

第9回 続高橋セミナー
最尤法によるポアソン回帰分析入門
2020年7月8日

第9章 花数を共変量とした種子数の探索的ポアソン回帰

ポアソン回帰に関連する論文を Web で検索し、データリストを含む探索的な解析が行われている論文として、下野 (2010) の「R を用いた一般化線形モデル (回帰係数編) : カウントデータを例に」を見出した。仮想データとして、3 地域 (Region : A, B, C) において 2 つの生育環境 (Habitat : Dry, Wet) に分布する植物種各 10 個体を選び、全体で 60 個のデータで、個体あたりの生産花数 (Flower No), 生産種子数 (Seed No) を調査した結果が示されている。生産種子数が地域や生育環境といった要因によって違いがあるか探索的な結果が示されている。このデータを用い、Excel, JMP, SAS を用いて別の切り口から探索的なポアソン回帰を試みる。

第9章 目次

9.	花数を共変量とした種子数の探索的ポアソン回帰	293
9.1.	データの概観	293
9.2.	JMP のポアソン回帰による探索的解析	297
	交互作用モデル, 主効果モデル	
9.3.	無償版 SAS の GENMOD プロシジャによる主効果モデル	304
9.4.	花数をオフセットとしたポアソン回帰	307
9.5.	花数をオフセットとした負の 2 項回帰の適用	310
	負の 2 項分布のパラメータ変換, Excel によるポアソン回帰, Excel によるガンマ・ ポアソン回帰 (負の 2 項回帰), SAS の GENMOD プロシジャによる負の 2 項分布を 用いた場合, 下野の R の glm.nb による結果, デビアンس (Deviance)	
	文献索引, 索引, 解析用ファイル一覧	323

第9回 続高橋セミナー 最尤法によるポアソン回帰分析入門

第9回 続高橋セミナー「最尤法によるポアソン回帰分析入門」は、ページ数が多いので章ごとに公開する。全体の章立てを次に示す。

目次

はじめに -----	1
1. ポアソン分布に従う各種のカウント・データ-----	7
2. ニュートン・ラフソン法によるポアソン回帰 -----	63
3. 尤度比検定のためのデザイン行列-----	95
4. デザイン行列を用いた回帰分析入門-----	135
5. 反復重み付き最尤法によるポアソン回帰 -----	175
6. 過分散・ゼロ過剰への対応 -----	207
7. 過分散がある場合の探索的ポアソン回帰 -----	237
8. 2本の回帰直線の比較-----	269
9. 花数を共変量とした種子数の探索的ポアソン回帰-----	293
10. オフセットを含む探索的ポアソン回帰-----	323
11. デビアンس・逸脱度・残差・テコ比・4種の残差 -----	359
12. パラメータの共分散行列の活用 -----	383
13. 最小2乗平均の謎を予測プロファイルで解く -----	421
文献, 文献索引, 索引, 解析用ファイル一覧 -----	461

9. 花数を共変量とした種子数の探索的ポアソン回帰

ポアソン回帰に関連する論文を Web で検索し、データリストを含む探索的な解析が行なわれている論文として、下野（2010）の「R を用いた一般化線形モデル（回帰係数編）：カウントデータを例に」を見出した。仮想データとして、3 地域（Region : A, B, C）において 2 つの生育環境（Habitat : Dry, Wet）に分布する植物種各 10 個体を選び、全体で 60 個のデータで、個体あたりの生産花数（Flower No）、生産種子数（Seed No）を調査した結果が示されている。生産種子数が地域や生育環境といった要因によって違いがあるか探索的な結果が示されている。このデータを用い、Excel, JMP, SAS を用いて別の切り口から探索的なポアソン回帰を試みる。

9.1. データの概観

下野（2010）は、種子数が花数によって影響を受けるので、花数をオフセットとした過分散を調整したポアソン回帰、さらに負の 2 項分布を使った解析結果を示し、カウント・データの解析について総合的に論じている。

第 1.5 節の「冠動脈心疾患の死亡者数」で示したように、オフセットは、観測された死亡者数の部分母集団の人数の対数として扱われている。花数は、種子数に対して影響を与えると推測されるので、オフセットとして扱うことに違和感がある。花数をオフセットとするモデルとすることは可能であるが、観測された反応変数としての側面もある。そこで、花数を共変量とした探索的な解析を行い、さらに、花数をオフセットとした Excel, JMP, SAS による解析結果を示すことにした。なお、R 言語のユーザで、無償の OnDemand SAS に興味を持たれた場合には、臨床評価研究会（ACE）基礎解析分科会（2017）、「新版 実用 SAS 生物統計ハンドブック」に、SAS と R による多くの解析例が並列的に示されプログラムもダウンロードできるので参考にしてみたい。

表 9.1 に下野で示されているデータを示す。生育環境（Habitat : Dry, Wet）を表側に、地域（Region : A, B, C）を表頭にし、植物種各 10 個体あたりの生産花数（Flower No）、および、生産種子数（Seed No）を示す。全体で 60 個のデータである。

表 9.1 生育環境と地域別の花数および種子数のデータ

	Region A		Region B		Region C		
Habitat	FlowerNo	SeedNo	FlowerNo	SeedNo	FlowerNo	SeedNo	
Dry	3	57	1	22	3	67	
	1	19	1	21	3	59	
	1	11	2	36	1	10	
	1	12	1	24	2	45	
	3	46	3	45	1	8	
	3	51	2	35	3	60	
	2	21	3	53	3	47	
	2	29	2	25	1	9	
	2	28	2	23	3	79	
	2	46	3	56	1	15	
	平均		32.00		34.00		39.90
	Wet	2	26	2	42	4	97
3		61	1	27	2	45	
2		35	3	101	2	38	
4		81	3	149	3	68	
3		85	2	56	1	28	
1		34	1	29	1	33	
1		25	1	35	2	55	
1		31	2	49	1	14	
2		48	3	86	4	129	
2		70	3	73	3	101	
平均			49.60		64.70		60.80

下野は、オフセットが有効なデータとして

- 1) ある面積の中に生えていた雑草個体数
- 2) ある土壌体積中に含まれていた埋土種子数
- 3) あるサイズの個体が生産した種子数
- 4) ある花数の個体が生産した種子数

などを挙げている。1) と 2) は、(ある面積中, ある土壌の体積中) は、その場所に存在する各個体の部分母集団と見なし、オフセットとして扱うことが有効と思われる。しかし、3) と 4) は、何らかの処置により (あるサイズの個体, ある花数) 自体も影響を受けられるので、部分母集団と見なすのではなく共変量としてモデルの中に入れることが望ましいと思われる。

下野は、図 9.1 に示すような生育環境別 地域別の種子数の平均の線グラフ、花数別の種子数のボックス・プロットを示している。なお、線グラフは、生育環境と地域について JMP の「モデルのあてはめ」で通常の線形モデルによる 2 元配置分散分析を行い「交互作用プロファイル」による結果であり、ボックス・プロットは、花数について「二変量の関係」で作成した結果である。

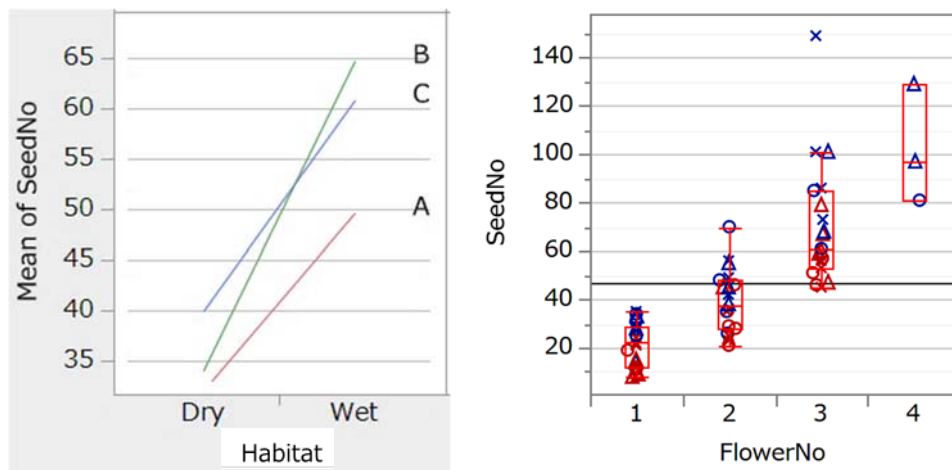


図 9.1 生育環境と種子数の関係（左）および花数と種子数の関係（右）

これらの図から、生育環境が Wet の場合は、Dry に比べて種子数は明らかに多く、地域 B と地域 C は、地域 A に比べて種子数が大きめであることが読み取れる。また、地域 B は、Wet 環境で他の地域に比べ種子数が多く、量的な交互作用が示唆されている。ボックス・プロットからは、花数が増えた場合に種子数の平均の増大に伴い、分散も増大する傾向が観察される。

なお、量的な交互作用とは、傾きが同じ方向を向いており、誤差的な変動が大きめに出ているとも解されるような傾きの違いである。量的な交互作用に対し質的な交互作用は、傾きがプラス方向とマイナス方向のような極端な場合も含め、偶然の変動とは思われないような傾きの差がある場合である。

JMP のグラフ・ビルダーを用い、図 9.1 に示した別々の図を一まとめにしたのが図 9.2 である。Y 軸方向に 2 水準の生育環境、X 軸方向に 3 水準の地域に分割し、分割された区画の中で、さらに X 軸方向に花数を、Y 軸方向に種子数とした散布図を描き、花数ごとにボックス・プロットを上書きし、さらに平滑線をあてはめた結果である。この結果から、花数の増加による種子数の増加は、やや指数的であることが読み取れる。

第 7.2 節で例示したように、S-PLUS にも同様な Trellis（格子）グラフ作成機能があり重宝していた。R 言語でも Trellis（格子）グラフのパッケージが提供されているようだが使用経験はない。詳細は、Murrell 著・久保訳（2009）、「R グラフィックス 第 4 章 Trellis 作図：lattice パッケージ」を参照のこと。

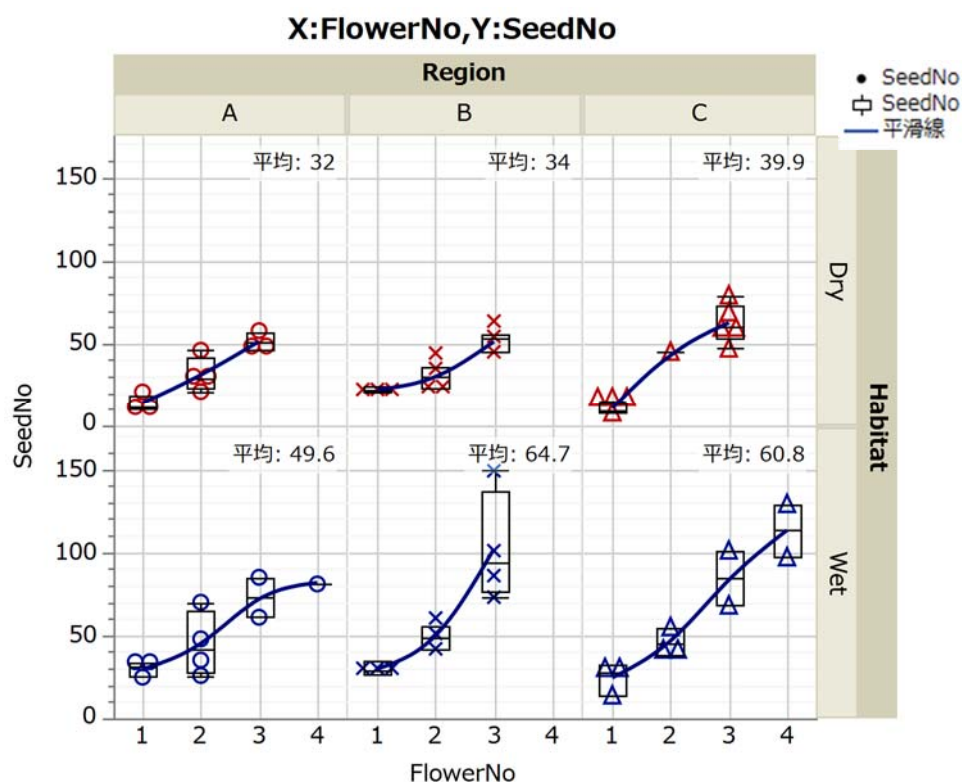


図 9.2 生育環境別 地域別 花数別の種子数に対する平滑線のあてはめ

JMP のグラフ・ビルダーによる作図は、基本の散布図 (X 軸: 花数, Y 軸: 種子数) を作成し、その上に「平滑線」を上書きさせ、「ボックス・プロット」も上書きさせる。さらに、X 軸のグループとして地域を、Y 軸のグループとして育成環境を設定する。なお、これらは、すべて画面上の操作 (GUI) で行うことができる。

