

Key-factor/key-stage 分析法による経時的測定データの解析

農業環境技術研究所 ^{やま}山 ^{むら}村 ^{こう}光 ^じ司

はじめに

薬剤散布が害虫個体数に与える影響を調べたり、施肥量が作物生長量に与える影響を調べたりするとき、実験を開始した後に複数の時点で計測を行うことが多いであろう。表-1 は計算例を示すために仮想的に作成したデータであるが、現実場面でもこのような形のデータが得られることが多いのではないだろうか。この仮想データでは、ある薬剤が害虫生存率に及ぼす影響を調べるために、薬剤量を 0, 10, 20, 40 と 4 段階に変えて実験を行った場面を想定している。各試験区において 120 匹の昆虫を供試し、実験開始後 1 日後の生存個体数 (N_1)、2 日後の生存個体数 (N_2)、3 日後の生存個体数 (N_3) を記録している。 N_1 , N_2 , N_3 を三つの发育ステージにおける個体数と見なせば、表-1 のデータは一種の生命表データのようにも見える。本稿では、このような経時的測定データを分析する際の注意点について述べ、その分析法として、生命表解析法の一つである key-factor/key-stage 分析法が活用できることを示したい。

I 誤った分析法

分散分析を始める場合に最初に行わなければならない

表-1 仮想データ
薬剤投与後に生存個体数を 1 日おきに 3 回調べたとする。

処理方法	薬剤の有無	薬剤量	個体数		
			N_1	N_2	N_3
A	なし	0	80	53	37
A	なし	0	85	58	40
B	あり	10	73	46	24
B	あり	10	81	51	25
C	あり	20	65	42	18
C	あり	20	63	39	15
D	あり	40	54	37	9
D	あり	40	58	35	12

Application of Key-factor/key-stage Analysis to Repeated Measurement Data. By Kohji YAMAMURA

(キーワード: Key-factor/key-stage 分析, 経時的測定データ, 繰り返し測定分散分析, 生命表, JMP)

のは変数変換である。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」の前提が必要とされる。しかし、実際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とが多い。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。個体数の場合には測定値の分散は平均値の二乗に近いオーダーで大きくなることが多いので、等分散性に近づけるための変換としては対数変換がしばしば用いられる。そのため本稿の分析例でも常用対数個体数 $\log(N_1)$, $\log(N_2)$, $\log(N_3)$ を用いることにする。最適変数変換式を決定するための手法については YAMAMURA (1999 a) などに解説がある。

変数変換式が決まったら、次にどのような分散分析法を採用するかを決めなければならない。経時的測定データを分析する際にはこの時点で間違った分析法が採用される可能性が高い。まず、第一に、表-1 のデータは 4×3 型の二元配置型実験計画と見なして分散分析を行うこともできる。「処理方法」(4水準)と「時間」(3水準)という二つの要因が存在し、それぞれの組み合わせごとに二つずつ反復が存在すると見なすわけである。また、表-1 のような経時的測定データは分割区法 (split plot design) のデータとして処理されていることも多い。データの持つ誤差は、各行の三つのデータ $\log(N_1)$, $\log(N_2)$, $\log(N_3)$ に共通の等分散誤差 (1次誤差) および相互に独立な等分散誤差 (2次誤差) の和からなるとみなすのである。しかし、各行の $\log(N_1)$, $\log(N_2)$, $\log(N_3)$ は同じ試験区の個体数を経時的に測定して得られたものであるから、2次誤差間の独立性が保たれていない。このため、これらの分析法を用いると、本当は差がないにもかかわらず間違って有意差を出してしまう可能性が高くなる。つまり、これらの検定では有意水準が所定の値、例えば 5% に保たれていないのである。繰り返し測定値を誤って反復 (replication) と見なして計算に用いた場合、それは一般に pseudoreplication と呼ばれている (山村, 1999)。

