一元配置多重比較の予備検定の改良

南山大学経営学部 松田 眞一

An Improvement of the Preliminary Test of Multiple Comparisons for One-way Layout

Shin-ichi MATSUDA

要 旨 一元配置多重比較で誤差に正規分布を仮定できるとき、処理群の等分散性に疑問のある場合がある。その場合は予備検定を行って等分散性をチェックし手法の選択を行うことが一般的によく用いられている方法である。しかし、その際に総合的な有意水準が保たれていないことはあまり知られていない。本稿の目的は従来の予備検定法を改良してシミュレーションでその動向を探ることにある。本稿で提案する不等分散の下での多重比較法と不等分散を重視した予備検定法は総合的な有意水準をほぼ保つことが確認された。

1. は じ め に

ここで取り上げるデータは一元配置であり、各群の例数は揃っていなくてもよい。特にデータ設計上は例数が揃っていたものの欠測値によって不揃いになったものを想定している。このようなデータは医薬試験などで見られるものであるが、その多くは誤差に独立で同一な正規分布を仮定する。しかし、中には誤差の等分散性に疑問のあるデータも見受けられ、実務家の多くはそのようなデータに対して予備検定を行って等分散性をチェックしている。

不等分散であることがはっきりしている場合の平均の比較に意味があるのかどうかはまだ議論のあるところであるが、実際は等分散であるかどうかが不明であるから不等分散に影響されない手法が必要である。また、あとの例で示すように平均自体にはっきりした意味をもつ場合もあり、このときはたとえ不等分散であっても平均の比較には意味がある。

松田 (1993) ではそのような状況で不等分散と判定されたときに用いる手法の性質を研究している。同様の研究は Dunnett (1980) でなされていて、GH (Games-Howell) 手順および Dunnett の提案した C 手順、T3 手順を比較して次の結果を得ている。

- ・ GH 手順は例数が少なくアンバランスだと有意水準を保てないことがある。
- ・ 例数が多いとき C 手順、少ないときは T3 手順がよい。

松田 (1993) ではそれに付加して次の結果を得た。

Key words: Monte Carlo simulation, Pairwise comparisons, Unbalanced data, Unequal variances.

- ・ T3 手順も例数が少なくアンバランスだと有意水準を保てないことがある。
- ・ C 手順と T3 手順を使い分けるなら例数が 40 ぐらいとなる。
- ・ GH 手順と C 手順の限界値を平均した方法である GH-C 手順は例数が少ないとき T3 手順より若干劣る方法である。

さらにその中で、予備検定を含めた下での有意水準の動向も調べられているが、その結果は以下の点で思わしくなかった。二標本問題においても予備検定の有意水準の設定について十分注意すべきであるとされているため（永田（1992）参照）、松田（1993）では多重比較の予備検定の有意水準を 50% まで上げる工夫を施した。それにもかかわらず、例数が少なくて不揃いである場合には不等分散に敏感に反応して有意水準が保たれない現象が観察されている。

本稿の目的は、有意水準の調整では抜本的改良にはつながらないと考え、新たな予備検定方式を提案してその有意水準の動向をシミュレーションで探ることにある。また、不等分散のときにはノンパラメトリックな手法を用いるという手順も広く普及しているがその問題点も考察する。

なお、本稿で考察する手法は、等分散性が仮定できるときの Tukey-Kramer の方法（以下では T-K 手順と呼ぶ）と等分散性が仮定できない場合に松田（1993）で提案した手法を改良した GH-C2 手順であり、ノンパラメトリックな手法としては Steel-Dwass の方法（以下では S-D 手順と呼ぶ）を取り上げる。T-K 手順と S-D 手順はそれぞれの状況でもっともポピュラーなものであり、GH-C2 手順は後で述べるように不等分散の場合に他の手順より適切なものとして提案されるものである。また、等分散性の予備検定は Bartlett 検定統計量を基礎として構築する。

2. モデルと多重比較法

次のような一元配置のデータを考える。

処理 1	$x_{11}, x_{12}, \dots, x_{1n_1}$
処理 2	$x_{21}, x_{22}, \dots, x_{2n_2}$
	
処理 k	$x_{k1}, x_{k2}, \dots, x_{kn_k}$

データ x_{ij} の構造は

$$x_{ij} = \mu_i + \varepsilon_{ij}$$

というように、処理の平均的効果を表す母数 μ_i と互いに独立に平均 0、分散 σ^2 の正規分布に従う誤差項 ε_{ij} とに分解できると仮定する。また、帰無仮説の族は以下で与えられる F とする。

$$F = \{\mu_i = \mu_j, 1 \leq i < j \leq k\}$$

すなわち、すべての処理の対に対して多重比較を行う場合を想定している。各手法は、帰無仮説の族 F に含まれる帰無仮説のうち一つでも誤って棄却する確率、すなわち Type I familywise error rate (FWE) を抑制することが要求される。（松田・永田（1990）参照）

名目有意水準を α として、本稿で取り扱う手法を簡潔に以下で示す。

どの手順でも検定統計量 T_{ih} および棄却限界値 ξ_{ih} について $T_{ih} \geq \xi_{ih}$ ならば、 μ_i と μ_h には差があると判定する。

2.1. T-K 手順

T-K 手順の検定統計量は次式で与えられる.

$$T_{ih} = \frac{|\bar{x}_{i.} - \bar{x}_{h.}|}{\sqrt{s^2(1/n_i + 1/n_h)}}$$

ただし,

$$\bar{x}_{i.} = \sum_{j=1}^{n_i} x_{ij}/n_i, \quad s^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2/\nu, \quad \nu = \sum_{i=1}^k n_i - k$$

である. 一方, 限界値は

$$\xi_{ih} = Q_{k,\nu}^{(\alpha)}/\sqrt{2}$$

で与えられる. ここで, $Q_{k,\nu}^{(\alpha)}$ は自由度 ν , 大きさ k のスチューデント化した範囲の分布の上側 100α パーセント点である. (Hochberg and Tamhane (1987) 参照)

この手順は処理群が等分散であるならば, Type I FWE を抑制する.