図 9.2 に対応して地域別・生育環境別・花数別の分散/平均の比を計算し過分散の程度を表 9.2 に示す。分散/平均の比は、1 以下の場合もあるが、それぞれの地域の分散/平均の比の平均は (3.41, 2.43, 3.26) と過分散となっている。

表 9.2 生育環境別 地域別 花数別の種子数の分散/平均の比

	Flower	Region A				Region B				Region C			
Habitat	No	<i>n</i>	平均	分散	分散/平均	<i>n</i>	平均	分散	分散/平均	<i>n</i>	平均	分散	分散/平均
Dry	1	3	14.00	19.00	.	3	22.33	2.33	0.10	4	10.50	9.67	0.92
	2	4	31.00	112.67	3.63	4	29.75	44.92	1.51	1	45.00	.	.
	3	3	51.33	30.33	0.59	3	51.33	32.33	0.63	5	62.40	137.80	2.21
Wet	1	3	30.00	21.00	0.70	3	30.33	17.33	0.57	3	25.00	97.00	3.88
	2	4	44.75	364.92	8.15	3	49.00	49.00	1.00	3	46.00	73.00	1.59
	3	3	73.00	288.00	3.95	4	102.25	1102.3	10.78	2	84.50	544.50	6.44
	4	1	81.00	2	113.00	512.00	4.53
(分散/平均) 比の平均					3.41								2.43
													3.26

9.2. JMP のポアソン回帰による探索的解析

実験計画法の観点からこのデータは、地域（A, B, C）を標示因子、生育環境（Dry, Wet）を制御的な因子、花数を共変量とする 3×2 の要因配置モデルと見なすことができる。共変量としての花数がなければ、ごく普通の繰り返し 10 の 2 元配置モデルとなる。地域と生育環境に交互作用が統計的に検出されたとしても地域は標示因子（変量効果、ランダムに選ばれた場所）であり、地域によって育成環境による種子数に差があった（交互作用あり）場合であっても、主たる解析では、交互作用を含まない解析を優先すべきである、ただし、副次的な解析としては、交互作用の確認を行うことは必要である。

交互作用モデル

JMP のポアソン回帰で過分散を考慮し、

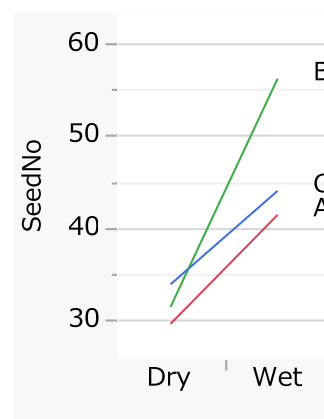
モデル： 種子数 = 生育環境 + 地域 + 生育環境×地域 + 花数

分布：Poisson, リンク関数：対数, 過分散：あり

による解析で交互作用（生育環境×地域）を確認する。表 9.3 に示すように、交互作用の p 値は 0.1565 であり、図 9.1 では、交互作用があるかと示唆したが、統計的には有意な差ではなかった。図 9.2 のボックス・プロットから、外れ値的なデータが示されており、これが平均値が大きくなっている原因である。

表 9.3 交互作用（生育環境×地域）を含めたポアソン回帰

モデル全体の検定					
モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値(Prob>ChiSq)	
差分	120.5426	241.0852	6	<.0001*	
完全	74.3134				
縮小	194.8559				
適合度統計量		カイ2乗	自由度	p値(Prob>ChiSq)	過分散
Pearson		181.6694	53	<.0001*	3.4277
デビアン		179.8630	53	<.0001*	
AICc					
167.4502					
効果の検定					
要因	自由度	尤度比カイ2乗	p値(Prob>ChiSq)		
Habitat	1	29.5774	<.0001*		
Region	2	4.0718	0.1306		
Habitat*Region	2	3.7089	0.1565		
FlowerNo	1	181.8276	<.0001*		



この図は、花数を共変量としてモデルに含めているので、図 9.1 とは形状が異なる

注) 過分散を反映。交互作用プロファイルは、花数を 2 と固定した場合である。

JMP のポアソン回帰では、「予測プロファイル」の出力に引き続き表 9.3 右に示すように「交互作用プロファイル」が出力され、どのような交互作用なのか手軽に可視化することができる。

表 9.4 に交互作用を含むパラメータの推定値および Excel で整形した共分散行列を示す。推定値に有意な差が有るか否かが分かるが、対数リンクでの結果なので、元の種子数に戻ないと理解しづらい。さらに、何らかのグラフ表示をしなければ結果の解釈は困難であり、どのような計算を JMP の内部で行っているのか、それをどのようにグラフ化しているのか Excel を用いた検討を行う。

表 9.4 交互作用含むモデルの推定値およびパラメータの共分散行列

パラメータ推定値							
項	推定値	標準誤差	尤度比カイ2乗	p値			
切片	2.6060	0.1063	423.6291	<.0001*			
Habitat[Dry]	-0.1970	0.0367	29.5774	<.0001*			
Region[A]	-0.0938	0.0528	3.2199	0.0727			
Region[B]	0.0893	0.0513	2.9845	0.0841			
Habitat[Dry]*Region[A]	0.0284	0.0527	0.2895	0.5905			
Habitat[Dry]*Region[B]	-0.0941	0.0510	3.4300	0.0640			
FlowerNo	0.5211	0.0398	181.8276	<.0001*			

パラメータの 共分散行列 $\Sigma(\hat{\beta})$	切片	Habitat [Dry]	Region [A]	Region [B]	Habitat [Dry] *Region[A]	Habitat [Dry] *Region[B]	Flower No
切片	0.0113	-0.0002	-0.0001	-0.0006	0.0000	0.0004	-0.0040
Habitat[Dry]	-0.0002	0.0013	0.0000	0.0001	0.0002	0.0000	0.0002
Region[A]	-0.0001	0.0000	0.0028	-0.0014	0.0006	-0.0004	0.0001
Region[B]	-0.0006	0.0001	-0.0014	0.0026	-0.0004	0.0007	0.0002
Habitat[Dry]*Region[A]	0.0000	0.0002	0.0006	-0.0004	0.0028	-0.0014	0.0000
Habitat[Dry]*Region[B]	0.0004	0.0000	-0.0004	0.0007	-0.0014	0.0026	-0.0001
FlowerNo	-0.0040	0.0002	0.0001	0.0002	0.0000	-0.0001	0.0016
JMPの結果をExcelに取り込み整形した表である							

JMP で質的変数を解析モデルに含めた場合には、対比型のデザイン行列の変数が生成される。表 9.5 に示すように 2 水準の生育環境の場合は、内部のコード順が (Dry, Wet) なので、Dry の場合に 1, Wet の場合に -1 となる Habitat[Dry]変数が生成される。地域のように 3 水準の場合は、内部コード順が (A, B, C) なので、A の場合 (1, 0), B の場合 (0, 1), C の場合 (-1, -1) となるような 2 組の Region[A], Region[B] 変数が生成される。交互作用は、取り上げた 2 つの名義尺度変数から生成される変数の総当たりによる Habitat[Dry]*Region[A], Habitat[Dry]*Region[B] 変数が生成される。なお、統計ソフトによってデフォルトで生成されるデザイン行列の変数は異なるので注意が必要である。

表 9.5 JMP の内部で生成される対比型デザイン変数

Habitat	Habitat [Dry]	Region	Region [A]	Region [B]	Habitat	Region	Habitat [Dry] *Region [A]	Habitat [Dry] *Region [B]
Dry	1	A	1	0	Dry	A	1	0
Wet	-1	B	0	1		B	0	1
		C	-1	-1		C	-1	-1
					Wet	A	-1	0
						B	0	-1
						C	1	1

表 9.6 に内部で生成されたデザイン行列の変数に対応する推定値 $\hat{\beta}$ ，花数は 2 に固定した生育環境と地域の組み合わせに対するデザイン行列の変数 \mathbf{x} を示す．生育環境が Dry，地域が A，花数が 2 の場合は，

$$\mathbf{x}_{\text{Dry},A,2} = [1 \ 1 \ 1 \ 0 \ 1 \ 0 \ 2]$$

であり，推定値 $\hat{\beta}$ を表 9.4 の列ベクトルを転置して行ベクトル

$$\hat{\beta}^T = [2.6060 \ -0.1970 \ -0.0938 \ 0.0893 \ 0.0284 \ -0.0941 \ 0.5211]$$

として示す．推定値 $\hat{y}_{\text{Dry},A,2}$ は，Excel の行列関数を用いて

$$\begin{aligned}\hat{y}_{\text{Dry},A,2} &= \exp(\mathbf{x}_{\text{Dry},A,2} \hat{\beta}) \\ &= \exp(\text{Mmult}(\mathbf{x}_{\text{Dry},A,2} \text{の範囲}, \text{Transpose}(\hat{\beta}^T \text{の範囲}))) \\ &= \exp(3.3858) = 29.5398\end{aligned}$$

表 9.6 交互作用を含むモデルの推定値

生育	環境	切片	Habitat	Region		交互作用		花数	対数	指数	分散
	地域	2.6060	Dry	A	B	Dry*A	Dry*B	0.5211	推定値	推定値	Var
Dry	A	1	1	1	0	1	0	2	3.3858	29.54	0.0109
Wet	"	1	-1	1	0	-1	0	2	3.7230	41.39	0.0075
Dry	B	1	1	0	1	0	1	2	3.4464	31.39	0.0102
Wet	"	1	-1	0	1	0	-1	2	4.0286	56.18	0.0056
Dry	C	1	1	-1	-1	-1	-1	2	3.5214	33.83	0.0090
Wet	"	1	-1	-1	-1	1	1	2	3.7840	43.99	0.0070

と計算している．図 9.3 に示すように JMP の予測プロファイルを用いて種子数 $\hat{y}_{\text{Dry},A,2} = 29.54$ が推定されている．推定値の 95%信頼区間 (24.08, 36.23) は，表 9.4 に示されているパラメータの共分散行列を $\Sigma(\hat{\beta})$ としたときに $\ln \hat{y}_{\text{Dry},A,2}$ の分散を，

$$\begin{aligned}\text{Var}(\ln \hat{y}_{\text{Dry},A,2}) &= \mathbf{x}_{\text{Dry},A,2} \Sigma(\hat{\beta}) \mathbf{x}_{\text{Dry},A,2}^T \\ &= \text{Mmult}(\text{Mmult}(\mathbf{x}_{\text{Dry},A,2} \text{の範囲}, \Sigma(\hat{\beta}) \text{の範囲}), \text{Transpose}(\mathbf{x}_{\text{Dry},A,2} \text{の範囲})) \\ &= 0.0109\end{aligned}$$

として，95%信頼区間は，

$$\begin{aligned}
 95\%CL &= \exp(\ln \hat{y}_{\text{Dry},A,2} \pm 1.96\sqrt{\text{Var}(\ln \hat{y}_{\text{Dry},A,2})}) \\
 &= \exp(3.8358 \pm 1.96\sqrt{0.0109}) \\
 &= (24.0829, 36.2312)
 \end{aligned}$$

