植物防疫基礎講座

正しい分散分析結果を導くための変数変換法

独立行政法人農業環境技術研究所 山村光司（やまむら こうじ）

は じ め に

表-1 は 1993 年に本誌の基礎講座で多重比較法の解説を行った際に用いたデータである（山村，1993 a；b；c)。当時は多重比較法に関して間違った手法が頻繁に使われていたようである。DUNCAN (1955) の多重範囲検定法はそれら間違った手法の中でも代表的な存在であった。これは DUNCAN (1955) が間違っていたというよりも，正確にはユーザーの誤用が広がっていたというべき現象である（山村，1998)。なぜ誤用がこれほど広く拡散したかという問題はそれ自体がたいへん面白い研究課題であるように思われる。ただし，2000，2001 年の日本応用動物昆虫学会誌 (Applied Entomology and Zoology) に掲載された論文を調べてみると，DUNCAN (1955) の多重範囲検定法はもはや全く見られなくなっている。ノンパラメトリック検定後に Dunn 法といった不適切な多重比較法が依然として一部で使用されてはいるものの，全体的には多重比較法の誤用はかなり減少したといえる。しかし，また別の面で不適切な検定処理がしばしば見られる。本稿では 1993 年の解説では十分には記述しきれなかった問題の中から，特に変数変換に関連した話題について解説してみたい。なお，本稿は 2002 年度の日本応用動物昆虫学会で開催された小集会「操作実験における方法論」でお話しした内容に手を加えたものである。本文に先立ち，小集会を企画された戒能洋一氏，高林純示氏に感謝したい。

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	トラップ番号	各月の誘殺個体数			
		5月	6月	7月	8月
A	1	10	26	45	356
A	2	8	16	55	341
B	3	16	48	112	874

Transformation Formulae for Performing Correct ANOVA. By Kohji YAMAMURA
(キーワード：変数変換，分散分析，等分散性，Taylor's power law，Box-Cox 変換，ノンパラメトリック検定，一般化線型モデル)

I なぜ変数変換が重要か

分散分析においては「測定値の分散（ばらつき）が平均値によらず一定である」という「等分散性」の仮定が必要とされる。しかし，例えば個体数データの場合には測定値の分散は平均値が大きくなるにつれて大きくなる。このため変数変換を行わずに分散分析を行うと有意差が検出できなくなることも多い。例えば表-1 のデータで三つのトラップに捕獲される個体数に差があるかどうかを検定してみよう。ここでは月をブロックと見なして「乱塊法の分散分析」を行うことにする。この場合，対数変換 $\ln(x)$ を施してから分析を行うと三つのトラップ間の有意確率は $P=0.0007$ であり非常に有意となる。しかし，変数変換を行わずに分散分析を行うと有意確率は $P=0.283$ であり，全く有意差は見られなくなる。また平方根変換 $\sqrt{x}$ を施してから分散分析を行うと有意確率は $P=0.0793$ であり，やはり 5%では有意差が見られない。この例からも，適切な変数変換法を選択することがいかに重要であるかが示唆される。

II 個体数の平均—分散関係に関する論争

一般に生物個体数の分散 s^2 はその平均 m とともに大きくなる。表-1 のデータにおいても，5 月から 8 月にかけて平均個体数が増加すると同時に，その個体数のバラツキも大きくなっている。イギリスのロザムステッド農業試験場の TAYLOR (1961) は個体数の平均と分散の間に次式が経験的に成立しやすいことを示唆した。

$$s^2=am^b \quad (1)$$

ここに a, b は定数である。この式の両辺の対数をとれば

$$\log(s^2)=\log(a)+b\log(m) \quad (2)$$

したがって，$\log(s^2)$ と $\log(m)$ の間に直線関係が成立する。この関係は Taylor's power law と呼ばれてきた。一方，京都大学の IWAO (1968) は次式が理論的かつ経験的に成立しやすいことを示唆した。

$$s^2=(\alpha+1)m+(\beta-1)m^2 \quad (3)$$

ここに α と β は定数である。平均こみあい度 m^* $(=m+s^2/m-1)$ を用いると上式は m^* と m の間の直線関係に帰着する。