

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 9 日

**第 10 章 オフセットを含む探索的ポアソン回帰**

ポアソン回帰の特徴的な事例は、観察対象の部分母集団のサイズが既知で、ある一定期間に発現する事象がカウントされるような場合である。これまでに取り上げた事例は、心疾患による死亡、癌の発生など人数で、対象となる母集団の人数が人口統計学的に得られる事例であった。本章で取り上げるのは、McCullagh and Nelder(1989) , *Generalized Linear Models 2<sup>nd</sup> ed.* の第 6.3.2 節で取り上げられている貨物船の損傷数データである。この事例は、 $5 \times 4 \times 2$  要因配置デザインであり、交互作用も含めて探索的な解析を試みる。

第 10 章 目 次

10. オフセットを含む探索的ポアソン回帰	323
10.1. 貨物船の損傷数 ( $5 \times 4 \times 2$ 要因配置, 対数リンク, オフセット)	323
10.2. 主効果モデルの適用	327
(0,1)型デザイン変数(最初の水準を基準), (1,-1)対比型デザイン行列, 予測プロファイル, 交互作用プロファイル	
10.3. Excel による予測プロファイル	337
10.4. 交互作用の検討	342
10.5. 主効果モデルを活用した新たな交互作用の可視化の試み	346
10.6. Excel のソルバーによるオフセットを含むポアソン回帰	351
10.7. SAS の GENMOD プロシジャを使った解析	354
SAS データセットの作成, 過分散を考慮したポアソン回帰, 負の 2 項回帰	
文献索引, 索引, 解析用ファイル一覧	359

## 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
<b>10. オフセットを含む探索的ポアソン回帰 -----</b>	<b>323</b>
11. デビアン스・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

## 10. オフセットを含む探索的ポアソン回帰

ポアソン回帰の特徴的な事例は、観察対象の部分母集団のサイズが既知で、ある一定期間に発現する事象がカウントされるような場合である。これまでに取り上げた事例は、心疾患による死亡、癌の発生など人数で、対象となる母集団の人数が人口統計学的に得られる事例であった。本章で取り上げるのは、McCullagh and Nelder(1989) , Generalized Linear Models 2<sup>nd</sup> ed. の第 6.3.2 節で取り上げられている貨物船の損傷数データである。この事例は、 $5 \times 4 \times 2$  要因配置デザインであり、交互作用も含めて探索的な解析を試みる。

### 10.1. 貨物船の損傷数 ( $5 \times 4 \times 2$ 要因配置, 対数リンク, オフセット)

McCullagh ら(1989) 貨物船の前方部の損傷数のデータを表 10.1 に示す。このデータは、ロイドの J.Crilley および L.N.Heminway によって提供されたもので、貨物船のタイプ (船種)、建造年、運行年の 3 因子の要因配置型で、それらの組み合わせセル毎に運航期間中に起きた貨物船の前方部への損傷数について集計された結果である。

それぞれの因子の水準は、

船種, 5 水準 : A, B, C, D, E

建造年, 4 水準 : 1960-64, 1965-69, 1970-74, 1975-79

運行年, 2 水準 : 1960-74, 1975-79

であり、 $5 \times 4 \times 2$  の 40 セルごとに、総運行月数に対する損傷数がカウントされている。なお、同一の船での複数回の損傷も含まれている。ただし、建造年が 1975-79 の場合に 1960-74 での運行年セルは、必然的に欠測値となる。船種が E タイプの建造年 1960-74 の場合の 1975-79 運行年のセルは、データが得られなかったための欠測値となっている。損傷数を運行月数で割り 1,000 月あたりの損傷数 (損傷千月比) を算出し加えてある。

表 10.1 に示した損傷数のカウント・データは、運行年 1960-74 と 1975-79 の分類は、総運行月数が同程度になるように分類されたのであろうか。建造年の区切りは、5 年単位となつて

おり、元々のデータは、船種、建造の年月、損傷があった年月、それまでの運行月数、廃船の年月、集計時までの運行月数などと思われるが、解析上の便宜を図るための集計結果が示されていると解釈される。

表 10.1 船舶の前方部への損傷数

船種	建造年	運行年 1960-74				運行年 1975-79			
		No	運行月数 $n_i$	損傷数 $y_i$	損傷千月比	No	運行月数 $n_i$	損傷数 $y_i$	損傷千月比
A	1960-64	1	127	0	0.00	2	63	0	0.00
	1965-69	3	1,095	3	2.74	4	1,095	4	3.65
	1970-74	5	1,512	6	3.97	6	3,353	18	5.37
	1975-79	7	-	-*	-	8	2,244	11	4.90
B	1960-64	9	44,882	39	0.87	10	17,176	29	1.69
	1965-69	11	28,609	58	2.03	12	20,370	53	2.60
	1970-74	13	7,064	12	1.70	14	13,099	44	3.36
	1975-79	15	-	-*	-	16	7,117	18	2.53
C	1960-64	17	1,179	1	0.85	18	552	1	1.81
	1965-69	19	781	0	0.00	20	676	1	1.48
	1970-74	21	783	6	7.66	22	1,948	2	1.03
	1975-79	23	-	-*	-	24	274	1	3.65
D	1960-64	25	251	0	0.00	26	105	0	0.00
	1965-69	27	288	0	0.00	28	192	0	0.00
	1970-74	29	349	2	5.73	30	1,208	11	9.11
	1975-79	31	-	-*	-	32	2,051	4	1.95
E	1960-64	33	45	0	0.00	34	-	-**	-
	1965-69	35	789	7	8.87	36	437	7	16.02
	1970-74	37	1,157	5	4.32	38	2,161	12	5.55
	1975-79	39	-	-*	-	40	542	1	1.85
		計	88,911	139		計	74,663	217	
千月比損傷数=1,000×( $y_i/n_i$ )				*必然的に空のセル, **誤って空のセル					

運行月数は、最も少ないもので 45 月、最大のもので 44,882 月であり、損傷数を運行月数で除して基準化し、通常の線形モデルを適用したくなる。図 10.1 に示すように 1,000 月あたりの損傷数を四捨五入し整数化して分布の形状を確認すると分布は、指数分布的である。また 1,000 月あたりの 34 セルの組み合わせの損傷数の平均は 3.1765、分散は 11.7861 であり、その比は、

$$\text{分散/平均}=11.7861/3.1765=3.71$$

3.71 と 1.0 よりもかなり大きく、過分散となっている。

また、2 値反応としてロジスティック回帰を適用することも可能であるが、2 項分布を仮定することも躊躇される。損傷率は、ほとんどが 1 パーセント以下で、延べ運行月数 163,574 月に対して延べ損傷数 356 件と、平均損傷率は 0.22 パーセントである。過分散を念頭におき、運行月数の対数をオフセットとしたポアソン回帰の適用が望ましいと思われる。



図 10.1 整数化した千月比の損傷数に対するポアソン分布のあてはめ

損傷数を  $y_i$  とし、デザイン行列の変数を  $(x_0, x_1, \dots, x_m)$  としたときに、通常の重回帰のモデル式は、

$$y_i = \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon_i \quad \varepsilon_i \sim \text{正規分布} \quad (10.1)$$

である。運行月数を  $n_i$  としたときに、対数リンクのポアソン回帰のモデル式は、

$$y_i = n_i \exp(\beta_0 x_0 + \beta_1 x_1 + \dots + \beta_m x_m) + \varepsilon_i \quad \varepsilon_i \sim \text{ポアソン分布} \quad (10.2)$$

となる。線形モデルにするために、両辺を  $n_i$  で除し、対数を取る。この際に誤差  $\varepsilon_i$  が和の形式で残り、扱いがややこしくなるので、推定値として、次の線形式

$$\ln\left(\frac{\hat{y}_i}{n_i}\right) = \hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m \quad (10.3)$$

とする。損傷数  $\hat{y}_i$  を運行数  $n_i$  で除し、月あたりの損傷数に対して、対数を取っている。左辺を差  $(\ln \hat{y}_i - \ln n_i)$  にして  $-\ln n_i$  を右辺に移項し、

$$\ln \hat{y}_i = \ln n_i + \hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m \quad (10.4)$$

を得る。ここでの  $\ln n_i$  が、いわゆるオフセット項である。  $\ln \hat{y}_i$  は、実数でカウント・データではないので、式 (10.4) を  $\hat{y}_i$  について解いた式

$$\hat{y}_i = n_i \cdot \exp(\hat{\beta}_0 x_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m) \quad (10.5)$$

にし、  $y_i$  に対する母数を  $\hat{y}_i$  としたポアソン分布の確率  $P_i$  を求め、対数尤度を  $\ln L_i$

$$\ln L_i = \ln P_i = \ln[\text{Poisson}(y_i; \hat{y}_i)] \quad (10.6)$$

を計算する。それらの和を取った対数尤度  $\ln L$

$$\ln L = \sum_{i=1}^n \ln L_i \quad (10.7)$$

を最大にするような  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$  を求める. 対数での式 (10.4) で  $(x_0=1, x_1=0, \dots, x_m=0)$  とすると  $\ln \hat{\beta}_0$  は, いわゆる切片であり, 式

$$\begin{aligned}\ln \hat{y}_i &= \ln n_i + (\hat{\beta}_0 \times 1 + \hat{\beta}_1 \times 0 + \dots + \hat{\beta}_m \times 0) \\ \ln \hat{y}_i &= \ln n_i + \hat{\beta}_0\end{aligned}\tag{10.8}$$

における  $\ln n_i$  は,  $\ln \hat{y}_i$  を推定する際に,  $\hat{\beta}_0$  に  $\ln n_i$  を加えた, あるいは, 基準となる切片を  $\hat{\beta}_0$  とし,  $\ln n_i$  ずらした (オフセットした) 切片を用いることになる.

第 2.6 節では, オーストラリアのある地方の冠動脈心疾患の死亡者数に対して, 対数目盛上の基準となるポアソン回帰直線の切片をオフセットしたグラフを例示したので, 参照してもらいたい.

推定値  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_m)$  を求めるための数値計算の方法は, 2 通りが定式化されている. 第 1 の方法は, 第 2.6 節で示した方法, すなわち式 (2.7) の対数尤度  $\ln L$  をパラメータで 2 階の偏微分行列 (ヘッセ行列, マイナスを付けて情報行列) を用いるニュートン・ラフソン法である. Excel のソルバーを用いることにより, 2 階の偏微分行列を用いずに対数尤度  $\ln L$  を最大化してパラメータの推定が行なえ, 第 10.3 節で例示する.

第 2 の方法は, 第 5.3 節で示した反復重み付き回帰を用いる方法である. この方法は, 式 (5.3) を用いる. 誤差  $\varepsilon_i$  がポアソン分布に従うとの仮定は, 反復計算のための回帰式と重みの計算式の中に組み込まれている. この方法は, 第 1 の方法に比べ技巧的であるが, 計算量が少ない利点がある.

なお, 第 2 の方法による反復重み付き回帰による方法が, 一般化線形モデルの由来である. ただし, モデル式を線形化できない打ち切りデータを含む寿命データのワイブル回帰などは, 線形化できないので, 第 1 の方法による方法で定式化されている. また, 第 7.3 節で扱ったゼロ過剰ポアソン回帰なども一般化線形モデルとはならないので, 第 1 の方法によっている.

## 10.2. 主効果モデルの適用

### (0, 1) 型デザイン変数（最初の水準を基準）

McCullagh ら(1989) では、運行月数の対数をオフセット、誤差分布をポアソン分布、リンク関数を対数としている。質的変数を表 10.2 に示すように最初の水準を基準とする (0, 1) 型のデザイン変数（ダミー変数）にし、過分散を考慮したポアソン回帰の結果が示されている。

表 10.2 質的変数の最初の水準を基準としたデザイン行列のための変数

船種	$x_B$	$x_C$	$x_D$	$x_E$	建造年	$x_{C65}$	$x_{C70}$	$x_{C75}$	運行年	$x_{Op75}$
A	0	0	0	0	60-64	0	0	0	60-74	0
B	1	0	0	0	65-69	1	0	0	75-79	1
C	0	1	0	0	70-74	0	1	0		
D	0	0	1	0	75-79	0	0	1		
E	0	0	0	1						

デザイン行列を用いたポアソン回帰のためのデータリストを表 10.3 に示す。運行月数の対数をオフセットとし、主効果モデル（船種、建造年、運行年）に対するデザイン変数（ $x_0, x_B, x_C, x_D, x_E, x_{C65}, x_{C70}, x_{C75}, x_{Op75}$ ）が示されている。

質的変数を含むポアソン回帰を統計ソフトで行う場合に、統計ソフトの内部で生成される量的変数について注意を払う必要がある。これは、統計ソフトが出力する結果の解釈に際して必要不可欠なためである。McCullagh らで使用されている統計ソフトは、彼らが推奨している（S, GLIM, Minitab）と思われる。これらの統計ソフトは、質的変数に対し、最初の水準を基準とした量的変数（ダミー変数）がデフォルトで使用されているのであろう。S の後の R も同様である。最初の水準を基準とした場合のデザイン変数は、推定されたパラメータが、「最初の水準（対照群）」との差となり、 $t$  検定あるいはカイ 2 乗検定の結果が、対照群とそれぞれの群の差についての検定統計量として使える利便性がある。

量的変数しか受け付けないポアソン回帰に対し、質的変数の各水準に対し、何らかの量的変数のセットを定義した対応が求められる。この量的変数が、通常ダミー変数と称されている。本書では、ポアソン回帰を行うために必要なデザイン行列（計画行列）に必要な変数なので、ダミー変数ではなくデザイン変数として統一的に使用してきた。連続変数と質的変数との交互作用は、(0, 1) ではなく、連続変数となりダミー変数と言い難いことも理由の一つである。また、デザイン行列のためのダミー変数と言うのも違和感があり、切片も含めて全てデザイン変数で統一してきた。

表 10.3 船舶の前方部への損傷数データに対するデザイン行列

No	船種	建造年度	運行年度	運行月数 $n_i$	損傷数 $y_i$	デザイン行列								
						$x_0$	$x_B$	$x_C$	$x_D$	$x_E$	$x_{C65}$	$x_{C70}$	$x_{C75}$	$x_{Op75}$
1	A	60-64	60-74	127	0	1	0	0	0	0	0	0	0	0
2			75-79	63	0	1	0	0	0	0	0	0	0	1
3		65-69	60-74	1,095	3	1	0	0	0	0	1	0	0	0
4			75-79	1,095	4	1	0	0	0	0	1	0	0	1
5		70-74	60-74	1,512	6	1	0	0	0	0	0	1	0	0
6			75-79	3,353	18	1	0	0	0	0	0	1	0	1
7		75-79	60-74	-	- *	1	0	0	0	0	0	0	1	0
8			75-79	2,244	11	1	0	0	0	0	0	0	1	1
9	B	60-64	60-74	44,882	39	1	1	0	0	0	0	0	0	0
10			75-79	17,176	29	1	1	0	0	0	0	0	0	1
11		65-69	60-74	28,609	58	1	1	0	0	0	1	0	0	0
12			75-79	20,370	53	1	1	0	0	0	1	0	0	1
13		70-74	60-74	7,064	12	1	1	0	0	0	0	1	0	0
14			75-79	13,099	44	1	1	0	0	0	0	1	0	1
15		75-79	60-74	-	- *	1	1	0	0	0	0	0	1	0
16			75-79	7,117	18	1	1	0	0	0	0	0	1	1
17	C	60-64	60-74	1,179	1	1	0	1	0	0	0	0	0	0
18			75-79	552	1	1	0	1	0	0	0	0	0	1
19		65-69	60-74	781	0	1	0	1	0	0	1	0	0	0
20			75-79	676	1	1	0	1	0	0	1	0	0	1
21		70-74	60-74	783	6	1	0	1	0	0	0	1	0	0
22			75-79	1,948	2	1	0	1	0	0	0	1	0	1
23		75-79	60-74	-	- *	1	0	1	0	0	0	0	1	0
24			75-79	274	1	1	0	1	0	0	0	0	1	1
25	D	60-64	60-74	251	0	1	0	0	1	0	0	0	0	0
26			75-79	105	0	1	0	0	1	0	0	0	0	1
27		65-69	60-74	288	0	1	0	0	1	0	1	0	0	0
28			75-79	192	0	1	0	0	1	0	1	0	0	1
29		70-74	60-74	349	2	1	0	0	1	0	0	1	0	0
30			75-79	1,208	11	1	0	0	1	0	0	1	0	1
31		75-79	60-74	-	- *	1	0	0	1	0	0	0	1	0
32			75-79	2,051	4	1	0	0	1	0	0	0	1	1
33	E	60-64	60-74	45	0	1	0	0	0	1	0	0	0	0
34			75-79	-	- **	1	0	0	0	1	0	0	0	1
35		65-69	60-74	789	7	1	0	0	0	1	1	0	0	0
36			75-79	437	7	1	0	0	0	1	1	0	0	1
37		70-74	60-74	1,157	5	1	0	0	0	1	0	1	0	0
38			75-79	2,161	12	1	0	0	0	1	0	1	0	1
39		75-79	60-74	-	- *	1	0	0	0	1	0	0	1	0
40			75-79	542	1	1	0	0	0	1	0	0	1	1
計				163,574	356									

