

## 順序尺度データにおける多様な多重比較法

林 智 幸

(2005年9月30日受理)

A variety of multiple comparison tests for ordinal scale data

Tomoyuki Hayashi

We often treat some of ordinal scale data as interval scale data, because we don't know enough how to test on ordinal scale data. We had better learn it. In this article for focusing on multiple comparison tests, on the basis of Hayashi(in press)'s framework("tightening reduplicative tests"), we redefined multiple comparison tests for ordinal scale data as, "tightening reduplicative Mann-Whitney tests" in between-subjects data, or "tightening reduplicative signed tests" in within-subjects data. And, using ranking and central limit theorem together, Tukey method or Scheffe method are valid to use on ordinal scale data.

Key words: multiple comparison tests on ordinal scale data, tightening reduplicative tests, statistical learning  
キーワード：順序尺度データにおける多重比較法, 厳格化された反復検定, 統計学習

心理学研究において、多重比較法は分散分析同様に頻繁に使用される統計解析法（以下、統計法）であり、両者は併用されることが多い。分散分析が間隔尺度データの統計法であることから、多くの者は、併用される多重比較法も間隔尺度データのみの統計法だと誤解している。しかし、多重比較法とは、複数集団の統計量を同時に比較する場合に用いられる分析手法である（森・吉田，1990）。この場合の統計量には、平均値を始めとして、順位和あるいは順位平均値、中央値や比、あるいは分散、相関係数などが含まれる。すなわち、多重比較法とは、間隔尺度データの代表値である平均値のみではなく、様々な統計量の同時比較を行う統計法を意味する。

ところで、心理学研究におけるデータは必ずしも間隔尺度版統計法が求める前提を満たしていない場合が多い。それにもかかわらず、多くの場合、分散分析などの間隔尺度版統計法を使っている。かつては、適切な分析手法が開発されていなかったという理由から、間隔尺度データでなくとも、とりあえず間隔尺度版統計法が使われていた（このような方法論の妥当性は、データが一定の条件を満たしていれば、後述する中心極限定理により保証される）。しかし、現代の統計学では、例えば順序尺度データを分析する場合には、専用の順序尺度版統計法を用いるべきだという共通認識

が、統計学者の間では、されるようになった。

しかし残念ながら、実際に研究で統計法を使う心理学者の間で、このような認識が普及しているとは言い難い。その理由には、非間隔尺度データ用の統計法の種類がまだ不十分であること、開発されていてもその統計法が十分に理解されていないことが挙げられる。実際、心理学研究で頻繁に使用されている統計ソフトSPSSにおいてさえも、順序尺度版多重比較法が使えない（小野寺・山本，2004）。

本稿では順序尺度データにおける多重比較法の使用手引きの解説を行うことで、その普及を目的としている。ただし、順序尺度版多重比較法は、間隔尺度版多重比較法に基礎を置いた統計法であり、本稿も、林（印刷中）で解説されている間隔尺度版多重比較法の知識に基づいて説明を行っている。

### 間隔尺度版多重比較法の基礎知識

本節では、順序尺度版多重比較法の基礎となる間隔尺度版多重比較法について厳格化の観点からの概説を行う。詳細は林（印刷中）を参照。

多重比較法とは、複数水準のデータが与えられたときに、どの水準間の統計量（多くの場合は代表値であり、間隔尺度データの代表値は平均値）に有意差があ

るかを統計的に判断するための検定法である。間隔尺度データにおいて、二水準の平均値の有意差を調べる場合に「 $t$ 検定」が使われるが、有意差を判断するために、 $t$ 値という統計量、 $t$ 分布、自由度、有意水準などの情報を用いる。多重比較法とは、厳格化された反復 $t$ 検定と定義することができるが、厳格化の種類で多重比較法を整理できる。すなわち、①統計量の観点から厳格化を行う統計量調整型多重比較法（Scheffe法など）、②分布の観点から厳格化を行う分布調整型多重比較法（Tukey HSD法など）、③有意水準の観点から厳格化を行う有意水準調整型多重比較法（Bonferroni法など）である。

多重比較法をどのような水準間比較を行うかによって、一対比較型と対比型とに大別できる（永田・吉田, 1997）。一対比較型は更に、①全ての水準間の比較（全比較）、②対照水準と各水準との比較（対照比較）、③単調性を想定した比較（傾向比較）とに小分類されるが、いずれにしろ選別した水準を単純に比較するタイプの多重比較法である。これに対して、対比型とは、複数の水準データを分析者が自在に組み合わせた複合水準データの比較ペアの有意差を調べる統計法である。

