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 6 月 26 日

第 7 章　過分散がある場合の探索的ポアソン回帰

第 7 章の最初の事例は，第 1.7 節の細菌を用いた 2×2 の実験結果を用い，要因配置型のカウント・データに対する探索的ポアソン回帰のアプローチの基本が示されている．第 2 の事例は，第 1.13 節のカブトガニの観察データについての事例で，2 変量ポアソン回帰，2 元配置ポアソン回帰，さらに共分散分析型ポアソン回帰を扱っている．第 3 の事例は，第 1 章で取り上げなかった事例で，2 群比較において，ポアソン分布が仮定できないカウント・データに対し，ゼロ過剰（Zero-Inflated）ポアソン分布などを扱っている．

第 7 章　目 次

連絡先:高橋 行雄, 〒105-0014 東京都 港区 芝 1―12－3－1005
takahashi.stat@nifty.com ，FAX : 03-342-8035

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

7. 過分散がある場合の探索的ポアソン回帰

第 7 章の最初の事例は，第 1.7 節の細菌を用いた 2×2 の実験結果を用い，要因配置型のカウント・データに対する探索的ポアソン回帰のアプローチの基本が示されている．第 2 の事例は，第 1.13 節のカブトガニの観察データについての事例で，2 変量ポアソン回帰，2 元配置ポアソン回帰，さらに共分散分析型ポアソン回帰を扱っている．第 3 の事例は，第 1 章で取り上げなかった事例で，2 群比較において，ポアソン分布が仮定できないカウント・データに対し，ゼロ過剰（Zero-Inflated）ポアソン分布などを扱っている．

7.1. ネズミチフス菌のコロニー数の事例

稀に起きる現象の観察から得られるカウント・データには，ポアソン分布が良くあてはまるが，稀とは言い難い事象のカウント・データの場合には，過分散となりがちでポアソン分布のあてはまりが悪くなることを第 6 章で示してきた．稀に起きる現象を対象にした実験的研究において，ポアソン分布があてはまることを期待していたが，過分散となってしまった場合を想定しよう．このような場合には，反応が大きめになるような複数の要因が内在していることが疑われる．

異なる実験条件データの併合

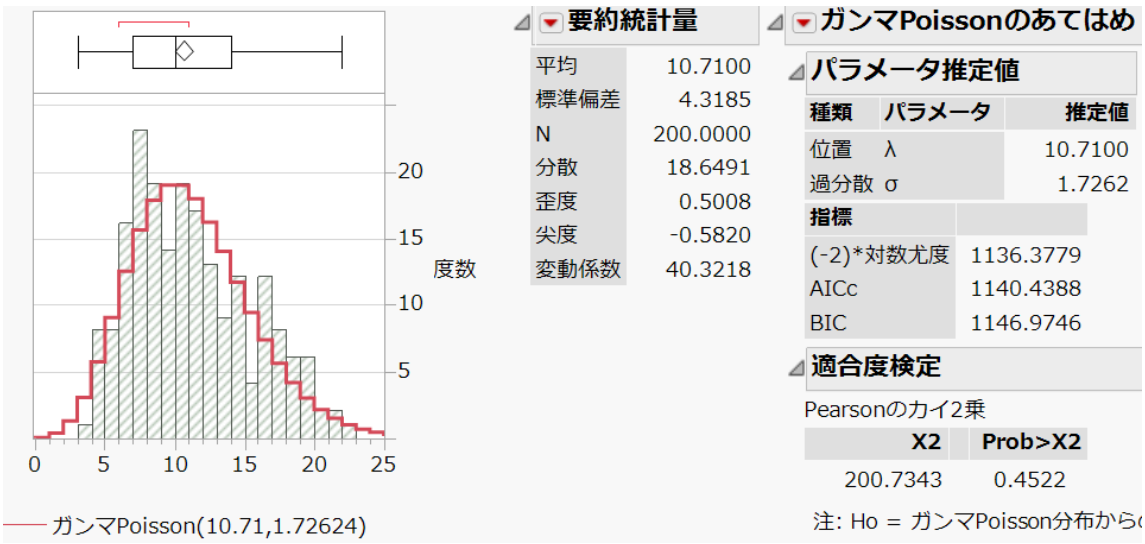
第 1.7 節で，ネズミチフス菌を用いた実験データの事例を示し，それぞれの条件下でポアソン分布があてはめることを示した［吉村ら（1992)］．実験は，2 つの要因に対する陰性対照群の特質を吟味するための実験的研究データであった．探索的なデータ解析を行う事例として表 7.1 に示すように実験は，あらかじめ設定した実験条件が結果におよぼす影響がないことを前提とし，200 枚のシャーレ全体の結果を得たとする．

表 7.1　ネズミチフス菌株に関するコロニー数（表 1.22：再掲）

		コロニー数																				
溶媒	活性化	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	計
A1	B1	0	1	0	0	2	1	2	3	5	3	2	3	3	6	6	4	5	1	2	1	50
	B2	1	5	4	10	7	6	5	6	3	1	1	1	0	0	0	0	0	0	0	0	50
A2	B1	0	0	1	1	2	2	3	4	7	7	5	5	1	6	2	2	1	1	0	0	50
	B2	0	2	3	5	12	10	4	6	2	2	1	3	0	0	0	0	0	0	0	0	50
	計	**1**	**8**	**8**	**16**	**23**	**19**	**14**	**19**	**17**	**13**	**9**	**12**	**4**	**12**	**8**	**6**	**6**	**2**	**2**	**1**	**200**

全 200 枚のシャーレ上のコロニー数についてのヒストグラム，要約統計量，ガンマ・ポアソン分布のあてはめ結果を表 7.2 に示す．平均=10.7100，分散=18.6491 とその比は 1.74 と過分散である．表には示していないが，ポアソン分布をあてはめた場合の適合度のカイ 2 乗値は，346.5154 であり，自由度 N-1=199 のカイ 2 乗に従うとして結果は $p < 0.0001$ でありポアソン分布のあてはめは棄却される．過分散を考慮したガンマ・ポアソン分布のあてはめに対する適合度のカイ 2 乗値は，200.7343， $p = 0.4522$ であり，あてはめは棄却されない．

表 7.2　全 200 シャーレ上のコロニー数に対するガンマ・ポアソン分布のあてはめ



説明変数ごとの層別

探索的な解析の第 1 歩は，得られている説明変数と応答変数間の関係の把握である．図 7.1 に示した A:溶媒および B:活性化で層別したヒストグラムを見ると，A:溶媒では，A1 も A2 も

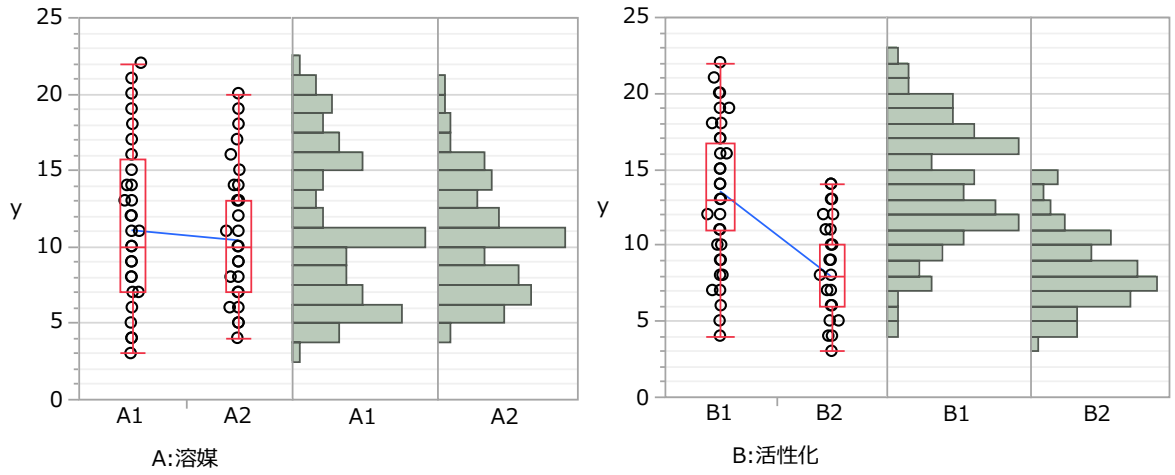


図 7.1　説明変数ごとの層別ヒストグラム

全体の分布と同様であり過分散は解消していない．B:活性化では，B1 と B2 の 2 つの平均値が異なる分布となり，過分散が解消することが期待される．

実際に分散/平均の比を計算した結果を表 7.3 に示す．B:活性化は（B1：1.14，B2：0.80）と全体の 1.74 に比べて縮小している．A:溶媒では（A1：2.19，A2：1.27）と過分散は解消していない．過分散の主な原因は，B:活性化であり，これを考慮すればピュアなポアソン分布と見なすことができそうである，ただし，図 7.1 から，実験条件 B1 に 2 つの山があり，更なる探索が必要とも思われれる．

表 7.3　説明変数ごとの分散/平均の比

要因	水準	N	平均	分散	比
A:溶媒	A1	100	11.0000	24.0388	2.19
	A2	100	10.4000	13.2279	1.27
B:活性化	B1	100	13.5000	15.3635	1.14
	B2	100	7.9000	6.2847	0.80
	全体	200	10.7000	18.6491	1.74

説明変数の組み合わせによる層別

A:溶媒および B:活性化の各 2 水準を組み合わせて図 7.2 に示すように層別ヒストグラムを作成する．A:溶媒の各水準を B:活性化の水準で分割すると過分散が解消さる．

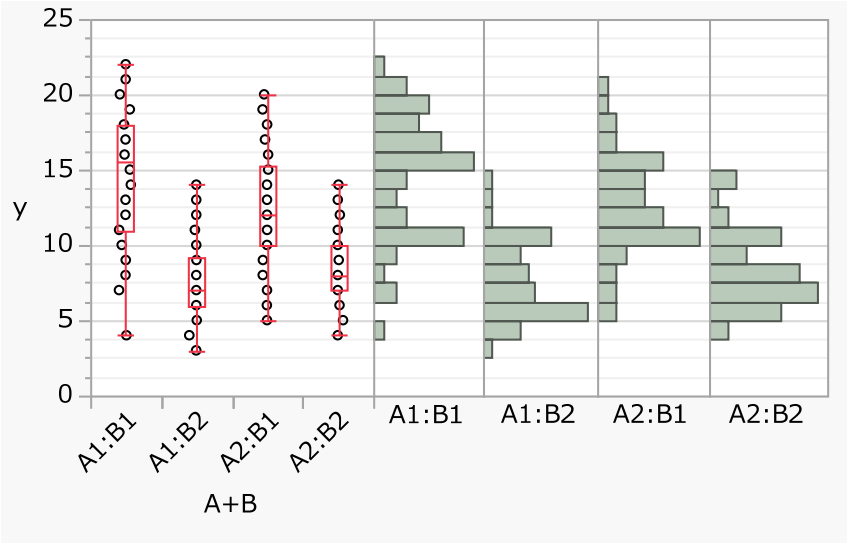


図 7.2　説明変数の組み合わせによる層別ヒストグラム

表 7.4 に示すように分散と平均の比が 1 を超えているのは，A1:B1 であり．平均=14.5000 に対して，分散=17.2331 やや過分散傾向であり，偶然とは思えない 2 つの山があり，まだ同定

できない原因が隠されているかもしれない．他の組み合わせについては，分散が平均より小さめなので過分散が解消されており，A:溶媒と B:活性化の 2 つの要因に対しポアソン分布を仮定した 2×2 の要因配置型の解析が可能となる．これについては，第 3.4～3.5 節に各種のデザイン行列を用いた解析方法と結果の見方についてすでに例示してある．

表 7.4　説明変数の組み合わせによる分散/平均の比

A:溶媒	B:活性化	N	平均	分散	比
A1	B1	50	14.5000	17.2331	1.19
	B2	50	7.5000	6.3351	0.84
A2	B1	50	12.5000	11.6424	0.93
	B2	50	8.3000	6.0833	0.73
	全体	200	10.7000	18.6491	1.74

適合度のカイ 2 乗検定

カウント・データなので，ポアソン分布があてはまることを期待したとしても，本節に示すように異なる発生原因が内在する場合には，過分散となりやすいことを例示した．ポアソン分布とガンマ・ポアソン分布に対する適合度のカイ 2 乗検定について Excel で計算した結果を表 7.5 に示す．コロニー数 y_i，$i=1,2,...,20$ に対し，その頻度 n_i が示されている．平均は 10.7100 なので，y_i に対するポアソン分布の確率は，

$$P_i=\frac{\hat{\mu}^{y_i}e^{-\hat{\mu}}}{y_i!}=\text{Poisson.dist}(y_i,\hat{\mu},\mathit{false})$$

によって計算されている．

ポアソン分布に対する適合度のカイ 2 乗検定は，第 1.7 節の表 1.22 で引用した吉村ら（1992）で使われていたカウント・データ y_i をそのまま使う方法を示した．ポアソン分布のあてはめの場合は，$Var(y_i)=\hat{\mu}$ なので平方和を $\hat{\mu}$ で割った

$$\chi^2=\sum_{i=1}^{20}\frac{n_i(y_i-\hat{\mu})^2}{\hat{\mu}}$$
$$=5.5503+33.6314+\cdots+11.9014=346.5154$$

によって計算されている．この χ^2 は，自由度 $N-1=199$ のカイ 2 乗分布に従うことから

$$p=\text{Chisq.dist.rt}(346.5154,199)=4.49\times10^{-10}$$

がえられ，ポアソン分布のあてはめは支持されない．ガンマ・ポアソン分布のあてはめは，式（6.6）から

$$P_{GPi}=\frac{\Gamma(y_i+1/\hat{\sigma})}{\Gamma(y_i+1)\Gamma(1/\hat{\sigma})}\frac{(\hat{\mu}\hat{\sigma})^{y_i}}{(1+\hat{\mu}\hat{\sigma})^{y_i+1/\hat{\sigma}}}$$