と推定されている。

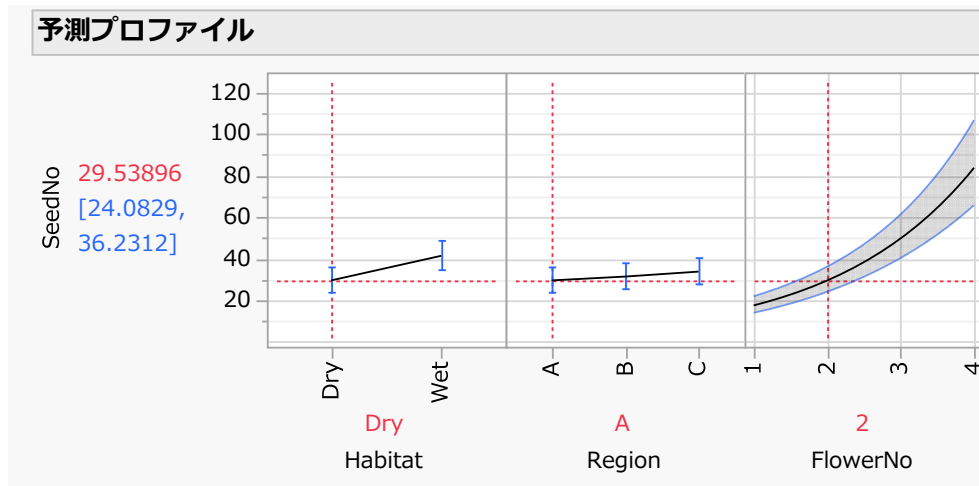


図 9.3 生育環境 Dry 地域 A 花数 2 に固定した場合の予測プロファイル

表 9.6 には、 $\hat{y}_{\text{Wet},A,2}$ 、 $\hat{y}_{\text{Dry},B,2}$ 、 $\hat{y}_{\text{Wet},B,2}$ 、 $\hat{y}_{\text{Dry},C,2}$ 、 $\hat{y}_{\text{Wet},C,2}$ についても推定されているので、これらの推定値を用いて Excel の「折れ線グラフ」により表 9.3 に示した交互作用プロファイルを図 9.4 (中) に示すように再現することができる。さらに、花数を変化させることにより図 9.4 (左) および (右) の交互作用プロファイルも容易に描くことができる。

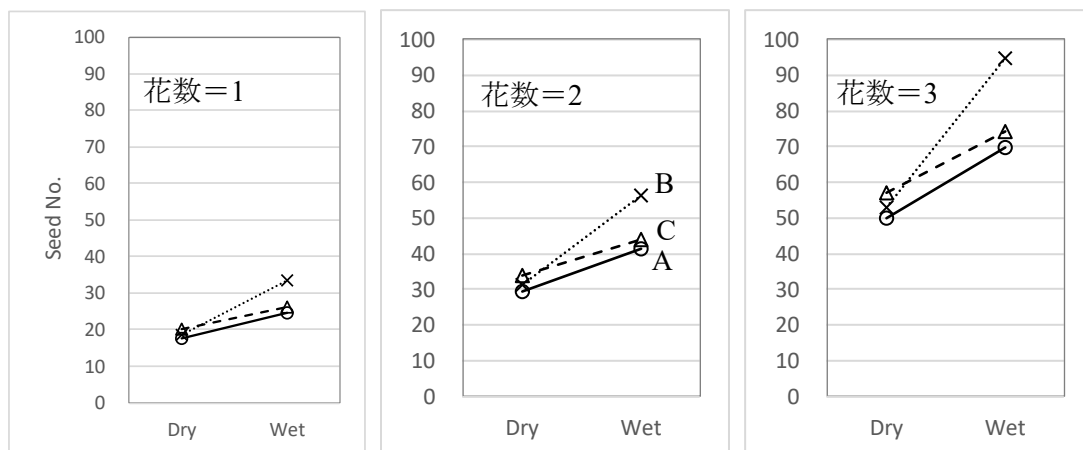


図 9.4 Excel の折れ線グラフによる生育環境による地域別の種子数の交互作用プロファイル

図 9.2 のグラフ・ビルダーで作成したグラフは、花数を X 軸にした種子数のデータのプロットであるが、因子間の関連については見えにくい。図 9.4 では、交互作用モデルに即した花数ごとの生育環境と地域における種子数の関係が明瞭に把握できる。

主効果モデル

交互作用が統計的には検出されなかったため、解析モデルから交互作用項を除いた主効果モデル

モデル： 種子数 = 生育環境 + 地域 + 花数

分布：Poisson, リンク関数：対数, 過分散：あり

による解析した結果を表 9.7 に示す.

表 9.7 過分散を反映した主効果モデルによる検討

モデル全体の検定				
モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値)
差分	113.2155	226.4310	4	<.0001*
完全	72.6557			
縮小	185.8712			
適合度統計量	カイ2乗	自由度	p値	過分散
Pearson	197.6378	55	<.0001*	3.5934
デビアン	192.5761	55	<.0001*	
AICc				
158.8964				
効果の検定				
要因	自由度	尤度比カイ2乗	p値	
Habitat	1	28.9356	<.0001*	
Region	2	5.7869	0.0554	
FlowerNo	1	171.8171	<.0001*	
パラメータ推定値				
項	推定値	標準誤差	尤度比カイ2乗	p値
切片	2.6197	0.1080	418.5343	<.0001*
Habitat[Dry]	-0.1984	0.0373	28.9356	<.0001*
Region[A]	-0.1016	0.0526	3.8146	0.0508
Region[B]	0.1126	0.0505	4.8943	0.0269*
FlowerNo	0.5160	0.0404	171.8171	<.0001*

地域 Region の p 値は、0.0554 と微妙な大きさであるのに対し、育成環境 Habitat には、明らかな統計的な差が検出されている。これらの要因を組み合わせた予測プロファイルおよび 2 つの要因を組み合わせた“交互作用”プロファイルを用いて結果の吟味を行う。

図 9.5 に示すように花数を 2 に固定した場合に、種子数が最も多くなるのは、育成環境が Dry で、地域が B の場合で、 $\hat{y}_{\text{Wet},B,2} = 52.5985$ であり、その 95%信頼区間は (45.8902, 60.2873) となる。この予測プロファイルの中で、ポインターで育成環境が Dry の位置で選択すると他の予測プロファイルが、下方にシフトし、推定値も $\hat{y}_{\text{Dry},B,2} = 35.3673$ と更新される。

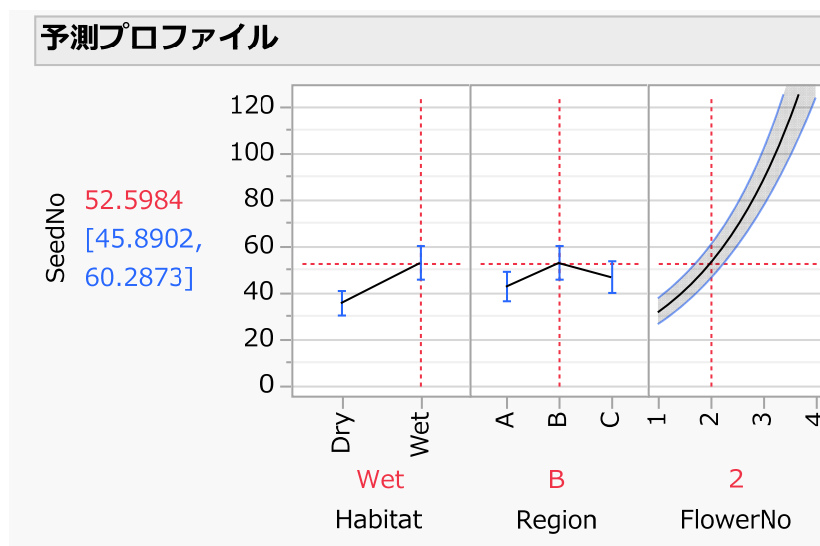


図 9.5 JMP による生育環境 Wet 地域 B 花数 2 の場合の種子数の推定

JMP の予測プロファイルの動的な機能は、探索的解析のための優れものであるので、Excel によって再現してみよう。表 9.8 に示すように、主効果モデルのパラメータの推定値、

$$\hat{\beta}^T = [2.6197 \quad -0.1984 \quad -0.1016 \quad 0.1126 \quad 0.5160]$$

および、パラメータの共分散行列 $\Sigma(\hat{\beta})$ を Excel に取り込んで、交互作用が有る場合と同様に対数推定値、対数分散、指数推定値、L95%、U95%を計算した結果を示す。

表 9.8 主効果モデルによる予測プロファイルの計算シート

				Habitat		Region		花数			指数			
				切片	Dry	A	B	2	対数	対数	指数	95%信頼区間		
作図用	環境	地域		2.6197	-0.1984	-0.1016	0.1126	0.5160	推定値	分散	推定値	L95%	U95%	
1	1	Dry	B	1	1	0	1	2	3.5658	0.0058	35.37	30.48	41.04	
2	2	Wet	"	1	-1	0	1	2	3.9627	0.0048	52.60	45.89	60.29	
1	1	Wet	A	1	-1	1	0	2	3.7484	0.0058	42.45	36.59	49.26	
2	2	"	B	1	-1	0	1	2	3.9627	0.0048	52.60	45.89	60.29	
3	3	"	C	1	-1	-1	-1	2	3.8391	0.0056	46.48	40.16	53.79	
		Wet	B	1	-1	0	1	1	3.4467	0.0081	31.40	26.33	37.44	
		"	"	1	-1	0	1	2	3.9627	0.0048	52.60	45.89	60.29	
		"	"	1	-1	0	1	3	4.4786	0.0049	88.12	76.83	101.06	
		"	"	1	-1	0	1	4	4.9946	0.0082	147.62	123.62	176.27	
				共分散行列 $\Sigma(\hat{\beta})$				切片	Habitat Dry	RegionA	RegionB	Flower No		
				JMPの結果を Excelに取り込み 整形した表である				切片	0.0117	-0.0002	-0.0001	-0.0008	-0.0041	
								HabitatDry	-0.0002	0.0014	0.0000	0.0000	0.0002	
								RegionA	-0.0001	0.0000	0.0028	-0.0014	0.0001	
								RegionB	-0.0008	0.0000	-0.0014	0.0025	0.0003	
								FlowerNo	-0.0041	0.0002	0.0001	0.0003	0.0016	

この予測プロファイルの計算シートは、Excel の行列関数を 1 行目で

$$\ln \hat{y} = \text{Mmult}(\text{デザイン変数 } x \text{ の範囲}, \text{Transpose}(\text{固定した } \hat{\beta}^T \text{ の範囲}))$$

$$= \text{Mmult} (\text{F8:J8}, \text{Transpose} (\$F\$7:\$J\$7)) = 3.5689$$

のように変数の範囲を選択し、2 行目以後にフィルハンドルで数式をコピーしている。回帰パラメータは、範囲を選択した後に「F4」キーでセルアドレスに \$ を付けて位置を固定している。このような計算に不慣れな場合には、第 4.5 節の「デザイン行列を用いた回帰分析の実際」を参照してもらいたい。対数分散は、1 行目で次のように計算式をセットし

$$\text{Var}(\ln \hat{y}) = \text{Mmult}(\text{Mmult}(x \text{ の範囲}, \text{固定した } \Sigma(\hat{\beta}) \text{ の範囲}), \text{Transpose}(x \text{ の範囲}))$$

$$= \text{Mmult} (\text{Mmult} (\text{F8:J8}, \$K\$18:\$O\$22), \text{Transpose} (\text{F8:J8})) = 0.0058$$

2 行目以後をフィルハンドルで数式をコピーしている。選択の範囲が異なるだけで通常の回帰分析と同じ計算式である。

図 9.5 に示した JMP の予測プロファイルと同様に、育成環境 Habitat を変化させ、推定値と 95%信頼区間を $\hat{y}_{\text{Dry}, B, 2} = 35.37$ (30.48, 41.04), $\hat{y}_{\text{Wet}, B, 2} = 52.60$ (45.89, 60.29) を計算する。次に、地域 Region を変化させ、さらに花数を変化させた推定値と 95%信頼区間を計算する。図 9.6 に示した予測プロファイルの作図は、「散布図」を使い、推定値を最初に描き、順次 95%信頼区間のための数値を「データの選択」機能を使い「追加」し、「データ系列の書式設定」で形を整える。