## 順位化の利点

心理学研究では、順序尺度版統計法は、間隔尺度版統計法の代用として用いられることが多い。順序尺度版統計法の計算手順を調べると、最初の段階で「データの順位化」が行われる点が特徴的である。岩原(1965)によると順位化には種類があるが、平均順位法が一般的である。平均順位法とは、例えば（8, 3, 3, 3, 2）のデータにおいて同順位3の処理として、（2位+3位+4位） $\div$ 3 = 3位のように平均順位を与えて、（1位, 3位, 3位, 3位, 5位）として処理する方法である。

順位化の利点を相関係数を例にして説明する。相関係数には幾つかの種類があるが、最も代表的なものがピアソンの積率相関係数である。二変数間の直線的関係を示すピアソンの積率相関係数だが、少数の外れ値によって大きく影響を受けるとされる（森・吉田, 1990）。外れ値への対処法として、多くの教科書では、外れ値の影響をほとんど受けないスピアマンの順位相関係数を利用することを推奨している。

この二種類の相関係数には密接な関係がある。森・吉田(1990)によると、スピアマンの順位相関係数とは、それぞれの変数における各ケースの順位を得点としてピアソンの積率相関係数を算出することによって求められる。すなわち、両者の違いである「データ

の順位化」こそが外れ値を持つデータに対して頑健性を付与する。

現在、順序尺度版統計法は、間隔尺度版統計法に比べてその種類が充分ではない。そのため、分析の目的に応じて、間隔尺度統計法の代用という措置が執られているが、最も簡便な対処法は、「データの順位化」を行った「順位化データ」についての間隔尺度用統計法を使うことだろう。

あるいは、統計法をある程度学習している者にとって、次のような反論がされるかもしれない。「間隔尺度版統計法は分析データが正規分布でなければならない。しかし順序尺度版データは正規分布とは限らない。順位化を行っても、順位化したデータが正規分布していなければ間隔尺度用統計法の使用は不適切ではないか。」結論から言えば、後述の中心極限定理により、非正規分布データに対しても、標本データ数が大きい場合には、使用可能であることが保証されている。

## 順序尺度における代表値の有意差検定法

順序尺度データを分析する簡便法は、前節で述べたように「順位化データの（間隔尺度用の）分析」であるが、現在では順序尺度版統計法が開発されており、専用の統計法を使うのがより適切であろう（森・吉田, 1990）。具体的には、対応なしの順序尺度データの分析法としては、三水準以上データ用の統計法がクラスカル・ウォリス検定、その二水準版のマン・ホイットニー検定（Mann-Whitney test, 以下 MW 検定）が開発されている。対応ありデータの統計法として、三水準以上データ用の統計法はフリードマン検定、その二水準版の符号検定（signed test）が開発されている。

林（印刷中）で述べたように、間隔尺度版多重比較法とは「厳格化された反復 $t$ 検定」である。間隔尺度データの二水準の代表値の有意差検定法として $t$ 検定を使うが、順序尺度データの二水準の代表値有意差検定として、対応なしの場合は MW 検定、対応ありの場合は符号検定を使う。上記の知識から類推できるように、順序尺度版多重比較法とは、対応なしの場合は「厳格化された反復 MW 検定」、対応ありの場合は「厳格化された反復符号検定」と再定義できる。間隔尺度版多重比較法では $t$ 検定の統計量及び確率を使って多重比較法を行うが、順序尺度版多重比較法では MW 検定及び符号検定の統計量及び確率を使うことになる。

ただし、MW 検定や符号検定は、直接の比較対象となる二水準データを分析対象とするため、一対比較型の多重比較法を行う場合の統計量計算式としてしか

使えないという問題点がある。対比型の多重比較法を行う場合、あるいは、関連する複数水準データ全体の情報を使った一対比較型の多重比較法を行う場合、後述する全体情報を考慮した統計量算出式が必要となる。

間隔尺度版多重比較法の統計量算出式は、最終的に一つの計算式に統一することができた(林, 印刷中)。しかし、順序尺度版多重比較法においては、対応なしの場合と対応ありの場合とでは、「データの順位化」段階において計算手順が異なるため、完全な統一化を図ることができない。以下、Table 1 の数値を例として説明する。

