

植物防疫基礎講座

多重比較法とその選び方（2）

多重比較法の主な種類

農林水産省農業環境技術研究所 やま山 むら村 こう光 じ司

はじめに

実験においてはデータはある程度は何らかの形で構造化、定量化されていることが多い。したがって、データ解析にあたっては、まず前回に述べたような構造化・定量化された分散分析を考慮してみるべきだと思われる。分散分析モデルの総平方和が、構造に基づいて自由度1の平方和に分割されれば、もはや多重比較法の出番はない。しかし、完全に等価な三つ以上の処理に出くわしたなら（すなわち自由度が2以上の平方和に出くわしたなら）、その場合の検定には多重比較法を用いることになる。例えば、前回用いたデータ（表-1）で、もしA地域に三つのトラップが設置されていたとしたら、この三つのトラップ間の差に関しては多重比較法で検定を行わなければならない。同様に、もし調査が三つの地域（A地域、B地域、C地域）で行われていたとしたら、この三つの地域の間の差に関しては多重比較法で検定を行わなければならない。

I 1段階検定による多重比較法

一般に、三つの値が「すべて同じもの」であるか否かを判定する際に二つの方針が考えられる。一つめの方針は、三つの値のばらつき（分散）を調べて、この分散が大きければ「この三つの値はすべて同じものとはいえない」と宣言するという方針である。もう一つの方針は、三つの値の最大値と最小値の差を計算し、この差が大きければ「この三つの値はすべて同じものとはいえない」と宣言するという方針である。前者の判定方針に基づい

て検定を行う方式がF検定である。一方で、後者の方針に基づく方法は「スチューデント化された範囲を用いる検定」である。分散分析法は前者の考え方を基幹とする分析体系であるが、多重比較の問題においては後者の考え方も力を発揮している。というのも、後者の方法は、その最大値・最小値のみを用いるという性質から、次のような重要な特徴を備えているからである。「 k 個の処理平均のすべての二つの組み合わせ（ ${}_kC_2$ 個）について、その二つの処理平均の差を比較するときに、本当はその二つに差がないにもかかわらず誤って有意差ありと判定してしまう確率を常に α におさえる。」すなわち、前回に述べた多重比較法の条件を満たしている。したがって、スチューデント化された範囲の検定を用いて、どの処理とどの処理の間に有意差があるかを判定することができる。この方法はTUKEYの方法といわれ、多重比較法の中で標準とされるものである。

いま、誤差の自由度を ν 、誤差平均平方を s_e^2 とし、 k 個の処理平均値 m_i （ $i=1, 2, \dots, k$ ）を比較する場合を考えよう。それぞれの処理が同じ数 n 個のデータを含むとする。あらゆるデータの真の誤差分散を σ^2 とすれば、おのおの処理平均値の真の分散はいずれも、 σ^2/n である。したがって、次の値は分散1の正規分布に従う。

$$\frac{m_i}{\sigma/\sqrt{n}} \quad (i=1, 2, \dots, k)$$

もし、 k 個の処理の真の平均値がすべて等しいならば、 k 個の $m_i/(\sigma/\sqrt{n})$ の値はすべて分散1の同じ正規分布に従う。分散1の同じ正規分布に従うならば、その k 個の値の最大値と最小値の差（範囲）が極端に大きくなるはずはない。したがって、あらかじめ「 k 個のサンプルを取り出したとき、 α 以下の確率でしか生じないような限界の範囲」を「分散1の正規分布」に関して計算して数表にしておけば、この表中の値と、実際の最大値と最小値の差を比較することによって検定ができる。 k 個の $m_i/(\sigma/\sqrt{n})$ の最大値と最小値の差が表中の限界値よりも大きければ「 k 個の処理の真の平均値はすべて等しい」という仮説を棄却するわけである。ただし、実際には σ^2 の値は不明なので平均誤差平方 s_e^2 を用い、 $m_i/(s_e/\sqrt{n})$ の最大値と最小値の差を使う。このため平均誤差平方の自