統計ソフト SAS では、質的変数に対してデフォルトの設定では、最後の水準を基準とした (1, 0) 型のデザイン変数が生成され、JMP では、対比 (1, -1) 型のデザイン変数が生成される。ただし、SAS では、内部で生成されるデザイン変数を意識しなくても、元の水準に対する推定値を最小 2 乗平均 (Lsmeans) としてたやすく推定できる。



デザイン変数を自らデータとして生成することは、表計算ソフト Excel が普及する以前には苦痛であったが、表 10.3 に示すように Excel のコピー&ペーストで比較的容易に (0, 1) から成るデザイン行列が作成できるようになった。このように自ら能動的に作成したデザイン行列を用い、統計ソフトによる解析をすることが、ブラックボックス的となりがちな統計ソフトの解析結果に対し理解を深めることが期待される。さらに、このデザイン行列を用い、Excel のみで統計ソフトと同様の解析を試みることは、さらなる理解を深めることになる。

表 10.3 のデザイン行列を用いた JMP によるポアソン回帰の結果を示す。適合度統計量の Pearson のカイ 2 乗統計量は 42.2753 で自由度が 25 なので、その比は 1.6910 と過分散となっている。パラメータの推定値は、運行月数をオフセットにしたので、月あたりの損傷数の対数となっている。

表 10.4 損傷数に対するポアソン回帰（最初を基準，過分散あり）

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	31.8251	63.6503	8	<.0001*
完全	40.3787			
縮小	72.2038			
適合度統計量	カイ2乗	自由度	p値	過分散
Pearson	42.2753	25	0.0168*	1.6910
デビアンس	38.6951	25	0.0395*	
AICc				
110.3226				
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	-6.4059	0.2828	958.7224	<.0001*
x_B	-0.5433	0.2309	5.0023	0.0253*
x_C	-0.6874	0.4279	2.8890	0.0892
x_D	-0.0760	0.3779	0.0408	0.8399
x_E	0.3256	0.3067	1.1062	0.2929
x_C65	0.6971	0.1946	13.4958	0.0002*
x_C70	0.8184	0.2208	13.9492	0.0002*
x_C75	0.4534	0.3032	2.1554	0.1421
x_Op75	0.3845	0.1538	6.3040	0.0120*

表 10.4 のパラメータの推定値は、月あたりの対数の損傷数なので解釈が困難である。表 10.5 に示すように基準となる最初の水準の（船種：A，建造年：60-64，運行年：60-74）の推定値は、切片の推定値  $\hat{\beta}_0 = -6.4059$  に等しくなる。他の水準は、最初の水準との差の推定値であり、

最初の水準の推定値を加えることにより，推定値が得られる．

$$\text{船種 B: 船種 A} + (\text{船種 B} - \text{船種 A}) = -6.4059 - 0.5434 = -6.9493$$

$$\text{船種 C: 船種 A} + (\text{船種 C} - \text{船種 A}) = -6.4059 - 0.6874 = -7.0933$$

：

$$\text{建造年 65-79: 建造年 65-79} + (\text{建造年 60-64} - \text{建造年 65-79}) = -6.4059 + 0.6971 = -5.7088$$

：

$$\text{運行年 75-79: 運行年 60-74} + (\text{運行年 75-79} - \text{運行年 60-74}) = -6.4059 + 0.3845 = -6.0214$$

元の水準での月当たりの対数の損傷数に，指数を取り月当たりの損傷数を計算する．少数点以下 2 桁がゼロなので 1,000 を掛けて損傷千月比の推定値を計算する．

表 10.5 パラメータの推定値に対する解釈

No	因子	パラメータ		JMP の結果 (表10.4)				Excelでの換算			
				推定値 $\beta$	標準 誤差	尤度比 カイ2乗	p 値	月あたり ln損傷数	月あたり 損傷数	損傷 千月比	
0		切片	$x_0$	-6.4059	0.2828	958.72	<.0001	-6.4059	0.0017	1.6518	
	船種	(A)	-	0				-6.4059	0.0017	1.6518	基準
1		B	$x_B$	-0.5433	0.2309	5.0023	0.0253	-6.9492	0.0010	0.9594	
2		C	$x_C$	-0.6874	0.4279	2.8890	0.0892	-7.0933	0.0008	0.8306	
3		D	$x_D$	-0.0760	0.3779	0.0408	0.8399	-6.4819	0.0015	1.5310	
4		E	$x_E$	0.3256	0.3067	1.1062	0.2929	-6.0803	0.0023	2.2874	
	建造	(60-64)	-	0				-6.4059	0.0017	1.6518	基準
5	年	65-69	$x_{C65}$	0.6971	0.1946	13.4958	0.0002	-5.7088	0.0033	3.3168	
6		70-74	$x_{C70}$	0.8184	0.2208	13.9492	0.0002	-5.5875	0.0037	3.7445	
7		75-79	$x_{C75}$	0.4534	0.3032	2.1554	0.1421	-5.9525	0.0026	2.5994	
	運行	(60-74)	-	0				-6.4059	0.0017	1.6518	基準
8	年	75-79	$x_{Op75}$	0.3845	0.1538	6.3040	0.0120	-6.0214	0.0024	2.4262	
オフセット: ln(運行月数),				Pearson 適度度 カイ2乗=42.23, df=25,				過分散 42.2753/25 = 1.6910			

船種 A の 1,000 月あたりの損傷数は，切片の対数の推定値が -6.4059 で，指数を取った 0.00165 件が月あたりの損傷数となり，

$$\begin{aligned}\hat{y}_{A,C60,Op60}^{(1000)} &= \exp(\hat{\beta}_0) \times 1000 \\ &= \exp(-6.4059) \times 1000 = 1.6518\end{aligned}$$

1,000 月あたりに換算すると 1.6518 件と推定される．

船種 B は，切片と  $x_B$  以外が全て 0 なので，

$$\begin{aligned}\hat{y}_{B,C60,Op60}^{(1000)} &= \exp(\hat{\beta}_0 + x_B \hat{\beta}_1) \times 1000 \\ &= \exp(-6.4059 - 1 \times 0.5434) \times 1000 = 0.9594\end{aligned}$$

と推定される．

建造年 65-69 の推定値は，デザイン変数の基準を最初の水準としているので，建造年 60-64 との差であり，船種 A，運行年 60-74 と固定した場合に，

$$\hat{y}_{A,C65,Op60}^{(1000)} = \exp(-6.4059 + 0.6971) \times 1000 = 3.3168$$

と推定される。

運行年 75-79 の推定値は、運行年 60-74 との差で、船種 A、運行年 60-74 とした場合に、

$$\hat{y}_{A,C60,Op75}^{(1000)} = \exp(-6.4059 + 0.3845) \times 1000 = 2.4262$$

と推定される。

これらの結果から、船種別の 1,000 月あたりの損傷数は、船種 C が 0.8306 件と最も少なく、船種 E が 2.2874 件と最も多いと推定される。建造年の場合は、60-64 年が 1.6518 件と最も少なく、70-74 年が 3.7445 件と最も多いことが分かる。運行年では、75-79 年の 2.4262 件が、60-74 年の 1.6518 件に比べて多くなっている。

多くの統計ソフトでは、質的変数を説明変数とした場合に水準数に応じたダミー変数を内部で自動生成し、それらを用いて何らかのデザイン行列を用いた計算が行なわれる。どのようなダミー変数を内部で生成するかは、開発者の考え方によって異なる。

SAS/GENMOD プロシジャでのデフォルトは、質的変数に対し最後の水準を基準とする (1, 0) 型のダミー変数であり、最初的水準に変更したい場合は、`classs` ステートメントのオプションで `ref=first` とすれば (0, 1) 型となる。ただし、デフォルト以外のダミー変数を生成すると、最小 2 乗平均 (Lsmeans) などの計算がサポートされなくなるデメリットがある。

JMP の一般化線形モデルでは、対比 (1, -1) 型のダミー変数がデフォルトであり、設定の変更はできない。そのため、最初的水準を基準にしたダミー変数にしたければ、これまでに示したように、自ら (0, 1) 型のダミー変数を生成しなければならない。その結果として、JMP が標準的に提供している予測プロファイルなどのグラフ表示ができなくなる。

### (1, -1) 対比型のデザイン行列

表 10.4 で示した結果は、質的変数に対し (0, 1) 型のデザイン変数を自ら生成して JMP のポアソン回帰を適用した結果であり、(0, 1) 型のデザイン変数のままでは、JMP の予測プロファイルがうまく機能しない。そこで、質的変数に対しては JMP の内部でデザイン変数を生成させることにより、予測プロファイルおよび交互作用プロファイルにおいて各因子ごとに取りまとめられて表示される。表 10.6 に (1, -1) の対比型のデザイン行列のための変数を示す。なお、対比型とは、1 の反対の性質を持つ数値として -1 が使われているためであり。したがって、対比型の変数は、足して 0 となる性質を持っている。

表 10.6 質的変数の (1, -1) 対比型のデザイン変数 (ダミー変数)

船種	$x'_A$	$x'_B$	$x'_C$	$x'_D$	建造年	$x'_{C60}$	$x'_{C65}$	$x'_{C70}$	運行年	$x'_{Op60}$
A	1	0	0	0	60-64	1	0	0	60-74	1
B	0	1	0	0	65-69	0	1	0	75-79	-1
C	0	0	1	0	70-74	0	0	1		
D	0	0	0	1	75-79	-1	-1	-1		
E	-1	-1	-1	-1						

JMP による主効果モデルは、表 10.7 に示すように因子ごとの尤度比カイ 2 乗検定に引き続き、(1, -1) 型のデザイン変数に対応する推定値が出力される。表 10.4 と比較すると「モデル全体の検定」は、差分の尤度比カイ 2 乗値が 63.6503 と同じ結果となっている。次の「効果の検定」は、質的変数ごとの尤度比カイ 2 乗検定で、要因配置実験で一般的な分散分析表と同様な形式となっている。

「パラメータ推定値」は、全く異なる。表 10.4 での切片 -6.4059 は、切片以外のデザイン変数が全て 0 の場合であり、船種が A、建造年が 1960-64、運行年が 1960-74 の場合の推定値となっている。表 10.7 の切片は、 $\hat{\beta}'_0 = -5.9176$  であり、(1, -1) 対比型のデザイン行列の場合には、(船種: A, 建造年: 60-64, 運行年: 60-74) のデザイン変数は ( $x'_A = 1, x'_B = 0, x'_C = 0, x'_D = 0, x'_{C60} = 1, x'_{C65} = 0, x'_{C75} = 0, x'_{Op60} = 1$ ) なので、表 10.4 の場合と同様の推定は、

$$\begin{aligned}\ln \hat{y}_{A,C60,Op60} &= \hat{\beta}'_0 + x'_A \hat{\beta}'_1 + x'_{C60} \hat{\beta}'_5 + x'_{Op60} \hat{\beta}'_8 \\ &= 5.9176 + 0.1962 - 0.4922 - 0.1922 \\ &= -6.4059\end{aligned}$$

とすることにより同じ結果が得られる。このように、推定値は名義尺度の各水準に与えるデザイン変数によって全く異なるので、パラメータの推定値を用いた結果の解釈には細心の注意が必要である。

表 10.8 に対比型のデザイン変数を使った場合のパラメータの推定値を用い、船種、建造年、および運行年の各水準の推定値 (月あたりの対数損傷数) を計算した結果を示す。切片 -5.9176 は、それぞれの質的変数の損傷数の対数の推定値の平均であり、1,000 月あたりの損傷数は、

$$\text{切片: } \exp(-5.9176) \times 1000 = 2.6915 \text{ 件}$$

となる。船種 E の推定値は、切片に船種 A～D の負の和を加え

$$\text{船種 E} = \text{切片} - \text{船種 A} - \text{船種 B} - \text{船種 C} - \text{船種 D} - \text{船種 E}$$

$$\begin{aligned}\ln \hat{y}_{E,..} &= \hat{\beta}'_0 - x'_A \hat{\beta}'_1 - x'_B \hat{\beta}'_2 - x'_C \hat{\beta}'_3 - x'_D \hat{\beta}'_4 \\ &= -5.9176 - 0.1962 - (-0.3471) - (-0.4912) - 0.1203 \\ &= -5.9176 + 0.5218 \\ &= -5.3958\end{aligned}$$

表 10.7 損傷数に対するポアソン回帰（対比型，過分散）

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	31.8251	63.6503	8	<.0001*
完全	40.3787			
縮小	72.2038			
適合度統計量	カイ2乗	自由度	p値(Prob>ChiSq)	過分散
Pearson	42.2753	25	0.0168*	1.6910
デビアンズ	38.6951	25	0.0395*	
AICc				
110.3226				

効果の検定			
要因	自由度	尤度比カイ2乗	p値
船種	4	13.9977	0.0073*
建造年	3	18.5735	0.0003*
運行年	1	6.3040	0.0120*

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	-5.9176	0.1280	57532.562	<.0001*
船種[A]	0.1962	0.1954	0.9809	0.3220
船種[B]	-0.3471	0.1459	5.2948	0.0214*
船種[C]	-0.4912	0.3148	2.8802	0.0897
船種[D]	0.1203	0.2782	0.1808	0.6707
建造年[60-64]	-0.4922	0.1505	11.2088	0.0008*
建造年[65-69]	0.2049	0.1181	2.9823	0.0842
建造年[70-74]	0.3262	0.1228	6.8496	0.0089*
運行年[60-74]	-0.1922	0.0769	6.3040	0.0120*

と推定される。損傷千月比は指数を取り 1,000 倍した結果である。建造年 および 運行年の場合も同様にして推定する。損傷数は、表 10.4 と全く異なるが、それぞれの水準の大小関係は保たれている。

月当たりの損傷数の対数について、対比型のデザイン変数なので船種 A～E の平均-5.7176 であり、切片と等しくなる。また、当然のことながら損傷千月比の幾何平均は、切片の損傷千月比と等しくなる。