2.2. GH-C2 手順

GH-C2 手順は Games and Howell (1976) が提案した GH 手順と Dunnett (1980) が提案した C 手順とを組み合わせたものである.

この 2 つの手順では同じ検定統計量

$$T_{ih} = \frac{|\bar{x}_{i.} - \bar{x}_{h.}|}{\sqrt{s_i^2/n_i + s_h^2/n_h}}$$

を用いている. ただし,

$$s_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2/\nu_i, \quad \nu_i = n_i - 1$$

である. 一方, 限界値 ξ_{ih} はそれぞれ GH 手順では

$$Q_{k,\nu_{ih}}^{(\alpha)}/\sqrt{2}$$

C 手順では

$$\frac{Q_{k,\nu_i}^{(\alpha)}(s_i^2/n_i) + Q_{k,\nu_h}^{(\alpha)}(s_h^2/n_h)}{\sqrt{2(s_i^2/n_i + s_h^2/n_h)}}$$

で与えられる. ただし,

$$\nu_{ih} = \frac{(s_i^2/n_i + s_h^2/n_h)^2}{s_i^4/n_i^2(n_i - 1) + s_h^4/n_h^2(n_h - 1)}$$

である. (Hochberg and Tamhane (1987) 参照)

松田 (1993) では漸近的には同等となるこの 2 つの方法の限界値の平均を新しく限界値とする方法を提案し, GH-C 手順と名付けた. 例数が少ないとき GH 手順は有意水準を保てず, C 手順は非常に保守的だったからである. ちなみに T3 手順は漸近的に劣るため限界値を平均化する手順には向かない. シミュレーションの結果 GH-C 手順は例数が少ないとき T3 手順より若干劣る

方法となることが分った。ここで提案する GH-C2 手順とは、GH-C 手順の限界値の組合せ方を工夫し、不等分散の場合に他の手順より適切な手法としたものである。

まず GH-C 手順を松田 (1993) で考察した範囲より広く調べた結果、有意水準を保てない場合も出てくることが分かった。単純な改良としては C 手順の重みを増すことが考えられるが、例数が揃っていて等分散の場合には実質有意水準がひどく落ち込み検出力の低下を免れない。そこで限界値の重み付けを動的にしてそれぞれの場合に対応することとした。

いくつかの場合についての C 手順、GH 手順、GH-C 手順の実質有意水準の結果が表 1 である。(名目有意水準は 5%) これから分かるように GH-C 手順の実質有意水準は近似的に C 手順と GH 手順の調和平均となっている。そのため各々の場合で実質有意水準を名目有意水準に揃えるために最適な重み付けを求めることができる。C 手順と GH 手順の限界値の比を $a:1$ とした場合の a の値が表 1 に記してある。

この a の値を事前に求めるための予測式を構成する。様々な例数や分散に対応する簡便な予測式を設計するのは難しい。ここではいくつかの指標を元に単純な線形回帰で予測することにする。まず a の変動に大きな影響があると考えられる指標の検討であるが、例数のばらつきが大きかったり例数の小さいところで分散が大きかったりすると a が大きくなることが分かるのでそれらを反映した指標をいくつか設定した。その後、表 1 のような様々な場合のデータを 400 個準備し、それらの結果がどの指標でよく説明できるかを指標の組合せも含めて重回帰分析で検討していった。最終的に比較的簡便で予測が有効な式として以下のものが導出された。

$$a = 5.12x$$

ただし、

$$x = \sum_{i=1}^k \left(\frac{n_i}{\bar{n}} - 1 \right)^2 \sqrt{\frac{1}{k} \sum_{i=1}^k \left(\frac{\sigma_i^2}{n_i \sigma^2} - \tau \right)^2}$$

$$\bar{n} = \frac{1}{k} \sum_{i=1}^k n_i, \quad \tau = \frac{1}{k} \sum_{i=1}^k \frac{\sigma_i^2}{n_i \sigma^2}, \quad \sigma^2 = \frac{\sum_{i=1}^k \nu_i \sigma_i^2}{\sum_{i=1}^k \nu_i}$$

その様子を示したのが図 1 である。導出された回帰直線は点線で示してある。ここで提案する GH-C2 手順はその結果を元に以下のような限界値を用いた手法である。

表 1. 各手順の実質有意水準 (名目有意水準 5%)

例 数	分 散	実 質 有 意 水 準			a
		C 手順	GH 手順	GH-C 手順	
2,2,2	1,1,1	0.007	0.027	0.012	-0.1387
2,2,2	1,1,3	0.007	0.029	0.011	-0.1179
4,4,2	1,1,1	0.019	0.057	0.027	0.0753
4,4,2	1,1,3	0.026	0.083	0.042	0.4307
4,4,2	3,3,1	0.015	0.042	0.022	-0.0816
4,4,2	1,1,10	0.039	0.101	0.055	1.7903
8,8,4	1,1,1	0.031	0.059	0.042	0.2489
8,8,4	1,1,10	0.035	0.059	0.045	0.3559
4,4,4,2	1,1,1,3	0.027	0.084	0.041	0.4752
4,4,4,2	1,1,1,10	0.035	0.103	0.055	1.2006

$$\xi_{ih} = \frac{1}{\sqrt{2}(\hat{a}+1)} \left(Q_{k, \nu_{ih}}^{(a)} + \hat{a} \frac{Q_{k, \nu_i}^{(a)}(s_i^2/n_i) + Q_{k, \nu_h}^{(a)}(s_h^2/n_h)}{s_i^2/n_i + s_h^2/n_h} \right)$$

