広島大学大学院教育学研究科紀要　第三部　第54号　2005　189－196

厳格化の観点からの多重比較法の整理

林　智幸・新見　直子

(2005年9月30日受理)

Tightening arrangement of multiple comparison tests

Tomoyuki Hayashi and Naoko Niimi

Multiple comparison tests are one of statistical tests that often used in amount of psychological studies. But there are too many kinds of multiple comparison tests to make proper use of the tests. In this article, we redefined multiple comparison tests for interval scale data as "tightening reduplicative *t*-tests". There are three lights for tightening *t*-tests; 1) level of significance, 2) distribution, and 3) statistic. Based on the each view many kinds of multiple comparison tests are developed. By setting information on each view for tightening *t*-tests before in this article, it will become possible to make proper use of multiple comparison tests.

Key words: multiple comparison tests on interval scale data, tightening reduplicative *t*-tests, statistical learning

キーワード：間隔尺度データにおける多重比較法，厳格化された反復 *t* 検定，統計学習

心理学が，実証的検証法を方法論として採用することによって，哲学から分離した以上，心理学研究に携わる者は，実証データを適切に分析する術を身につけなければならない。特に，数量的データを分析する方法論は「統計解析法」として，多くの心理学教育機関において必須の教授内容に盛り込まれている。いわゆる「心理統計法」などの授業では，心理学研究に必要な，使用頻度の高い，代表的な統計解析法を学習する機会が学生に与えられる。しかし，このような授業が開講されているにもかかわらず，必ずしも大学生の統計解析法（以下，統計法）の習得率は高くなく，残念ながら大学院生のみならず，心理学研究を行う現役の大学教員においてさえも充分な習得がされているとはいえない。

この習得率の低さは，自分にとって使用頻度の低い統計法を習得していないこと（量的な習得率の低さ）だけではなく，日頃から頻繁に使用している統計法を十分に理解していないこと（質的な習得率の低さ），質と量の両方において指摘できる。例えば，*t* 検定や分散分析及び多重比較法などは，心理学研究に携わる者にとって必須の統計法である。しかし，これらの統計法をデータに応じて適切に使いこなしていない者が大変多く，特に，多重比較法について顕著である。その原因として，多重比較法の解説書がこれまで充分に浸透していなかったことも挙げられるが（大変優れた解説書として永田・吉田（1997）がある），それらの優れた解説書が必ずしも統計法の使用者にとって「分かり易く」解説されていないことも大きな原因と筆者は考えている。

本稿は，統計法の使用者の立場から書かれた「多重比較法の使用手引き」となるものを目指した。ただし詳細な計算手順ではなく，計算手順の背景にある思想の解説を試みている。そのため，幾つかの数式を紹介しているが，数学的素養のない心理学者でも納得して多重比較法を使用できるような枠組みを提供することを目的とする。

多重比較法の意味

多重比較法とは，*t* 検定や分散分析と同様に，水準間の統計量（主に代表値）に有意な差があるかどうかを調べるための統計法である。例えば，中学生（a1），高校生（a2），大学生（a3），大学院生（a4）の四水準データが与えられたとしよう。*t* 検定とは，これら

の中から任意の二水準のみを抽出して，その他の水準情報を無視して，平均値の有意差を調べる統計法である（その他にも異なった使用法もあるが，心理学研究において「t 検定」という名称から一般的に連想されるのは二水準間の有意差検定としての使用法である）。分散分析とは，四つの水準全体の平均値に有意差があるかどうかを調べる統計法である。そして本稿で扱う多重比較法は，四水準データから，二水準間の比較ペアを複数設定して，複数ペアの平均値の有意差を同時に調べる統計法である。比較ペアの設定法として，多重比較法は一対比較型と対比型の二種類に分類される。一対比較型とは，(a1－a2)，(a1－a3)，(a1－a4）のように，水準データを単独で使ってペアを設定するタイプである。これに対して対比型とは，(大学受験前［a1＆a2］－大学受験後［a3＆a4］）のように，複数の水準データを組み合わせてペアを設定するタイプである。

これらの手法は本来互いに独立しているが，心理学研究の文脈において，多重比較法は，分散分析である要因の主効果が有意であった場合，その要因のどの水準間の代表値に有意差があるかを調べるため，下位検定の一種として使われている。このような従属的な使用法が定着したため，いずれも同じ代表値の有意差を調べる統計法であるにもかかわらず，二水準の有意差を調べる t 検定と，三水準以上の有意差を調べる多重比較法は全く異質な統計法であるという印象を与えることになった。

