

植物防疫基礎講座

多重比較法とその選び方(3)

ノンパラメトリック検定で用いる多重比較法

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

近年、「ロバストな推定量」ということばを耳にする機会が増してきたように思われる。前々回に述べたように、分散分析では「誤差が正規分布に従い、かつその分散が処理などに依存せずに常に一定である」と仮定して分析を行っている。しかし、実際にはこの仮定は満たされていないことも多い。本稿で取り上げてきたフェロモントラップ誘殺数のデータもその一つである(表-1)。このように仮定が満たされていないにもかかわらず分散分析を行うと、いろいろと具合の悪いことが生じてくる。トラップデータの例では、変換を行わずに分散分析を行うと、本当はトラップ間に非常に有意な差があるにもかかわらず、それを検出することができなかった。このような間違った結果を導かないためにも、多少は仮定が満たされていなくても、正しい検定結果を出してくれるような推定量「頑健な(ロバストな)推定量」があればたいへん有難い。

ロバストな推定量の一種として昔から使われていたものに「ランク検定」と呼ばれる1群がある。これは分布を仮定しないことから「ノンパラメトリック検定法」と呼ばれる検定法の一部分に属する。このランク検定では、データの値そのものを用いないで、まずデータをその大小関係から順位値に置き換える。そして、この順位に関して検定を行うというものである。元の数値自体は用いないのであるから、これはずいぶんいい加減な方法であるように感じられるかもしれない。ところが、いい加減な分散分析を行うよりは、ランク検定のほうがずっと正しい結果を出してくれるのである。むしろラン

ク検定の最も大きな欠点は、適用できる実験計画の範囲が狭いという点にあると思われる。分散分析は、基本的にいかなる複雑なモデルにでも適用できるのに対して、ランク検定は、主として1元配置の実験計画しか取り扱えない。今回は、このようなランク検定法に関する多重比較法のいくつかを紹介したい。

I ランク検定

再び、フェロモントラップのデータを取り上げて計算してみよう。今回は説明のため、三つのトラップが同じ地域に設置されていたとみなし(データ構造を無視して)分析を行う。前々回はBox-Cox変換という面倒な方法によって苦労して変換式を決定してから分散分析を行った。しかし、ランク検定では何も仮定を置かないので、データを変換せずにそのまま分析にとりかかれる。今のデータの形の場合、ランク検定ではFriedman検定という検定法を使うことができる。ランク検定の詳しい計算法やその背景についてはほかの参考書にまかせるが、ここで、簡単に計算法を追ってみたい。まず、今のデータは月ごとにブロック化されているので、月ごとの三つのデータについて順位(1, 2, 3)をつける。同じ値が存在した場合には「中間順位」をつける。例えば、順位1番と2番になるべき値が等しい値であったなら、その値にはそれぞれ $(1+2)/2=1.5$ という値をつける。次に、各トラップごとに、それぞれの四つの順位値を合計する。今の場合、トラップA1では9, A2では11, B3では4となる。そして、これらの値の偏差平方和 S を計算する。(いまの場合、偏差平方和と残差平方和は同じである。)この S を検定量に用い、Friedman検定用のしかるべき数表と比べて検定を行う。今の場合、 $S=(9-8)^2+(11-8)^2+(4-8)^2=26$ 。数表から、この値は0.042レベルで有意であることがわかる。対数変換後の分散分析では、0.0007レベルで有意であったのに比べると、検出力はきわめて低いが、何とか有意差は出してくれる。今のデータの場合、変換せずに分散分析を行うと有意差は検出できなかったのだが、Friedman検定を使えば、変数変換方法で何ら頭を悩ませることなく、有意差に到達できたというわけである。さらに重要なのは、この検定では「三つのトラップが同じ母集団に属する」という帰無仮説以外には何も仮定を置いていないため、そういう意

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

味で、この Friedman 検定による結論は信頼性が高いという点である。

一般にランク検定の手順には次のような二つの側面がある。

- ① データ数があまり多くなく、同順位 (タイ) がない場合には、数表から正確な判定基準を読み取って検定を行う。
- ② データ数が多い場合には近似理論により、正規分布やカイ 2 乗分布を用いて検定する。