II 繰り返し測定分散分析

表-1のようなデータを分析するための定番の方法としては繰り返し測定分散分析 (repeated measurement ANOVA) と呼ばれる手法がある。この方法においては、時間的な変化が存在するか否かを検定する際に、まず測定値の時間的な差を計算する。今の場合 $d_1 = \log(N_2) - \log(N_1)$, $d_2 = \log(N_3) - \log(N_2)$ を計算する。そしてこの二つの変量 d_1 , d_2 について多変量分散分析 (MANOVA) を行う (Cole and Grizzle, 1966)。そこで「切片=0」の検定を行って有意差が見いだされたならば、それは「 d_1 , d_2 のいずれかがゼロではない」ことを示している。これは「時間的な差がゼロではない」ということであるから「時間的な変化が有意に存在する」ことを意味しているわけである。また「 d_1 , d_2 に対して処理方法が有意に影響を及ぼしている」という結果が出たならば、それは「時間変化が処理方法によって有意に異なる」ということであるから「処理方法と時間の交互作用が有意である」ことを意味している。論理的な解釈が遠回りでありわかりづらいが、これは正しい検定になっている。

表-1のデータを分析すると表-2のようになる。処理方法と時間の交互作用が有意である ($P=0.021$) ことか

表-2 対数個体数に関する繰り返し測定分散分析

要因	F	df1	df2	P
処理方法	47.2	3	4	0.001
時間	1609.7	2	3	<0.001
処理方法×時間	6.2	6	6	0.021

処理方法×時間の F 値は WILKS の Λ 統計量の F 近似。その他の F 値は正確な値。

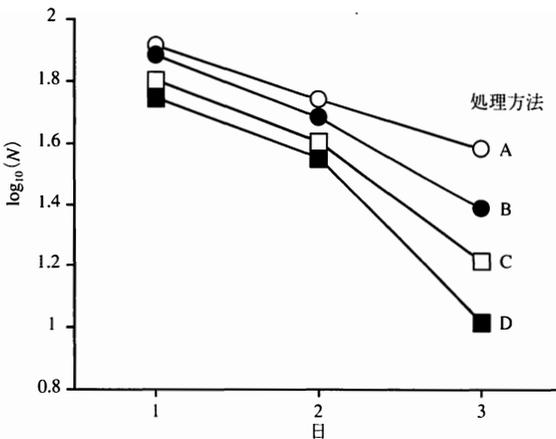


図-1 薬剤投与後の生存個体数の時間的変化
表-1の各処理方法の平均値を示す。

ら、処理方法の効果が時間によって異なることがわかる。分散分析において交互作用が有意になった場合には平均値のグラフを描いてみるのが定石である。今の場合、そのグラフ (図-1) を見ると、1日後にすでに処理方法間にある程度の差が出ていること、1~2日目の間にはあまり差が広がっていないこと、そして2~3日目の間にまた差が広がっていることがわかる。このグラフからも、確かに処理方法の効果は日によって異なることが読みとれる。ここで、次に知りたくなるのは「薬剤の効果が何日目に関何パーセント生じているか」といった点ではないであろうか。このような設問に対して繰り返し測定分散分析は十分に答えてくれないようである。しかし、後述する key-factor/key-stage 分析法を用いれば、このような設問にも対応することができる。

III Key-factor 分析法

個体群動態学分野では、生命表を分析する際に key-factor 分析法という手法がしばしば用いられてきた。もともとこの分析法は「個体数の世代間変動がどの要因によって主としてもたらされているか」といった問題を扱っていた。個体数の世代間変動は生存率や増殖率の変動によってもたらされており、これらのコンポーネントはかけ算の形で作用する。対数スケールにすればそれらは足し算の形になるので変動を定義するうえで都合がよい。たとえば幼虫期の生存率が一定量だけ増加したとき、対数スケールでは、個体数はもとの個体数レベルに依存せずに一定量だけ増加する。このためほとんどの key-factor 分析では対数スケールで変動を議論してきた。前節では「等分散性を満たすため」という統計学上の都合から対数変換を採用した。しかし、key-factor 分析法ではそれとは別の生物学上の理由から対数変換を採用してきたわけである。ただし、このように足し算の形になるようなスケールを選べば、各コンポーネントに付随する誤差も足し算の形で作用するので、個体数の誤差が個体数レベルに依存しなくなり等分散性が満たされやすい。したがって、両者が一致しているのは必ずしも偶然ではない。