心理学以外の分野ではもはや多重比較法は分散分析と併用すべきではないという意見もある(永田・吉田，1997)。同データに分散分析と多重比較法を適用した場合に結果が一致しないことを考えると，両者は別物であると認識すべきである（ただし，現時点の心理学研究では，併用がマナーとなっているので，これに従うべきだろう）。多重比較法が分散分析の従属的な使用の枷から解放されたとすれば，(反復) t 検定と多重比較法をどのように使い分けるかを考えなければならない。

分散分析や多重比較法をまだ学習していない，t 検定のみを使える統計法の初心者がいると仮定する。この初心者に複数水準のデータを与え，各水準間のどこに有意差があるかを調べさせるとしよう。おそらくは，試行錯誤を繰り返しながら，この初心者はデータから複数ペアを設定して，t 検定を単純に繰り返し適用すると推測できる。実際，初期に開発された多重比較法の一種である Fisher の（無制限）LSD 法も同様の発想から開発されたものと推測できる。

しかし，このような「単純な反復 t 検定」には問題がある（以下，3変数 u1，u2，u3を例とする）。3変数から全ての比較ペアを設定すると，u1－u2（u1とu2の比較ペア），u2－u3，u3－u1のように3ペアとなる。u1－u2ペアの有意差を t 検定で調べる場合，その有意性を有意水準（以下「α」とする）＝5％として判断することにしよう。α とは，ごく大雑把に言えば「結論が間違っている可能性」となり，$(1-\alpha)$ は「結論の信頼性」と言えるだろう[1]。同様に，u2－u3ペア，u3－u1ペアに $\alpha=5\%$ の t 検定を適用することになる。今，3ペアの有意性検定が同時に成立することが重要となり，確率論的には「同時に」はかけ算で表現される。結論の信頼性は，3ペア同時に対しては，「95（＝100－5）％×95％×95％≒85.7％」と低下してしまう。このように単純な反復 t 検定は，結論の信頼性を低下させてしまうため，三水準以上の有意差の検討には用いてはならず，別の統計法である多重比較法を使用することが望ましいと考えられる。

ただし，単純な反復 t 検定の問題点が「結論の信頼性の低下」であるならば，低下させない工夫を行えばよい。結論の信頼性の低下は，個々のペアの有意水準を5％としたことに由来する。ならば全体の結論の信頼性の低下を極力抑えるために，個々のペアの有意水準をより厳しくすればよい。例えば $\alpha=1\%$ とした場合，「99（＝100－1）％×99％×99％≒97.0％」と，全体の結論の信頼性が妥当な水準に収まる。このように，反復 t 検定を単純に使用するには問題があるが，α を厳しく調整することで対応できる。

有意水準を厳しく調整する方法は発想としては非常に単純であり，初心者にとっても理解がしやすいだろう。実は，有意水準を調整するという発想が多重比較法の原点となっており，Bonferroni 法と呼ばれる多重比較法は有意水準を（直接）厳しくする方法として最も基本的な手法である。

Bonferroni 法による多重比較法の手順　1）比較するペアを設定，そのペア数を a とする。2）全体の α を設定する。多くの場合 $\alpha=5\%$ が採用される。3）各ペアの有意性検定について，「調整された有意水準」あるいは「名義水準（α'）」を設定して基準とする。Bonferroni 法では，一律に同一基準を適用するが，式（1）から α' を算出する。4）個々のペアに対して t 検定を行い，確率を求める。5）「算出された確率＜α'」であればそのペアに有意差ありと判定する。

$$\alpha'=\frac{\alpha}{a} \quad \cdots\cdots \quad (1)$$

多くの統計法の教科書では，Bonferroni 法のような有意水準を調整するタイプの多重比較法が説明され

ていないか，説明が後回しにされている。しかし，Bonferroni法を最初の多重比較の手法として説明した方が，学習者にとっても，教育者にとっても論理展開が明瞭である。このタイプの多重比較法を最初に学習することが，多重比較法学習の基盤となるだろう。

以上の説明からわかるように，t検定と多重比較法とは非常に関連が深く，間隔尺度における多重比較法とは，単純な反復t検定ではなく，繰り返しにおけるt検定を厳格化した，厳格化された反復t検定と定義できる。

多重比較法を整理する視点

前節で，多重比較法とは「厳格化された反復t検定」であることを説明した。以下，多重比較法を整理する視点を述べる。すなわち，①厳格化の方法はどのようなものか，前述の有意水準の厳格化以外にはないのか，②対応なし要因，対応あり要因によって多重比較法の計算手順は異なるのか，③単独の変数の比較だけではなく，複数の変数の合成変数の比較もできないだろうか，である。

