植物防疫基礎講座：

一般化線型モデルとモデル選択

―統計解析の新しい流れ―

農業環境技術研究所　山村光司（やまむらこうじ）

はじめに

かつては分散分析と回帰分析は別々の手法として扱われてきたが，現在の統計解析ソフトウエアでは，これらはいずれも「一般線形モデル（general linear model）」の一部として扱われることが多くなっている。これをさらに拡張した手法である「一般化線型モデル（generalized linear model）」も現在では頻繁に用いられつつある。また，解析の結果は以前は有意確率（P値）によって議論されていたが，最近では赤池情報量基準（AIC）などのモデル選択基準によって議論されることも多くなっている。本稿では，これらの比較的新しいアプローチの必要性とその使用上の留意点について解説を試みたい。

I　一般線形モデル（線形モデル）

表-1は山村（2002）で用いたハスモンヨトウの誘殺数データの一部である。このデータには二つの要因（トラップと月）が含まれている。それらの要因が，ハスモンヨトウの対数誘殺数に影響を与えているかどうかを調べたいとする。二つの要因のうち，トラップは定性的要因であり，月は定量的要因であるとする。昔の解析方針に従えば，定性的要因は分散分析で分析を行い，定量的要因は回帰分析で分析を行うことになる。しかし，定性的要因であれ定量的要因であれ，観測値を「要因効果の和」として説明するという概念は共通している。このことから，定性的要因と定量的要因の両方を含めた分析が可能であり，それらは一般線形モデルあるいは単に線形モデルと呼ばれている。

例えば，単回帰分析において，n個のデータに$y = a + bx$という直線式をあてはめるとする。これはデータに次のモデルを想定することと同じである。

$$y_i = a + bx_i + e_i \quad (i = 1, 2, \cdots, n) \qquad (1)$$

ここにy_iは第i番目の観測値であり，x_iは説明変数，e_iは正規分布に従う誤差である。これは次のように行列形式で書き表すことができる。

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad (2)$$

これは次のような形をしている。

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \qquad (3)$$

ここに$\mathbf{y}$は観測値を縦に並べたベクトルであり，$\mathbf{X}$はデザイン行列と呼ばれる行列である。$\mathbf{b}$は推定すべきパラメーターを縦に並べたベクトルであり，$\mathbf{e}$は正規分布に従う誤差e_iを縦に並べたベクトルである。次に分散分析の場合を考えてみよう。一元配置分散分析において，二つの処理水準を設けて完全無作為配置でそれぞれの処理を2回ずつ繰り返したとき，モデルは

$$y_{ij} = T_i + e_{ij} \quad (i = 1, 2\,;\, j = 1, 2) \qquad (4)$$

ここにy_{ij}は第i番目の処理の第j番目の観測値である。T_iは第i番目の処理の効果であり，e_{ij}は正規分布に従う誤差である。このモデルは次のように書かれることのほうが多い。

$$y_{ij} = \mu + T_i + e_{ij} \quad (i = 1, 2\,;\, j = 1, 2) \qquad (5)$$

ここにμは2処理の基準となる効果である。この式は次のように書き表すことができる。

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ T_1 \\ T_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \end{bmatrix} \qquad (6)$$

これもやはり$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$の形をしている。デザイン行列$\mathbf{X}$を比べると，回帰分析の場合は値がギッシリと詰まっているのに対して，分散分析の場合は0と1からなるスカスカの行列になっている点だけが異なる。このため，回帰分析も分散分析も基本的には同様の操作で推定・検定を行うことができる。また，回帰分析と分散分

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

トラップ番号	各月の誘殺個体数			
	5月	6月	7月	8月
1	8	16	55	341
2	16	48	112	874

Generalized Linear Models and Model Selection.　By Kohji YAMAMURA

（キーワード：一般線形モデル，一般化線型モデル，尤度比検定，過分散，モデル選択，予測力）

析を混在させることも容易であることがわかる。

表−1のデータをいくつかのモデルで分析した結果を表−2に示した。モデルAではトラップに関する分散分析，モデルBでは月に関する回帰分析を行っている。モデルCではトラップと月の分析を同時に行っており，モデルDではトラップと月の交互作用までを含めた分析を行っている。平方和の部分を見ると，このType I分散分析の場合には，該当する要因に対応する平方和はモデルA～Dのいずれでも同じになっている。しかし，有意確率P値はモデルによって大きく異なっている。モデルCおよびDではトラップ間に5%水準で有意差が見られるが，モデルAでは有意差は見られない。