Key-factor 分析法を用いて個体数変動を分析する場合には、例えば以下のような手順となる。いま生命表が s 個の発育ステージからなるとしよう。第 i ステージにおける対数生存率を $f_i = \log(N_{i+1}) - \log(N_i)$ とし、最初のステージの対数個体数を $f_0 = \log(N_1)$ と記すと、第 s ステージの対数個体数は次のように表すことができる。

$$\log(N_s) = \sum_{i=0}^{s-1} f_i \quad (1)$$

この分散 $\text{var}[\log(N_s)]$ は次のように分割することができる。

$$\text{var}[\log(N_s)] = \sum_{i=0}^{s-1} \text{cov}[f_i, \log(N_s)] \quad (2)$$

ここに cov は共分散を示す。したがって、右辺の $\text{cov}[f_i, \log(N_s)]$ の大きさを相互に比較することによって、「どの発育ステージが $\log(N_s)$ の変動に何パーセント寄与しているか」を判定することができる。従来の key-factor 分析法では、最も寄与の大きいステージのことを key-factor (変動主要因) と呼んでいた。しかし、混乱を避けるために、これは key-stage と呼んでおくべきであろう。

IV Key-factor/key-stage 分析法

Key-factor 分析法と分散分析法を組み合わせれば「どの要因がどのステージに働くことによって個体数を決定しているか」を定量的に判定することができるようになる。YAMAMURA (1999 b) はこの方法を key-factor/key-stage 分析法と呼んだ。計算方法は後で述べることにして、まずこの方法を表-1 の3日後の個体数対数値 $\log(N_3)$ に応用するとどのような結果が得られるかを例示してみよう。表-3 がその分析結果であり、この表では key-factor/key-stage 分析表と分散分析表を横方向につなげた形で示してある。表中の「合計」の列は $\log(N_3)$ に関する分散分析表の「平方和」の列である。ただし、比較をしやすくするために数値を10,000倍してある。その右側の F および P の列は $\log(N_3)$ に関する分散分析における検定結果である。表-3の中央の3列は(2)式を用いて平方和を三つのステージ成分に分解したものである。2~3日目の生存率(f_2)を通じた寄与が2,330であり最も大きい。したがって、処理方法は主として2~3日目の生存率を通じて $\log(N_3)$ の個体数に影響を及ぼしていることがわかる。0~1日目の生存率(f_0)を通じた寄与は1,096であり、2~3日目の生存率を通じて寄与と比べると半分程度である。また、1~2日目

の生存率(f_1)を通じた寄与はわずか107であり、他のステージの寄与と比べると1割以下でしかない。

ただし、表-3の分析では実験デザインの持っているさまざまな情報が十分に生かされているとはいえない。一般に、手持ちの情報を十分に活用した分散分析を行うと、計算方法は複雑になるものの、結論はシンプルかつクリアーになることが多い(山村, 1993; 1998)。今の場合は四つの処理群は薬剤量という定量的な情報を持っている。このような定量的な情報を持ったデータの場合には回帰分析型の分析が薦められる。また、今の場合には表-4に示したような形の分析も有効であろう。ここでは、まず四つの処理方法を「薬剤の有無」によって二つのグループに分けている。薬剤量0の処理方法はいわば「薬剤無処理グループ」に属しており、薬剤量10, 20, 40の三つの処理方法は「薬剤処理グループ」の中に属する。つまり処理方法が入れ子構造になっているのである。表-4ではさらに薬剤量10, 20, 40の三つの処理方法の間で薬剤量の常用対数値に関する2次式回帰を行っている。このため全体としては「狭義の分散分析」と回帰分析が混ざった形になっている。分散分析表の P 値を見ると「薬剤の有無」に関しては $P=0.001$ であり、薬剤を使用することによって有意に $\log(N_3)$ が小さくなるのがわかる。また「 $\log(\text{薬剤量})$ 1次項」について $P=0.002$ であることから、 $\log(\text{薬剤量})$ を増すことによって有意に $\log(N_3)$ が小さくなるのがわかる。また「 $\log(\text{薬剤量})$ 2次項」については $P=0.798$ であり有意ではない。したがって、 $\log(\text{薬剤量})$ の増加とともに $\log(N_3)$ は曲線的ではなく直線的に減少していると(暫定的に)見なすことができる。ステージ間を比較すると、「薬剤の有無」については、 f_2 成分が f_0 成分の2倍以上である。「 $\log(\text{薬剤量})$ 1次項」に関して、 f_2 成分は f_0 成分よりもかなり大きい。また「 $\log(\text{薬剤量})$ 1次項」の f_1 成分は負の値である。今の場合