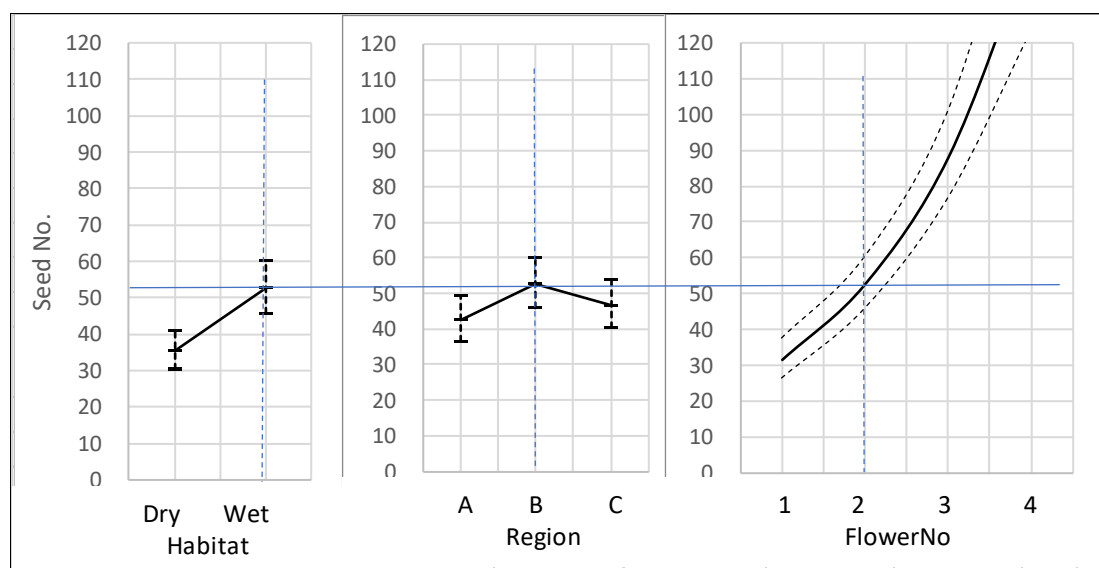


図 9.6 Excel による予測プロファイル（生育環境 Wet 地域 B 花数 2）に固定

2 つの要因を組み合わせた JMP の交互作用プロファイルを図 9.7 に示す。交互作用が、解析モデルに含まれていなくても 2 つの要因の組み合わせた場合の予測値を推定している。推定結果は予測プロファイルで設定した要因の水準の値を反映している。

環境要因と地域の組み合わせは、花数が 2 の場合の予測値となっている。花数を変化させると、その花数を反映した推定値が示される。環境要因と花数の組み合わせは、地域 B に固定した推定値であるので、他の地域を選択すると上下に変化する。地域と花数の組み合わせは、育成環境 Wet に固定した場合の推定値で Dry に変化させると下方に移動する。

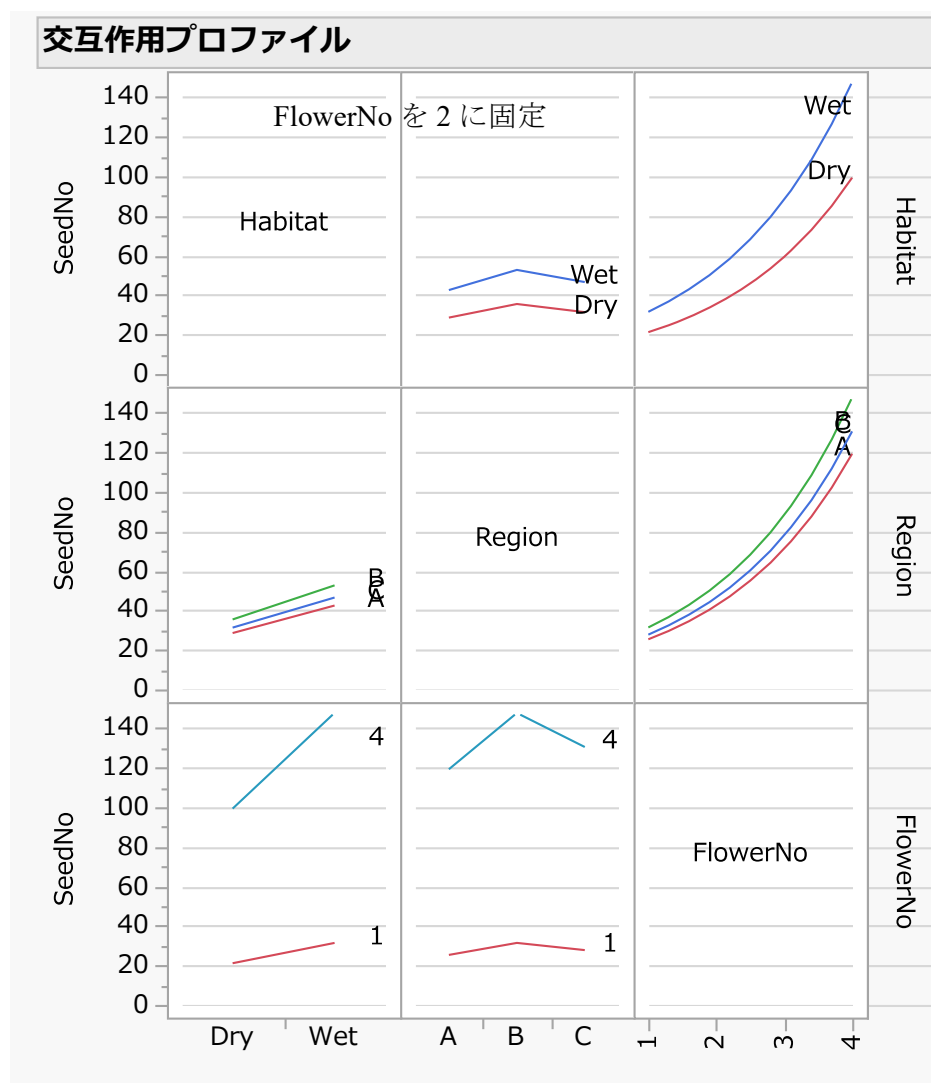


図 9.7 主効果モデルにおける交互作用プロフィール
連続変数の花数と質的変数間の交互作用の表示は、(最小値 1 と最大値 4) となっている。変更はできるが、表示できるのは 2 本に限定されている。

9.3. 無償版 SAS の GENMOD プロシジャによる主効果モデル

無償版 OnDemand SAS の GENMOD プロシジャを用い、JMP でのポアソン回帰の結果を検証しつつ、使用法の解説を行う。SAS は、DATA ステップで SAS データセットを作成し、PROC ステップで各種の GENMOD などのプロシジャを起動する。表 9.1 のデータリストをできるだけコンパクトにするために、(育成環境別 地域別) を先頭に、

(花数₁ 種子数₁), (花数₂ 種子数₂), ..., (花数₁₀ 種子数₁₀)

のように横並びにすると `datalines` 以下に示した 6 行で全データを表すことができる。

```
Title "Poisson_S9_1_F_Seed.sas    2019/12/05 Y.Takahashi" ;
data d01 ;
  input Habitat$ Region$ @@ ;
  do No=1 to 10 ;
    input  FlowerNo SeedNo @@ ;
    ln_FloweNo = log(FlowerNo); output ;
  end;
datalines ;
Dry A   3 57   1 19   1 11   1 12   3 46   3 51   2 21   2 29   2 28   2 46
Dry B   1 22   1 21   2 36   1 24   3 45   2 35   3 53   2 25   2 23   3 56
Dry C   3 67   3 59   1 10   2 45   1  8   3 60   3 47   1  9   3 79   1 15
Wet A   2 26   3 61   2 35   4 81   3 85   1 34   1 25   1 31   2 48   2 70
Wet B   2 42   1 27   3 101  3 149  2 56   1 29   1 35   2 49   3 86   3 73
Wet C   4 97   2 45   2 38   3 68   1 28   1 33   2 55   1 14   4 129  3 101
;
proc print data=d01 ; run;
```

- 1) 横並びのデータリストを縦方向に並べ直すために、まず、`input Habitat$ Region$ @@ ;`により、文字型の変数として `Habitat` と `Region` を読み込み、`@@ ;` により改行せずにその位置で読み込みポインターを待機させる。`Do No=1 to 10 ;` により、`end;` まで 10 回繰返しが行なわれる。なお、`Habitat$` の `$` は、文字型のデータとの定義である。
- 2) `Do No=1 to 10 ;` の次の `Input FlowerNo SeedNo @@ ;` により データリストから `FlowerNo` `SeedNo` を読み込み `@@ ;` により「読み込みポインター」改行せずにその位置で待機させる。
- 3) 読み込んだ `FlowerNo` を用いて オフセット `ln_FloweNo = log(FlowerNo);` を計算する。
- 4) `Output ;` により、既に読み込んだ `Habitat`, `Region` に加え、制御変数としての `No` , 読み込んだ `FlowerNo`, `SeedNo` 計算された `ln_FloweNo` を SAS データセット `d01` に出力する。
- 5) `Do No=1 to 10 ;` で `No=2` として繰り返し、`No=10` を実行すると最初のデータリストが尽きる。

- 6) 次の行に読み込みポインターが移動し、次のデータリストに対し、1) から 5) を繰り返す。
- 7) これをデータリストが尽きるまで繰り返す。
- 8) Proc print data=d01 ; run; により SAS データセット d01 の 60 行分が表 9.9 に示すよう
に出力される。

表 9.9 作成された SAS データセットを proc print で出力した結果

Obs	Habitat	Region	No	FlowerNo	SeedNo	In_FloweNo
1	Dry	A	1	3	57	1.09861
2	Dry	A	2	1	19	0.00000
3	Dry	A	3	1	11	0.00000
4	Dry	A	4	1	12	0.00000
:						
60	Wet	C	10	3	101	1.09861

SAS の GENMOD プロシジャは、ありとあらゆる一般化線形モデルによる解析が行なえるようになっている。JMP の一般化線形モデルは、GENMOD プロシジャに限らず 10 分の 1 程度の解析機能に絞られているが、GUI 操作で扱えるグラフ表示に優れている。

```
Title2 '<<< poisson 過分散 >>>' ;
proc genmod data=d01 ;
  class Habitat Region / param=effect ;
  model SeedNo = Habitat Region FlowerNo
    / dist=poisson link=log scale=Pearson type3;
run ;
```

- 9) Proc genmod data=d01 ; は、SAS データセット d01 に対しての解析の指示である。
- 10) Class Habitat Region は、名義尺度変数として扱うことを宣言している。
- 11) Param=effect ; は、対比型の変数の内部生成を指示しているオプションである。
- 12) Param オプションにより名義尺度の最初または最後の水準を基準として、他の水準との「差」が得られるような変数を生成することも可能である。
- 13) Model SeedNo = Habitat Region FlowerNo; により回帰式を設定し、後のオプションで各種の解析モデルを規定する。
- 14) Dist=poisson によりポアソン回帰を設定している。JMP ではサポートされていない（負の 2 項分布，ゼロ過剰ポアソン，ゼロ過剰負の 2 項分布）なども扱える。
- 15) Link=log は対数リンクの設定である。

16) Scale=Pearson は、過分散の調整にピアソン残差を使う指示で、SAS では、デビエンス残差を指定することもできる。

17) Type3 オプションは、JMP と同様の「効果の検定」を出力するためのオプションである。

表 9.10 に示すのは、JMP と同様の「効果の検定」であり、表 9.7 に示した JMP の結果に相当する。JMP の出力に加えて、誤差の自由度を考慮した F 検定が追加されている。

表 9.10 SAS による JMP と同様の「効果の検定」

Type 3 分析の LR 統計量						
要因	分子の自由度	分母の自由度	F 値	Pr > F	カイ 2 乗	Pr > ChiSq
Habitat	1	55	28.94	<.0001	28.94	<.0001
Region	2	55	2.89	0.0639	5.79	0.0554
FlowerNo	1	55	171.82	<.0001	171.82	<.0001

表 9.11 に示す「パラメータの推定値」は、表 9.7 に示した JMP の結果に相当する。大きく異なるのは、SAS では推定値の標準誤差を用いた Wald 検定となっていて、JMP の尤度比検定の結果とは微妙に異なることに注意が必要である。過分散の調整のための推定値は、「尺度」の行の推定値の欄の 1.8956 である。JMP では、表 9.7 の「適合度の統計量」の過分散 3.5934 の平方根 1.8956 に一致する。

表 9.11 SAS によるポアソン回帰の対比型デザイン変数による推定値

最大尤度パラメータ推定値の分析								
パラメータ		自由度	推定値	標準誤差	Wald 95% 信頼限界		Wald カイ 2 乗	Pr > ChiSq
Intercept		1	2.6197	0.1080	2.4080	2.8314	588.43	<.0001
Habitat	Dry	1	-0.1984	0.0373	-0.2716	-0.1253	28.24	<.0001
Region	A	1	-0.1016	0.0526	-0.2047	0.0014	3.74	0.0532
Region	B	1	0.1126	0.0505	0.0137	0.2116	4.98	0.0257
FlowerNo		1	0.5160	0.0404	0.4368	0.5951	163.25	<.0001
尺度		0	1.8956	0.0000	1.8956	1.8956		

