

原

著

医学における統計学の応用について

— (第4編) 分散分析と多重比較法について —

新潟大学医学部衛生学教室 (主任: 山本正治教授)

遠藤 和男・片桐 幹雄

An Application of Statistical Methods to Medical Science
— (Part IV) on Analysis of Variance and Multiple Comparison Methods —

Kazuo ENDOH and Mikio KATAGIRI

Department of Hygiene and Preventive Medicine

Niigata University School of Medicine

(Director: Prof. Masaharu YAMAMOTO)

Analysis of variance is a popular and important method to evaluate the experimental results. In case there is a statistical significance recognized by analysis of variance, then a researcher tends to use computer programs for multiple comparison methods. He or she, however, can't understand how to apply those methods properly. As a reference for those who need analysis of variance and multiple comparison methods, we present interpretations of the statistical methods by making use of some hypothetical examples.

Key words: analysis of variance, one-way layout, mutiple comparison methods,
non-parametric methods

分散分析, 一元配置, 多重比較法, ノンパラメトリック技法

はじめに

分散分析は、実験結果を評価する方法として、非常に有名かつ重要な方法である。コンピュータ・プログラムの発達によって、分散分析で有意差が認められた場合に、各水準間の平均値の差を検出するのも容易となった。しかしながらかえって、どの方法を用いてよいか分からな

い、といった声をよく耳にする。疑問の一部については簡潔に解説¹⁾したことがあるが、その他にも様々な相談を受けてきた。

分散分析も非常に多種多様であるが、一元配置の繰り返しを個体差と考えれば、繰り返しのない二元配置とみなせる。前者における多重比較法の多くはそのまま後者にも適用できる。そこで本編では、一元配置に焦点をあ

Reprint requests to: Kazuo ENDOH,
Department of Hygiene and Preventive Medicine,
Niigata University School of Medicine,
Niigata City, 951, JAPAN.

別刷請求先: 〒951 新潟市旭町通1番町
新潟大学医学部衛生学教室

遠藤 和男

て、分散分析及び多重比較法の適用について、仮想的な例を掲げながら解説を加えていく。検出力²⁾を検討するため、有意性はできる限り p 値で示すよう心掛けた。ただし本来、 p 値から判断するのではなく、有意水準 α は検定の始めに設定しておくべきである。

I. 分散分析の適用の前に

1. 正規性 (normality) について

過去の文献または自分のデータから、ある測定値が正規分布に従うことが明らかにされていれば、標本数がたとえ少なくとも、標本平均は正規分布に従うことが知られている。これを中心極限定理²⁾という。また、右にスソをひく (skewed to the right) 分布でも、対数をとると正規分布に従うようになるデータも多いので、検定の前に十分検討する必要がある。なお、正規性の検定方法については、拙著²⁾または第2編³⁾を参照されたい。

2. 分散の一樣性 (同質性: homogeneity) について

分散の一樣性の検定方法として、① 一般的なパートレット (Bartlett) の方法⁴⁾、② 最大と最小との分散比を検討するハートレイ (Hartley) の方法⁵⁾、③ 1つだけ大きな不偏分散に注目するコクラン (Cochran) の方法⁶⁾が知られている。ただし、1. 正規性と比較した場合、F検定法の頑健性 (robustness) から、一樣性についてはそれほど厳密でなくともよいと考えられている⁷⁾。

また、③のように1つだけ不偏分散が大きな場合、飛び離れた値が不自然に含まれている可能性もある。グラブズ・スミルノフ (Grubbs-Smirnov) の棄却検定法⁸⁾を適用し、飛び離れた値を除外したあとで再び計算すると、不偏分散が安定することも多い²⁾。データの信頼性も増すため、ぜひ適用について検討されたい。

3. 臨床データの取扱いについて

分散分析は、もともと実験データの評価方法であり、臨床データではそのまま適用できない場合が多い。その理由として、以下の3点があげられる。

- ① 遺伝的素因が全く同等と考えられるマウス等の動物では、個体差を無視して繰り返しの実験と考えてよいが、ヒトでは個人差を無視できない。したがって、一元配置では誤差分散 V_E が大きくなってしまい、F検定で有意差が認められないことが多い。また②、③の理由によって、二元配置も適用しにくい。
- ② 因子水準は研究者自身ではなく、患者の背景因子によって決定される。患者側の因子が名義尺度²⁾であ

表 1 穀物摂取頻度とローレル指数との関係⁹⁾

摂取頻度	男子	女子
	平均±標準偏差 (人)	平均±標準偏差 (人)
人より多い	126.9±18.18(21)*	133.2±24.45(20)*
普通	118.5±14.55(61)	125.8±20.99(84)
人より少ない	118.7±6.37(5)	119.1±11.36(12)