Table 1. 順序尺度版多重比較法の例題データ

a1	9	7	8	8	12	11	8	13
a2	6	5	6	3	6	7	10	9
a3	10	13	8	13	12	14	14	16
a4	9	11	13	14	16	12	15	14

①対応なしデータの計算式 データの順位化を行う場合、関連する水準データ全体の情報を使うか、あるいは直接対比対象となる二水準データのみの情報を使うかによって順位化の詳細が異なる。全体情報を考慮する場合は、分析対象となる全データについて順位化を行う (Table 2)。

Table 2. 「対応なしデータ」として処理した場合の全データの順位化

a1	13	6.5	9.5	9.5	20	17.5	9.5	23.5
a2	4	2	4	1	4	6.5	15.5	13
a3	15.5	23.5	9.5	23.5	20	27.5	27.5	31.5
a4	13	17.5	23.5	27.5	31.5	20	30	27.5

対比型、あるいは全体情報を使った一対比較型の多重比較法を使う場合、Table 2 のデータを式 (1) の統計量計算式に入力することになる。

$$\chi^2 = \frac{\left( \sum_{i=1}^a C_i \times \frac{TR_i}{N_i} \right)^2}{V \times \sum_{i=1}^a \frac{C_i}{N_i}} \dots\dots (1)$$

C<sub>i</sub> … 対比計算における重み係数  
 TR<sub>i</sub>, N<sub>i</sub> …… i水準における順位和TR, 人数N  
 V … 対応なしデータにおける誤差分散

式 (1) には誤差分散 V の数値を入力しなければならないが、誤差分散 V の計算式をクラスカル・ウォリス検定を参考にして導く。クラスカル・ウォリス検定の計算式には幾つかの種類があるが、森・吉田 (1990) の式 [4-2-12] を参考にする。以下の式 (2) の展開を行うことで、誤差分散 V (式 (2) の分母のイ

タリック部分) が算出される。

$$\begin{aligned} \chi^2(H) &= \frac{\sum_{i=1}^a N_i \times (TR_i - TR'')^2}{\frac{N(N+1)}{12}} \\ &= \frac{\sum_{i=1}^a N_i \times \sum_{i=1}^i (TR_i - TR'')^2}{\frac{N(N+1)}{12}} \times \frac{1}{\frac{1}{N}} \\ &= \frac{N \times \frac{1}{N} \times \sum_{i=1}^a (TR_i - TR'')^2}{\frac{N(N+1)}{12} \times \frac{1}{N}} \\ &= \frac{\sum_{i=1}^a (TR_i - TR'')^2}{\frac{N(N+1)}{12} \times \frac{1}{N}} \dots\dots(2) \end{aligned}$$

TR<sub>i</sub>, N<sub>i</sub> …… i水準における順位和TR, 人数N  
 N …… 全人数。N =  $\sum_{i=1}^i N_i$   
 TR'' …… 順位データの総平均値。TR'' =  $\frac{N+1}{2}$

ただし、式 (2) は同順位なしデータにおける計算式である。下記に同順位がない場合の式 (3-1)、同順位がある場合の式 (3-2) の誤差分散 V の計算式を示す。

$$\begin{aligned} V &= \frac{N(N+1)}{12} \dots\dots (3-1) \\ &= \frac{1}{N-1} \times \sum_{i=1}^a \sum_{j=1}^k \left( R_{ij} - \frac{N+1}{2} \right)^2 \dots\dots (3-2) \end{aligned}$$

R<sub>ij</sub> …… セル<sub>ij</sub>における順位値 (1 < i < a, 1 < j < k)

一対比較型多重比較法を行う場合、全情報を順位化した Table 2 を計算対象とする方法以外にも、直接比較する二水準データのみの情報を利用こともできる。その場合は、比較対象となる水準データに限定して「データの順位化」を行う。例えば a1 と a2 の比較を行う場合、Table 3 のような順位化を行う。

Table 3. 「対応なしデータ」として処理した場合の限定データにおける順位化

a1	11.5	6.5	9	9	15	14	9	16
a4	4	2	4	1	4	6.5	13	11.5

Table 3 に対して、式 (1) 及び式 (3-1)、式 (3-2) を使って統計量を計算してもよし、SPSS などの統計ソフトを使って MW 検定を実行、算出された統計量を使ってもよい。