表 7.5　ポアソン分布およびガンマ・ポアソン分布のあてはめ

	コロニー			ポアソン分布				ガンマ･ポアソン分布		
	数	頻度		確率	適合度	推定値		確率	適合度	推定値
i	y	n		P	カイ2乗	n^		P_{GP}	カイ2乗	n^
1	3	1		0.0046	5.5503	0.9		0.0154	3.2153	3.1
2	4	8		0.0122	33.6314	2.4		0.0287	19.4825	5.7
3	5	8		0.0262	24.3541	5.2		0.0453	14.1082	9.1
4	6	16		0.0468	33.1415	9.4		0.0627	19.1987	12.5
5	7	23		0.0716	29.5588	14.3		0.0782	17.1232	15.6
6	8	19		0.0958	13.0287	19.2		0.0894	7.5475	17.9
7	9	14		0.1140	3.8224	22.8		0.0951	2.2143	19.0
8	10	19		0.1221	0.8943	24.4		0.0950	0.5181	19.0
9	11	17		0.1189	0.1335	23.8		0.0899	0.0773	18.0
10	12	13		0.1061	2.0199	21.2		0.0812	1.1701	16.2
11	13	9		0.0874	4.4068	17.5		0.0703	2.5528	14.1
12	14	12		0.0669	12.1278	13.4		0.0586	7.0256	11.7
13	15	4		0.0478	6.8736	9.6		0.0472	3.9818	9.4
14	16	12		0.0320	31.3547	6.4		0.0369	18.1636	7.4
15	17	8		0.0201	29.5530	4.0		0.0281	17.1199	5.6
16	18	6		0.0120	29.7726	2.4		0.0209	17.2471	4.2
17	19	6		0.0068	38.5009	1.4		0.0151	22.3033	3.0
18	20	2		0.0036	16.1165	0.7		0.0107	9.3362	2.1
19	21	2		0.0018	19.7729	0.4		0.0075	11.4543	1.5
20	22	1		0.0009	11.9014	0.2		0.0051	6.8944	1.0
計	20	200	μ^=	10.7100	**346.5154**		μ^=	**10.7100**	**200.7343**	
区分数		N			199		σ^=	**0.0678**		
					0.0000		σ'^=	1.7262		
		$Var(y)$=μ^=		10.7100	$Var(y)$=μ^(μ^σ^+1))=			18.4880		

である．ガンマ・ポアソン分布のパラメータは，Excel のソルバーで対数尤度を最大化するように $\hat{\mu}$ と $\hat{\sigma}$ を変化させて推定する．表 7.2 の JMP で求められた過分散パラメータ $\hat{\sigma}'=1.7262$ は，$\sigma'=1+\mu\sigma$ の関係から $\hat{\sigma}$ を

$$\hat{\sigma}=\frac{\hat{\sigma}'-1}{\hat{\mu}}=\frac{1.7262-1}{10.7100}=0.0678$$

と換算することができる．適合度のカイ 2 乗検定は，式（6.10）で示したガンマ・ポアソン分布の分散

$$\begin{aligned} Var(\hat{y}) &= \hat{\mu}(\hat{\mu}\hat{\sigma}+1) \\ &= 10.7100\times(10.7100\times 0.0678+1)=18.4880 \end{aligned}$$

を用いて，

$$\begin{aligned} \chi^2_{GP} &= \sum_{i=1}^{20}\frac{n_i(y_i-\hat{\mu})^2}{\hat{\mu}(\hat{\mu}\hat{\sigma}+1)} \\ &= 3.2153+19.4825+\cdots+6.8944=200.7343 \end{aligned}$$

となる．χ^2_{GP} は，自由度 $N-1=198$ のカイ 2 乗分布に従うことから

$$p=\text{Chisq.dist.rt}(200.7343,198)=0.4522$$

がえられ，ガンマ・ポアソン分布のあてはめは棄却されない．

このように，統計ソフトの出力結果を鵜呑みにすることなく，Excel を用いて再現できるかを確認することは，統計モデルの理論の理解をより確実なものとし，統計ソフトが対応していない課題に対し，応用できる力の根源となる．

JMP のバージョン 14 をベースにした解析例を示してきたのであるが，バージョン 15 が新たに提供されたので，使い始めると結果がまったく異なる事態に直面する．ポアソン分布のあてはめに対する適合度の検定は，表 7.6 に示すようにバージョン 15 では，カイ 2 乗=75.1451 であるのに対し，表 7.7 に示すようにバージョン 14 では，カイ 2 乗=346.5154 とまったく異なる．少なくともバージョン 14 についての適合度の検定は表 7.5 に示すように Excel で再現でるが，バージョン 15 の自由度 12 となっていることも不可解である．新しい理論に基づいているのかも知れないが，調査中である．

表 7.6　JMP バージョン 15 のポアソン分布の適合度の検定

要約統計量

平均	10.71
標準偏差	4.3184657
平均の標準誤差	0.3053616
平均の上側95%	11.31216
平均の下側95%	10.10784
N	200

Poisson分布のあてはめ

パラメータ		推定値	標準誤差	下側95%	上側95%
平均	λ	10.71	0.2314087	10.262827	11.169978

指標	
(-2)*対数尤度	1172.297
AICc	1174.3172
BIC	1177.5954

適合度検定

	X2	自由度	Prob>X2
Pearsonのカイ2乗	75.145111	12	<.0001*

表 7.7　JMP バージョン 14 のポアソン分布の適合度の検定

要約統計量

平均	10.71
標準偏差	4.3184657
平均の標準誤差	0.3053616
平均の上側95%	11.31216
平均の下側95%	10.10784
N	200

Poisson分布のあてはめ

パラメータ推定値

種類	パラメータ	推定値	下側95%信頼限界	上側95%信頼限界
尺度	λ	10.71	10.262827	11.169978

指標	
(-2)*対数尤度	1172.297
AICc	1174.3172
BIC	1177.5954

適合度検定

Pearsonのカイ2乗

X2	Prob>X2
346.515406	<.0001*

注: Ho = Poisson分布からのデータ。 p値が小さい場合はHoを棄却。

7.2. カブトガニのサテライト数に対する探索的解析

得られたカウント・データが過分散である場合，第 7.1 節では，名義尺度の 2 つの説明変数で層別することにより，分散と平均の比が小さくなることを示した．第 1.13 節の表 1.45 に示したデータには，説明変数として順序尺度（甲羅の色，後体部の棘の状態）の 2 変数，連続尺度（甲羅の幅，体重）の 2 変数も含まれている［アグレスティ（2003)］．これらの説明変数が，応答変数であるサテライト数にどのような関わり合いがあるのか探索的解析を行う．なお，この探索的解析については，高橋（2019a)「最尤法による探索的ポアソン回帰」で詳細に示している．

甲羅の色・後体部の棘

最初の一歩は，説明変数が名義あるいは順序尺度の場合には，それらのカテゴリ（水準）で層別してサテライト数の分布の形状を観察し，過分散の程度を把握することである．JMP の「二変量の関係」でサテライト数を説明変数で層別した結果を図 7.3 に示す．

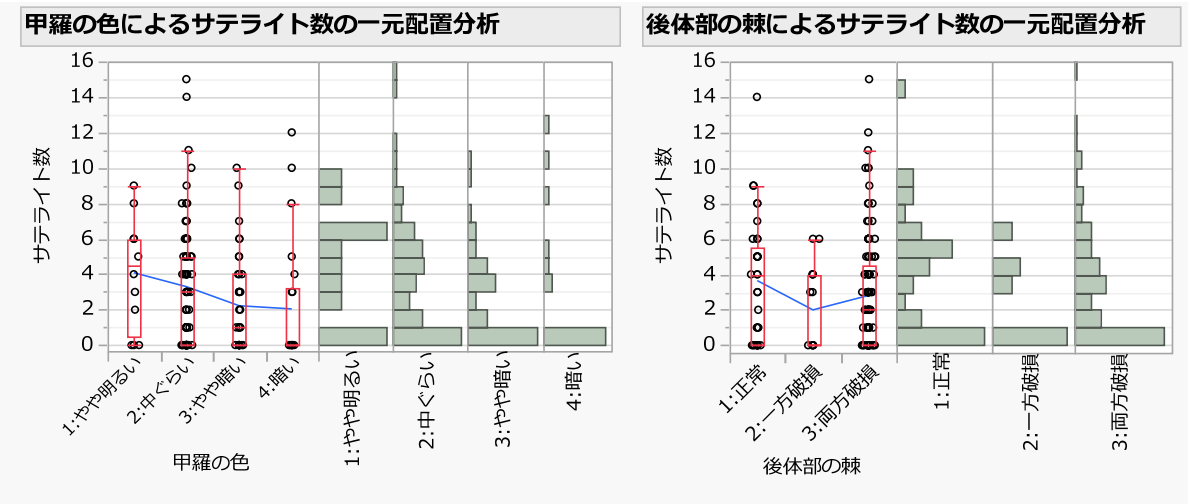


図 7.3　甲羅の色および後体部の棘の状態とサテライト数の関連

雌の甲羅の色については，暗くなるに従いゼロ・カウントの割合が増えサテライト数の平均値が減少傾向であることが読み取れる．雌の後体部の棘の状態については，正常の場合には，サテライト数の 5 匹あたりに山があり，雄が連結する割合が多いようであるが，サテライト数の平均値は同程度である．

表 7.8 に甲羅の色と後体部の棘の状態を組み合わせた場合のサテライト数につての N，平均，分散，および，分散と平均の比を示す．甲羅の色が暗くなるにつれて後体部の棘は，正常

から片方破損，さらに両方破損へ移行するが，ある程度のサテライト数が観察されている場合には，分散と平均の比が 2 以上であり過分散が解消する様子はない．

表 7.8　甲羅の色別 後体部の棘別 の分散/平均の比

甲羅の色	棘の状態	N	平均	分散	分散/平均
1:やや明るい	**1:正常**	**9**	**4.44**	**10.53**	**2.37**
	2:一方破損	2	4.50	4.50	1.00
	3:両方破損	1	0.00	-	-
2:中ぐらい	**1:正常**	**24**	**3.29**	**12.13**	**3.68**
	2:一方破損	8	1.75	6.21	3.55
	3:両方破損	63	3.49	10.03	2.87
3:やや暗い	**1:正常**	**3**	**5.33**	**10.33**	**1.94**
	2:一方破損	4	1.75	4.25	2.43
	3:両方破損	37	2.03	6.25	3.08
4:暗い	**1:正常**	**1**	**0.00**	-	-
	2:一方破損	1	0.00	-	-
	3:両方破損	20	2.25	13.99	6.22
	全体	173	2.92	9.91	3.40

甲羅の幅・体重

雌の甲羅の幅とサテライト数の関連については，第 1.13 節で対数リンクによるポアソン回帰の結果を示した．第 6.6 節および第 6.7 節では，甲羅の幅とサテライト数の関連についてガンマ・ポアソン回帰およびゼロ過剰ガンマ・ポアソン回帰の結果を示した．図 7.4 に示すように，甲羅の幅が小さい場合，および，体重が軽い場合にはサテライト数がゼロの場合が多く，甲羅の幅あるいは体重が大きくなるに従い，サテライト数が急激に増えている．

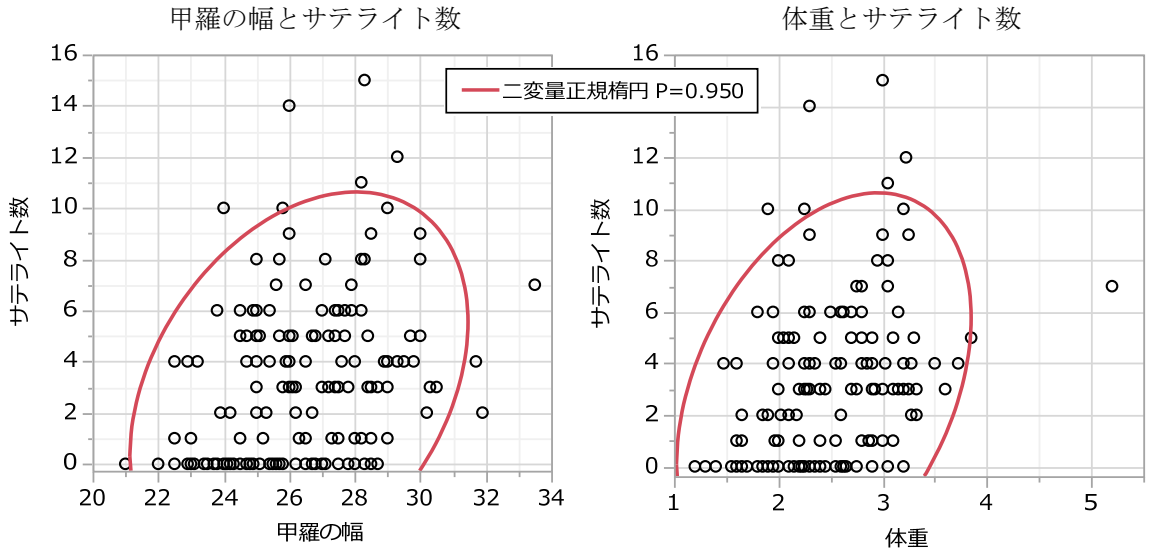


図 7.4　甲羅の幅および体重の増加に伴うとサテライト数の変化

探索的な解析を行う際には，きめ細かなデータの吟味が大切である．サテライト数が甲羅の幅に対して指数関数的に増加するという仮定をしたのだが，散布図を仔細に見れば，幅が 29 cm を超えると，ゼロ・カウントもなくなり，逆にサテライト数は減少傾向となる．体重とサテライト数の関係も，3.0 kg を超えるとサテライト数も同様に減少傾向となる．