表-1 ハスモンヨトウのフェロモントラップ誘殺実験結果

地域	トラップ名	各月の総誘殺数			
		5月	6月	7月	8月
A	A1	10	26	45	356
	A2	8	16	55	341
B	B3	16	48	112	874

On the choice of multiple comparison procedures(2).
By Kohji YAMAMURA

由度ごとに別の限界値を検定に用いなければならない。結局、この限界値は有意水準 α 、処理個数 k 、誤差平均平方の自由度 ν の三つの値に依存する。この限界値を $Q(\alpha, k, \nu)$ と記し、「 k 個の正規確率変数の範囲を自由度 ν でスチューデント化した分布の α 水準限界値」と表現する。

この方式を利用し、結局、処理 i と処理 j の平均の差、 $m_i - m_j$ が次の不等式を満たせば処理 i と処理 j には α 水準で有意差があると判定できることになる。なぜなら、「処理 i と処理 j の効果が真に等しい」という条件を含むあらゆる帰無仮説のうちで「 k 個の処理の効果が真にすべて等しい」という帰無仮説が最も大きな有意確率を示すからである。

$$|m_i - m_j| > \left(\frac{s_e}{\sqrt{n}} \right) Q(\alpha, k, \nu) \quad (3)$$

先のフェロモントラップのデータについて、三つのトラップが同じ地域に設置されていた場合を想定して、TUKEYの方法で多重比較を行ってみよう。前回にならって、対数変換後に分散分析を行う。その分散分析表を表-2に示してある。今の場合、二つの処理を比較する組み合わせは₃C₂=3通りある。つまりA1とA2、A1とB3、A2とB3である。このような二つごとの比較は「対比較（ついひかく、pairwise comparison）」と呼ばれている。今の場合3通りの pairwise comparison が存在するわけである。 $\alpha = 0.05$ 、 $k=3$ 、 $\nu=6$ を用いると、 $Q(0.05, 3, 6)$ は 4.339。処理ごとのデータ数は4、 $s_e = \sqrt{0.0282} = 0.1679$ である。したがって上式(3)の右辺の値は 0.3644 となる。今は比較したい平均値は、トラップA1について $m_1 = 3.8106$ 、トラップA2について $m_2 = 3.6728$ 、トラップB3について $m_3 = 4.5338$ であるから、差が 0.3644 以上となる組み合わせを考えれば結論は次のようになる。「0.05 水準において、トラップA1とA2の間には差がない。トラップA1とB3の間には差がある。トラップA2とB3の間にも差がある。」このことから、いまの場合、間接的に「A地域とB地域には有意差がある」ことがいえる。厳密に計算すると、トラップ

A1とトラップB3との間の有意確率は 0.0021 である。前回の分散分析によれば本当はA地域とB地域には 0.0003 水準で差があるのだが、データを持つ構造を無視したために、検出力がやや落ちていることがわかる。

TUKEY法は、もともと処理ごとのデータ数が等しい場合を想定した方法であるが、処理ごとのデータ数が異なる場合には、次の判定基準を用いることにより、同じように検定を行うことができる。第 i 処理のデータ数を n_i とすれば、

$$|m_i - m_j| > s_e \sqrt{\frac{1}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} Q(\alpha, k, \nu) \quad (4)$$

これは、上の(3)式の n の代わりに二つの処理データ数 (n_i, n_j) の調和平均を使った式である。この式を用いた場合、厳密には検定基準は α よりも少し小さくなり、帰無仮説をやや棄却しにくくなる、つまりやや「保守的」な検定となる (HAYTER, 1984)。しかし、有意水準は α 以下に保たれているには違いないので、一応問題はない。この(4)式を用いる方法は TUKEY-KRAMER法あるいは KRAMER 法と呼ばれる。