Note: 尺度パラメータは Pearson カイ 2 乗/DOF の平方根により推定されています。

SAS の GENMOD プロシジャでは、これまで示してきた予測プロファイル、交互作用プロファイルの作図が行なえないのは残念なことであるが、表 9.12 に示すように covb オプションで共分散行列を得ることができるので、Excel に取り込んで予測プロファイルおよび交互作用プロファイルを作成することができる。

表 9.12 GENMOD プロシジャによるパラメータの共分散行列 (covb オプション)

推定値の共分散行列					
	Prm1	Prm2	Prm3	Prm4	Prm5
Prm1	0.01166	-0.000243	-0.000069	-0.000764	-0.004096
Prm2	-0.000243	0.001394	0.0000101	9.4424E-6	0.0002299
Prm3	-0.000069	0.0000101	0.002764	-0.001375	0.0000977
Prm4	-0.000764	9.4424E-6	-0.001375	0.002549	0.0002708
Prm5	-0.004096	0.0002299	0.0000977	0.0002708	0.001631

このパラメータの共分散行列は、JMP の出力を Excel に取り込んだ表 9.8 の結果と一致することの確認のためである。SAS/GENMOD プロシジャの Lsmeans ステートメントなどの各種の推定機能を発揮させるためには、デザイン変数の設定を対比型からデフォルトの最後の水準を基準とする (1, 0) 型でないと利用できない。また、表 9.8 および図 9.6 で示した予測プロファイルは、共変量としての FlowerNo も含めている。SAS/GENMOD プロシジャの推定は、質的因子の各水準を対象としているために、共変量についての推定は限定されている。

ただし、表 9.13 に示すように推定したい (生育環境, 地域, 花数) のカテゴリを 61 行目以後に追加し、output ステートメントで、表 9.14 に示すように推定値および 95%信頼区間を出力することができる。この出力を用いて予測プロファイルを作成することはできるが、パラメータの共分散行列を使って Excel で計算の方が生産的である。表 9.8 の下段に示した Excel での計算を検証するためには有益である。

表 9.13 推定したい予測プロファイルデータの追加

Obs	Habitat	Region	No	FlowerNo	SeedNo	In_FloweNo
:						
60	Wet	C	10	3	101	1.09861
61	Wet	B	1	1	.	
62	Wet	B	1	2	.	
63	Wet	B	1	3	.	
64	Wet	B	1	4	.	

表 9.14 追加データに対する予測値と 95%信頼区間の計算結果

Obs	Habitat	Region	No	FlowerNo	SeedNo	In_FloweNo	y_est	ln_y_est	L95	U95
61	Wet	B	1	1	.	0.00000	31.397	3.44673	26.330	37.440
62	Wet	B	1	2	.	0.69315	52.598	3.96269	45.890	60.287
63	Wet	B	1	3	.	1.09861	88.115	4.47865	76.831	101.057
64	Wet	B	1	4	.	1.38629	147.615	4.99461	123.619	176.269

9.4. 花数をオフセットとしたポアソン回帰

ポアソン回帰でのオフセットは、第 1.5 節の冠動脈心疾患の死亡者数で示したように、年齢階層別の部分母集団の大きさが人口統計学的に既知である場合に、稀に発生する死亡者数を統計モデルによって調整するためである。共変量は、反応変数に明らかに影響を及ぼすことが分かっているが、事前にコントロールできない変数であり、種子数に対して花数は、共変量との位置づけられる。花数をオフセットとしても扱えるが、単位花数に換算した場合には、2×3 の要因配置型のポアソン回帰となり、花数と種子数の関連が見えにくくなる。

下野（2010）と同様に、過分散を無視し、花数の対数をオフセットとした JMP によるポアソン回帰を行い結果の吟味を行う。表 9.15 に示すように地域は（A, B, C）の 3 水準なので、*RegionB* 変数は「B」の場合に 1、それ以外は 0、*RegionC* 変数は「C」の場合に 1、それ以外は 0、生育環境は（Dry, Wet）の 2 水準なので、*HabitatWet* 変数は「Wet」の場合に 1、それ以外は 0 とする。

表 9.15 最初の水準を基準とするデザイン変数

Region	<i>RegionB</i>	<i>RegionC</i>	Habitat	<i>HabitatWet</i>
A	0	0	Dry	0
B	1	0	Wet	1
C	0	1		

モデル式は

$$y_i = Flower_i \cdot \exp[\beta_0 + \beta_1 \cdot (RegionB_i) + \beta_2 \cdot (RegionC_i) + \beta_3 \cdot (HabitatWet_i)] + \varepsilon_i$$

ただし、 $\varepsilon_i \sim \text{Poisson}$ 分布

であり、両辺に対数を取り推定値の形式として

$$\ln \hat{y}_i = \ln(Flower_i) + \hat{\beta}_0 + \hat{\beta}_1 \cdot (RegionB_i) + \hat{\beta}_2 \cdot (RegionC_i) + \hat{\beta}_3 \cdot (HabitatWet_i)$$

であり、表 9.16 に過分散を無視した結果を示す。推定値に対する標準誤差は、過分散が全く考慮されておらず、*RegionC* の推定値は、*RegionA*（切片）との対数での差で $\hat{\beta}_2 = 0.1352$ に対

表 9.16 過分散を無視したポアソン回帰の結果

パラメータ推定値						
項	推定値	標準誤差	尤度比カイ2乗	p値	下側信頼限界	上側信頼限界
切片	2.7422	0.0425	2463.6554	<.0001*	2.6581	2.8247
RegionB	0.1903	0.0473	16.2424	<.0001*	0.0976	0.2831
RegionC	0.1352	0.0471	8.2693	0.0040*	0.0430	0.2277
HabitatWet	0.4390	0.0389	130.8425	<.0001*	0.3629	0.5156

して $p=0.0040$ と有意な差となっている．ところが，図 9.2 で Dry と Wet それぞれの RegionA と RegionC を比べても，平均的な差はデータの変動に比べても有意な差とは，とても見えない．これは，ポアソン回帰の場合に分散を推定値と同じと見なしているために，過分散を全く考慮しないことにある．

表 9.17 に示すように，適合度統計量の Pearson カイ 2 乗値は 183.5650 となり，自由度 56 のカイ 2 乗分布から $p<0.0001$ となり，過分散であることが分かる．Pearson カイ 2 乗を自由度で割った過分散の係数 $\phi=3.2779$ を考慮した結果が示されている．結果として RegionC の RegionA に対する差の標準誤差は，

$$SE = \sqrt{3.2779 \times 0.0471} = 0.0853$$

と 2 倍弱大きくなり $p=0.1122$ と有意な差ではなくなっている．

表 9.17 過分散を考慮したポアソン回帰の結果

モデル全体の検定				
モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	22.5930	45.1861	3	<.0001*
完全	77.8913			
縮小	100.4843			
適合度統計量	カイ2乗	自由度	p値	過分散
Pearson	183.5650	56	<.0001*	3.2779
デビアン	181.0588	56	<.0001*	
AICc				
166.8937				

パラメータ推定値						
項	推定値	標準誤差	尤度比カイ2乗	p値	下側信頼限界	上側信頼限界
切片	2.7422	0.0770	751.5849	<.0001*	2.5887	2.8905
RegionB	0.1903	0.0857	4.9550	0.0260*	0.0227	0.3587
RegionC	0.1352	0.0853	2.5227	0.1122	-0.0316	0.3029
HabitatWet	0.4390	0.0705	39.9160	<.0001*	0.3016	0.5780

推定値は，対数変換された結果であり，このままでは見通しが悪いので，Excel の行列関数を用いて生育環境別 地域別 花数別 の対数での推定値を計算し，指数を取って元のスケールに戻す必要がある．表 9.18 に示すように，対数推定値の Excel での計算は， $\ln \hat{y}_{\text{dry},A,l}$ の場合， $\text{Region}B_i = 0$ ， $\text{Region}C_i = 0$ ， $\text{HabitatWet}_i = 0$ なので，

$$\begin{aligned}\ln \hat{y}_{\text{dry},A,l} &= \ln(1) + 2.7422 + 0.1903 \times 0 + 0.1352 \times 0 + 0.4390 \times 0 \\ &= 2.7422\end{aligned}$$

であり，指数を取って

$$\hat{y}_{dry,A,1} = \exp(2.7422) = 15.52$$

が計算されている．表の中の花数は，「1」としているが，これを変えれば他の花数についての計算が自動的に得られるようにしてある．

表 9.18 各種の推定値の計算

生育 環境	地域	offset 花数	切片	Region		Habitat	対数 推定値	対数 分散	指数 推定値		
		1	2.7422	B	C	Wet				L95%	U95%
Dry	A	1	1	0	0	0	2.7422	0.0059	15.52	13.35	18.05
	B	1	1	1	0	0	2.9325	0.0052	18.77	16.29	21.63
	C	1	1	0	1	0	2.8774	0.0052	17.77	15.42	20.47
Wet	A	1	1	0	0	1	3.1812	0.0047	24.08	21.04	27.55
	B	1	1	1	0	1	3.3715	0.0040	29.12	25.71	32.99
	C	1	1	0	1	1	3.3164	0.0039	27.56	24.37	31.17

表 9.19 パラメータの共分散行列（JMP の出力を整形）

共分散	切片	RegionB	RegionC	HabitatWet
切片	0.005924	-0.004017	-0.003987	-0.003078
RegionB	-0.004017	0.007338	0.004017	0.000000
RegionC	-0.003987	0.004017	0.007273	-0.000049
HabitatWet	-0.003078	0.000000	-0.000049	0.004968

育成環境が Dry で地域が A の対数推定値は，2.7422 であり，その分散は，デザイン変数のベクトルを $\mathbf{x}_{Dry,A,1} = [1 \ 0 \ 0 \ 0]$ とし，パラメータの共分散行列を $\Sigma(\hat{\beta})$ としたときに，次の 2 次形式で

$$\begin{aligned} Var(\ln \hat{y}_{dry,A,1}) &= \mathbf{x}_{Dry,A,1} \Sigma(\hat{\beta}) \mathbf{x}_{Dry,A,1}^T \\ &= \text{Mmult}(\text{Mmult}(\mathbf{x}_{Dry,A,1} \text{ の範囲}, \Sigma(\hat{\beta}) \text{ の範囲}), \text{Transpose}(\mathbf{x}_{Dry,A,1} \text{ の範囲})) \end{aligned}$$

=	1	0	0	0	0.005924	-0.004017	-0.003987	-0.003078	1	=	0.0059
					-0.004017	0.007338	0.004017	0.000000	0		
					-0.003987	0.004017	0.007273	-0.000049	0		
					-0.003078	0.000000	-0.000049	0.004968	0		

と計算されている．指数推定値

$$\begin{aligned} \hat{y}_{dry,A,1} &= \exp(\ln \hat{y}_{dry,A,1}) \\ &= \exp(2.7422) = 15.52 \end{aligned}$$

に対して、95%信頼区間は、

$$95\%CI = \exp \left[\ln \hat{y}_{\text{dry},A,1} \pm 1.96 \sqrt{\text{Var}(\ln \hat{y}_{\text{dry},A,1})} \right]$$

$$= \exp \left[2.7422 \pm 1.96 \sqrt{0.0059} \right] = (13.35, 18.05)$$

として計算されている。

花数=1 の場合について、Excel の「折れ線グラフ」を用いて地域 A に対する生育環境の折れ線を描き、「データの選択」により地域 B および地域 C を上書きした結果を図 9.8（左）に示す。表 9.18 の花数の欄を「花数=2」および「花数=3」に変更して得られた折れ線グラフを図 9.8（中）、図 9.8（右）に示す。このように何らかの統計モデルで推定された結果についてグラフ化することは、結果の解釈するために役に立つ。

表 9.20 花数を変え場合の推定値

生育 環境	地域	offset	指数	offset	指数	offset	指数
		花数	推定値	花数	推定値	花数	推定値
Dry	A	1	15.52	2	31.04	3	46.56
	B	1	18.77	2	37.55	3	56.32
	C	1	17.77	2	35.54	3	53.30
Wet	A	1	24.08	2	48.15	3	72.23
	B	1	29.12	2	58.24	3	87.37
	C	1	27.56	2	55.12	3	82.68