注) * : 「普通」の群より大きい ($p < 0.05$)
 * : 「少ない」群より大きい ($p < 0.05$)

れば、2群がほとんどなので多重比較を行うまでもない。

- ③ 因子が順序尺度による場合も、研究者が予め各水準の人数を揃えるのは至難のわざである。したがって、(a) 各水準の等間隔性が保証されないばかりか、(b) それぞれの例数もバラバラになってしまう。

例として、アンケート調査をもとにローレル指数の平均値を比較した、筆者らのデータ⁹⁾を表1に示した。レフェリーから一元配置の適用について指摘されたが、①~③の理由で反論し、スチューデント (Student) またはウェルチ (Welch) のt検定を適用した。アンケートの回答から分けた3つの群は、互いに独立とみなしてよい。分散分析では各水準が独立ではないし、ウェルチ法を用いるようでは、分散の一樣性自体に疑問が残る。

II. 多重仮説過誤について

全標本数がN、因子水準の数がk、繰り返し数lが一定である場合の一元配置を例とする。分散分析表は表2のとおりである。有意水準 α は事前に設定しておく。自由度 $\phi_A = k - 1$, $\phi_E = N - k$, 因子間と誤差との分散比を計算し、 $F_0 = V_A / V_E > F_{N-k}^{k-1}(\alpha)$ であれば、母平均がすべて等しいとする帰無仮説: $m_1 = m_2 = \dots = m_k$ を棄却し、有意差ありと判断できる。

したがって、上記の等号のどこかは棄却され、差が認められるはずで、多重比較 (multiple comparison) 法

表 2 一元配置分散分析表

要因	偏差平方和	自由度	不偏分散	分散比
因子間	S_A	$\phi_A = k - 1$	$V_A = S_A / \phi_A$	$F_0 = V_A / V_E$
誤差	S_E	$\phi_E = N - k$	$V_E = S_E / \phi_E$	
計	SS	$\phi = N - 1$		

注) 全標本数 $N =$ 因子水準 $k \times$ 繰り返し数 l ,
 繰り返し数が異なる場合 $N = \sum_{i=1}^k l_i$

の適用となる。しかし、2群の平均値の差を検出する目的で、スチューデントのt検定を繰り返して用いるのは誤りである。各水準のデータは独立とみなせないため、一連のデータとして評価する必要がある。

t検定の場合、「2群の平均値に差は認められない」という帰無仮説 H_0 が正しい時に、これを棄却してしまう危険率を α (第1種の過誤) とする。因子水準の数が k 個の場合、平均値を比較する組合せの数 r は、 $r = {}_kC_2 = k(k-1)/2$ となる。t検定を r 回繰り返した場合の危険率を α' とするのは早計に過ぎる。

前述の H_0 を r 回とも正しく採用する確率は $(1-\alpha)^r$ 、逆に1回でも棄却してしまう過誤の確率は $1-(1-\alpha)^r \approx r\alpha$ となる。これを多重仮説過誤 (family-wise error, 以下 FWE と略す)¹⁰⁾ と呼ぶ。したがって、種々の多重比較法では FWE を小さく抑えて、有意差ありとする対立仮説の拾い過ぎを戒めるように働いている。

III. すべての組合せについての比較

【例1】拙著²⁾にも掲載した例の一部に変更を加えて表3に示した。分散分析表は表4のとおりである。パートレットの検定について、繰り返し数 l が一定でない場合も汎用できる手順を下方に示した。一様性ありとする帰無仮説を棄却できない。以下この例を用いて、種々の多重比較法について検討を加えてみたい。

1. フィッシャーの最小有意差検定法

多重比較に際して最も有意差の出やすい検定法は、拙著²⁾にも紹介した近似法である。正式にはフィッシャーの最小有意差検定法 (Fisher's Least Significance Difference procedure, 以下 LSD 法と略す)¹¹⁾ と呼ばれている。Macintosh の StatView にも搭載され、解説¹²⁾ もあるが十分とは言いがたい。スチューデントの t 検定式 I) において、共通の分散 u^2 の代わりに表2の誤差分散 V_E を用いたのが式 II) である。繰り返し数 l が一定の場合、分母のカッコ内は $2/l$ となる。

$$t_0 = \frac{|\bar{x}_i - \bar{x}_j|}{u \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} > t_{df} \left(\frac{\alpha}{2}\right) \quad \dots\dots\dots \text{式 I)}$$

$$df = n_i + n_j - 2$$

$$t_0 = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{V_E \left(\frac{1}{l_i} + \frac{1}{l_j}\right)}} > t_{N-k} \left(\frac{\alpha}{2}\right) \quad \dots\dots\dots \text{式 II)}$$