このように、統計ソフトから出力されたポアソン回帰の結果は、Excel に取り込み何らかの計算を加えて結果の解釈をする必要が常にある。

表 10.8 対比型デザイン変数による主効果モデル

変数番号		パラメータ	JMP推定値	月あたりln損傷数		月あたり損傷数	損傷千月比	
0		切片 $x_0$	<b>-5.9176</b>	<b>-5.9176</b>		0.0027	<b>2.6915</b>	
1	船種	A $x'_A$	0.1962	-5.7214		0.0033	3.2751	幾何
2		B $x'_B$	-0.3471	-6.2648	平均	0.0019	1.9022	平均
3		C $x'_C$	-0.4912	-6.4088	<b>-5.9176</b>	0.0016	1.6470	<b>2.6915</b>
4		D $x'_D$	0.1203	-5.7974		0.0030	3.0355	
		(E) 負の和	<b>0.5218</b>	-5.3958		0.0045	4.5354	
5	建造年	60-64 $x'_{C60}$	-0.4922	-6.4099		0.0016	1.6452	幾何
6		65-69 $x'_{C65}$	0.2049	-5.7128	平均	0.0033	3.3036	平均
7		70-74 $x'_{C70}$	0.3262	-5.5915	<b>-5.9176</b>	0.0037	3.7296	<b>2.6915</b>
		(75-79) 負の和	<b>-0.0388</b>	-5.9565		0.0026	2.5890	
8	運行年	60-74 $x'_{Op60}$	-0.1922	-6.1099	平均	0.0022	2.2208	幾何平均
		(75-79) 負の和	<b>0.1922</b>	-5.7254	<b>-5.9176</b>	0.0033	3.2620	<b>2.6915</b>

### 予測プロファイル

JMP のデフォルトの対比 (1, -1) 型のダミー変数でポアソン回帰を行った場合には、主効果に関する予測プロファイルおよび交互作用プロファイルが出力される。まず、予測プロファイルの結果を図 10.2 に示す。この図は、船種を A、建造年 60-64、運行年 60-74 を基準とし

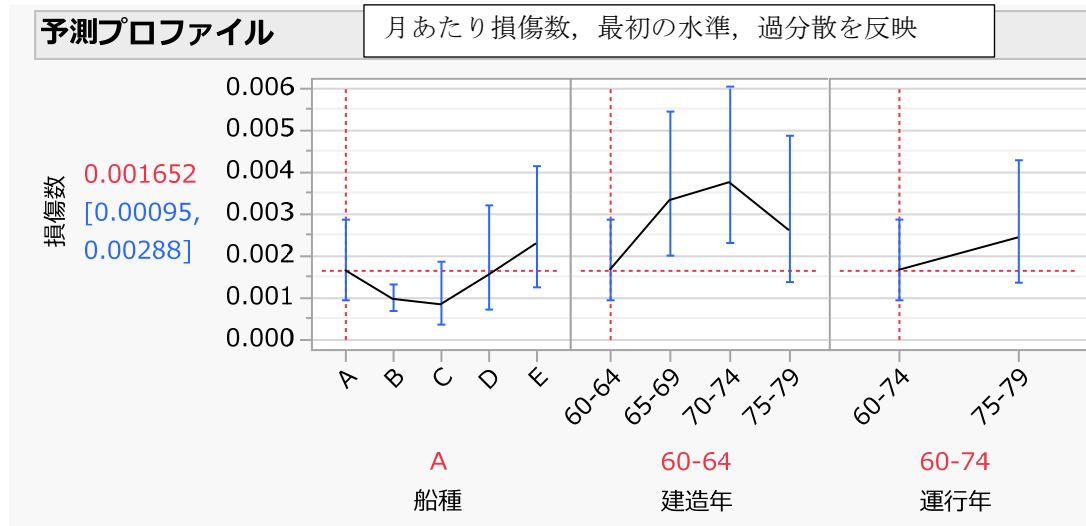


図 10.2 JMP の予測プロファイル（船種を A、建造年 60-64、運行年 60-74）を基準

た場合が示されている。損傷数の推定値は 0.001652、95%信頼区間が (0.00095, 0.00288) と推定されている。これは、推定された月あたりの損傷数の対数に対し、指数を取ったもので、 $\exp(-6.4059)=0.001652$  であり、1,000 月あたりでは 1.6552 件となる。なお、図 10.2 の予測プロファイルでの船種 (B, C, D, E) の推定値は、表 10.5 に示すように (船種 A、建造年 60-

64, 運行年 60-74) を基準とした推定値 (0.0010, 0.0008, 0.0015, 0.0023) が表示されている。同様に建造年 (65-69, 70-74, 75-79) の推定値は (0.0033, 0.0037, 0.0026) が、運行年 (75-79) の推定値は (0.0024) である。

JMP で出力されている予測プロファイル上で、マウス操作により表示内容を変更することができる。それぞれの変数の最大となっている水準 (船種 E, 建造年 70-74, 運行年 75-79) にマウスを移動し選択すると図 10.3 が得られる。推定値として 0.007617, 1,000 月あたり 7.62 件が得られている。また、最小となる水準の組み合わせとなる推定値を得ることも容易にできる。

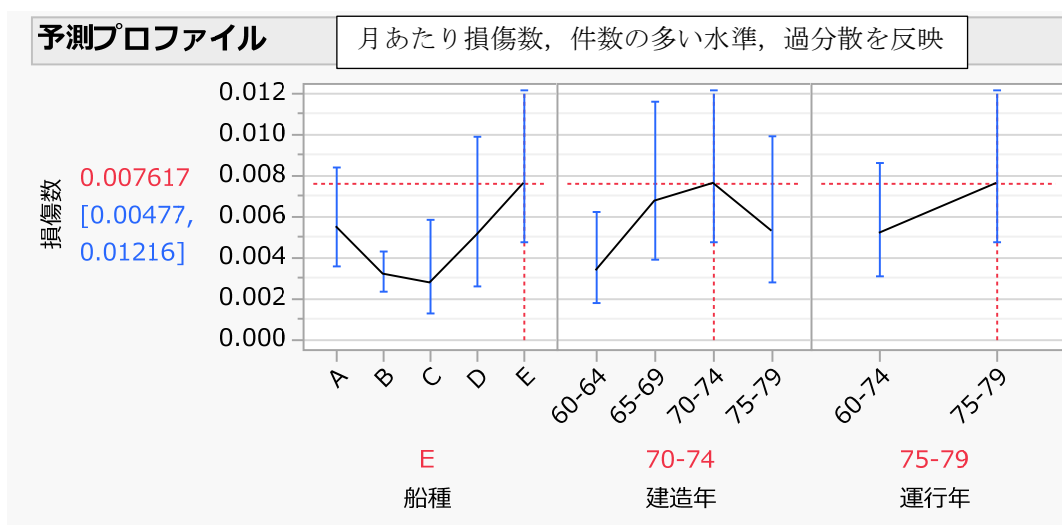


図 10.3 JMP の予測プロファイル (船種 E, 建造年 70-74, 運行年 75-79) を基準

図 10.3 の予測プロファイルでの船種 (A, B, C, D) の推定値は, (船種 E, 建造年 70-74, 運行年 75-79) に固定した場合のそれぞれの船種の推定値 0.007617 となっていて, 図 10.2 とは異なる。このように主効果の推定値は, 固定されるものではなく, 他の条件により変化する相対的なものであることに注意が必要である。

### 交互作用プロファイル

予測プロファイルに引き続き, 交互作用プロファイルを作成する。交互作用を含まない主効果モデルなので, “交互作用” はないが, 図 10.4 に示すように 3 つの質的を組み合わせた推定値のプロファイルが得られる。全体を俯瞰するためには有益であり, 主効果に対する予測プロファイルと合わせて使うと効果的である。

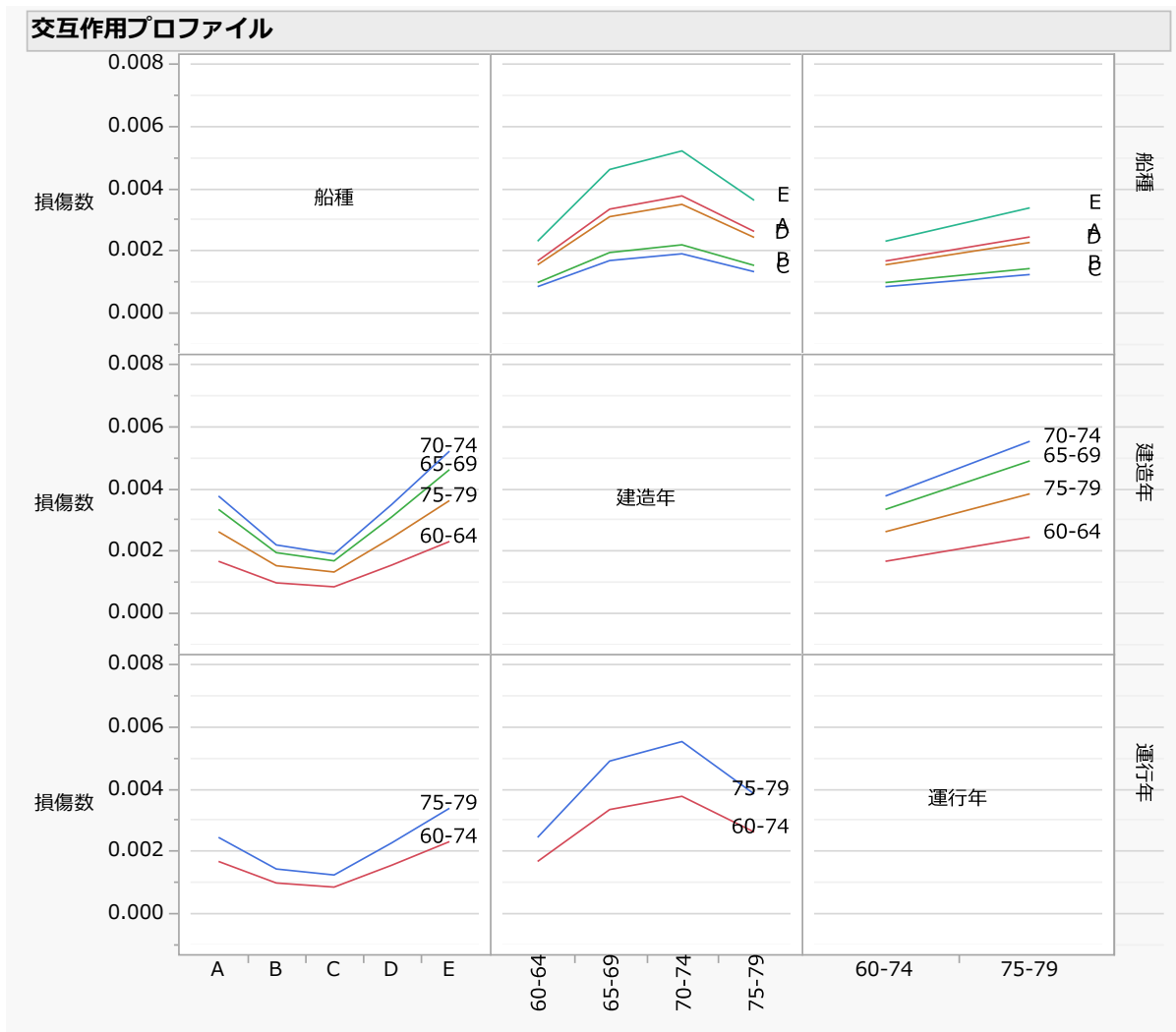


図 10.4 交互作用プロファイル（船種を A，建造年 60-64，運行年 60-74）を基準



### 10.3. Excel による予測プロファイル

JMP の予測プロファイルは、結果を解釈するために有益であるが、JMP 以外では見かけないことがないので、Excel による作図法を示す。95%信頼区間も含めるためには、パラメータの共分散行列を用いた計算が必要となるのであるが、まず、推定値のみの折れ線グラフを示す。

表 10.5 から因子と水準、損傷千月比を抜き出して作図用の Excel データとする。3 因子の水準と推定値を選択し、「折れ線」グラフを選択する。2 つの因子をまたいで繋がっている線を因子の切れ目のポイントのみを選択し「線なし」とする。船種は質的変数なので点線に変更した結果を図 10.5 に示す。Excel のグラフなので推定値を変えれば、自動的に更新される。

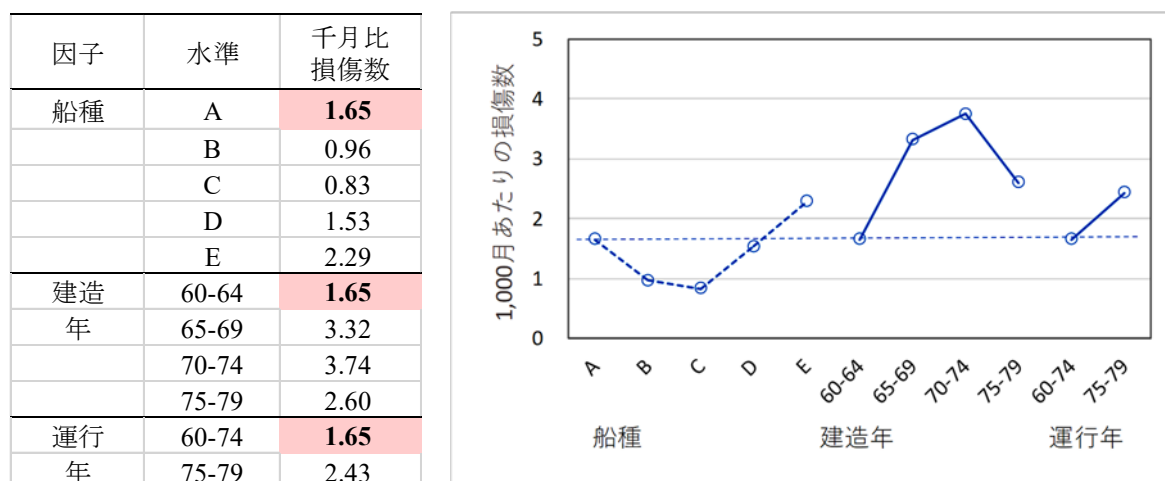


図 10.5 Excel による予測プロファイル（船種を A、建造年 60-64、運行年 60-74）を基準

図 10.5 の予測プロファイルに 95%信頼区間の重ね書きを試みよう。JMP では、自動的に 95%信頼区間を表示しているが、どのような計算を行っているのであろうか。表 10.3 に示したデザイン行列  $\mathbf{X}$  の各行のベクトル  $\mathbf{x}_i$  とパラメータの推定値  $\hat{\boldsymbol{\beta}}$  の積から推定値は、次のように

$$\ln(\hat{y}_i / n_i) = \mathbf{x}_i \hat{\boldsymbol{\beta}}$$

求められる。この  $\ln(\hat{y}_i / n_i)$  の分散  $Var[\ln(\hat{y}_i / n_i)]$  は、パラメータの推定値  $\hat{\boldsymbol{\beta}}$  のパラメータの共分散行列  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$  を使って

$$Var[\ln(\hat{y}_i / n_i)] = \mathbf{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{x}_i^T$$

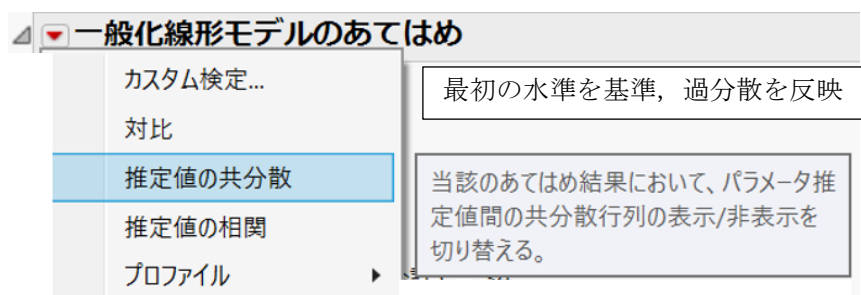
によって推定することができる。パラメータの共分散行列は、JMP のプルダウンメニューで「推定値の共分散」を選択することによって得られる。SAS の GENMOD プロシジャの場合であれば、Model ステートメントの covb オプションで得られる。使っている統計ソフトでパ