表-3 $\log(N_3)$ に関する key-factor/key-stage 分析表および分散分析表

一元配置型の計算例。数値間の比較を容易にするために各要素は10,000倍にしてある。

要因	自由度	ステージ			合計	F	P
		f_0	f_1	f_2			
処理方法	3	1,096	107	2,330	3,532	40.4	0.002
誤差	4	33	-25	108	117		
合計	7	1,129	82	2,438	3,648		

表-4 $\log(N_3)$ に関する key-factor/key-stage 分析表および分散分析表

データの持つ「構造」と「量定性」の両方を組み入れた Type I 分析の例。数値間の比較を容易にするために各要素は10,000倍にしてある。

要因	自由度	ステージ			合計	F	P
		f_0	f_1	f_2			
薬剤の有無	1	584	140	1,419	2,143	73.5	0.001
$\log(\text{薬剤量})$ 1次項	1	514	-33	906	1,387	47.5	0.002
$\log(\text{薬剤量})$ 2次項	1	-2	0	4	2	0.1	0.798
誤差	4	33	-25	108	117		
合計	7	1,129	82	2,438	3,648		

は絶対値が小さいのでこれは「ふれ」と解釈すべきだが、もし絶対値が大きかったならば「この時期には薬剤量の増加は $\log(N_3)$ を逆に増加させる方向に働く」と解釈することができる。Key-factor/key-stage 分析法を使えば、このような複雑な相互作用も明らかにすることが可能になる。

V 計算手順

Key-factor/key-stage 分析表は、多変量分散分析 (Multivariate Analysis Of VAriance: MANOVA) をサポートする統計解析ソフトウェアを用いれば簡単に作成することができる。多変量分散分析の計算の中間段階として平方和積和行列 (Sum of Squares and Products Matrix: SSPM) が計算される (CHATFIELD and COLLINS, 1980)。Key-factor/key-stage 分析表の中の各数値は、この SSP 行列の各行を合計したものと同じなのである。ここでは SAS Institute のパソコン向け統計解析ソフトウェアである JMP を例にとりて表-3 の計算手順を追ってみたい。

まず表-1 のデータを入力した後に、 $f_0, f_1, f_2, \log(N_3)$ の欄を作り、計算式を入力することによりこれらの値を計算しておく。「処理方法」と「薬剤の有無」の二つの要因はいずれも類別変数であるので、その指定を忘れずに行っておく。次にメニューバーの Analyze メニューから Fit Model を選択すると、モデル指定のために、ここで図-2 のような画面が現れる。左ボックス中の変数一覧から f_0, f_1, f_2 の三つを選んで Y ボタンを押すことにより、これらを Pick Role Variables のボックスに入れる。また、変数一覧から「処理方法」を選んで Add ボタンを押すことにより、これを Construct Model Effects のボックスに入れる。次に右上の Per-

