

農業研究と多重比較手法

三輪哲久
農業環境技術研究所

1 はじめに

筆者は30年以上にわたって、農業技術開発（および品質管理）の現場で統計手法の開発と適用に携わってきた。その間、現場の農業技術研究者には極めて好評であるにもかかわらず、統計理論家や論文のレフリーには不評である（好まれない）手法が2つあることに気づいた。それは

(1)「Duncan (1955)の多重範囲検定」と(2)「直交表による実験計画」

である。Duncanの方法はDMRT (Duncan's Multiple Range Tests) という略称で親しまれている。Ryan & Woodall (2005)によると、Duncan (1955)は、2005年までの被引用回数のランク付けで第3位である。最近15年間のみに限定するとDuncan (1955)は姿を消し、一方、Benjamini & Hochberg (1995)のFDR法が第7位と上昇中である。Duncanの方法がファミリー単位の過誤率 $FWER$ を保障しないことはいまや良く知られている。果たして、Duncanの方法は使ってはいけないのであろうか。

一方、 $FWER$ を保障する方法としては、Tukey (1953)の方法、Dunnett (1955)の方法、Scheffé (1959)の方法が有名である。これらの方法はどのように使い分けるべきなのか。

本稿では、農業技術研究の現場における多重比較手法の適用に関して、筆者が日ごろ感じていることについて述べ、また筆者が取り組んできた順序制約のもとでの多重比較手法について紹介する。統一的な内容とはなっていないことをお許しいただきたい。

2 多重比較の問題のタイプ

Duncan (1955)は、表1の大麦のデータを用いて処理の対比較 (pairwise comparisons) を行なうための多重範囲検定を提案した。

表 1. Barley yields (bushels/acre)

品種	A_1	A_2	A_3	A_4	A_5	A_6	A_7
平均収量	49.6	71.2	67.6	61.5	71.3	58.1	61.0

表 2. Analysis of Variance

変動因	自由度	平方和	分散	F比	p値
品種 A	6	2201.73	366.96	4.61**	0.002
ブロック R	5	709.60	141.92		
誤差 E	30	2387.43	79.58		
全体	41	5298.77			

一般に、 k 個の処理 A_1, \dots, A_k からなる一元配置分散分析モデルの処理平均値と分散推定値を

$$\bar{y}_i \sim N(\mu_i, \sigma^2/n_i) \quad (1 \leq i \leq k), \quad \hat{\sigma}^2 \sim \sigma^2 \chi^2(\nu)/\nu$$

とする。上記の例では $n_i \equiv n = 6$, $\hat{\sigma}^2 = 79.58$ ($\nu = 30$) である。

このようなデータが与えられたとき，Tukey の方法，Dunnett の方法，Scheffé の方法のどれを使うべきかという質問を受けることがよくある。混乱の原因の一つは，いくつかの教科書で，同じデータに対して複数の解析を求めていることである。たとえば，次のような具合である。

- (1) 表 1 のデータに対して，全てのペアの対比較を行なえ。
- (2) 表 1 で処理 A_1 が対照品種であるとして Dunnett の検定を行なえ。
- (3) 品種群 $\{A_1, A_2, A_3\}$ と $\{A_4, A_5, A_6, A_7\}$ とは親の系統が異なっている。対比

$$\frac{\mu_1 + \mu_2 + \mu_3}{3} - \frac{\mu_4 + \mu_5 + \mu_6 + \mu_7}{4} \quad (1)$$

に対して Scheffé の検定を実行せよ。

しかし，これらの方法は使う目的が異なっている。実験を行なう前に研究の目的が決まっているならば，「Tukey 法を使うべきか，Dunnett 法を使うべきか」を迷うことはないはずである。処理群間の対比較 (pairwise comparisons) を行なうことが実験の目的であれば，対比較のための多重比較手法の中から，適切に過誤率を制御する方法を選択すればよい。

表 3. 多重比較の問題のタイプと手法

対比較	対照との比較	対比	要因実験	回帰分析
最小有意差法	最小有意差法	t 検定	主効果	多項式回帰
Duncan 法	Dunnett 法 (両側)	Scheffé 法	交互作用	成長曲線回帰
Newman-Keuls 法	Dunnett 法 (片側)	Bartholomew 法	直交表	多次元回帰
R-E-G-W 法	逐次 Dunnett 法
Tukey 法	(修正) Williams 法			
Hayter 法				

なお，最初は全ての処理の対比較を行なう目的で実験を計画し，Tukey 法で解析したとする。しかし，データを見た後で (1) 式のような対比 (contrast) に興味を持つことがあるかもしれない。このとき，(1) 式の対比を Scheffé 法で検定するとファミリー単位の過誤率 $FWER$ は正しく保障されない。対比に対する Tukey 法を用いなければならない。