多重比較の計算でよく出てくるもう一つの方法は、SCHEFFÉ法と呼ばれる方法である。いま、 k 個の処理の中から二つのグループを取り出して、その平均値の違いを比較してみよう。それぞれ処理の真の値を $\mu_1, \mu_2, \dots, \mu_k$ とすれば、例えば μ_1, μ_8 の平均値と μ_2, μ_3, μ_5 の平均値を比較する場合、次の値がゼロであるか否かを検定することと同じである。

$$\frac{\mu_1 + \mu_8}{2} - \frac{\mu_2 + \mu_3 + \mu_5}{3}$$

この推定値は、

$$\frac{m_1 + m_8}{2} - \frac{m_2 + m_3 + m_5}{3} \quad (5)$$

このような比較を日本語では「対比（たいひ）」と呼ぶが、先ほどの「対比較（ついひかく）」という言葉と1字違いで非常にまぎらわしい。これは「二つのグループ間の比較」であるから、原語の「contrast」という言葉のほうがずっとわかりやすいように思われる。

第 i 処理のデータ数を n_i とかけば、上の(5)の値の分散推定値 s_c^2 は、各 m_i の分散推定値 s_e^2/n_i に係数の2乗をかけて足せば良いので

$$s_c^2 = s_e^2 \left\{ \left(\frac{1}{2} \right)^2 \left(\frac{1}{n_1} + \frac{1}{n_8} \right) + \left(\frac{1}{3} \right)^2 \left(\frac{1}{n_2} + \frac{1}{n_3} + \frac{1}{n_5} \right) \right\}$$

である。いま、自由度 $(k-1, \nu)$ の F 分布の上側確率 α の値を $F(\alpha, k-1, \nu)$ で表そう。SCHEFFÉは、「 k 個のすべての処理平均値に差がない」という帰無仮説の基で、このようなグループ分け（重み付き平均を含む）のいか

表-2 データ構造を無視した分散分析表（対数変換後）

変動要因	平方和	自由度	平均平方	F	確率
モデル全体	25.095	5	5.019	177.96	<0.0001
ブロック(月)	23.384	3	7.795	276.37	<0.0001
トラップ	1.711	2	0.856	30.34	0.0007
誤差	0.169	6	0.0282		
全体	25.264	11			

なる組み合わせを考えても、「その平均値の差が $s_c\sqrt{(k-1)F(\alpha, k-1, \nu)}$ を超える確率」が α 以下になることを証明した。この関係を利用し、例えば上の (5) 式の絶対値が $s_c\sqrt{(k-1)F(\alpha, k-1, \nu)}$ より大きければ、「第 1, 8 処理の平均値は、第 2, 3, 5 処理の平均値と α 水準で有意に異なる」ということができる。一般に (5) 式のようなグループ間の比較 (contrast) を形式的に表現すれば、

$$\sum_{i=1}^k c_i m_i$$

ただし、ここに

$$\sum_{i=1}^k c_i = 0$$

が成り立っているはずである。 $\sum_{i=1}^k c_i m_i$ の推定分散 s_c^2 は

$$s_c^2 = s_e^2 \sum_{i=1}^k \frac{c_i^2}{n_i}$$

$\sum_{i=1}^k c_i m_i$ が次の不等式を満たせば、二つのグループの間に有意な差があると判定する。

$$\left| \sum_{i=1}^k c_i m_i \right| > s_c \sqrt{(k-1)F(\alpha, k-1, \nu)} \quad (6)$$

ただし、SCHEFFÉ (1953) 自身も触れているように、SCHEFFÉ の方法は、二つの処理の比較 (pairwise comparison) のみを問題にする場合には、検出力が弱いので使うべきではない。SAS も含め、統計パッケージで SCHEFFÉ の方法の出力があるときは、それは pairwise comparison に SCHEFFÉ 法を使った計算であることが多い。この場合、TUKEY 法あるいは TUKEY-KRAMER 法が出力できるならば SCHEFFÉ 法の出力はまず不用であると考えてよい。なお、ソフトウェアによっては、BONFERRONI 法や DUNN-SIDÁK の方法による多重比較の結果も同時に出力してくれるものがある。しかし、これらの方法も SCHEFFÉ 法と同じく検出力が弱いだけなので、やはり使用は薦められない。