「モデルAの分散分析とモデルBの回帰分析を別々に行えばよいから，線形モデルの概念などは不要である」という意見もあるかもしれない。しかし，そのように一部の要因だけを取り出して分析を行うと結果が誤ったものになる。一般に，観測値は次のような成分からなっている。

$$(\text{観測値}) = (\text{要因効果による期待値}) + (\text{誤差成分}) \quad (7)$$

右辺の第1項は観測した要因効果のすべてを正確に用いたときの期待値を表している。手持ちの要因をすべて正確に組み込んだとしても制御できない変動成分があり，これが右辺の第2項である。モデルAの分散分析では要因効果のうちのトラップの成分だけを取り込んで，残りは誤差成分とみなしてしまっている。これでは誤差成分を過大に推定してしまう。モデルAの分散分析で有意差が見られなかったのはそのためである。誤差成分の推定に関しては，誤差成分の自由度の問題もあるが，できる限り多くの要因を考慮して推定を行うべきである。今のデータの場合は，すべての要因を組み込むと，例えば表−2のモデルDのような分析となり，分散分析と回

表−2　ハスモンヨトウのデータの対数変換値 $\log_e(x+0.5)$ に5種類のモデルを当てはめた分析結果

変動因	df	SS	F値	P値	$\hat{\phi}$	R_D	AICc
A. トラップに関する分散分析（$k=2$）							
トラップ	1	1.44	0.5	0.50			
誤差	6	16.36			2.73	0.06	40.43
B. 月に関する回帰分析（$k=2$）							
月	1	15.68	44.5	＜0.01			
誤差	6	2.11			0.35	0.85	24.06
C. トラップと月に関する分析（$k=3$）							
トラップ	1	1.44	10.6	0.02			
月	1	15.68	115.4	＜0.01			
誤差	5	0.68			0.14	0.91	24.31
D. 交互作用までを入れた分析（$k=4$）							
トラップ	1	1.44	8.5	0.04			
月	1	15.68	93.1	＜0.01			
月×トラップ	1	0.01	0.0	0.87			
誤差	4	0.67			0.17	0.89	42.91
E. 切片のみを当てはめた分析（$k=1$）							
誤差	7	17.80			2.54	0	35.50

kはモデルに含まれるパラメーター数，dfは自由度，SSはType I平方和，$\hat{\phi}$は誤差分散の推定値を示す．正規分布誤差の場合には逸脱度Dは残差平方和であり，これは誤差の行のSSの欄に記載されている．R_Dの計算では，検定の場合と同様に，最も多くの要因効果を含むモデル（モデルD）から計算された不偏分散推定値（$\hat{\phi}=0.17$）を使用する．例えば，モデルCでは$R_D = 1-(0.68+2\times3\times0.17)/(17.80+2\times0.17) = 0.91$．一方，AICcの計算ではモデルごとに別々の$\phi$を最尤推定するため，パラメーター数$k$は一つ多くなる．

帰分析を組み合わせた線形モデルとなる。

Ⅱ　一般化線型モデル

表−3はアブラムシが存在すると葉に穴ができやすいか否かを調べるための実験の結果である（Crawley, 1993）。アブラムシが存在すると，葉の化学組成が変化して，後に他の食葉性昆虫に食害されにくくなるという「誘導防御」の可能性を検証しようとしている。2本の木のそれぞれにアブラムシの存在する葉と存在しない葉を設け，そのうちの穴のある葉数と穴のない葉数を計測している。表−1のハスモンヨトウのデータでは観測値は誘殺個体数という定量値で与えられていたのに対して，この実験では観測値は「穴がある」,「穴がない」という二つの定性値（二値データ）で与えられている。説明する対象が定量値の場合と全く異なるように見えるが，一般化線型モデルの概念を用いれば，このような定性値の場合も定量値とほぼ同様の感覚で分析することが可能となる。

一般化線型モデルでは，式（7）の右辺の二つの項を次のように拡張する。

（1）「誤差成分」は正規分布に限らず「指数分布族」の分布のいずれかであれば構わない。例えば，正規分布のほかに二項分布，ポアソン分布，ガンマ分布でも構わない。

（2）「要因効果による期待値」は，要因効果の和の形に限らず，要因効果の和を単調関数で変換したものでも構わない。例えば，要因効果の和を指数関数やロジスティック関数で変換したものでも構わない。