(一定の場合 $2/l$)

式 II) の t_0 値は汎用性が高く、他の検定法でも計算されている。不等式が成立すれば、有意水準 α で有意と

表3 【例1】一元配置の仮想例

因子水準 \ 繰り返し	繰り返し					平均
	1	2	3	4	5	
A ₁ (6:00)	12	11	12	13	12	12.0
A ₂ (12:00)	12	13	14*	14	12	13.0*
A ₃ (18:00)	14	15	14	14	14	14.2
						13.067*

注) *: 拙著【例12-2】²⁾と異なる数値

表4 【例1】の分散分析表

要因	偏差平方和	自由度	不偏分散	分散比
因子間	12.133	2	6.0667	$F_0 = 10.706$
誤差	6.800	12	0.5667	$(p = 0.002146)$
計	18.933	14	1.3524	

【Bartlett の検定の手順】

- ①各水準の不偏分散 u^2 を算出する。
 $u^2_1 = 0.5, u^2_2 = 1.0, u^2_3 = 0.2$ である。
- ② $\bar{u}^2 = (u^2_1 + u^2_2 + u^2_3)/3 = 0.5667 = V_E$
- ③自由度 $\phi_E = 12$ 、各水準の繰り返し数 l_i から、
$$c = 1 + \frac{1}{3(k-1)} \left[\sum \frac{1}{l_i - 1} - \frac{1}{\phi_E} \right] = 1.111$$
- ④ $\chi^2_0 = [\phi_E \ln \bar{u}^2 - \sum (l_i - 1) \ln u^2_i] / c$
 $= 2.155 < \chi^2_{k-1}(0.05) = 5.991 (N.S.)$
- ⑤帰無仮説は棄却できず、一様性ありとみなせる。

A ₁	0.105322 (N.S.)	0.000370 ($p < 0.001$)
A ₂		0.039968 ($p < 0.05$)
A ₃		

注) A₁:A₂ などはそれぞれの交差した所を見る。

図1 Student 法による検定

判断できる。キレがよいとの評価がある¹³⁾ 反面、FWE を犯す可能性を避けることができない。スチューデント法による結果を図1に、LSD 法などの結果とともに表5にも示した。自由度は $n_i + n_j - 2 < N - k$ であり、 V_E が u^2 より小さい場合には、後者の方が p 値が小さくなるが、【例1】では有意性に変わりがない。

2. ボンフェローニの不等式

ボンフェローニ (Bonferroni) の不等式¹³⁾ も有名で、StatView にも搭載されている。FWE $\approx r\alpha$ と考えられるので有意水準を α ではなく、 α/r とする考え方である。式 II) が用いられることに変わりはない。通常、 α/r

を与える%点と、 r 倍する前の p 値の両方が表示される。初心者では p 値を鵜呑みにして有意差を拾い過ぎないよう、十分に注意する必要がある。

むしろ、LSD法による直接確率 p 値を、 r 倍して α と比較する方がわかりやすい。 $k=3$ の時 $r=kC_2$ も3となるので、3倍して表5に示した。スチューデント法やLSD法と比較すると、 A_1-A_3 の有意性は0.1%未満から1%未満に変わる。 A_2-A_3 については、5%未満で有意から有意差なしと判断される。

$FWE = \alpha r$ と近似してよいのは、水準数 r が4以下と小さい場合であり、 k が大きくなるほど近似が悪くなる。例えば $\alpha=5\%$ の場合について、 $k=3$ なら r も3なので、 $FWE=14.3\%$ に対して $\alpha r=15\%$ であるが、 $k=5$ では $r=10$ 、 $FWE=40.1\%$ に対して $\alpha k=50\%$ と差が大きくなる。したがってボンフェローニ法は厳しすぎる(保守的ともいう)と言われている⁷⁾¹⁴⁾。

3. シェフェの方法

シェフェ(Scheffé)¹⁵⁾の方法も有名で、StatMateやStatViewにも搭載されている。母平均 m_i に対して $\sum c_i m_i = 0$ 、 $\sum c_i = 0$ となるように係数 c_i を設定する。2群の検定の場合は $c_1=1$ 、 $c_2=-1$ 、 $c_3=\dots=c_k=0$ とおけばよい。

$$S_{12} = \frac{(\bar{x}_1 - \bar{x}_2)^2}{\left(\frac{1}{l_1} + \frac{1}{l_2}\right)(k-1)V_E} > F_{N-k}^{k-1}(\alpha) \quad \dots\dots\dots \text{式III}$$