ラメータの共分散行列が得られない場合は、第 12 章でデザイン行列および重みベクトルを用いた Excel による計算方法を示すので、参考にしてもらいたい。

デザイン行列の各行の  $\mathbf{x}_i$  と  $\hat{\boldsymbol{\beta}}$  の積から推定値  $\ln(\hat{y}_i / n_i) = \mathbf{x}_i \hat{\boldsymbol{\beta}}$  を求め、その分散をパラメータの共分散行列  $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$  を使って  $\mathbf{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{x}_i^T$  によって推定する方法は、第 4.5 節「デザイン行列を用いた回帰分析の実際」で丁寧に示した。この方法は、複数の変数に対する回帰分析にも、ポアソン回帰にも適用できる基本的な方法である。第 1.4 節「人工データ (恒等リンク, 3 水準)」の「共分散行列を用いた 95%信頼区間の計算」は、恒等リンクのポアソン回帰の場合で活用事例である。

推定値  $\ln(\hat{y}_i / n_i)$  の分散が求まれば、95%信頼区間の推定は、容易である。表 10.9 に示すように共分散行列を JMP から Excel に取り込む。デフォルトの結果の表示では、小数点以下 4 桁程度であるが、8 桁程度に変更し、Excel に取り込んだ後に桁数の表示を縮めることが、計算精度を保つために必要である。

表 10.9 推定値 (パラメータ) の共分散行列の獲得



#### 推定値の共分散

共分散									
	切片	x_B	x_C	x_D	x_E	x_C65	x_C70	x_C75	x_Op75
切片	0.0800	-0.0530	-0.0458	-0.0396	-0.0408	-0.0266	-0.0343	-0.0344	-0.0094
x_B	-0.0530	0.0533	0.0428	0.0390	0.0404	0.0038	0.0138	0.0160	0.0009
x_C	-0.0458	0.0428	0.1831	0.0384	0.0412	0.0030	0.0043	0.0126	-0.0002
x_D	-0.0396	0.0390	0.0384	0.1428	0.0387	0.0020	0.0024	-0.0111	-0.0003
x_E	-0.0408	0.0404	0.0412	0.0387	0.0941	-0.0002	-0.0025	0.0049	0.0013
x_C65	-0.0266	0.0038	0.0030	0.0020	-0.0002	0.0379	0.0272	0.0281	-0.0036
x_C70	-0.0343	0.0138	0.0043	0.0024	-0.0025	0.0272	0.0487	0.0367	-0.0090
x_C75	-0.0344	0.0160	0.0126	-0.0111	0.0049	0.0281	0.0367	0.0919	-0.0147
x_Op75	-0.0094	0.0009	-0.0002	-0.0003	0.0013	-0.0036	-0.0090	-0.0147	0.0237

ここでのデザイン行列のベクトル  $\mathbf{x}_i$  は 1 行 9 列, パラメータの共分散行列  $\Sigma(\hat{\beta})$  は 9 行 9 列, 転置した  $\mathbf{x}_i^T$  は 9 行 1 列であり, 共分散行列の要素を  $C_{j,k}$  としたときに,  $Var[\ln(\hat{y}_i / n_i)]$  は,

$$\begin{aligned} Var[\ln(\hat{y}_i / n_i)] &= \mathbf{x}_i \Sigma(\hat{\beta}) \mathbf{x}_i^T \\ &= \begin{bmatrix} x_{i,0} & x_{i,1} & \cdots & x_{i,8} \end{bmatrix} \begin{bmatrix} C_{0,0} & C_{0,1} & \cdots & C_{0,8} \\ C_{1,0} & C_{1,1} & \cdots & C_{1,8} \\ \vdots & \vdots & \ddots & \vdots \\ C_{8,0} & C_{8,1} & \cdots & C_{8,8} \end{bmatrix} \begin{bmatrix} x_{i,0} \\ x_{i,1} \\ \vdots \\ x_{i,8} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{j=0}^8 x_{i,j} C_{j,0} & \sum_{j=0}^8 x_{i,j} C_{j,1} & \cdots & \sum_{j=0}^8 x_{i,j} C_{j,8} \end{bmatrix} \begin{bmatrix} x_{i,0} \\ x_{i,1} \\ \vdots \\ x_{i,8} \end{bmatrix} \\ &= \sum_{j=0}^8 x_{i,j} C_{j,0} x_{i,0} + \sum_{j=0}^8 x_{i,j} C_{j,1} x_{i,1} + \cdots + \sum_{j=0}^8 x_{i,j} C_{j,8} x_{i,8} \\ &= \sum_{j=0}^8 \sum_{k=0}^8 x_{i,j} C_{j,k} x_{i,k} \end{aligned}$$

で与えられる. このように, 共分散行列の全ての要素  $C_{j,k}$  に対して, その添え字に対応する  $x_{i,j}$  と  $x_{i,k}$  を掛けて加えたものになる. この式は,  $x_{i,j}$  についての 2 次の項の和の形式となっているので, 「2 次形式」と言われている. Excel での行列計算が, 実際にはどのようなものなのか, シグマを用いた計算式と対比することによって理解を深めてもらいたい.

表 10.10 の上側に (0, 1) 型のデザイン行列から得られたポアソン回帰パラメータの推定値  $\hat{\beta}^T$  および共分散行列  $\Sigma(\hat{\beta})$  を示す. 表 10.10 の下側には, 船種 A, 建造年 60-64, 運行年 60-74 を基準とした場合の各因子の水準の 1,000 月当たりの損傷数の推定値と 95%信頼区間を計算した結果である.

大変な計算のように思われるかもしれないが, 第 4.5 節では  $2 \times 2$  のパラメータの共分散行列に対して示した Excel の行列計算の計算式と, 全く同じ形式であり, 推定値  $\ln(\hat{y}_i / n_i)$  の分散が

$$\begin{aligned} Var[(\ln(\hat{y}_i / n_i))] &= \mathbf{x}_i \Sigma(\hat{\beta}) \mathbf{x}_i^T \\ &= \text{Mmult}(\text{Mmult}(\mathbf{x}_i \text{の範囲}, \Sigma(\hat{\beta}) \text{の範囲}), \text{Transpose}(\mathbf{x}_i \text{の範囲})) \end{aligned}$$

にて求まれば, 95%信頼区間の推定値の計算は, 選択の範囲を  $2 \times 2$  から  $9 \times 9$  に代えるだけで  $2 \times 2$  の場合と同様に容易に計算ができる.

95 信頼区間を求めるためには, 推定値の分散を求めることが必須であるが, 一般的に計算が煩雑になるために例示されてるのは, 1 次式のあてはめ場合のみで,  $3 \times 3$  のパラメータの共分散行列を必要とする 2 次式の 95%信頼区間の例示は, 極めてまれである. 2 次式は, 切

表 10.10 予測プロファイルの 95%信頼区間の計算シート

		----- 船種-----					--- 建造年 --- 運行年				月あたり	対数	---- 千月比 ----		
因子	水準	切片	B	C	D	E	65-69	70-74	75-79	75-79	対数	分散	損傷数	信頼区間	
		$x_0$	$x_B$	$x_C$	$x_D$	$x_E$	$x_{C65}$	$x_{C70}$	$x_{C75}$	$x_{Op75}$	損傷数	Var	$y^{(1000)}$	L95%	U95%
船種	A	1	0	0	0	0	0	0	0	0	-6.4059	0.0800	1.6518	0.9490	2.8750
	B	1	1	0	0	0	0	0	0	0	-6.9493	0.0273	0.9593	0.6939	1.3263
	C	1	0	1	0	0	0	0	0	0	-7.0933	0.1715	0.8306	0.3689	1.8702
	D	1	0	0	1	0	0	0	0	0	-6.4819	0.1435	1.5309	0.7285	3.2170
	E	1	0	0	0	1	0	0	0	0	-6.0803	0.0925	2.2874	1.2601	4.1522
建造年	60-64	1	0	0	0	0	0	0	0	0	-6.4059	0.0800	1.6518	0.9490	2.8750
	65-69	1	0	0	0	0	1	0	0	0	-5.7088	0.0645	3.3168	2.0160	5.4569
	70-74	1	0	0	0	0	0	1	0	0	-5.5875	0.0600	3.7444	2.3167	6.0521
	75-79	1	0	0	0	0	0	0	1	0	-5.9525	0.1030	2.5994	1.3856	4.8765
運行年	60-74	1	0	0	0	0	0	0	0	0	-6.4059	0.0800	1.6518	0.9490	2.8750
	75-79	1	0	0	0	0	0	0	0	1	-6.0214	0.0848	2.4262	1.3710	4.2936

		----- 船種-----					--- 建造年 ---			運行年
		切片	B	C	D	E	65-69	70-74	75-79	75-79
		$x_0$	$x_B$	$x_C$	$x_D$	$x_E$	$x_{C65}$	$x_{C70}$	$x_{C75}$	$x_{Op75}$
		-6.406	-0.543	-0.687	-0.076	0.326	0.697	0.818	0.453	0.384
共分散		$x_0$	$x_B$	$x_C$	$x_D$	$x_E$	$x_{C65}$	$x_{C70}$	$x_{C75}$	$x_{Op75}$
	$x_0$	0.0800	-0.0530	-0.0458	-0.0396	-0.0408	-0.0266	-0.0343	-0.0344	-0.0094
	$x_B$	-0.0530	0.0533	0.0428	0.0390	0.0404	0.0038	0.0138	0.0160	0.0009
	$x_C$	-0.0458	0.0428	0.1831	0.0384	0.0412	0.0030	0.0043	0.0126	-0.0002
	$x_D$	-0.0396	0.0390	0.0384	0.1428	0.0387	0.0020	0.0024	-0.0111	-0.0003
	$x_E$	-0.0408	0.0404	0.0412	0.0387	0.0941	-0.0002	-0.0025	0.0049	0.0013
	$x_{C65}$	-0.0266	0.0038	0.0030	0.0020	-0.0002	0.0379	0.0272	0.0281	-0.0036
	$x_{C70}$	-0.0343	0.0138	0.0043	0.0024	-0.0025	0.0272	0.0487	0.0367	-0.0090
	$x_{C75}$	-0.0344	0.0160	0.0126	-0.0111	0.0049	0.0281	0.0367	0.0919	-0.0147
	$x_{Op75}$	-0.0094	0.0009	-0.0002	-0.0003	0.0013	-0.0036	-0.0090	-0.0147	0.0237

月当たり対数推定値： $\ln(\hat{y}_i / n_i) = \mathbf{x}_i \hat{\boldsymbol{\beta}} = \text{Mmult}(\mathbf{x}_i \text{の範囲}, \text{Transpose}(\hat{\boldsymbol{\beta}}^T \text{の範囲}))$

分散： $\text{Var}(\ln \hat{y}_i / n_i) = \mathbf{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{x}_i^T = \text{Mmult}(\text{Mmult}(\mathbf{x}_i \text{の範囲}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \text{の範囲}), \text{Transpose}(\mathbf{x}_i \text{の範囲}))$

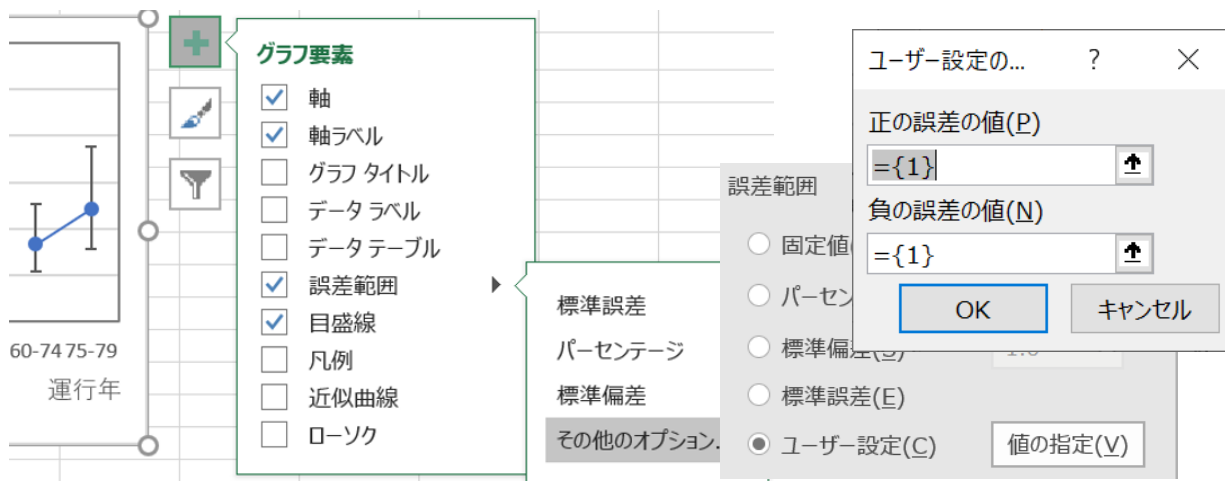
月当たり推定値： $\hat{y}_i^{(1000)} = \exp[(\ln(\hat{y}_i / n_i)] \times 1,000$

1,000 月あたりの 95%信頼区間： $95\% \text{CL} = \exp\{\ln(\hat{y}_i / n_i) \pm 1.96 \sqrt{\text{Var}[(\ln(\hat{y}_i / n_i))]} \} \times 1,000$

片を含めて 3 変数であるのに対し、ここでの事例は 9 変数となっていて、Excel 行列関数なしには、各種の条件における 95%信頼区間の計算は絶望的である。なお、2 次式の 95%信頼区間については、第 12.4 節を参照のこと。

95%信頼区間の計算は、それらの結果のグラフ表示とセットでなければ、計算した苦労が浮かばれない。また、グラフ化に手間がかかると避けたくなる。なお、Excel の折れ線グラフで誤差範囲の設定の「標準誤差」は、表示されているデータ全体から計算された標準誤差を一律に表示するのでまったく使いものにならない。「その他のオプション」で「ユーザ設定」を使う必要がある。

Excel の折れ線グラフに 95%信頼区間を上書きするためには、「グラフの要素→誤差範囲→その他のオプション→ユーザ設定→値の推定→[正の誤差の値(P), 負の誤差の値(N)]」による設定を行う。



Excel の誤差範囲の設定は、範囲の中心点からの距離で与える仕様になっているので、図 10.6 (左) に示すように

正の誤差範囲：  $U = U_{95\%} - \text{推定値}$

負の誤差範囲：  $L = \text{推定値} - L_{95\%}$

を改めて計算する必要がある。この結果を用いて図 10.6 (右) に 95%信頼区間付き予測プロファイルを示す。この図は、図 10.2 の JMP での予測プロファイルと同等の結果を表している。ただし、図 10.2 の Y 軸の目盛りは、月あたりの損傷数であるのに対し、図 10.6 では、1,000 月あたりの損傷数となっている。慣れるまでは煩わしいが出来上がりは上々である。

因子	水準	推定値 $y^{(1000)}$	$y^{(1000)}$ からの距離	
			$L$	$U$
船種	A	1.65	0.70	1.22
	B	0.96	0.27	0.37
	C	0.83	0.46	1.04
	D	1.53	0.80	1.69
	E	2.29	1.03	1.86
建造年	60-64	1.65	0.70	1.22
	65-69	3.32	1.30	2.14
	70-74	3.74	1.43	2.31
	75-79	2.60	1.21	2.28
運行年	60-74	1.65	0.70	1.22
	75-79	2.43	1.06	1.87