甲羅の幅を Y 軸，体重を X 軸，サテライト数をラベルとした散布図を図 7.5 に示す．相関が高いことは自明である．観察すべきことは，サテライト数に対する甲羅の幅と体重の相互関係を読み解くことにある．まず，甲羅の幅を固定して左から右に水平方向にサテライト数の変化を追と増加傾向が読み取れる．次に体体重固定して下から上に垂直方向にサテライト数の変化を追うと少なくなったり多くなったりし，はっきりした傾向が読み取れない．左下から右上は，明らかな増加傾向となってる．

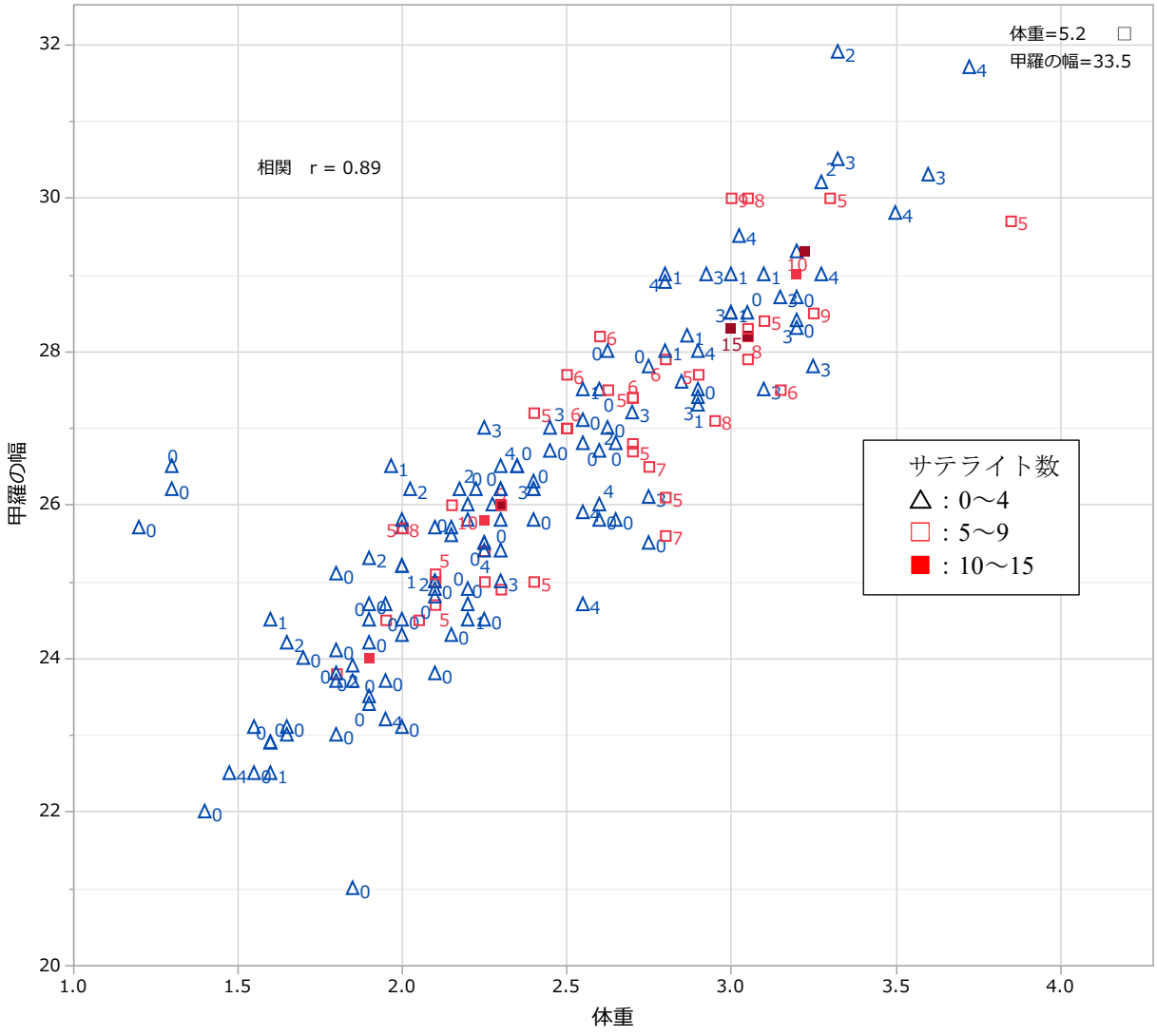


図 7.5　甲羅の幅および体重の散布図にサテライト数と

ポアソン重回帰

過分散を承知で，対数リンクによるポアソン重回帰を行い，甲羅の幅か体重か，どちらがサテライト数との関連が高いか検討する．表 7.9 に示すように，甲羅の幅の推定値は，0.0461，体重の推定値は，0.4470 であり，尤度比検定の結果は，体重のみが有意な差であった．

表 7.9　対数リンクによるポアソン重回帰

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	-1.2952	0.8989	2.0691	0.1503
甲羅の幅	0.0461	0.0467	0.9658	0.3257
体重	0.4470	0.1586	7.9780	0.0047*

共分散

	切片	甲羅の幅	体重
切片	0.8080	-0.0412	0.1156
甲羅の幅	-0.0412	0.0022	-0.0067
体重	0.1156	-0.0067	0.0252

この結果は，甲羅の幅が増加した場合にサテライト数が増大することを否定しているのではなく，図 7.5 で示したように，体重が同じ場合に，甲羅の幅が大きくなる縦方向にサテライト数の増加が見いだされにくいことを反映している．従って，体重が増えれば，甲羅の幅も広くなり，サテライト数も増えるが，体重を固定した場合に甲羅の幅が広くなってもサテライト数はさほど増えないと解釈される．図 7.6 は，JMP による対数リンクでのポアソン 2 重回帰に引き続き「予測プロファイル」の機能を用い，体重を（2，3，4 kg）と変化させた場合

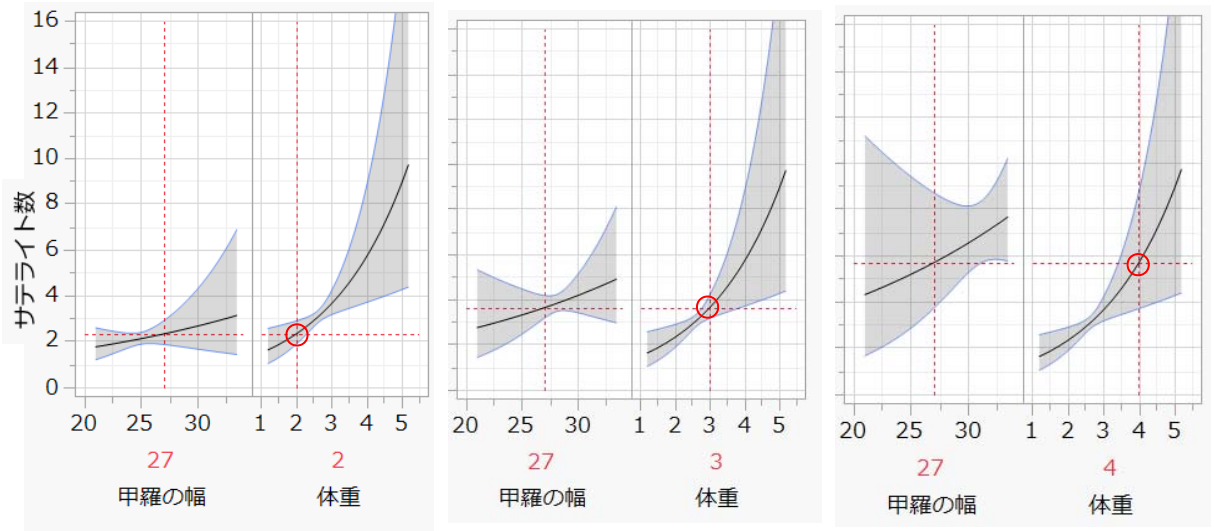


図 7.6　体重を変化させた場合の甲羅の幅とサテライト数との関連

に，甲羅の幅がサテライト数に及ぼす影響を図示したものである．甲羅の幅は体重の増加に

伴いサテライト数も全体的には増加しているが，95%信頼区間の表示から，傾きがマイナスになる可能性が読み取れ，このことが表 7.9 に示したように p=0.3257 であることに対応する．

JMP の予測プロファイルは，ポアソン重回帰に限らず通常の重回帰の場合でも推定結果の可視化に役立つ．これは，1 変数の場合の回帰の推定値に対して 95%信頼区間あるいは 95%予測区間を図示することは一般的であるが，2 変数の場合にはどうしたら良いのだろうか．そこで，2 つの変数の片方を固定し他方を変化させた場合，回帰の推定値および 95%信頼区間を示すことにより可視化する方法 JMP で「予測プロファイル」と称している．

Excel による量的変数に対する予測プロファイル

単回帰分析において回帰直線の 95%信頼区間を散布図上に重ね書きすることは，結果の解釈をするために一般的に行われている．ほとんどの回帰分析の入門書には，95%信頼区間の描画のための式が示されている．ポアソン回帰でも同様に 95%信頼区間の描画のための式を第 1.4 節で示した．

さて，説明変数が 2 以上ある場合には，どのように描いたら良いのであろうか．図 7.6 左に示した JMP の「予測プロファイル」では，甲羅の幅を 27 cm，体重を 2 kg に固定した場合にそれぞれのポアソン回帰曲線と 95%信頼区間が示されている．更に体重を 3 kg に変化させた場合は図 7.6 中，4 kg に変化させた場合は図 7.6 右に図示されている．SAS を含む他の統計ソフトで，このような結果の解釈に有益な表示にこれまで出会ったことがない．

Excel を用いて，JMP による「予測プロファイル」を再現することにより，JMP 内部での計算方法を再現してみる．表 7.9 から推定された回帰パラメータ $\hat{\boldsymbol{\beta}}$ は，

$$\hat{\boldsymbol{\beta}} = [-1.2952 \quad 0.0461 \quad 0.4470]^T$$

である．切片を $x_0 = 1$，甲羅の幅を $x_1 = 27$ とした場合に，体重を x_2 =（1，2，3，4，5）と変化させた場合の推定値 $\hat{y}$ は，

$$\hat{y}_{x_2=1} = \exp[1\times(-1.2952) + 27\times 0.0461 + 1\times 0.4470] = \exp(0.3958) = 1.4856$$
$$\hat{y}_{x_2=1} = \exp[1\times(-1.2952) + 27\times 0.0461 + 2\times 0.4470] = \exp(0.8428) = 2.3228$$
$$;$$
$$\hat{y}_{x_2=1} = \exp[1\times(-1.2952) + 27\times 0.0461 + 5\times 0.4470] = \exp(2.1837) = 8.8791$$

として計算される．それぞれの推定値の分散 $Var(\hat{y})$ は，パラメータの共分散行列を $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ としたときに表 7.9 から

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = \begin{array}{|rrr|}\hline 0.8080 & -0.0412 & 0.1156 \\ -0.0412 & 0.0022 & -0.0067 \\ 0.1156 & -0.0067 & 0.0252 \\ \hline\end{array}$$

なので, 体重が x_2=1 の場合に, $\boldsymbol{x}$=[1　27　1] として, 次の 2 次形式で計算することができる.

$$Var[\ln(\hat{y})] = \boldsymbol{x}\boldsymbol{\Sigma}\boldsymbol{x}^T = 0.0703$$

推定値 $\hat{y}_{x_2=1}$ =1.4856 の 95%信頼区間は,

$$(U95\%\ L95\%,) = \exp\left\{\ln(\hat{y}_{x_2=1}) \pm 1.96\sqrt{Var[\ln(\hat{y}_{x_2=1})]}\right\}$$
$$= \exp\left\{0.3958 \pm 1.96\sqrt{0.0703}\right\} = (0.8836, 2.4978)$$

として計算される. 逆に体重を x_2=2 kg に固定して甲羅の幅を x_1=(21，24，27，30，33)と変化させた場合も同様の計算方法によって推定値および 95%信頼区間を計算することができる.

表 7.10 に予測プロファイルを計算するための Excel シートを示す. 推定値 $\hat{\boldsymbol{\beta}}$ を列ベクトル, パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を与えて任意の $\boldsymbol{x}$=[x_0　x_1　x_2] に対する計算シートとなっている. 甲羅の幅を x_1=27 とし, x_2=（1，2，3，4，5）と変化させた場合，体重を x_2=2 と固定して x_1=（21，24，27，30，33）と変化させた場合の計算結果が示されている. 体重を x_2=3 と変える場合には，Excel シー上で（3，3，3，3，3）と上書きすれば，再計算が行われる.

表 7.10　予測プルファイルの計算のための Excel シート

$\boldsymbol{x}$		推定値 $\boldsymbol{\beta}$	共分散 $\boldsymbol{\Sigma}$	切片	甲羅の幅	体重	
x_0	切片	-1.2952	β_0	**0.8080**	-0.0412	0.1156	
x_1	甲羅の幅	0.0461	β_1	-0.0412	**0.0022**	-0.0067	
x_2	体重	0.4470	β_2	0.1156	-0.0067	**0.0252**	
x_0	x_1	x_2	ln(y^)	*Ver* (ln(y^)	y^	L95%	U95%
1	**28.3**	**3.05**	1.3720	0.0026	**3.9433**	3.5670	4.3592
1	**27**	1	0.3958	0.0703	1.4856	0.8836	2.4978
1	**27**	2	0.8428	0.0125	2.3228	1.8653	2.8927
1	**27**	3	1.2898	0.0051	3.6319	3.1574	4.1778
1	**27**	4	1.7367	0.0480	5.6788	3.6963	8.7245
1	**27**	5	2.1837	0.1412	8.8791	4.2511	18.5457
1	21	**2**	0.5663	0.0391	1.7618	1.1959	2.5954
1	24	**2**	0.7046	0.0061	2.0230	1.7352	2.3584
1	27	**2**	0.8428	0.0125	2.3228	1.8653	2.8927
1	30	**2**	0.9810	0.0583	2.6672	1.6618	4.2809
1	33	**2**	1.1193	0.1434	3.0626	1.4581	6.4324

最初の行の $\boldsymbol{x}$=[1　28.3　3.05] は，表 1.45 で与えられたデータリストの最初のデータである．推定値と 95%信頼区間は，JMP による予測プロファイルの結果と一致する.

