R による多重比較法の研究

2000MM015 堀内　賢太郎
指導教員　松田 眞一

1　はじめに

医学・薬学に関する統計的データ解析の分野において、統計的多重比較法は必要不可欠のものになっている。そこで、その考え方を少しでも理解すると同時に、R 上で簡単な検定のプログラムを作成することを目的とした。特に私は、すべての対比較を一度におこなうテューキー (Tukey) の方法と対比較を段階的におこなうテューキー・ウェルシュ (Tukey-Welsch) の方法についてプログラミングをおこなう。

2　多重比較法とは

吉村 [2] では、「いくつかの群または水準がある。その相互に、平均値的な値（パラメータとして数値で実現できればそれでよい）において、差異があるかどうかを、検定という推測形式で確認したい。群の各対で検定をおこなうと、多重性のために公称の有意水準に比べて、第 1 種の過誤 (type I error) の確率が大きくなる。この現象を防ぐために多重性を考慮した公称の有意水準、つまり棄却限界値あるいは棄却域の調整をおこなって検定する。このやり方を多重比較法という。」と述べられている。

3　テューキー (Tukey) の方法と手順

Tukey の方法は、母平均について群間ですべての対比較を同時に検定するための多重比較法である。また、この方法は、母集団分布に対してある特定の確率分布を仮定して、その確率分布を規定する母数（パラメータ）について推測をおこなう統計的方法（パラメトリック法）である。そのため、前提条件として母集団は正規分布であるとし、すべての群で母分散は等しいと仮定する。その下で a 個の群の母平均についてすべての対比較を考えていく。記号 $H_{\{i,j\}}$ は帰無仮説 $H_{\{i,j\}}:\mu_i=\mu_j$ を表すものとして、推測の対象となるファミリー（帰無仮説族）を $\mathcal{F}=\{H_{\{1,2\}},\cdots,H_{\{1,a\}},H_{\{2,3\}},\cdots,H_{\{2,a\}},\cdots,H_{\{a-1,a\}}\}$ と設定する。ここでは、帰無仮説に対する対立仮説として $H^A_{\{i,j\}}:\mu_i\neq\mu_j$ の形のものだけを考える (永田・吉田 [1] 参照)。

手順 1
推測の対象となるファミリーを明示する。

$$\mathcal{F}=\{H_{\{1,2\}},\cdots,H_{\{1,a\}},H_{\{2,3\}},\cdots,H_{\{2,a\}},\cdots,H_{\{a-1,a\}}\}$$

手順 2
有意水準 α を定める。$\alpha=0.05$ または 0.01 と定めることが多い。

手順 3
データに対して、それぞれの群ごとに平均 $\bar{x}_i$ および分散 V_i を計算する。

$$\bar{x}_i=\frac{\sum_{j=1}^{n_i}x_{ij}}{n_i},V_i=\frac{\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2}{n_i-1}\quad(i=1,2,\cdots,a)$$

手順 4
誤差自由度 ϕ_e および誤差分散 V_E を計算する。

$$\phi_e=N-a=n_1+n_2+\cdots+n_a-a$$

$$V_E=\frac{\sum_{i=1}^{a}(n_i-1)V_i}{\phi_E}$$

手順 5
すべての i と j の組み合わせに対して検定統計量 t_{ij} を計算する。

$$t_{ij}=\frac{\bar{x}_i-\bar{x}_j}{\sqrt{V_E(\frac{1}{n_i}+\frac{1}{n_j})}}\quad(i,j=1,2,\cdots,a;i<j)$$

手順 6
$\max_{i,j}|t_{ij}|$ の分布に対する $|t_{ij}|$ の p 値（有意確率）を用いて検定をおこなう。p 値が有意水準 α より小さいならば帰無仮説 $H_{\{i,j\}}$ は棄却され、μ_i と μ_j には差があると判断する。逆に p 値が有意水準 α 以上ならば、その帰無仮説を保留する。「保留する」という意味は「$H_{\{i,j\}}$ が成り立っているとみなす」という意味ではなく、「μ_i と μ_j には差（有意差）があるとは言えない」という程度の意味である。

