第 X 部　分散分析

目次

1　一元配置分散分析

1.1　はじめに

- 3 組以上の母平均の差を比較します。
- 観測されたばらつきが、母平均の差によるものか、単なる誤差かを判定します。
- やっていることはシンプルです。
- 説明は複雑です。

1.1.1　ポイント

- 水準
- 変動
- 変動の分解
- 変動が従う確率分布

1.1.2　和の確率変数の期待値と分散

X, Y を独立な確率変数とする。

$$\mathsf{E}\left(X+Y\right)=\mathsf{E}\left(X\right)+\mathsf{E}\left(Y\right) \tag{1}$$

$$\mathsf{V}\left(X+Y\right)=\mathsf{V}\left(X\right)+\mathsf{V}\left(Y\right) \tag{2}$$

が成り立つ。

1.1.3 総和

$n \times m$ 個のデータがあり、n に関してデータを識別する添え字に i、m に関して識別する添え字に j を採用し、ij 番目の値を x_{ij} とあらわす。i に関する総和は

$$\sum_{i=1}^{n} x_{ij} = x_{1j} + x_{2j} + \cdots + x_{nj} \tag{3}$$

であり、j に関する総和は

$$\sum_{j=1}^{m} x_{ij} = x_{i1} + x_{i2} + \cdots + x_{im} \tag{4}$$

そして、i および j に関する総和は

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} =& x_{11} + x_{12} + \cdots + x_{1m} \\ &+ x_{21} + x_{22} + \cdots + x_{2m} \\ &+ \cdots \\ &+ x_{n1} + x_{n2} + \cdots + x_{nm}\end{aligned} \tag{5}$$

である。

(3)(4)(5) の関係は

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \\
=&x_{11}+x_{12}+\cdots+x_{1m} \\
+&x_{21}+x_{22}+\cdots+x_{2m} \\
+&\cdots \\
+&x_{n1}+x_{n2}+\cdots+x_{nm} \qquad (5)
\end{aligned}
$$

$$
\begin{aligned}
&=(x_{11}+x_{12}+\cdots+x_{1m}) \\
&+(x_{21}+x_{22}+\cdots+x_{2m}) \\
&+\cdots \\
&+(x_{n1}+x_{n2}+\cdots+x_{nm}) \qquad (6) \\
&=\sum_{j=1}^{m} x_{1j}+\sum_{j=1}^{m} x_{2j}+\cdots+\sum_{j=1}^{m} x_{nj} \qquad (7) \\
&=\sum_{i=1}^{n} x_{i1}+\sum_{i=1}^{n} x_{i2}+\cdots+\sum_{i=1}^{n} x_{im} \qquad (8)
\end{aligned}
$$

表 1 ij に関する総和

x_{11}	x_{12}	$\cdots$	x_{1m}	$\sum_{j=1}^{m} x_{1j}$
x_{21}	x_{22}	$\cdots$	x_{2m}	$\sum_{j=1}^{m} x_{2j}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_{n1}	x_{n2}	$\cdots$	x_{nm}	$\sum_{j=1}^{m} x_{nj}$
$\sum_{i=1}^{n} x_{i1}$	$\sum_{i=1}^{n} x_{i2}$	$\cdots$	$\sum_{i=1}^{n} x_{im}$	$\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}$

1.1.4 標準正規分布

平均 0, 分散 1 の正規分布 $N(0,\ 1)$ を『**標準正規分布**』といい、標準正規分布に従う確率変数を $Z = z_i$ であらわす。標準正規分布の確率密度関数は

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{9}$$

である。

平均 μ, 分散 σ^2 の正規分布 $N\left(\mu, \sigma^2\right)$ に従う確率変数を $X = x_i$ とする。この x_i を標準化

$$z_i = \frac{x_i - \mu}{\sigma} \tag{10}$$

すると、標準正規分布 $N(0,\ 1)$ にしたがう確率変数 z_i を得ることができる。

1.1.5 χ^2 分布

標準正規分布 $N(0,1)$ に従う n 個の独立な確率変数 z_1, z_2, $\cdots$, z_n の平方和から統計量を作り $\chi^2_{(0)}$ とあらわす。

$$\chi^2_{(0)} = \sum_{i=1}^{n} {z_i}^2 \tag{11}$$

このとき n を『**自由度**』といい ν であらわす。$\chi^2_{(0)}$ が従う確率分布を『**自由度 $\boldsymbol{\nu = n}$ の $\boldsymbol{\chi^2}$ 分布**』という。