Excel シートでの計算式を次に示す.

$\ln(\hat{y}_i)$ = Mmult($\boldsymbol{x}_i$の範囲, $\hat{\boldsymbol{\beta}}$の範囲)

$Var(\ln(\hat{y}_i))$ = Mmult(Mmult($\boldsymbol{x}_i$の範囲, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$の範囲), Transpose($\boldsymbol{x}_i$の範囲))

$$\hat{y}_i = \exp(\ln(\hat{y}_i))$$
$$L95\% = \exp(\ln(\hat{y}_i)) - 1.96\sqrt{Var(\hat{y}_i)}$$
$$U95\% = \exp(\ln(\hat{y}_i)) + 1.96\sqrt{Var(\hat{y}_i)}$$

これらの予測プロファイルを Excel の散布図の機能を使って作図した結果を図 7.7 に示す．この図は，図 7.6 の左端の図に対応する．

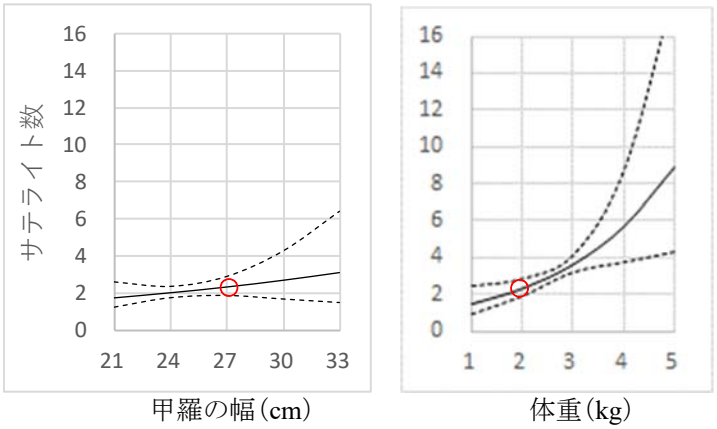


図 7.7　体重を 2 kg に固定した場合の甲羅の幅のプロファイルと
甲羅の幅を 27 cm に固定した場合の体重のプロファイル

交互作用（甲羅の色×体重）を含めたポアソン重回帰

さて，甲羅の色が暗くなるにつれて棘の破損が多くなり，サテライト数の平均が減ることを表 7.8 で示した．では，甲羅の色と体重を組み合わせた場合に，何らかの関連が見出されるのであろうか．名義あるいは順序尺度の甲羅の色と連続尺度の体重の 2 変数がサテライト数に及ぼす影響を観察すためには，図 7.8 に示すように JMP の「二変量の関係」を用いて，「甲羅の色」で「グループ」化し，「層別確率楕円」を描くことにより概観できる．

甲羅の色が「やや明るい」場合には，確率楕円に左右の振れがないので，サテライト数は体重に関連しないようであり，「中ぐらい，やや暗い，暗い」場合は，やや正の相関を持つように傾いており，体重が増えればサテライト数が増えるようである．

この様な関連を，ポアソン回帰で見出すためには，甲羅の色について何らかの数値を使いデザイン行列化し，体重との交互作用を含めたポアソン重回帰を行う必要がある．JMP の一般化線形モデルでは，名義尺度に対しては対比型のデザイン行列を自動生成するので，「モデル効果の構成」で表 7.11 に示すように（甲羅の色，体重，甲羅の色*体重）を設定すればよい．なお，順序尺度として JMP で設定した場合には，対比型とはならないので注意が必要である．

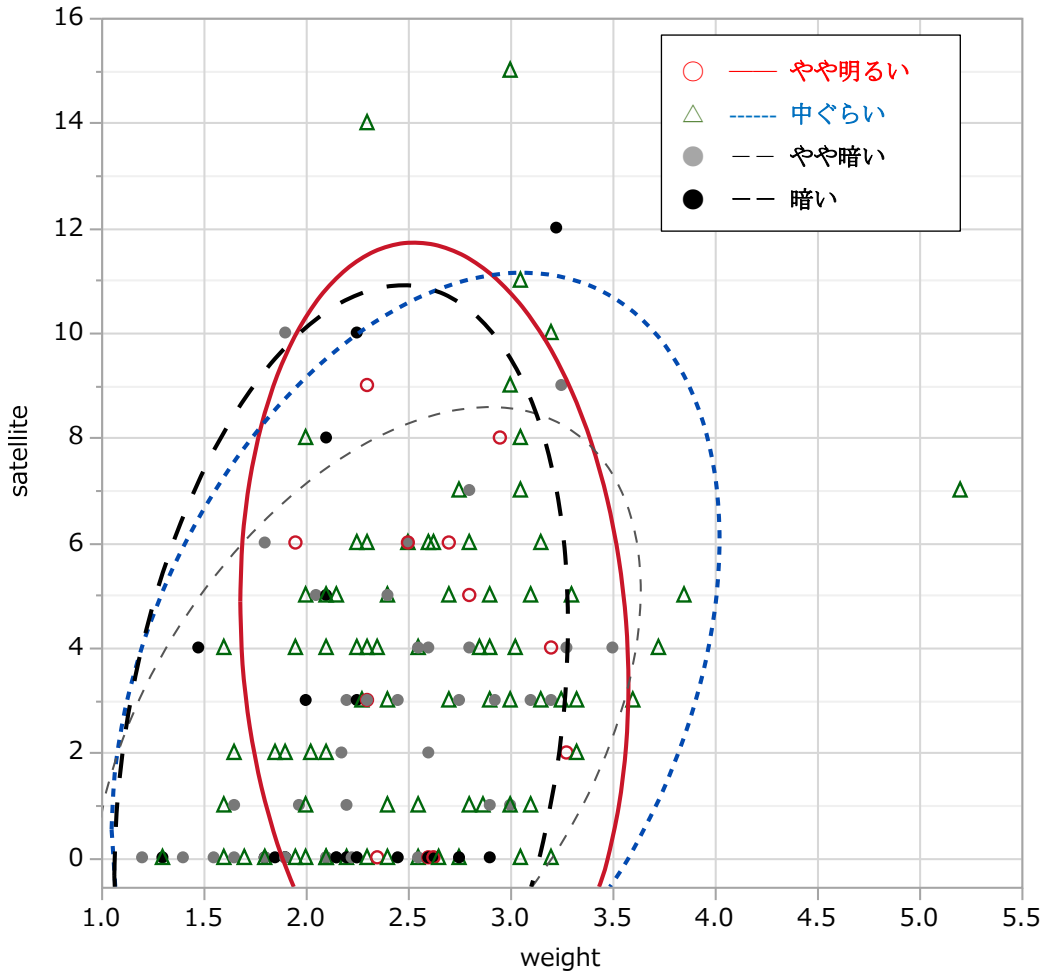
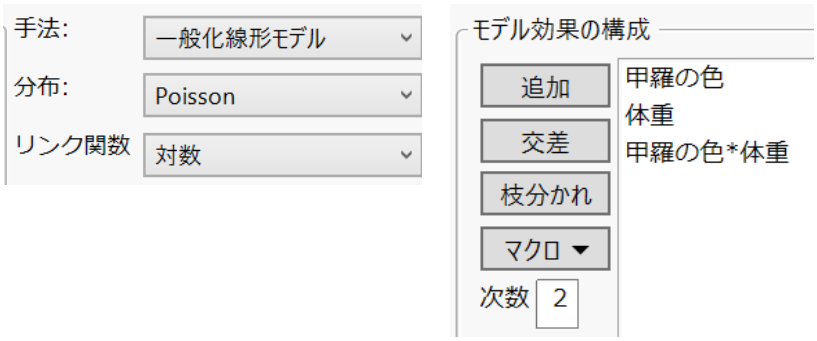


図 7.8　体重とサテライト数に対する甲羅の色による層別確率楕円

表 7.11　JMP の一般化線形モデルにおける交互作用の設定



解析モデルに多水準の名義尺度が含まれ，さらに交互作用が含めると，解析モデルのデザイン変数が膨張する．パラメータの推定結果を表 7.12 に示すが，切片を含めて 1+3+1+3=8 変数となり，このままでは，結果の解釈は困難を極める．そこで，「予測プロファイル」の機能を用いて，図 7.9 に示すように甲羅の色ごとの体重の増加によるサテライト数との関連を概観する．

表 7.12　甲羅の色と体重の交互作用を含めた対数リンクでのポアソン重回帰

項	推定値	標準誤差	尤度比カイ2乗	p値
切片	-0.2778	0.3450	0.6530	0.4191
甲羅の色[1:やや明るい]	2.2221	0.7978	7.5086	0.0061*
甲羅の色[2:中ぐらい]	0.2010	0.3797	0.2812	0.5959
甲羅の色[3:やや暗い]	-1.1855	0.4865	6.0352	0.0140*
体重	0.5463	0.1344	16.0804	<.0001*
甲羅の色[1:やや明るい]*体重	-0.7518	0.3050	6.1530	0.0131*
甲羅の色[2:中ぐらい]*体重	-0.0646	0.1456	0.1967	0.6574
甲羅の色[3:やや暗い]*体重	0.3820	0.1870	4.2010	0.0404*

過分散の調整を行っていないので p 値は小さ目になっている

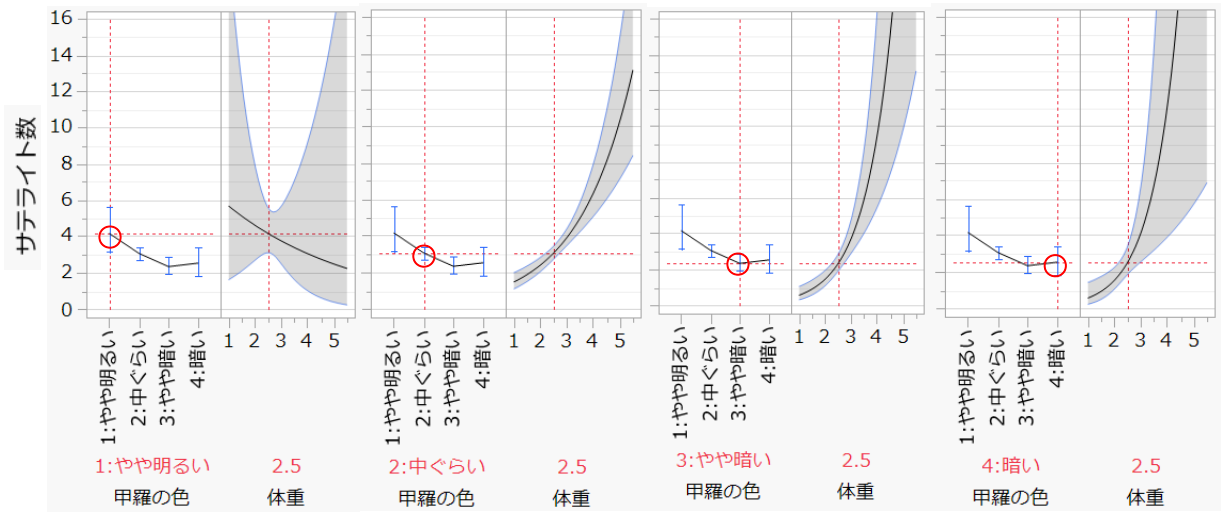


図 7.9　甲羅の色別の体重とサテライト数の関連

この結果は，図 7.8 の層別確率楕円で観察し見出した結果を支持する結果となっている．甲羅の色が「やや明るい」場合は，体重とサテライト数の関連は，マイナスの傾きで，95%信頼区間からプラスの傾きも起こりえる結果となっており，一定の傾向は見いだせない．「中ぐらい」以上では，体重とサテライト数との関連はプラスの傾きで，95%信頼区間からも明らかにプラスの傾きが支持される．

Excel による質的変数を含む予測プロファイル

JMP の名義尺度のダミー変数（デザイン変数）は，表 7.13 に示すように対比型（ダミー変数 x_i について足し算すると 0 となる）とする．体重と甲羅の色の交互作用は，体重とそれぞれのダミー変数の積となる．表 7.12 に対数リンクによるポアソン回帰の推定結果，図 7.9 に予測プロファイルを示した．変数の数は多くなるが，予測プロファイル作成のための計算方法は，2 変数のポアソン回帰の場合と考え方は同じである．

表 7.13　対比型のダミー変数(デザイン変数)

甲羅の色	x_1	x_2	x_3
1:やや明るい	1	0	0
2:中ぐらい	0	1	0
3:やや暗い	0	0	1
4:暗い	-1	-1	-1

Excel による予測プロファイルも変数が増えると煩雑になるが，基本は 2 変数の場合と同じである．表 7.12で得られた推定値，さらに JMP で共分散行列を出力した結果をコピーし，表 7.14 に示すように共分散行列の枠にペーストする．

甲羅の色が「1:やや明るい」場合には，表 7.13 から（$x_1=1$，$x_2=0$，$x_3=0$），甲羅の色が「2:中ぐらい」場合には，（$x_1=0$，$x_2=1$，$x_3=0$），となり，体重を $x_4=$（1，2，3，4，5 kg）と変化させ，交互作用（$x_5=x_1x_4$，$x_6=x_2x_4$，$x_7=x_3x_4$）を計算した推定値結果が示されている．さらに，体重を 2.5 kg に固定し，甲羅の色を（1，2，3，4）と変化させた場合，甲羅の色が「4:暗い」場合は，（$x_1=-1$，$x_2=-1$，$x_3=-1$）であるが，それらの交互作用が計算されている．