式IIIが成立すれば、有意水準 α で $m_1 - m_2 \neq 0$ と判断できる。 $S_{12}=2.206$ 、 $S_{13}=10.676$ 、 $S_{23}=3.176$ を $F_{12}^{2}(0.05)=3.885$ などと比較すればよい。F分布の逆関数から直接確率 p 値を算出して表5に示した。有意性はボンフェローニ法による場合と同様で、シェフェ法も保守的である⁷⁾¹⁴⁾と言われている。

4. チューキー・クレイマーの方法

チューキー(Tukey)の方法¹⁶⁾は繰り返し数 l が一定の場合に適用される。一定でない場合は修正法である、チューキー・クレイマー(Tukey-Kramer, 以下TKと略す)法¹⁷⁾が用いられる。後者を式IV)に示したが、不等号の右側 q は t 分布ではなく、スチューデント化した範囲の%点¹⁸⁾であることに注意する。

似たような方法として、スチューデント・ニューマン・クルーズ(Student-Neuman-Keuls, 以下SNKと略す)法¹⁹⁾を式V)に示した。%点側を $\sqrt{2}$ で割るか、逆に t_{ij} 側を $\sqrt{2}$ 倍するかの違いであり、両式は本質的に

表5【例1】の多重比較法のまとめ

多重比較	A_1-A_2	A_1-A_3	A_2-A_3
検定法/差	-1.0	-2.2	-1.2
Student法	0.105322	0.000370	0.039968
FisherのLSD法	0.057503	0.000589	0.026885
Bonferroni不等式	0.172509	0.001767	0.080655
Scheffé法	0.152812	0.002169	0.078136
TK・SNK法*	0.131764	0.001548	0.064742

注) *: Tukey-Kramer法及びStudent-Newman-Keuls法

は変わらない。したがって以下、TK・SNK法と総称する。式IV)の t_{ij} 値は式II)の t_0 値と等しく、また、式V)による t_{ij} 値は式IV)の $\sqrt{2}$ である。

$$t_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{V_E \left(\frac{1}{l_i} + \frac{1}{l_j}\right)}} > q(k, N-k, \alpha) / \sqrt{2} \quad \dots\dots\dots \text{式IV}$$

$$t_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{V_E}{2} \left(\frac{1}{l_i} + \frac{1}{l_j}\right)}} > q(k, N-k, \alpha) \quad \dots\dots\dots \text{式V}$$

【例1】では $k=3$ 、 $N-k=12$ の%点を用いる。5%有意点 $=3.7729$ 、1%有意点 $=5.0459$ である。式V)により $t_{12}=2.970$ 、 $t_{13}=6.535$ 、 $t_{23}=3.564$ から、それぞれ $N.S.$ 、 $p < 0.01$ 及び $N.S.$ と判定される。なお、簡易統計数値表¹⁸⁾により自由度 $1=\nu$ 、自由度 $2=N-k$ とおくと、 $F_0 = (t_{ij}/c)^2$ がF分布に近似できるので、直接確率 p 値を算出して表5に示した。

A_2-A_3 は、スチューデント法及びLSD法により5%有意であったが、ボンフェローニ法、シェフェ法及びTK・SNK法では有意差なしと判断され、FWEを避けることができた。後3者で検討すると、 A_1-A_2 及び A_1-A_3 についても p 値がそれぞれ最も小さいため、TK・SNK法の検出力が最高と考えられる。

【結論1】すべての水準の組合せについて平均値の差を検討する場合、TK・SNK法が最も優れている。

IV. 対照群と実験群との比較

IIIで述べた方法は、すべての因子水準の組合せに対して、平均値の差を検定している。しかし、対照群に対して $k-1$ 個の実験群を設定した場合、すべての組合せに興味があるわけではない。したがってIIIに掲げた方法の他に、対照群と実験群のみの平均値に注目する、多重

比較法が考案されている。

【例2】因子水準 $k=4$ で、各水準間でデータの個数にバラツキが認められる表6のデータについて検討する。分散分析表は省略したが有意 ($p<0.01$) である。また、前述した1~4の方法による直接確率 p 値も計算して表7に示した。以下の2つの方法による結果と比較検討してもらいたい。

