多重比較法における修正Scheffe法の提案

稲葉，太一
吉田，亮太

(Citation)
神戸大学大学院人間発達環境学研究科研究紀要,17(2):39-42

(Issue Date)
2024-03-31

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100487733

(URL)
https://hdl.handle.net/20.500.14094/0100487733

神戸大学大学院人間発達環境学研究科研究紀要
第17巻第2号　2024

研究論文

多重比較法における修正Scheffe法の提案

Proposal of Modified Scheffe Method in Multiple Comparison Procedure

稲葉　太一*　　吉田　亮太**
Taichi INABA*　Ryota YOSHIDA**

要約：多重比較法における対比の検定にはScheffe法がある。Scheffe(1953)によると、一元配置データにおける任意の対比はF検定によって包括的に検定することができる。しかし、この方法には2つの欠点がある。1つ目の欠点は、群の数を a とおいたとき少数の対比であっても第1自由度が(a-1)と過大になり検出力が下がることである。2つ目は、多重比較法において重要な概念であるType I FWE(Family-Wize-Error-Rate)が保証されていないことである。Type I FWEとは、多重比較法において興味の対象となる帰無仮説の族に対して、正しい帰無仮説のうちの1つ以上が棄却される誤りの確率である。今回、提案する修正Scheffe法は、利用者が興味を持つ有限個の対比に対して、本質的な対比の数を求めることで、1つ目の欠点を回避する。このことで、対比の検定における最適な手法として、高い検出力が確保される。また同時に、本質的な帰無仮説のうちの正しい帰無仮説を考えることで、2つ目の欠点であったType I FWEも保証できる。

キーワード：多重比較法、シェッフェ法、Type I FWE、対比、任意の対比

0．扱うデータと帰無仮説の族

多重比較法とは、多項目や多変量のデータに対して、その多重性に着目する手法群のことである。大きく分けて2つの手法群がある。1つ目は、分布の情報を用いない手法群で、ボンフェローニの不等式を用いることをスタートに、ホルム法などがある。もう1つの手法群は、分布の情報を用いる手法群で、チューキー法、ダネット法、ウィリアムズ法に代表される。シェッフェ法は、後者の群に属し、F検定を用いる手法である。対比較に関しては、チューキー法の改良版であるチューキー・クレーマー法や、ペリ法が用いられる。

多重比較法では、「**複数の帰無仮説に対して、これらの内で正しいものを1つ以上棄却する誤りの確率**」Type I FWE（Type I Family-Wize-Error rate）を制御することが重要な目標である。この論文では、一元配置データ

$x_{ij} \sim N(\mu_i, \sigma^2), i = 1 \sim a, j = 1 \sim n_i$

を扱い、これらの母平均に関する帰無仮説として、任意の対比

$H_0: \sum_{i=1}^{a} c_i \mu_i = 0$,　ただし、$\sum_{i=1}^{a} c_i = 0$

を考える。対比とは、2つの母平均が等しい仮説や、2つの母平均の平均値と3つ目の母平均が等しい、といった仮説を含む、非常に汎用性のある仮説のことである。このとき、Scheffe（1953）より、Scheffe法という手法が提案されている。

Hochberg and Tamhane（1987）や永田＆吉田（1997）によると、この手法は、すべての母平均が等しいという包括的な帰無仮説が成り立つ時には、任意の対比に対する統計量が、常に一元配置分散分析におけるF統計量以下であることから、Type I FWEが保証されることが示されている。

しかし、母数の配置が一般的な場合には、包括的な帰無仮説が正しいとは限らないため、Type I FWEが保証されていない。また、通常のScheffe法では、F検定の限界値 $F(\phi_A, \phi_E; \alpha)$ を用いる際に、分子の自由度として

$$\phi_A = a - 1 \qquad (0.1)$$

を採用している。しかし、対比を少数個のみ検定する際に、この自由度は過大である。そこで、この論文では、有限個の対比について検定する際には、まず本質的な対比の数を求め、これが（0.1）式の自由度を超えない時には採用することを提案する。また、これに伴って、Type I FWEも常に保証される

*　　神戸大学大学院　発達環境学研究科　准教授
**　大阪府庁

(2023年9月29日　受付
2023年12月14日　受理)