要因効果の和の変換関数の逆関数は「リンク関数」と呼ばれている。上の（2）を逆方向から言い換えると「要因効果による期待値をリンク関数で変換したものが要因効果の和で表される」ということである。

表−3のデータの場合は，穴のある葉数は穴の発生確率のもとで二項分布に従って変動していると考えられる。そこで式（7）の「誤差成分」として正規分布の代わりに二項分布を仮定する。さらに式（7）の「要因効果による期待値」として「要因効果の和をロジスティック関数で変換したもの」を考える。これは一般化線型モデルの一種である「ロジスティック回帰分析」であり，発生確率を分析する際に普通に用いられてきた解析手法である。ロジスティック関数の逆関数はロジット関数であるから，ここではリンク関数としてロジット関数を用いていることになる。

ロジスティック回帰分析の結果は表−4に示されている。線形モデルの分析が表−2のような分散分析（ANOVA）となるのに対して，一般化線型モデルの分析は表−4のような逸脱度分析（analysis of deviance）とよばれるものになる。この「逸脱度（D）」は正規分布誤差の場合の「残差平方和」に相当するものであり，次式で定義される（McCullagh and Nelder, 1989）。

$$D = 2\phi(l_{\mathrm{Max}} - l) \qquad (8)$$

lは用いたモデルのもとでの最大対数尤度，l_{Max}は最大モデル（maximal model）を当てはめたときの最大対数尤度である。ここに，最大モデルとは，データ数と固定効果パラメーター数が等しく，完全に適合するモデルのことである。ϕはdispersion parameterと呼ばれ，誤差分布の分散をϕ倍に拡大させる効果をもつパラメーターである。正規分布誤差の場合は，既に表−2で用いたようにϕは分散そのものであり，伝統的にはσ^2と表記されてきた。ポアソン分布と二項分布の場合は，その分布の性質からϕは1に固定されている。

あるモデルから要因を一つ除去すると，モデルの当てはまり具合が悪くなり逸脱度が増加する。このときの逸脱度の増加量が大きければ，除去した要因は有意であったと判断することができる。具体的には，スケール化された逸脱度（D/ϕ）の差が漸近的にカイ二乗分布に従うという性質を用い，表−4にあるようなカイ二乗検定を行う。式（8）からわかるように，D/ϕの差は対数尤度の差の2倍，つまり尤度比の対数の2倍なので，このカイ二乗検定は「尤度比検定」と呼ばれている。ただし，カイ二乗検定のかわりにF検定を用いるべき場面もある。表−4の二項分布の場合にはϕは1に固定されているので，カイ二乗検定をそのまま用いた。しかし，ϕが決まっておらず，検定の際にϕを推定して用いる場合に

表−3　アブラムシと葉の穴との関係に関するデータ（Crawley, 1993）

木番号	アブラムシ	穴あり	穴なし	合計
1	なし	35	1,750	1,785
1	あり	23	1,146	1,169
2	なし	146	1,642	1,788
2	あり	30	333	363

表−4　アブラムシと葉の穴との関係に関するTypeⅡ逸脱度分析．dfは自由度

要因	df	カイ二乗値	P値
木	1	104.0266	＜0.01
アブラムシ	1	0.0033	0.95
木×アブラムシ	1	0.0008	0.98

は，ϕの推定値の変動を考慮して，カイ二乗検定の代わりにF検定を用いるべきであろう。特に，正規分布誤差の場合にはF検定は正確な検定となるため，その場合は従来どおりにF検定を用いる。

尤度比検定にせよF検定にせよ，その検定統計量は二つのモデルの逸脱度の差から計算される。そのため，その差を計算する際の順序が問題となる。「要因Bは要因Aで説明できない残りの部分を説明し，要因Cは要因A，Bで説明できない残りの部分を説明する」というような意味で，要因が優劣の順に並んでいる場合がある。このときには，後ろの劣位の要因から順番に除去して逸脱度の差を計算して検定を行う。これはType I検定と呼ばれている。表−2では便宜上Type I検定を採用した。表−3の葉の穴に関するデータの場合は典型的な二元配置実験の形をしている。こうした場合はType IIと呼ばれる除去順序が一般に推薦される。交互作用項は主効果で説明できない残りを説明するためのものである。このため，まず主効果と交互作用を含むモデルから交互作用のみを除去し，その際の逸脱度の増加量から交互作用の検定を行う。そして，ここで交互作用が有意になったのならば主効果の検定へは原則として進まない。交互作用が有意でない場合には主効果の検定に進むが，このときすべての主効果を平等に扱う。つまり，交互作用を除去したモデルから主効果を一つだけを除去し，その際の逸脱度の増加量から主効果の検定を行う。この操作をすべての主効果について平等に行う。なお，すべての要因を平等に扱いたい場合にはType IIIと呼ばれる除去順序になる。ある要因の検定を行う場合には，すべての要因を含むモデルからその要因だけを除去したときの逸脱度の増加量を計算して検定を行う。重回帰分析のように，要因間に優劣関係が全く存在しない場合には，このType III検定を用いる。