5. ボンフェローニのt検定

ボンフェローニの不等式の変法である。すべての組合せの数は $r=4 \cdot 3/2=6$ であるが、対照群との比較では $r=k-1=3$ となる²⁰⁾。したがって、【例2】では表7に示したLSD法による直接確率 p 値を、6倍ではなく3倍すればよいことになる。

6. ダネットの方法

ダネット (Dunnett) の方法²¹⁾ は、対照群と実験群とを比較する場合、最も検出力が高いと言われている¹⁰⁾。原表では各水準の繰返し数 l が一定の場合を想定しているが、多少異なっても構わないとされている。

- ① 自由度 ϕ_E 、誤差分散 V_E は表6下方の数値を用いる。
- ② 対照群の平均値 \bar{x}_1 と、その差が最大である水準 i の平均値 \bar{x}_i とを比較する。式II) から t_k 値を算出する。
- ③ 有意水準が α の場合、片側の%点 $d(k, \phi_E, \alpha)$ ¹⁰⁾ よりも t_k の方が大きければ、有意差ありと判断する。ただし、実験が終了するまで対照群と各実験群との大小が不明であれば、両側の%点を用いること。
- ④ 次に差の大きい水準について、 $k-1$ から 2 になるまで②~④を繰り返すが、有意差なしなら終了する²⁰⁾。
- ⑤ ただし、逐次検定によらない考え方¹⁰⁾ もある。なお、 $k=2$ の時は t 分布の片側の%点と一致する。

【例2】でダネットの片側の%点を用いると、

$$\begin{cases} t_4=3.45424 > d(4, 16, 0.01)=3.0521, \\ t_3=2.95701 > d(3, 16, 0.01)=2.8835, \\ t_2=0.62579 < d(2, 16, 0.05)=1.7458 \end{cases}$$

逐次検定によらず k を一定にすると、 t_3 は5%有意と判断される。なお、繰返し数が一定ではなくとも、直接確率 p 値を正確に計算するプログラムが開発されている²²⁾ が、FORTRANのサブルーチンが必要である。本編でも5%及び1%の有意性の表示にとどめた。

A_1-A_3 について、スチューデント法及びLSD法では有意差が認められたがボンフェローニの不等式及びシェフェ法では消失し、 A_1-A_4 の有意性は1%未満から5%未満に変わった。保守的であることが理解できよう。

表6 【例2】対照群を設けた場合の仮想例

繰返し	1	2	3	4	5	6	平均
因子水準							
$A_1(0)$	205	206	161	190	194	—	191.20
$A_2(I)$	201	221	197	185	—	—	201.00
$A_3(II)$	248	265	191	220	212	262	233.00
$A_4(III)$	202	276	237	254	230	—	242.20
							218.45

注) A_1 を対照群とし、—は欠損値を示す。
 $\phi_A=k-1=3, \phi_E=N-k=16,$
 $V_A=3007.12, V_E=544.975$ から、
 $F_0=5.518 > F_{16}^3(0.01)=5.292 (p<0.01)$

表7 対照群と実験群との多重比較

多重比較	A_1-A_2	A_1-A_3	A_1-A_4
検定法 / 差	-9.8	-41.8	-51.0
Student 法	0.416376	0.023590	0.005163
Fisher の LSD 法	0.570274	0.009274	0.003264
Bonferroni 不等式	>1	0.055644	0.019584
Scheffé 法	0.940519	0.066363	0.027071
TK・SNK 法*	0.920805	0.041267	0.015210
Bonferroni の t 法	>1	0.027822	0.009792
Dunnett 法	N.S.	$p<0.05^*$	$p<0.01$

注) * : Tukey-Kramer 法及び Student-Newman-Keuls 法
 * : 逐次検定法を採用した場合は $p<0.01$ と判断できる。

TK・SNK 法では両者とも5%未満、ボンフェローニのt検定及びダネット法では、前者が5%、後者は1%未満である。ただし、2種類のボンフェローニ法では、 p 値が1を超える場合があるので不合理である。

【結論2】すべての組合せではなく、対照群と実験群のみの比較については、ダネット法が最も優れている。

V. 特殊な多重検定法について

7. 用量効果を見るためのウィリアムズ法

【例2】の因子水準が、例えば(0)0, (I)50, (II)100及び(III)200 mg/Kgのように、ある薬剤の用量になっている場合を想定する。当然、用量が増加するにしたがって、平均値も上昇するかどうかを検討したい。直線的な関係を見たい場合、もはや多重比較ではなく、直交多項式⁷⁾や共分散分析¹⁰⁾の対象となる。