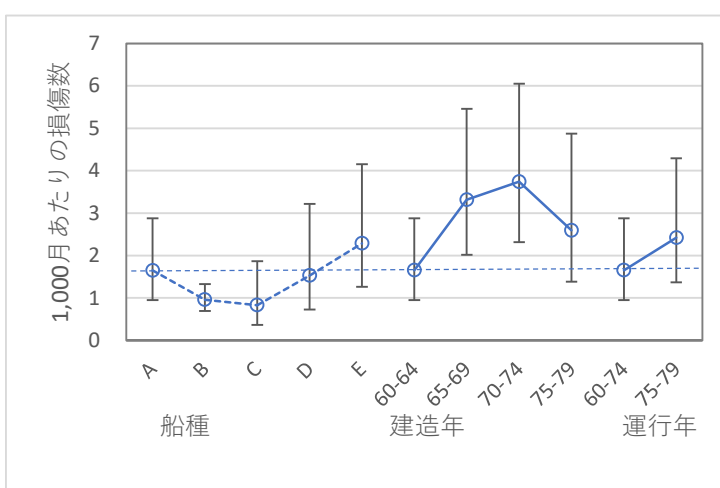


図 10.6 Excel の折れ線グラフによる 95%信頼区間付き予測プロファイル

## 10.4. 交互作用の検討

観察データに対して、主効果モデルで検討した結果、過分散であるということは、ある船種の、ある建造年の、ある運行年にかぎって損傷を受けやすい建造方法が取られている可能性が McCullagh ら (1989) で指摘されている。このような疑問に対して、交互作用を含めたポアソン回帰を適用して、結果を吟味する必要がある。交互作用として（船種×建造年）と（船種×運行年）を含めたモデルを適用する。なお、（建造年×運行年）は、組み合わせがない場合（ミッシング・セル）があり、交互作用としては含めないことにする。なお、交互作用として（建造年×運行年）をモデルに含めた場合には、推定値に「バイアスあり」との警告が出される。

解析モデルは、次の

$$\ln \text{ 損傷数} = \ln \text{ 運行月数} + \ln (\text{切片} + \text{船種} + \text{建造年} + \text{運行年} + \text{船種} \times \text{建造年} + \text{船種} \times \text{運行年})$$

とする。交互作用の自由度は、船種が 4、建造年が 3 なので、船種×建造年は  $4 \times 3 = 12$  となり、運行年の自由度は 1 なので、船種×運行年は  $4 \times 1 = 4$  となる。データ数が 34 なので、切片と主効果の自由度は 9 で、交互作用の自由度 16 を加えると 25 となり、誤差の自由度は  $34 - 25 = 9$  となる。JMP での結果を表 10.4 に示す。Pearson のカイ 2 乗が 9.8680 と自由度 9 とほぼ等しくなり過分散が解消している。

表 10.11 交互作用モデル 1（過分散なし）

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	68.9038	137.8075	24	<.0001*
完全	53.1936			
縮小	122.0974			
適合度統計量		カイ2乗	自由度	p値
Pearson		9.8680	9	0.3613
デビアン		8.5208	9	0.4826
AICc				
318.8873				
効果の検定				
要因	自由度	尤度比カイ2乗	p値	
船種	4	1.7671	0.7785	
建造年	3	13.2437	0.0041*	
運行年	1	0.1822	0.6695	
船種*建造年	12	25.2354	0.0137*	
船種*運行年	4	6.0661	0.1943	

交互作用が、統計的に有意となったのは、(船種×建造年)であり、有意ではない(船種×運行年)をモデルから除いて、ただし、(運行年)と(船種)の主効果も有意ではないからとの理由で除いてはならない。これは、交互作用(船種×運行年)をモデルに入れた結果として、主効果が有意ではなくなったとも考えられるからである。同様に、交互作用(船種×建造年)が有意であるが、主効果(船種)は有意ではないからといってモデルから除いてしまうと、交互作用が適切に評価されなくなる恐れがある。

表 10.12 に交互作用(船種×建造年)と主効果(船種, 建造年, 運行年)に対するポアソン回帰の結果を示す。Pearson のカイ 2 乗値は自由度 13 にたいして 17.3680 とやや大きくなるが、統計的には有意ではない。

表 10.12 交互作用モデル 2 (過分散なし)

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	65.8707	131.7415	20	<.0001*
完全	56.2267			
縮小	122.0974			
適合度統計量	カイ2乗	自由度	p値	
Pearson	17.3680	13	0.1830	
デビアンス	14.5869	13	0.3338	
AICc				
231.4534				
効果の検定				
要因	自由度	尤度比カイ2乗	p値	
船種	4	1.9918	0.7373	
建造年	3	12.8690	0.0049*	
運行年	1	10.6215	0.0011*	
船種*建造年	12	24.1082	0.0197*	

図 10.7 に(船種を E, 建造年 70-74,, 運行年 75-79)とした予測プロファイルを示す。予測プロファイルに引き続き交互作用プロファイルを描くこともできるが、水準の数が多いこともあり、煩雑で読み取りにくいので、図 10.7 の予測プロファイル上で、船種を A から E まで順次選択した時の建造年の予測プロファイルの変化を詳細に検討する。

図 10.8 に示すように、船種 A, B, C では、建造年の予測プロファイルは、月あたりの損傷数は、若干の上下はあるもののほぼ同様の推移であり、交互作用はないものと推測される。船種 D では、建造年 70-74 に損傷数が急上昇し、建造年 75-79 には、船種 A, B, C と同程度



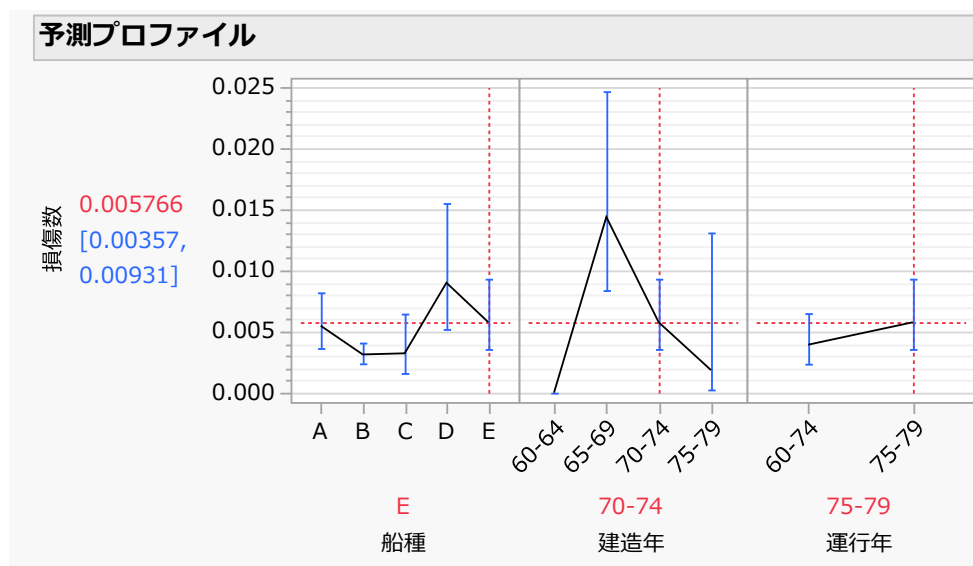


図 10.7 予測プロファイル（船種をE, 建造年 70-74, 運行年 75-79）を基準

のレベルに落ち着いている．船種 E は，建造年 65-69 に驚異的な損傷数の急上昇があり，建造年 70-74 には，半減平均的な損傷数となっている．建造年 75-79 では，さらに低いレベルまで損傷数が減っている．

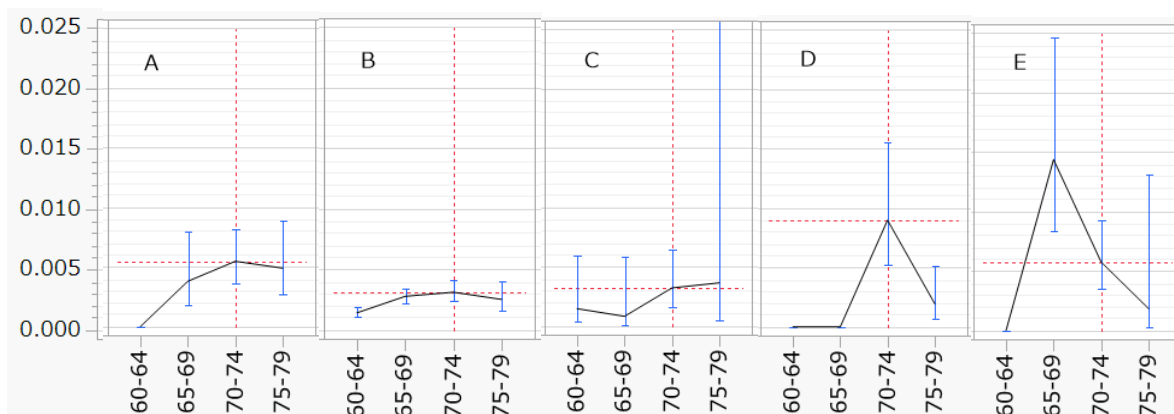


図 10.8 船種を変化させた場合の建造年の予測プロファイル

交互作用を含むモデルの予測プロファイルの推定値を報告書に表として示したい．主効果モデルの場合は，第 10.3 節で「Excel による予測プロファイル」で，共分散行列を JMP から取得して推定値および 95%信頼区間の計算を行い，折れ線グラフにより作図した結果を示してきたが，交互作用を含む場合には，デザイン変数が 21 なので， $21 \times 21$  の行列となり，Excel による対応は非現実的である．

図 10.8 で示した予測プロファイルは，運行年 75-79 に固定し，船種を A～E と変化させつつ建造年別の推定値と 95%信頼区間を描いている．これらの推定値を得るためには，JMP の



ポアソン回帰の追加機能である「列の保存➡（予測式，平均の信頼区間）」を使うことにより得ることができる。

表 10.13 に示すように，解析用のデータの末尾に推定したい条件についてのデータを追加することにより，自動的に推定値，および 95%信頼区間を得ることができる．JMP の予測プロファイルは，月あたりの故障数となっているので，運行月数として 1 とする．もちろん対数は 0.0 になる．表 10.13 には，追加レコードも含めて JMP ファイルを Excel に取り込み，折れ線グラフ用の長さ（L，U）を計算した結果が示されている．これらを用いて図 10.9 に示すように，Excel による交互作用プロファイルを得る．

表 10.13 追加レコードに対する JMP による予測値と 95%信頼区間の計算

No	船種	建造年	運行年	運行月数	ln運航月数	損傷数	目盛り	交互作用予測値	L 95%	U 95%	長さL	長さU
1	A	60-64	60-74	127	4.84	0		0.00	0.00			
2	A	60-64	75-79	63	4.14	0		0.00	0.00			
3	A	65-69	60-74	1095	7.00	3		2.83	1.33	6.02		
:												
38	E	70-74	75-79	2161	7.68	12		12.46	7.72	20.13		
39	E	75-79	60-74	.								
40	E	75-79	75-79	542	6.30	1		1.00	0.14	7.10		
追加レコード	A	60-64	75-79	1	0.00		A60	<b>0.0000</b>	0.0000		<b>0.0000</b>	
	A	65-69	75-79	1	0.00		A65	<b>0.0038</b>	0.0018	0.0080	<b>0.0020</b>	<b>0.0042</b>
	A	70-74	75-79	1	0.00		A70	<b>0.0055</b>	0.0037	0.0082	<b>0.0018</b>	<b>0.0027</b>
	A	75-79	75-79	1	0.00		A75	<b>0.0049</b>	0.0027	0.0089	<b>0.0022</b>	<b>0.0040</b>
	B	60-64	75-79	1	0.00		B60	<b>0.0014</b>	0.0011	0.0019	<b>0.0003</b>	<b>0.0005</b>
コド	B	65-69	75-79	1	0.00		B65	<b>0.0028</b>	0.0022	0.0035	<b>0.0006</b>	<b>0.0007</b>
	:						:					
	E	70-74	75-79	1	0.00		E70	<b>0.0058</b>	0.0036	0.0093	<b>0.0022</b>	<b>0.0035</b>
	E	75-79	75-79	1	0.00		E75	<b>0.0018</b>	0.0003	0.0131	<b>0.0015</b>	<b>0.0113</b>

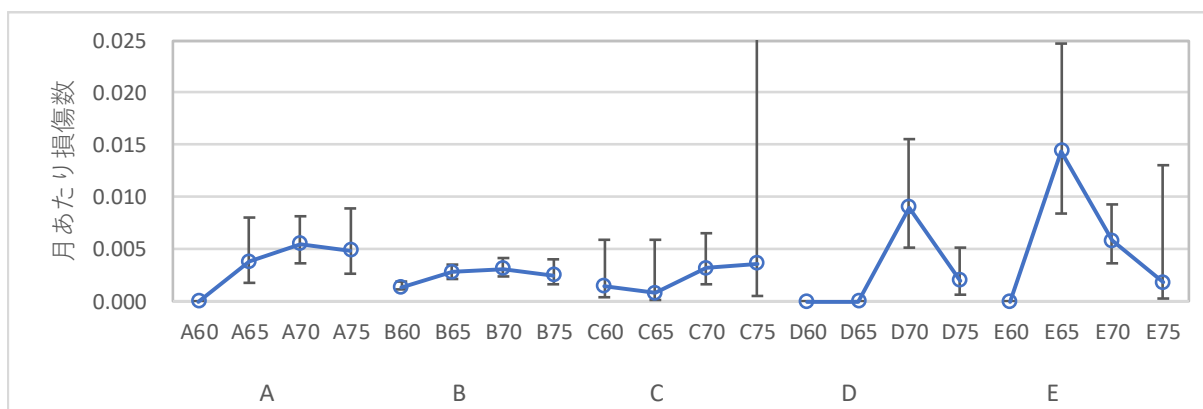


図 10.9 Excel による船種を変化させた場合の建造年の交互作用予測プロファイル

## 10.5. 主効果モデルを活用した新たな交互作用の可視化の試み

観察データに対する交互作用を含めたポアソン回帰は、因子間の特異的な関連を見極めることが必要である。しかし、船種に対する損傷数の解析の本来の目的は、保険会社の視点からは、直近の損傷数のデータから、近未来の損傷数を予測し、保険料を適切に算定するのが目的と思われる。

ある年に建造されたある船舶の種類によって、実際の運航年に起きた損傷数が、平均的な損傷数に比較して多くなった場合には、次年度の保険料を高く設定する必要があり、損傷数が平均的な損傷数に比べて減少傾向にあるならば、保険料を安く設定することも必要と思われる。

船舶の種類と建造年との間に交互作用があるということは、ある船種の損傷数について特異的な変動があることが示唆されたことになる。どのような特異的な変動かは、図 10.8 の船種別の建造年による損傷数の推移によって、船種 D と船種 E に特異的な変動があることが認識されるが、この図は予測プロファイルを手作業で変化させて、切り貼り作業で完成させたものであり、全体を概観するには難点がある。

伝統的に交互作用の検出は、分散分析表の分散比の  $F$  検定で行ない、有意ならばグラフに示し目視による解釈をしてきた。ポアソン回帰の場合であれば、表 10.11 および表 10.12 に示したように、尤度比カイ 2 乗検定の結果を分散比の  $F$  検定と同様に示してきたのであるが、グラフ表示を試みるも隔靴搔痒のごとくである。

どのような交互作用なのかを手軽に認識するために、主効果モデルでの予測値  $\ln \hat{y}_i$  を、損傷千月比に換算した予測値  $\hat{y}_i^{(1,000)}$ （以下、主効果予測値とする）を計算する。元データの損傷数から計算した損傷千月比  $y_i^{(1,000)}$ （以下、損傷千月比とする）と対比することにより、交互作用が浮き彫りできるのではないかと考えた。

主効果モデルの予測値  $\hat{y}_i^{(1,000)}$  は、表 10.3 のデザイン行列に対し、表 10.4 に示した JMP で得られたのポアソン回帰のパラメータの推定値  $\hat{\beta}$ 、デザイン行列の  $i$  行目のベクトル  $\mathbf{x}_i$  を用い、表 10.14 に示すように

$$\hat{y}_i^{(1,000)} = \frac{\exp[\ln(n_i) + (\mathbf{x}_i \hat{\beta})]}{n_i} \times 1000$$

