

二元配置分散分析（繰り返しのある）

2つの要因 A、B が関係するデータの分散分析を考慮する。

要因Aの水準の個数を n_A 、要因Bの水準の個数を n_B 、

要因Aの i 番目の水準を A_i ($i = 1 \sim n_A$)、要因Bの j 番目の水準を B_j ($j = 1 \sim n_B$)

と表わす。

以降、 i は要因Aの水準の添字、 j は要因Bの水準の添字 とする。

また、要因Aの水準が A_i 、要因Bの水準が B_j での標本値をセル(i, j)の標本値と呼ぶ。

各セルの標本数はすべて n とし、セル(i, j)の k 番目 ($k = 1 \sim n$) の標本値を $X_{i,j,k}$ とする。

全データの平均を
$$\mu = \frac{\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n X_{i,j,k}}{n * n_A * n_B}$$

要因Aの水準 A_i の平均を
$$\mu_i^A = \frac{\sum_{j=1}^{n_B} \sum_{k=1}^n X_{i,j,k}}{n * n_B} \quad (i = 1 \sim n_A)$$

要因Bの水準 B_j の平均を
$$\mu_j^B = \frac{\sum_{i=1}^{n_A} \sum_{k=1}^n X_{i,j,k}}{n * n_A} \quad (j = 1 \sim n_B)$$

セル(i, j)の平均を
$$\mu_{i,j}^{AB} = \frac{\sum_{k=1}^n X_{i,j,k}}{n} \quad \text{とする。}$$

標本値は 以下のように表示できる。

$$\begin{aligned} X_{i,j,k} = & \mu + \\ & (\mu_i^A - \mu) + \\ & (\mu_j^B - \mu) + \\ & (\mu_{i,j}^{AB} - \mu - (\mu_i^A - \mu) - (\mu_j^B - \mu)) + \\ & X_{i,j,k} - \mu_{i,j}^{AB} \end{aligned} \quad \dots (1)$$

ここで 以下のように置き換える。

$$\begin{aligned} a_i & \equiv (\mu_i^A - \mu) \\ b_j & \equiv (\mu_j^B - \mu) \\ (ab)_{i,j} & \equiv \mu_{i,j}^{AB} - \mu - (\mu_i^A - \mu) - (\mu_j^B - \mu) \\ \varepsilon_{i,j,k} & \equiv X_{i,j,k} - \mu_{i,j}^{AB} \end{aligned}$$

式 (1) は次のように記述できる。

$$X_{i,j,k} = \mu + a_i + b_j + (ab)_{i,j} + \varepsilon_{i,j,k} \quad \dots (2)$$

a_i , b_j , $(ab)_{i,j}$ については以下が成り立つ。

$$\bullet \sum_{i=1}^{n_A} a_i = 0 \quad \dots (3)$$

$$\bullet \sum_{j=1}^{n_B} b_j = 0 \quad \dots (4)$$

$$\bullet \sum_{i=1}^{n_A} (ab)_{i,j} = 0 \quad \dots (5)$$

$$\bullet \sum_{j=1}^{n_B} (ab)_{i,j} = 0 \quad \dots (6)$$

(上記については 後段で証明)

$X_{i,j,k} - \mu = a_i + b_j + (ab)_{i,j} + \varepsilon_{i,j,k}$ であるから

$$\begin{aligned}(X_{i,j,k} - \mu)^2 &= (a_i + b_j + (ab)_{i,j} + \varepsilon_{i,j,k})^2 \\ &= a_i^2 + b_j^2 + (ab)_{i,j}^2 + \varepsilon_{i,j,k}^2 + \\ &\quad 2 * a_i * b_j + \\ &\quad 2 * a_i * (ab)_{i,j} + \\ &\quad 2 * a_i * \varepsilon_{i,j,k} + \\ &\quad 2 * b_j * (ab)_{i,j} + \\ &\quad 2 * b_j * \varepsilon_{i,j,k} + \\ &\quad 2 * (ab)_{i,j} * \varepsilon_{i,j,k}\end{aligned}$$

従って

$$\begin{aligned}\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (X_{i,j,k} - \mu)^2 &= \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (a_i + b_j + (ab)_{i,j} + \varepsilon_{i,j,k})^2 \\ &= \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (a_i^2 + b_j^2 + (ab)_{i,j}^2 + \varepsilon_{i,j,k}^2) + \dots \quad (7) \\ &\quad \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (2a_i * b_j) + \dots \quad (8) \\ &\quad \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (2 * a_i * (ab)_{i,j}) + \dots \quad (9) \\ &\quad \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (2 * a_i * \varepsilon_{i,j,k}) + \dots \quad (10) \\ &\quad \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (2 * b_j * (ab)_{i,j}) + \dots \quad (11) \\ &\quad \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (2 * b_j * \varepsilon_{i,j,k}) + \dots \quad (12) \\ &\quad \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (2 * (ab)_{i,j} * \varepsilon_{i,j,k})\end{aligned}$$

