二元配置モデルにおける交互作用の統計的検定法

2014SS019 生田将人

指導教員：白石高章

1 はじめに

繰り返しのある場合の二元配置分散分析モデルにおいて交互作用の検定法について考察する.

2 二元配置モデルの設定

要因 A の各水準 A_i $(i = 1, \cdots, r)$, 要因 B の各水準 B_j $(j = 1, \cdots, s)$ および, 水準 A_i と B_j の交互作用効果 γ_{ij} について, 次の構造を仮定する.

$$X_{ijk} \equiv \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$
$$(i = 1, \cdots, r\,;\, j = 1, \cdots, s\,;\, k = 1, \cdots, t)$$
$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{s} \beta_j = \sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{s} \gamma_{ij} = 0$$

このとき, 実験の繰り返し数はすべての水準において等しいことを仮定する. ただし, μ は総平均, α_i は要因 A の主効果, β_j は要因 B の主効果, γ_{ij} は A と B の交互作用とする. $\{e_{ijk} \mid i = 1, \cdots, r\,;\, j = 1, \cdots, s\,;\, k = 1, \cdots, t\}$ は誤差を表す変量で, 互いに独立であり, $e_{ijk} \sim N(0, \sigma^2)$ である. また, $(\alpha_1 \leqq \cdots \leqq \alpha_r)$ または $(\beta_1 \leqq \cdots \leqq \beta_s)$ という制約条件を課してもよい.

表 1 二元配置分散分析

	要因 B_1	$\cdots$	要因 B_s
要因 A_1	$X_{111}, \cdots, X_{11t}$	$\cdots$	$X_{1s1}, \cdots, X_{1st}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
要因 A_r	$X_{r11}, \cdots, X_{r1t}$	$\cdots$	$X_{rs1}, \cdots, X_{rst}$

各標本平均は

$$\bar{X}_{ij\cdot} \equiv \frac{1}{t} \sum_{k=1}^{t} X_{ijk} \quad (i = 1, \cdots, r\,;\, j = 1, \cdots, s)$$
$$\bar{X}_{i\cdot\cdot} \equiv \frac{1}{st} \sum_{j=1}^{s} \sum_{k=1}^{t} X_{ijk} \quad (i = 1, \cdots, r)$$
$$\bar{X}_{\cdot j\cdot} \equiv \frac{1}{rt} \sum_{i=1}^{r} \sum_{k=1}^{t} X_{ijk} \quad (j = 1, \cdots, s)$$
$$\bar{X}_{\cdots} \equiv \frac{1}{rst} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} X_{ijk}$$

と表現できる. また,

$$n \equiv rst - rs$$

とおく.

3 検定統計量

帰無仮説 H_{01} : すべての i, j で $\gamma_{ij} = 0$

対立仮説 H_{11} : ある i, j が存在して $\gamma_{ij} \neq 0$

を考える.

また, 誤差平方和 S_E と要因 A と要因 B の交互作用による平方和 S_I, 誤差分散 $\hat{\sigma}^2$ をそれぞれ

$$S_E \equiv \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} \left(X_{ijk} - \bar{X}_{ij\cdot} \right)^2$$
$$S_I \equiv t \sum_{i=1}^{r} \sum_{j=1}^{s} (\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots})^2$$
$$\hat{\sigma}^2 \equiv \frac{S_E}{n}$$

とおく.

4 交互作用の検定

F 検定統計量は,

$$F_I \equiv \frac{\frac{S_I}{\sigma^2} \frac{1}{(r-1)(s-1)}}{\frac{S_E}{\sigma^2} \frac{1}{n}} = \frac{\frac{S_I}{(r-1)(s-1)}}{\frac{S_E}{n}}$$

である. このとき文献 [2] の説明と同様の証明方法により次の定理を得る.

定理 1 帰無仮説 H_{01} の下で, F_I は自由度 $(r-1)(s-1), n$ の F 分布に従う.

証明 $\hat{\gamma}_{ij} = \bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots}$ とすると, H_{01} の下で,

$$\frac{t}{\sigma^2} Cov(\hat{\gamma}_{ij}, \hat{\gamma}_{i'j'}) = \begin{cases} \dfrac{(r-1)(s-1)}{rs} & (i = i',\, j = j') \\ \dfrac{1-r}{rs} & (i = i',\, j \neq j') \\ \dfrac{1-s}{rs} & (i \neq i',\, j = j') \\ \dfrac{1}{rs} & (i \neq i',\, j \neq j') \end{cases}$$

が成立する. 計算の簡易化のため, それぞれを a, b, c, d とおく.

$$a \equiv \frac{(r-1)(s-1)}{rs},\ b \equiv \frac{1-r}{rs},\ c \equiv \frac{1-s}{rs},\ d \equiv \frac{1}{rs}$$

$$\mathbf{Z} \equiv (Z_1, \cdots, Z_{rs})^T$$
$$\equiv (\bar{X}_{11\cdot} - \bar{X}_{1\cdot\cdot} - \bar{X}_{\cdot 1\cdot} + \bar{X}_{\cdots}, \cdots, \bar{X}_{rs\cdot} - \bar{X}_{r\cdot\cdot} - \bar{X}_{\cdot s\cdot} + \bar{X}_{\cdots})^T$$

に対して $\mathbf{Z}$ の分散共分散行列 A を求める.