ことを示す。

１節では、従来の Scheffe 法の基本的な考え方を紹介する。２節では、改良した「修正 Scheffe 法」を提案する。３節では、修正 Scheffe 法が Type I FWE を保証することを示す。４節では、まとめを述べる。

１．従来の Scheffe 法

Scheffe 法とは、任意の対比

$H_0 : \sum_{i=1}^{a} c_i \mu_i = 0$, ただし、$\sum_{i=1}^{a} c_i = 0$

について、次の検定統計量

$$F = \frac{\left\{\sum_{i=1}^{a} c_i \bar{x}_i\right\}^2 / \phi_A}{V_E \sum_{i=1}^{a} c_i^2 / n_i} \tag{1.1}$$

を考え、棄却限界値としては、一元配置分散分析での $F(\phi_A, \phi_E; \alpha)$ を用いる手法である。

一元配置分散分析とは、すべての母平均が一致するという帰無仮説（包括的帰無仮説という）

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a \tag{1.2}$$

に対して、検定統計量

$$F_0 = \frac{V_A}{V_E} = \frac{S_A/\phi_A}{S_E/\phi_E} \tag{1.3}$$

が棄却限界値 $F(\phi_A, \phi_E; \alpha)$ を越えると帰無仮説を棄却するという手法である。ここで、

$$\text{要因平方和：} S_A = \sum_{i=1}^{a} n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2 \tag{1.4}$$

$$\text{誤差平方和：} S_E = \sum_{i=1}^{a} \sum_{j=1}^{n_i} \left(\bar{x}_{ij} - \bar{x}_i\right)^2 \tag{1.5}$$

$$\phi_A = a - 1,\ \phi_E = \sum_{i=1}^{a} n_i - a \tag{1.6}$$

である。

ここで、包括的帰無仮説（1.2）の下で、Scheffe 法の Type I FWE が α 以下になる理由を説明する。Scheffe 法の検定統計量（1.1）の分子に、コーシー・シュワルツの不等式を用いると、

$$\left\{\sum_{i=1}^{a} \frac{c_i}{\sqrt{n_i}} \sqrt{n_i} \left(\bar{x}_i - \bar{\bar{x}}\right)\right\}^2 \le \left(\sum_{i=1}^{a} \frac{c_i^2}{n_i}\right) \left\{\sum_{i=1}^{a} n_i \left(\bar{x}_i - \bar{\bar{x}}\right)^2\right\}$$

であるから、（1.1）式の F が、常に（1.3）式の F_0 以下であることが示される。

つまり､包括的帰無仮説（1.2）が成り立つ時には、F 検定の理論によって、F が $F(\phi_A, \phi_E; \alpha)$ 以上になる確率は、α 以下であることが確認できる。

したがって、任意の対比に対して、Type I FWE が α 以下だと思われるかもしれないが、これは誤解である。あくまでも、包括的な帰無仮説の下で成り立つにすぎない。これを以下の例で確認する。

１つの対比に興味を持った場合、（1.1）式の分子にある $\sum_{i=1}^{a} c_i \bar{x}_i$ は正規分布に従うと考えられる。このとき、これを標準化して分母に誤差分散を代入すると、この２乗である F は自由度 $(1, \phi_E)$ の F 分布に従うと考えられる。つまり、従来の Scheffe 法の $\phi_A = a - 1$ は大きすぎる。次節では、有限個の対比について、本質的な対比の数を自由度として利用することを考える。

２．修正 Scheffe 法のアルゴリズム

まず、すべての対比に興味を持つ場合には、$\phi_A^* = a - 1$ とする。以下、この節では、興味を持つ対比が有限個（ m^* 個）の場合を考える。最初に、これらの族を設定する。

$$\left\{H_0^{(k)} : \sum_{i=1}^{a} c_i^{(k)} \mu_i = 0, ただし \sum_{i=1}^{a} c_i^{(k)} = 0, (k = 1, \cdots, m^*)\right\}$$

これらの対比の係数 $c_i^{(k)}$ を並べた次の行列を考える。

$$\begin{pmatrix} c_1^{(1)} & \cdots & c_a^{(1)} \\ \vdots & \ddots & \vdots \\ c_1^{(m^*)} & \cdots & c_a^{(m^*)} \end{pmatrix}$$