①厳格化の方法　二水準の平均値の差を調べるためにt検定を用いる場合，二水準のズレの大きさを意味するt値（統計量）が，t分布において，どのような有意水準の下で，有意差があるかを判断する。すなわちt検定においては，(1) 統計量，(2) 分布，(3) 有意水準の情報が重要な要素となる。多重比較法を厳格化された反復t検定と定義するならば，これら3つの観点から，厳格化ができることが推測できよう。

(1) 統計量の観点から厳格化を行う多重比較法（以下「統計量調整型多重比較法」）の基本はScheffe法である。Scheffe法では，統計量（例えばt値）を「変数−1」で割ることによって，厳格化を行う。本来Scheffe法はF分布を使って有意性を検定するが，F分布＝(t分布)^{2}であるため，t分布を使って検定を行っても良い。

(2) 分布の観点から厳格化を行う多重比較法（「分布調整型多重比較法」）の基本はTukey HSD法である。t検定では「(student化された) t分布」を用いて有意判定を行うが，Tukey法では，t分布の代わりに「student化された範囲qの分布」を用いる。

(3) 有意水準の観点から厳格化を行う多重比較法（「有意水準調整型多重比較法」）の基本は，既に紹介したBonferroni法である。全体の有意水準を「比較の総ペア数」で割ることによって厳格化を行う。

上記のScheffe法，Tukey HSD法，Bonferroni法は多重比較の基本的な手法であり，最も厳しい基準を全ての比較ペアに適用する「シングルステップ法」に分類される手法である（永田・吉田，1997)。多重比較には基本的なシングルステップ法の他にも，比較ペアの特性に応じて基準の厳しさを調整する「ステップダウン法」が開発されている。先に挙げた，(2) 分布調整型多重比較法，(3) 有意水準調整型多重比較法についてはステップダウン法の手法が開発されており，次節以降にて解説を行う。

②対応なし／ありの観点　多くの統計解析法では，統計量を算出するための計算手順が，データ対応の有無により異なっているが，多重比較法でも同様に「対応なし／あり」によって使用される手法を区別しなければならない。「対応なし／あり」は，計算式においては，主として水準間の相関の有無によって表される。「対応なし」とは水準間の無相関性を意味しており，「対応あり」の特殊形と位置づけることができる。

③一対比較／対比の視点　多重比較法とは，その名の通り「比較」を行うものだが，比較の種類は二つに大別できる（森・吉田，1990)。一つは，複数変数からにに選択された二変数の代表値（平均値）を比較する「一対比較 (pair comparisons)」であり，多くの多重比較法はこの一対比較型の比較を行っている。もう一つは，一対比較のように選択された変数を単純に使うのではなく，分析者によって自在に組み合わされた複合変数の比較を行う「対比 (contrasts)」である。本稿では一対比較型を対比型の特殊形と位置づける。

有意水準調整型多重比較法の発展

原理が最も理解しやすく，また応用範囲の高いものが，有意水準調整型多重比較法である。多重比較法を使う場合，有意差を，偶然誤差としてではなく，正しく有意差と判断する「検定力」を考慮する必要がある。有意水準調整型多重比較法の基本がBonferroni法であるが，比較ペアの数によっては検定力が顕著に低下してしまうので，検定力を改善した手法が多数開発されている。ここでは，Sidak法とRyan法とHolm法を紹介する。

Sidak法もBonferroni法と同様に，一律の厳しい名義水準α'を全ての比較ペアに適用するシングルステップ法に分類される。名義水準は以下の公式で計算する。具体的手順は，比較する総ペア数をkと定め，下記の式 (2-1) に代入する。

$$\alpha' = 1 - (1 - \alpha)^{\left(\frac{1}{k}\right)} \quad \cdots\cdots \quad (2\text{-}1)$$

Sidak法で統制された名義水準α'はk数のペアに

対して適用されるので，全体では，式（2-2）のようにαとして復元される。Bonferroni法が全ペアの有意水準を全体有意水準（α）を復元しないのに対して，Sidak法は完全な復元を行う（Table 1）。

$$
\begin{aligned}
\alpha' &= 1-(1-\alpha)^{\left(\frac{1}{k}\times k\right)} \\
&= 1-(1-\alpha)^{1} \\
&= 1-1+\alpha \\
&= \alpha \qquad \cdots\cdots \ (2\text{-}2)
\end{aligned}
$$