ただし,

$$\hat{a} = 5\bar{x} + 0.6$$

ここで $\bar{x}$ は上の x の σ_i に標本標準偏差を代入したものである。なお、有意水準を確実に保つために回帰残差の標準偏差の約 3 倍の値を定数項として付加し、データ依存度を軽減するために係数の有効桁数は 1 桁とした。その様子は図 1 の実線で示してある。この改良により例数が少数でも T3 手順とほぼ同様の有意水準を保ちながら非常にアンバランスな場合でも有意水準を保てるようになった。

図 1. CH-C2 手順のための比率の予測

2.3. S-D 手順

S-D 手順は、分布の仮定を要しないノンパラメトリック手法として Hochberg and Tamhane (1987) に紹介されている。したがって、誤差項 ε_{ij} は独立同一分布に従っていれば正規分布でなくとも構わない。もちろん、同一分布である以上等分散は仮定されているはずで、不等分散のときに用いるのは誤用である。ただし、この仮定に対してパラメトリック手法よりロバストであることも考えられる。

この手順の検定統計量は以下で与えられる。

$$T_{ih} = \frac{|R_{ih} - n_i(n_i + n_h + 1)/2| - 1/2}{\sqrt{n_i n_h V_{ih} / (n_i + n_h - 1)}}$$

ただし、 R_{ih} は第 i 群と第 h 群を併合して順位付けしたときの第 i 群の順位和で、 V_{ih} はその順位の分散である。なお、同順位がない場合は $V_{ih} = (n_i + n_h - 1)(n_i + n_h + 1)/12$ となる。一方、限界値は

$$\xi_{ih} = Q_{k, \infty}^{(a)} / \sqrt{2}$$

で与えられる。

3. 等分散性の予備検定法

従来の予備検定は等分散であることを帰無仮説にした通常の検定を用い、棄却されれば等分散を仮定しない手法へと進み、棄却されなければ等分散を仮定した手法へと進むという形式で行なわれてきた。実際予備検定が行っていることは場合分けであるにもかかわらず通常の検定という枠組みで行なわれていることは必ずしも適当なことではない、との指摘が最近はなされてきている。(永田 (1992) 参照) しかし、その考えはまだ広く浸透していない。それは予備検定を含めた手順の動向が詳しく調べられていないからであるが、それを理論的に行うのは難しい。本稿では

シミュレーションの立場からその検証を行いその改良を試みる。

松田 (1993) で示されているように多重比較問題では永田 (1992) が用いたような有意水準による調整では欠点を回避できない。それは等分散を仮定した手法が不等分散であることに敏感に反応して有意水準が保たれなかったり多重決定における群の扱いを著しく不平等にするせいである。これらを解決するには帰無仮説中心の従来の検定方式を取りやめ、もっと対立仮説を重視した新たな方法を作り上げるよりほかはないと考えられる。

まず、予備検定に必要と思われる性質を列挙してみよう。

- ・第二種の過誤を重視すること。(特に例数の少ないとき)
- ・例数が多いとき帰無仮説が真なら棄却されにくいこと。
- ・例数が多いとき対立仮説が真なら帰無仮説が棄却されやすいこと。

通常の検定方式は第1の性質を全く無視して第2の性質を固定的に棄却されにくくすることで実現している。もちろん、第3の性質は満たされている。実際は場合分けであるにもかかわらず第1の性質を無視しているのが問題であった。予備検定を行いたい場合はたいいてい次のような構図が成りたっているはずだからである。

等分散の仮定外では有意水準を保てない方法 vs. 制限はないが検出力の低い方法

そこで永田 (1992) のように第1の性質を考慮して第2の性質を緩める操作をしても多重比較問題ではうまくいかない。松田 (1993) では予備検定の有意水準を50%まで譲歩してもなお総合的に有意水準が保たれない現象が見られたのである。これ以上の有意水準の引き上げは第2の性質を無視することにほかならず初めから等分散を仮定しない方法を用いることとほぼ同等である。検出力の低下をいとわないのであればそれは一つの選択である。ここではあくまで3つの性質にこだわって改良を試みる。

本稿で提案する予備検定方式はBartlett検定を基礎とした、対立仮説に基づいた検定である。それは第二種の過誤の確率が一定となるように限界値を決定する方法であるが、通常、等分散性に関する対立仮説は複合仮説であるため、そのすべてについて第二種の過誤を抑制しようとするとは帰無仮説をほとんど採用しないのと同じになる。また、適当な単純対立仮説を定めて検定を行うとそれより帰無仮説に近い対立仮説については第3の性質が満たされず、それゆえ単純対立仮説の選定が大きく影響する。用いる検定統計量は非心カイ二乗分布に従うものとして検定するため、どのように対立仮説を扱うかは非心母数をどのように決定するかに集約される。その非心母数の定め方によって第3の性質を満たすように工夫した。

初めに対立仮説の下でのBartlett検定について述べておく。採用した検定統計量は修正済みの

$$B = \frac{\nu \log s^2 - \sum_{i=1}^k (\nu_i \log s_i^2)}{1 + \frac{\sum_{i=1}^k \nu_i^{-1} - \nu^{-1}}{3k-3}}$$

である。ただし、 s^2, s_i^2, ν, ν_i は前節で用いた記号であり、 $\log$ は自然対数である。対立仮説における非心母数は、Kendall and Stuart (1979) を参考にとすると以下の式で与えられることが分かる。

$$\lambda = \sum_{i=1}^k \frac{\nu_i}{2} \left(\frac{\sigma_i^2}{\sigma^2} - 1 \right)^2$$