3 FDR 法，Fisher の PLSD 法，Duncan の方法

3.1 ファミリー単位の過誤率

多重比較 (多重検定) とは，その名のとおり多重の (複数の) 検定を行なうことである。いま， m 個の未知母数 $\theta_1, \dots, \theta_m$ に興味があり， m 個の帰無仮説

$$H_i^0: \theta_i = \theta_i^0, \quad 1 \leq i \leq m$$

に対する検定を考える。たとえば，上記 2 節に例示した対比較の場合は， $m = k(k-1)/2$ であり，

$$H_1^0: \mu_1 - \mu_2 = 0, H_2^0: \mu_1 - \mu_3 = 0, \dots, H_m^0: \mu_{k-1} - \mu_k = 0$$

と表わされる。 m 個の帰無仮説の集合 $\mathcal{H} = \{H_1^0, \dots, H_m^0\}$ を帰無仮説のファミリー (family) とよぶ。 L を添え字の部分集合 $L \subset \{1, \dots, m\}$ として， L に含まれる添え字をもつ帰無仮説 H_i^0 ($i \in L$) が全て成り立っているという積仮説を $H_L^0 = \bigcap_{i \in L} H_i^0$ と表わす。 $L = \{1, \dots, m\}$ の場合，すなわちファミリー \mathcal{H} の全ての帰無仮説が成り立っている場合を完全帰無仮説 (complete null hypothesis)，

L が $\{1, \dots, m\}$ の部分集合の場合を部分帰無仮説 (partial null hypothesis) とよぶ。 k 水準の分散分析モデルに対比較を行なう場合、完全帰無仮説は $H_{1\dots k}: \mu_1 = \dots = \mu_k$ である。

特定の帰無仮説 H_i^0 に対しては、 H_i^0 が成り立っているときに H_i^0 を棄却することを第 I 種の過誤 (type I error) という。多重比較において、真に成り立っている帰無仮説が $H_L^0 = \cap_{i \in L} H_i^0$ のとき、 H_i^0 ($i \in L$) のうち、少なくとも 1 つ以上で第 I 種の過誤を犯す確率をファミリー単位の (第 I 種の) 過誤率 $FWER$ (family-wise error rate) と定義する。

実際には、仮説のファミリー \mathcal{H} のうち、どの帰無仮説が成り立っているかは未知である。完全帰無仮説 $H_{\{1, \dots, m\}}^0$ に対してのみ $FWER$ を α 以下に保障することを、弱い意味 (weakly) で $FWER$ を制御する (あるいは保障する) という。完全帰無仮説も含めて、あらゆる部分帰無仮説のもとで $FWER$ を α 以下に保障することを、強い意味 (strongly) で $FWER$ を制御するという。

対比較における Tukey 法や、対照との比較における Dunnett 法は、強い意味で $FWER$ を制御する方法である。

3.2 FDR 法

Benjamini & Hochberg (1995) は FDR (False Discovery Rate) とよばれる過誤率を提案した (エフディーアール, 誤検出率?, 偽検出率?)。

m 個の帰無仮説のファミリー $\mathcal{H} = \{H_1^0, \dots, H_m^0\}$ のうち、 m_0 個が真に成り立っているとすると ($0 \leq m_0 \leq m$)、検定結果は表 4 のように要約される。

表 4. m 個の帰無仮説に対する検定結果 (判定の数)

	有意ではないと 判定	有意と 判定	計
真の帰無仮説	U	V	m_0
偽の帰無仮説	T	S	$m - m_0$
計	$m - R$	R	m

R は棄却された帰無仮説の数で、観測可能な確率変数である。一方、帰無仮説 H_1^0, \dots, H_m^0 のうち、どの m_0 個が真に成り立っているかは未知 (m_0 の値自体も未知) であるから、確率変数 U, V, S, T は観測不能である。

間違っ棄却された帰無仮説の割合 (False Discovery Proportion)

$$Q = V/(V + S) = V/R \quad (R = V + S = 0 \text{ のときは, } Q = 0 \text{ とする})$$

を考える。この Q も確率変数であり、その期待値

$$Q_e = E\{Q\} = E\{V/(V + S)\} = E\{V/R\}$$

を FDR (False Discovery Rate) とよぶ。 FDR は次の性質を持つ。

(a) $m_0 = m$ (完全帰無仮説) のとき、 $FDR = FWER$ 。

(b) $m_0 < m$ のとき、 $FDR \leq FWER$ 。

FDR を α 以下に制御する方法は、性質 (a) により、完全帰無仮説のもとで (つまり、弱い意味で) $FWER$ を制御する。しかし一般の部分帰無仮説のもとでは、性質 (b) により、 $FDR \leq \alpha$ であっても、 $FWER \leq \alpha$ とは限らない。すなわち、 FDR 法は強い意味で $FWER$ を制御する方法ではない。そして、強い意味で $FWER$ を制御する方法よりも、 FDR 法は検出力が高くなる。