いまの Friedman 検定の場合、処理数、ブロック数が十分に多ければ、多変量中心極限定理により、処理ごとの順位合計の値は多変量正規分布に従う。したがって、この場合は正規分布理論に基づく検定を行うことができる。正規分布理論に基づく検定では通常は F 検定や t 検定を使うのだが、ランク検定の場合はデータを順位に置き換えた時点で誤差分散が自動的に決まってしまうので、 F 検定ではなくカイ 2 乗検定、 t 検定ではなく正規分布検定をそれぞれ用いることになる。これら正規近似はデータ数が少ないときには精度がよくない。しかし、SAS を含め多くの統計パッケージでは、データ数の大小によらず、正規近似を使って検定を行うしくみになっている。したがって、データ数が少なく、かつ同順位が存在しない場合には、統計量だけを統計パッケージで計算させ、最終的な検定は数表を開いて行うのがよい。なお、SAS の場合、ランク検定の計算サブルーチン (プロシジャ) は 1 元配置無作為化法に関しては準備されている (プロシジャ `upar1way`)。しかし、今のデータのように 1 元配置乱塊法に関しては用意されていない。SAS のマニュアルでは、乱塊法の場合は、「プロシジャ `rank`」なるものを使ってデータを順位に置き換えてから、「プロシジャ `glm`」などを使って普通の分散分析 (F 検定) を行うことを推薦している (SAS Institute, 1990)。これは、正確な誤差分散がわかっているのに、それを用いずに普通の F 検定を行うという形の方式である。この方法は、Friedman 検定の代用として使うことができる。IMAN and DAVENPORT (1980) は Friedman のオリジナルな統計量をカイ 2 乗近似して検定するよりは、ランク変換後の F 検定のほうがむしろ近似精度がよいと報告している。しかし、カイ 2 乗近似は有意水準を多くの場合 α 以下に保つことができるのに対し、 F 近似は有意水準を α 以下に保てないことが多いので、あまり薦めることはできないであろう。なお、ランク変換後の F 検定が許されるのは、限られた分散分析モデルだけであることには十分注意しておくべきである (THOMPSON, 1991)。

II 乱塊法の場合の多重比較

さて、先の Friedman 検定で有意差が出たので、次にはやはり、どのトラップとどのトラップの間に有意差があるかを調べたい。今までに述べた Tukey 法や Scheffé 法は、誤差が正規分布に従うことを前提しているため、これらの方法をランク検定にそのまま使うことはできない。そこで、ここで少し多重比較法の原点に戻って問題を考え直してみよう。

一般に K 個の処理平均がある場合、そのうちの二つを取り出して比較する組み合わせは ${}_k C_2$ とおりある。この ${}_k C_2$ とおりの検定をそれぞれ α 水準で行うとどのようなことになるであろうか? K 個の処理に全く差がないという帰無仮説の下で、この ${}_k C_2$ 回の検定のそれぞれが α の確率で間違っして有意差を出してしまう。この ${}_k C_2$ 回の検定のうち「少なくとも 1 回以上間違っして有意差を出してしまう確率」は、1.0 から「 ${}_k C_2$ 回の検定のいずれでも有意差を出さない確率」を引いたものである。仮りに ${}_k C_2$ 回の検定が独立であるとした場合には、「 ${}_k C_2$ 回の検定のいずれでも有意差を出さない確率」は乗法定理より $(1 - \alpha)$ を ${}_k C_2$ 回掛けたものとなるから、結局「少なくとも 1 回以上間違っして有意差を出してしまう確率」の上限は、

$$1 - (1 - \alpha)^{{}_k C_2} \quad (7)$$

となる。いま、五つの処理のすべての組合せについてそれぞれ $\alpha = 0.05$ 水準で検定を行うと、少なくとも一つ以上の検定で間違っして有意差を出してしまう確率の上限は $1 - (1 - 0.05)^{{}_5 C_2} = 0.40$ である。したがって、このような検定を行うと実際の有意水準は 0.05 ではなく、0.40 を上限とする値にまで上昇してしまっている。このような問題は「多重性 (multiplicity)」の問題と呼ばれる。