ただし、 σ_i^2 は対立仮説の下での分散で、 $\sigma^2 = \sum_{i=1}^k \nu_i \sigma_i^2 / \nu$ である。しかし、実際シミュレーション

を行ってみると省いた項の影響で例数が多くなって非心母数が大きくなると当てはまりがよくなる。そこで Bartlett 検定統計量の形を保ったままの

$$\lambda = - \sum_{i=1}^k \nu_i \log(\sigma_i^2/\sigma^2)$$

を原形として用いることにした。(付録参照)

この λ を用いると特定の単純対立仮説 $H_1: \lambda = \lambda_0$ に対し第二種の過誤の確率を一定にした検定を行える。第3の性質を満たすためにはこの非心母数のオーダー ν_i での増加を抑制すればよいことになる。すなわち、 λ を定数に固定すれば帰無仮説の立場、オーダー ν_i で増加すれば対立仮説の立場であるのでその間のオーダーで増加する設定にすれば第3の性質を満たすのである。

単純な設定として $\sqrt{\nu_i}$ オーダーで増加する次の形のものがシミュレーションで比較的良好な有意水準の動向を示すことが確認された。

$$\lambda_0 = -\sqrt{n_{\min}-1} \sum_{i=1}^k \log(\sigma_{0i}^2/\sigma_0^2)$$

ただし、 $n_{\min}$ は n_i の最小値であり、 $\sigma_{0i}^2 = 28i - 27(i=1, 2, \dots, k)$, $\sigma_0^2 = \sum_{i=1}^k \sigma_{0i}^2/k$ である。計算の基礎となる σ_{0i}^2 は、まず、いくつかのパターンの中から処理群の数 k の影響の小さいものとして等差型に設定し、次に松田(1993)で有意水準50%の通常の予備検定を行ったとき例数10のあたりで実質有意水準が5%となったのに合わせてその差を決定した。そのときの実際の棄却率の変化の例を図2に示す。例数が少ないうちは対立仮説に重きを置いているが、例数が増えるにしたがって等分散を検出しやすくなっている様子が分かり、第2の性質を満たしている。結果に当てはめた曲線は、次の関数型で定めたものである。

$$a + \frac{b}{\sqrt{n}} + \frac{c}{n} + \frac{d}{n\sqrt{n}}$$

ここで、 n は処理1の例数で、最小二乗法により係数 a, b, c, d を決定した。

また、例数が10のときの帰無仮説の棄却率は処理群 k に対して以下のようになり、安定していることが分かる。

k	3	4	5	6	7
棄却率	0.51	0.52	0.51	0.49	0.48

なお、 λ_0 と元の λ を対比させると実際は $k=3$ のとき $n_i=10$ で(1, 4.5, 8), $n_i=20$ で(1, 3, 5)を限界値決定のための単純対立仮説の分散として設定していることになり、例数が増えるにつれて

図2. 新しい予備検定の棄却率
Sample size (n, n, n)

帰無仮説に近づくことが分かる。

最終的に本稿で考察した予備検定による手順は次のようになる。

手順1 Bartlett 検定統計量 B を計算する。

手順2 非心母数 λ_0 を計算する。

手順3 自由度 $k-1$ 非心母数 λ_0 の非心カイ二乗分布の下側 1% 点 B_0 を求める。

手順4 $B < B_0$ なら T-K 手順, $B \geq B_0$ なら GH-C2 手順で多重比較を行う。

4. シミュレーションの手順

シミュレーションは、以下の手順で行った。

手順1 処理数 k , 各処理における例数 $(n_1, n_2, \dots, n_k)$, 各処理の平均 $(\mu_1, \mu_2, \dots, \mu_k)$, 各処理の誤差項の分散 $(\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$ を定める。

手順2 正規乱数を発生させ、第2節のような一元配置データを構成する。

手順3 各手法に必要な検定統計量と限界値を計算する。(名目有意水準 α を 0.05 とする。) また、新しい予備検定方式により等分散かどうかを判定する。

手順4 各処理対の検定結果と Type I error の状況 (および対立仮説では検出状況) を調べる。

手順5 手順2から手順4までを 10,000 回繰り返し、各処理対の棄却率, Type I FWE, (対立仮説ではさらに各検出力) を求める。なお、シミュレーションの状況を確認するため 1,000 回ごとの結果に基づいてそれらの標準偏差も求めた。

シミュレーションにおいて求めた 1,000 回ごとの標準偏差はどの場合も妥当な値を示し、シミュレーションの方法自体に問題がなかったことを示唆していた。

なお、非心カイ二乗分布による検定は Patnaik (1949) のカイ二乗近似を用いて行った。すなわち、非心カイ二乗分布 $\chi^2_{m,\lambda}$ を $\rho\chi^2_{m'}$ で近似した。ただし、

$$\rho = 1 + \frac{\lambda}{m + \lambda}$$

$$m' = m + \frac{\lambda^2}{m + 2\lambda}$$

である。

本稿で調査した各パラメータは、まず、1 割程度の欠測を考えて $k=3, 4, 5$ で欠測値を 2 個まで、使い分けが重視される分散比として 3 倍までの整数比のすべての組み合わせを設定した。さらに、計画的な実験で得られたものでないデータを考慮して一群が他群の半分の例数であるデータや $k=6, 7$ のデータ、および、分散比 10 倍までのいくつかの配置についても検討した。以下では $k=3$ の場合について考察する。他の場合も傾向はほぼ同様であった。