1.1.6 χ^2 分布の和

二つの確率変数 u_1, u_2 がそれぞれ独立に χ^2 分布に従うとする。

$$u_1 = \sum_{i=1}^{n} {z_i}^2 \tag{12}$$

$$u_2 = \sum_{j=1}^{m} {z_j}^2 \tag{13}$$

u_1 と u_2 の和を u_3 とあらわす。

$$\begin{aligned} u_3 &= u_1 + u_2 && (14)\\ &= \sum_{i=1}^{n} {z_i}^2 + \sum_{j=1}^{m} {z_j}^2 && (15)\\ &= \sum_{i=1}^{n} {z_i}^2 + \sum_{i=n+1}^{n+m} {z_i}^2 && (16)\\ &= \sum_{i=1}^{n+m} {z_i}^2 && (17) \end{aligned}$$

u_3 は自由度 $n+m$ の χ^2 分布に従う。

1.1.7 χ^2 分布の自然数倍

確率変数 w_n が自由度 $\nu = n$ の χ^2 分布に従うとする。

$$w_n = \sum_{i=1}^{n} {z_i}^2 \tag{18}$$

すると、

$$2w_n = w_n + w_n \tag{19}$$

は、自由度 $\nu = n + n = 2n$ の χ^2 分布に従う。

これに w_n を加えると、

$$2w_n + w_n = (w_n + w_n) + w_n \tag{20}$$

$$= w_n + w_n + w_n = 3w_n \tag{21}$$

は、自由度 $\nu = 2n + n = 3n$ の χ^2 分布に従う。これを一般化する。

a を自然数とすると

$$aw_n = w_n + w_n + \cdots + w_n \tag{22}$$

は、自由度 $\nu = an$ の χ^2 分布に従う。

1.1.8 母平均 μ が既知の場合の χ^2 分布

母平均 μ が既知 $(\mu = \mu^*)$ であるとする。正規母集団 $n\left(\mu^*,\ \sigma^2\right)$ からサンプル・サイズ $n > 0$ の標本 A を抽出し、i 番目の値を x_i であらわす。x_i を標準化した値の平方和を $\chi^2_{(\mu^*)}$ であらわす。この統計量

$$\chi^2_{(\mu^*)} = \sum_{x_i \in A}^{n} \left(\frac{x_i - \mu^*}{\sigma} \right)^2 \tag{23}$$

は、『**自由度 $\boldsymbol{\nu = n}$ の $\boldsymbol{\chi^2}$ 分布**』に従う。

1.1.9 母平均 μ が未知の場合の χ^2 分布

標本 A に属する要素 x_i の平均を『**標本平均**』といい、$\bar{X}$ であらわす。

$$\bar{X} = \frac{1}{n} \sum_{x_i \in A}^{n} x_i \tag{24}$$

標本平均 $\bar{X}$ を使って標準化した値の平方和を $\chi^2_{(\bar{X})}$ であらわす。この統計量

$$\chi^2_{(\bar{X})} = \sum_{x_i \in A}^{n} \left(\frac{x_i - \bar{X}}{\sigma} \right)^2 \tag{25}$$

は、『**自由度 $\boldsymbol{\nu = n - 1}$ の $\boldsymbol{\chi^2}$ 分布**』に従う。

1.1.10 標本平均の分布

正規母集団 $N\left(\mu,\ \sigma^2\right)$ からサンプル・サイズ n の標本を抽出し、標本平均 $\bar{X}$ を求めると、その期待値と分散は

$$\mathsf{E}\left(\bar{X}\right) = \mu \tag{26}$$

$$\mathsf{V}\left(\bar{X}\right) = \frac{\sigma^2}{n} \tag{27}$$

であり標本平均 $\bar{X}$ は

$$\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{n}\right) \tag{28}$$

である。

1.2 用語

- 結果に影響を及ぼすと考えられるものを『**因子**』という。
- 因子の条件を『**水準**』という。
- 観測において、全ての水準に対して 1 回ずつ観測することを 1 単位とする。つまり 1 単位の観測において水準の数だけデータが取得される。この観測単位を便宜上『**実験**』とよぶ。実験は何度も繰り返してデータが収集される。
- 1 因子による複数の水準の母平均に差があるかどうかを検定する分析を『**1 元配置分散分析**』という。