によって計算した結果である。

表 10.14 主効果モデルにおける予測値の計算

	船	建造	運行	運行	損傷	損傷千 月比	デザイン行列										予測値	主効果 予測値
No	種	年度	年度	月数 $n_i$	数 $y_i$	$y_i^{(1000)}$	$x_0$	$x_B$	$x_C$	$x_D$	$x_E$	$x_{C65}$	$x_{C70}$	$x_{C75}$	$x_{Op75}$	$\beta^{\wedge}$	$y_i^{\wedge(1000)}$	
1	A	60-64	60-74	127	0	0.0000	1	0	0	0	0	0	0	0	0	-6.4059	= 1.6518	
2			75-79	63	0	0.0000	1	0	0	0	0	0	0	0	1	-0.5433	2.4262	
3		65-69	60-74	1,095	3	2.7397	1	0	0	0	0	1	0	0	0	-0.6874	3.3168	
4			75-79	1,095	4	3.6530	1	0	0	0	0	1	0	0	1	-0.0760	4.8718	
5		70-74	60-74	1,512	6	3.9683	1	0	0	0	0	0	1	0	0	0.3256	3.7445	
6			75-79	3,353	18	5.3683	1	0	0	0	0	0	1	0	1	0.6971	5.5000	
7		75-79	60-74	-	-	-	1	0	0	0	0	0	0	1	0	0.8184	-	
8			75-79	2,244	11	4.9020	1	0	0	0	0	0	0	1	1	0.4534	3.8181	
9	B	60-64	60-74	44,882	39	0.8689	1	1	0	0	0	0	0	0	0	0.3845	0.9594	
10			75-79	17,176	29	1.6884	1	1	0	0	0	0	0	0	1		1.4091	
:																		
33	E	60-64	60-74	45	0	0.0000	1	0	0	0	1	0	0	0	0		2.2874	
34			75-79	-	-	-	1	0	0	0	1	0	0	0	1		-	
35		65-69	60-74	789	7	8.8720	1	0	0	0	1	1	0	0	0		4.5932	
36			75-79	437	7	-	1	0	0	0	1	1	0	0	1		6.7466	
37		70-74	60-74	1,157	5	4.3215	1	0	0	0	1	0	1	0	0		5.1855	
38			75-79	2,161	12	5.5530	1	0	0	0	1	0	1	0	1		7.6166	
39		75-79	60-74	-	-	-	1	0	0	0	1	0	0	1	0		-	
40			75-79	542	1	1.8450	1	0	0	0	1	0	0	1	1		5.2874	
$y_i^{\wedge(1000)}=(\exp(\ln(n_i)+\text{Mmult}(\boldsymbol{x}_i\text{の範囲},\boldsymbol{\beta}^{\wedge}\text{の範囲})))/n_i*1000$																		

JMP のデータセットにもこれと同じ結果が出力されており、図 10.10 に示すようにグラフ・ビルダーで「船種別・運行年別・建造年別の主効果予測値  $\hat{y}_i^{(1,000)}$  の推移図を描き、さらに損傷千月比  $y_i^{(1,000)}$  を上書きした結果である。平均的な損傷数に対応するのが主効果予測値  $\hat{y}_i^{(1,000)}$  の推移であり、実際の損傷千月比  $y_i^{(1,000)}$  により、特異的な変動が浮き彫りされている。

- 1) 多くの場合、1,000 月あたり 5 件以下の損傷数である。
- 2) 船種 C タイプの建造年 1970-74 の 1960-74 の運行年の千月比の損傷数は、7.66 件となるが、1975-79 の運行年では、1.03 件と減少している。
- 3) 船種 D タイプの建造年 1970-74 は、運行年に関わらず千月あたりの損傷数が多い。
- 4) 船種 E タイプは、建造年 1965-70 は、運行年に関わらず千月あたりの損傷数が 8.87, 16.02 と多く、船種 E タイプの建造年 1970-74 は、建造年 1965-70 に比べて減少しているが、4.32, 5.55 と少なくはない。しかし、船種 E タイプの建造年 1975-79 の場合には、1.85 と減少している。

保険会社の観点から考察してみよう。

- 1) 船種 A は、建造年 75-79 の運行年 75-79 の  $y_i^{(1,000)}$  が「平均」より僅かに多くなっている。次年度にさらに多くなるか注意が必要であるが、保険料は据え置きとする。
- 2) 船種 B は、常に「平均」的な推移であり、通常の保険料で据え置きとする。

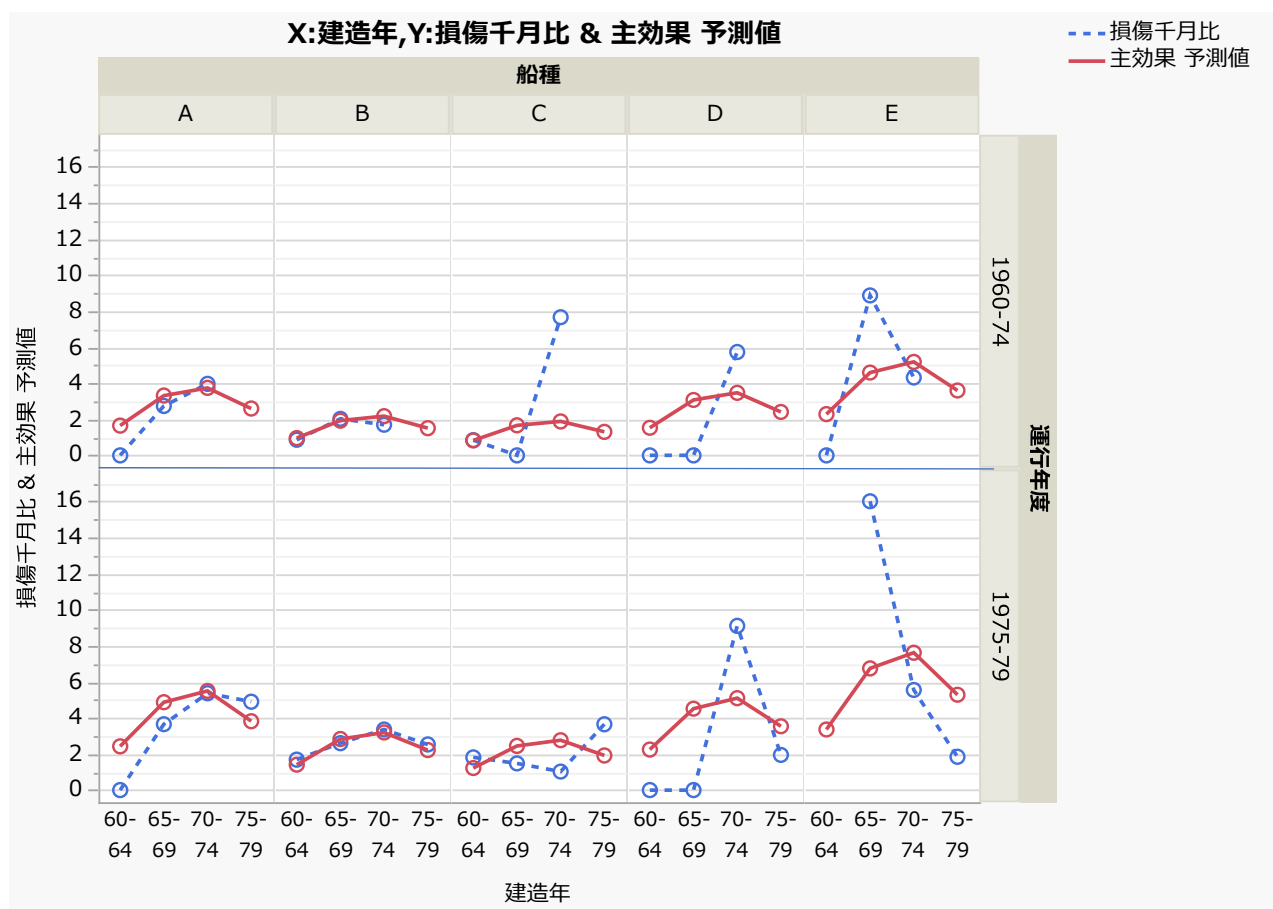


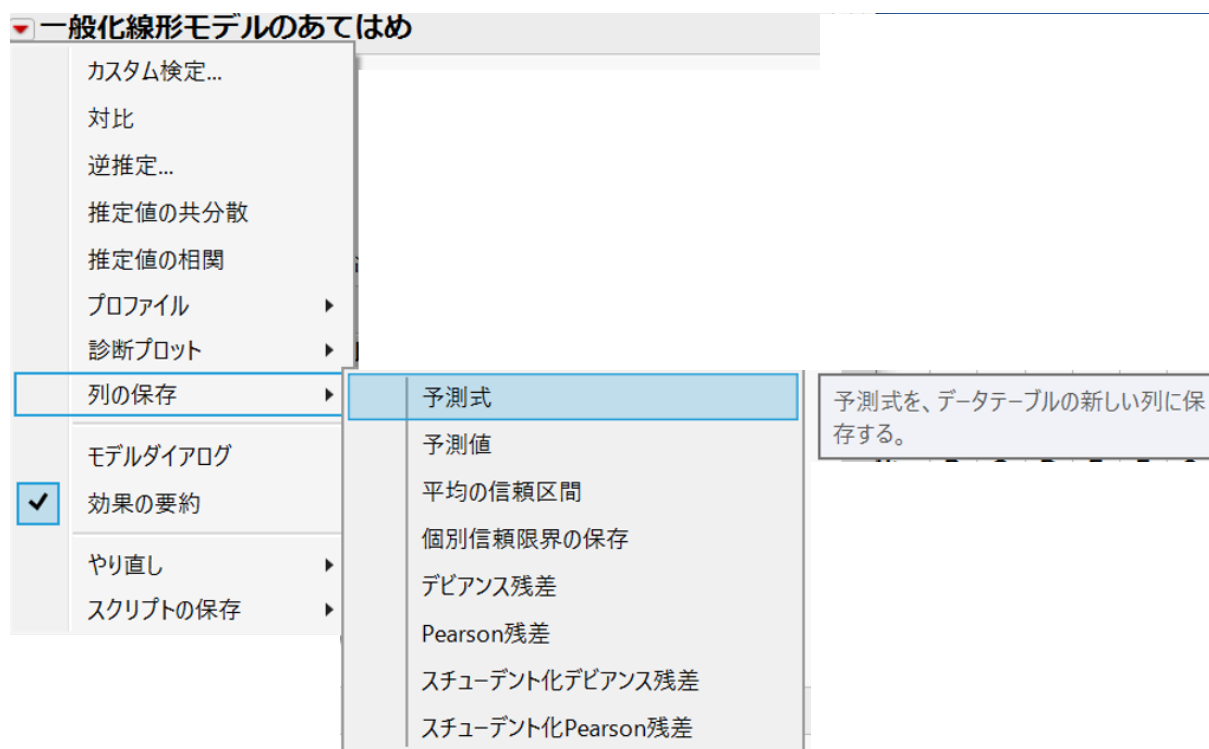
図 10.10 船種別、建造年別の運行年による 1,000 月あたりの損傷数の変化

- 3) 船種 C は、建造年 70-74 の場合に運行年 60-74 で「平均」よりも明らかに  $y_i^{(1,000)}$  が多くなったが、運行年 70-79 で  $y_i^{(1,000)}$  が「平均」以下となっているので、何が原因なのかを検討の上、慎重に保険料の算定が必要であろう。
- 4) 船種 D は、建造年 70-74 の場合の運行年 60-74 で  $y_i^{(1,000)}$  が「平均」より明らかに多くなり、運行年 70-79 でも同様の傾向であり、保険料を高く設定してあるか、確認を要する。
- 5) 船種 E は、建造年 65-69 の  $y_i^{(1,000)}$  が運行年 60-74 でも「平均」より明らかに多く、運行年 70-79 では、さらに増加しており、次年度以降さらなる保険料の増額が必要である。建造年 70-74、建造年 75-79 については、「平均」的な保険料での継続としてよいだろう。

JMP による主効果モデルのポアソン回帰の予測値から  $\hat{y}_i^{(1,000)}$  を求め、グラフ・ビルダーを用いて実際の損傷数  $y_i^{(1,000)}$  を重ね合わせた推移図が、結果を考察する際に有益であると思われる。統計ソフト S-PLUS では、JMP のグラフ・ビルダーに相当する Trellis (格子) グラフ

機能が備わっているが、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。なお、SAS には、グラフ・ビルダーに代わるものがない。

Excel の折れ線グラフを用いて、図 10.10 に匹敵するものができることを示そう。まず、主効果モデルで推定されたパラメータを用いた「予測値」を外部ファイルに出力する。JMP では、「列の保存→予測式」で推定値  $\hat{y}_i$  が得られる。



元のデータと予測値を Excel に取り込み、損傷数  $y_i$  を  $y_i^{(1,000)}$  に、予測値  $\hat{y}_i$  を  $\hat{y}_i^{(1,000)}$  に、次式で、

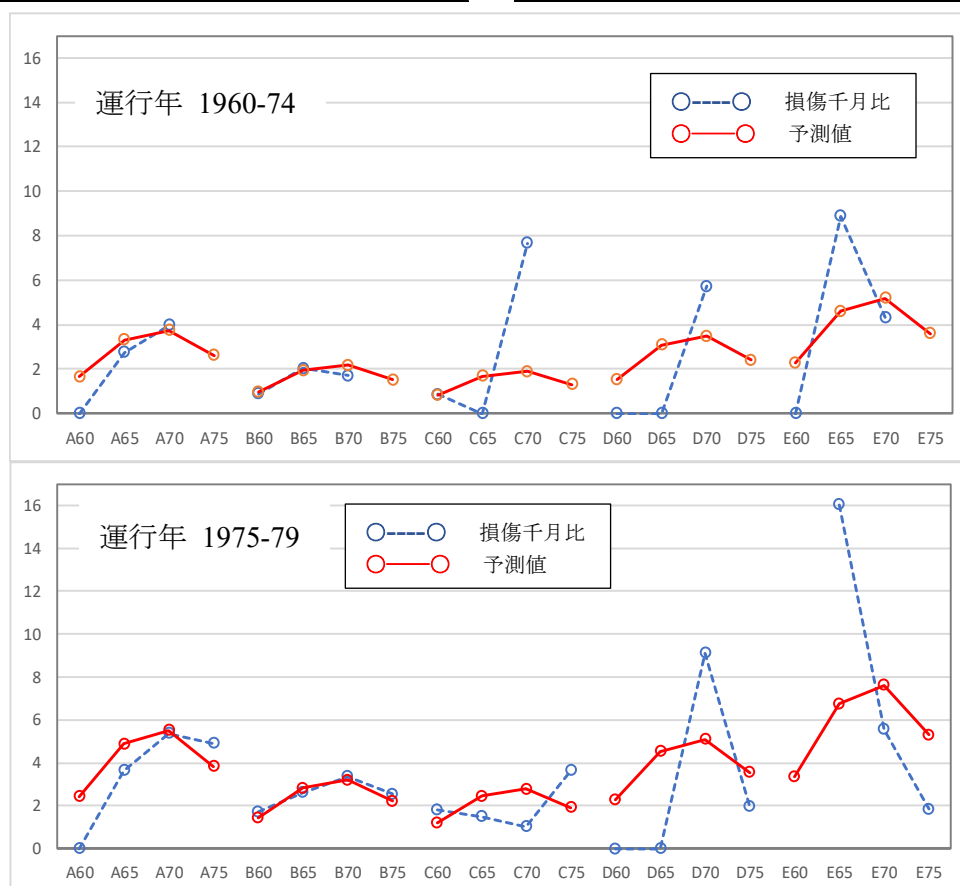
$$y_i^{(1,000)} = \frac{y_i}{n_i} \times 1,000, \quad \hat{y}_i^{(1,000)} = \frac{\hat{y}_i}{n_i} \times 1,000$$