ただし, $Z_1, \cdots, Z_{rs}$ は互いに独立で同一の $N(0,1)$ に従う.

$$A \equiv \frac{t}{\sigma^2} V(\mathbf{Z})$$

とおくと, A^2 の各要素は次の 4 つに場合分けができる.
$(i = i', j = j')$ のとき

$$a^2 + b^2(s-1) + (r-1)\left\{c^2 + d^2(s-1)\right\} = \frac{(r-1)(s-1)}{rs}$$

$(i = i', j \neq j')$ のとき

$$2ab + b^2(s-2) + (r-1)\left\{2cd + d^2(s-2)\right\} = \frac{1-r}{rs}$$

$(i \neq i', j = j')$ のとき

$$2\{ac + bd(s-1)\} + (r-2)\left\{c^2 + d^2(s-1)\right\} = \frac{1-s}{rs}$$

$(i \neq i', j \neq j')$ のとき

$$2\{ad + bc + bd(s-2)\} + (r-2)\left\{2cd + d^2(s-2)\right\} = \frac{1}{rs}$$

したがって, A^2 と A のすべての要素は等しくなるため,

$$A^2 = A \quad \cdots ①$$

が成立する. また, A は分散共分散行列であるため, 対称行列である. すなわち,

$$A^T = A \quad \cdots ②$$

も成立する. なお, A の階級 $\mathrm{tr}(A)$ は,

$$\mathrm{tr}(A) = (r-1)(s-1) \quad \cdots ③$$

となり, ①, ②, ③ より A は階数 $(r-1)(s-1)$ の rs 次の対称な巾等行列である.

$\hat{\boldsymbol{\gamma}} \equiv (\hat{\gamma}_{11}, \cdots, \hat{\gamma}_{rs})^T$ とおくと, $(\sqrt{t}\,\hat{\boldsymbol{\gamma}})/\sigma = A\mathbf{Z}^T$ と表現できる.
ここで, S_I/σ^2 は,

$$\frac{S_I}{\sigma^2} = \mathbf{Z}^T A \mathbf{Z}$$

と表現できるため, H_{01} の下で, 文献 [1] の定理 3.22 より,

$$\frac{S_I}{\sigma^2} \sim \chi^2_{(r-1)(s-1)}$$

が成立することが示される.
さらに, S_E/σ^2 は常に χ^2 分布に従うことがいえるため, H_{01} の下で, 文献 [1] の定理 3.20 より,

$$F_I \sim F_n^{(r-1)(s-1)}$$

が成立する. □

F_I を検定統計量として, 自由度 $(r-1)(s-1), n$ の F 分布の上側 $100\alpha\%$ 点を $F((r-1)(s-1), n; \alpha)$ とすると, 水準 α の検定方式は検定関数 $\phi(\cdot)$ を用いて,

$$\phi(\mathbf{X}) = \begin{cases} 1 & (F_I > F((r-1)(s-1), n; \alpha) \text{ のとき}) \\ 0 & (F_I < F((r-1)(s-1), n; \alpha) \text{ のとき}) \end{cases}$$

と表現される. すなわち, $F_I > F((r-1)(s-1), n; \alpha)$ のとき H_{01} を棄却し, $F_I < F((r-1)(s-1), n; \alpha)$ のとき H_{01} を棄却しない.

5 C 言語プログラムの解説

交互作用の統計解析法による検定結果を出力するプログラムを C 言語により作成した. 本論文では交互作用の統計解析法での main 関数プログラムのみを以下に記載する. 交互作用の統計解析法の詳細なプログラムについては, 本稿に記載した.

```
int  main(void){
    int r,s,t;
    int NU1,NU2;
    float Y[100][100][100];
    float Q,X;
    printf("要因 A の数を入力:");
    scanf("%d",&r);
    printf("要因 B の数を入力:");
    scanf("%d",&s);
    printf("繰り返し数を入力:");
    scanf("%d",&t);
    load(Y,r,s,t);
    output(Y,r,s,t);
    return 0;
}
```

5.1 朝食と身長のデータ

年齢と朝食に身長が関係しているかどうかを調べるにあたり, 2011 年から 2016 年の 6 年間のデータを使用した. 年齢は 16 歳から 19 歳までの 4 年間とし, 男性と女性で朝食と身長のデータを集めた.

5.2 解析結果

男性は F_I=3.877569, p 値= 0.002457 となり, α=0.01 で帰無仮説 H_{01} は棄却された. 女性は F_I=0.825476, p 値 =0.555948 となり, α=0.05 で棄却されなかった. この検定により男性の身長は身長に対する年齢と朝食の有無の効果は一様ではないことがわかる.

6 おわりに

本論文では二元配置モデルにおいての検定法式を提案した. 検定を行うための C 言語プログラムを作成し, 実際のデータを用いることによって交互作用の統計解析法をより理解を深めることができた.

参考文献

[1] 白石高章 :『統計科学の基礎』. 日本評論社, 東京, 2012.

[2] Shiraishi, T. (1991) Statistical inference based on aligned ranks for two-way manova with interaction. Ann.Inst. Statist. Math Vol.43, No.4, 715-734

[3] 総務省:統計局. 朝食の摂取状況別体格測定.
http://www.stat.go.jp/library/faq/faq21/faq21a06.htm
2018/1/25 閲覧