1.3 構造モデル

水準を識別する添え字を i とし水準の総数を n とする。実験を識別する添え字を j とし実験の総数を m とする。第 i 水準、第 j 回目の実験の結果を y_{ij} であらわす。μ_i を『**第 i 水準の母平均**』とする。ε_{ij} を『**誤差**』とし、独立に標準正規分布 $N(0,\ 1)$ に従うと仮定する。水準数 n の一元配置分散分析のデータの構造を

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad ;1 \leqq i \leqq n,\ 1 \leqq j \leqq m \tag{29}$$

とあらわす。[1)]

1） 分散分析では誤差 ε_{ij} が独立に同一の $N(0,\ 1)$ に従うと仮定しているので各水準の分散は等しいことが前提となる。

ここで μ を『**一般平均**』といい

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\mu_i \tag{30}$$

と定義する。

α_i を『**第 i 水準の効果**』といい

$$\alpha_i = \mu_i - \mu \tag{31}$$

と定義する。すると、第 i 成分の母平均 μ_i は

$$\mu_i = \mu + \alpha_i \tag{32}$$

である。これを (29) に代入すると

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{33}$$

を得る。ここで

$$\begin{align} \sum_{i=1}^{n}\alpha_i &= \sum_{i=1}^{n}(\mu_i - \mu) \tag{34} \\ &= \sum_{i=1}^{n}\mu_i - n\mu \tag{35} \\ &= 0 \tag{36} \end{align}$$

である。

1.3.1 仮説

水準の母平均と一般平均との差があるかどうかの検定は、水準の効果 $\alpha_1,\ \alpha_2,\ \cdots,\ \alpha_n$ が 0 かどうかの検定である。そこで、仮説を

$$H_0: \quad \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0 \tag{37}$$

$$H_1: \quad \alpha_1,\ \alpha_2,\ \cdots,\ \alpha_n \text{のうち少なくとも 1 つは 0 でない。} \tag{38}$$

とする。

1.3.2 全体平均と水準平均

構造モデル

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{29}$$

に対して y_{ij} の水準 i および実験 j に関する平均を『**全体平均**』といい $\bar{y}$ であらわし、水準 i に関する平均を『**水準平均**』といい $\bar{y}_i$ であらわす。

$$\bar{y} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} \tag{39}$$

$$\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij} \quad ; i = 1, 2, \cdots, n \tag{40}$$

数値例

全体平均

$$\bar{y} = \frac{75.5 + 82.0 + 75.0 + 67.5}{20} = 15.0 \tag{41}$$

水準平均

$$\bar{y}_1 = \frac{75.5}{5} = 15.1 \tag{42}$$

$$\bar{y}_2 = \frac{82.0}{5} = 16.4 \tag{43}$$

$$\bar{y}_3 = \frac{75.0}{5} = 15.0 \tag{44}$$

$$\bar{y}_4 = \frac{67.5}{5} = 13.5 \tag{45}$$

表 2　数値例 1

実験	A_1	A_2	A_3	A_4
第 1 回目	15.2	16.8	15.0	14.6
第 2 回目	15.7	16.2	13.9	12.8
第 3 回目	13.9	16.1	14.7	14.3
第 4 回目	16.2	17.6	16.3	12.1
第 5 回目	14.5	15.3	15.1	13.7
合計	75.5	82.0	75.0	67.5

1.3.3 変動の分解

一般平均の推定値を $\hat{\mu}$ であらわし、その推定値を全体平均 $\bar{y}$ で定義する。

$$\hat{\mu} = \bar{y} \tag{46}$$

第 i 水準の効果 α_i の推定値を $\hat{\alpha}_i$ であらわし、水準平均と全体平均の差で定義する。

$$\hat{\alpha}_i = \bar{y}_i - \bar{y} \tag{47}$$

誤差 ε_{ij} の推定値を $\hat{\varepsilon}_{ij}$ であらわし観測値 y_{ij} と水準平均 $\bar{y}_i$ の差で定義する。