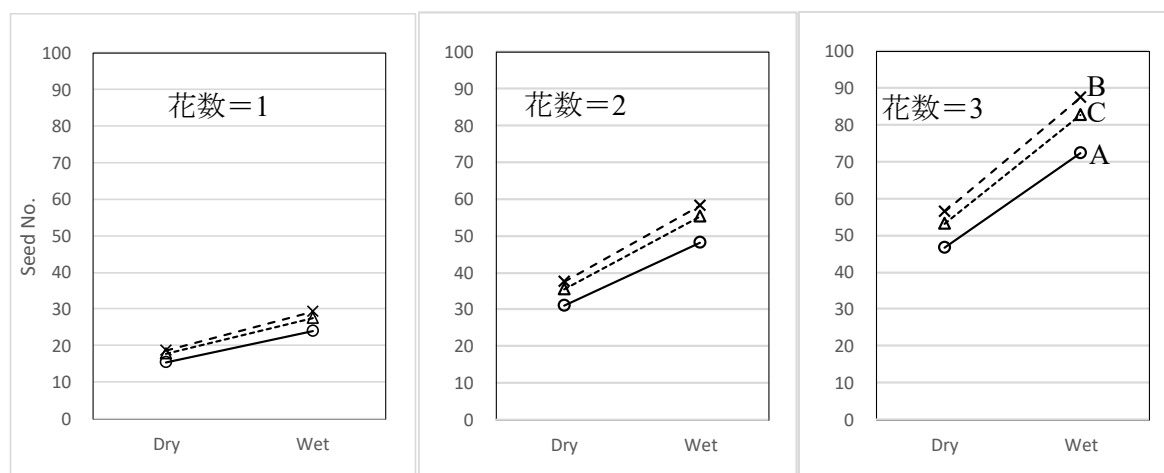


図 9.8 花数オフセットを考慮した組み合わせプロファイル

9.5. 花数をオフセットとした負の2項回帰への適用

下野（2010）は、負の2項分布をあてはめた一般化線形モデルの解析結果も示しているで、Excelでの負の2項回帰およびSASのGENMODプロシジャを使い、結果の吟味を行う。

負の2項分布のパラメータ変換

負の2項分布は、第6.1節で示したように、成功の確率を π としたときに一定の成功数 k が得られるまでの試行を行った場合の失敗の数 y の分布として定義されている。ポアソン分布のパラメータは、平均 μ であり過分散を考慮した場合に、負の2項回帰が適していると言っても理解に苦しむのではないだろうか。

第6.1節では、負の2項分布の整数で定義されている成功数 k を実数化し、組合せ数の計算で用いられている階乗をガンマ関数に置き換え、成功の確率を平均 μ と k で

$$\pi = \frac{k}{\mu + k}, \quad k = \frac{1}{\sigma}$$

置き換え、ポアソン分布のパラメータの平均 μ と整合性が取れるようにする。さらに、 k を $k=1/\sigma$ で置き換えることにより過分散の尺度パラメータ σ が定義される。これにより、ポアソン回帰に後付けで過分散の係数でパラメータの標準誤差を割り増しする便宜的な方法に代え、回帰パラメータとして推定が可能となる。

位置パラメータを μ 、尺度パラメータ σ を用いて負の2項分布（ガンマ・ポアソン分布）を表すことができることを式（6.6）で示した。ガンマ・ポアソン分布の計算で、 σ が小さくなると $\Gamma(1/\sigma)$ が増大するために計算不能となるので、Excelの対数ガンマ関数`Gammaln()`関数を使って、計算する必要がある。

$$\begin{aligned} \text{GammaPoisson}(y; \mu, \sigma) &= \frac{\Gamma(y+1/\sigma)}{\Gamma(y+1)\Gamma(1/\sigma)} \left(\frac{1/\sigma}{\mu+1/\sigma} \right)^{1/\sigma} \left(1 - \frac{1/\sigma}{\mu+1/\sigma} \right)^y \\ &= \frac{\Gamma(y+1/\sigma)}{\Gamma(y+1)\Gamma(1/\sigma)} \cdot \frac{(\mu\sigma)^y}{(1+\mu\sigma)^{y+1/\sigma}} \\ &= \exp \left[\ln \Gamma(y+1/\sigma) - \ln \Gamma(y+1) - \ln \Gamma(1/\sigma) + y \ln(\mu\sigma) - (y+1/\sigma) \ln(1+\mu\sigma) \right] \end{aligned}$$

従って、一般化線形モデルで分布を負の2項分布に設定した場合には、過分散を考慮したガンマ・ポアソン回帰を適用したことになる。

負の2項回帰で、分布を負の2項分布に設定したといっても、元々のパラメータとしての成功確率 π 、成功数 k としたときの失敗数 y の分布との理解することに違和感を持つ。そも

とも成功数 k は、整数として定義されていて Excel の負の 2 項分布 `NegBinom.dist()` 関数で $k=1/\sigma$ などの実数での計算は不能である。位置パラメータを μ 、尺度パラメータを σ としたときの負の 2 項分布に代わる分布としてガンマ・ポアソン分布が知られている。したがって、負の 2 項回帰よりも、ガンマ・ポアソン回帰とすべきと思うが、これらが混在して使われている。

負の 2 項回帰の応用に関しては、南・Lennert-Cody (2013), 「ゼロの多いデータの解析：負の 2 項回帰モデルによる傾向の過大推定」を参照のこと。

Excel によるポアソン回帰

Excel により、一足飛びにガンマ・ポアソン回帰を行う前に、最初にポアソン回帰を行うことにする。その Excel シートに追加変更をし、負の 2 項分布を仮定したガンマ・ポアソン回帰を行う。第 7.3 節で事例を示したように、対数尤度の推定結果により AIC あるいは AICc を求め、モデルのあてはまりについて比較検討も容易にできる。

表 9.1 に示したデーリストを表 9.21 に示すように行方向に展開し、表 9.15 に示した最初の水準を基準とするデザイン変数とする。Region は、[A : (0, 0), B : (1, 0), C : (0, 1)] とするデザイン変数 (*RegioB*, *RegioC*) とし、Habitat は、(Dry : 0, Wet : 1) とするデザイン変数 *HabitatWet* とし、切片 1 を含めてデザイン行列を生成する。

対数推定値を

$$\begin{aligned}\ln \hat{y}_i &= \ln(\text{FlowerNo}_i) + \mathbf{x}_i \hat{\boldsymbol{\beta}} \\ &= \ln(\text{FlowerNo}_i) + \hat{\beta}_0 + \hat{\beta}_1(\text{RegionB}_i) + \hat{\beta}_2(\text{RegionC}_i) + \hat{\beta}_3(\text{HabitatWet}_i)\end{aligned}$$

で計算する。指数推定値を $\hat{y}_i = \exp(\ln \hat{y}_i)$ で計算し、 $y_i = \text{SeedNo}_i$ としてポアソン確率を Excel の関数

$$\begin{aligned}P_i &= \frac{\hat{y}_i^{y_i} e^{-\hat{y}_i}}{y_i!} \\ &= \text{Poisson.Dist}(y_i, \hat{y}_i, \text{false})\end{aligned}$$

を使い計算する。適当に設定した切片モデル $\hat{\boldsymbol{\beta}}^T = [3 \ 0 \ 0 \ 0]^T$ を初期値に与え、対数尤度 $\ln L$ を計算すると

$$\ln L = \sum_{i=1}^{60} \ln(P_i) = -344.2485$$

となるので、 $\ln L$ を最大化するように $\hat{\boldsymbol{\beta}}^T = [3 \ 0 \ 0 \ 0]^T$ を変化させると、

$$\hat{\boldsymbol{\beta}}^T = [2.7422 \ 0.1903 \ 0.1352 \ 0.4390]^T$$

画得られ、表 9.17 に示した JMP で推定されてパラメータに一致する。

過分散パラメータを推定するためにピアソンのカイ 2 乗の計算を

$$\text{Pearsonカイ2乗} = \sum_{i=1}^{60} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} = 183.5651$$

計算し、自由度 (60-4) で割り、過分散パラメータ 3.2779 を計算する。ここまでの、ガンマ・ポアソン回帰を Excel で行うための準備であり、表 9.17 の JMP での結果に一致する。

表 9.21 ポアソン回帰による過分散パラメータの推定

						$\hat{\beta}_0$	切片		2.7422	変化させるセル			過分散
						$\hat{\beta}_1$	RegionB		0.1903			対数尤度	3.2779
						$\hat{\beta}_2$	RegionC		0.1352			ln L	カイ2乗
						$\hat{\beta}_3$	HabitatWet		0.4390		最大化	-255.3235	183.5651
No	Habitat	Region	切片	RegionB	RegionC	HabitatWet	FlowNo	SeedNo	対数推定値	指数推定値	Poisson確率	対数尤度 i	Poissonカイ2乗
1	Dry	A	1	0	0	0	3	57	3.8408	46.56	0.0177	-4.0331	2.3398
2	Dry	A	1	0	0	0	1	19	2.7422	15.52	0.0633	-2.7593	0.7799
3	Dry	A	1	0	0	0	1	11	2.7422	15.52	0.0573	-2.8591	1.3168
4	Dry	A	1	0	0	0	1	12	2.7422	15.52	0.0741	-2.6018	0.7987
5	Dry	A	1	0	0	0	3	46	3.8408	46.56	0.0585	-2.8385	0.0068
:													
58	Wet	C	1	0	1	1	1	14	3.3164	27.56	0.0018	-6.3222	6.6719
59	Wet	C	1	0	1	1	4	129	4.7027	110.24	0.0077	-4.8621	3.1922
60	Wet	C	1	0	1	1	3	101	4.4150	82.68	0.0060	-5.1216	4.0590

Excel によるガンマ・ポアソン回帰（負の 2 項回帰）

ポアソン回帰のための Excel の計算シートがあれば、若干の追加でガンマ・ポアソン回帰の計算シートを作成することができる。過分散パラメータ σ を追加し、ポアソン確率の計算をガンマ・ポアソン確率に変更する。ガンマ・ポアソン確率は、

$$\begin{aligned} \text{GammaPoisson}(y; \mu, \sigma) \\ = \exp \left[\ln \Gamma(y+1/\sigma) - \ln \Gamma(y+1) - \ln \Gamma(1/\sigma) + y \ln(\mu\sigma) - (y+1/\sigma) \ln(1+\mu\sigma) \right] \end{aligned}$$

の式に変更する。

ピアソンのカイ 2 乗値は、分散がポアソン分布の場合は $\text{Var}(\hat{y}_i) = \hat{y}_i$ であるが、ガンマ・ポアソン分布の場合は、式 (6.10) に示したように

$$\text{Var}(\hat{y}_i) = \hat{y}_i(1 + \hat{y}_i\hat{\sigma})$$

となるので、分母を

$$\text{Pearsonカイ2乗} = \sum_{i=1}^{60} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i(1 + \hat{y}_i\hat{\sigma})}$$

のように変更する。

初期値を適当に設定した切片モデル $[\hat{\beta} \ \hat{\sigma}]^T = [3.00 \ 0.00 \ 0.00 \ 0.00 \ 0.10]^T$ とすると、対数尤度 $\ln L$ は、

$$\ln L = \sum_{i=1}^{60} \ln(P_i^{(GP)}) = -246.1822$$

となるので、 $\ln L$ 最大するように $[\hat{\beta} \ \hat{\sigma}]^T$ を変化させると

$$[\hat{\beta} \ \hat{\sigma}]^T = [2.7477 \ 0.1629 \ 0.0927 \ 0.4530 \ 0.0429]^T$$

と、表 9.22 に示すような結果を得る。対数尤度は、 $\ln L = -226.2417$ とポアソン回帰の場合に比べ 29.0818 大きくなっている。ピアソンのカイ 2 乗も 60.6647 となり、自由度とほぼ同じ大きさとなり、あてはまりは極めてよい。

表 9.22 ガンマ・ポアソン（負の 2 項）回帰

					$\hat{\beta}_0$		切片=	2.7477	変化させるセル				
					$\hat{\beta}_1$		RegionB=	0.1629					
					$\hat{\beta}_2$		RegionC=	0.0927				対数尤度	
					$\hat{\beta}_3$		HabitatWet=	0.4530				$\ln L$	カイ2乗
							$\hat{\sigma}$ =	0.0429			最大化	-226.2417	60.6647
No	Habitat	Region	切片	RegionB	RegionC	HabitatWet	FlowNo	SeedNo	対数推定値	指数推定値	ガンマPoisson	対数尤度 <i>i</i>	Poissonカイ2乗
1	Dry	A	1	0	0	0	3	57	3.8463	46.82	0.0204	-3.8917	0.7363
2	Dry	A	1	0	0	0	1	19	2.7477	15.61	0.0552	-2.8961	0.4422
3	Dry	A	1	0	0	0	1	11	2.7477	15.61	0.0611	-2.7946	0.8146
4	Dry	A	1	0	0	0	1	12	2.7477	15.61	0.0701	-2.6577	0.4993
5	Dry	A	1	0	0	0	3	46	3.8463	46.82	0.0339	-3.3844	0.0048
:													
58	Wet	C	1	0	1	1	1	14	3.2935	26.94	0.0120	-4.4258	2.8833
59	Wet	C	1	0	1	1	4	129	4.6797	107.74	0.0098	-4.6245	0.7465
60	Wet	C	1	0	1	1	3	101	4.3921	80.81	0.0105	-4.5603	1.1304

SAS の GENMOD プロシジャによる負の 2 項分布を用いた場合

SAS の GENMOD プロシジャで、名義尺度の最初の水準を基準とするデザイン変数を作成のために、class ステートメントのオプションを param=ref ref=first と設定する。また、負の 2 項分布の設定は、dist=negbin によって設定できる。

```
Title2 ' <<< 負の 2 項分布 offset >>> ' ;
proc genmod data=d01 ;
  class Habitat Region / param=ref ref=first ;
  model SeedNo = Region Habitat
    / dist=negbin link=log type3 offset=ln_FloweNo covb ;
run ;
```

SAS による実行結果を次に示す.

