

一元配置分散分析

一元配置分散分析では、違いをもたらす要因は1つ。

そして、状況に変化をもたらす条件を水準と言う。

水準の個数を n とし、 i 番目の水準を $A_i (i = 1 \sim n)$ と表わす。

水準 A_i の標本個数を n_i 、平均を μ_i 、 k 番目の標本値を $X_{i,k}$ とする。

つまり $\mu_i = \frac{1}{n_i} \sum_{k=1}^{n_i} X_{i,k}$ である。

また標本総数を $N = \sum_{i=1}^n n_i$ 、全体平均値を $\mu = \frac{1}{N} (\sum_{i=1}^n \mu_i * n_i)$ とする。

標本値 $X_{i,k}$ は、一般に $X_{i,k} = \mu + (\mu_i - \mu) + (X_{i,k} - \mu_i)$ と表せるので、

$X_{i,k} - \mu = (\mu_i - \mu) + (X_{i,k} - \mu_i)$ となる。

$$\begin{aligned} \sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu)^2 &= \\ \sum_{i=1}^n \sum_{k=1}^{n_i} (\mu_i - \mu)^2 &+ 2 \sum_{i=1}^n \sum_{k=1}^{n_i} (\mu_i - \mu) (X_{i,k} - \mu_i) \quad \dots (A) \\ + \sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu_i)^2 & \end{aligned}$$

上式の中間の (A) は 0 である。

$$\begin{aligned} (A) &= 2 \sum_{i=1}^n \sum_{k=1}^{n_i} (\mu_i - \mu) (X_{i,k} - \mu_i) \\ &= 2 \sum_{i=1}^n (\mu_i - \mu) \sum_{k=1}^{n_i} (X_{i,k} - \mu_i) \\ &= 2 \sum_{i=1}^n (\mu_i - \mu) \left(\sum_{k=1}^{n_i} X_{i,k} - \sum_{k=1}^{n_i} \mu_i \right) \\ &= 2 \sum_{i=1}^n (\mu_i - \mu) (n_i \mu_i - n_i \mu_i) \\ &= 0 \end{aligned}$$

したがって、以下となる。

$$\begin{aligned} \sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu)^2 &= \sum_{i=1}^n \sum_{k=1}^{n_i} (\mu_i - \mu)^2 + \sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu_i)^2 \\ &= \sum_{i=1}^n n_i (\mu_i - \mu)^2 + \sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu_i)^2 \\ &\quad (B) \qquad (C) \end{aligned}$$

- ・ 左辺 $\sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu)^2$ は一定。
- ・ 右辺の左側 (B) $= \sum_{i=1}^n n_i (\mu_i - \mu)^2$ は 水準間のばらつきを表す。
- ・ 右辺の右側 (C) $= \sum_{i=1}^n \sum_{k=1}^{n_i} (X_{i,k} - \mu_i)^2$ は 水準内のばらつきを表す。

(B) を自由度 $n - 1$ で割った値を $R = (B) / (n - 1)$ 、

(C) を自由度 $N - n$ で割った値を $E = (C) / (N - n)$ とする。

$F = R / E$ は 自由度 $(n - 1, N - n)$ の F 分布に従うので、

この F 値に基づく p 値が有意水準以下であれば、有意性が認められる。