本稿の最初に述べたように、 FDR 法は最近急速に使われるようになった。本学会の昨年度の年会でも特別セッションとして取り上げられている。 FDR 法の基本的考え方、応用、参考文献については、昨年度年会の予稿集を参考にされたい。

3.3 Fisher の PLSD 法

対比較での $FWER$ の制御に関して、強い意味と弱い意味との中間に属する伝統的な方法が 2 つある。保護つき LSD (Protected LSD) 法と保護つき Duncan 法 (Protected Duncan) である。

処理 A_i と A_j の平均値の差 $\mu_i - \mu_j$ ($1 \leq i < j \leq k$) に対して、個別に t 検定

$$|\bar{y}_i - \bar{y}_j| > \hat{\sigma} \sqrt{1/n_i + 1/n_j} \cdot t(\nu; \alpha/2)$$

を実行する方法を最小有意差法 (Least Significant Difference, LSD 法) という。特定の帰無仮説 $H_{ij}^0: \mu_i = \mu_j$ に対する第 I 種の過誤率 (比較単位の過誤率, comparison-wise error rate, $CWER$) は α 以下になる。しかし、この検定を $m = k(k-1)/2$ 回実行すれば、そのどこかで第 I 種の過誤をおかす確率 $FWER$ は α よりはるかに高くなる。

まず最初に分散分析の F 検定で、完全帰無仮説

$$H_{1\dots k}^0: \mu_1 = \dots = \mu_k$$

を検定し、この完全帰無仮説が棄却された場合のみ、LSD 法を実行する方法を保護つき LSD 法 (protected LSD, PLSD, Fisher's PLSD) という。一方、完全帰無仮説の検定を実行することなく、LSD 法による検定を行なう方法を保護なし LSD 法 (unprotected LSD) という。PLSD 法は、弱い意味で $FWER$ を制御する。すなわち、完全帰無仮説 $H_{1\dots k}^0: \mu_1 = \dots = \mu_k$ が成り立っている場合、 F 検定で有意になる確率は α であるから、さらにいずれかのペアで、 $\mu_i = \mu_j$ が棄却される確率は明らかに α 以下である。しかし、たとえば $H_{2\dots k}^0: \mu_1 \gg \mu_2 = \dots = \mu_k$ のような部分帰無仮説のもとでは、最初の F 検定は、ほとんど確実に棄却されるので、 $\mu_2 = \dots = \mu_k$ のいずれかで間違っただけで第 I 種の過誤をおかす確率は、やはり α より大きくなってしまふ。つまり、PLSD 法は強い意味では $FWER$ を保障しない。

当然 Fisher も多重性には気づいており (Fisher, 1935)、次のような Bonferroni 型の調整について言及している。

Thus, in comparing the best with the worst of ten tested varieties, . . . out of 45 pairs We might require . . . the probability . . . as small as 1 in 900, instead of 1 in 20,

対比較に対して、強い意味で $FWER$ を保障する方法としては、Tukey 法 (アンバランストモデルでの Tukey-Kramer 法) および閉検定手順に基づく REGW 法がある (詳細については Hsu (1996), Hochberg & Tamhane (1987), 永田・吉田 (1997) などを参照)。

3.4 Duncan 法

Duncan (1955) の方法では、大きさ p の添え字の集合 $P = \{i(1), \dots, i(p)\}$ に対して、 p 個の処理平均が等しいという帰無仮説 $H_P^0: \mu_{i(1)} = \dots = \mu_{i(p)}$ を、水準

$$\alpha_p^D = 1 - (1 - \alpha)^{p-1} \tag{2}$$

で検定する。帰無仮説 $H_{ij}^0: \mu_i = \mu_j$ は、添え字 i と j をともに含む P ($i, j \in P$) に対する $H_P^0: \mu_{i(1)} = \dots = \mu_{i(p)}$ が全て棄却されたときに、棄却される。

(2) 式によると、帰無仮説 $H_P^0: \mu_{i(1)} = \dots = \mu_{i(p)}$ に対して、過誤をおかさない確率が $1 - \alpha$ の $p - 1$ 乗である。 α_p^D は $p > 2$ のとき α より大きくなり、 p とともに増加する。たとえば表 1 の大麦実験の例で、 $p = k = 7$ の場合、 $\alpha = 0.05$ とすると、 $\alpha_7^D = 1 - 0.95^6 = 0.264$ となる。