この問題を解決する、最も安易な方法は、上の (7) 式の値自体を α に調整するというものである。先ほどの例では、 ${}_5 C_2$ 回のそれぞれの検定を 0.05 水準で行ったために実際の有意水準が (7) 式に従って 0.40 を上限とする値にまで上昇してしまっした。それならば、むしろ居直って、(7) 式の値が 0.05 になるように α を調整すればいいのではないか、という考え方が出てくる。つまり、(7) 式より、

$$\alpha = 1 - (1 - \alpha')^{{}_k C_2}$$

を満たす α' を求めて、その α' を用いて ${}_k C_2$ 回の検定を行えば文句はなかろう、というわけである。上式を変形すれば、

$$\alpha' = 1 - (1 - \alpha)^{1/kC_2} \quad (8)$$

である。この α' を使って検定を行う方式を「Dunn-Šidák の不等式を用いた検定」という。

(7) 式を導くときには、 kC_2 回の検定が独立であると仮定することにより有意確率の上限を求めた。しかし、検定の種類によっては、ひょっとしたら、これらの検定で有意となる確率は、相互に排他的かもしれない。その場合には (7) 式の上限は満たされなくなる。そのような心配症の人は (8) 式の代わりに最も安心・確実な次式を用いるとよい。

$$\alpha' = \alpha / kC_2 \quad (9)$$

この式の意味は集合論で使う「ベン図」なるものを描いてみれば理解できる。この (9) 式の α' を用いて検定を行う方式を「Bonferroni の不等式を用いた検定」という。

これら Dunn-Šidák の不等式や Bonferroni の不等式は、いかなる多重比較問題にも使える便利な式である。正規分布を仮定した分散分析の場合にももちろん活用することができる。しかし、その場合 Dunn-Šidák 法や Bonferroni 法は Tukey 法や Tukey-Kramer 法といった巧妙な方法に比べてかなり検出力が低い。統計パッケージによっては分散分析の後に Tukey-Kramer 法と同時に Dunn-Šidák 式や Bonferroni 式を用いた検定結果を出力するものがあるが、これら二つの出力は Scheffé 法と同様、基本的に不十分なものである。ただし、ランク検定のような正規分布を仮定しない検定においてはこの Dunn-Šidák 法や、Bonferroni 法が重要な意味を持つ。

先ほどのフェロモントラップのデータの場合には、処理数が 3 であるから、全部で ${}_3C_2 = 3$ 通りの組合せがあり、それぞれで検定を行うことができる。Bonferroni 式より、この 3 回の検定のそれぞれを $0.05/3 = 0.0167$ 水準で検定を行えば、全体としての有意水準は間違いなく 0.05 に保たれることがわかる。この方式で検定を実際に行ってみよう。二つの処理の比較の際に Friedman 検定を用いると「符号検定」というものになるが、これは一般に検出力がかなり小さい。もし正規分布の仮定が成り立っている場合には、 t 検定に比べるとその漸近的な効率率は約 64% しかない。そこで、ここでは「Wilcoxon の符号化順位検定」を用いる。この検定は t 検定に比べて漸近効率は約 95% である。この検定においては、まず二つの処理の一方から他方を引き、その差の絶対値について大きさの順に順位をつける。次に、差が正のもの順位と差が負のもの順位をそれぞれ合計する。この順位合計のうち小さいほうの値 R を検定量に用い、数表を開いて検定を行う。今のデータの場合、この検定を行う

と 3 通りの検定のいずれもが $\alpha = 0.0167$ 水準で有意ではない。実はいまの場合、データ数は 4 であるから、仮に片方の処理のすべてのデータがもう一方のすべてのデータよりも大きい場合であっても、そのようなことがたまたま生じる確率は $(1/2)^4 = 0.0625$ もある。したがって、データ数が四つの場合の「Wilcoxon の符号化順位検定」では 0.0167 水準はおろか 0.05 水準でも絶対に有意差は出ないのである。

いまの Bonferroni 式を用いた検定では、二つの組合せごとに順位を付け直して検定を行った。これに対して、順位を付け直さずに (joint ranking のままで) 行う、もっと計算の楽な検定方法も存在する。すべての処理に真に差がないとき、先ほど述べたように、データ数が大きい場合には順位合計は多変量正規分布に従う。したがって、このような場合には、正規分布に基づく多重比較法を使っても許されるかのようにみえる。いま、処理数を k 、繰返し数 (ブロック数) を b としよう。Friedman 検定の場合、順位の変異分散 s_e^2 としては次の式を用いる。

$$s_e^2 = \frac{ss}{b(k-1)}$$

ただし、ここに ss はすべてのデータにつけられた順位の偏差平方和、つまり bk 個の (順位 - (すべての bk 個の順位の平均))² をすべて足したものである。順位に同順位 (タイ) がない場合には、上の分散はもっと簡便な式 $k(k+1)/12$ でも計算できるが、タイのある場合には直接上の式で計算を行うのがよいであろう。さて、Tukey 型の判定基準は、前回の (3) 式と同様に、

$$|m_i - m_j| > \left(\frac{s_e}{\sqrt{b}}\right) Q(\alpha, k, \infty) \quad (10)$$