ダネット法により、対照群と高用量の実験群との間で

有意差が認められるなら、用量効果ありと判定してよいが、全体として判定するには、次に述べるウィリアムズ (Williams)²³⁾ の方法による方がよい。原表は繰り返し数 l が一定の場合であるが、多少異なっても構わないと言われている¹⁰⁾。

- ① 自由度 ϕ_E 、誤差分散 V_E は表 6 下方の数値を用いる。
- ② $\bar{x}_1 \leq \bar{x}_2 \leq \dots \leq \bar{x}_k$ が成立していることを確認する。成立していない場合には、1 つ下の因子水準と合併させて新しい平均値を求め、途中では $\bar{x}_i = \bar{x}_j$ になっていると考える。ただし、対照群の平均値が最大の場合は、上記の式で \leq を $>$ と読み替える。
- ③ 最高用量の平均値 \bar{x}_k と最低用量の平均値 \bar{x}_1 について、式 II) の t_k 値 (ダネット法と同値) を計算する。
- ④ 有意水準が α の場合、片側の % 点 $w(k, \phi_E, \alpha)$ ¹⁰⁾ より t_k の値が大きい時には、平均値 \bar{x}_k と \bar{x}_1 との間に有意差ありと判断する。有意差なしなら⑤に進まないで終了する、逐次検定法の一つである。
- ⑤ $k-1$ から 2 になるまで、③~④を繰り返す。

【例 2】 $t_4 \sim t_2$ はダネット法の場合と同値である。

$$\begin{cases} t_4 = 3.45424 > w(4, 16, 0.01) = 2.66, \\ t_3 = 2.95701 > w(3, 16, 0.01) = 2.64, \\ t_2 = 0.62579 < w(2, 16, 0.05) = 1.75 \end{cases}$$

前 2 者は 1% 未満で有意、最後は有意差なしとなる。逐次でないダネット法では t_3 は 5% 有意であった。

【結論 3】 用量効果の判定にダネット法を用いてもよいが、ウィリアムズ法の方がやや優れている。

8. クラスカル・ウォリス法

正規分布に従わないかまたは例数が小さくて正規性が不明な場合や、分散の一樣性が否定された場合、ノン・パラメトリック技法 (non-parametric methods) の一つである、クラスカル・ウォリス (Kruskal-Wallis, 以下 KW と略す) 法²⁴⁾ を用いることになる。ただし前述したように、過去の文献を検索する必要があるし、対数変換についても適用を検討してほしい。

KW 法で有意差が認められた場合、従来はやはりノン・パラメトリック技法の一つである、マン・ホイットニー (Mann-Whitney, 以下 MW と略す)²⁵⁾ 法を用いて、繰り返して検定を行うのが一般的であった。しかしながら、この手順でもやはり FWE¹⁰⁾ を犯す可能性が高い。KW 法に続いて、前述したシェフェ法及びチューキー法を適用するソフト²⁶⁾ が知られている。後者の直接確率はやや甘いので、F 近似による修正を加えた。

表 8 に掲げた【例 3】について検討する。全標本 N の中で小さい方からの順位を表 9 に示した。繰り返し数が一定でなくとも汎用できる、KW 法の検定手順を下方に示した。 $H=12.517$ と有意 ($p < 0.01$) である。種々の多重比較法について、検討した結果を表 10 に示

表 8 【例 3】 Kruskal-Wallis 法の仮想例

因子	1	2	3	4	5	平均
A	7	9	10	8	8	8.4
B	11	10	14	13	12	12.0
C	10	13	18	17	14	14.4
D	13	16	13	18	15	15.0
						12.5

注) 一元配置分散分析法を適用すると、
 $F_0 = 9.645 > F_{16}^3(0.001) = 9.006$
 $(p < 0.001)$

表 9 【例 3】 の順位表

因子	1	2	3	4	5	平均順位
A	1	4	6	2.5	2.5	3.2
B	8	6	14.5	11.5	9	9.8
C	6	11.5	19.5	18	14.5	13.9
D	11.5	17	11.5	19.5	16	15.1
						10.5

【Kruskal-Wallis 法の手順】

- ① 全標本の中で小さい方から順位 r_{ij} をつける。
- ② タイのものがあつたら、平均的順位とする。
- ③ 各水準について、平均順位 \bar{r}_i を算出する。全体の平均順位 \bar{R} は $(N+1)/2$ である。
- ④ 分散を V 、自由度 $k-1$ の統計量を H とすると、

$$V = \sum \sum (r_{ij} - \bar{R})^2 = \sum \sum r_{ij}^2 - \frac{N(N+1)^2}{4}$$

$$H = (N-1) \sum_i (\bar{r}_i - \bar{R})^2 / V = 12.517 > \chi^2_3(0.01) = 11.345 (p < 0.01)$$