ここで、 $p-1$ は仮説 $H_p^0: \mu_{i(1)} = \dots = \mu_{i(p)}$ の自由度である。Duncan の方法は、つぎのような考え方に基づいている。たとえば、 $k=4$ とおりの処理が、品種 V について 2 水準 (V_1, V_2)、窒素 N について 2 水準 (N_1, N_2) の要因実験から構成されているとする。

処理	母平均	水準組み合わせ
A_1	μ_1	V_1, N_1
A_2	μ_2	V_1, N_2
A_3	μ_3	V_2, N_1
A_4	μ_4	V_2, N_2

このとき、処理間の 3 つの自由度に対応する 3 つの対比

$$\begin{aligned} V \text{ の主効果: } & \mu_1 + \mu_2 - \mu_3 - \mu_4 \\ N \text{ の主効果: } & \mu_1 - \mu_2 + \mu_3 - \mu_4 \\ V \times N \text{ の交互作用: } & \mu_1 - \mu_2 - \mu_3 + \mu_4 \end{aligned}$$

をそれぞれ α (たとえば 5%) の水準で検定すると、Duncan 法と同じ $FWER$ となる。

なお Duncan 法においても、通常は、先ず完全帰無仮説 $H_{1\dots k}^0: \mu_1 = \dots = \mu_k$ に対する F 検定が行なわれる。すなわち、弱い意味で $FWER$ は保障されている。

(2) 式 α_p^D の代わりに $\alpha_p^S \equiv \alpha$ を用いる方法が Student-Newman-Keuls 法 (SNK 法) であり、 $\alpha_p^R = 1 - (1 - \alpha)^{p/k}$ を用いる方法が Ryan-Einot-Gabriel-Welsch 法 (REGW 法) である。SNK 法は弱い意味で $FWER$ を制御し、REGW 法は強い意味で $FWER$ を制御する。

これらの弱い意味で $FWER$ を制御する方法は、 FDR の制御する方法と似た性質をもつ。たとえば Shaffer (2007) は、SNK 法が FDR を制御することを示している。

4 順序制約のもとでの多重比較

筆者は最近、順序制約のもとでの多重比較手法の研究に取り組んできた。以下では、筆者と共同研究者による最近の成果について紹介する。

帰無仮説 $H^0: \mu_1 = \dots = \mu_k$ に対して、処理平均 μ_i ($1 \leq i \leq k$) の間に何らかの順序関係が存在することが事前に分かっていたり、あるいは実験者が一方向のみに興味をもっていたりする場合は多い。たとえば用量反応関係において応答に単調性が仮定できる場合 (3) 式や、第 1 処理が対照処理 (control) で片側に興味がある場合 (4) 式などが考えられる。

$$H^A : \mu_1 \leq \dots \leq \mu_k \text{ (少なくとも 1 つは厳密な不等号)} \quad (3)$$

$$H^A : \mu_1 \leq \mu_i, \quad 2 \leq i \leq k \text{ (少なくとも 1 つは厳密な不等号)} \quad (4)$$

4.1 片側推測と両側推測の融合

$k=2$ の場合、帰無仮説 $H^0: \mu_1 = \mu_2$ と対立仮説 $H^A: \mu_1 < \mu_2$ に対して、片側検定の方が両側検定よりも検出力が高い。しかし、片側検定に対応する信頼区間の上限は無量大である。仮に対立仮説は片側であっても、両側信頼区間が求められる場合は多い。たとえば A_1 が毒性物質への無暴露群 (対照群)、 A_2 が暴露群である ($A_1 =$ 低暴露群、 $A_2 =$ 高暴露群でもよい) とすると、毒性を検出するためには片側検定が適切である。このとき、片側信頼区間は $\mu_2 - \mu_1$ の下限を与える。しかし、毒性は最大限どれくらい高くなるのかという $\mu_2 - \mu_1$ の上限に関する情報は全く与えられないことになる (単純に、論文のレフリーが検定だけではなく、両側信頼区間を示すことを要求することもある)。