5. 全体帰無仮説における結果の考察

各手法の性質を見るには、まず処理の平均がすべて等しい全体帰無仮説 $\mu_1 = \mu_2 = \dots = \mu_k$ における Type I FWE の動向を確認しなければならない。たとえ検出力の高い手法であっても有意水準を保てないのであれば使用に耐えないからである。本稿では新たに提案した GH-C2 手順と予

備検定による手順での有意水準に注目する。

結果のうち一部を図示したのが図3～10である。なお、これらの図は各シミュレーションにおける Type I FWE の値のプロットとそれに当てはめた曲線からなっている。当てはめた曲線は試行錯誤の末、次の関数型に定めた。

$$a + \frac{b}{n} + \frac{c}{n^2}$$

ここで、 n は処理1の例数で、最小二乗法により係数 a, b, c を決定した。この当てはめは各図の最小例数の近辺を除いて全体の傾向をつかむのに十分であると判断した。

なお、図の記号のうち PT と表示したものが予備検定による手順での結果である。

5.1. 等分散の場合

この場合は GH-C2 手順、予備検定による手順ともにほぼ有意水準を保っている様子が図3～4で確認できる。予備検定による手順は、組み合わせたそれぞれの手順が有意水準を保っているにもかかわらず、有意水準を保てないことがあるが、その破れは大きくても 0.3% 程度である。松田 (1993) の結果と比較すると GH-C2 手順は T3 手順とほぼ同等以上の有意水準を保っていて単独で用いるのに十分な手順といえる。

5.2. 不等分散の場合

例数が揃っている場合が図5, そうでない場合が図6～10である。図5のように例数が揃っている場合は T-K 手順の有意水準の破れが小さいため GH-C2 手順、予備検定による手順とも問題なく有意水準を保っている。一方、例数が揃っていない場合は、図6のように例数の少ないところの分散が大きい場合に予備検定による手順で若干の有意水準が保たれない現象がみられるものの松田 (1993) で行った有意水準 50% の通常の予備検定の場合より良好であり、実用上ほとんど問題ないと思われる。以下では実用上の適用範囲として予備検定による手順の実質有意水準が 6% 以下となる状況を検討していく。

図8のように例数が2倍程度のアンバランスな場合は T-K 手順の有意水準の破れは激しくなり、予備検定による手順も若干それに引きずられる。しかし、提案した GH-C2 手順は有意水準を

図3. Type I FWE
Sample size (n, n, n), Variance (1, 1, 1)

図4. Type I FWE
Sample size ($n, n, n-2$), Variance (1, 1, 1)

保っており、予備検定による手順の実質有意水準も6%は超えない。これ以上アンバランスでもGH-C2手順はそれほど問題はないと思われるが、予備検定による手順の実質有意水準が6%を超える可能性が高い。したがって、そのような場合は予備検定を行わずにGH-C2手順を用いるのが適当である。

一方不等分散が強い図9, 10の設定では例数が2の群がある場合のみGH-C2手順の有意水準を保てず、予備検定による手順の実質有意水準も6%を超えている。よって、例数が2の群があってアンバランスな場合は予備検定を行わずGH-C2手順を用いるのが適当である。

なお、松田(1993)で注意したように例数が揃っている場合でもT-K手順を安易に用いてはいけない。全体の有意水準の破れは小さくても処理対を対等に見ない欠点があるからである。その様子は後の例でも触れる。

6. 検出力の動向

松田・永田(1990)で平均の配置の種類の影響が少なく多重比較の総合的判断に適しているとした平均棄却数比で予備検定による手順の検出力を考察する。平均棄却数比とは母平均間に差のあるすべての対のうち検出された対の割合である。

図11~12は等間隔の平均の配置(EQ)に対する等分散のときの検出力の動向でT-K手順が最も適している場合の結果である。(fは平均の配置の標準偏差を表す。)例数が少ないときはT-K手順とGH-C2手順との差は大きく、図11では18%程度も離されている。予備検定による手順ではその差を4%ほど縮めることができ、効果が上がっていることが分かる。図12のように例数が多くなると3つの手順の差はほとんどないが、予備検定による手順の改善率は図11の場合より相対的に上がっている。

一方、不等分散の場合の結果はほとんど変わらないため図は省略した。T-K手順は有意水準を保てなくなるためその分だけ検出力が増加することは観察された。

図5. Type I FWE
Sample size (n, n, n), Variance (1, 1, 3)

図6. Type I FWE
Sample size ($n, n, n-2$), Variance (1, 1, 3)

7. S-D 手順との比較

特に注意を要する比較のために図 6, 8, 10, 12 に S-D 手順の結果を示してある。

まず, 全体帰無仮説の下での傾向であるが, 図 6 を見ても分かるように分散比 3 倍程度では有意水準が保たれている。しかし, 例数の少ないデータでは実質有意水準が非常に小さくなり, 全く検出できないこともあることが分かる。この点は漸近結果を用いずに離散確率を正確に計算すれば改善の余地があるが, 例数が非常に少ない場合はそれでも実質有意水準は 0 になる。

次に図 8, 10 より, 大きくアンバランスなときは T-K 手順と同様有意水準が保てなくなる様子が見えてくる。その程度は不等分散がひどくなるにつれて大きくなるが, T-K 手順と比較するとロバストである。

最後に図 12 で検出力の状況を考察する。S-D 手順は GH-C2 手順より常に下回っていることが分かる。データの分布が正規分布である場合, あるいはそうでなくても平均が正規分布に十分近い分布の場合は GH-C2 手順が有利である。ちなみに図 11 の設定では S-D 手順は全く何も検出できないということを付け加えておく。