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_i \tag{48}$$

そして観測された y_{ij} をそれぞれの推定値の和であらわすと

$$\begin{aligned} y_{ij} &= \hat{\mu} + \hat{\alpha}_i + \hat{\varepsilon}_{ij} && (49) \\ &= \bar{y} + (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i) && (50) \end{aligned}$$

右辺第 1 項を左辺へ移項すると

$$y_{ij} - \bar{y} = (\bar{y}_i - \bar{y}) + (y_{ij} - \bar{y}_i) \tag{51}$$

(51) 左辺は、「各測定結果 y_{ij} と全体平均 $\bar{y}$ の差」であり、(51) 右辺は、「水準平均 $\bar{y}_i$ と全体平均 $\bar{y}$ の差」と「各測定結果 y_{ij} と水準平均 $\bar{y}_i$ の差」の和である。つまり (51) は $y_{ij} - \bar{y}$ を $\bar{y}_i - \bar{y}$ と $y_{ij} - \bar{y}_i$ に分解したことを意味する。

1.3.4 全変動

(51) を二乗し、i および j に関して総和をとる[2)]。

$$\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y})^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}\left((\bar{y}_i-\bar{y})+(y_{ij}-\bar{y}_i)\right)^2 \tag{52}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}\left((\bar{y}_i-\bar{y})^2+2(\bar{y}_i-\bar{y})(y_{ij}-\bar{y}_i)+(y_{ij}-\bar{y}_i)^2\right) \tag{53}$$

$$= m\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{m}(\bar{y}_i-\bar{y})(y_{ij}-\bar{y}_i) + \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2 \tag{54}$$

2）分散分析では平均からの偏差の平方和を『**変動**』という。

ここで (54) 第 2 項は

$$2\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\bar{y}_i-\bar{y}\right)\left(y_{ij}-\bar{y}_i\right)=2\sum_{i=1}^{n}\left(\left(\bar{y}_i-\bar{y}\right)\sum_{j=1}^{m}\left(y_{ij}-\bar{y}_i\right)\right) \tag{55}$$

$$=2\sum_{i=1}^{n}\left(\left(\bar{y}_i-\bar{y}\right)\times 0\right) \tag{56}$$

$$=0 \tag{57}$$

よって (52) は

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\bar{y}\right)^2=m\sum_{i=1}^{n}\left(\bar{y}_i-\bar{y}\right)^2+\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\bar{y}_i\right)^2 \tag{58}$$

(58) 左辺は、全ての y_{ij} の全平均からのばらつきの程度をあらわし右辺第 1 項は m 個の水準平均の全平均からのばらつきの程度をあらわす。右辺第 2 項は水準内における y_{ij} の誤差によるばらつきの程度を意味する。

分散分析では、

- $\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y})^2$ を『**全変動**』といい、S_T であらわす。
- $m\sum_{i=1}^{n}(\bar{y}_i-\bar{y})^2$ を『**級間変動**』といい、S_A であらわす。
- $\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2$ を『**誤差変動**』[3)]といい、S_e であらわす。

3) 『**級内変動**』ともいう。

(58) は

$$S_T = S_A + S_e \tag{59}$$

であって、全変動を級間変動と誤差変動に分解しているので**『変動の分解』**という。

$$S_T = \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} - \bar{y})^2 \tag{60}$$