再びフェロモントラップのデータを取り上げ、SCHEFFÉ の方法を使ってみよう。いかなる contrast を検定してもよいが、先ほどの分析と比較するために、例として A 地域の平均値と B 地域での平均値の間の差を取り上げて検定してみよう。調べたい差の推定値は

$$\frac{3.8106 + 3.6728}{2} - 4.5338 = -0.7921$$

である。0.05 水準での判定基準は $s_c\sqrt{(k-1)F(\alpha, k-1, \nu)}$ に $s_c = \sqrt{s_e^2} = 0.10284$, $k=3$, $\alpha=0.05$, $\nu=6$ を入れて 0.3298 である。0.7921 はこの値よりも大きいので、「A 地域と B 地域は 0.05 水準で有意差がある」と結論できる。正確に計算すると、この値の

有意確率は 0.0008 である。前回の分散分析によれば本当は A 地域と B 地域の間には 0.0003 水準で有意差があったのだが、データ構造を無視したために、検出力が低下している。

II 多段階検定 (step down 型) による多重比較法

処理平均値を大きさの順にならべ、両端から検定範囲を狭めていきながら、逐次に検定を行うという形の検定法がある。このような検定手法は step down 型の多段階検定と呼ばれており、手続きは面倒ではあるが、その代わり 1 段階検定法に比べてやや高い検出力を持っている。例えば、 k 個の処理平均の間の多重比較を行うとき、次のようなステップをたどる。

- ① k 個の平均値の間で検定を行う。検定結果が有意ならば片方どちらかの端のデータを除いて $(k-1)$ 個の中で検定を行い、これら $(k-1)$ 個の処理間に有意差があるかどうかを調べる。また、逆の端のデータを除いて同様に検定を行う。
- ② 有意差の検出されたグループについては、さらに端のデータを除いて $(k-2)$ 個の間の検定を行う。もし有意差がなければ、そのグループについてはさらに分割せず、検定をストップする。
- ③ ②の操作を繰り返す。

いま 10 個の処理平均値を大きさの順にならべて 6 個と 4 個に区分けしたとしよう。このとき、6 個の間の有意性の検定と、4 個の間の有意性の検定を同時に行うことになる。仮に有意水準を α でなくて、それぞれ $\alpha_1 = 1 - (1-\alpha)^{(6/10)}$ と $\alpha_2 = 1 - (1-\alpha)^{(4/10)}$ として検定を行ってみよう。このとき「本当は、この 6 個内に差はなく、かつ 4 個内にも差がない」にもかかわらず、どちらかの検定でたまたま有意差が出てしまう確率というのは、1.0 から「どちらの検定でも有意差が出ない確率」を引いたものである。仮に二つの検定が独立であるならば、その確率は、

$$\begin{aligned} & 1 - (4 \text{ 個の間に有意差が出ない確率}) \times (6 \text{ 個の間に有意差の出ない確率}) \\ & = 1 - (1-\alpha_1)(1-\alpha_2) \end{aligned}$$

これは書き直すと

$$\begin{aligned} & 1 - \left\{ 1 - \left[1 - (1-\alpha)^{(4/10)} \right] \right\} \\ & \quad \left\{ 1 - \left[1 - (1-\alpha)^{(6/10)} \right] \right\} \\ & = \alpha \end{aligned}$$

つまり、「どちらかの検定でたまたま有意差が出てしまう確率」は見事に α に一致する。このことから一般に p 個

の比較を行う際に有意水準を

$$\begin{aligned} \alpha' &= 1 - (1 - \alpha)^{(p/k)} & p < k-1 \\ &= \alpha & p \geq k-1 \end{aligned}$$