Table 1 各ペア数における全体有意水準（α）=5%の場合の個別ペア及び復元された有意水準

		Bonferroni		Sidak	
変数の数	全ペア数	個別ペア	全ペア	個別ペア	全ペア
2	1	5.0000%	5.0000%	5.0000%	5.0000%
3	3	1.6667%	4.9171%	1.6952%	5.0000%
4	6	0.8333%	4.8970%	0.8512%	5.0000%
5	10	0.5000%	4.8890%	0.5116%	5.0000%
6	15	0.3333%	4.8850%	0.3414%	5.0000%
7	21	0.2381%	4.8827%	0.2440%	5.0000%
8	28	0.1786%	4.8813%	0.1830%	5.0000%
9	36	0.1389%	4.8804%	0.1424%	5.0000%
10	45	0.1111%	4.8797%	0.1139%	5.0000%

Ryan法は，一対比較型の比較の場合に使われる方法で，「順位化された平均値の差の距離」の概念を用いる（「分布調整型多重比較法の発展」参照，あるいは森・吉田(1990)参照）。変数の数をa，距離をステップ数rとして表し，各ペアのステップ数に応じて，式(3)を使って名義水準α'を計算，ステップ数が2になるまで比較を進めていく。なお，最も厳しい基準はr＝aの場合のα'だが，この数値はBonferroni法のα'と同値である。

$$\alpha' = \frac{2\times\alpha}{a\times(r-1)} \quad \cdots\cdots \ (3)$$

Holm法は，ペアの確率の小さい順番に厳しい基準を適用していく方法であり，ペアごとに名義水準が異なる。この方法は，Bonferroni法やSidak法同様，どんなタイプの多重比較法でも適用できる。具体的な手順は以下の通り。最初に比較する総ペア数kを設定する。各ペアの検定統計量に対応する確率を求め，確率の小さい順に並び替えたものを，P(i)とする（1番小さいものをP(1)，2番目に小さいものをP(2)）。P（i）に対応するペアに適用する名義水準α(i)'を以下の式（4)で計算する。最も厳しい基準はi＝1の場合だが，この時のα'はBonferroni法のα'と一致する。

$$\alpha' = \frac{\alpha}{k-i+1} \quad \cdots\cdots \ (4)$$
$$(1 \leqq i \leqq k)$$

なお，SPSSに搭載されている「TamhaneのT2」とは，Sidak式の調整された有意水準を用いたWelch検定を意味する（小野田・山本，2004）。

分布調整型多重比較法の発展

分布調整型多重比較法とは，t分布の代わりに，多重性が考慮された別の確率分布を用いて検定を行う手法である。代表的な確率分布が「順位化された平均値の差の距離」概念を導入した「student化された範囲q分布」であり，このタイプの最も基本的な手法が既に説明したTukey HSD法である。しかし，この方法は以下のような制約を持つ。その特徴とは，①等分散であること，②データ数が等しいこと，③最も厳しい基準を全比較ペアに一律に適用すること。Tukey型の多重比較法は多数開発されたが，それらはこれらの欠点を改善していく中で生まれていった。しかし，類書には明確には記載されていないが，最重要な特徴として，④Tukey型多重比較法は本来「対応なし」データ用の多重比較であることは忘れてはならないだろう。

Student化された範囲q分布　Tukey型多重比較法を支える「Student化された範囲q分布」とは，「確率変数z1，z2，…，zaが互いに独立に標準正規分布N（0,1^2）に従い，それらとは独立に確率変数$\chi^2(\phi)$が自由度ϕのχ^2分布に従っているとする。このとき，確率変数Qの確率分布を群の数a，自由度ϕのStudent化された範囲qの分布」と呼ぶ（一石，2004）。すなわちq分布とは，複数の独立した（対応のない）z分布と，χ^2分布から作成されたものである。また，χ^2分布も独立した複数のz分布と関連がある。χ^2分布とは，標準正規z分布する確率変数の二乗の変数がどんな分布をするだろうか，という動機がχ^2分布作成の契機となっており，z分布から，独立して抽出したn個のxi（i＝1，2，…，N）から作った，$\Sigma(x_i)^2$の分布は，自由度ϕのχ^2分布に従う（一石，2004）。

$$Q(a,\ \phi) = \max \frac{|Z_i - Z_j|}{\sqrt{\left(\frac{\chi^2(\phi)}{\phi}\right)}} \quad \cdots\cdots \ (5)$$

式(5)のように，q分布は，分母分子共々に，複数の独立した・標準正規分布をベースに作成していることから，①適用するデータは基本的に正規分布であること，②データは独立（対応なし）であるという制約がある。上記の①と②の前提はq分布，χ^2分布のみならず，t分布やF分布にも当てはまる。そのため，t検定や分散分析（F分布を利用）の前提にも先程の①