$$S_A = m \sum_{i=1}^{n} (\bar{y}_i - \bar{y})^2 \tag{61}$$

$$S_e = \sum_{i=1}^{n} \sum_{j=1}^{m} (y_{ij} - \bar{y}_i)^2 \tag{62}$$

数値例

表 2 をもとに変動を分解する。

全体平均

$$\bar{y} = 15.0 \tag{41}$$

水準平均

$$\bar{y}_1 = 15.1 \tag{42}$$

$$\bar{y}_2 = 16.4 \tag{43}$$

$$\bar{y}_3 = 15.0 \tag{44}$$

$$\bar{y}_4 = 13.5 \tag{45}$$

水準平均は水準の総数 $n = 4$ だけある。

そして、水準 i の効用の効果 $\hat{\alpha}_i$ は水準平均 μ_i と全体平均 μ の差 $\bar{y}_i - \bar{y}$ なので

$$\hat{\alpha}_1 = 15.1 - 15.0 = 0.1 \tag{63}$$

$$\hat{\alpha}_2 = 16.4 - 15.0 = 1.4 \tag{64}$$

$$\hat{\alpha}_3 = 15.0 - 15.0 = 0.0 \tag{65}$$

$$\hat{\alpha}_4 = 13.5 - 15.0 = -1.5 \tag{66}$$

である。これらの二乗和の $m = 5$ 倍が S_A である。

表 3 $\boldsymbol{y_{ij} - \bar{y}}$

実験	A_1	A_2	A_3	A_4
第 1 回目	0.2	1.8	0.0	−0.4
第 2 回目	0.7	1.2	−1.1	−2.2
第 3 回目	−1.1	1.1	−0.3	−0.7
第 4 回目	1.2	2.6	1.3	−2.9
第 5 回目	−0.5	0.3	0.1	−1.3

これらの二乗和が S_T である。

表 4 $\boldsymbol{y_{ij} - \bar{y}_i}$

実験	A_1	A_2	A_3	A_4
第 1 回目	0.1	0.4	0	1.1
第 2 回目	0.6	−0.2	−1.1	−0.7
第 3 回目	−1.2	−0.3	−0.3	0.8
第 4 回目	1.1	1.2	1.3	−1.4
第 5 回目	−0.6	−1.1	0.1	0.2

これらの二乗和が S_e である。

問題 X − 1 − 1

S_T, S_A, S_e をもとめなさい。

数値例 表 2 をもとに S_T, S_A, S_e を求めると

$$
\begin{aligned}
S_T = &\quad 0.04+3.24+0.00+0.16+0.49+1.44+1.21+4.84+1.21+1.21 \\
&+0.09+0.49+1.44+6.76+1.69+8.41+0.25+0.09+0.01+1.69 \\
=&\ 34.76 \qquad (67)
\end{aligned}
$$

$$
S_A = 5 \times (0.01+1.96+0.00+2.25) = 21.10 \qquad (68)
$$

$$
\begin{aligned}
S_e = &\quad 0.01+0.16+0.00+1.21+0.36+0.04+1.21+0.49+1.44+0.09 \\
&+0.09+0.64+1.21+1.44+1.69+1.96+0.36+1.21+0.01+0.04 \\
=&\ 13.66 \qquad (69)
\end{aligned}
$$

である。

1.4　$\boldsymbol{F}$ 検定

第 i 水準の水準平均の推定値 $\hat{\mu}_i = \bar{y}_i$ は、

$$\bar{y}_i = \mu + \alpha_i + \bar{\varepsilon}_i \tag{70}$$

であって、全体平均の推定値 $\hat{\mu} = \bar{y}$ は

$$\bar{y} = \mu + \bar{\varepsilon} \tag{71}$$

である。

そして第 i 水準の効果 $\hat{\alpha}_i$ は

$$\begin{aligned}
\hat{\alpha}_i &= \bar{y}_i - \bar{y} && (47)\\
&= (\mu + \alpha_i + \bar{\varepsilon}_i) - (\mu + \bar{\varepsilon}) && (72)\\
&= \mu + \alpha_i + \bar{\varepsilon}_i - \mu - \bar{\varepsilon} && (73)\\
&= \alpha_i + \bar{\varepsilon}_i - \bar{\varepsilon} && (74)
\end{aligned}$$

である。

水準の母平均に差が無く

$$\alpha_i = 0 \tag{75}$$

の場合であっても、(74) より、

$$\hat{\alpha}_i = \bar{\varepsilon}_i - \bar{\varepsilon} \tag{76}$$

なので、水準 i の効果 $\hat{\alpha}_i$ は 0 になるとは限らない。

そこで、水準 i

$$\hat{\alpha}_i = \bar{y}_i - \bar{y} \tag{47}$$

のばらつきの程度が、水準 i の母平均 μ_i の違いによるものなのか、単に誤差 ε_{ij} だけによるものなのかを、級間変動 S_A と誤差変動 S_e を比較して判断する。