計算する。更に運用年ごとに上下に配置したいので、運用年別・船種別・建造年別にソートした結果を表 10.15 に示す。運行年毎に  $y_i^{(1,000)}$  と  $\hat{y}_i^{(1,000)}$  を選択し、折れ線グラフを作成し、線の色とマーカの書式を設定し、船種間の切れ目の線を消して、図 10.11 が完成する。

細々した交互作用の検討をせずに、主効果モデルによる予測値  $\hat{y}_i^{(1,000)}$  が図示されているので、損傷千月比  $y_i^{(1,000)}$  が予測値  $\hat{y}_i^{(1,000)}$  同等レベルなのか、あるいは、大きく外れているが一目で把握でき、交互作用の解析結果と同様な判断が行なえる。私にとっても、このような予測プロファイルの作成は、これが初めてであり、新たな解析法として有益であることを認識した。

表 10.15 Excel の折れ線グラフのための Excel シート

運行年	船種	運行年	損傷数 $y_i^{(1000)}$	推定値 $\hat{y}_i^{(1000)}$
Op60-74	A	A60	0.00	1.65
		A65	2.74	3.32
		A70	3.97	3.74
		A75		2.60
	B	B60	0.87	0.96
		B65	2.03	1.93
		B70	1.70	2.17
		B75		1.51
	C	C60	0.85	0.83
		C65	0.00	1.67
		C70	7.66	1.88
		C75		1.31
	D	D60	0.00	1.53
		D65	0.00	3.07
		D70	5.73	3.47
		D75		2.41
	E	E60	0.00	2.29
		E65	8.87	4.59
		E70	4.32	5.19
		E75		3.60
Op75-79	A	A60	0.00	2.43
		A65	3.65	4.87
		A70	5.37	5.50
		A75	4.90	3.82
	B	B60	1.69	1.41
		B65	2.60	2.83
		B70	3.36	3.19
		B75	2.53	2.22
	C	C60	1.81	1.22
		C65	1.48	2.45
		C70	1.03	2.77
		C75	3.65	1.92
	D	D60	0.00	2.25
		D65	0.00	4.52
		D70	9.11	5.10
		D75	1.95	3.54
	E	E60		3.36
		E65	16.02	6.75
		E70	5.55	7.62
		E75	1.85	5.29



## 10.6. Excel のソルバーによるオフセットを含むポアソン回帰

Excel のソルバーを用いてオフセットを含む場合のポアソン回帰については、これまでも取り上げてきたが、変数の数は切片を含め 3 変数までであり、更に変数が多い場合の適用可能性を含め検討する。

第 2 章では、対数尤度関数について 2 階の偏微分行列を用いたニュートン・ラフソン法を用いた反復計算を示した。ただし、変数が多くなると 2 階の偏微分行列の計算が変数の 2 乗に比例して増えるので、3 変数までは  $3 \times 3$  の内 6 変数の計算であるが、4 変数となると  $4 \times 4$  の 16 変数となり、Excel での対応は限界を超えている。

第 5 章では、反復重み付き回帰による方法を示したが、変数の数が増えてもデザイン行列  $X$  に対する種々の行列計算なので、行列のサイズを変更するだけで対応可能であり、変数が増えたとしても対応は可能ではあるが、5 変数となると  $5 \times 5$  の行列計算が必要となり、限界に近い。

第 2 章でも第 5 章でもパラメータの推定だけでなく、パラメータの共分散行列の推定も Excel で行っており、このために計算手順が込み入っている。これらは、ポアソン回帰を通じて一般化線形モデルの理論をきちっと学習するために適しているが、変数が多くなった場合には、勧められない。

第 10.5 節で示したように、95%信頼区間の算出せずに、モデルによる予測値と観測値の対応を主体にするような場合には、Excel のソルバーを用いることにより容易に各種のポアソン回帰のパラメータの推定を行うことができる。この方法ならば、これまでも示してきたガンマ・ポアソン回帰（負の 2 項分布）、ゼロ過剰ポアソン回帰、ゼロ過剰ガンマ・ポアソン回帰のパラメータの推定でも、Excel のソルバーによる主効果モデルに対応可能である。

第 10.2 節の「主効果モデル」で質的変数に対するデザイン行列を使い、表 10.14 に示すように JMP で推定したパラメータを用いたオフセットがある場合のポアソン回帰の予測値の計算過程を示した。統計ソフトによるポアソン回帰を行う際に、どのようなデザイン行列が内部で生成されているのかを正確に知ることが、推定された結果の解釈に不可欠であることを繰返し示してきた。

生成されたデザイン行列を使い、Excel ソルバーを使ってポアソン回帰に限らず各種の回帰モデルに対してパラメータの推定を行うことに煩雑さはなく、生成されたデザイン行列に対応したパラメータの推定値が、どのような性質を持つのを理解し、結果の解釈に役立つ各種のグラフを試行錯誤的に作成することは、現実の問題に対して統計ソフトから出力される各種の結果を、解釈しやすいグラフを自在に作成するために役に立つ。

第 10.2 節の表 10.3 をベースに、Excel ソルバーを用い最尤法によるポアソン回帰を行うために拡張する。表 10.16 の右側の「ポアソン確率  $p_i$ 」の列、「 $\ln p_i$ 」の列が表 10.13 に付け加えたものであり、ポアソン確率は、「 $\hat{y}_i$ 」を母数とした場合の「 $y_i$ 」のポアソン分布の確率で Excel の Poisso.dist() 関数で計算した結果であり、その対数を取った「 $\ln p_i$ 」がそれぞれの対数尤度となっている。「 $\ln p_i$ 」を全て加えた結果が対数尤度「 $\ln L = -125.245$ 」であり、これを最大にするようにパラメータ  $\beta$  を Excel のソルバーで変化させることにより最尤解が得られる。

初期値としてパラメータ (-6, 0, 0, 0, 0, 0, 0, 0, 0) を与えている。これは、 $\beta_0 = -7$  とすると  $\ln L = -224.95$ 、 $\beta_0 = -5$  とすると  $\ln L = -465.94$  となり、 $\beta_0 = -6$  としたときが  $\ln L = -125.245$  と大きくなったためである。さらに、その前後 (-5.5, -6.5) などと動かして  $\ln L$  が大きくなるか試し、適当な初期値を設定する。Excel のソルバーで  $\ln L = -125.245$  のセルを「目的のセルの設定」とし、「変数セルの変更」欄で、パラメータ (-6, 0, 0, 0, 0, 0, 0, 0, 0) のセルを選択し、「解決」により、対数尤度  $\ln L$  を最大にするようなパラメータを瞬時に求めてくれる。

表 10.16 初期値に対する対数尤度

	建 船	運 造	運行 月数	運行 月数	損傷 数	---船種---					--建造--				Op	変化 セル	$\ln n_i +$ $x\beta =$	$\exp(\ln$ $y_i)$	ポアソ ン確率	$\ln L$ -125.245	最大 化セル
						x	B	C	D	E	65	70	75	65							
No	種	年	年	$n_i$	$y_i$	0	1	2	3	4	5	6	7	8	$\beta$	$\ln y_i$	$y_i$	$P_i$	$\ln P_i$		
1	A	60-	60-	127	0	1	0	0	0	0	0	0	0	0	-6.0000	-1.16	0.3	0.730	-0.315		
2			75-	63	0	1	0	0	0	0	0	0	0	1	0.0000	-1.86	0.2	0.855	-0.156		
3		65-	60-	1,095	3	1	0	0	0	0	1	0	0	0	0.0000	1.00	2.7	0.221	-1.510		
4			75-	1,095	4	1	0	0	0	0	1	0	0	1	0.0000	1.00	2.7	0.150	-1.898		
5		70-	60-	1,512	6	1	0	0	0	0	0	1	0	0	0.0000	1.32	3.7	0.091	-2.400		
6			75-	3,353	18	1	0	0	0	0	0	1	0	1	0.0000	2.12	8.3	0.001	-6.590		
8		75-	75-	2,244	11	1	0	0	0	0	0	0	1	1	0.0000	1.72	5.6	0.015	-4.188		
9	B	60-	60-	44,882	39	1	1	0	0	0	0	0	0	0	0.0000	4.71	111.3	0.000	-34.123		
10			75-	17,176	29	1	1	0	0	0	0	0	0	1	0.0000	3.75	42.6	0.006	-5.045		
11		65-	60-	28,609	58	1	1	0	0	0	1	0	0	0		4.26	70.9	0.015	-4.205		
:																					
37		70-	60-	1,157	5	1	0	0	0	1	0	1	0	0		1.05	2.9	0.092	-2.387		
38			75-	2,161	12	1	0	0	0	1	0	1	0	1		1.68	5.4	0.005	-5.204		
40		75-	75-	542	1	1	0	0	0	1	0	0	1	1		0.30	1.3	0.351	-1.048		



表 10.18 に対数尤度を最大化するようなパラメータ (-6.4059, -0.5433, -0.6874, -0.0759, 0.3255, 0.6972, 0.8185, 0.4534, 0.3845) が推定されている。この結果が正しいことは、表 10.4 の JMP でのポアソン回帰の結果と一致することにより検証される。得られた  $\ln L = -68.2808$  は、JMP の結果の対数のサマリー表の「完全」に一致している。

表 10.17 JMP による主効果モデルの対数尤度（過分散なし）

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	53.8166	107.6333	8	<.0001*
完全	68.2808			
縮小	122.0974			
適合度統計量		カイ2乗	自由度	p値(Prob>ChiSq)
Pearson		42.2753	25	0.0168*
デビアン		38.6951	25	0.0395*
<b>AICc</b>				
162.0615				

表 10.18 ソルバーによる対数尤度の最大化（過分散なし）

	建	運	運行	損傷	---船種---					--建造--				Op	変化	$\ln n_i +$	$\exp(\ln$	ポアソ	$\ln L$	最
	船	造	行	月数	数	x	B	C	D	E	65	70	75	65	セル	$x\hat{\beta} =$	$y_i \hat{)}$	ン確率	-68.2808	大
No	種	年	年	$n_i$	$y_i$	0	1	2	3	4	5	6	7	8	$\hat{\beta}$	$\ln y_i \hat{)}$	$y_i \hat{)}$	$P_i$	$\ln P_i$	化
1	A	60-	60-	127	0	1	0	0	0	0	0	0	0	0	-6.4059	-1.56	0.2	0.811	-0.210	セ
2			75-	63	0	1	0	0	0	0	0	0	0	1	-0.5433	-1.88	0.2	0.858	-0.153	ル
3		65-	60-	1,095	3	1	0	0	0	0	1	0	0	0	-0.6874	1.29	3.6	0.211	-1.554	
4			75-	1,095	4	1	0	0	0	0	1	0	0	1	-0.0759	1.67	5.3	0.163	-1.816	
5		70-	60-	1,512	6	1	0	0	0	0	0	1	0	0	0.3255	1.73	5.7	0.159	-1.839	
6			75-	3,353	18	1	0	0	0	0	0	1	0	1	0.6972	2.91	18.4	0.093	-2.374	
8		75-	75-	2,244	11	1	0	0	0	0	0	0	1	1	0.8185	2.15	8.6	0.087	-2.442	
9	B	60-	60-	44,882	39	1	1	0	0	0	0	0	0	0	0.4534	3.76	43.1	0.052	-2.950	
10			75-	17,176	29	1	1	0	0	0	0	0	0	1	0.3845	3.19	24.2	0.047	-3.052	
11		65-	60-	28,609	58	1	1	0	0	0	1	0	0	0		4.01	55.1	0.049	-3.025	
:																				
37		70-	60-	1,157	5	1	0	0	0	1	0	1	0	0		1.79	6.0	0.161	-1.829	
38			75-	2,161	12	1	0	0	0	1	0	1	0	1		2.80	16.5	0.059	-2.836	
40		75-	75-	542	1	1	0	0	0	1	0	0	1	1		1.05	2.9	0.163	-1.813	

Excel のソルバーを使うことにより、推定値は得られるが信頼区間を計算するために必須のパラメータの共分散行列が得られないのが残念である。ただし、主効果モデルでの予測値に対する観測値を同一の折れ線グラフで表わすことの有用性を第 10.5 節で示したが、Excel のソルバーの結果を用いても容易に実現できる。

## 10.7. SAS の GENMOD プロシジャを使った解析

SAS/GENMOD プロシジャが本格的に提供されたのは、2000 年ごろのバージョン 8 からである。その当時は、ロジスティック回帰に対する適用を主体にしており、高橋 (2002)、「GENMOD プロシジャによる計数データの解析」で、LOGISTIC プロシジャとの使い分けに興味があり、ポアソン回帰について関心はなかった。無償版の OnDemand SAS が提供されるようになったので、これまでの章でも GENMOD プロシジャによるポアソン回帰の使い方を示してきた。ここでは、第 10.5 節で示した主効果モデルによる探索的な解析について GENMOD プロシジャを主体にし、負の 2 項回帰 (ガンマ・ポアソン回帰) への拡張も示す。無償版の SAS の利用に関しては、高波洋平・舟尾暢雄 (2016)、「SAS Studio によるやさしい統計データ分析」が詳しい。SAS と R と用いた各種の統計解析については、臨床評価研究会 (ACE) 基礎解析分科会 (2017) に詳しく示されているが、残念ながらポアソン回帰については記述はない。

### SAS データセットの作成

SAS は、DATA ステップおよび PROC ステップの繰返しを基本としている。DATA ステップとは、「DATA」で始まるステップで、SAS データセットを作成するためのプログラミング言語である。PROC ステップとは、GENMOD プロシジャなど数百の解析用のプロシジャを起動するステップである。DATA ステップで解析用の SAS データセットを作成し、PROC ステップで解析した結果を SAS データセットへ出力し、DATA ステップで加工し、また別の PROC ステップでの解析を行う。このような SAS データセットを介した繰返が特徴である。

SAS の DATA ステップでは、表 10.3 に示すような Excel ファイルを読み込むこともできるが、必要最小限のデータから、プログラミング機能を用いて解析用の SAS データセットを作成するのが通常である。必要最小限  $h$  なのは、運行月数  $n_i$  および損傷数  $y_i$  である。他の因子の水準は、プログラミング的に内部生成する。実際のデータリストは、`datalines` ステートメント以下の 5 行で、最初の行は船種 A で、その後に  $(n_1, y_1, n_2, y_2, \dots, n_8, y_8)$  が続いている。船種 (A, B, C, D, E) も内部で生成できるが、行の識別子として付けてある。

SAS の DAT ステップの基本は、1 行分のデータを読み込み、何らかの処理をした後に最初のステートメントに戻り、読み込むデータが尽きるまで続ける。1 行の最初の船種の文字データを読み込み、建造年 `Cons` とし 4 年度分、運行年度 `Oper` とし 2 年度分、合計 8 セットの運行月数 `Month`、損傷数 `Inci` を順次読み込んで `output` ステートメントで、それ以前に設定され

た変数名とそのデータを SAS データセット d01 に順次出力する。1 行分の入力データリストから 8 行分の SAS データセットができあがり、5 行分のデータリストから、全体で 40 行分の SAS データセットが生成される。