総合的に判断すると S-D 手順は例数が少ないとき全く検出できず, 例数が多いときは T-K 手順よりロバストであるものの有意水準が保てなくなる方法である。さらに, 適用してもさしつかえない範囲であっても検出力は GH-C2 手順より若干劣るので正規性があるデータに対する多重比較法としては勧められない。

また, T-K 手順との使い分けはよく行なわれているような有意水準 5% での予備検定は論外として, もう少し高い有意水準を設定して行うとしても松田 (1993) の GH-C 手順と同様で例数が少ない場合は有意水準を保てない。本稿で提案した新しい予備検定方式を用いるなら総合的有意水準がほぼ保たれているものの図 8 のように非常にアンバランスだと全体的に有意水準を保てない。さらに, 図 6 のようなデータで例数が少ないときは予備検定の境界付近で T-K 手順と S-D 手順の検出状況に非常なアンバランスが生じる。すなわち, 不等分散と判断されたら平均の違いはほとんど (あるいは全く) 検出してくれないが, そうでないなら有意水準の倍近くも検出するといった状態になる。GH-C2 手順の場合も若干のアンバランスがあるが, S-D 手順よりは改善され

図 7. Type I FWE
Sample size ($n, n, n-2$), Variance (3, 3, 1)

図 8. Type I FWE
Sample size ($n, n, n/2$), Variance (1, 1, 3)

ている。

8. 例

この節では不等分散の場合でも平均の比較を行う必要がある例を取り上げ、予備検定による手順を実際に適用する。

渡井 (1993) は雑誌ごとの広告料とその反応からなるデータを回帰分析を用いて解析している。その結果は雑誌の効率を示す傾きに集約されるが、その比較はまさに不等分散の場合も考慮した平均の多重比較の枠組みに入る。本稿では元データのうち掲載回数が多い雑誌に限定し、渡井 (1993) が回帰の当てはまりのよくないものとして示した雑誌を除き、広告料の値上げの影響を除くため期間を短くしたデータで解析を行うこととした。比較した雑誌数は7である。

解析したデータを表2に示す。ここで、資料請求件数とは渡井 (1993) が示した方法で推定したもので、雑誌は特定されてもどの号に対する反応か不明のデータから相関の高くなるように旬ごとの資料請求件数を重み付けたものである。資料請求が雑誌だけでなく号までも特定できるならこの部分は正確に改善できる。なお、広告料は雑誌が特定されることのないように雑誌1の1ページ広告料で基準化してある。

渡井 (1993) は広告料について次の変換を施した後、原点を通る直線回帰で雑誌の効率を導きだしている。

$$\tilde{x}_i = x_i^2 / x_0$$

ここで、 x_0 とはその雑誌の1ページの広告料であり、この変換は1ページの広告料は変えずに半ページの広告料の効果を1/4に見積もったものである。先に回帰の当てはまりがよくない雑誌があると述べたが、それは半ページの広告料が1ページの半額から大きくずれているものであり、渡井 (1993) は広告スペースのほうからの変換を考えるとそれらの雑誌のほとんどこの関係が成り立つことを示している。そのような雑誌に対し個別の変換を施して他のものと同時に解析することも考えられるが、本稿ではそれらは対象外とすることにした。

図9. Type I FWE
Sample size ($n, n, n-2$), Variance (1, 1, 10)

図10. Type I FWE
Sample size ($n, n, n/2$), Variance (1, 1, 10)

変換した広告料と資料請求件数に対する回帰分析の様子を図 13 で示す。また、各雑誌の効率を示す傾きとその分散をまとめたものが表 3 である。このとき等分散性を残差に関してではなく、傾きに関して考えるなら一元配置と同様の状況となる。しかし、先のシミュレーションではこのデータのような設定ではどうなるか不明であるため次のデータでシミュレーションしてみた。

例数: (19, 19, 19, 19, 19, 19, 11)

分散: (1, 1, 1, 1, 1, 1, 45)

これはこのデータ解析で出てくる数値を最も影響の大きい場合に直したものであるが、予備検定による手順の Type I FWE は 4.0% であった。(ちなみに T-K 手順は 30% まで跳ね上がっている。)したがって、先の手順はこのデータ解析では問題なく効力を発揮するものと思われる。そこでこれを予備検定による手順で検定してみよう。

まず、等分散性の判断であるが、第 3 節に示した非心母数を計算すると $\lambda_0 = 14.04$ である。したがって、非心カイ二乗分布の限界値は $\chi^2_{11}(0.99) = 5.88$ で与えられる。表 3 のデータに基づく Bartlett 検定統計量の値は 86.60 であるので棄却され、不等分散の場合の GH-C2 手順に進む。第 2.2 節に示した GH-C2 手順により各対の検定統計量と限界値および検定結果は表 4 のようになる。*印のついているところが棄却された対である。この表では参考として T-K 手順だと検定結果がどのようなになるかも示してある。

得られた結果は少々解釈しづらい。例数が不揃いのデータや不等分散のデータに見られる現象であるが、平均の順序と棄却結果が必ずしも一致しないのである。今回のデータでは雑誌 3 が非常に大きな分散を持っているためずいぶん傾きの小さい雑誌の方まで棄却できない状態になっている。雑誌 3 は明かな外れ値を含んでいるためこのような結果になったのであるが、背景が不明のためあえてそのまま解析した。GH-C2 手順はその場合でもその雑誌の解析のみに注意すればよいだけで他の雑誌間の比較には影響せずその点でロバストといえる。T-K 手順では先に指摘した多重決定の不平等さがその際に見受けられる。雑誌 3 の分散の大きさが他に影響して棄却できなくなる処理対があり、この対の検出力が下がっていることになる。また、雑誌 3 も一見他の雑

図 11. 平均棄却数比 [EQ]
Sample size (4, 4, 4), Variance (1, 1, 1)