4　テューキー・ウェルシュ (Tukey-Welsch) の方法と手順

Tukey-Welsch の方法は、対比較をおこなう際にステップダウン法（下降手順）を用いた多重比較法である。手順が異なるだけで、求めたいものは Tukey の方法と同じであり、パラメトリック法を用いた多重比較法である。

ここでは、帰無仮説 $H_{\{1,2\}},H_{\{2,3\}},H_{\{1,2,3\}},H_{\{2,4,5\}}$, $H_{\{1,2\cdots,5\}}$ などを H_P と表現する。ここで P は群番号を表す数字の集合で、これらの帰無仮説については、それぞれ $P=\{1,2\},\{2,3\},\{1,2,3\},\{1,2,\cdots,5\}$ が対応する。また、集合 P に含まれる要素の個数を p（小文字）で表し、p を H_P の大きさとよぶ。さらに、帰無仮説 H_P の検定のために用いられる検定統計量としてステューデント化された範囲の統計量 (Q 統計量) を用いる。

$$Q_P=\max_{i,j\in P}\sqrt{2}|t_{ij}|=\max_{i,j\in P}\frac{\sqrt{2}|\bar{x}_i-\bar{x}_j|}{\sqrt{V_E\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}$$

ここでの $|t_{ij}|$ は $H_{\{i,j\}}$ の検定統計量であり、これは、μ_i と μ_j の差を検出するためのものである。したがって、$Q_P=\max_{i,j\in P}\sqrt{2}|t_{ij}|$ は μ_i と $\mu_j(i,j\in P)$ の組み合わせの中で一番大きな違いを検出するために構成されている。つまり、Q_P は H_P の検定統計量と考えることができる。

また、有意確率 (p 値) を min_P と表す。これは、Q 統計量の場合 $\max_{i,j}|t_{ij}|$ の分布に対する $|t_{ij}|$ の p 値（有

意確率）を意味する (永田・吉田 [1] 参照)。

手順 1

推測の対象となるファミリーを明示する。

$$\begin{aligned}\mathcal{F} &= \{\text{ すべての } H_P, P \subseteq \{1,2,\cdots,a\}\} \\ &= \{H_{\{1,2\}},\cdots,H_{\{a-1,a\}},H_{\{1,2,3\}},\cdots,H_{\{1,2,\cdots,a\}}\}\end{aligned}$$

手順 2

有意水準 α を定める。$\alpha = 0.05$ または 0.01 と定めることが多い。プログラムでは min_P と有意水準 α_p を比較することで検定をおこなう。ここで、α_p は

$$\begin{aligned}&\alpha_p = 1-(1-\alpha)^{p/a} \quad (2 \le p \le a-2 \text{ の時}) \\ &\alpha_{a-1} = \alpha_a = \alpha\end{aligned}$$

と定める。明らかに、$\alpha_2 \le \alpha_3 \le \cdots \le \alpha_{a-1} = \alpha_a = \alpha$ という関係がある（例えば、$a = 4$ で有意水準 $\alpha = 0.05$ とすると $\alpha_2 = 0.0253, \alpha_3 = 0.0500, \alpha_4 = 0.0500$ となる)。これらはテューキー・ウェルシュの有意水準の配分とよばれ、有意水準 α を個々の帰無仮説 H_P の大きさ p に応じ、シダック (Šidák) の不等式に基づいて配分したものである。

手順 3

検定統計量 Q_P を定める。

手順 4

与えられた形式のデータをランダムダンプリングし、それぞれの群ごとに平均 $\bar{x}_i$ および分散 V_i を計算する。

手順 5

誤差自由度 ϕ_E および誤差分散 V_E を計算する。

手順 6

大きさ $p = a$ すなわち $P = \{1,2,\cdots,a\}$ とおく。検定統計量 Q_P の値を計算する（**手順 3** で決めたもの)。そして、その Q_P から min_P を求める（**3 節 (手順 6)** 参照)。α_p は**手順 2** で求めたものである。