表 10 ノン・パラメトリック法での多重比較

因子水準	MW 法*	Scheffé 法	Tukey 法	B-不等式#
A-B	<i>N.S.</i>	0.3688	0.3218	0.4551
A-C	$p < 0.05$	0.0405	0.0481	0.0240
A-D	$p < 0.01$	0.0166	0.0254	0.0082
B-C	<i>N.S.</i>	0.7491	0.6931	> 1
B-D	<i>N.S.</i>	0.5657	0.5038	0.9239
C-D	<i>N.S.</i>	0.9913	0.9872	> 1

注) *Mann-Whitney 法, #Bonferroni の不等式

表 11 多重比較法のまとめ

検定方法	特徴及び適用
Fisher の最小有意差検定法	有意差は出やすいが、多重仮説過誤を犯す恐れが大きい。 水準数 k が大きくなるほど、有意差が出にくくなる。 繰り返し数が一定でなくてよい。保守的と言われている。
Bonferroni の不等式	
Scheffé 法	
Tukey 法	} Kruskal-Wallis 法に続いて適用できる。 繰り返し数が一定でなければならない。
Tukey-Kramer 法	
Student-Newman-Keuls 法	Tukey の修正法で、すべての組合せの比較に適用される。
Bonferroni の t 検定	原理、適用は Tukey-Kramer 法と同じである。
Dunnett 法	対照群との比較に用いるが、不合理な場合もある。
Williams 法	対照群との比較に際して、最良の方法と言われている。
Duncan の多範囲検定法	用量効果を判定するには、最良の方法と言われている。 生物学の分野で多く用いられている。

因子水準	A	B	C	D
A		0.075870	0.002167	0.000882
B	-3.6		0.328908	0.166212
C	-6.0	-2.4		0.969490
D	-6.6	-3.0	-0.6	

注) 太枠内は直接確率 p 値, ゴチックは平均値の差を示す。

図 2 【例 3】に Tukey-Kramer 法を適用した場合 (または Student-Newman-Keuls 法)

した。

A-C の有意性はいずれの方法でも 5% 未満である。A-D は MW 法によると 1%, シュフェ法及びチューキー法では 5% 未満で有意と判断される。MW 法と KW 法では順位の付け方が異なるため、FWE を犯している可能性を否定できない。また、HALBAU²⁷⁾ ではボンフェローニの不等式による結果も打ち出され、直接確率 p 値を 6 倍して示した。A-D は 1% 有意と判断できるものの、1 を超える確率もあるので不合理である。

なお、念のため TK・SNK 法も適用し、その結果を図 2 に示した。A-C は 1%, A-D も 0.1% 未満で有意と判断できる。ノン・パラメトリック技法は、検出力²⁾ に劣ることが理解できるであろう。

【結論 4】KW 法で有意差が認められた場合、シュフェ法やチューキー法を応用して多重比較を行うが、正規分布に従う場合より検出力が落ちる。

9. 生物学で適用される多重比較法

医学の分野ではあまり適用されないが、生物学ではダ

ソカンの多範囲検定 (Duncan's multiple range test)²⁸⁾ がよく知られている。SNK 法をさらに修正した方法と言われている。繰り返し数が一定でなく、しかも 3~5 例と極少数の場合にも適用できる。

ただし、正規性及び分散の一様性が仮定できる場合に限られるのが欠点である。パートレットの検定に続いて多重比較を実施できるソフト²⁹⁾ が利用できる。かつてレフェリーからこの方法が指定された例³⁰⁾ もあるので、参考までに紹介した。

おわりに

種々の検定法は絶対的なものではない。あくまでも、いかに少ない誤差で自分のデータをまとめられるか、にかかっていると断言できる。日本医事新報社¹⁾ からの依頼を契機に、これまで筆者らも疑問を感じていた、種々の多重比較法について、直接確率 p 値による検討を試みた。すべての方法について網羅したとはいいがたいが、種々の方法について表 11 にまとめを示した。

実用ソフトもできるだけ多く紹介したが、Macintosh の StatMate は、F 分布の % 点などの精度が低く、使い勝手にも工夫の余地が残る。他のプログラムでも一長一短があり、本編の例を用いて事前に検討されたい。

なお、検定方法の名称は専門の辞典³¹⁾ に準じてなるべくカタカナ表記とし、できるだけ原著を示した。若手研究者の参考になれば幸いである。

参考文献

1) 遠藤和男: 一元配置分散分析法と多重比較の適用について。日本医事新報, No. 3719: 169~170, 1995.