表 7.14　予測プルファイルの計算のための Excel シート

		推定値	共分散		甲羅の色			体重	甲羅の色×体重		
	項	$\boldsymbol{\beta}$^	$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^)	切片	A_1	A_2	A_3	W	A_1×W	A_2×W	A_3×W
x_0	切片	-0.278	β_0	**0.119**	0.140	-0.106	-0.060	-0.046	-0.052	0.041	0.024
x_1	A_1	2.222	β_1	0.140	**0.637**	-0.152	-0.199	-0.052	-0.241	0.056	0.074
x_2	A_2	0.201	β_2	-0.106	-0.152	**0.144**	0.048	0.041	0.056	-0.054	-0.020
x_3	A_3	-1.185	β_3	-0.060	-0.199	0.048	**0.237**	0.024	0.074	-0.020	-0.089
x_4	W体重	0.546	β_4	-0.046	-0.052	0.041	0.024	**0.018**	0.019	-0.017	-0.010
x_5	A1×W	-0.752	β_5	-0.052	-0.241	0.056	0.074	0.019	**0.093**	-0.021	-0.028
x_6	A2×W	-0.065	β_6	0.041	0.056	-0.054	-0.020	-0.017	-0.021	**0.021**	0.008
x_7	A3×W	0.382	β_7	0.024	0.074	-0.020	-0.089	-0.010	-0.028	0.008	**0.035**

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	$\ln(y$^)	$Ver(\ln(y$^))	y^	L95%	U95%
1	**1**	0	0	1	1	0	0	1.739	0.405	5.690	1.635	19.803
1	**1**	0	0	2	2	0	0	1.533	0.075	4.633	2.713	7.913
1	**1**	0	0	3	3	0	0	1.328	0.044	3.773	2.498	5.699
1	**1**	0	0	4	4	0	0	1.122	0.314	3.072	1.024	9.212
1	**1**	0	0	5	5	0	0	0.917	0.883	2.501	0.396	15.786
1	0	**1**	0	1	0	1	0	0.405	0.022	1.499	1.119	2.008
1	0	**1**	0	2	0	2	0	0.887	0.007	2.427	2.068	2.848
1	0	**1**	0	3	0	3	0	1.368	0.004	3.928	3.492	4.419
1	0	**1**	0	4	0	4	0	1.850	0.013	6.359	5.082	7.957
1	0	**1**	0	5	0	5	0	2.332	0.035	10.293	7.130	14.861
1	1	0	0	**2.5**	2.5	0	0	1.431	0.022	4.181	3.128	5.589
1	0	1	0	**2.5**	0	2.5	0	1.127	0.004	3.087	2.746	3.471
1	0	0	1	**2.5**	0	0	2.5	0.857	0.010	2.357	1.930	2.879
1	**-1**	**-1**	**-1**	**2.5**	-2.5	-2.5	-2.5	0.936	0.024	2.550	1.887	3.447
	甲羅の色			体重	甲羅の色×体重					推定値	95%信頼区間	

推定値，分散，95%信頼区間の計算は，表 7.10 の Excel シートで示したと同様に次に示す計算式が用いられている．

$$\ln(\hat{y}_i) = \mathrm{Mmult}(\boldsymbol{x}_i\text{の範囲}, \hat{\boldsymbol{\beta}}\text{の範囲})$$
$$(1\times 8) \qquad (8\times 1)$$
$$Var(\ln(\hat{y}_i)) = \mathrm{Mmult}(\mathrm{Mmult}(\boldsymbol{x}_i\text{の範囲}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\text{の範囲}), \mathrm{Transpose}(\boldsymbol{x}_i\text{の範囲}))$$
$$(1\times 8) \qquad (8\times 8) \qquad (8\times 1)$$
$$\hat{y}_i = \exp(\ln(\hat{y}_i))$$
$$L95\% = \exp(\ln(\hat{y}_i)) - 1.96\sqrt{Var(\hat{y}_i)}$$
$$U95\% = \exp(\ln(\hat{y}_i)) + 1.96\sqrt{Var(\hat{y}_i)}$$

甲羅の色を「1 ：やや明るい」に固定し，体重を（1，2，3，4，5）kg と変化させた場合について図 7.10(左)，甲羅の色を「2：中ぐらい」に固定し，体重を（1，2，3，4，5）kg と変化させた場合について，図 7.10(中)，体重を 2.5 kg に固定した場合の甲羅の色のプロファイ（1，2，3，4）を変化させた場合について図 7.10(右）に示す．

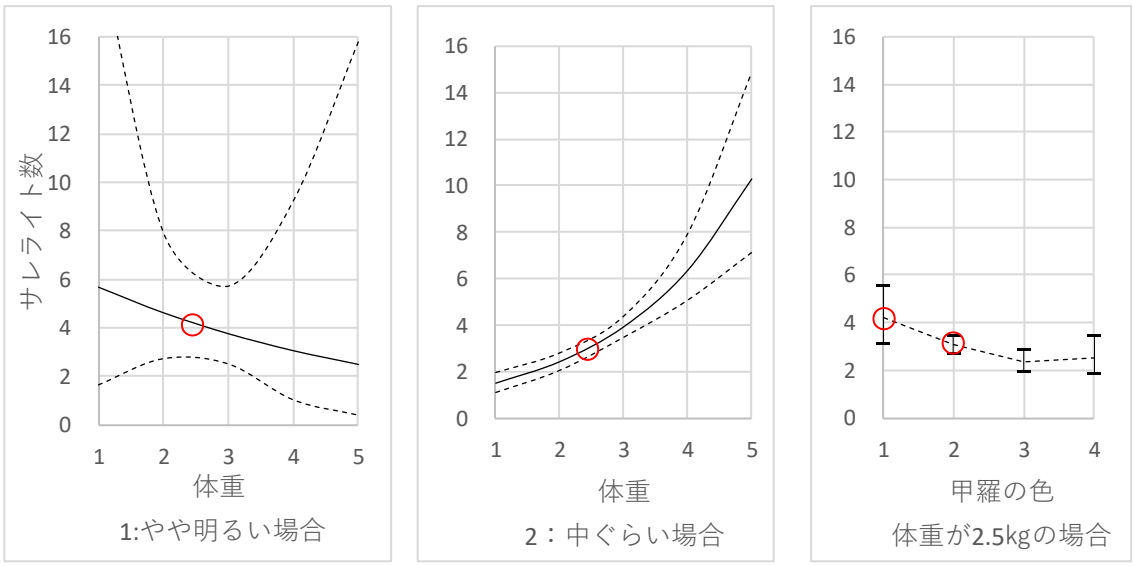


図 7.10　甲羅の色を「1：やや明るい」，「2：中ぐらい」に固定した場合のた場合の体重のプロファイル，体重を 2.5 kg に固定した場合の甲羅の色のプロファイル

交互作用（後体部の棘×体重）を含めたポアソン重回帰

後体部の棘の状態は，甲羅の色によって破損が進行することを表 7.8 で明らかにした．甲羅の色が「中くらい」の場合には，後部の棘が「正常」と「両方破損」に分かれているので，サテライト数との関連を甲羅の色が「中くらい」に限定して関連を調べた結果を図 7.11 に示す．

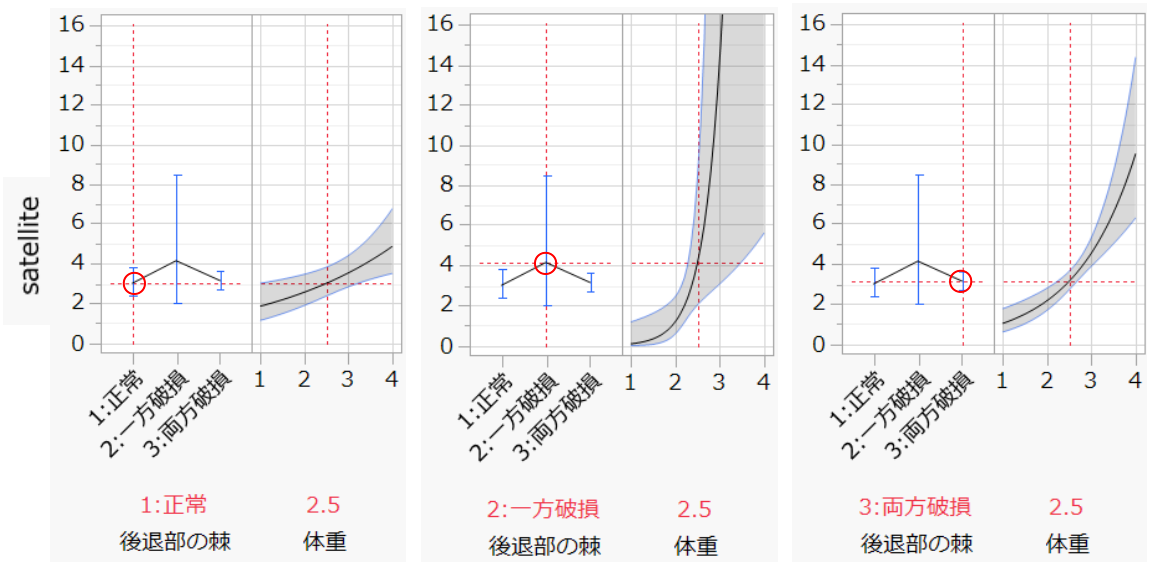


図 7.11　甲羅の色が「中ぐらい」の後部の棘の状態別の体重とサテライト数の関連

甲羅の色が「中ぐらい」で後体部の棘が「正常」の場合に体重が増えればサテライト数も微増する．「一方破損」および「両方破損」では，体重が増えた場合にサテライト数が急増する．表 7.8 から，甲羅の色が「やや明るい」場合には，後体部の棘は 12 匹中 9 匹が「正常」で，図 7.9 から体重が増えてもサテライト数は増えない．甲羅の色が「中ぐらい」に変化すると，体重が増加するとサテライト数も大幅に増える．更に色が「やや暗い，暗い」場合には，更に体重が増えるにつれて，サテライト数が増えるとも言えるが，体重が小さい場合には，サテライト数が減少することが読み取れる．

グラフ・ビルダーによる探索解析的

交互作用が疑われるような観察データに対し，探索的な解析を行うためには，各種のグラフ表示が欠かせない．これまでも JMP の多彩なグラフ表示を活用し，カブトカニの各種の変数とサテライト数の関連を浮き彫りにしてきたが，満足できるものではなかった．全体を俯瞰できるように結果を 1 枚のグラフで表わすことは，可能なのだろうか．JMP の新しい作図機能である「グラフ・ビルダー」を用いた結果を図 7.12 に示す．

この図から，これまでの探索的解析の結果がより鮮明に浮彫される．サテライト数は，甲羅の色が暗くなるにつて後方の棘の破損が進み，それに伴い，体重の軽い雌ほど連結する雄のサテライト数が減少することが読み取れる．甲羅の色が暗くなり，後部の棘の状態が悪くなる加齢現象により，体重の軽い雌ほど連結する雄のサテライト数が減少すると解される．そのため，ゼロ・カウントが多い過分散となったと推測される．

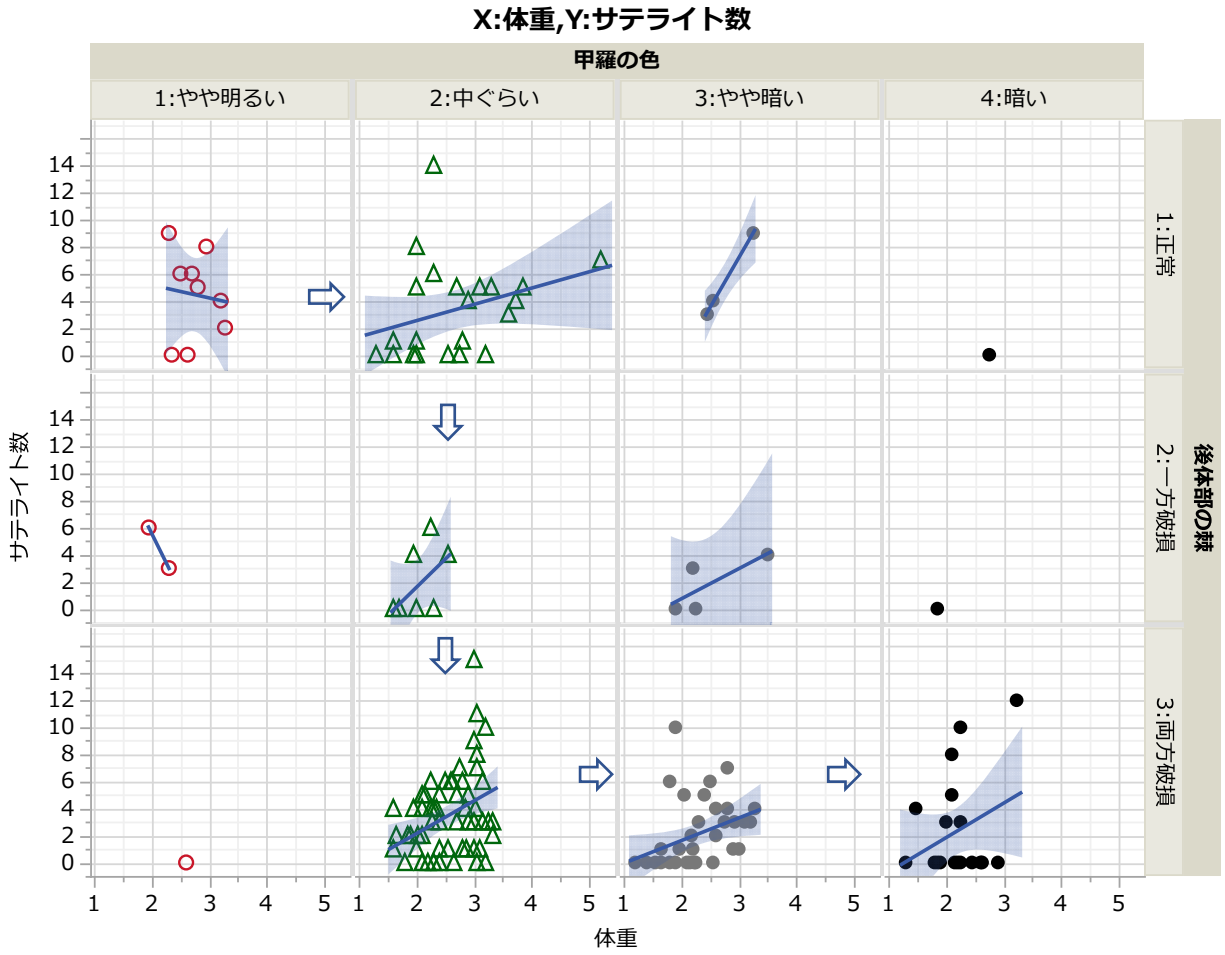


図 7.12　甲羅の色・棘の状態による層別散布図

各セルの中の回帰直線と 95%信頼区間の表示は，通常の回帰分析の結果で，ポアソン回帰の結果ではない．

全データに対するサテライト数の分布について，第 6.6 節で示したようにゼロ過剰ポアソン分布よりも，さらに第 6.7 節で示したようにゼロ過剰ガンマ・ポアソン分布のあてはめが良好であったが，それらの分布を用いた回帰分析には難点がある．これは，図 7.4 にも示したように，甲羅の幅大きい場合，および，体重が重い場合にはサテライト数のゼロが存在しなくなるので，ゼロ過剰ガンマ・ポアソン分布を仮定して回帰分析を行うと，体重が重い場合にも過剰なゼロが存在を仮定することになり，現実のデータとの乖離を無視できなくなるためである．