この行列を行に関してシュミットの直交化を行う。直交化した後の行の数(対比の数)を ϕ_A^* とおく。更に、残った対比に関して、

$$F^{(k)} = \frac{\left\{\sum_{i=1}^{a} c_i^{(k)} \bar{x}_i\right\}^2 / \phi_A^*}{V_E \sum_{i=1}^{a} \left[c_i^{(k)}\right]^2 / n_i} \tag{2.1}$$

を計算し、$F(\phi_A, \phi_E; \alpha)$ 以上であれば、その対比を棄却し、そうでなければ保留する。この手法を修正 Scheffe 法と呼ぶ。次節では、この修正 Scheffe 法の Type I FWE が α 以下であることを示す。

３．修正 Scheffe 法の Type I FWE

多重比較において、Type I FWE を考えるときには、正しい仮説が何であるかに注目する。最初に、興味を持つ有限個（ m^* 個）の対比を考え、このうちの最初の m_0 個の対比が正しいとする。

$H_0^{(k)} : \sum_{i=1}^{a} c_{0i}^{(k)} \mu_i = 0$, ただし $\sum_{i=1}^{a} c_{0i}^{(k)} = 0, (k = 1, \cdots, m_0)$

これを書き改めると、対比全体は、以下のように表現できる。

$$C_0 \mu = \begin{pmatrix} c_{01}^{(1)} & \cdots & c_{0a}^{(1)} \\ \vdots & \ddots & \vdots \\ c_{01}^{(m_0)} & \cdots & c_{0a}^{(m_0)} \end{pmatrix} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_a \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

次に、この C_0 を用いて、以下の統計量を構成する。

$$C_0 \bar{x} = \begin{pmatrix} c_{01}^{(1)} & \cdots & c_{0a}^{(1)} \\ \vdots & \ddots & \vdots \\ c_{01}^{(m_0)} & \cdots & c_{0a}^{(m_0)} \end{pmatrix} \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_a \end{bmatrix}$$

$$= \begin{pmatrix} c_{01}^{*(1)} & \cdots & c_{0a}^{*(1)} \\ \vdots & \ddots & \vdots \\ c_{01}^{*(m_0)} & \cdots & c_{0a}^{*(m_0)} \end{pmatrix} \begin{bmatrix} \bar{x}_1^* \\ \vdots \\ \bar{x}_a^* \end{bmatrix} =: C_0^* \bar{x}^*$$

ただし、$c_{0i}^{*(k)} = \sigma \frac{c_{0i}^{(k)}}{\sqrt{n_i}}$, $\bar{x}_i^* = \frac{\bar{x}_i - \mu_i}{\sqrt{\sigma^2/n_i}} \sim N(0,1)$ とする。

この構成した C_0^* について、次のように下側に単位行列を追加し、行ベクトルに関してシュミットの直交化した行列を C_1^* とする。

$$\begin{bmatrix} C_0^* \\ I \end{bmatrix} \rightarrow C_1^* = \begin{bmatrix} C^* \\ J \end{bmatrix}$$

このとき、C_1^* は直交行列であるから、$T_1 = C_1^* \overline{x}^*$ は a 次元の標準正規分布に従う。また、この行列の上側には正しい対比に関する係数が並んでいる。これを C^* とする。この行数を m とおき、次の統計量を考える（この m は本質的な正しい対比の数である）。

$$T = C^* \overline{x}^*$$

すると、T の各成分が互いに独立に標準正規分布に従っているため、これら成分の2乗和 $T'T$ は自由度 m のカイ二乗分布に従う。

ここで、任意の対比を1つ固定する。この対比を1番目として直交化しなおしても、上記の理論は成り立つ。つまり、最初に選ばれた対比による統計量を t_1^2 と表すと $t_1^2 \leq T'T$ が常に成り立つ。

したがって、Type I FWE の定義に戻って、正しい帰無仮説の内で1つ以上を棄却する確率を考えると、正しい対比に対する検定統計量の分子の最大値が $T'T$ を越えないことがわかる。