このような問題に対して，Miwa & Hayter (1999) は次のような片側推測と両側推測の利点を融合した方法を提案した。

$$\bullet t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} < \bar{y}_2 - \bar{y}_1 \quad (\text{両側検定で有意}) \Rightarrow \bar{y}_2 - \bar{y}_1 - t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} \leq \mu_2 - \mu_1 \leq \bar{y}_2 - \bar{y}_1 + t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} \quad (5)$$

$$\bullet t(\nu; \alpha) \cdot \hat{\sigma} \sqrt{2/n} < \bar{y}_2 - \bar{y}_1 \leq t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} \quad (\text{片側検定で有意}) \Rightarrow 0 < \mu_2 - \mu_1 \leq \bar{y}_2 - \bar{y}_1 + t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} \quad (6)$$

$$\bullet -t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} < \bar{y}_2 - \bar{y}_1 \leq t(\nu; \alpha) \cdot \hat{\sigma} \sqrt{2/n} \Rightarrow \bar{y}_2 - \bar{y}_1 - t(\nu; \alpha) \cdot \hat{\sigma} \sqrt{2/n} \leq \mu_2 - \mu_1 \leq \bar{y}_2 - \bar{y}_1 + t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} \quad (7)$$

$$\bullet \bar{y}_2 - \bar{y}_1 \leq -t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n} \Rightarrow \bar{y}_2 - \bar{y}_1 - t(\nu; \alpha) \cdot \hat{\sigma} \sqrt{2/n} \leq \mu_2 - \mu_1 \leq 0 \quad (8)$$

(6) 式において，信頼下限の符号は厳密に成り立つ。すなわち，ゼロを含まない。

この方法は，竹内 (1973) の多重信頼方式に基づくものであり，帰無仮説 $H_0: \mu_2 - \mu_1 = \delta$ に対する検定において，次の受容域を用いることによって得られる (図 1)。

- $\mu_2 - \mu_1 = \delta \leq 0$ に対しては $\bar{y}_2 - \bar{y}_1 - (\mu_2 - \mu_1) \leq t(\nu; \alpha) \cdot \hat{\sigma} \sqrt{2/n}$
- $\mu_2 - \mu_1 = \delta > 0$ に対しては $|\bar{y}_2 - \bar{y}_1 - (\mu_2 - \mu_1)| \leq t(\nu; \alpha/2) \cdot \hat{\sigma} \sqrt{2/n}$

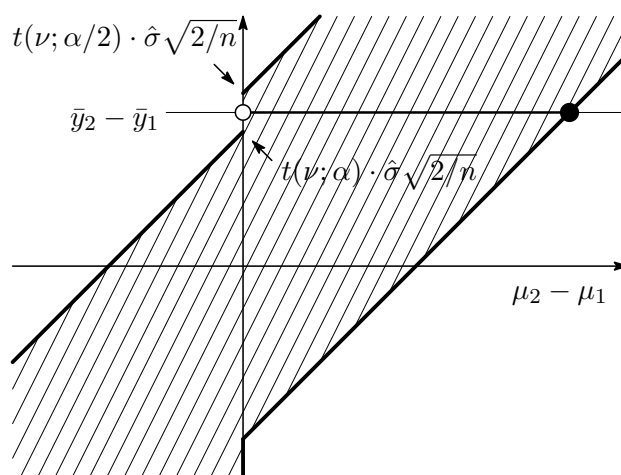


図 1. 片側と両側信頼区間の融合

(5) 式は通常の両側検定で有意となる場合である。この場合には通常の両側信頼区間が得られる。(6) 式は，片側検定では有意となり，両側検定では有意とならない場合である (図 1 参照)。この場合でも，上記の手順は，2つの処理に差があることは主張できる。つまり有意差の検出 (directional decision) に関しては，片側検定と同じ検出力をもつ。そして，この方法では常に両側信頼区間が得られる。

$k \geq 3$ の場合については，対比較 (Miwa & Hayter, 1999; Hayter *et al.*, 2001)，対照との比較 (Hayter *et al.*, 2000)，隣り合う処理の比較 (Somerville *et al.* 2001) に対して，新たな数表とともに手順が示されている。

計算例

Steel *et al.* (1997) は、表 5 の estrogen assay の例を与えている。この例はまた、Hochberg & Tamhane (1987) でも取り上げられている。

表 5. Uterine weights (mg) of mice from an estrogen assay
(means of $n = 4$ mice)

	Control		Test solutions				
	A_1	A_2	A_3	A_4	A_5	A_6	A_7
\bar{y}_i	96.15	88.25	75.40	68.45	84.90	78.90	70.20
$\hat{\sigma}^2 = 145.84$ ($\nu = 21$)							

実験者は $\mu_1 > \mu_i$ ($2 \leq i \leq 7$) かどうかに興味があるものとする。 $\mu_1 - \mu_i$ について、Dunnett 法による 90%片側同時信頼区間、両側信頼区間、および融合型同時信頼区間は次のようになる。