図 12. 平均棄却数比 [EQ]
Sample size (10, 10, 10), Variance (1, 1, 1)

表 2. 広告料と資料請求件数

雑 誌 1	雑 誌 2	雑 誌 3	雑 誌 4	雑 誌 5	雑 誌 6	雑 誌 7
0.51, 7.00	0.44, 3.75	0.27, 3.00	0.58, 3.75	0.42, 3.00	0.42, 6.00	0.60, 14.50
0.51, 6.67	0.44, 3.00	0.27, 2.25	0.58, 2.00	0.42, 1.50	0.42, 7.50	0.60, 13.00
0.51, 9.67	0.44, 4.00	0.27, 1.88	0.58, 1.25	0.42, 0.00	0.42, 2.50	0.60, 9.00
0.51, 9.00	0.44, 10.50	0.27, 1.13	0.58, 6.00	0.42, 0.75	0.42, 3.50	0.95, 43.00
0.51, 1.33	0.44, 6.00	0.27, 5.63	0.58, 0.00	0.42, 0.38	0.42, 8.25	0.95, 39.75
0.51, 4.67	0.44, 1.50	0.27, 0.00	0.58, 6.25	0.42, 1.75	0.42, 5.50	0.95, 35.50
1.00, 39.67	0.89, 26.25	0.27, 0.00	0.95, 24.25	0.73, 4.50	0.42, 0.00	0.95, 37.00
0.51, 10.00	0.44, 9.75	0.27, 1.88	0.95, 11.00	0.73, 15.75	0.77, 14.50	0.95, 34.50
1.00, 25.33	0.89, 22.50	0.27, 6.00	0.95, 12.25	0.73, 11.25	0.77, 20.25	0.95, 39.75
1.00, 41.67	0.89, 21.00	0.53, 45.00	0.95, 18.50	0.73, 19.50	0.77, 18.75	0.95, 27.25
1.00, 15.00	0.89, 12.00	0.53, 19.88	0.58, 9.00	0.42, 1.50	0.77, 16.25	0.95, 37.50
1.00, 35.33	0.89, 15.00	0.53, 51.75	0.58, 35.25		0.77, 19.25	0.95, 32.00
1.00, 27.67	0.89, 30.00	0.53, 16.88	1.20, 22.00		0.77, 12.25	0.95, 33.00
1.00, 52.00	0.89, 17.25	0.53, 18.38	0.95, 10.50		0.77, 16.50	0.60, 26.50
1.00, 20.00	0.89, 27.75	0.53, 4.88	0.95, 10.50		0.77, 17.25	0.95, 40.00
1.00, 35.00	0.89, 15.00	0.53, 10.13	0.95, 12.25		0.77, 19.75	0.95, 42.00
1.00, 33.00	0.89, 21.00	0.27, 4.13				
1.00, 30.00	0.89, 24.00	0.27, 4.50				
1.00, 21.00		0.53, 13.88				

期間は1990年8月～1992年7月、掲載しなかった号や広告料改定前のデータを除いたため不揃い。

誌と変わらぬ動向を示しその特異なところを見逃してしまうこととなるとともに、本当に分散が大きいなら他の雑誌間より有利に検出し過ぎている。

さて、GH-C2 手順に基づく結果を考察してみると効率のよい雑誌としては雑誌7があげられ、ついで雑誌1がよいといえる。雑誌3は雑誌7と同じぐらいの平均であるが、雑誌2,6との差が見いだせず、大きな効果を示した号がある原因が分からない限り中程度の効率であると判断される。

9. お わ り に

本稿の目的は要求される機能を果たしていない従来の予備検定方式による多重比較を改め、有意水準の抑制と検出力の低下のバランスを取ることであったのだが、提案した方法は十分な効力があると考えられる。等分散性の検定は非心カイ二乗分布によるものとなり複雑になっているが、カイ二乗近似を用いるならば計算の手間は僅かに増大するだけと思われる。

また、不等分散のときにS-D手順を用いることは場合によっては有意水準自身を保持できない

図 13. 修正した広告料と資料請求件数

表 3. 広告効率を示す傾きと分散

	傾	き	分	散
雑 誌 1	31.13		106.4/19	
雑 誌 2	23.73		47.3/18	
雑 誌 3	40.73		900.7/19	
雑 誌 4	15.61		125.4/16	
雑 誌 5	15.50		78.8/11	
雑 誌 6	22.23		20.1/16	
雑 誌 7	38.85		37.8/16	

表 4. GH-C2 手順における検定統計量と限界値, および T-K 手順の結果

対	GH-C2 手順		T-K 手順
	検定統計量	限 界 値	
(1,2)	2.5798	3.2032	
(1,3)	1.3186	3.2595	
(1,4)	4.2338*	3.2189	棄却
(1,5)	4.3749*	3.3098	
(1,6)	3.3990*	3.2371	
(1,7)	2.7359	3.2152	
(2,3)	2.4034	3.2802	棄却
(2,4)	2.5100	3.2635	
(2,5)	2.6301	3.4005	
(2,6)	0.7611	3.2250	
(2,7)	6.7685*	3.2163	棄却
(3,4)	3.3797*	3.2501	棄却
(3,5)	3.4154*	3.2680	棄却
(3,6)	2.6520	3.2943	棄却
(3,7)	0.2665	3.2837	
(4,5)	0.0284	3.3085	
(4,6)	2.1953	3.3114	
(4,7)	7.2767*	3.2774	棄却
(5,6)	2.3193	3.4880	
(5,7)	7.5653*	3.4197	棄却
(6,7)	8.7368*	3.2488	棄却