III　過分散の問題

一般化線型モデルを適用する上で，最も気をつけなければならないのは過分散の問題であろう。表−5は山村・鈴木（2006）で用いた薬剤処理試験データである。処理区，対照区，無処理区のそれぞれで3回の無作為反復を行い，害虫の発生個体数を調べたとする。これは計数データ（count data）であるから，ここでは定石に従って誤差にポアソン分布を仮定しよう。また，昆虫の増殖率や生存率は掛け算の形で作用するため，各要因効果は昆虫個体数の期待値に対して掛け算の形で影響すると考えるのが妥当であろう。つまり，昆虫個体数の期待値の対数値に対して要因効果が和の形で作用すると考えるのが妥当であろう。そのため，リンク関数として対数関数を使用する。このとき，処理区と対照区における害虫の密度指数の対数値は次のように推定される（山村・鈴木，2006）。ここでは，無処理区の効果を0とおく「端点制約」によるパラメーター定義を活用している。

表−5　薬剤処理試験の数値例.「新農薬実用化試験実施の手引き」のデータの一部を改変したもの

試験区	無作為反復		
	1	2	3
処理区	25	18	24
対照区	18	13	11
無処理区	153	148	73

	推定値	標準誤差
（切片）	4.83	0.05
処理区	− 1.72	0.13
対照区	− 2.19	0.16

同じ実験処理を施した区であっても，ポアソン分布の期待値は式（7）中の「要因効果による期待値」とピッタリとは一致せずに反復ごとに変動しているかもしれない。そのような場合にこのような普通の「ポアソン回帰」を行うと精度を誤って高く見積もってしまう。ポアソン分布の場合にはもともとϕは1に固定されていた。ここではϕが1よりも大きい可能性を考慮し，期待値の変動によって誤差分布の分散がポアソン分布の分散のϕ倍に増えているとして近似的に表現してみよう。これは「過分散（over−dispersion）」と呼ばれる状況である。事前に証拠がない限りは常に過分散を考慮して分析を行うべきであろう（McCullagh and Nelder, 1989, p. 90）。過分散を考慮した場合には対数尤度ではなくquasi−likelihoodと呼ばれるものを扱っていることになる。ϕの推定には一般化Pearsonカイ二乗統計量が用いられる。表−5のデータで過分散を考慮してϕを推定すると$\hat{\phi} = 5.89$であり，結果は次のように変わる。

	推定値	標準誤差
（切片）	4.83	0.13
処理区	− 1.72	0.32
対照区	− 2.19	0.40

密度指数の推定値は変わらないが，その標準誤差（SE）は$\sqrt{\hat{\phi}}$倍に大きくなる。先ほどのポアソン回帰では推定値の精度を誤って過大に評価していたことがわかる。

期待値の変動を考慮した推定法としては，安易に上のように定数倍の過分散を仮定して推定を行う方式のほかに，その期待値の変動を明示的にモデル化する「一般化

線形混合モデル（generalized linear mixed model : GLMM）」も最近ではしばしば用いられるようになってきた。これは一般化線型モデルをさらに拡張したモデルである。昆虫個体数の期待値の対数値に対して要因効果が和の形で作用する場合には，各要因にかかわる誤差も対数スケールで和の形になる。したがって，期待値が対数スケールで等分散正規分布に従って変動しており，観測値はその期待値のまわりにポアソン分布に従って変動していると仮定するのが妥当であろう。この仮定のもとで，一般化線形混合モデルにより推定を行うと以下のようになる（山村・鈴木，2006）。

	推定値	標準誤差
（切片）	4.79	0.14
処理区	− 1.69	0.23
対照区	− 2.17	0.25

期待値の変動が大きい場合には，この一般化線形混合モデルの近似として，対数変換値 $\log_e(x + 0.5)$ に関する線形モデルを用いることができる。その推定値と標準誤差は次のようであり，上の結果とおおむね同じである（山村・鈴木，2006）。