片側 90%同時信頼区間	両側 90%同時信頼区間	融合型 90%同時信頼区間
$-9.83 \leq \mu_1 - \mu_2 < \infty$	$-13.03 \leq \mu_1 - \mu_2 \leq 28.83$	$-9.83 \leq \mu_1 - \mu_2 \leq 29.98$
$3.02 \leq \mu_1 - \mu_3 < \infty$	$-0.18 \leq \mu_1 - \mu_3 \leq 41.68$	$0 < \mu_1 - \mu_3 \leq 42.83$
$9.97 \leq \mu_1 - \mu_4 < \infty$	$6.77 \leq \mu_1 - \mu_4 \leq 48.63$	$5.62 \leq \mu_1 - \mu_4 \leq 49.78$
$-6.48 \leq \mu_1 - \mu_5 < \infty$	$-9.68 \leq \mu_1 - \mu_5 \leq 32.18$	$-6.48 \leq \mu_1 - \mu_5 \leq 33.33$
$-0.48 \leq \mu_1 - \mu_6 < \infty$	$-3.68 \leq \mu_1 - \mu_6 \leq 38.18$	$-0.48 \leq \mu_1 - \mu_6 \leq 39.33$
$8.22 \leq \mu_1 - \mu_7 < \infty$	$5.02 \leq \mu_1 - \mu_7 \leq 46.88$	$3.87 \leq \mu_1 - \mu_7 \leq 48.03$

提案する方法では、片側検定と同様に、処理 A_3, A_4, A_7 は対照処理よりも重量が低いと判定でき、なおかつ、常に上側信頼限界が与えられる。

4.2 アンバランスモデルでの Bartholomew 検定

Bartholomew (1959) は、 σ^2 が既知の場合、(3) 式の対立仮説に対して、尤度比検定統計量

$$\bar{\chi}^2 = \sum_{i=1}^k n_i (\hat{\mu}_i - \bar{y})^2 / \sigma^2$$

を与えた。その帰無仮説のもとでの分布は

$$\Pr\{\bar{\chi}^2 > c\} = \sum_{l=2}^k P(l, k; \mathbf{n}) \Pr\{\chi^2(l-1) > c\}, \quad c > 0 \quad (9)$$

と表わされる。ここで、 $\hat{\mu}_1, \dots, \hat{\mu}_k$ は、順序制約 $\mu_1 \leq \dots \leq \mu_k$ の下での最尤推定量である。 $P(l, k; \mathbf{n})$ は、 $\hat{\mu}_1, \dots, \hat{\mu}_k$ が l 個の異なる値 (すなわち l 個のレベル) から構成される確率であり、レベル確率と呼ばれる。レベル確率はサンプルサイズ $\mathbf{n} = \{n_1, \dots, n_k\}$ に依存する。 σ^2 が未知の場合の検定統計量については、 σ^2 を推定値で置き換えた $\bar{B}^2 = \sum_{i=1}^k n_i (\hat{\mu}_i - \bar{y})^2 / \hat{\sigma}^2$ など、いくつかの方法が提案されている。

Bartholomew 検定は比較的早い時期に開発された。その検出力は非常に高く、実用性の高い手法であるにもかかわらず、あまり広く普及してこなかった。その理由は、サンプルサイズ n_i が不揃いの場合に、レベル確率 $P(l, k; \mathbf{n})$ の計算が極めて困難であり、従って、検定等計量 \bar{B}^2 の p 値の計算も困難であったためである。 F 検定と検定統計量の形はほとんど同じでありながら、アンバランスモデルでに対応できないことが、Bartholomew 検定の最大の難点であった。

Miwa *et al.* (2000) は、サンプルサイズが不揃いの場合にも、レベル確率 $P(l, k; \mathbf{n})$ を高速に精度良く計算する方法を開発し Bartholomew 検定を実行することが可能となった。

計算例

Banno & Yamagami (1989) は, *Eupromus ruber* の幼虫および成虫における ECI (conversion efficiency of ingested food) を研究している (表 6)。ECI は, 幼虫の各ステージ, および成虫について

$$\text{ECI (\%)} = 100 \times \frac{\text{虫の乾物重}}{\text{食べた木の乾物重}}$$

によって定義される。

表 6. *Eupromus ruber* の ECI (%)

Stage	3rd	4th	5th	6th	7th	Adult
ECI (%)	2.415	2.411	2.129	2.009	1.923	1.669
n_i	4	21	17	15	10	21

$\hat{\sigma}^2 = 0.578$ ($\nu = 82$)
 $\bar{B}^2 = 11.43$ ($p = 0.00455$), $F = 2.286$ ($p = 0.0535$)