誤差分散 V を統計ソフトなどで算出する方法を説明する。手順は以下の通り。①対象となるデータの順位化を行う (同順位の場合は平均順位値を求める)。②変換した順位化データに対して「対応なし一要因分

散分析」を実行し、「要因変動の平方和 (SS)」と「全体変動 (要因変動+誤差変動) の平均平方和 (MS)」を求める。SS, MS について森・吉田 (1990) を参照。

③式 (4) に代入して  $\chi^2 (H)$  を求める。

$$\begin{aligned} \chi^2(H) &= \frac{\text{要因変動の平方和(SS)}}{\text{全体変動の平均平方和(MS)}} \\ &= \frac{\text{要因変動の平方和(SS)}}{\frac{\text{全体変動の平方和(SS)}}{N-1}} \dots\dots (4) \end{aligned}$$

式 (4) から分かるように、クラスカル・ウォリス検定とは、分散分析の情報を利用して計算することができる。このため「クラスカル・ウォリスの順位の一元配置分散分析 (Kruskal-Wallis one-way analysis of variance by ranks)」という別称が与えられている (松原, 2000)。

②対応ありデータの計算式 全体情報を考慮した場合の統計量計算式の説明を行う。データの順位化を行うが、対応ありデータの場合と、対応なしデータの場合とでは順位化の詳細が異なる。順位化の作業は、同一被験者データ内で行うため、Table 4 のように4水準の場合、1, 2, 3, 4 の順位点を付与することとなる。

Table 4. 「対応ありデータ」として処理した場合の全データの順位化

a1	2.5	2	2.5	2	2.5	2	1	2
a2	1	1	1	1	1	1	2	1
a3	4	4	2.5	3	2.5	4	3	4
a4	2.5	3	4	4	4	3	4	3

全体情報を計算対象とする場合の統計量計算式は式 (5) となる。

$$\chi^2 = \frac{\left( \sum_i^a C_i \times \frac{TR_i}{N} \right)^2}{V \times \sum_i^a \frac{C_i}{N}} \dots\dots (5)$$

$C_i$  … 対比計算における重み係数

$TR_i$  ……  $i$ 水準における順位和TR

$N$  …… 人数

$V$  … 対応ありデータにおける誤差分散

式 (5) に入力する誤差分散  $V$  の計算式をフリードマン検定を参考にして導く。参考にした計算式は森・吉田 (1990) の式 [4-2-15] である。以下の式 (6) の展開を行うことで、誤差分散  $V$  (式 (6) の分母のイタリック部分) が算出される。

$$\begin{aligned} \chi^2(S) &= \frac{12 \times \sum_i^a (TR_i - TR')^2}{na(a+1)} \\ &= \frac{\sum_i^a (TR_i - TR')^2}{\frac{Na(a+1)}{12}} \\ &= \frac{\sum_i^a (TR_i - TR')^2}{\frac{Na(a+1)}{12}} \times \left( \frac{1}{N} \right)^2 \\ &= \frac{\sum_i^a \left\{ \frac{1}{n} \times (TR_i - TR') \right\}^2}{\frac{Na(a+1)}{12} \times \left( \frac{1}{N} \right)^2} \\ &= \frac{\sum_i^a \left( \frac{TR_i}{N} - \frac{TR'}{N} \right)^2}{\frac{Na(a+1)}{12N} \times \frac{1}{N}} \dots\dots (6) \end{aligned}$$

$TR_i$  ……  $i$ 水準における順位和TR

$TR'$  …… 順位和の総平均値。  $TR' = \frac{N(a+1)}{2}$

$N$  …… 人数

$a$  …… 水準の数

ただし、式 (6) は同順位なしデータにおける計算式である。下記に同順位がない場合の式 (7-1)、同順位がある場合の式 (7-2) の誤差分散  $V$  の計算式を示す。

$$V = \frac{a(a+1)}{12} \dots\dots (7-1)$$

$$= \frac{\sum_i^a \sum_j^N \left( R_{ij} - \frac{a+1}{2} \right)^2}{(a-1) \times N} \dots\dots (7-2)$$