ここに m_i は第 i 処理の順位合計 R_i をその繰返し数 b で割ったもの (順位平均) である。前回の (3) 式と違い、スチューデント化範囲の自由度が ∞ になっているのは、ランク検定の場合に誤差分散が正確に決まってしまうことと関連している。データ数が少ない場合には、正規近似の精度が悪いので、その場合は数表を使って検定すべきである。その数表は MacDONALD and THOMPSON (1967) に掲載されている。二つのグループの比較 (contrast) $\sum_{i=1}^k c_i m_i$ に関しては、前回の (6) 式と同様に、Scheffé 型の検定を行うことができる。判定基準は、

$$\left| \sum_{i=1}^k c_i m_i \right| > s_e \sqrt{\chi^2(\alpha, k-1)} \quad (11)$$

ここに s_e は問題とする contrast $\sum_{i=1}^k c_i m_i$ の推定分散 s_e^2 の平方根である。つまり、

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

前回の(6)式をそのまま使えば、判定基準としては $s_c \sqrt{(k-1)F(\alpha, k-1, \infty)}$ の値を使うことになるはずだが、 F 分布とカイ2乗分布の関係から、この値は $s_c \sqrt{\chi^2(\alpha, k-1)}$ と同じであるので、普通こちらの表現を使う。

ランク検定における多重比較法として、いくつかの教科書や統計パッケージでは、これら順位付け直しなしの(joint rankingによる)Tukey型検定法やScheffé型検定法が取り上げられている(例えば、米澤ら, 1988; 白旗, 1987)。しかし、これら joint rankingによる検定は、「すべての処理効果が真に等しい」という仮説(完全帰無仮説と呼ばれる)にのみ基づいている。ランク検定の場合、「処理*i*と処理*j*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「すべての処理効果が真に等しい」という帰無仮説が最も大きな有意確率を与えるとは限らない。したがって、これらの joint rankingによる方法は厳密な意味では多重比較法ではなく、有意水準 α の値を必ずしも保証しない。このため、これらの方法を使用することはあまり推薦できない(HOCHBERG and TAMHANE, 1987)。むしろ、先ほど行ったように、二つの処理ごとに順位を付け直して Bonferroni 型で有意水準を調整しながら(あるいは Dunn-Šidák 型で調整しながら)Wilcoxon の符号化順位検定を行う方式のほうが、検出力は弱い、確実であり信頼できる。

ちなみに、もう一度フェロモントラップのデータを用い、joint ranking のままで検定を行ってみよう。いまの場合、データ数が少ないので、MacDONALD and THOMPSON (1967)の数値を読みとれば、 $\alpha = 0.05$ では限界となる順位合計 R_i の差は7である。いまのデータでは、トラップ A1 と A2 の R_i の差は $|9 - 11| = 2$ 、A1 と B3 の差は $|9 - 4| = 5$ 、A2 と B3 との差は $|11 - 4| = 7$ である。したがって、「トラップ A2 と B3 の差は 0.05 水準で有意であり、その他の二つの組合せの差は有意でない」と結論される。実はいまの場合、完全帰無仮説の下での正確な有意水準は、数表をみなくても手計算で簡単に求めることができる。「すべてのトラップ間で差がない」という帰無仮説の下での(3!)⁴と通りの組合せのうち、7以上の差が生じるのは3!×9とおりあるので、この有意確率は正確には0.0417である。