一方、一元配置分散分析での理論により、平均値 $\overline{x}_i$ と誤差分散の推定量 V_E は、互いに独立である。よって、平均値 $\overline{x}_i$ の一次結合である T と誤差分散推定量 V_E も独立であり、

$$\frac{T'T/m}{V_E}$$

は自由度 (m, ϕ_E) の F 分布に従う。つまり、正しい対比を1つ以上棄却する確率は自由度 (m, ϕ_E) の F 分布の棄却限界値を用いれば制御できる。

ここで、m には複数の可能性があり、データを分析する立場ではわからない情報である。ただし、3節で求めた自由度 m は、2節で求めた ϕ_A^* を越えることはない。ここで、次の性質に注目しよう。

補題 3.1　F 分布の分位点 $F(m, n; \alpha)$ は、$m_1 < m_2$ に対して、次の不等式が成り立つ。

$$m_1 F(m_1, n; \alpha) < m_2 F(m_2, n, \alpha) \qquad (3.1)$$

証明　3つの確率変数 (X_1, Z, Y) が、互いに独立で、自由度がそれぞれ $(m_1, m_2 - m_1, n)$ のカイ二乗分布に従うとする。ここで、$X_2 = X_1 + Z$ は自由度 m_2 のカイ二乗分布に従う。このとき、$i = 1,2$ に対して、$F_i = \frac{X_i/m_i}{Y/n}$ は自由度対 (m_i, n) の F 分布に従う。よって、

$$\alpha = Pr\{F_i \geq F(m_i, n; \alpha)\} \ (i = 1,2) \qquad (3.2)$$

ここで、$Z > 0$ であるから $X_1 < X_2$ ゆえ、次の関係が成り立つ。

$$m_1 F_1 = \frac{X_1}{Y/n} < \frac{X_2}{Y/n} = m_2 F_2 \qquad (3.3)$$

(3.2)，(3.3) 式より、(3.1) 式が成り立つ。

証明終

修正 Scheffe 法の判定基準である (2.1) 式の両辺を ϕ_A^* 倍すると、統計量と「F 分布の分位点の第1自由度倍」を比較すると読み取れる。補題 3.1 より、小さい自由度を採用することで棄却しやすくなることがわかる。

これは、より小さい第1自由度で検定することが望ましいという直感とも一致する。この補題 3.1 により、2節の手法の Type I FWE が保証されることもわかる。したがって、次の定理が示される。

定理 3.1：修正 Scheffe 法の Type Ⅰ FWE は α 以下である。

なお、補題 3.1 によると、従来の Scheffe 法は修正 Scheffe 法よりも、常に棄却されにくいことがわかる。したがって、従来の Scheffe 法の Type Ⅰ FWE も α 以下であることがわかる。ただし、補題 3.1 は、従来の Scheffe 法よりも、修正 Scheffe 法の検出力が常に高いことを示している。

4．まとめ

まず、従来「Scheffe 法」は包括的な帰無仮説が正しい時には、Type Ⅰ FWE が α 以下になっているが、これ以外の母数の配置の際には、α 以下であるとは示されていなかった。今回、少数個の対比しか興味がない時でも、分子の自由度を $\phi_A = a - 1$ としていることで、最適性もないことが確認できた。

更に、今回提案した「修正 Scheffe 法」は、Type Ⅰ FWE が常に α 以下であることも示されており、分子の自由度も、対比の本質的な本数以下に設定できる、最適な手法であることが示された。

なお、蛇足ながら、今回の修正 Scheffe 法の提案と、その Type Ⅰ FWE の証明によって、従来の Scheffe 法には最適性はないものの、Type Ⅰ FWE は制御されていることがわかった。

今後の、修正 Scheffe 法の、多方面での利用に期待したい。

5．謝辞

著者等は、査読者の親切なコメントに助けられました。誠にありがとうございました。

6．参考文献

Hochberg, Y. and Tamhane, A.C.（1987）: Multiple Comparison Procedures, John Wiley & Sons.

Scheffe, H. (1953) : "A method for judging all contrasts in the analysis of variance", Biometrika 40, pp.87-104.
永田靖＆吉田道弘（1997）：統計的多重比較法の基礎、サイエンティスト社