$R_{ij}$  …… セル $ij$ における順位値

$\frac{a+1}{2} = R_i$  ……  $i$ 水準における順位データの平均値

一対比較型多重比較法を行う場合、全情報を順位化した Table 4 を計算対象とする方法以外にも、直接比較する二水準データのための情報を利用することもできる。その場合は、比較対象となるデータに限定して順位化を行う。a1 と a2 の比較を行う場合、Table 5 のようになる。

Table 5. 「対応ありデータ」として処理した場合の限定データにおける順位化

a1	2	2	2	2	2	2	1	2
a4	1	1	1	1	1	1	2	1

Table 5 に対して、式 (6) 及び式 (7-1)、式 (7-2) を使って統計量を計算してもよし、SPSS などの統計ソフトを使って符号検定を実行、算出された統計量を

使ってもよい。ただし統計ソフトの種類によっては、(標準化された) 統計量が算出されない場合がある。その場合は式 (8) を使って統計量を求め、自由度 1 の  $\chi^2$  分布を使って確率を求める。

$$\chi^2(S) = \frac{\{(+の数) - (-の数)\}^2}{(+の数) + (-の数)} \dots\dots (8)$$

誤差分散 V を統計ソフトなどで算出する方法は以下の通りである。①対象となるデータに対して、同一参加者内で順位化を行う (同順位の場合は平均順位値とする)。②変換した順位化データに対して「対応なし一要因分散分析」を行い、「要因変動の平方和 (SS)」と「全体変動 (要因変動 + 誤差変動) の平方和 (SS)」を求める。SS について森・吉田 (1990) を参照。③式 (9) に代入して  $\chi^2 (S)$  を求める。

$$\chi^2(S) = \frac{\text{要因変動の平方和(SS)}}{\frac{\text{全体変動の平方和(SS)}}{n(a-1)}} \dots\dots (9)$$

フリードマン検定も、クラスカル・ウォリス検定と同様に分散分析の情報を利用して計算できる。このため「フリードマンの順位の二元配置分散分析 (Friedman two-way analysis of variance by ranks)」という別称が与えられている (松原, 2000)。

③自由度の設定 本節で説明した統計量計算式は  $\chi^2$  統計量を導くので、 $\chi^2$  分布を使って確率を求めることになる。 $\chi^2$  分布には自由度  $df$  のパラメーターがあるが、 $df$  の設定法は一对比較型か対比型かによって異なる。一对比較型の場合は  $df=1$ 、対比型の場合は  $df = (\text{水準数}-1)$  とする。なお、MW 検定や符号検定は、教科書によっては  $\chi^2$  統計量ではなく、 $z$  統計量を導く計算式が紹介されていることがある。その場合は、 $z$  分布を使って確率を求めればよい。

## 順序尺度版の有意水準調整型多重比較法

具体的な多重比較法の説明に移る。多重比較法の厳格化の原理の中で最も理解が容易なのが、有意水準を直接調整するタイプであろう。このタイプの代表的なものとして Bonferroni 法, Sidak 法, Ryan 法, Holm 法が挙げられる (森・吉田, 1990; 永田・吉田, 1997)。

有意水準調整型多重比較法では、比較ペアの有意差を厳格化された有意水準、すなわち名義水準  $\alpha'$  を使って判定を行う。各手法によって名義水準の設定式が異なるが、いずれにしても、比較ペアから算出された統計量とそれに基づく確率を求め、両者を比較するという

パターンで処理を進めていけばよい。各手法の具体的な計算方法については、森・吉田 (1990)、永田・吉田 (1997)、林 (印刷中) を参照。なお、Ryan 法で多重比較を行う場合には、ステップ数を求める必要があるが、各水準の順位とあるいは順位と平均を使ってステップ数を求めればよい。

## 順序尺度版の分布調整型多重比較法

分布調整型多重比較法の特徴は、統計量から確率を算出するために使う分布として、多重比較法のために開発された特殊な分布を用いる点に特徴がある。間隔尺度における分布調整型多重比較法の基本手法である Tukey HSD 法は、算出された統計量を、本来の  $t$  分布の代わりに「Sutudent 化された範囲  $q$  分布」を使って検定を行う。同様に、順序尺度における分布調整型多重比較法の基本手法である Steel-Dwass 法も  $q$  分布を用いて検定を行うことになる。