III 完全無作為化法での多重比較

完全無作為化法の場合には、全体の検定は、Kruskal-Wallis 検定というもので普通行う。Kruskal-Wallis 検定においては、 k 個の処理のすべてのデータを込みにして順位をつけ、各処理ごとの順位合計 R_i ($i = 1, 2, \dots$,

k)を計算する。そしてこの R_i を各処理のデータ数 n_i で割った順位平均 $m_i = R_i/n_i$ を計算する。どの処理にも本当に差がない場合には、この順位平均 m_i はどの処理でも似たような値をとるであろうから、順位平均の処理間のばらつきは小さくなるはずである。そこで、「順位平均の分散がある値よりも大きければ、この k 個の処理集団は均質ではないと判断する」という方式の検定方法が考えられる。これが Kruskal-Wallis 検定である(この発想は Friedman 検定でも全く同じである)。

全体の検定で有意差が出た場合には、次に ${}_k C_2$ とおりのすべての二つの組合せごとに有意差を検定する。このとき2組のデータだけで順位を付け直してから、Wilcoxon の順位和検定あるいは Mann-Whitney の U 検定を行う。ただし有意水準は Bonferroni の式あるいは Dunn-Šidák の式で調整しなければならない。Wilcoxon の順位和検定は Kruskal-Wallis 検定の2標本の場合に相当するもので、順位合計 R_i ($i = 1, 2$)を用いて検定を行う方式である。Mann-Whitney の U 検定は $U_i = R_i - n_i(n_i + 1)/2$ ($i = 1, 2$)を用いて検定を行う点が前者と異なるが、内容はやはり同じである。

乱塊法の場合と違い、完全無作為化法の場合には、2組ごとに順位を付け直してから Tukey 法と同様の発想で検定を行うこともできる。これは Steel-Dwass の方法と呼ばれ、Bonferroni 式や Dunn-Šidák 式を用いるよりも検出力が高い。この Steel-Dwass の方法のための正確な数表はあまり完備されていないが、データ数が大きい場合には、容易に近似的に計算を行うことができる。まず、処理ごとのデータ数が等しく n であり、かつ同順位(タイ)のない場合を考えよう。このとき、データ数が大きければ中心極限定理により、 $R_i - R_j$ ($i \neq j$)は平均ゼロ、分散 $V(R_i - R_j) = n^2(2n + 1)/3$ の正規分布に従う。前回述べたように、標準正規分布 $N(0, 1)$ から k 個のサンプルを取り出したときの、その最大値と最小値の差(範囲)の α 限界点は「スチューデント化された範囲のパーセント点 $Q(\alpha, k, \infty)$ 」として数表化されている。これは、 k 個の独立サンプルを取り出したとき、任意の二つの処理平均値の差 $m_i - m_j$ が $N(0, 2)$ に従うならば、その $m_i - m_j$ の最大絶対値の α 限界点が $Q(\alpha, k, \infty)$ になることを意味している。 $R_i - R_j$ を $\sqrt{V(R_i - R_j)}/2$ で割ったものは $N(0, 2)$ に従い、かつ、 ${}_k C_2$ 個のそれらの値の相互の共分散構造も上の状況に一致することから、結局 $|R_i - R_j| / \sqrt{V(R_i - R_j)}/2$ の α 限界点は $Q(\alpha, k, \infty)$ となる。この関係をそのまま用いて検定してもよいし、あるいは、 R_i の平均が $E(R_i) = n(2n + 1)/2$ 、分散が $V(R_i) = n^2(2n + 1)/12$ であることから、少し表現を変えて、次の不等式が成り立つときに「二つの処理は α 水準

で有意差がある」と判定すればよい。

$$|R_i - E(R_i)| - 0.5 > \sqrt{\frac{V(R_i)}{2}} Q(\alpha, k, \infty) \quad (12)$$

ここで左辺から 0.5 を引いているのは、離散分布を連続分布で近似する際にその近似をよくするためによく用いる「連続修正」である。 $|R_i - R_j|$ を用いて検定を行う場合には -0.5 でなく -1.0 により修正するのがよいであろう。なお、処理ごとのデータ数が等しくない場合には、Tukey-Kramer 法のような修正を考慮することができる。この場合、第 i 処理と第 j 処理を比較するとき、 R_i の平均は $E(R_i) = n_i(n_i + n_j + 1)/2$ 、分散は $V(R_i) = n_i n_j(n_i + n_j + 1)/12$ である。そこで、この値を用いて上の (12) 式で検定することができる。同順位(タイ)がある場合には、実際の $V(R_i)$ は $n^2(2n + 1)/12$ や $n_i n_j(n_i + n_j + 1)/12$ より小さくなるので、(12) 式の基準はやや保守的な方向に偏るが、有意水準 α が保たれるには違いないので一応問題は無いであろう。もし気になるようであれば、いちいち正確な $V(R_i)$ を用いればよい。

