.8	83.4	総合計：T=246.6	
平均$\bar{x}_{\cdot j\cdot}$	13.07	14.13	13.90	全平均：x=13.70	

表 8 $\bar{x}_{ij}$ **表**

	B_1	B_2	B_3
A_1	13.45	14.65	14.1
A_2	12.5	13.6	13.55
A_3	13.25	14.15	14.05

2）平方和を求める

修正項：$\mathrm{CT} = \dfrac{T^2}{n} = \dfrac{246.6^2}{18} = 3378.42$

総平方和：$S_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}\left(x_{ijk} - \bar{\bar{x}}\right)^2 = \sum\sum\sum x_{ijk}{}^2 - CT$

$$= \left(13.6^2 + 13.3^2 + \cdots + 14.2^2 + 13.9^2\right) - 3378.42 = 3384.98 - 3378.42 = 6.56$$

AB 間平方和：$S_{AB} = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(T_{ij}.)^2}{r} - CT = \sum\sum\frac{(A_iB_j\text{水準でのデータの和})^2}{A_iB_j\text{水準でのデータ数}} - CT$

$$= \left(\frac{26.9^2 + 25.0^2 + \cdots + 27.1^2 + 28.1^2}{2}\right) - 3378.42 = 3384.67 - 3378.42 = 6.25$$

A 間平方和：$S_A = \sum_{i=1}^{a}\frac{(T_{i}..)^2}{b \times r} - CT = \frac{84.4^2 + 79.3^2 + 82.9^2}{6} - 3384.67 = 3380.71 - 3378.42 = 2.29$

B 間平方和：$S_B = \sum_{j=1}^{b}\frac{(T_{\cdot j}.)^2}{a \times r} - CT = \frac{78.4^2 + 84.8^2 + 83.4^2}{6} - 3384.67 = 3382.20 - 3378.42 = 3.77$

AB 間平方和：$S_{A\times B} = S_{AB} - S_A - S_B = 6.25 - 2.29 - 3.77 = 0.19$
誤差平方和：$S_E = S_T - S_{AB} = 6.56 - 6.25 = 0.31$

3）自由度を求める

$\phi_T = n - 1 = 18 - 1 = 17\,, \phi_A = a - 1 = 3 - 1 = 2\,, \phi_B = b - 1 = 3 - 1 = 2$
$\phi_{A\times B} = \phi_A \times \phi_B = 2 \times 2 = 4\,, \phi_E = \phi_T - \phi_A - \phi_B - \phi_{A\times B} = 17 - 2 - 2 - 4 = 9$

4）分散分析表にまとめる

F_0=V/E

表 9　分散分析表(1)

要因	平方和 S	自由度 φ	平均平方 V	F_0
A	2.29	2	1.145	33.24
B	3.77	2	1.885	54.73
A×B	0.19	4	0.048	1.38
E	0.31	9	0.034	
計	6.56	17		

F(2 , 9 ; 0.05)=4.26，F(2 , 9 ; 0.01)=8.02
F(4 , 9 ; 0.05)=3.63，F(4 , 9 ; 0.01)=6.42

5）A×B の扱いを決める。

$V_{A\times B}/V_E \geqq F(\phi_{A\times B}, \phi_E; 0.20)$なら A×B を無視しないで推定手順へ移る。
$V_{A\times B}/V_E < F(\phi_{A\times B}, \phi_E; 0.20)$なら A×B を無視して、A×B を誤差項 E へプールした（加えた）分散分析表を作り直し，その後推定手順へ移る。

$F(4, 9; 0.20)=1.87 > V_{A\times B}/V_E = 1.38$ なので、分散分析表を作り直す。

5）A×B をプールした分散分析表を作る

$S_E' = S_{A\times B} + S_E$，$\phi_E' = \phi_{A\times B} + \phi_E$，

表 10　分散分析表(2)

要因	平方和 S	自由度 φ	平均平方 V	F_0
A	2.29	2	1.145	29.77
B	3.77	2	1.885	49.01
E	0.50	13	0.038	
計	6.56	17		

F(2 , 13 ; 0.05)=3.81，F(2 , 13 ; 0.01)=6.70

以上より、主効果 A、主効果 B とも 1%有意となったが、交互作用効果 A×B は 5%有意とはならなかった。

6）推定（A×B を無視する場合）

最適水準：表 6 より、A については$\bar{x}_{i..}$，B については $\bar{x}_{.j.}$の値を見比べて定める。

ここでは、A_1，B_2を最適水準とする。

点推定：$\hat{\mu}_{ij} = \widehat{\mu + \alpha_i + \beta_j} = \widehat{\mu + \alpha_i} + \widehat{\mu + \beta_j} - \hat{\mu} = \bar{x}_{i..} + \bar{x}_{i..} - \bar{\bar{x}}$

$$\hat{\mu}_{12} = \bar{x}_{1..} + \bar{x}_{.2.} - \bar{\bar{x}} = 14.1 + 14.1 - 13.7 = 14.5$$

区間推定：信頼率 $100(1-\alpha)\%$ の $\mu_i\mathrm{j} = \mu + \alpha_i + \beta_j$ の信頼区間

$$\left(\left(\bar{x}_{i..} + \bar{x}_{i..} - \bar{\bar{x}}\right) - t(\phi_E{}', \alpha)\sqrt{\frac{V_E{}'}{n_e}}, \left(\bar{x}_{i..} + \bar{x}_{i..} - \bar{\bar{x}}\right) + t(\phi_E{}', \alpha)\sqrt{\frac{V_E{}'}{n_e}}\right)$$

ここでn_eは有効反復数と呼ばれ、$n_e = \frac{n}{a+b-1}$ で求められる。

$$n_e = \frac{18}{3+3-1} = \frac{18}{5} = 3.6$$

$$\left(14.5 - t(13, 0.05)\sqrt{\frac{0.038}{3.6}}, 14.5 + t(13, 0.05)\sqrt{\frac{0.038}{3.6}}\right) = \left(14.5 - 2.16\sqrt{\frac{0.038}{3.6}}, 14.5 - 2.16\sqrt{\frac{0.038}{3.6}}\right)$$

=(14.28 , 14.72)