本来  $q$  分布が間隔尺度版多重比較法のために開発されたことを考えると、順序尺度版多重比較法に転用するやり方に疑問を感じるかもしれない。しかし、転用についての理論的根拠が「中心極限定理 (central limit theorem)」によって与えられている。中心極限定理とは、確率変数 (連続量でも離散量でもよい)  $x$  が、母平均  $\mu$ 、母分散  $\sigma^2$  をもつある分布に従うとき、これから無作為に抽出した大きさ  $n$  の標本平均  $x'$  の分布は、 $n$  が大きくなるにつれて、母平均  $\mu$ 、母分散  $\sigma^2/n$  の正規分布に近似する、という定理である (松原, 2000)。

心理統計法の教科書には中心極限定理が充分に解説されていないことも多いが、その効果は大きい。統計的研究法を行う心理学者にとって、平均値を対象とする間隔尺度版統計法は馴染みが深い。この平均値を対象とする解析法は、標本サイズが十分に大きければ、母集団の分布が正規分布でなくとも、すなわち、正規分布を仮定しない順序尺度データに対しても、間隔尺度版統計法を適用することの妥当性を与えてくれる。なぜならば、順序尺度版多重比較法では、順位和平均 (あるいはその派生である順位和) を計算対象としているため、中心極限定理が作用するからだ。

なお、中心極限定理の効果は多重比較法に限定されるものではなく、 $t$  検定や分散分析をはじめとする多くの統計法に当てはまる。統計法の教科書の多くには、 $t$  検定や分散分析の分析データは非正規分布データには適用してはならない、という記述されている。確かに  $t$  検定、分散分析、多重比較法は標本データの正規分布性を仮定して開発された手法であるが、直接分析対象となるものが標本データの分布ではなく、標本平

均の分布であることから、中心極限定理が作用する。そのため、少なくともデータ数が多い場合には、これらの統計法を用いても問題はない。

ただし、順序尺度データにおいて  $q$  分布を用いた多重比較法を行う場合、有意判定には注意を要する。そもそも、本来  $q$  分布は  $t$  分布の改良版として開発されたため、 $t$  分布を使う場面で  $q$  分布が代用される。しかし、順序尺度版統計法で用いられる本来の分布は  $\chi^2$  分布（あるいは、その開平である  $z$  分布）であり、 $t$  分布ではないため、そのままでは  $q$  分布を使えない。

ここで  $z$  分布、 $t$  分布、 $q$  分布、 $\chi^2$  分布の関係を整理する。これらの分布の中で、基本となる分布が  $z$  分布であり、 $z$  分布を二乗化したものが  $\chi^2$  ( $df=1$ ) 分布である。 $z$  分布は大標本（標本サイズ $=\infty$ ）データのために開発された分布であり、小標本データにも使えるように改良された分布が  $t$  分布である。このことから、 $df=\infty$  と設定したとき  $t$  分布  $= z$  分布となる（森・吉田，1990）。また、多重比較法のために  $t$  分布を改良したものが  $q$  分布である。ゆえに、 $z$  分布  $= q$  分布 ( $df=\infty$ ) となるので、 $z$  統計量を  $q$  分布を使って検定を行う場合、 $z$  統計量  $>$  「 $q$  分布表の棄却臨界値」 であれば有意差ありと見なす。また、 $\chi^2$  分布  $= (z$  分布) $^2 = (q$  分布 ( $df=\infty$ )) $^2$  となるので、 $\chi^2$  統計量を  $q$  分布を使って検定を行う場合、 $\chi^2$  統計量  $>$  「 $q$  分布表の棄却臨界値」 であれば有意差ありと見なす。

このように順序尺度データに対して  $q$  分布を使う場合  $df=\infty$  を仮定する必要がある。 $df=\infty$  とは、本来標本サイズが無限数であることを意味するが、実際の研究場面において標本サイズは有限でしかありえないので、一般には、標本サイズがある程度の大きさであれば  $df=\infty$  とみなしても構わない。永田・吉田(1997)は「標本サイズ  $n$  は10以上」を一つの目安としている（但し、 $n=10$  が十分な大きさを意味するわけではない）。結局、順序尺度データに対して  $q$  分布に基づく多重比較法を使う場合、中心極限定理と、 $df=\infty$  の仮定の二つの理由から、標本サイズにある程度の大きさが要求されることになる。