上記の (7) ~ (12) はいずれも 0 である。したがって以下となる。

$$\begin{aligned}\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (X_{i,j,k} - \mu)^2 &= \\ \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n a_i^2 + \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n b_j^2 + \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (ab)_{i,j}^2 + \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n \varepsilon_{i,j,k}^2\end{aligned}$$

$$\begin{aligned}S_T &\equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (X_{i,j,k} - \mu)^2, \quad S_A \equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n a_i^2, \quad S_B \equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n b_j^2, \\ S_{A \times B} &\equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n (ab)_{i,j}^2, \quad S_E \equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \sum_{k=1}^n \varepsilon_{i,j,k}^2 \quad \text{とすると}\end{aligned}$$

$$S_T = S_A + S_B + S_{A \times B} + S_E \quad \text{となる。}$$

上記のそれぞれの自由度を

$$df_T = n * n_A * n_B - 1, \quad df_A = n_A - 1, \quad df_B = n_B - 1, \quad df_{A \times B} = (n_A - 1)(n_B - 1)$$

$$df_E = df_T - df_A - df_B - df_{A \times B} \quad \text{として。}$$

$$SS_A \equiv S_A / df_A, \quad SS_B \equiv S_B / df_B, \quad SS_{A \times B} \equiv S_{A \times B} / df_{A \times B}, \quad SS_E \equiv S_E / df_E \quad \text{とすると、}$$

$\frac{SS_A}{SS_E}$ は自由度(df_A、df_E) の F 分布に従う。

$\frac{SS_B}{SS_E}$ は自由度(df_B、df_E) の F 分布に従う。

$\frac{SS_{A \times B}}{SS_E}$ は自由度(df_{AxB}、df_E) の F 分布に従う。

上記 3 つの F 値に基づく p 値が有意水準以下であれば、それぞれにつき有意性が認められる。

以降では (3) ~ (12) が 0 であることを示す。

(数式の表示の仕方が異なるので注意)

式 (3) $\sum_i a_i = 0$ を示す。

$$\begin{aligned}\sum_i (\mu^{A_i} (n^* n_B)) &= (\sum_i \sum_j \sum_k X_{ijk}) = \mu^* (n^* n_A^* n_B) \text{ ということは} \\ (n^* n_B)^* \sum_i \mu^{A_i} &= (n^* n_B)^* \mu^* n_A \text{ となるので、} \\ \sum_i \mu^{A_i} &= \mu^* n_A \text{ つまり } \sum_i (\mu^{A_i} - \mu) = 0. \text{ したがって } \sum_i a_i = 0.\end{aligned}$$

式 (4) $\sum_j b_j = 0$ を示す。

$$\begin{aligned}\sum_j (\mu^{B_j} (n^* n_A)) &= (\sum_i \sum_j \sum_k X_{ijk}) = \mu^* (n^* n_A^* n_B) \text{ ということは} \\ (n^* n_A)^* \sum_j \mu^{B_j} &= (n^* n_A)^* \mu^* n_B \text{ となるので、} \\ \sum_j \mu^{B_j} &= \mu^* n_B \text{ つまり } \sum_j (\mu^{B_j} - \mu) = 0. \text{ したがって } \sum_j b_j = 0.\end{aligned}$$

式 (5) $\sum_i (a b)_{ij} = 0$ を示す。

$$\begin{aligned}(a b)_{ij} &= (\mu^{AB_{ij}} - \mu - (\mu^{A_i} - \mu) - (\mu^{B_j} - \mu)) = (\mu^{AB_{ij}} - \mu) - a_i - b_j \\ \sum_i (a b)_{ij} &= \sum_i (\mu^{AB_{ij}} - \mu) - \sum_i a_i - \sum_i b_j \\ \sum_i (a b)_{ij} &= \sum_i (\mu^{AB_{ij}} - \mu) - \sum_i b_j \quad (\because \sum_i a_i = 0) \\ \sum_i (a b)_{ij} &= \sum_i (\mu^{AB_{ij}} - \mu) - \sum_i (\mu^{B_j} - \mu) \\ \sum_i (a b)_{ij} &= \sum_i (\mu^{AB_{ij}} - \mu^{B_j})\end{aligned}$$

$$\text{一方 } \sum_i (n^* \mu^{AB_{ij}}) = \sum_i \sum_k X_{ijk} = n^* n_A^* \mu^{B_j} \text{ なので}$$

$$n^* \sum_i \mu^{AB_{ij}} - n^* n_A^* \mu^{B_j} = 0$$

$$n^* \sum_i \mu^{AB_{ij}} - n^* \sum_i \mu^{B_j} = 0$$

$$\sum_i \mu^{AB_{ij}} - \sum_i \mu^{B_j} = 0$$

$$\sum_i (\mu^{AB_{ij}} - \mu^{B_j}) = 0$$

となるから、

$$\sum_i (a b)_{ij} = \sum_i (\mu^{AB_{ij}} - \mu^{B_j}) = 0 \text{ となる。}$$

式 (6) $\sum_j (a b)_{ij} = 0$ を示す。式 (5) の場合と同様に、

$$\sum_j (a b)_{ij} = \sum_j (\mu^{AB_{ij}} - \mu) - \sum_j a_i - \sum_j b_j$$

$$\sum_j (a b)_{ij} = \sum_j (\mu^{AB_{ij}} - \mu) - \sum_j a_i \quad (\because \sum_j b_j = 0)$$