や②が含まれることになる。確かに，ある程度の補足仮定（「水準間の相関が等しい」などの球面性の仮定）を置くことで「対応あり」データにも適用することができ，分散分析も対応ありデータにも適用できるような修正式が教科書に解説されている。しかし，本来は「対応なし」であることを留意すべきだろう。特に q 分布は，t 分布や F 分布に比べて，複数の独立した z 分布の利用度が t 分布や F 分布に比べて高いため，q 分布を対応あり要因に使うのは現在の統計学では許容されないかもしれない。例えば，最も信頼度の高い統計ソフト SPSS において，対応あり要因の分散分析を「反復測定」プロシジャーで実行した場合，「対応なし」要因の分析の場合に比べて，多重比較法の種類が少ない（Tukey 型多重比較法が使用不可）のは上記の理由によると推測できる。

分散等質を仮定する Tukey 型　分散が等質の場合の Tukey 型多重比較法の基本が既述の Tukey HSD 法である。この発展型として Tukey-Kramer 法，Student-Newman-Keuls 法（SNK 法），Tukey WSD 法，Tukey-Welsch（Q）法が挙げられる（永田・吉田，1997）。

Tukey HSD 法は2水準の標本サイズ n が等しい場合にのみ使えるとされたが，現在では n が非等質の場合にも使えるように拡張されている（前者と区別する場合には拡張版を Tukey-Kramer 法と呼ぶ）。一律に厳しい基準を適用するシングルステップの Tukey HSD 法から，比較するペアの特性に応じて基準を適切に緩和していくステップダウン的な Tukey 法が開発されている。最も単純なものが，ステップ数（比較ペアの距離）に応じて q 分布の臨界値を使い分ける手法としてかつて使用されてきたものが SNK 法であるが，現在ではこの方法は認められていない。SNK 法を改良したものが Tukey WSD 法や Tukey Welsch（Q）法だが，前者は「厳しすぎる」Tukey HSD 法と「やや甘い」SNK 法の平均化を図った方法であり，後者はステップ数応じて Sidak の不等式に有意水準を調整した q 分布の臨界値を使う方法である。

分散非等質を仮定する Tukey 型　分散が非等質の場合は，統計量としては通常の対応なし t 検定の代わりに Welch 検定を用いる。両者の大きな違いは，後者が前者よりも自由度が厳しく調整されていることである。具体的な手法としては，Games-Howell 法，Dunnett の C 法，Games-Howell-C 法が挙げられる（小野田・山本，2004）。Games-Howell 法は，Welch 検定により算出された統計量と，より厳しくされた自由度を使った q 分布の臨界値にて有意性を判定する手法である。ただし，この手法では厳格化が充分でないために，更に厳しく調整された自由度を用いるものが Dunnett の C 法である。また，後者の手法は厳しすぎるために，両者を平均化したものとして Games-Howell-C 法が開発されている。

なお，分布調整型の多重比較法には，「Student 化された範囲 q 分布」だけではなく，その他の分布を用いた検定法も開発されている。「Student 化された最大絶対値」分布を用いる多重比較法としては，分散が等質の場合には Hochberg の GT2法，Gabriel 法，分散が非等質の場合には Dunnett の T3法が開発されている。また，対照群と各郡との比較を行う Dunnett 法や，対照群と各郡の単調増減の傾向性を比較する Williams 法なども特殊な分布を用いる（小野田・山本，2004）。

統計量の計算式

多重比較法が「厳格化された反復 t 検定」であるならば，当然のことながら比較ペアごとに統計量が計算されるはずである。本節では，多重比較法における統計量の計算式を解説する。

多重比較法の統計量計算式はこれから解説するように色々な種類があるが，①対応なし版一対比較式，②対応あり版一対比較式，③対応なし版対比式，④対応あり版対比式の四種類に大別される。データの種類（対応の有無），多重比較法の目的（一対比較型か対比型か）に応じて，各種の統計量計算式を使えばよい。

しかし，既に述べたように，対応なし版の計算式は対応あり版の計算式の特殊形であり，また一対比較の計算式は対比の計算式の特殊版と見なすことができる。ここから，基本的には対応あり版対比式の式（6）を汎用式と位置づけることができる。しかし，式（6）は多くの情報を必要とし，計算も煩雑になるので，実際の計算では幾つかの仮定を設けた簡略化された計算式が用意されている。

本来一種類の複雑な統計量計算式から複数の簡略版計算式を導くためには，式の中に含まれる変数に仮定を置く必要がある。その仮定とは，対応なしの場合では「複数変数の分散（標準偏差）が等しい（分散の等質性）」，

$$t = \frac{\sum_{i}^{a} (CiX'i)}{\sqrt{\sum_{i}^{a} \sum_{j}^{Ni} (CiSi \times CjSj \times Rij \div N'ij)}} \quad \cdots\cdots \quad (6)$$