また  $q$  分布を使う場合、分析対象となるデータは水準間に独立であることが前提となる。この前提条件から、間隔尺度データにおいて  $q$  分布型多重比較法は、対応ありデータに対して使うことが不適となる（林，印刷中）。しかし、本稿で説明した式 (5)～式 (8) 使った順序尺度版多重比較法を行う場合、対応ありであっても  $q$  分布を使ってもよい。式 (5)～式 (8) の計算対象となるデータは Table 4 のように、同一被験者内データに対して順位化を行っているため、ある被験者がどの水準においても高い順位値を得るという相関の

あるデータにならないためである。順序尺度においては  $q$  分布は、対応なしデータと対応ありデータのいずれに対しても使える。

本節では、 $q$  分布を使った多重比較法を中心に説明したが、分布調整型多重比較法には別の分布を使ったものも開発されている。詳細については永田・吉田(1997)を参照してもらいたい。

## 順序尺度版の統計量調整型多重比較法

統計量を厳格化する多重比較法として Scheffe 法があるが、順序尺度版多重比較法としてこの手法を用いる場合には有意判定には注意を要する。そもそも間隔尺度版 Scheffe 法は、「統計量の縮小化（統計量  $\div$  [変数  $-1$ ]）」の操作を行った上で、 $F$  分布で有意性判定を行っている。間隔尺度版統計法の多くは  $t$  分布を用いて有意性判定を行うが、自由度  $a$  の  $t$  分布の棄却値の二乗は自由度 ( $1, a$ ) の  $F$  分布と一致することを利用して、 $F$  分布を代用している。しかし、順序尺度版分析法の多くは  $t$  分布ではなく  $z$  分布（あるいはその二乗の  $\chi^2$  分布）で有意判定を行っている。先述したように、 $t$  分布において  $df=\infty$  を仮定した場合  $z$  分布と同じになるため、順序尺度で  $F$  分布を代用する場合、自由度は ( $1, \infty$ ) と設定しなければならない。

あるいは  $F$  分布の代わりに  $\chi^2$  分布を使って検定を行う方法もある。 $F$  分布と  $\chi^2$  分布との関連性から、 $\chi^2$  分布で検定を行う場合、「統計量の縮小化」を行わずに、 $df = (\text{変数} - 1)$  の  $\chi^2$  分布を使って有意判定を行うことになる（永田・吉田，1997）。

$$\begin{aligned} & \text{棄却臨界値} \\ & = (\text{変数} - 1) \times F(\text{変数} - 1, \infty; \text{有意水準}) \\ & = \chi^2(\text{変数} - 1; \text{有意水準}) \end{aligned} \quad \dots\dots (10)$$

## どの手法を用いればよいか

順序尺度版多重比較法は SPSS などの統計ソフトに搭載されておらず、実施する場合には（ある程度）手計算で行う必要がある。しかし高度な多重比較法ほど計算が複雑となるので、ここでは比較的簡便でなるべく検定力の高いものを以下推奨していく。

① **一対比較型** 統計量の計算には式 (1) あるいは式 (5) を使うよりも、MW 検定や符号検定の計算式を使うのがよい。ただし符号検定で統計量が必要になる場合は式 (8) を使う。

統計ソフトを使う場合は、確率が自動的に表示され

Table 6. 例題を「対応なしデータ」とみなして、有意水準を厳格化した反復符号検定としての多重比較法の結果

比較ペア	統計量及び確率	step数	確率順位	Bonferroni	判定	Sidak	判定	Ryan	判定	Holm	判定
a1 vs a2	$\chi^2(1)=5.338, p=0.0209$	2	4	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0250	<i>s.</i>	0.0167	<i>n.s.</i>
a1 vs a3	$\chi^2(1)=4.864, p=0.0274$	2	5	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0250	<i>n.s.</i>	0.0250	<i>n.s.</i>
a1 vs a4	$\chi^2(1)=6.353, p=0.0117$	3	3	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0125	<i>s.</i>	0.0125	<i>s.</i>
a2 vs a3	$\chi^2(1)=9.598, p=0.0019$	3	2	0.0083	<i>s.</i>	0.0085	<i>s.</i>	0.0125	<i>s.</i>	0.0100	<i>s.</i>
a2 vs a4	$\chi^2(1)=10.260, p=0.0014$	4	1	0.0083	<i>s.</i>	0.0085	<i>s.</i>	0.0083	<i>s.</i>	0.0083	<i>s.</i>
a3 vs a4	$\chi^2(1)=0.176, p=0.6744$	2	6	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0250	<i>n.s.</i>	0.0500	<i>n.s.</i>