$$\sum_j (a b)_{ij} = \sum_j (\mu^{AB_{ij}} - \mu) - \sum_j (\mu^{A_i} - \mu)$$

$$\sum_j (a b)_{ij} = \sum_j (\mu^{AB_{ij}} - \mu^{A_i})$$

$$\text{一方 } \sum_j (n^* \mu^{AB_{ij}}) = \sum_j \sum_k X_{ijk} = n^* n_B^* \mu^{A_i} \text{ なので}$$

$$n^* \sum_j \mu^{AB_{ij}} - n^* n_B^* \mu^{A_i} = 0$$

$$n^* \sum_j \mu^{AB_{ij}} - n^* \sum_j \mu^{A_i} = 0$$

$$\sum_j \mu^{AB_{ij}} - \sum_j \mu^{A_i} = 0$$

$$\sum_j (\mu^{AB_{ij}} - \mu^{A_i}) = 0$$

$$\text{となるから、} \sum_j (a b)_{ij} = \sum_j (\mu^{AB_{ij}} - \mu^{A_i}) = 0 \text{ となる。}$$

$$(7) \text{ 式 } \sum_i \sum_j \sum_k a_i * b_j = 0$$

$$\begin{aligned} \therefore \sum_i \sum_j \sum_k a_i * b_j &= \sum_i \sum_j (n * a_i * b_j) \\ &= n * \sum_i \sum_j (a_i * b_j) \\ &= n * \sum_i a_i (\sum_j b_j) = 0 \end{aligned}$$

$$(8) \text{ 式 } \sum_i \sum_j \sum_k a_i * (a b)_{ij} = 0$$

$$\begin{aligned} \therefore \sum_i \sum_j \sum_k a_i * (a b)_{ij} &= \sum_i \sum_j (n * a_i * (a b)_{ij}) \\ &= n * \sum_i \sum_j (a_i * (a b)_{ij}) \\ &= n * \sum_i a_i (\sum_j (a b)_{ij}) = 0 \end{aligned}$$

$$(9) \text{ 式 } \sum_i \sum_j \sum_k (b_j * (a b)_{ij}) = 0$$

$$\begin{aligned} \therefore \sum_i \sum_j \sum_k (b_j * (a b)_{ij}) &= \sum_j \sum_i (n * b_j * (a b)_{ij}) \\ &= n * \sum_j \sum_i (b_j * (a b)_{ij}) \\ &= n * \sum_j b_j (\sum_i (a b)_{ij}) = 0 \end{aligned}$$

$$(10) \text{ 式 } \sum_i \sum_j \sum_k (a_i * \varepsilon_{ijk}) = 0$$

$$\begin{aligned} \therefore \sum_i \sum_j \sum_k (a_i * \varepsilon_{ijk}) &= \sum_i \sum_j \sum_k (a_i * (X_{ijk} - \mu^{AB}_{ij})) \\ &= \sum_i a_i \sum_j \sum_k (X_{ijk} - \mu^{AB}_{ij}) \\ &= \sum_i a_i \sum_j (\sum_k X_{ijk} - \sum_k \mu^{AB}_{ij}) \\ &= \sum_i a_i \sum_j (n * \mu^{AB}_{ij} - n * \mu^{AB}_{ij}) \\ &= 0 \end{aligned}$$

$$(11) \text{ 式 } \sum_i \sum_j \sum_k (b_j * \varepsilon_{ijk}) = 0$$

$$\begin{aligned} \therefore \sum_i \sum_j \sum_k (b_j * \varepsilon_{ijk}) &= \sum_i \sum_j \sum_k (b_j * (X_{ijk} - \mu^{AB}_{ij})) \\ &= \sum_i \sum_j b_j \sum_k (X_{ijk} - \mu^{AB}_{ij}) \\ &= \sum_i \sum_j b_j (\sum_k X_{ijk} - \sum_k \mu^{AB}_{ij}) \\ &= \sum_i \sum_j b_j (n * \mu^{AB}_{ij} - n * \mu^{AB}_{ij}) \\ &= 0 \end{aligned}$$

$$(12) \text{ 式 } \sum_i \sum_j \sum_k ((a b)_{ij} * \varepsilon_{ijk}) = 0$$

$$\begin{aligned} \therefore \sum_i \sum_j \sum_k ((a b)_{ij} * \varepsilon_{ijk}) &= \sum_i \sum_j \sum_k ((a b)_{ij} * (X_{ijk} - \mu^{AB}_{ij})) \\ &= \sum_i \sum_j (a b)_{ij} \sum_k (X_{ijk} - \mu^{AB}_{ij}) \\ &= \sum_i \sum_j (a b)_{ij} (\sum_k X_{ijk} - \sum_k \mu^{AB}_{ij}) \\ &= \sum_i \sum_j (a b)_{ij} (n * \mu^{AB}_{ij} - n * \mu^{AB}_{ij}) \\ &= 0 \end{aligned}$$