(1)$\alpha_p \le min_P$ となるなら、ファミリーに含まれるすべての仮説を保留して検定作業を終了する。

(2)$\alpha_p > min_P$ となるなら H_P を棄却して手順 7 へ。

手順 7

大きさ $p = a-1$ であるすべての P に対して H_P を検定する。このとき min_P は P に応じた値を用いる。

(1)$\alpha_p \le min_P$ となるなら、$R \subseteq P$ となるすべての H_R（H_P が誘導 (imply) するすべての H_R）を保留する。

(2)$\alpha_p > min_P$ となるなら H_P を棄却する。

$p = a-1$ となるすべての H_P が保留されるか $a-1 = 2$ なら、検定作業を終了する。一方、$a-1 > 2$ であり、棄却された大きさ $a-1$ の H_P があれば手順 8 へ。

手順 8

ある段階において大きさ p の H_P が棄却されたら、大きさ $r = p-1$ の $R(\subseteq P)$ のうちで以前の段階で保留と判定されていない H_R を検定する。このとき、min_R は R に応じた値を用いる。

(1)$\alpha_r \le min_R$ となり H_R が保留されるなら、H_R とそれが誘導するすべての仮説を保留する。

(2)$\alpha_r > min_R$ となるなら H_R を棄却する。

大きさ $r = p-1$ となるすべての H_R が保留されるか $p-1 = 2$ なら、検定作業を終了する。一方、$p-1 > 2$ であり、棄却された大きさ $r = p-1$ の H_R があれば、r をあらためて p とおいてさらに手順 8 を続け、そうでなければ終了する。

5 テューキー・ウェルシュの必要性

ここではすべての対比較をおこなう多重比較法の中で、Tukey と Tukey-Welsch の方法を取り上げた。前者はシングルステップ法で後者はステップダウン法であるので、両者を比べると後者のほうが対比較の仮説を多く棄却する可能性がある。実際、本論文中の数値結果においても、サンプルサイズが等しい場合、前者は 4 つ、後者は 5 つ、サンプルサイズが異なる場合、前者は 5 つ、後者は 6 つ対比較の仮説が棄却されている（これは、前者の場合に棄却される仮説は、後者の場合にも棄却されることを意味する)。

つまり、後者は有意水準の配分をおこなうことによって、棄却限界値が前者よりも小さくなるため、より細かく検定することができる。従って、後者のほうが有意になりやすい（検出力が高い）といえる。

また、一般的なソフトには Tukey の方法をおこなう多重比較法（前者）が入っているが、Tukey-Welsch の方法をおこなう多重比較法（後者）は入っていないことのほうが多い。これは、後者が最良な方法でないことが分かっている（対比較において後者よりも検出力の高い方法が存在する）ことや、知名度の問題があるといえる。しかし、私が今回プログラミングをおこなった方法以外にも、基本的な多重比較法を応用し、より検出力の高い多重比較法へと発展させた方法が存在する。従って、前者だけでなく、後者の方法も理解し、プログラミングすることが必要であるといえる。

6 おわりに

テューキーの方法をおこなうプログラムは、検定統計量、有意確率、検定結果を 1 つの while 文にまとめることができたので、満足している。また、両者ともサンプルサイズが異なるデータでも検定できるようにデータの読み込ませ方を工夫した。

テューキー・ウェルシュの方法をおこなうプログラムでは、階乗と組み合わせを求める 2 つのプログラムも外部関数として作成できた。

苦労した点は、有意確率を求める際に使用した ptukey コマンドである。ここに検定統計量として、あらかじめ $\sqrt{2}$ 倍したものを代入すること（基準化すること）がなかなか理解できなかった。

参考文献

[1] 永田靖・吉田道弘, 「統計的多重比較法の基礎」, サイエンティスト社,1997.

[2] 吉村功,“多重比較方式の問題点”, 統計数理研究所共同研究レポート 18,「多重比較方式の諸問題」, pp.58-71,1989.