<<< 負の 2 項分布 offset >>> GENMOD プロシジャ モデルの情報	
データセット	WORK.D01
分布	Negative Binomial
リンク関数	Log
従属変数	SeedNo
オフセット変数	ln_FloweNo
読み込んだオブザベーション数	60
使用されたオブザベーション数	60

分類変数に対してデザイン変数が、指示通り最初の水準を基準とした変数になっていることが確認できる.

分類変数の水準の情報			
分類	値	デザイン変数	
Habitat	Dry	0	
	Wet	1	
Region	A	0	0
	B	1	0
	C	0	1

各種の適合度の基準が出力されている. Pearson のカイ 2 乗は, 60.6645 と Excel の結果 60.6647 と計算誤差範囲である. 完全対数尤度は, -226.2417 と一致している.

適合度評価の基準			
基準	自由度	値	値/自由度
デビアンس	56	61.2206	1.0932
Scaled デビアンス	56	61.2206	1.0932
Pearson カイ 2 乗	56	60.6645	1.0833
Scaled Pearson カイ 2 乗	56	60.6645	1.0833
対数尤度		8440.6031	
完全対数尤度		-226.2417	
AIC (小さいほどよい)		462.4833	
AICC (小さいほどよい)		463.5945	
BIC (小さいほどよい)		472.9551	

パラメータの推定値, 標準誤差などが出力されている. 表 9.22 に示した Excel の推定値と一致している. Dispersion の推定値 0.0429 は, Excel の過分散パラメータ σ に一致している.

最大尤度パラメータ推定値の分析								
パラメータ		自由度	推定値	標準誤差	Wald 95% 信頼限界	Wald カイ 2 乗	Pr > ChiSq	
Intercept		1	2.7477	0.0699	2.6106 2.8847	1544.19	<.0001	
Region	B	1	0.1629	0.0831	-0.0000 0.3258	3.84	0.0500	
Region	C	1	0.0927	0.0833	-0.0705 0.2560	1.24	0.2655	
Habitat	Wet	1	0.4530	0.0678	0.3202 0.5859	44.66	<.0001	
Dispersion		1	0.0429	0.0124	0.0243 0.0757			

Note: 負の 2 項の分散性パラメータは最尤法により推定されました。

Type3 の対数尤度は、Region の場合は、全体の尤度から Region を除いた（切片，Habitat）での尤度との差から計算され、全体の尤度から Habitat を除いた（切片，Region）での尤度との差から計算されている。

Type 3 分析の LR 統計量			
要因	自由度	カイ 2 乗	Pr > ChiSq
Region	2	3.71	0.1565
Habitat	1	33.87	<.0001

パラメータの共分散行列を covb オプションで出力している。この行列を Excel に取り込むことにより各種の推定が行なえ、種々のグラフの作成を可能にする。パラメータ推定値の Habitat の標準誤差 0.0678 の 2 乗が、Prm4 の対角要素 0.004596 になる。

推定値の共分散行列					
	Prm1	Prm2	Prm3	Prm4	Dispersion
Prm1	0.004889	-0.003502	-0.003589	-0.002467	-4.521E-6
Prm2	-0.003502	0.006910	0.003568	-0.000116	-0.000013
Prm3	-0.003589	0.003568	0.006937	0.0000375	-0.000065
Prm4	-0.002467	-0.000116	0.0000375	0.004596	0.0000318
Dispersion	-4.521E-6	-0.000013	-0.000065	0.0000318	0.0001544

下野の R の glm.nb による結果

下野（2010）で示されている R の glm.nb による結果を次に引用し、Excel および SAS の結果と対比する。

Excel と SAS での結果と照合を試みる。Deviance 残差は、-2.1987～2.5954 の範囲で良くあてはまっていることが分かる。SAS では、

```
output out=out1 stdresdev=stdresdev ;
```

をステートメントで挿入し、out1 ファイルを用いて再計算することで結果が得られる。結果の表示は、R の方がコンパクトである。

```
Call:
glm.nb(formula = SeedNo ~ Region + Habitat + offset(log(FlowerNo)),
       data = seed, init.theta = 23.32892933, link = log)

Deviance Residuals:
      Min       1Q   Median       3Q      Max
-2.1987  -0.7527  -0.0828   0.6518   2.5954

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.74768    0.07013   39.178  < 2e-16 ***
RegionB       0.16289    0.08303    1.962   0.0498 *
RegionC       0.09274    0.08335    1.113   0.2658
HabitatWet    0.45303    0.06787    6.675  2.47e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(23.3289) family taken to be 1)

Null deviance: 110.540 on 59 degrees of freedom
Residual deviance: 61.221 on 56 degrees of freedom
AIC: 462.48

Number of Fisher Scoring iterations: 1

              Theta:      23.33
             Std. Err.:      6.74

2 x log-likelihood: -452.483
```

下野（2010）：第 6 図 負の 2 項分布を当てはめた一般化線形モデルの解析結果を引用。

パラメータの推定値は、一致しているが、標準偏差は、小数点 4 桁目でわずかの違いであるが、切片は、（R : 0.7031, SAS : 0.0699）と計算誤差とはいえない違いがある。

R の Dispersion parameter for Negative Binomial (23.3289) は、SAS では見いだせない。Null deviance:110.540 も SAS で見いだせないが、Residual deviance:61.221 は、SAS の「適合度評価の基準」の中のデビアンズに一致する。AIC:462.48 は、SAS の AIC (小さいほどよい):462.4833 に一致する。また、 $2 \times \log\text{-likelihood} : -452.483$ は、「完全対数尤度」 -226.2417 の 2 倍に等しい。SAS と Excel で得られた過分散パラメータ $\sigma : 0.0429$ は、R の出力には見いだせない。

R でも SAS などの統計ソフトで出力された結果に基づいて何らかの考察をしようと思ったときに、統計ソフトの中でどのようなデザイン変数（ダミー変数）が生成されているのかを把握し、出力結果を Excel に取り込み、対数リンクの場合であれば指数を計算するなどの計算を行い、グラフ化することが望まれる。私も多くの統計ソフトを使ってきたが、常に苦勞したのは、結果のグラフ化であった。これまでも多数の JMP のグラフと Excel のグラフを示してきたので、参考にしてもらいたい。

デビアンس (Deviance)

R の出力に、幾つかの Deviance が登場する。最初は、「Deviance Residuals」で、モデルのあてはまりの検討のために、四分位統計量などが出力されている。最後の方に「Residual deviance」が現われる。その前には、「Null deviance」があり、これらが何を意味しているのか、説明だけ聞いてもなかなか理解しづらい。第 11 章で「デビアンス・逸脱度・残差・テコ比」についてあらためて取り上げる。第 1.9 節の「デビアンス」をベースに若干の解説を加えると、「Null deviance」は、切片のみのモデルでの場合のマイナス 2 倍の対数尤度を意味していて、これまでは、「縮小モデル」で統一し、「Null deviance」を使わないようにしてきた。

通常の回帰分析では、回帰直線のあてはまりの検討に「残差プロット」が使われる。Excel の分析ツールの回帰分析に「残差のグラフ作成」のほかに「標準化された残差」などもあるが、Deviance Residuals (デビアンス残差) とは、いったい何なのだろうか。通常の残差ならば、観測データと推定値の差であることは自明であるが、「標準化された残差」も何となく理解できる範囲ではあるが、「デビアンス残差」は、残差らしきものらしいが、「デビアンス」とは、何なのか。「逸脱度」と言われても、ますます理解に苦しむことになる。なぜ「通常の残差」ではいけないのだろうか。「デビアンス残差」の計算方法を通じて理解しなければ、身につかない。

ピアソン残差は、観測データ $y_1 = 57$ に対し、推定値 $\hat{y}_1 = 46.82$ の差に対し、 $Var(\hat{y}_1)$ の平方根で除した

$$\begin{aligned}\text{Pearson残差}_1 &= \frac{y_1 - \hat{y}_1}{\sqrt{\hat{y}_1(1 + \hat{y}_1\hat{\sigma})}} \\ &= \frac{57 - 46.82}{\sqrt{46.82 \times (1 + 46.82 \times 0.0429)}} = 0.8581\end{aligned}$$

として定義されており、この 2 乗が Pearson カイ 2 乗 0.7363 であるが、ポアソン分布は右に裾を引いているので、推定値より小さい場合に残差が過小評価されやすくなる欠点が内在する。

デビアンズ残差は、観測データ $y_1 = 57$ に対し、推定値も $\hat{y}_1 = 57$ とした飽和モデルの場合の対数尤度と、完全モデルでの対数尤度の差に着目した方法である。表 9.23 に飽和モデルの場合の対数尤度の計算結果を示す。

表 9.23 Deviance Residuals vs. Residual deviance

					$\hat{\beta}_0$		切片	2.7477	変化させるセル						
					$\hat{\beta}_1$		RegionB	0.1629							デビ
					$\hat{\beta}_2$		RegionC	0.0927							アンズ
					$\hat{\beta}_3$		HabitatWet	0.4530							差の2倍
							σ	0.0429							
											最大化	-226.24		-195.63	61.2208
No	Habitat	Region	切片	Region B	Region C	Habitat Wet	Flow er	See dNo	対数推定値	指数推定値	ガンマ Poisson	対数尤度 i	ガンマ Poisson	対数尤度 i	デビアンズ残差
								y_i		y_i^{\wedge}		$\ln L_{完全}$		$\ln L_{飽和}$	
1	Dry	A	1	0	0	0	3	57	3.85	46.82	0.02	-3.89	0.03	-3.56	0.8112
2	Dry	A	1	0	0	0	1	19	2.75	15.61	0.06	-2.90	0.07	-2.70	0.6341
3	Dry	A	1	0	0	0	1	11	2.75	15.61	0.06	-2.79	0.10	-2.32	-0.9745
4	Dry	A	1	0	0	0	1	12	2.75	15.61	0.07	-2.66	0.09	-2.38	-0.7492

58	Wet	C	1	0	1	1	1	14	3.29	26.94	0.01	-4.43	0.08	-2.48	-1.9723
59	Wet	C	1	0	1	1	4	129	4.68	107.74	0.01	-4.62	0.01	-4.29	0.8171
60	Wet	C	1	0	1	1	3	101	4.39	80.81	0.01	-4.56	0.02	-4.07	0.9934

表 9.22 の完全モデルの結果に飽和モデルの結果を加えている。

「飽和モデル」の 1 行目は、観測データ $y_1 = 57$ に対し、推定値も $\hat{y}_1 = 57$ とした場合であり、対数尤度 $\ln L_{飽和,1} = -3.56$ が計算されている。デビアンズ残差 *Deviance Residuals* は、これらの差から、第 11 章で示す計算式を用い、

$$\text{式 (11.1)} \quad d_1 = 2[\ln(L_{飽和,1}) - \ln(L_{完全,1})] = 2 \times [-3.56 - (-3.89)] = 0.66$$

$$\text{式 (11.2)} \quad \varepsilon_1^{(D)} = \text{Sign}(y_1 - \hat{y}_1) \sqrt{d_1} = \text{Sign}(57 - 46.82) \sqrt{0.66} = 0.8112$$