Ci, Cj … 対比計算における重み係数
X'i …… i水準におけるXの平均値
N'ij …… 水準iの人数Nと水準jの人数Nの調和平均
Si, Sj … i, j水準の標準偏差

対応ありの場合では，分散の等質性に加えて，「各変数間の相関が等しい（相関の等質性）」である。この仮定の根拠は，多重比較法を分散分析と併用する場合，分析対象データには分散分析の計算上の仮定が必要となるためである。分散分析の分析対象データには先に挙げた「分散の等質性」「相関の等質性」が必要となる。もしあるデータに分散分析を使ったならば，当然，そのデータは「分散／相関の等質性」を満たしていると仮定することになる。このような分散分析における仮定は「球面性の仮定」と呼ばれ，この仮定から導かれた分散（分散分析から導かれる誤差分散）を利用した簡略版統計量計算式が開発されている（以下分散分析から導かれた分散を「誤差分散」と呼ぶ）。

①対応なし版一対比比較式　式（6）を二水準版に限定した計算式を用いるが，代入する標準偏差（＝√分散）の種類によって計算式が異なる。分散の等質性を仮定するならば式（7）を利用することになるが，この式は計算が容易であるため，多くの多重比較法の場合に使われる。誤差分散を使う場合は，誤差分散の自由度を使用する。

分散の非等質を仮定すれば各変数の分散情報を利用した式（8）を用いる。分散がそれほど大きく異なっていなければ統計量は式（8）から計算し，自由度は（N1+N2−2）とする。分散が大きく異なっているならばWelchの自由度調整式により計算された値を使う（具体的な計算式は森・吉田（1990）を参照）。なお，森・吉田（1990）から，式（7）は分散を等質と仮定した場合の対応なし *t* 検定，式（8）はWelch修正による式であることが分かる。

$$t = \frac{X'1-X'2}{\sqrt{V\times\left(\frac{1}{N1}+\frac{1}{N2}\right)}} \quad \cdots\cdots (7)$$

X'i, Ni …… i水準におけるXの平均値，人数N
V … 分散分析から導かれた誤差分散

$$t = \frac{X'1-X'2}{\sqrt{\left(\frac{V1}{N1}+\frac{V2}{N2}\right)}} \quad \cdots\cdots (8)$$

X'i, Ni, Vi … i水準におけるXの平均値，人数N，分散V

②対応あり版一対比較式　対応あり版一対比較式も代入する分散情報の種類により計算式が異なる。対応あり版多重比較法では，対応なし版に比べて分散が小さくなる。分散の等質性を仮定すれば，二水準間の相関情報を利用して分散を式（9）の分母のように小さくすることができる。もし，対応あり分散分析から導かれた誤差分散を利用することができれば，その誤差分散は既に相関情報を利用して小さくなった分散であるため，式（7）を用いることができる（自由度は誤差分散の自由度とする）。分散を非等質だと仮定すれば，式（10）を用いることになる（自由度はN−1とする）。なお，式（10）は，森・吉田（1990）から，対応あり *t* 検定の計算式であることが分かる。

$$t = \frac{X'1-X'2}{\sqrt{V\times\left(\frac{1}{N1}+\frac{1}{N2}\right)\times(1-R)}} \quad \cdots\cdots (9)$$

X'i, Ni …… i水準におけるXの平均値，人数N
V … 等分散と仮定した場合の分散情報
R … 二変数の相関係数

$$t = \frac{X'1-X'2}{\sqrt{\left(\frac{S1^2+S2^2-2\times R\times S1\times S2}{N}\right)}} \quad \cdots\cdots (10)$$

X'i, Si …… i水準におけるXの平均値，標準偏差S
R … 二変数の相関係数
N … 人数

③対応なし版対比式　分散分析の誤差分散を利用するなど分散を等質だと仮定する場合は式（11）を用いる（自由度は誤差分散の自由度を参照）。分散を非等質だと仮定する場合は式（12）を用いる（自由度はΣ（Ni−1）とする）。なお，森・吉田（1990）をはじめとする多くの教科書に解説されている対比計算式は式（11）である。

$$t = \frac{\sum_{i}^{a}(CiX'i)}{\sqrt{V\times\sum_{i}^{a}\left(\frac{Ci}{Ni}\right)}} \quad \cdots\cdots (11)$$

Ci … 対比計算における重み係数
X'i, Ni …… i水準におけるXの平均値，人数N
V … 誤差分散

$$t = \frac{\sum_{i}^{a}(CiX'i)}{\sqrt{\sum_{i}^{a}\left(\frac{Ci\times Vi}{Ni}\right)}} \quad \cdots\cdots (12)$$