完全無作為化法において、順位の付け直しなしの(つまり、joint ranking による)検定を行う場合には、誤差分散 s_e^2 としては、データ総数を N とするとき次の式を用いる。

$$s_e^2 = \frac{SS}{N-1}$$

タイがない場合には、より簡便な式 $N(N+1)/12$ を用いればよい。先ほどの乱塊法の場合と同様に、この誤差分散を前回の式(4)、(6)に代入し、順位平均の差を検定すれば、それぞれ Tukey-Kramer 型検定、Scheffé 型検定となる。このときの自由度 ν は乱塊法の場合と同じく ∞ である。処理ごとのデータ数が等しく、かつ小さいときの Tukey 型検定の比較的正確な棄却限界値は、MacDONALD and THOMPSON (1967) に掲載されている。また、この誤差分散を用いて joint ranking のまま Dunn-Šidák 型や Bonferroni 型の検定を行うという検出力の弱い方式も存在する(高木, 1985; 米澤ら, 1988; HETTMANSPERGER, 1984; ZAR, 1984 で取り上げられている)。いずれにせよ、これら一連の joint ranking による検定法は、完全帰無仮説にのみ基づくため、乱塊法の場合と同様に厳密な意味で多重比較法ではない(有意水準 α を必ずしも保証できない)。したがって、乱塊法の場合と同じく、一般にあまり推薦できないであろう。やは

り 2 組ごとに順位を付け直して、Steel-Dwass 検定を行うか、あるいは、同じく 2 組ごとに順位を付け直して Bonferroni 不等式や Dunn-Šidák 不等式を用いて検定を行うほうが、検出力は弱いが確実である。なお、教科書によっては、Kruskal-Wallis 検定や Friedman 型検定の後に、joint ranking のまま「LSD 法」と呼ばれる方法に相当する検定を行っているものがある(CONOVER, 1980; SPRENT, 1993)。しかし、もともと LSD 法自体が多重比較法ではないので、この方法は論外とすべきであろう。

おわりに

最近の統計ソフトは分散分析と同時に、ほとんど自動的に多重比較の計算結果も出力してくれるため、多重比較法の利用例はますます多くなりつつあるようである。しかし、データが構造化されているにもかかわらず、無批判に多重比較法を用いて分析を行ったために、あまりクリアでない結果しか得られないという事態も頻繁に生じているように見受けられる。データ分析にあたっては、まず分散分析モデル自体に十分工夫をこらした後に慎重に正しい多重比較法を活用するべきだと思われる。

本連載にあたっては農業環境技術研究所の三輪哲久博士にいろいろとご教示いただいた。ここに記して感謝したい。

引用文献

- 1) CONOVER, W. J. (1980) : Practical Nonparametric Statistics. 2nd edition. Wiley.
- 2) HETTMANSPERGER, T. P. (1984) : Statistical Inference Based on Ranks. Wiley.
- 3) HOCHBERG, Y. and A. C. TAMHANA (1987) : Multiple Comparison Procedures. Wiley.
- 4) IMAN, R. L. and J. M. DAVENPORT (1980) : Comm. Statist. A9 : 571~595.
- 5) MACDONALD, B. J. and W. A. THOMPSON (1967) : Rank sum multiple comparison in one and two-way classifications. Biometrika 54 : 487~497.
- 6) SPRENT, P. (1993) : Applied Nonparametric Statistical Methods. 2nd edition. Chapman & Hall.
- 7) SAS Institute (1990) : SAS/プロシジャリフェレンスガイド Release 6.03 Edition. SAS 出版局.
- 8) 白旗慎吾 (編) (1987) : 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.
- 9) THOMPSON, G. L. (1991) : Biometrika 78 : 697~701.
- 10) 米澤勝衛・佐々木義之・今西茂・藤井宏一 (1988) : 生物統計学, 朝倉書店.
- 11) 高木正見 (1985) : 植物防疫 39 : 487~491.
- 12) ZAR, J. H. (1984) : Biostatistical Analysis. 2nd edition. Prentice-Hall.