表 5 データの構造

実験	A_1	A_2	$\cdots$	A_n
第 1 回目	$y_{11} = \mu + \alpha_1 + \varepsilon_{11}$	$y_{21} = \mu + \alpha_2 + \varepsilon_{21}$	$\cdots$	$y_{n1} = \mu + \alpha_n + \varepsilon_{n1}$
第 2 回目	$y_{12} = \mu + \alpha_1 + \varepsilon_{12}$	$y_{22} = \mu + \alpha_2 + \varepsilon_{22}$	$\cdots$	$y_{n2} = \mu + \alpha_n + \varepsilon_{n2}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
第 m 回目	$y_{1m} = \mu + \alpha_1 + \varepsilon_{1m}$	$y_{2m} = \mu + \alpha_2 + \varepsilon_{2m}$	$\cdots$	$y_{nm} = \mu + \alpha_n + \varepsilon_{nm}$

1.4.1 S_A が従う分布

天下り的に

$$\sum_{i=1}^{n}(\bar{y}_i - \bar{y}) \tag{77}$$

を考える。

ここで

$$\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij} \tag{39}$$

であって、

$$\begin{align} \bar{y} &= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} \tag{40} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m} y_{ij}\right) \tag{78} \\ &= \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i \tag{79} \end{align}$$

である。

構造モデル

$$y_{ij} = \mu_i + \varepsilon_{ij} \tag{29}$$

において、右辺第１項の μ_i は第 i 水準の母平均なので定数である。右辺第二項 ε_{ij} が標準正規分布に従うから、

$$\varepsilon_{ij} \sim N\left(0,\ 1\right) \quad \Longrightarrow \quad y_{ij} \sim N\left(\mu_i,\ 1\right) \tag{80}$$

よって

$$\mathsf{V}\left(y_{ij}\right) = 1 \tag{81}$$

である。

そして水準平均 $\bar{y}_i$ は $N(\mu_i,\ 1)$ に従う母集団から抽出されたサンプル・サイズ m の標本における標本平均

$$\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij} \tag{39}$$

の分散だから、

$$\mathsf{V}\left(\bar{y}_i\right) = \frac{1}{m} \tag{82}$$

従って標本 i における標本平均 $\bar{y}_i$ は

$$\bar{y}_i \sim N\left(\mu_i,\ \frac{1}{m}\right) \tag{83}$$

である。

この $\bar{y}_i$ を $\bar{y}_i$ の標本平均と標準偏差

$$\bar{y} = \frac{1}{m}\sum_{j=1}^{m}\bar{y}_i \tag{84}$$

$$\sigma = \sqrt{\frac{1}{m}} \tag{85}$$

を使って標準化した値の平方和を $\chi^2_{(\bar{y})}$ とあらわすと

$$\chi^2_{(\bar{y})} = \sum_{i=1}^{n}\left(\frac{\bar{y}_i - \bar{y}}{\sqrt{\frac{1}{m}}}\right)^2 \tag{86}$$

$$= m\sum_{i=1}^{n}(\bar{y}_i - \bar{y})^2 = S_A \tag{87}$$

は自由度

$$\nu = n - 1 \tag{88}$$

の χ^2 分布に従う。

1.4.2　S_e が従う分布

次に S_e が従う分布を考察する。

$$S_e = \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2 \tag{62}$$

$$= \sum_{i=1}^{n}\left(\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2\right) \tag{89}$$

ここで括弧内の総和部分

$$\sum_{j=1}^{m}(y_{ij}-\bar{y}_i)^2 \tag{90}$$

は自由度 $\nu = m-1$ の χ^2 分布にしたがうから、(89) 全体として、つまり、S_e は、自由度

$$\nu = n(m-1) = nm-n \tag{91}$$

の χ^2 分布に従う。

1.4.3 S_T が従う分布

級間変動 S_A の自由度は

$$\nu_A = n - 1 \tag{92}$$

誤差変動 S_e の自由度は

$$\nu_e = nm - n \tag{93}$$

したがって、$S_A + S_e$ の自由度は

$$\begin{aligned} \nu_A + \nu_e &= (n-1) + (nm - n) &(94)\\ &= nm - 1 &(95) \end{aligned}$$