	推定値	標準誤差
（切片）	4.78	0.17
処理区	− 1.66	0.24
対照区	− 2.12	0.24

実は，表−2のハスモンヨトウの個体数の分析では最初からこの近似方式を採用していたわけである。

Ⅳ　モデル選択と予測力

20世紀には，統計学に関する考え方の違いからBayesian，Fisherian，frequentistという3種の人類が存在してきたとされる（Efron, 1998；芝村，2004）。表−2と表−4では有意確率 P 値を用いて仮説の信頼性を表現した。これはFisherianの立場である。ところがFisherianには大きな弱点があった。それは「P 値が大きくて仮説が棄却されなかった場合には何も言えないし何も採用できない」という弱点である。この弱点を克服するため，Neyman−Pearsonは帰無仮説と対立仮説を設けて二者択一の問題として扱うことを提案した。品質管理における生産者危険率 α と消費者危険率 β をそれぞれ第1種過誤率，第2種過誤率と表現しなおし，両方の過誤率を考慮しつつ，必ずどちらかの仮説を採用する。P 値と α 値はしばしば混同されるが，両者は全く別のものである（Hubbard and Bayarri, 2003）。この立場はfrequentistと呼ばれている。ところが，実際の世界では，そのように二者択一の問題に単純化できることはほとんどなく，多くの仮説の中から最も妥当な仮説を選ばなければならない。その点ではFisherianもfrequentistにも限界がある。最近では，統計処理の際に P 値のみに頼るのではなく，多くの候補モデルの中から最適なモデルを選択するという推論が多く見られるようになってきた。モデルの優劣を判断する基準として，赤池情報量基準（AIC）をはじめ，いくつかの基準が使用されてきている。こうしたアプローチを採用する人々は，従来の3種の人類には収まらない新人類と言えるかもしれない。

最適なモデルを選択する際の最も自然な方針は「予測力」を基準にするという方針であろう。つまり，予測がよく当たる仮説やモデルを採用する。ここで「予測がよく当たる」という状況を厳密に考えると，それは「既存のデータ（データA）から構築した予測モデルのもとで，次回のデータ（データB）が発生する確率が高い」ということである。ここでは，何度も予測操作を繰り返した際の全体の発生確率の大小を問題にすべきであるから，独立事象の乗法定理により「発生確率の幾何平均」の大小を問題にすべきであろう。数値の大小関係は対数値に変換しても変化しないため，「発生確率の幾何平均」の大小関係は，その対数値すなわち「対数確率の期待値」の大小関係と同一である。いま，既存のデータにモデルを当てはめた際の最大対数尤度を l とし，モデル中のパラメーター数を k とすると「対数確率の期待値」は漸近的に $l - k$ で与えられる。この−2倍つまり $-2l + 2k$ がAICと定義されている。AICを最小にするモデルは予測力を漸近的に最大化するモデルであることがわかる。

表−2で線形モデルの検定を議論した際に，できるだけ多くの要因効果を入れたモデルのもとで推定した分散 $\hat{\phi}$ を用いて検定を行うべきであると述べた。同様のことは予測場面についても言える。いま，同じ ϕ のもとで切片だけを含むモデル（null model）をあてはめた場合の最大対数尤度を l_{Null}，その逸脱度を D_{Null} とする。正規分布誤差の場合の古典的な適合度指標である「寄与率」は次式で計算されていた。

$$R^2 = 1 - \frac{D}{D_{\mathrm{Null}}} \qquad (9)$$

式（8）にあるように，逸脱度 D は対数確率の差の定数倍で定義されている。上述のように「対数確率の期待値」は予測力を示すから「逸脱度の期待値」は予測力の差を示す。したがって，式（9）の分母分子をその期待値で置き換えたもの，つまり期待逸脱度の比（R_D）を「予測力の改善割合」を示す値として用いることができる。

$$R_D = 1 - \frac{E(D)}{E(D_{\mathrm{Null}})} = 1 - \frac{l_{\mathrm{Max}} - l + k}{l_{\mathrm{Max}} - l_{\mathrm{Null}} + 1} = 1 - \frac{D + 2k\phi}{D_{\mathrm{Null}} + 2\phi} \quad (10)$$