sonality プルダウンメニューから Manova を選択し、Run Model ボタンで実行する。出力結果の中の Overall E & H Matrices を開くと、そこには図-3 に示されるような SSP 行列が現れる。このうち Whole Model II と Intercept の行列は使わない。残りの行列について、各行の三つの数値を足しあわせる。たとえば「処理方法」の SSP 行列の f_0 の行については $0.03509+0.00262+0.07185=0.10956$ である。これを 10,000 倍すると 1096 となる。この値を表-3 の「処理方法」の f_0 の欄に書き込む。行列の他の行についても同じような計算を行って表-3 の対応する欄に書き込む。誤差に関する成分については図-3 の一番上の「E」の SSP 行列から計算する。最後に表の各行、各列の合計値とその総合計値を計算することによって key-factor/key-stage 分析表を完成させる。表-3 の右部分の分散分析表については、図-2 において f_0, f_1, f_2 の代わりに $\log(N_3)$ を Pick Role Variables ボックスに入れ、Personality プルダウンメニューから Standard Least Squares を選択して Run Model ボタンを押すことにより計算することができる。

複数の要因を分析する場合も上と同様の手順によって表を作成することができる。ただし、要因効果が相互に「直交しない」という条件の場合には少し工夫が必要である。反復数の等しい実験では一般に要因効果は直交している。しかし定量的な要因が二つ以上存在する場合 (重回帰分析型の場合) には要因効果は直交していないことが多い。表-4 の分析の場合にも「 \log (薬剤量) 1 次項」と「 \log (薬剤量) 2 次項」という二つの定量的要因が存在しており、要因効果が直交していない。この

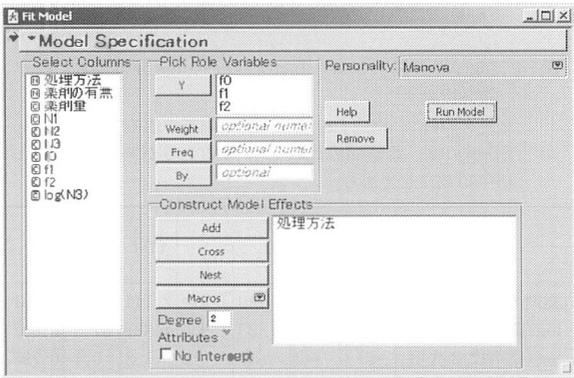


図-2 統計解析ソフトウェア JMP におけるモデルの指定画面

表 1: Fit Manova

Manova Fit			
Overall E&H Matrices			
E			
	f0	f1	f2
f0	0.00194008	-0.00056688	0.00195084
f1	-0.00056688	0.00177723	-0.0037039
f2	0.00195084	-0.0037039	0.0125964
Whole Model H			
	f0	f1	f2
f0	0.03509217	0.00262515	0.07184581
f1	0.00262515	0.0010097	0.00701872
f2	0.07184581	0.00701872	0.15408856
Intercept			
	f0	f1	f2
f0	27.0568115	-2.8096567	-5.0967682
f1	-2.8096567	0.29176279	0.52926299
f2	-5.0967682	0.52926299	0.96009267
処理方法			
	f0	f1	f2
f0	0.03509217	0.00262515	0.07184581
f1	0.00262515	0.0010097	0.00701872
f2	0.07184581	0.00701872	0.15408856