Table 7. 例題を「対応ありデータ」とみなして、有意水準を厳格化した反復符号検定としての多重比較法の結果

比較ペア	統計量及び確率	step数	確率順位	Bonferroni	判定	Sidak	判定	Ryan	判定	Holm	判定
a1 vs a2	$\chi^2(1)=4.500, p=0.0339$	2	4.5	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0250	<i>n.s.</i>	0.0200	<i>n.s.</i>
a1 vs a3	$\chi^2(1)=4.500, p=0.0339$	2	4.5	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0250	<i>n.s.</i>	0.0200	<i>n.s.</i>
a1 vs a4	$\chi^2(1)=6.125, p=0.0133$	3	3	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0125	<i>n.s.</i>	0.0125	<i>n.s.</i>
a2 vs a3	$\chi^2(1)=8.000, p=0.0047$	3	1.5	0.0083	<i>s.</i>	0.0085	<i>s.</i>	0.0125	<i>s.</i>	0.0091	<i>s.</i>
a2 vs a4	$\chi^2(1)=8.000, p=0.0047$	4	1.5	0.0083	<i>s.</i>	0.0085	<i>s.</i>	0.0083	<i>s.</i>	0.0091	<i>s.</i>
a3 vs a4	$\chi^2(1)=0.000, p=1.0000$	2	6	0.0083	<i>n.s.</i>	0.0085	<i>n.s.</i>	0.0250	<i>n.s.</i>	0.0500	<i>n.s.</i>

るので、有意水準調整型多重比較法の Holm 法を使うのがよいだろう。有意水準を自由に指定できる  $q$  分布臨界値表があるのなら Tukey-Welch 法を行うのが良い (例えば永田・吉田 (1997) を参照)。有意水準が  $\alpha=0.05$  など固定されている  $q$  分布臨界値表を使う場合は、順序尺度版 Tukey WSD 法を行っても良い。ただし、Tukey 型の多重比較法は中心極限定理を援用するために、ある程度の標本サイズが必要となる点に注意せよ (標本サイズが小さいならば、Holm 法を用いた方がよい)。統計量調整型の Scheffe 法については、一対比較型では検定力がその他の手法に比べて小さいため、他の手法を用いた方がよい。

②対比型 本節で紹介した式 (1) あるいは式 (5) を使って統計量を求める。 $q$  分布調整型多重比較法は、統計量調整型の Scheffe 法に比べて検定力が低いので、Scheffe 法がよい。有意水準調整型多重比較法も使えるが、比較ペアの数に応じて検定力が低くなるので、比較ペア数が多いならば Scheffe 法を用いた方がよいだろう。

③計算例 最後に、Table 1 の例題データを使った順序尺度版多重比較法の計算例を示す。一対比較型の比較を行い、有意水準の観点で厳格化を行っている。対応なしデータとみなした計算例を Table 6, 対応ありデータとみなした計算例を Table 7 で示した。

対応ありデータとみなした場合の多重比較法では、比較ペアの確率の順位に同値が見られた。このような場合の対処法は教科書には言及されていないが、Table 7 のように平均順位値を与える方法が考えられる。Ryan 法や  $q$  分布型多重比較法で重要なパラメーター

である step 数についても同順位となった場合、同様の処理を行えばよいだろう。

本稿では、順序尺度版多重比較法が間隔尺度版多重比較法と関連があることを強調している。学習現象とは、一般に、学習者が持つ既得知識を基礎として新しい知識が獲得されていく。統計的研究を行う心理学者の多くは、間隔尺度版統計法の知識を持っていても、順序尺度版統計法を学んでいない。本稿は、比較的馴染みのある間隔尺度版多重比較法を既得知識として、新しい知識である順序尺度版多重比較法をどのように学習すればよいかの学習過程の一例を示した。効果的な統計法の教授学習法を考えるための資料になるだろう。

## 【引用文献】

林 智幸 (印刷中). 厳格化の観点からの多重比較法の整理 広島大学大学院教育学研究科紀要第三部 (教育人間科学関連領域), 54.

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

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

松原 望 (2000). 改訂版 統計の考え方 放送大学教育振興会

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

小野寺孝義・山本嘉一郎 (2004). SPSS 事典 BASE 編 ナカニシヤ出版