対数リンクによるポアソン回帰は，元データには指数曲線のあてはめ，両辺に対数を取り線形化するモデルであり，ゼロ・データに対しては対数変換が行われないように調整する仕組みになっている．この仕組みは，一般化線形モデルで分布を正規とし，対数リンクとした

場合でも適用され，ゼロを含むようなデータに対し指数曲線をあてはめることが可能となる．なお，ポアソン回帰を行っても過分散が解消されないような場合に，正規分布を仮定し，対数リンクによる指数曲線をあてはめる場合にも，ゼロ・データに対する調整が行われる．

過剰なゼロが，どのような状況で発生するかを念頭にし，「甲羅の色」，「後体部の棘」とサテライト数の関係から，甲羅の色が暗くなるにつれゼロ・カウントが増加するが，後体部の棘については，関連が見いだされなかった．さらに，甲羅の色と後体部の棘を組み合わせても過分散は解消しなかった．

甲羅の幅と体重の 2 変数間には 0.89 と高い相関があり，2 変数のポアソン回帰に引き続き，図 7.6 に示したように体重を段階的に変化させた場合の甲羅の幅の推定曲線と 95%信頼区間のプロファイルから，甲羅の幅をポアソン回帰の説明変数に加える必要がないことが，視覚的にも見いだされた．もちろん，2 変数のポアソン回帰の尤度比検定で，甲羅の幅の p 値は 0.3257 と有意ではないことからも推測されることではあるが，JMP の予測プロファイル機能は，視覚的に変数相互の関連を見い出し，より具体的な相互関係の理解するために有益である．

この予測プロファイル機能により，図 7.9 に示したように 4 水準の甲羅の色と体重の 2 変数に交互作用を加えたポアソン回帰で，甲羅の色が「やや明るい」場合に，体重が増えてもサテライト数が増えないことが図示され，甲羅の色が「中ぐらい，やや暗い，暗い」場合とは，全く異なるプロファイルであることが明示された．他方，図 7.11 に示すように後体部の棘と体重の関連には，交互作用を示唆するような兆候は見いだせなかった．

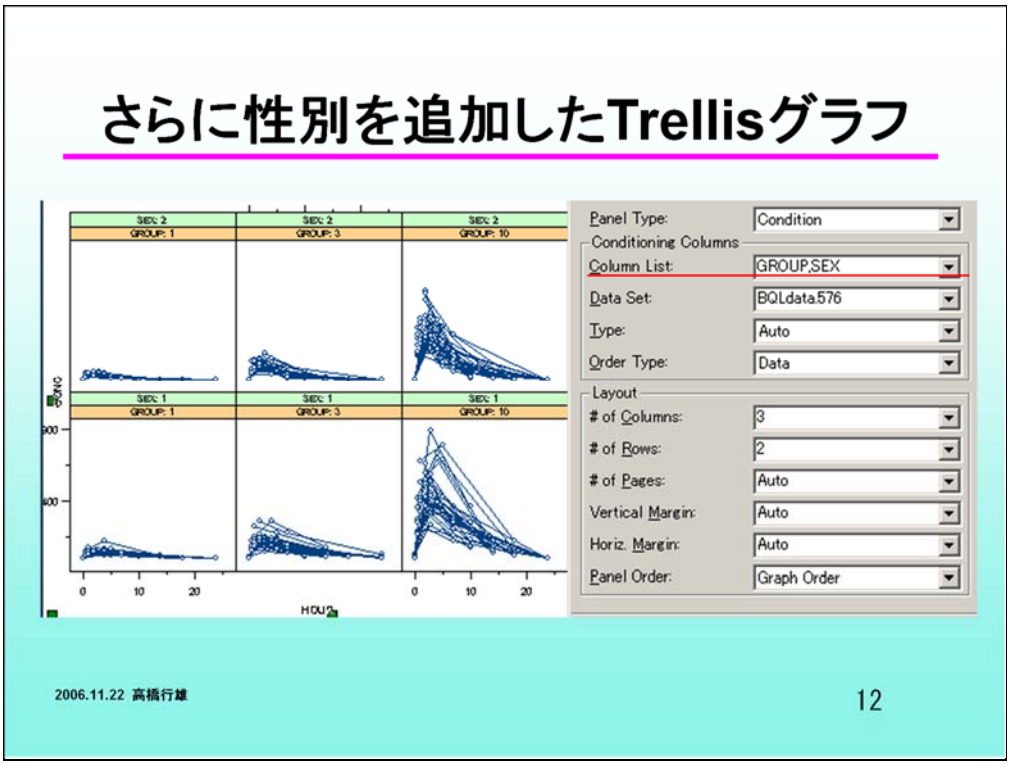
甲羅の色と後体部の棘に体重，さらにそれらの交互作用を含めたポアソン回帰は，観察データなので，データが不均一であり，解を得ることができなかった．これらの変数とサテライト数の関連を見出すためには，図 7.12 に示すように JMP のグラフ・ビルダーが役に立つ．最初に体重とサテライト数の散布図を描き，回帰直線と 95%信頼区間を上書きする．ここまでならば，JMP の伝統的な二変量の関係での対応と同じであるが，これに 4 水準の甲羅の色，3 水準の後体部を組み合わせた 4×3 の場合についてタイル状に体重とサテライト数の回帰直線と 95%信頼区間を並べて表示できた．

グラフ・ビルダーで対数リンクのポアソン回帰が実施できれば申し分ないのであるが，残念ながら現在のバージョン 14 では対応していない．伝統的な回帰分析であっても，名義尺度の水準ごとの散布図行列上に回帰直線の 95%信頼区間が表示されるだけでも，結果を総合的

に俯瞰するために有益である．これに類似する機能が S プラスにあり，以前は愛用していたのであるが，JMP グラフ・ビルダーは，S プラスの機能を大幅に凌駕する探索的な統計解析を支援するツールとして優れている．

S-PLUS の Trellis（格子）グラフ

JMP でグラフ・ビルダーが提供されたときに，S-PLUS の Trellis（格子）グラフを思い出した．Trellis（格子）グラフの有用性については，2006 年の S-PULS ユーザ・コンファレンスで「SAS ユーザのための S-Plus 活用術」を発表した．Web で検索すると（株）NTT データ数理システムのサイトに当時の資料が掲載されているのが見出された．ポアソン回帰の事例ではないが，得られたデータおよび結果のグラフ化の方法について参考にしてもらいたい．久保訳（2009），「R グラフィックス，第 4：lattice パーケージ」も同様と思われる．



http://www.msi.co.jp/splus/usersCase/medical/　2020/05/15 アクセス

株式会社中外臨床研究センター 様
SASユーザのためのS-PLUS活用術で新薬のスピーディーな臨床開発に役立てる

高橋（2006），「SAS ユーザのための S-Plus 活用術」
http://www.msi.co.jp/splus/usersCase/medical/pdf/06takappt.pdf　2020/05/15 アクセス

7.3. 殺人被害者数に関する AICc を用いた分布の同定

Agresti (2013),「Categorical Data Analysis 3rd ed.」の「Section 14.4　Negative Binomial Regression」には,「殺人被害者数に関する調査データ」に対して過分散を考慮した解析結果が示されている．また，このデータを引用して，蓑谷 (2013),「一般線形モデルと生存時間解析」の「第 6.5 節　負の 2 項回帰モデル」で，このデータを引用して論じている．どちらの著書でも，各種の過分散モデルをあてはめ，期待度数と回帰パラメータを主体にした記述がなされている．

JMP によるポアソン回帰

調査は，被検者 1,308 人に対して,「過去 12 か月以内に、殺人の被害者であることを個人的に何人知っていますか」と質問した結果である．表 7.15 は，被検者を（黒人と白人）に層別した結果である．分散と平均の比が（2.2027，1.6828）と 1 を大きく超えていることから調査データに特有の過分散が，全体でも層別した場合でも起きている．

表 7.15　何人の被害者を知っていますか

被害者数	白人	黒人	全体
y	*n*	*n*	*n*
0	1070	119	1189
1	60	16	76
2	14	12	26
3	4	7	11
4	0	3	3
5	0	2	2
6	1	0	1
計	1149	159	1308
平均	0.0923	0.5220	0.1445
分散	0.1552	1.1498	0.2951
分散/平均	1.6828	2.2027	2.0423
1以上の割合	6.9%	25.2%	9.1%

白人を $x_1 = 0$，黒人を $x_1 = 1$ とする標示型デザイン行列とし，分布を Poisson，恒等リンクによるポアソン回帰の結果を表 7.16 に示す．切片の推定値 $\hat{\beta}_0 = 0.0923$ は，表 7.15 で示した白人の平均値であり，x に対する推定値 $\hat{\beta}_1 = 0.4298$ は，黒人の平均値 0.5220 人から白人の平均値 0.0923 との差である．もちろん，推定値 $\hat{y}$ は，白人の場合は，$\hat{\beta}_0 = 0.0923$，黒人の場合は,

$$\hat{y}_{\text{黒人}} = \hat{\beta}_0 + 1 \times \hat{\beta}_1 = 0.0923 + 1 \times 0.4298 = 0.5220 \text{ 人}$$

である．適合度の Pearson のカイ 2 乗は 2279.8732 と自由度 1306 に対して大きく $p < 0.0001$ となり，過分散であることが確認される．

表 7.16　白人 vs 黒人に関するポアソン回帰

手法: 一般化線形モデル
分布: Poisson
リンク関数 恒等
☐ 過分散に基づく検定と信頼区間

適合度統計量	カイ2乗	自由度	p値(Prob>ChiSq)
Pearson	2279.8732	1306	<.0001*
デビアンス	844.7073	1306	1.0000

AICc
1121.9990

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	0.0923	0.0090	106.0000	<.0001*
x	0.4298	0.0580	118.0931	<.0001*

そこで，パラメータの推定値の標準誤差を過分散の調整により大きくする．JMP の「過分散に基づく検定と信頼区間」を考慮した解析を行った結果を表 7.17 に示す．過分散は，Pearson のカイ 2 乗値 2279.8732 を自由度 1306 で割った 1.7457 と推定されている．

表 7.17　白人 vs 黒人に関する過分散を考慮したポアソン回帰

手法: 一般化線形モデル
分布: Poisson
リンク関数 恒等
☑ 過分散に基づく検定と信頼区間

適合度統計量	カイ2乗	自由度	p値	過分散
Pearson	2279.8732	1306	<.0001*	1.7457
デビアンス	844.7073	1306	1.0000	

AICc
646.4465

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(Prob>ChiSq)
切片	0.0923	0.0118	60.7209	<.0001*
x	0.4298	0.0766	67.6483	<.0001*

この過分散 1.7457 を使い表 7.16 のパラメータ推定値 $\hat{\beta}_1 = 0.4298$ の標準誤差 $SE(\hat{\beta}_1) = 0.0580$ を元の分散に戻し，標準誤差を計算し直した結果が表 7.17 に

$$SE(\hat{\beta}_1') = \sqrt{0.0580^2 \times 1.7457} = 0.0766$$

と計算され，尤度比カイ 2 乗=118.0931 は，過分散で除した

$$\text{尤度比カイ2乗}' = 118.0931 / 1.7457 = 67.6483$$

結果となっている．過分散を調整しても白人と黒人間に知っている殺人の被害者数には明らかな差である．

ポアソン回帰における過分散の調整は，簡便な方法で魅力的ではあるが，元の分布をポアソン分布と仮定したままであり，便宜的な方法である．そこで，ゼロ過剰ポアソン分布，ガンマ・ポアソン分布（負の二項分布），ゼロ過剰ガンマ・ポアソン分布（負の二項分布）を仮定する回帰分析を行い，ポアソン分布を仮定した回帰分析の場合と比較する．

Excel によるポアソン回帰

まず，表 7.17 に示した JMP でのポアソン回帰を Excel によって再現する．白人を $x_1=0$，黒人を $x_1=1$ とする標示型デザイン行列とし，ポアソン回帰によるあてはめを行った結果を表 7.18 に示す．切片の推定値 $\hat{\beta}_0=0.0923$ は，表 7.15 で示した白人の平均値であり，x_1 に対する推定値 $\hat{\beta}_1=0.4298$ は，黒人の平均値 0.5220 から白人の平均値 0.0923 との差である．もちろん，推定値 $\hat{y}$ は，白人の場合は，$\hat{\beta}_0=0.0923$，黒人の場合は，

$$\hat{y}_{\text{黒人}}=\hat{\beta}_0+1\times\hat{\beta}_1=0.0923+1\times0.4298=0.5220$$