図-3 平方和積和行列 (SSP 行列) の出力画面

ように要因が直交していない場合には「Type I の SSP 行列」なるものを用いて計算を行う必要がある（山村, 2000）。この Type I というのは、要因に優先順位を仮定し、ある要因は別の要因が働いた後の残りに対して働くと仮定して平方和を分割してゆく方式である。表-4 の場合には、三つの要因間の優先関係は比較的明らかである。「薬剤の有無」は薬剤量の変化の影響よりも優先する要因であるし、また「log (薬剤量) 2 次項」は「log (薬剤量) 1 次項」では説明できない残りの部分を説明するためのものである。JMP ではいわゆる「Type III の SSP 行列」しか計算してくれないため、表-4 を JMP で計算する場合には、例えば以下のような手順をとる必要がある。まず、データシートに「log (薬剤量) 1 次項」と「log (薬剤量) 2 次項」の欄を追加して、そこに薬剤量の常用対数値とその二乗値をそれぞれ入力する。薬剤量が 0 のデータについては log (薬剤量) が計算できないのでその欄には 0 か何か適当な値を入れておく。薬剤量の要因は「薬剤の有無」に入れ子構造になっているので、どのような値を入れても結果は同じである。次に、これら三つの要因すべてを入れた多変量分散分析によって SSP 行列を計算し、これを用いて表-4 の「log (薬剤量) 2 次項」と「誤差」の行の要素を計算する。次に「薬剤の有無」と「log (薬剤量) 1 次項」だけを入れた多変量分散分析によって SSP 行列を計算し、そこから「log (薬剤量) 1 次項」の行の要素を計算する。そして最後に「薬剤の有無」だけを入れて SSP 行列を計算し、そこから「薬剤の有無」の行の要素を計算する。このように今の場合には多変量分散分析を 3 回行う必要がある。なお、統計パッケージの SAS を用いれば Type I の SSP 行列を一回の計算で求めることができ

る。

おわりに

分散分析は「どの要因が有意に影響しているか」を示してくれるが「どの要因がどのステージを通じて影響しているか」については定量的には示してくれない。その欠点を補うための一つの方法として key-factor/key-stage 分析法は役に立つのではないと思われる。上の例では個体数を扱ったため対数変換後の測定値を分析したが、測定対象によっては変換せずに分析を行うべき場合もあるであろう。その場合、例えば表-1 においては、まず f_0, f_1, f_2 の代わりに $N_1, (N_2 - N_1), (N_3 - N_2)$ の三つの値を計算する。そして、これらの値に関する SSP 行列を計算して key-factor/key-stage 分析表を構築することになる。また、その際も、表-4 の例に見られるように、できる限り手持ちの情報を生かした解析を行うように心がけるべきであろう。「要因 (factor)」と「ステージ (stage)」の両方の情報を適切な形で分析に組み込むことができれば、そこから得られる結論は飛躍的に面白くなる場合も多いのではないと思われる。

引用文献

- 1) CHATEFIELD, C. and A. J. COLLINS (1980): Introduction to Multivariate Analysis (福場庸, 大沢豊, 田畑吉雄訳「多変量解析入門」). 1986. 培風館
- 2) COLE, J. W. L. and J. E. GRIZZLE (1966): Biometrics 22: 810~828.
- 3) 山村光司 (1993): 植物防疫 47: 370~375.
- 4) ——— (1998): 日本土壤肥料学雑誌 69: 649~653.
- 5) ——— (1999): 同上 70: 84~89.
- 6) ——— (2000): 個体群生態学会会報 57: 15~21.
- 7) YAMAMURA, K. (1999 a): Res. Popul. Ecol. 41: 229~234.
- 8) ——— (1999 b): Ecology 80: 533~537.

主な次号予告

次号 9 月号に掲載が予定されている記事は次のとおりです。

静岡県におけるメロン黄化えそ病根絶の経過
池田二三高他

グラウンドカバープランツに発生する病害
竹内 純

パセリの土壌病害
清水時哉

エンドファイト研究の現状と展望
古賀博則

わが国におけるトマト萎凋病菌レース 3 の新発生
石井貴明

我が国のリンゴ・ナシに発生するウイロイド病
伊藤 伝他

新生物農薬タイリクヒメハナカメムシの使い方)
土屋 亨

植物防疫基礎講座
ミダレカクモンハマキの周年飼育法 佐藤 威

談話室
クリタマバチの生物的防除—熊本県における成功への軌跡— 村上陽三

トビックス
農薬ラベル表示の変更—MSDS に対応して— 平井 勇

臭化メチル削減計画の現状 東 義弘

定期講読者以外のお申込みは至急前金にて本会へ
定価 1 部 920 円 送料 76 円