ため避けるべきであると結論される。提案した GH-C2 手順を用いるほうがよい。

今後の課題は予備検定による場合分けで生じる結果の不連続性の回避であろう。T-K 手順と S-D 手順の使い分けよりも GH-C2 手順を用いたほうが差が小さいといえるが、どの程度予備検定の境界付近で結果が食い違うのかは分かっていない。それらを研究すること、また、総合的な方法を開発して連続な一方式に統一することを考えていきたい。

謝 辞 例におけるデータの提供, および有益な助言を賜った東京理科大学工学部の吉村 功教授, 貴重なコメントをいただいた広島大学原医研の大瀧 慈先生, ならびに丁寧なコメントをいただいた査読者各位に感謝致します。

参 考 文 献

- Dunnett, C.W. (1980): Pairwise multiple comparisons in the unequal variance case. *J. Amer. Statist. Assoc.* 75, 796-800.
- Games, P.A. and Howell, J.F. (1976): Pairwise multiple comparison procedures with unequal N 's and/or variances: A Monte Carlo study, *J. Educ. Statist.* 1, 113-125.
- Hochberg, Y. and Tamhane, A.C. (1987): *Multiple comparison procedures*. John Wiley, New York.
- Kendall, M.G. and Stuart, A. (1979): *The advanced theory of statistics*. Vol. 2. *Inference and relationship*, 4th ed, Charles Griffin, London.
- 永田 靖 (1992): 入門統計解析法. 日科技連.
- 松田眞一・永田 靖 (1990): 多重比較における新たな検出力の提案と各手法の特徴比較. 応用統計学 19, 93-113.
- 松田眞一 (1993): 不等分散一元配置における多重比較法. 南山経営研究 7, 397-413.
- Patnaik, P.B. (1949): The non-central χ^2 - and F -distributions and their applications. *Biometrika* 36, 202-232.
- 渡井 真 (1993): 広告の効果の評価法の研究, 東京理科大学工学部経営工学科卒業研究論文.

付 録

Bartlett 検定統計量の対立仮説の下での評価

一元配置モデルに対する確率密度は母数の与え方を少し変えると次式で表せる.

$$L(X | \mu, \sigma^2, r) = \prod_{i=1}^k \left[(2\pi\sigma^2 r_i)^{-n_i/2} \exp \left\{ -\frac{\sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2}{2\sigma^2 r_i} \right\} \right]$$

ただし, $X = \{x_{ij}\}$ で, 本文の母数との関連は $\sigma^2 = \sum n_i \sigma_i^2 / \sum n_i$, $r_i = \sigma_i^2 / \sigma^2$ であり, $\sum n_i r_i / \sum n_i = 1$ となる. 従って, 帰無仮説 $H_0: r_i = 1 (i=1, 2, \dots, k)$ に対する Bartlett 検定統計量の元となる尤度比検定統計量は

$$\ell = \frac{L(X | \bar{x}, v, 1)}{L(X | \bar{x}, v, t)}$$

となる. ただし, $\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$, $v_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / n_i$, $v = \sum_{i=1}^k n_i v_i / \sum n_i$, $t_i = v_i / v$, $1 = (1, \dots, 1)$ である.

先の式を代入して少し計算すると

$$-2 \log \ell = \sum_{i=1}^k n_i (t_i - 1 - \log t_i)$$

が得られる. 対数を展開すると

$$= \sum_{i=1}^k \left\{ \frac{n_i(t_i-1)^2}{2} - \frac{n_i(t_i-1)^3}{3} + \frac{n_i(t_i-1)^4}{4} - \dots \right\}$$

となり, 帰無仮説が真だと 3 次以降は漸近的に無視できる.

対立仮説が真の場合, 3 次以降が無視できるなら非心母数は,

$$\lambda = \sum_{i=1}^k \frac{n_i(r_i-1)^2}{2}$$

で与えられる. しかし, 固定した単純対立仮説に対してはその影響を無視できず, シミュレーションで誤差の増大が見られた. そこで対数を展開する前の式に戻って考えてみる. それはある単純対立仮説の母数を用いて次のように変形できる.

$$-2 \log \ell = \sum_{i=1}^k \{ n_i(\bar{t}_i - 1 - \log \bar{t}_i) - n_i \log r_i + n_i(t_i - \bar{t}_i) \}$$

ただし, $\bar{t}_i = v_i / (v r_i)$ である. 式の前半はその単純対立仮説を真としたときの対数尤度比統計量にほかならず, 漸近的期待値を一致させる非心母数が

$$\lambda = - \sum_{i=1}^k n_i \log r_i$$

で得られることが分かる. 本論で述べた非心母数は Bartlett の行った補正により n_i を ν_i に置き換えたものである. シミュレーションによるとこの非心母数を用いた近似は先のものよりかなり改善されていることが確認できた.

定理: Bartlett の検定統計量

$$B = \frac{\nu \log s^2 - \sum_{i=1}^k (\nu_i \log s_i^2)}{1 + \frac{\sum_{i=1}^k \nu_i^{-1} - \nu^{-1}}{3k-3}}$$

は, 対立仮説の下で自由度 $k-1$, 非心母数

$$\lambda = - \sum_{i=1}^k \nu_i \log (\sigma_i^2 / \sigma^2)$$

の非心カイ二乗分布で近似される. ただし, $s_i^2 = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / \nu_i$, $s^2 = \sum_{i=1}^k \nu_i s_i^2 / \nu$, $\sigma^2 = \sum \nu_i \sigma_i^2 / \sum \nu_i$ である.

(1994 年 4 月 12 日受付 1995 年 1 月 13 日最終修正)

著者連絡先: 〒 466 名古屋市昭和区山里町 18 南山大学経営学部情報管理学科