Ci … 対比計算における重み係数
X'i, Ni, Vi …… i水準におけるXの平均値，人数N，分散V

④対応あり版対比式　対比式も一対比較式と同様に，相関情報を利用して分散を小さくする必要がある。等分散かつ各水準間の相関も等質と仮定する場合に

Table 2 例題データ

データ	a1	a2	a3	a4
	9	6	10	9
	7	5	13	11
	8	6	8	13
	8	3	13	14
	12	6	12	16
	11	7	14	12
	8	10	14	15
	13	9	16	14
平均値	9.5	6.5	12.5	13.0
標準偏差	2.20	2.20	2.51	2.27

Table 3 各種の誤差分散と自由度

誤差分散の種類	誤差項	自由度
対応なし分散分析の誤差分散	5.286	28
対応あり分散分析の誤差分散		
球面性の仮定	3.381	21
Greenhouse-Geisser	3.439	20.6
Huynh-Feldt	3.381	21

Table 4 対応なしデータとしての処理した場合の統計量

比較ペア	誤差分散に基づく統計量	個別の分散情報に基づく統計量	個別の分散情報に基づくWelch統計量
a1 vs a2	$t(28)=2.610, p=0.0144$	$t(14)=2.722, p=0.0165$	$t(14.00)=2.722, p=0.0165$
a1 vs a3	$t(28)=2.610, p=0.0144$	$t(14)=2.542, p=0.0235$	$t(13.77)=2.542, p=0.0246$
a1 vs a4	$t(28)=3.045, p=0.0050$	$t(14)=3.130, p=0.0074$	$t(13.99)=3.130, p=0.0080$
a2 vs a3	$t(28)=5.219, p=0.0000$	$t(14)=5.084, p=0.0002$	$t(13.77)=5.084, p=0.0002$
a2 vs a4	$t(28)=5.654, p=0.0000$	$t(14)=5.814, p=0.0000$	$t(13.99)=5.814, p=0.0001$
a3 vs a4	$t(28)=0.435, p=0.6669$	$t(14)=0.418, p=0.6820$	$t(13.86)=0.418, p=0.6825$

Table 5 対応ありデータとしての処理した場合の統計量

比較ペア	球面性の仮定に基づく統計量	Greenhouse-Geisser修正の統計量	個別の分散情報に基づく統計量
a1 vs a2	$t(21)=3.263, p=0.0037$	$t(20.65)=3.235, p=0.0041$	$t(7)=3.464, p=0.0105$
a1 vs a3	$t(21)=3.263, p=0.0037$	$t(20.65)=3.235, p=0.0041$	$t(7)=3.384, p=0.0117$
a1 vs a4	$t(21)=3.807, p=0.0010$	$t(20.65)=3.775, p=0.0012$	$t(7)=3.862, p=0.0062$
a2 vs a3	$t(21)=6.526, p=0.0000$	$t(20.65)=6.471, p=0.0000$	$t(7)=6.620, p=0.0003$
a2 vs a4	$t(21)=7.070, p=0.0000$	$t(20.65)=7.010, p=0.0000$	$t(7)=6.745, p=0.0003$
a3 vs a4	$t(21)=0.544, p=0.5923$	$t(20.65)=0.539, p=0.5957$	$t(7)=0.509, p=0.6263$

Table 6 各種の有意水準調整型多重比較法における名義水準 α'

比較ペア	step数[1]	確率順位[1]	Bonferroni	Sidak	Ryan	Holm
a1 vs a2	2	4	0.0083	0.0085	0.0250	0.0167
a1 vs a3	2	5	0.0083	0.0085	0.0250	0.0250
a1 vs a4	3	3	0.0083	0.0085	0.0125	0.0125
a2 vs a3	3	2	0.0083	0.0085	0.0125	0.0100
a2 vs a4	4	1	0.0083	0.0085	0.0083	0.0083
a3 vs a4	2	6	0.0083	0.0085	0.0250	0.0500

1：step数，確率順位は，対応なし版及び対応あり版の「個別の分散情報に基づく統計量」を参考にしたもの。
今回のデータでは，両者から導かれたstep数，確率順位は偶然一致している。

は，対応あり分散分析から導かれた誤差分散を利用して式（11）を用いるする方法がある。分散分析を利用しない場合には計算手順が幾つかある。その場合，分散と相関の等質性を仮定する場合は式（13）を用いると良い。また，相関の等質性を仮定するが，分散を非等質と仮定する場合は，式（14）を用いる。そして，相関も分散（標準偏差）も非等質であるとするならば式（6）を用いることになる。