S_A は自由度 $\nu = n - 1$ の χ^2 分布に従う統計量であり、S_e は自由度 $\nu = nm - n$ の χ^2 分布に従う統計量だから、その和

$$S_T = S_A + S_e \tag{59}$$

は自由度 $\nu = nm - 1$ の χ^2 分布に従う統計量である。

1.4.4 $\boldsymbol{F}$ 分布

S_A は自由度 $\nu = n - 1$ の χ^2 分布に従い、S_e は自由度 $nm - n$ の χ^2 分布に従うから

$$V_A = \frac{S_A}{n-1} \tag{96}$$

$$V_e = \frac{S_e}{nm-n} \tag{97}$$

から統計量

$$F_0 = \frac{V_A}{V_e} \tag{98}$$

をつくると、統計量 $F_{(0)}$ は自由度 $(n-1,\ nm-n)$ の F 分布に従う。

H_1 の下では、

$$\mathsf{E}\left(V_A\right) > \mathsf{E}\left(V_e\right) \tag{99}$$

なので、有意水準 5% として、

$$F(n-1,\ nm-n,\ 5\%) \tag{100}$$

で片側検定をおこなう。

1.4.5　分散分析表

分散分析に必要な統計量をまとめた表を分散分析表という。数値例表 2 を例にすると

$$S_A = 21.10 \tag{68}$$

$$S_e = 13.66 \tag{69}$$

$$\nu_A = n - 1 = 4 - 1 \tag{101}$$

$$\nu_e = n(m-1) = 4(5-1) \tag{102}$$

$$F_0 = \frac{S_A/\nu_A}{S_e/\nu_e} = \frac{7.033333}{0.853750} = 8.2381650 \tag{103}$$

表 6　分散分析表

項目	変動	自由度	不偏分散	F 値
級間	S_A	$n-1$	$\frac{S_A}{n-1}$	$\frac{\left(\frac{S_A}{n-1}\right)}{\left(\frac{S_e}{nm-n}\right)}$
誤差	S_e	$nm-n$	$\frac{S_e}{nm-n}$	

$$F(3,\ 15,\ 0.05) = 3.287382 \tag{104}$$

- $F_0 > F(3,\ 15,\ 0.05)$ なので H_0 は棄却され、「全ての水準の効果は 0 である」といえない。
- つまり、水準の効果は少なくとも 1 つは 0 ではないと、95% の確率で断言することができる。
- 一定の有意水準のもとで帰無仮説が棄却されることを『**有意になる**』という[4)]。
- 「有意に 0 と異なる」等と表現されることもある。
- F 検定の場合、「F 検定で有意となる」等と表現される。

4) 検定全般で使われる表現であり、帰無仮説が棄却されることを『**有意**』と表現する。帰無仮説が棄却できないときを『**非有意**』という。

1.5　まとめ

- 1 因子による複数水準の効果の有無を検定する分析手法が一元配置分散分析。
- 構造モデルの誤差に標準正規分布を仮定すると、級間変動と誤差変動が χ^2 分布に従う。
- 級間変動と誤差変動による F 検定が一元配置分散分析である。

参考文献

1.『経済数学教室 1 巻』小山昭雄著「岩波書店」1994 年 5 月 30 日
2.『新体系 大学数学 入門の教科書 上下』吉沢光雄著「講談社」2022 年 6 月 16 日
3.『技術者のための高等数学 7 確率と統計（原書代 8 版)』E・クライツィグ著　田栗正章訳「培風館」2004 年 11 月 30 日第 8 版
4.『データ解析のための数理統計入門』久保川達也著「共立出版」2023 年 10 月 15 日初版
5.『統計学入門』東京大学教養学部統計学教室篇「東京大学出版会」1991 年 7 月 9 日初版
6.『行動科学における統計解析法」芝祐順・南風原朝和著 「東京大学出版会」1990 年 3 月 20 日　初版
7.『計量経済学の理論』A.S. ゴールドバーガー著, 福地崇生・森口親司訳「東洋経済新報社」

1970 年 9 月 10 日 第 1 刷

8.『計量経済学序説』R.J. ウォナコット・T.H. ウォナコット著, 国府田恒夫・田中一盛訳「培風館」 1975 年 6 月 10 日初版
9.『マーケティング・サイエンス』片平秀貴著 「東京大学出版会」1987 年 4 月 20 日初版
10.『回帰分析』佐和光男著 「朝倉書店」1979 年 4 月 20 日初版
11.『完全独習 統計学入門』小島寛之著「ダイヤモンド社」2006 年 9 月 28 日 初版
12.『分散分析の基礎』高橋敬子著「プレアデス出版」2009 年 10 月 1 日