として検定を行うと、全体の有意水準は常に正しく α に保たれることがわかる。結局、step down 法で多段階検定を行う際にはこの有意水準を用いればほぼ正しい検定法となる。これは RYAN 法と呼ばれているものである。従来使われていた DUNCAN 法、NEWMAN-KEULS 法等の step down 法は、この観点からすれば「誤った方法」であり、これらの方法を使った論文は現在ではもはや受理されない可能性が高いであろう。ちなみにこれらの方法では次の有意水準を用いていた。

$$\begin{aligned} \text{DUNCAN 法} & \quad \alpha' = 1 - (1 - \alpha)^{p-1} \\ \text{NEWMAN-KEULS 法} & \quad \alpha' = \alpha \end{aligned}$$

上の RYAN の α' 有意水準を用いて、繰り返し「スチューデント化された範囲を用いた検定」を行うことにより、先の TUKEY 法よりも検出力の高い検定を行うことができる。SAS ではこの RYAN 法を REGWQ 法と呼んでいる。ただし、SAS の REGWQ では、処理ごとのデータ数が一定でないときにデータ数の調和平均を用いて計算を行うようになっている (SAS Institute, 1990)。これはかなり具合が悪いので、SAS の REGWQ は「すべての処

理ごとのデータ数が同じ」場合に限り使用すべきである。また、「スチューデント化された範囲を用いた検定」を行う代わりに、それぞれの段階で F 検定を行うのもよい。こちらの方式は SAS では REGWF と呼んでいる。REGWQ と REGWF のどちらのほうが検出力が高いかは、真の処理平均の位置関係によって異なるであろう。両者の優劣は、検出力を評価する基準によっても異なるようである (HOCHBERG and TAMHANE, 1987)。

なお、RYAN 法とは逆に、範囲を広めてゆきながら多段階検定を行うやり方もある (step up 型)。WELSCHE (1977) が開発した step up 法は、有意水準を正確に α に調整しながら、かつ RYAN 法よりも少し高い検出力を保証する。しかし、WELSCHE 法は、その限界値の計算がやっかいであり、SAS では対応していない。数表としては佐々木 (1987) に掲載されているものが入手しやすいであろう。

引 用 文 献

- 1) HAYTER, A. J. (1984): Ann. Statist. 12: 61~75.
- 2) HOCHBERG, Y. and A. C. TAMHANE (1987): Multiple Comparison Procedures. Wiley.
- 3) 佐々木昭博 (1987): 植物防疫 41: 289~294.
- 4) SAS Institute (1990): SAS/STAT ユーザーズガイド Release 6.03 Edition. SAS 出版局.
- 5) SCHEFFÉ, H. (1953): Biometrika 40: 87~104.
- 6) WELSCHE, R. E. (1977): J. Amer. Statist. Assoc. 72: 566~575.

学 界 だ よ り

○第 11 回「植物保護とバイオテクノロジー」シンポジウム開催のお知らせ

主 催：日本農薬学会・農薬バイオテクノロジー研究会
日 時：平成 5 年 10 月 22 日 (金) 10:30~17:00

場 所：明治大学大学会館・6 階、JR「お茶の水」駅お茶の水橋口 (新宿寄り) 下車、坂を 3 分下った左側。電話 03-3296-4276, 4277

テーマ：「生物機能利用の最前線」

プログラム：

1. 寒地型芝草のエンドファイト・その特性と利用
(株)ジャパン・ターフグラス) 笠原春夫氏
2. 微生物除草剤の現状

(三井東圧化学(株)ライフサイエンス研)

郷原雅敏氏

3. 生物農薬の開発過程における諸問題—病害防除を中心として—

(日本植物防疫協会研究所) 田中 薫氏

4. 寄生蜂による寄生制御の分子機構
(北大、低温科学研) 早川洋一氏

5. 昆虫の共生微生物
(蚕糸・昆虫農業技術研) 野田博明氏

6. コガネムシに殺虫活性を持つ新規 Bt
(東京農工大、農、応用生物科学) 佐藤令一氏

参加費：3,000 円

連絡先：〒305 つくば市観音台 2-1-2

農業生物資源研究所 宇垣正志氏
TEL 0298-38-7424, FAX 0298-38-7408