この値が最大となるモデルを選択する。この選択方式は，正規分布誤差の場合は Mallows の C_p 基準による選択と同一である。ポアソン分布誤差や二項分布誤差の場合で過分散を考慮しない場合には AIC 基準による選択と同一である。過分散を考慮する場合は Burnham and Anderson（2002）の QAIC 基準と同一である。正規分布誤差の場合にせよ過分散の場合にせよ，ϕ はできるだけ多くの要因を組み込んだモデルのもとで推定する。そして，推定された単一の $\hat{\phi}$ をすべてのモデルで用いてモデルを比較する（Burnham and Anderson, 2002, p. 69 ; McCullagh and Nelder, 1989, p. 91）。ϕ は「局外パラメーター（nuisance parameter）」の一種であり，要因効果の最尤推定とは別途に推定されるため，パラメーター数 k の中にはパラメーター ϕ の数 1 は加えない。

表−2 の R_D を比較すると，トラップと月の主効果だけを組み込んだモデルが最も R_D が大きく，したがって予測力が高い。式（7）の要因効果のうちの一部のマイナーな要因効果をあえて 0 と置いて予測を行うことにより予測力を高めていることになる。参考までに，表−2 には AICc の値も同時に示した。AICc は，式（7）のように観測値を手持ちの要因効果による期待値と残りの誤差成分に分けるのではなく，統計モデル自体を真であるとする誤った思想に基づいている。正規分布誤差を仮定したモデル選択において，別のモデルを推定する際に毎回その別のモデルが真であると仮定しなおして，誤差分散 ϕ をモデルごとに最尤推定しなおす。AICc はそういう誤った手順の場合の − 2（対数確率の期待値）の不偏推定値である。表−2 では，AICc は月だけを組み込んだモデルで最小となっている。なお，アブラムシと葉の穴の関係に関する分析の場合には，ϕ は 1 に固定されていたため，R_D による選択と AIC による選択は同一である。木の要因だけを含むモデルで R_D が最大となり AIC が最小となる（$R_D = 0.96$）。

予測力を最大化するモデルはデータ量に依存して変化する。データ量が増えるとパラメーターの推定精度が高まるために，より複雑なモデルの予測性能が高くなる。そのため，原則として，より複雑なモデルが選択されるようになる。それに伴って R_D の最大値もある程度まで大きくなる。もし十分に大きな R_D を示すモデルが存在しないならば，それはデータが不十分であることを意味しているため，データ量を増やしたり観測する要因数を増やしたりして改善を図らなければならないであろう。ただし，予測が目的ではなく，野外現象を要約して，現象が生じる「主なメカニズム」を把握するのが目的の場合には，R_D が大きすぎると逆に具合が悪いかもしれない。例えば，仮に $R_D = 0.8$ を「最適な要約度合い」と考えてみよう。表−2 の場合には，月の効果だけを組み込んだモデルの R_D が最も 0.8 に近いことから，このときは月の効果だけを含むモデルを「要約モデル」として採用すべきことになる。

おわりに

本稿では具体的な統計解析ソフトウエアの操作法については触れなかったが，一般線形モデルの計算を行う場合には，まず要因ごとにその性質（定性的要因か定量的要因か）を指定する。それらをモデルに含めると，多くのソフトウエアは定性的要因については分散分析型，定量的要因については回帰分析型と解釈して計算を行ってくれる。一般化線型モデルの場合は，さらに誤差分布型とリンク関数，そして過分散の有無を指定する。分散分析や逸脱度分析を行う際にはそのタイプ（Type I，II，III など）を指定する。また，推定値パラメーターの定義（端点制約，ゼロ和制約，多項式中心化など）の違いから，ソフトウエアによって推定値が異なる場合があるため注意が必要である。統計解析ソフトウエアも進歩を続けている。最新版のソフトウエアのマニュアルを参照しながら，具体的に様々な計算を行って比較を行っていただきたいと思う。

引用文献

1）Burnham, K. P. and D. R. Anderson（2002）: Model selection and multimodel inference : a practical information theoretic approach, 2nd edn., Springer, New York, 488 pp.
2）Crawley, M. J.（1993）: GLIM for ecologists, Blackwell, New York, 379 pp.
3）Efron, B.（1998）: Statist. Sci. **13** : 95 ～ 122.
4）Hubbard, R. and M. J. Bayarri（2003）: Am. Stat. **57** : 171 ～ 178.
5）McCullagh, P. and J. A. Nelder（1989）: Generalized linear models, 2nd edn., Chapman and Hall, London, 511 pp.
6）芝村　良（2004）: R. A. フィッシャーの統計理論，九州大学出版会，福岡，181 pp.
7）山村光司（2002）: 植物防疫 **56** : 436 ～ 441.
8）————・鈴木芳人（2006）: 同上 **60** : 112 ～ 116.