である．ポアソン確率は，$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1+x_{1,i}$ として Excel の関数を使い，

$$Poisson_i=\text{Poisson.Dist}(y_i,\hat{y}_i,false)$$

で計算されている．この確率を用いて，推定人数

$$\hat{n}_i=n_iPoisson_i$$

を計算し，$n_i-\hat{n}_i$ により，推定人数の偏差によりあてはめの性能を可視化している．

表 7.18　ポアソン分布を仮定した回帰

				G_0	β_0^=	**0.0923**							
				G_1-G_0	β_1^=	**0.4298**						飽和モデル	
人種	i	x_0	x_1	y	n	y^	*Poisson*	$\ln L$	n^	n - n^	χ^2	*Poisson*	$\ln L$
白人	1	1	0	0	1070	0.0923	0.9119	-98.71	1047.74	22.26	98.71	1.0000	0.00
G_0	2	1	0	1	60	0.0923	0.0841	-148.53	96.66	-36.66	535.91	0.3679	-60.00
	3	1	0	2	14	0.0923	0.0039	-77.73	4.46	9.54	552.31	0.2707	-18.30
	4	1	0	3	4	0.0923	0.0001	-36.13	0.14	3.86	366.60	0.2240	-5.98
	5	1	0	4	0	0.0923	0.0000	0.00	0.00	0.00	0.00	0.1954	0.00
	6	1	0	5	0	0.0923	0.0000	0.00	0.00	0.00	0.00	0.1755	0.00
	7	1	0	6	1	0.0923	0.0000	-20.97	0.00	1.00	378.32	0.1606	-1.83
黒人	8	1	1	0	119	0.5220	0.5933	-62.12	94.34	24.66	62.12	1.0000	0.00
G_1	9	1	1	1	16	0.5220	0.3097	-18.75	49.25	-33.25	7.00	0.3679	-16.00
	10	1	1	2	12	0.5220	0.0808	-30.18	12.85	-0.85	50.22	0.2707	-15.68
	11	1	1	3	7	0.5220	0.0141	-29.85	2.24	4.76	82.34	0.2240	-10.47
	12	1	1	4	3	0.5220	0.0018	-18.90	0.29	2.71	69.52	0.1954	-4.90
	13	1	1	5	2	0.5220	0.0002	-17.12	0.03	1.97	76.83	0.1755	-3.48
	14	1	1	6	0	0.5220	0.0000	0.00	0.00	0.00	0.00	0.1606	0.00
				N=	1308		(-2) $\ln L$=	**1117.99**		*Pearson* χ^2=	2279.87	(-2) $\ln L$=	**273.28**
				k=	2		AICc=	1122.00	df=1306,	p=	0.0000	AICc=	273.28
												デビアンス=	**844.71**

統計的な評価としては，それぞれの対数尤度

$$\ln L_i=n_i\ln(Poisson_i)$$

を求め，それらの合計 $\ln L$ の負の 2 倍 $(-2)\ln L$ は，

$$(-2)\ln L=\sum_i\ln(L_i)=1117.99$$

として計算され，AICc は，$N=1{,}308$，$k=2$ として

$$\begin{aligned}\text{AICc} &= (-2)\ln L + 2k + 2k(k+1)/(N-k-1)\\ &= 1117.99 + 2\cdot 2 + 2\cdot 2\cdot(2-1)/(1308-2-1)\\ &= 1122.00\end{aligned}$$

であり，表 7.16 に示した JMP での結果に一致する．Pearson のカイ 2 乗値

$$\text{Pearsonのカイ2乗}=\sum_i \chi_i^2 = \sum_i n_i \frac{(y_i-\hat{y})^2}{\hat{y}} = 2279.8732$$

によって計算されている．これに対し，デビアンスが 844.7073 と全く異なり，過分散ではないとの判断になる．第 11.4 節で詳細に示すが，飽和モデルのマイナス 2 倍の対数尤度を計算すると 273.28 となり，完全モデルの 1117.99 との差が 844.71 と Excel での計算結果と一致する．SAS/GENMOD を使う場合には，どちらの過分散を使うか注意が必要である．

ゼロ過剰ポアソン回帰

ゼロ過剰ポアソン回帰は, ゼロ人（$y_i=0$）場合の過剰な割合を ω とし, ゼロ人でない（$y_i\neq 0$）場合の割合 $(1-\omega)$ に対してポアソン分布を次のように

$$\begin{aligned} y_i=0 &: \quad P_i^{\text{ゼロ}} = \hat{\omega} + (1-\hat{\omega})\cdot \text{Poisson.Dist}(y_i, \hat{y}_i, \mathit{false})\\ y_i\neq 0 &: \quad P_i^{\text{ゼロ}} = (1-\hat{\omega})\cdot \text{Poisson.Dist}(y_i, \hat{y}_i, \mathit{false})\end{aligned}$$

過程して計算する．推定したいパラメータ（$\hat{\omega}$，$\hat{\beta}_0$，$\hat{\beta}_1$）は，適当な初期値を設定し，Excel

表 7.19　Excel によるゼロ過剰ポアソン回帰

					ω^=	**0.7708**	*poisson zero*				
				G_0	β_0^=	**0.4167**	*Intercept*				
				G_1-G_0	β_1^=	**1.4101**	*x*				
人種	*i*	x_0	x_1	*y*	*n*	*y*^	$P^{\text{ゼロ}}$	ln *L*	*n*^	*n* - *n*^	χ^2
白人	1	1	0	0	1070	0.4167	0.9219	-87.03	1059.25	10.75	445.86
G_0	2	1	0	1	60	0.4167	0.0630	-165.91	72.35	-12.35	48.99
	3	1	0	2	14	0.4167	0.0131	-60.67	15.07	-1.07	84.23
	4	1	0	3	4	0.4167	0.0018	-25.23	2.09	1.91	64.06
	5	1	0	4	0	0.4167	0.0002	0.00	0.22	-0.22	0.00
	6	1	0	5	0	0.4167	0.0000	0.00	0.02	-0.02	0.00
	7	1	0	6	1	0.4167	0.0000	-13.72	0.00	1.00	74.81
黒人	8	1	1	0	119	1.8268	0.8077	-25.42	128.42	-9.42	217.39
G_1	9	1	1	1	16	1.8268	0.0674	-43.16	10.71	5.29	5.99
	10	1	1	2	12	1.8268	0.0616	-33.45	9.79	2.21	0.20
	11	1	1	3	7	1.8268	0.0375	-22.99	5.96	1.04	5.27
	12	1	1	4	3	1.8268	0.0171	-12.20	2.72	0.28	7.76
	13	1	1	5	2	1.8268	0.0063	-10.15	0.99	1.01	11.02
	14	1	1	6	0	1.8268	0.0019	0.00	0.30	-0.30	0.00
				N=	1308		(-2) ln *L*=	**999.86**		*Pearson* χ^2=	965.58
				k=	3		AICc=	**1005.88**		*df*=1305, *p*=	1.0000

のソルバーにて，$(-2)\ln L$ を最小化するように（$\hat{\omega}$，$\hat{\beta}_0$，$\hat{\beta}_1$）を変化させて求める．表 7.19 に示すように，（$\hat{\omega}$=0.7708，$\hat{\beta}_0$=0.4167，$\hat{\beta}_1$=1.4101）が得られる．マイナス 2 倍の対数尤度 $(-2)\ln L$ は，

$$(-2)\ln L = \sum_i \ln(L_i) = 999.86$$

として計算され，AICc は，$N = 1{,}308$，$k = 3$ として

$$\begin{aligned} \mathrm{AICc} &= (-2)\ln L + 2k + 2k(k+1)/(N-k-1) \\ &= 1117.99 + 2\cdot 3 + 2\cdot 3\cdot(3-1)/(1308-3-1) \\ &= 1005.88 \end{aligned}$$

とポアソン回帰の場合の $\mathrm{AICc}(Poisson) = 1122.00$ に比べて大幅に減少している．

SAS/GENMED によるゼロ過剰ポアソン回帰

JMP の一般化線形モデルでは，ゼロ過剰ポアソン回帰がサポートされていないので，SAS の GENMOD プロシジャにより結果の検証を行う．分布の設定は，dist=zip を使う．

```
Title2 '<<< ゼロ過剰 Poisson >>>' ;
proc genmod data=d01 ;   /* zero Poisson */
     freq  n ;
     zeromodel  ;
     model y = x / dist=zip link=identity ;
     output out=out03 xbeta=xbeta pzero=poisson_zero ;  run ;
     proc print data=out03 ; run ;
```

表 7.20　SAS GENMOD によるゼロ過剰ポアソン分布を仮定した回帰

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	0.4167	0.0730	0.2737	0.5597	32.61	<.0001
x	1	1.4101	0.2303	0.9587	1.8615	37.48	<.0001
尺度	0	1.0000	0.0000	1.0000	1.0000		
AICC（小さいほどよい）			1005.8798				

Obs	y	G	x	n	xbeta	poisson_zero
1	0	G0	0	1070	0.41669	0.77078
:						
8	0	G1	1	119	1.82678	0.77078
:						

推定値は，$\hat{\beta}_0$=0.4167，$\hat{\beta}_1$=1.4101 と Excel の結果と一致し，ゼロ過剰割合は pzero オプションの出力で *poisson_zero*=0.77078 であり，Excel の $\hat{\omega}$=0.7708 に一致する．AICc も 1005.88 と一致することが確認された．

ガンマ・ポアソン回帰（負の 2 項回帰）

負の二項分布は，第 6.1 節の式（6.2）で出現確率π，および，成功数kとしたときに，失敗数yの分布とし，次のようにガンマ関数を用い

$$NegBinom(y;k,\pi)=\frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)}\pi^k(1-\pi)^y \tag{7.1}$$

と定義されていた．この式のパラメータを負の二項分布の期待値μ（位置パラメータ）および分散に関連する過分散σ（形状パラメータ）となるように変換する．成功の確率πを期待値μとkで，

$$\pi=\frac{k}{\mu+k}$$

で置き換え，

$$GammaPoisson(y;k,\mu)=\frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)}\left(\frac{k}{\mu+k}\right)^k\left(1-\frac{k}{\mu+k}\right)^y \tag{7.2}$$

さらに，kを$1/\sigma$で置き換え，整理すると

$$GammaPoisson(y;\mu,\sigma)=\frac{\Gamma(y+1/\sigma)}{\Gamma(y+1)\Gamma(1/\sigma)}\cdot\frac{(\mu\sigma)^y}{(1+\mu\sigma)^{y+1/\sigma}} \tag{7.3}$$

が得られる．ここで，パラメータμをガンマ・ポアソン回帰の場合に，

$$\mu=\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1x_{1,i}$$

とする．

表 7.18 に示したポアソン回帰をガンマ・ポアソン回帰となるように分布の確率計算を変更した結果を表 7.21 に示す．白人を$x_1=0$，黒人を$x_1=1$とする標示型デザイン行列とし，切片の推定値$\hat{\beta}_0=0.0923$は，表 7.15 で示した白人の平均値であり，x_1に対する推定値$\hat{\beta}_1=0.4298$は，黒人の平均値 0.5220 から白人の平均値 0.0923 との差である．もちろん，推定値$\hat{y}$は，白人の場合は，$\hat{\beta}_0=0.0923$，黒人の場合は，

$$\hat{y}_{黒人}=\hat{\beta}_0+1\times\hat{\beta}_1=0.0923+1\times0.4298=0.5220$$

である．ここまでは，ポアソン回帰の結果と同じである．

ガンマ・ポアソン回帰の場合のパラメータは，ポアソン回帰の場合の（β_0，β_1）に加えてσが加わる．表 7.21 には，マイナス 2 倍の対数尤度$(-2)\ln L$が最小になるように求めた

$\hat{\sigma}=4.9429$ が結果として示されている．白人の $y_1=0$ の場合については，$\hat{y}_1=0.0923$ なので，ガンマ・ポアソン分布の確率は，

$$GammaPoisson(y_1=0;\hat{y}_1=0.0923,\sigma=4.9429)=\frac{\Gamma(y+1/\sigma)}{\Gamma(y+1)\Gamma(1/\sigma)}\cdot\frac{(\mu\sigma)^y}{(1+\mu\sigma)^{y+1/\sigma}}$$

$$=\frac{\Gamma(0+1/4.9429)}{\Gamma(0+1)\Gamma(1/4.9429)}\cdot\frac{(0.0923\times4.9429)^0}{(1+0.0923\times4.9429)^{0+1/4.9429}}$$

$$=\frac{4.5354}{1\times4.5354}\cdot\frac{1}{1.0790}=0.9268$$

として計算されている．この確率を用いて，推定人数

$$\hat{n}_i=n_i GammaPoisson_i$$

を計算し，$n_i-\hat{n}_i$ により，推定人数の偏差によりあてはめの性能を可視化している．

表 7.21　ガンマ・ポアソン分布を仮定した回帰

					β_0^=	**0.0923**	*Intercept*				
				G_0	β_1^=	**0.4298**	*x*				
				G_1-G_0	σ^=	**4.9429**	*Dispersion*				
人種	*i*	x_0	x_1	*y*	*n*	*y*^	*GP*	ln *L*	*n*^	*n* - *n*^	χ^2
白人	1	1	0	0	1070	0.0923	0.9268	-81.33	1064.90	5.10	67.80
G_0	2	1	0	1	60	0.0923	0.0587	-170.09	67.47	-7.47	368.07
	3	1	0	2	14	0.0923	0.0111	-63.07	12.70	1.30	379.34
	4	1	0	3	4	0.0923	0.0025	-23.90	2.92	1.08	251.78
	5	1	0	4	0	0.0923	0.0006	0.00	0.73	-0.73	0.00
	6	1	0	5	0	0.0923	0.0002	0.00	0.19	-0.19	0.00
	7	1	0	6	1	0.0923	0.0000	-10.00	0.05	0.95	259.84
黒人	8	1	1	0	119	0.5220	0.7726	-30.71	122.84	-3.84	17.35
G_1	9	1	1	1	16	0.5220	0.1126	-34.94	17.91	-1.91	1.96
	10	1	1	2	12	0.5220	0.0488	-36.24	7.76	4.24	14.03
	11	1	1	3	7	0.5220	0.0258	-25.60	4.11	2.89	23.00
	12	1	1	4	3	0.5220	0.0149	-12.62	2.37	0.63	19.42
	13	1	1	5	2	0.5220	0.0090	-9.42	1.43	0.57	21.46
	14	1	1	6	0	0.5220	0.0056	0.00	0.90	-0.90	0.00
				N=	1308		(-2) ln *L*=	**995.80**		*Pearson* χ^2=	**1424.03**
				k=	3		AICc=	**1001.82**		*df*=1305, *p*=	0.0114