として計算されている。「飽和モデル」が常に「完全モデル」の場合よりも確率が高い（等しい場合もある）ので、 $\text{Sign}(y_i - \hat{y}_i)$ によって観測値と予測値の符号でプラスマイナスを $\sqrt{d_1}$ に付けている。このデビアンズ残差の中央値は -0.0828 となり R での結果に一致する。

Residual deviance : 残差デビアンズ（逸脱度）は、飽和モデルの対数尤度 $\ln L_{飽和} = -195.63$ と完全モデルの対数尤度 $\ln L_{完全} = -226.24$ の差の 2 倍で定義されている。実際に計算すると

$$\begin{aligned} \text{Residual deviance} &= 2(\ln L_{飽和} - \ln L_{完全}) \\ &= 2 \times (-195.6313 - 226.2417) = 61.2208 \end{aligned}$$

となる。もちろんデビアンズ残差 *Deviance Residuals* の平方和

$$\begin{aligned} \text{Residual deviance} &= \sum_{i=1}^{20} (\varepsilon_i^{(D)})^2 \\ &= 0.8112^2 + 0.6341^2 + \dots + 0.9934^2 = 61.2208 \end{aligned}$$

に等しい。いずれにしても、対数尤度を自ら計算することが、理解の向上となる。

第 9.4 節で、過分散を考慮した場合の各種の推定値の計算を表 9.6 に示した。負の 2 項回帰の場合も同様に推定可能である。回帰パラメータは、Excel, SAS, および, R で同じ推定値

$$\hat{\beta} = [2.7477 \quad 0.1629 \quad 0.0927 \quad 0.4530]^T$$

が得られている。パラメータの共分散行列は、SAS での結果を Excel に取り込み、花数が 1 とした場合について、表 9.24 に示すように推定値と 95%信頼区間が求められる。

表 9.24 負の 2 項回帰での各種の推定結果

			offset		Region		Habitat					
	生育		花数	切片	B	C	Wet	対数	対数	指数		
x	環境	地域	1	2.7477	0.1629	0.0927	0.4530	推定値	分散	推定値	L95%	U95%
1	Dry	A	1	1	0	0	0	2.7477	0.0049	15.61	13.61	17.90
2	"	B	1	1	1	0	0	2.9106	0.0048	18.37	16.04	21.04
3	"	C	1	1	0	1	0	2.8404	0.0046	17.12	14.98	19.57
4	Wet	A	1	1	0	0	1	3.2007	0.0046	24.55	21.51	28.02
5	"	B	1	1	1	0	1	3.3636	0.0042	28.89	25.44	32.82
6	"	C	1	1	0	1	1	3.2935	0.0044	26.94	23.66	30.67
共分散 $\Sigma(\hat{\beta})$				切片	RegionB	RegionC	HabitatWet					
				切片	0.0049	-0.0035	-0.0036	-0.0025				
				RegionB	-0.0035	0.0069	0.0036	-0.0001				
				RegionC	-0.0036	0.0036	0.0069	0.0000				
				HabitatWet	-0.0025	-0.0001	0.0000	0.0046				

(生育環境 : Dry, 地域 : A) の場合の対数推定値のデザイン変数 x_1 は,

$$x_1 = [1 \quad 0 \quad 0 \quad 0]$$

なので、花数を 1 とした場合の対数推定値は,

$$\ln \hat{y}_{Dry,A} = \ln(1) + x_1 \hat{\beta} = 2.7477$$

と計算され、対数分散は,

$$Var(\ln \hat{y}_{Dry,A}) = x_1 \Sigma(\hat{\beta}) x_1^T = 0.0049$$

であり、指数を取った推定値および 95%信頼区間は,

$$\begin{aligned} \hat{y}_{Dry,A} &= \exp(\ln \hat{y}_{Dry,A}) = 15.61 \\ (L95\%, U95\%) &= \exp \left[\ln \hat{y}_{Dry,A} \pm 1.96 \sqrt{Var(\ln \hat{y}_{Dry,A})} \right] \\ &= (13.61, 17.90) \end{aligned}$$

として計算されている。

過分散を考慮した簡便的な表 9.18 の推定結果と負の 2 項回帰での表 9.24 の推定結果を比較すると、ほぼ同程度の推定値と分散が得られていることが確認される。

第9章 文献索引










久保訳・Murrell著(2009) - R グラフィックス - Rで思い通りのグラフを作図するためにー	295
下野(2010) - Rを用いた一般化線形モデル(回帰係数編): カウントデータを例に	293
南・Cheridy(2013) - ゼロの多いデータの解析: 負の2項回帰モデルによる傾向の過大推定	314
臨床評価研究会(ACE)基礎解析分科会(2017) - 新版 実用SAS生物統計ハンドブック	293

第9章 索引

あ Rグラフィックス - Trellis(格子)グラフ	295	か 組合せ数 - ガンマ関数	313
Rのglm.nb - 下野(2010)	318	グラフ・ビルダー - JMP	295
生育環境別 - 種子数	294	- 平滑線	296
位置パラメータ μ - 負の2項分布	313	結果のグラフ化 - 統計ソフト	319
逸脱度 - デビアン	320	交互作用 - 質的	295
- Residual deviance	320	- 推定値	299
- 残差デビアン	321	- デザイン行列	299
Excel - 折れ線グラフ	312	- 副次的な解析	297
- Gammaln() 関数	313	- ポアソン回帰	297
- ガンマ・ポアソン回帰	314	- 量的	295
- ガンマ・ポアソン確率	315	交互作用プロファイル - 折れ線グラフ	300
- 対数ガンマ関数	313	- JMP	297
- NegBinom.dist() 関数	313	covbオプション - 共分散行列	307
- ポアソン回帰	314	- GENMOD	307
- 予測プロファイル	303	さ 最初の水準を基準 - デザイン変数	309
S-PLUS - Trellis(格子)グラフ	295	SAS - GENMODプロシジャ	305
オフセット - 共変量	309	- DATAステップ	305
- 土壌体積中	294	- Proc genmod	306
- 花数	293, 309	- PROCステップ	305
- 負の2項回帰	313	SASデータセット - proc print	306
- ポアソン回帰	309	SASとR - 臨床評価研究会(2018)	293
- 面積の中	294	残差デビアン - 逸脱度	321
折れ線グラフ - Excel	312	GENMOD - covbオプション	307
- 交互作用プロファイル	300	- Scale=Pearson	306
OnDemand SAS - 無償版	305	- Type3	306
カイ2乗 - ピアソン	314	- 対比型デザイン変数	307
か カウントデータ - 下野(2010)	293	- dist=negbin	316
各種の推定 - 負の2項回帰	322	- Dist=poisson	306
過分散 - 尺度	307	- param=ref ref=first	316
- 調整	310	- Link=log	306
- ピアソンのカイ2乗	315	- Wald検定	307
- 分散/平均の比	296	GENMODプロシジャ - SAS	305
- 無視	309	指数推定値 - 対数推定値	314
過分散を調整 - ポアソン回帰	293	質的 - 交互作用	295
過分散を反映 - 主効果モデル	301	実験計画法 - 要因配置型	297
Gamma Poisson Probability() - JMP	316	実数化 - 成功数	313
Gammaln() 関数 - 対数ガンマ関数	313	失敗の数の分布 - 負の二項分布	313
ガンマ・ポアソン回帰 - Excel	314	下野(2010) - Rのglm.nb	318
ガンマ・ポアソン確率 - Excel	315	- カウントデータ	293
ガンマ・ポアソン分布 - 負の2項分布	313	- 負の二項分布	316, 318
ガンマ関数 - 組合せ数	313	尺度 - 過分散	307
共分散行列 - covbオプション	307	尺度パラメータ σ - 負の2項分布	313
- パラメータ	298, 311	JMP - グラフ・ビルダー	295
共変量 - オフセット	309	- ボックス・プロット	295
- 花数	294	- Gamma Poisson Probability()	316
- 部分母集団	294	- 交互作用プロファイル	297
行列関数 - 推定値を計算	310	- 予測プロファイル	297
久保訳(2009) - Trellis作図	295	縮小モデル - Null deviance	320
- latticeパッケージ	295	主効果モデル - 過分散を反映	301

さ	- 予測プロファイル	302	は	分散/平均の比 - 過分散	296
	種子数 - 生育環境別	294		分類変数 - デザイン変数	317
	- 地域別	294		平滑線 - グラフ・ビルダー	296
	推定値 - 交互作用	299		変量効果 - 標示因子	297
	推定値を計算 - 行列関数	310		ポアソン回帰 - Excel	314
	Scale=Pearson - GENMOD	306		- オフセット	309
	成功数 - 実数化	313		- 過分散を調整	293
	生物統計ハンドブック - 臨床評価研究会 (20	293		- 交互作用	297
た	Type3 - GENMOD	306		- 探索的解析	297
	対数ガンマ関数 - Gammaln() 関数	313		飽和モデル - 完全モデル	321
	対数推定値 - 指数推定値	314		ボックス・プロット - JMP	295
	対比型 - デザイン行列	298	ま	南ら(2013) - 負の2項回帰	314
	対比型デザイン変数 - GENMOD	307		無視 - 過分散	309
	探索的解析 - ポアソン回帰	297		無償版 - OnDemand SAS	305
	地域別 - 種子数	294		面積の中 - オフセット	294
	調整 - 過分散	310		要因配置型 - 実験計画法	297
	dist=negbin - GENMOD	316	や	予測プロファイル - Excel	303
	Dist=poisson - GENMOD	306		- JMP	297
	デザイン行列 - 交互作用	299		- 主効果モデル	302
	- 対比型	298		読み込みポインター - DATAステップ	305
	デザイン変数 - 最初の水準を基準	309		latticeパッケージ - 久保訳(2009)	295
	- 分類変数	317	ら	量的 - 交互作用	295
	DATAステップ - SAS	305		Link=log - GENMOD	306
	- 読み込みポインター	305		臨床評価研究会(2017) - 生物統計ハンドブック	293
	Deviance - デビアンス	320		臨床評価研究会(2018) - SASとR	293
	Deviance Residuals - デビアンス残差	320		Residual deviance - 逸脱度	320
	Deviance Residuals - Residual deviance	321		- Deviance Residuals	321
	デビアンス - 逸脱度	320	わ	Wald検定 - GENMOD	307
	デビアンス残差 - Deviance Residuals	320			
	- ピアソン残差	320			
	統計ソフト - 結果のグラフ化	319			
	土壌体積中 - オフセット	294			
	Trellis(格子)グラフ - Rグラフィックス	295			
	- S-PLUS	295			
	Trellis作図 - 久保訳(2009)	295			
な	Null deviance - 縮小モデル	320			
	NegBinom.dist() 関数 - Excel	313			
は	花数 - オフセット	293, 309			
	- 共変量	294			
	param=ref ref=first - GENMOD	316			
	パラメータ - 共分散行列	298, 311			
	ピアソン - カイ2乗	314			
	ピアソンのカイ2乗 - 過分散	315			
	ピアソン残差 - デビアンス残差	320			
	標示因子 - 変量効果	297			
	副次的な解析 - 交互作用	297			
	負の2項回帰 - オフセット	313			
	- 各種の推定	322			
	- 南ら(2013)	314			
	負の2項分布 - 位置パラメータ μ	313			
	- ガンマ・ポアソン分布	313			
	- 尺度パラメータ σ	313			
	負の二項分布 - 失敗の数の分布	313			
	- 下野(2010)	316, 318			
	部分母集団 - 共変量	294			
	Proc genmod - SAS	306			
	proc print - SASデータセット	306			
	PROCステップ - SAS	305			

第9章 解析用ファイル一覧

	7 KB	第9章01_花数	JMP Data Table
	18 KB	第9章01_花数	Microsoft Excel ワークシート
	12 KB	第9章02_花数_交互作用	JMP Data Table
	40 KB	第9章02_花数_交互作用	Microsoft Excel ワークシート
	1 KB	第9章03_花数.sas	テキスト ドキュメント
	14 KB	第9章04_花数_オフセット	JMP Data Table
	22 KB	第9章04_花数_オフセット	Microsoft Excel ワークシート
	1 KB	第9章04_花数_負の2項回帰.sas	テキスト ドキュメント
	79 KB	第9章05_花数_負の2項	Microsoft Excel ワークシート

非売品, 無断複製を禁ずる

第 9 回 続高橋セミナー

最尤法によるポアソン回帰分析入門<第 9 章>

第 9 章 花数を共変量とした種子数の探索的ポアソン回帰

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3 の 1005

2020 年 7 月 8 日 高橋 行雄