式（13）と式（14）を用いる場合には，球面性の仮定が守られている場合には各種の「誤差分散」を用い，自由度は「対応あり版誤差分散の自由度」を用いる。仮定が守られていない場合は，Greenhouse-Geisserの修正式，あるいはHuynh-Feldtの修正式による誤差分散，自由度を用いる（南風原，2002）。最も厳密に計算を行いたい場合は，式（6）を用いた計算を行う方がよい。式（6）を用いる場合には，自由度は（N－1）となる。なお，対応あり版対比式は森・吉田（1990）をはじめとする多くの教科書には説明されていない。式（13）は，岩原（1965）に説明されている計算式だが，その他の式はほとんど解説されていない。

$$t = \frac{\sum_{i}^{a}(CiX'i)}{\sqrt{V \times \sum_{i}^{a}\left(\frac{Ci}{Ni}\right) \times (1-R)}} \quad \cdots\cdots (13)$$

Ci … 対比計算における重み係数
X'i, Ni …… i水準におけるXの平均値，人数N
V … 誤差分散
R … 各水準間の相関（等質だと仮定）

$$t = \frac{\sum_{i}^{a}(CiX'i)}{\sqrt{\sum_{i}^{a}\left(\frac{Ci \times Vi}{Ni}\right) \times (1-R)}} \quad \cdots\cdots (14)$$

Ci … 対比計算における重み係数
X'i, Ni, Vi …… i水準におけるXの平均値，人数N，分散V
R … 各水準間の相関（等質だと仮定）

計算例

最後に，一対比較型の多重比較法を手計算で行う場合の手順を確認する。例題データをTable 2に示す。

1）どのような比較ペアを設定するかを考える。一対比較型か，対比型かによって統計量計算式の種類が異なるので，目的に応じて適切な計算式を使う。有意水準調整型多重比較法を使う場合，比較ペア数が重要となる。

統計量計算式には，関連する水準データ全体の情報から導かれた誤差分散を使うものがある。その誤差分散にも種類があるが，代表的なものをTable 3に示した。統計量の計算例として，対応なしの場合については，（a1）対応なし分散分析の誤差分散，（a2）個別の分散情報（分散等質），（a3）個別の分散情報（分散非等質）を使った場合の t 値及び確率を示している。対応ありの場合については，（b1）球面性の仮定に基づく誤差分散，（b2）Greenhouse-Geisser修正を行った誤差分散，（b3）個別の分散情報を使った場合の t 値及び確率を示している（Tabel 4, Table 5）。

2）どのような厳格化を使った多重比較法を用いるかを選択する。一対比較型の比較においては，対応なしの場合はTukey（分布調整）型多重比較法，対応ありの場合は有意水準調整型多重比較法を用いた方がよく，検定力の観点からはScheffe（統計量調整）型多重比較法は推奨しない。対比型の比較においては，Scheffe法か有意水準調整型を用いる。有意水準調整型は比較ペアの数に応じて検定力が低下するので，比較ペアが多いようであればScheffe法を用いた方がよいだろう。対比型においてTukey型多重比較法を用いる場合は，シングルステップ法であるTukey HSD法を使う。

Table 6において，（a2）及び（b3）の統計量及び確率を使った，有意水準調整型多重比較法の名義水準 α' の値を計算している。参考にしてもらいたい。

本稿では，多重比較に焦点を当てて，統計解析法の計算手順の背景にある「計算原理（思想）」の解説を行った。多くの統計法の教科書には，具体的な計算手順は解説されているものの，「どうしてそのような計算を行うのか」という計算原理が解説されていない。心理学研究に携わる者には，統計法が難しいと嘆く者も多いが，このような計算原理を充分理解することが必要となるだろう。また，統計教育・学習においても，機械的に計算手順を教育学習するのではなく。計算原理に力点を置いた教授学習法が必要になると思われる。

【注】

1）正確には結果の信頼性は，「α：本当は差がないのに差があると判断する誤り（第一種の過誤）」と「β：本当は差があるのに差がないと判断する誤り（第二種の過誤）」のを総合的に判断しなければならないが，ここでは，初心者にも分かり易いように，第一種の過誤のみを取り上げて簡略化して説明をしている。

【引用文献】

南風原朝和　(2002)．心理統計学の基礎　総合的理解のために　有斐閣
一石　賢　(2004)．道具としての統計解析　日本実業出版社
岩原信九郎　(1965)．新訂版　教育と心理のための推測学　日本文化科学社
森　敏昭・吉田寿夫　(1990)．心理学のためのデータ解析テクニカルブック　北大路書房
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
小野寺孝義・山本嘉一郎　(2004)．SPSS事典BASE編　ナカニシヤ出版