統計的な評価としては，それぞれの対数尤度

$$\ln L_i=n_i\ln(GP_i)$$

を求め，それらの合計 $\ln L$ の負の 2 倍 $(-2)\ln L$ は，

$$(-2)\ln L=\sum_i \ln(L_i)=995.80$$

として計算され，AICc は， $N=1{,}308$， $k=3$ として

$$\begin{aligned}\mathrm{AICc} &= (-2)\ln L + 2k + 2k(k+1)/(N-k-1)\\ &= 995.80 + 2\cdot 3 + 2\cdot 3\cdot(3-1)/(1308-3-1)\\ &= 1001.82\end{aligned}$$

となる．ポアソン回帰の場合の $\mathrm{AICc}(Poisson)=1122.00$ に比べて大幅な減少となっている．

SAS の GENMOD プロシジャにより結果の検証を行う．分布の設定は，負の二項分布 negbin を使う．

```
Title2 '<<< 負の二項分布 ガンマ・ポアソン >>>' ;
proc genmod data=d01 ;   /* negbin */
     freq  n ;
     model y = x / dist=negbin  link=identity ;
     run ;
```

表 7.22　SAS GENMOD による負の二項分布（ガンマ・ポアソン分布）を仮定した回帰

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	0.0923	0.0108	0.0711	0.1134	72.80	<.0001
x	1	0.4298	0.1090	0.2162	0.6433	15.56	<.0001
Dispersion	1	4.9429	1.0005	3.3242	7.3497		
AICC（小さいほどよい）			1001.8163				

推定値は， $\hat{\beta}_0$=0.0923， $\hat{\beta}_1=0.4298$ と Excel の結果と一致し，Dispersio=4.9429 は，Excel の $\hat{\sigma}=4.9429$ に一致する．AICc も 1001.82 と一致することが確認された．

ゼロ過剰ガンマ・ポアソン回帰

ゼロ過剰ガンマ・ポアソン回帰は，ゼロ人（$y_i=0$）場合の過剰な割合を ω とし，ゼロ人でない（$y_i\neq 0$）場合の割合 $(1-\omega)$ に対してガンマ・ポアソン分布を次のように

$$y_i=0 \ :\ GP_i^{\text{ゼロ}} = \hat{\omega}+(1-\hat{\omega})\cdot GammaPoisson(y_i;\hat{y}_i,\hat{\sigma})$$
$$y_i\neq 0 \ :\ P_i^{\text{ゼロ}} = (1-\hat{\omega})\cdot GammaPoisson(y_i;\hat{y}_i,\hat{\sigma})$$

過程して計算する．推定したいパラメータ（$\hat{\omega}$， $\hat{\beta}_0$， $\hat{\beta}_1$， $\hat{\sigma}$）は，適当な初期値を設定し，Excel のソルバーにて，$(-2)\ln L$ を最小化するようにパラメータを変化させて求める．表 7.23

に示すように，（$\hat{\omega}$=0.6152， $\hat{\beta}_0$=0.2423， $\hat{\beta}_1$=1.0172， $\hat{\sigma}$=1.0192）が得られ，$(-2)\ln L$ は，

$$(-2)\ln L = \sum_i \ln(L_i) = 994.74$$

として計算され，AICc は，$N = 1{,}308$， $k = 4$ として

$$\begin{aligned}\mathrm{AICc} &= (-2)\ln L + 2k + 2k(k+1)/(N-k-1)\\ &= 994.74 + 2\cdot 4 + 2\cdot 4\cdot(4-1)/(1308-4-1)\\ &= 1002.77\end{aligned}$$

とガンマ･ポアソン回帰の場合の AICc(*GammaPoisson*) = 1001.82 に比べてわずかに大きくなっている．

表 7.23　ゼロ過剰ガンマ・ポアソン分布を仮定した回帰

					ω^=	**0.6152**	*pzero*				
				G_0	β_0^=	**0.2424**	*Intercept*				
				G_1-G_0	β_1^=	**1.0172**	*x*				
					σ^=	**1.0190**	*Dispersion*				
人種	*i*	x_0	x_1	*y*	*n*	*y*^	$GP^{ゼロ}$	$\ln L$	*n*^	*n* - *n*^	χ^2
白人	1	1	0	0	1070	0.2424	0.9251	-83.34	1062.91	7.09	207.96
G_0	2	1	0	1	60	0.2424	0.0602	-168.59	69.19	-9.19	113.96
	3	1	0	2	14	0.2424	0.0118	-62.14	13.58	0.42	143.11
	4	1	0	3	4	0.2424	0.0023	-24.26	2.67	1.33	100.65
	5	1	0	4	0	0.2424	0.0005	0.00	0.53	-0.53	0.00
	6	1	0	5	0	0.2424	0.0001	0.00	0.10	-0.10	0.00
	7	1	0	6	1	0.2424	0.0000	-10.93	0.02	0.98	109.69
黒人	8	1	1	0	119	1.2596	0.7864	-28.60	125.03	-6.03	65.64
G_1	9	1	1	1	16	1.2596	0.0944	-37.77	15.01	0.99	0.37
	10	1	1	2	12	1.2596	0.0526	-35.35	8.36	3.64	2.29
	11	1	1	3	7	1.2596	0.0294	-24.70	4.67	2.33	7.37
	12	1	1	4	3	1.2596	0.0164	-12.33	2.61	0.39	7.83
	13	1	1	5	2	1.2596	0.0092	-9.38	1.46	0.54	9.73
	14	1	1	6	0	1.2596	0.0052	0.00	0.82	-0.82	0.00
				N=	1308		(-2) $\ln L$=	**994.74**		*Pearson* χ^2=	**768.61**
				k=	4		AICc=	**1002.77**		*df*=1304, *p*=	1.0000

SAS の GENMOD プロシジャにより結果の検証を行う．分布の設定は，ゼロ過剰負の二項分布 zinbo オプションを使う．

```
Title2 '<<< ゼロ過剰 負の二項分布 ゼロ過剰 ガンマ・ポアソン >>>' ;
proc genmod data=d01 ;   /* zero negbin */
     freq  n ;
     zeromodel  ;
     model y=x / dist=zinb link=identity ;
     output out=out04 pred=pred pzero=pzero ; run ;
     proc print data=out04 ; run ;
```

表 7.24　SAS GENMOD によるゼロ過剰負の二項分布を仮定した回帰

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	0.2424	0.1239	−0.0004	0.4851	3.83	0.0504
x	1	1.0172	0.4483	0.1386	1.8959	5.15	0.0233
Dispersion	1	1.0190	1.1771	0.1059	9.8044		
AICC (小さいほどよい)			1002.7735				

Obs	y	G	x	n	pred	pzero
1	0	G0	0	1070	0.24236	0.61524
:						
8	0	G1	1	119	1.25960	0.61524
:						

推定値は，$\hat{\beta}_0$=0.2424，$\hat{\beta}_1$=1.0172 と Excel の結果と一致し，dispersio=1.0190 は，Excel の $\hat{\sigma}$=1.0190 に一致する．AICc も 1002.77 と一致することが確認された．

仮定した分布間の比較

これまでに取り上げたポアソン分布，ゼロ過剰ポアソン分布，ガンマ・ポアソン分布，ゼロ過剰ガンマ・ポアソン分布を仮定した回帰分析のあてはめの良さについて検討する．表 7.25 に示すように，ポアソン分布を仮定した場合の $(n_i - \hat{n}_i)$ は，$y_i = 0$ の場合にプラス 22.26 人であり，ゼロ過剰ポアソン分布を仮定した場合には，プラス 10.75 人と精度が向上し，AICc での比較でも 1122.00 から 1105.88 と大幅に減少し，あてはめ精度の向上が図られた．

ガンマ・ポアソン分布を仮定した場合は，$y_i = 0$ の場合の $(n_i - \hat{n}_i)$ は，プラス 5.10 人とさらに小さくなり，AICc も 1105.88 から 1001.821 と 4.06 の減少となっている．ガンマ・ポアソン分布を仮定した場合に比べ，ゼロ過剰ガンマ・ポアソン分布を仮定した場合には，マイナス 2 倍の対数尤度 $(-2)\ln L$ は，わずかに増えるが，パラメータ数が 4 となり，AICc は増加している．これらの結果から，ガンマ・ポアソン分布を仮定した回帰が尤もあてはまりがよいとの結果となる．

表 7.25　仮定した 4 分布の性能比較

			Poisson		**ゼロ過剰*Poisson***		**ガンマ*Poisson***		**ゼロ過剰*GP***	
人種	*y*	*n*	*n*^	*n* - *n*^	*n*^	*n* - *n*^	*n*^	*n* - *n*^	*n*^	*n* - *n*^
白人	0	1070	1047.74	22.26	1059.25	10.75	1064.90	5.10	1062.91	7.09
G0	1	60	96.66	-36.66	72.35	-12.35	67.47	-7.47	69.19	-9.19
	2	14	4.46	9.54	15.07	-1.07	12.70	1.30	13.58	0.42
	3	4	0.14	3.86	2.09	1.91	2.92	1.08	2.67	1.33
	4	0	0.00	0.00	0.22	-0.22	0.73	-0.73	0.53	-0.53
	5	0	0.00	0.00	0.02	-0.02	0.19	-0.19	0.10	-0.10
	6	1	0.00	1.00	0.00	1.00	0.05	0.95	0.02	0.98
黒人	0	119	94.34	24.66	128.42	-9.42	122.84	-3.84	125.03	-6.03
G1	1	16	49.25	-33.25	10.71	5.29	17.91	-1.91	15.01	0.99
	2	12	12.85	-0.85	9.79	2.21	7.76	4.24	8.36	3.64
	3	7	2.24	4.76	5.96	1.04	4.11	2.89	4.67	2.33
	4	3	0.29	2.71	2.72	0.28	2.37	0.63	2.61	0.39
	5	2	0.03	1.97	0.99	1.01	1.43	0.57	1.46	0.54
	6	0	0.00	0.00	0.30	-0.30	0.90	-0.90	0.82	-0.82
			(-2) ln *L*=	1117.99		999.86		995.798		994.743
			AICc=	**1122.00**		**1005.88**		**1001.82**		**1002.77**
	AICc順位			**4**		**3**		**1**		**2**
	パラメータ数			2		3		3		4

どのような分布を仮定してポアソン回帰をしたらよいのだろうか. なやましい問題である. 目の前にあるデータだけで決めたとすると，同様な調査を再度行った場合に，そのたびごとに分布の同定を行なうのであろうか．データには誤差的変動が付きまとっていることも考慮すると，目の前のデータだけで分布の同定は，不確性に振り回されることになる.

これまでも種々のカウント・データの例示から，調査データから得られるカウント・データは，過分散になりがちであり，単純にポアソン回帰をあてはめると p 値を低めに推定するバイアスが入り込むことを注意してきた. この例のようにデータ数が多ければ, p 値を出すまでもなく明らかな差であるような場合に，どのような分布があてはまるのかの議論は非生産的である.

調査データは，常に探索的解析の要素があり，この例であれは，性別・年齢階層などにより，知っている被害者の平均値がどの様に変化するのか，その変化は統計的に意味のあるものなのか，あるいは，誤差変動の範囲内なのか，その判断に際し過分散の大きさを考慮するのが現実的と思われる.

第7章 文献索引

アグレスティ著, 渡邉・菅波・吉田・角野・寒水・松永 訳(2003) - カテゴリカルデー解析入門 243
Agresti(2013) - Categorical Data Analysis 3rd ed. 258
久保訳, Murrell著(2009) - Rグラフィックス 257
高橋(2006) - SASユーザのためのS-Plus活用術 257
高橋(2019a) - 最尤法による探索的ポアソン回帰 243
蓑谷(2013) - 一般線形モデルと生存時間解析 258
吉村・大橋 責任編集(1992) - 毒性試験データの統計解析 237

第7章 索引

第 7 章　解析用ファイル一覧

12 KB	第7章01_細菌2x2	JMP Data Table
26 KB	第7章01_細菌2x2	Microsoft Excel ワークシート
11 KB	第7章02_カブトガニ_	Microsoft Excel ワークシート
47 KB	第7章02_カブトガニ_プロファイル	Microsoft Excel ワークシート
16 KB	第7章02a_カブトガニ_クロス表	JMP Data Table
26 KB	第7章02b_カブトガニ_回帰	JMP Data Table
9 KB	第7章02c_カブトガニ_甲羅色_中	JMP Data Table
18 KB	第7章02d_カブトガニ_グラフ・ビルダー	JMP Data Table
5 KB	第7章03_被害者	JMP Data Table
2 KB	第7章03_被害者.sas	テキスト ドキュメント
60 KB	第7章03_被害者	Microsoft Excel ワークシート

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 7 章＞＞

第 7 章　過分散がある場合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 6 月 26 日　　高橋 行雄