若い幼虫ほど, より多くの食物を必要とするので, $\mu_1 \geq \mu_2 \geq \mu_3 \geq \mu_4 \geq \mu_5 \geq \mu_6$ の関係が想定される。実際, 標本平均値 \bar{y}_i はこの順に並んでいる。したがって, 順序制約のもとでの最尤推定値 $\hat{\mu}_i$ は標本平均値 \bar{y}_i で与えられる。検定統計量は $\bar{B}^2 = \sum_{i=1}^6 n_i \cdot (\hat{\mu}_i - \bar{y})^2 / \hat{\sigma}^2 = 11.43$ ($p = 0.00455$) となり, 高度に有意である。通常の F 検定を実施すると $F = \sum_{i=1}^6 n_i \cdot (\bar{y}_i - \bar{y})^2 / (k-1) \hat{\sigma}^2 = 2.286$ ($p = 0.0535$) であり, 5%水準でも有意とはならない。

なお, レベル確率 $P(l, k; n)$ の計算は多くの分野に登場し, Miwa *et al.* (2000) のアルゴリズムを使用することが可能である。

4.3 多重対比検定と多変量正規関数

(3) 式の単調順序対立仮説の検定には, 尤度比検定のほかに, 有限個の対比の最大値を検定統計量として用いる方法もいくつか提案されている。すなわち, (1) 式の一元配置モデルにおける処理平均値 \bar{y}_j ($1 \leq j \leq k$) に対して, m 個の対比 (contrast) を考え

$$z_i = \sum_{j=1}^k a_{ij} \bar{y}_j, \quad \sum_{j=1}^k a_{ij} = 0 \quad (1 \leq i \leq m) \quad (10)$$

あるいは, その標準化した値

$$t_i = \sum_{j=1}^k a_{ij} \bar{y}_j / \sqrt{\sum a_{ij}^2 / n_j} \quad (11)$$

の最大値を検定統計量として用いるものである。たとえば, 次のような検定がある。

(1) Williams の検定 (Williams, 1971)

$$z_i = -\bar{y}_1 + \frac{n_{i+1} \bar{y}_{i+1} + \cdots + n_k \bar{y}_k}{n_{i+1} + \cdots + n_k}, \quad 1 \leq i \leq k-1 = m$$

(2) max- t 法 (広津, 1992; 広津ら, 1997)

$$z_i = -\frac{n_1 \bar{y}_1 + \cdots + n_i \bar{y}_i}{n_1 + \cdots + n_i} + \frac{n_{i+1} \bar{y}_{i+1} + \cdots + n_k \bar{y}_k}{n_{i+1} + \cdots + n_k}, \quad 1 \leq i \leq k-1 = m$$

$$t_i = z_i / \sqrt{1/(n_1 + \cdots + n_i) + 1/(n_{i+1} + \cdots + n_k)}$$

(3) 隣り合う処理の比較 (Lee and Spurrier, 1995)

$$t_i = (-\bar{y}_i + \bar{y}_{i+1}) / \sqrt{1/n_i^2 + 1/n_{i+1}^2}, \quad 1 \leq i \leq k-1 = m$$

このような多重対比検定を実行するためには、次の確率を計算する必要がある。

$$\Pr \left\{ \max_{1 \leq i \leq m} \frac{t_i}{\hat{\sigma}} \leq \delta \right\} = \Pr \{t_i \leq \delta \cdot \hat{\sigma}, 1 \leq i \leq m\} \quad (12)$$

ここで、ベクトル (t_1, \dots, t_m) は m 次元多変量正規分布に従う確率変数であり、(12) 式は、その同時分布関数の計算に他ならない。

Miwa *et al.* (2003) は、任意の多次元正規ベクトル $\mathbf{x} = (x_1, \dots, x_m)^T \sim N_m(\boldsymbol{\mu}, \Sigma)$ に対して、非心象限確率

$$P_m(\boldsymbol{\mu}, \Sigma) = \Pr\{x_i \geq 0, 1 \leq i \leq m\} = \int_0^\infty \cdots \int_0^\infty \phi_m(\mathbf{x}; \boldsymbol{\mu}, R) dx_1 \cdots dx_m \quad (13)$$

を極めて正確に計算するアルゴリズムを与えた。現在、このアルゴリズムはRのパッケージ `mvtnorm` に組み込まれている (Genz *et al.*, 2009)。

謝辞

筆者は昨年度、図らずも日本計量生物学会賞を受賞する栄誉を与えられた。30 有余年、農業技術開発の分野で研究を続けてきているので、Fisher の創始した計量生物学会から表彰されることは、ことさら喜ばしく感じられる。これまで、ここには書ききれないくらい多くの人に支えられてきた。学会においても、多くの指導を受けてきた。心よりお礼を申し上げたい。