DATA ステップで生成された SAS データセット d01 を確認するためには、proc print で SAS データセット名を data=d01 と指定して出力する。なお、Data ステップの詳細は、第 9.3 節を参照のこと。

```
Title "Poisson_S10_Ship_Incidents 2019/12/011 Y.Takahashi" ;
data d01 ;
  input Type$ @@ ;
  do Cons="60-64","65-69","70-74","75-79" ;
    do Oper="60-74","75-79" ;
      input Month Inci @@ ;
      ln_Month=log(Month) ; output ;
    end;
  end ;
datalines ;
A 127 0 63 0 1095 3 1095 4 1512 6 3353 18 . . 2244 11
B 44882 39 17176 29 28609 58 20370 53 7064 12 13099 44 . . 7117 18
C 1179 1 552 1 781 0 676 1 783 6 1948 2 . . 274 1
D 251 0 105 0 288 0 192 0 349 2 1208 11 . . 2051 4
E 45 0 0 . 789 7 437 7 1157 5 2161 12 . . 542 1
;
proc print data=d01 ; run;
```

Proc print による 40 行分の出力						
Obs	Type	Oper	Cons	Month	Inci	ln_Month
1	A	60-64	60-74	127	0	4.8442
2	A	60-64	75-79	63	0	4.1431
3	A	65-69	60-74	1095	3	6.9985
4	A	65-69	75-79	1095	4	6.9985
:						
39	E	75-79	60-74	.	.	.
40	E	75-79	75-79	542	1	6.2953

## 過分散を考慮したポアソン回帰

GENMOD プロシジャで、分布を dist=poisson、リンク関数を link=log、過分散の指定を scale=pearson、オフセットを offset=ln\_Month とし、推定値  $\ln \hat{y}_i$  を xbaata=xbeta、推定値  $\hat{y}_i$  を predicted=pred、スチューデント化デビアンズ残差を stdresdev=s\_dev を SAS データセット output

out=out01 で作成している。再度 proc print でこれらの統計量を出力し、proc univariate で s\_dev についての基本統計量を計算している。

```
Title2 '<<< poisson offset ref=first 過分散 >>>' ;
proc genmod data=d01 ;
  class Type Cons Oper / ref=first ;
  model Inci = Type Cons Oper
    / dist=poisson link=log type3
    scale=Pearson offset=ln_Month ;
  output out=out01 xbeta=xbeta predicted=pred
    stdresdev=s_dev ;
run ;
proc print data=out01 ; run ;
proc univariate data=out01 ;
  var s_dev ; run;
```

**<<< poisson offset ref=first 過分散 >>>  
GENMOD プロシジャ**

適合度評価の基準			
基準	自由度	値	値/自由度
デビアンس	25	38.6951	1.5478
Scaled デビアンス	25	22.8828	0.9153
Pearson カイ 2 乗	25	42.2753	1.6910
Scaled Pearson カイ 2 乗	25	25.0000	1.0000
対数尤度		454.1742	
完全対数尤度		-68.2808	
AIC (小さいほどよい)		154.5615	
AICC (小さいほどよい)		162.0615	
BIC (小さいほどよい)		168.2988	

アルゴリズムは収束しました。

注) デビアンス=38.6951 は、過分散なしの場合。表 10.17 に一致。

Scaled デビアンスは、デビアンスを尺度 1.3004 の 2 乗=1.6910 で除した結果

Pearson カイ 2 乗=42.2753 は、過分散なしの場合。表 10.17 に一致。

Scaled Pearson カイ 2 乗=25.0000 は、42.2753/1.6910

対数尤度=454.1742 は、調査中

完全対数尤度=-68.2808 は、過分散なしの場合。表 10.17 に一致。

AICC (小さいほどよい)=162.0615 は、過分散なしの場合。表 10.17 に一致。

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	-6.4059	0.2828	-6.9601 -5.8517	513.24	<.0001	
Type	B	1	-0.5433	0.2309	-0.9960 -0.0907	5.54	0.0186	
Type	C	1	-0.6874	0.4279	-1.5260 0.1512	2.58	0.1082	
Type	D	1	-0.0760	0.3779	-0.8166 0.6646	0.04	0.8407	
Type	E	1	0.3256	0.3067	-0.2756 0.9268	1.13	0.2885	
Type	A	0	0.0000	0.0000	0.0000 0.0000	.	.	
Cons	65-69	1	0.6971	0.1946	0.3157 1.0785	12.83	0.0003	
Cons	70-74	1	0.8184	0.2208	0.3857 1.2511	13.74	0.0002	
Cons	75-79	1	0.4534	0.3032	-0.1409 1.0477	2.24	0.1348	
Cons	60-64	0	0.0000	0.0000	0.0000 0.0000	.	.	
Oper	75-79	1	0.3845	0.1538	0.0830 0.6859	6.25	0.0124	
Oper	60-74	0	0.0000	0.0000	0.0000 0.0000	.	.	
尺度		0	1.3004	0.0000	1.3004 1.3004			

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。

注) 推定値は、過分散の（なし，あり）に関係なく一致，表 10.4 および表 10.17 に一致。  
標準偏差は過分散ありで調整，表 10.4 に一致，尺度 1.3004 で除せば過分散なしの場合となる。  
尺度は，過分散パラメータ 1.6910 平方根．．

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
Type	4	25	3.50	0.0212	14.00	0.0073
Cons	3	25	6.19	0.0027	18.57	0.0003
Oper	1	25	6.30	0.0189	6.30	0.0120

注) カイ 2 乗値は，表 10.7 の JMP の過分散がある場合の結果と一致。  
P 値も表 10.7 に一致。

PROC GENMOD から出力された予測値など

Obs	Type	Cons	Oper	Month	Inci	ln_Month	pred	xbeta	s_dev
1	A	60-64	60-74	127	0	4.8442	0.2098	-1.56171	-0.50059
2	A	60-64	75-79	63	0	4.1431	0.1528	-1.87830	-0.42682
3	A	65-69	60-74	1095	3	6.9985	3.6319	1.28975	-0.28332
4	A	65-69	75-79	1095	4	6.9985	5.3346	1.67422	-0.51881
39	E	75-79	60-74	.	.	.	.	.	.
40	E	75-79	75-79	542	1	6.2953	2.8658	1.05284	-1.07973

Proc GENMOD の Output ステートメントで出力された xbeta は推定損傷数の対数，pred は xbeta の指数で推定損傷数，s\_dev は，スチューデント化デビアンズ残差である．このデ

ータを Excel に取り込み、損傷数千月比を計算することにより、第 10.5 節の表 10.15 および図 10.11 で示した交互作用の新しい検討のためのグラフの作成が可能となる。

スチューデント化デビアンズ残差 (s\_dev) についての統計量を、Proc UNIVARIATE で計算した結果の一部を次に示す。これにより、過分散を設定した主効果モデルに対する残差の評価が行なえる。最大値が 2.31407、最小値が-1.76687 となり、外れ値はないものと確認される。

s\_dev についての PROC UNIVARIATE から出力されたリスト

極値			
最小値		最大値	
値	Obs	値	Obs
-1.76687	22	1.01410	10
-1.45675	32	1.32054	35
-1.33261	38	1.65314	36
-1.32681	19	1.76877	30
-1.07973	40	2.31407	21

## 負の 2 項回帰

GENMOD プロシジャで負の 2 項回帰を行うためには、分布の設定で dist=negbin とするだけで実行できる。

```
Title2 '<<< 負の 2 項回帰 offset >>>' ;
proc genmod data=d01 ;
  class Type Cons Oper / ref=first ;
  model Inci = Type Cons Oper
    / dist=negbin link=log type3 offset=ln_Month ;
run ;
```

残念ながら、以下のメッセージがあり、結果を示すことができない変数を絞ると収束するので、データ数に対してパラメータ数が多いためと考えられる。

WARNING: 相対的な Hessian 収束基準 5.8797664802 は限界値 0.0001 を超えています。%w 収束は疑わしいです。

## 第10章 文献索引

高橋(2002) - GENMODプロシジャによる計数データの解析	354
高波・舟尾(2016) - SAS Studioによるやさしい統計データ分析	354
McCullagh and Nelder(1989) - Generalized Linear Models 2nd ed.	323

## 第10章 索引

あ R and SAS - 臨床評価研究会(ACE) (2017)	354	か covbオプション - GENMODプロシジャ	337
R言語 - Trellis(格子)グラフ	349	さ 最後の水準を -1 - 対比型デザイン行列	332
浮き彫り - 特異的な変動	347	最後の水準を基準 - SAS	331
運行数 ni - オフセット	325	最初的水準 - ref=first	331
Excel - 折れ線グラフ	337, 350	最初的水準を基準 - (0,1)型デザイン変数	327
- 交互作用	345	最小2乗平均 - Lsmeans	329
- デザイン行列	329	最大化 - ソルバー	353
- 予測プロファイル	337, 345	最尤法 - ソルバー	352
S - 統計ソフト	327	SAS - OnDemand SAS	354
S-PLUS - Trellis(格子)グラフ	348	- 最後の水準を基準	331
Lsmeans - 最小2乗平均	329	- GENMODプロシジャ	354
オフセット - 運行数 ni	325	- デザイン変数	328
- 切片	326	- DATAステップ	354
- ソルバー	351	- 統計ソフト	328
オフセット offset - GENMOD	355	- PROCステップ	354
折れ線グラフ - Excel	337, 350	- ref=first	331
- 95%信頼区間	341	SAS and R - 臨床評価研究会(ACE) (2017)	354
OnDemand SAS - 無償版	354	SAS無償版 - 高波・舟尾(2016)	354
か 解釈 - パラメータの推定値	330	3因子 - 要因配置型	323
外部ファイル - 予測値	349	GENMODプロシジャ - オフセット offset	355
カウント・データ - 損傷数	323	- 過分散 scale= pearson	355
過分散 scale= pearson - GENMODプロシジャ	355	- covbオプション	337
過分散 - 分散/平均	324	- SAS	354
- ポアソン回帰	355	- 高橋(2002)	354
過分散なし - 交互作用	342	- 負の2項回帰	358
貨物船 - McCullaghら(1989)	323	- 分布 dist=negbin	358
貨物船の前方部の損傷数 - ロイド	323	- 分布 dist=poisson	355
船舶の前方部 - 損傷数	324	- ポアソン回帰	356
95%信頼区間 - 折れ線グラフ	341	- リンク link=log	355
- 分散	340	質的変数 - ダミー変数	327
- 予測プロファイル	340	- デザイン変数	327
共分散行列 - JMP	338	JMP - 共分散行列	338
- 2次形式	339	- グラフ・ビルダー	347
- パラメータ	337	- 交互作用プロファイル	336
グラフ・ビルダー - JMP	347	- 推定値の共分散	337
- 主効果予測値	348	- 対比(1, -1)型	331
- 損傷千月比	348	- デザイン変数	328
GLIM - 統計ソフト	327	- 予測プロファイル	334
交互作用 - Excel	345	重回帰のモデル式 - ポアソン回帰	325
- 過分散の解消	342	主効果モデル - 交互作用	343
- 主効果モデル	343	- 主効果予測値	346
- 分散分析表	346	- ソルバー	351
- ポアソン回帰	342	- 対比型デザイン変数	334
- McCullaghら(1989)	342	- 探索的な解析	346
- 目視的に解釈	346	- ポアソン回帰	327
- 予測プロファイル	344	主効果予測値 - グラフ・ビルダー	348
交互作用プロファイル - JMP	336	- 主効果モデル	346
- 予測プロファイル	334	- 損傷千月比	346
誤差範囲 - 中心点からの距離	341	初期値 - 対数尤度	352

信頼区間付き - 予測プロファイル	341	は 分布 dist=negbin - GENMODプロシジャ	358
推定値の共分散 - JMP	337	- 負の2項回帰	358
推定値の分散 - 2次形式	339	分布 dist=poisson - GENMODプロシジャ	355
stdresdev=s - スチューデント化デビアン	355	分散 - 95%信頼区間	340
スチューデント化デビアン - stdresdev=s	355	分散/平均 - 過分散	324
切片 - オフセット	326	分散分析表 - 交互作用	346
(0,1)型デザイン変数 - 最初の水準を基準	327	ポアソン回帰 - 過分散	355
ソルバー - オフセット	351	- 交互作用	342
- 最大化	353	- GENMODプロシジャ	356
- 最尤法	352	- 主効果モデル	327
- 主効果モデル	351	- ソルバー	352
- 対数尤度	353	- 損傷数	329
- ポアソン回帰	352	- 対比型, 過分散	333
損傷数 - カウント・データ	323	- デザイン行列	329
- 船舶の前方部	324	- ロジスティック回帰	324
- ポアソン回帰	329	ポアソン回帰のモデル式 - 重回帰	325
損傷数データ - デザイン行列	328	ま McCullaghら(1989) - 貨物船	323
損傷千月比 - グラフ・ビルダー	348	- 交互作用	342
- 主効果予測値	346	- ダミー変数	327
た 対数尤度 - 初期値	352	- 統計ソフト	327
- ソルバー	353	Minitab - 統計ソフト	327
対比(1, -1)型 - JMP	331	無償版 - OnDemand SAS	354
対比型, 過分散 - ポアソン回帰	333	目視的に解釈 - 交互作用	346
対比型デザイン行列 - 最後の水準を -1	332	や 要因配置型 - 3因子	323
対比型デザイン変数 - 主効果モデル	334	予測プロファイル - Excel	337, 345
高波・舟尾(2016) - SAS無償版	354	- 95%信頼区間	340
高橋(2002) - GENMODプロシジャ	354	- 交互作用	344
ダミー変数 - 質的変数	327	- 交互作用プロファイル	334
- デザイン変数	327	- JMP	334
- McCullaghら(1989)	327	- 信頼区間付き	341
探索的な解析 - 主効果モデル	346	予測値 - 外部ファイル	349
中心点からの距離 - 誤差範囲	341	ら リンク link=log - GENMODプロシジャ	355
デザイン行列 - Excel	329	臨床評価研究会(ACE)(2017) - R&SAS	354
- 損傷数データ	328	- SAS&R	354
- ポアソン回帰	329	ref=first - 最初の水準	331
デザイン変数 - SAS	328	- SAS	331
- 質的変数	327	ロイド - 貨物船の前方部の損傷数	323
- JMP	328	ロジスティック回帰 - ポアソン回帰	324
- ダミー変数	327		
DATAステップ - SAS	354		
- プログラミング機能	354		
統計ソフト - S	327		
- GLIM	327		
- SAS	328		
- McCullaghら(1989)	327		
- Minitab	327		
特異的な変動 - 浮き彫り	347		
Trellis(格子)グラフ - R言語	349		
- S-PLUS	348		
な 2次形式 - 共分散行列	339		
- 推定値の分散	339		
は パラメータ - 共分散行列	337		
パラメータの推定値 - 解釈	330		
負の2項回帰 - GENMODプロシジャ	358		
- 分布 dist=negbin	358		
PROCステップ - SAS	354		
プログラミング機能 - DETAステップ	354		



## 第 10 章 解析用ファイル一覧表

	8 KB	第10章01_Ship Damage_データ	JMP Data Table
	18 KB	第10章01_Ship Damage_データ	Microsoft Excel ワークシート
	12 KB	第10章02_Ship Damage_主効果	JMP Data Table
	30 KB	第10章02_Ship Damage_主効果	Microsoft Excel ワークシート
	13 KB	第10章03_Ship Damage_共分散	JMP Data Table
	27 KB	第10章03_Ship Damage_予測プロファイル	Microsoft Excel ワークシート
	18 KB	第10章04_Ship Damage_交互作用	JMP Data Table
	21 KB	第10章04_Ship Damage_交互作用	Microsoft Excel ワークシート
	12 KB	第10章05_Ship Damage_主効果_予測	JMP Data Table
	30 KB	第10章05_Ship Damage_主効果_予測値	Microsoft Excel ワークシート
	20 KB	第10章06_Ship Damage_ソルバー	Microsoft Excel ワークシート
	10 KB	第10章06_Ship Damage_主効果	JMP Data Table
	2 KB	第10章07_Ship Damage_SAS	テキストドキュメント

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 <10 章>

**第 10 章 オフセットを含む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 9 日 高橋 行雄