

二元配置分散分析（繰り返しのない）

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1とし、セル(i, j)の標本値を X_{ij} とする。

$$\text{全データの平均を} \quad \mu = \frac{\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} X_{ij}}{n_A * n_B}$$

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

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

標本値は 以下のように表示できる。

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

ここで 以下のように置き換える。

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

式 (1) は次のように記述できる。

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

$a_i, b_j, \varepsilon_{i,j}$ については以下が成り立つ。

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

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

$$\cdot \sum_{i=1}^{n_A} \varepsilon_{i,j} = 0 \quad \dots (5)$$

$$\cdot \sum_{j=1}^{n_B} \varepsilon_{i,j} = 0 \quad \dots (6)$$

(上記については 後段で証明)

$X_{i,j} - \mu = a_i + b_j + \varepsilon_{i,j}$ であるから

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

従って

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

上記の (7) ~ (9) はいずれも 0 である。したがって以下となる。

$$\begin{aligned}\sum_{i=1}^{n_A} \sum_{j=1}^{n_B} (X_{i,j} - \mu)^2 &= \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} a_i^2 + \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} b_j^2 + \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \varepsilon_{i,j}^2 \\ S_T &\equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} (X_{i,j} - \mu)^2, \quad S_A \equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} a_i^2 \\ S_B &\equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} b_j^2, \quad S_E \equiv \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \varepsilon_{i,j}^2\end{aligned}$$

とすると

$$S_T = S_A + S_B + S_E \quad \text{となる。}$$

上記のそれぞれの自由度を

$$\begin{aligned}df_T &= n_A * n_B - 1, \quad df_A = n_A - 1, \quad df_B = n_B - 1, \quad df_E = df_T - df_A - df_B \\ \text{更に } SS_A &\equiv S_A / df_A, \quad SS_B \equiv S_B / df_B, \quad SS_E \equiv S_E / df_E \quad \text{とすると、}\end{aligned}$$

$\frac{SS_A}{SS_E}$ 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。

上記 2 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) ~ (9) 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。

$$\begin{aligned}\sum_i (\mu^{A_i} * n_B) &= \sum_i \sum_j X_{ij} = \mu * (n_A * n_B) \text{ であり} \\ n_B * \sum_i \mu^{A_i} &= n_B * \mu * n_A \text{ となるので、} \\ \sum_i \mu^{A_i} &= \mu * n_A \text{ つまり } \sum_i (\mu^{A_i} - \mu) = 0. \text{ したがって } \sum_i a_i = 0.\end{aligned}$$

式 (4) $\sum_j b_j = 0$ を示す。

$$\begin{aligned}\sum_j (\mu^{B_j} * (n_A)) &= \sum_i \sum_j X_{ij} = \mu * (n_A * n_B) \text{ であり} \\ n_A * \sum_j \mu^{B_j} &= n_A * \mu * n_B \text{ となるので、} \\ \sum_j \mu^{B_j} &= \mu * n_B \text{ つまり } \sum_j (\mu^{B_j} - \mu) = 0. \text{ したがって } \sum_j b_j = 0.\end{aligned}$$

式 (5) $\sum_i \varepsilon_{ij} = 0$ を示す。

$$\begin{aligned}\varepsilon_{ij} &= (X_{ij} - \mu - (\mu^{A_i} - \mu) - (\mu^{B_j} - \mu)) = (X_{ij} - \mu) - a_i - b_j \\ \sum_i \varepsilon_{ij} &= \sum_i (X_{ij} - \mu) - \sum_i a_i - \sum_i b_j \\ \sum_i \varepsilon_{ij} &= \sum_i (X_{ij} - \mu) - \sum_i b_j \quad (\because \sum_i a_i = 0) \\ \sum_i \varepsilon_{ij} &= \sum_i (X_{ij} - \mu) - \sum_i (\mu^{B_j} - \mu) \\ \sum_i \varepsilon_{ij} &= \sum_i (X_{ij} - \mu^{B_j}) \\ \text{一方 } \sum_i X_{ij} &= n_A * \mu^{B_j} \text{ なので} \\ \sum_i X_{ij} - n_A * \mu^{B_j} &= \sum_i X_{ij} - \sum_i \mu^{B_j} = \sum_i (X_{ij} - \mu^{B_j}) = 0 \\ \text{となるから、} \sum_i \varepsilon_{ij} &= \sum_i (X_{ij} - \mu^{B_j}) = 0 \text{ となる。}\end{aligned}$$

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

$$\begin{aligned}\sum_j \varepsilon_{ij} &= \sum_j (X_{ij} - \mu) - \sum_j a_i - \sum_j b_j \\ \sum_j \varepsilon_{ij} &= \sum_j (X_{ij} - \mu) - \sum_j a_i \quad (\because \sum_j b_j = 0) \\ \sum_j \varepsilon_{ij} &= \sum_j (X_{ij} - \mu) - \sum_j (\mu^{A_i} - \mu) \\ \sum_j \varepsilon_{ij} &= \sum_j (X_{ij} - \mu^{A_i}) \\ \text{一方 } \sum_j X_{ij} &= n_B * \mu^{A_i} \text{ なので} \\ \sum_j X_{ij} - n_B * \mu^{A_i} &= \sum_j X_{ij} - \sum_j \mu^{A_i} = \sum_j (X_{ij} - \mu^{A_i}) = 0 \\ \text{となるから、} \sum_j \varepsilon_{ij} &= \sum_j (X_{ij} - \mu^{A_i}) = 0 \text{ となる。}\end{aligned}$$

式 (7) $\sum_i \sum_j (a_i * b_j) = 0$ を示す。

$$\because \sum_i \sum_j (a_i * b_j) = \sum_i a_i (\sum_j b_j) = 0 \quad (\because \sum_j b_j = 0)$$

式 (8) $\sum_i \sum_j (a_i * \varepsilon_{ij}) = 0$ を示す。

$$\because \sum_i \sum_j (a_i * \varepsilon_{ij}) = \sum_i a_i (\sum_j \varepsilon_{ij}) = 0 \quad (\because \sum_j \varepsilon_{ij} = 0)$$

式 (9) $\sum_i \sum_j (b_j * \varepsilon_{ij}) = 0$ を示す。

$$\because \sum_i \sum_j (b_j * \varepsilon_{ij}) = \sum_j b_j (\sum_i \varepsilon_{ij}) = 0 \quad (\because \sum_i \varepsilon_{ij} = 0)$$