参考文献

- Banno, H. and Yamagami, A. (1989) Food consumption and conversion efficiency of the larvae of *Eupromus ruber* (Dalman) (Coleoptera: Cerambycidae). *Appl. Ent. Zool.*, **24**, 174–179.
- Bartholomew, D. J. (1959) A test of homogeneity for ordered alternatives. *Biometrika*, **41**, 36–48.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, **57**, 289–300.
- Duncan, D. B. (1955) Multiple range and multiple F tests. *Biometrics*, **11**, 1–42.
- Dunnett, C. W. (1955) A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.*, **50**, 1096–1121.
- Fisher, R. A. (1935) *The Design of Experiments*. Edinburgh and London: Oliver & Boyd.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. and Hothorn, T. *mvtnorm: Multivariate Normal and t Distributions* (2009). R package version 0.9-5.
- Hayter, A. J., Miwa, T. and Liu, W. (2000) Combining the advantages of one-sided and two-sided procedures for comparing several treatments with a control. *J. Statist. Plann. Inference*, **86**, 81–99.
- Hayter, A. J., Miwa, T. and Liu, W. (2001) Efficient directional inference methodologies for the comparisons of three ordered treatment effects. *Journal of the Japan Statistical Society*, **31**, 153–174.
- 広津千尋 (1992) 実験データの解析—分散分析を超えて—. 東京: 共立出版.

- 広津千尋, 西原健自, 杉原正顕 (1997) 最大 t 法のための有意確率, 検出力, 例数設計計算アルゴリズム. *応用統計学*, **26**, 1–16.
- Hochberg, Y. and Tamhane, A. C. (1987) *Multiple Comparison Procedures*. New York: Wiley.
- Hsu, J. C. (1996) *Multiple Comparisons: Theory and Methods*. London: Chapman and Hall.
- Lee, R. E. and Spurrier, J. D. (1995) Successive comparisons between ordered treatments. *J. Statist. Plann. Inference*, **43**, 323–330.
- 三輪哲久 (1997) 農業研究分野における多重比較論争. *応用統計学*, **26**, 99–109.
- 三輪哲久 (1998) Bartholomew's test as a multiple contrast test and its application. *Japanese Journal of Biometrics*, **19**, 1–9.
- 三輪哲久 (2008) 多重検定の基礎理論. *計量生物学*, **29**, 5–14.
- Miwa, T. and Hayter, A. J. (1999) Combining the advantages of one-sided and two-sided test procedures for comparing several treatment effects. *J. Amer. Statist. Assoc.*, **94**, 302–307.
- Miwa, T., Hayter, A. J. and Kuriki, S. (2003) The evaluation of general non-centred orthant probabilities. *J. Roy. Statist. Soc. Ser. B*, **65**, 223–234.
- Miwa, T., Hayter, A. J. and Liu, W. (2000) Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models. *Comput. Statist. Data Anal.*, **34**, 17–32.
- 三輪哲久, 佐々木昭博, 大塚雅雄 (1988) 多処理実験における対比較-各手法の特徴と適用上の問題点-. *植物防疫*, **42**, 351–356.
- 永田靖, 吉田道弘 (1997) *統計的多重比較法の基礎*. 東京: サイエントリスト社.
- Nelder, J. A. (1971) Discussion to the paper by O'Neill and Wetherill. *J. Roy. Statist. Soc. Ser. B*, **36**, 244–246.
- Ryan, T. P. and Woodall, W. H. (2005) The most-cited statistical papers. *Journal of Applied Statistics*, **32**, 461–474.
- Shaffer, J. P. (2007) Controlling the false discovery rate with constraints: The newman-keuls test revisited. *Biometrical Journal*, **49**, 136–143.
- Somerville, P., Miwa, T., Liu, W. and Hayter, A. J. (2001) Combining one-sided and two-sided confidence interval procedures for successive comparisons of ordered treatment effects. *Biometrical Journal*, **43**, 533–542.
- Steel, R. G. D., Torrie, J. H. and Dickey, D. A. (1997) *Principles and Procedures of Statistics: A Biometrical Approach*. New York: McGraw-Hill, 3rd ed. edition.
- 竹内啓 (1973) *数理統計学的方法的基礎*. 東京: 東洋経済新報社.
- Tukey, J. W. (1953) The problem of multiple comparisons. *Unpublished report*. Princeton University.
- Williams, D. A. (1971) A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, **27**, 103–117.

連絡先: 三輪哲久

〒 305-8604 つくば市観音台 3-1-3

農業環境技術研究所 生態系計測研究領域

E-mail: miwa@niaes.affrc.go.jp; Tel: 029-838-8221; Fax: 029-838-8199