- 2) 遠藤和男, 山本正治: 医統計テキスト, 西村書店(新潟), 1992.
- 3) 遠藤和男: 医学における統計学の応用について(第2編) 平均値の差の検定を中心として. 新潟医誌, **104(2)**: 78~85, 1990.
- 4) **Bartlett, M.S.**: Properties of sufficiency and statistical tests. Proc. Royal Soc., **A. 160**: 268~282, 1937.
- 5) **Hartley, H.O.**: The maximum F -ratio as a short-cut test for heterogeneity of variance. Biometrika, **37**: 308~312, 1950.
- 6) **Cochran, W.G.**: The distribution of the largest of a set of estimated variances as a fraction of their total. Ann. Eug., **11**: 47~52, 1941.
- 7) 田中 豊, 垂水共之, 編: パソコン統計解析ハンドブック, Ⅲ実験計画法編, 共立出版(東京), 1986.
- 8) **Grubbs, F.E.**: Sample criteria for testing outlying observations. Ann. Math. Stat., **21**: 27~58, 1950.
- 9) 角田正史, 中平浩人, 遠藤和男, 山本正治: 小児肥満の要因に関する検討—肥満度の親子相関および環境の影響について—, 日本公衛誌, **41**: 1032~1036, 1994.
- 10) 吉村 功, 大橋靖雄, 編: 毒性試験講座⑩毒性試験データの統計解析, 地人書館(東京), 1992.
- 11) **Fisher, R.A.**: The design of experiment. (8th ed.) Oliver and Boyd (Edinburgh), 1966.
- 12) 榎 日出夫: 応用マッキントッシュ1, StatViewによる解析, 金原出版(東京), 1994.
- 13) 丹後敏郎(古川俊之監修): 新版医学への統計学, 朝倉書店(東京), 1993.
- 14) 竹内 啓, 編: 統計学辞典, 東洋経済(東京), 1989.
- 15) **Scheffé, H.**: A method for judging all contrasts in the analysis of variance. Biometrika, **40**: 87~104, 1953.
- 16) **Tukey, J.W.**: Comparing individual means in the analysis of variance. Biometrics, **5**: 99~114, 1949.
- 17) **Kramer, C.Y.**: Extension of multiple range tests to group means with unequal number of replications. Biometrics, **12**: 307~310, 1956.
- 18) 統計数値表編集委員会, 編: 簡約統計数値表, 日本規格協会(東京), 1977.
- 19) **Keuls, M.**: The use of the 'studentized range' in connection with an analysis of variance. Euphytica, **1**: 112~122, 1952.
- 20) **Glantz, S.A.**: Primer of biostatistics (3rd ed.), McGraw-Hill Inc. (New York), pp91~106, 1992.
- 21) **Dunnnett, C.W.**: New tables for multiple comparisons with a control. Biometrics, **20**: 482~491, 1964.
- 22) 佐野正樹, 岡山佳弘: Dunnnett の片側・両側検定のP値計算プログラムおよび高精度パーセント点. 医薬安全性研究会会報, **No. 32**: 21~31, 1990.
- 23) **Williams, D.A.**: A test for differences between population means when several dose levels are compared to a zero dose control. Biometrics, **27**: 103~117, 1971.
- 24) **Kruskal, W.H. and Wallis, W.A.**: Use of ranks in one-criterion variance analysis. J. Amer. Stat. Assoc., **47**: 782~790, 1975.
- 25) **Mann, H.B. and Whitney, D.R.**: On a test of whether one of two random variables in stochastically larger than the other. Ann. Math. Stat., **18**: 50~60, 1947.
- 26) 白旗慎吾: パソコン統計解析ハンドブック, IVノンパラメトリック編, 共立出版(東京), 1987.
- 27) 高木廣文: HALBAU-4 マニュアル, II. 基礎統計学, pp135~143, 現代数学社(京都), 1994.
- 28) **Duncan, O.D.**: Multiple range and multiple F test. Biometrics, **11**: 1~42, 1955.
- 29) 石居 進: ライフサイエンス・パソコンシリーズ3, BASICによる統計処理, pp133~143, 培風館(東京), 1983.
- 30) **Hanyu, S., et al.**: Distribution of serotonin-immunoreactive paraneurons in the lower urinary tract of dogs. Amer. J. Anatomy, **180**: 349~356, 1987.
- 31) 千葉大学統計グループ, 訳: ケンドール統計学用語辞典, 丸善(東京), 1987.

(平成7年8月31日受付)