

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5月8日

第3章 尤度比検定のためのデザイン行列

ポアソン分布を仮定した2群間の尤度比検定については、第1.6節でポアソン回帰による方法を示した。尤度比検定は、身近な 2×2 の分割表に対する検定の一つとして良く知られおり、ポアソン分布を仮定した2群間の尤度比検定の考え方と対比する。ポアソン回帰を使用して様々な解析を行うためには、回帰分析の基礎となるデザイン行列について習熟し、目的とする比較を行うために適合するデザイン行列を自ら生成することが必要である。そのために、 2×2 の要因配置型のデータに対してポアソン回帰を行うために必要な各種のデザイン行列について基礎的な考え方に基づいた応用方法を示す。さらに、2本のポアソン回帰直線をあてはめる場合に、切片が共通の場合、傾きが共通の場合、交互作用を検討する場合、別々の直線をあてはめる場合、などに必要なデザイン行列の型の選択方法について示す。

第3章 目次

3.	尤度比検定のためのデザイン行列	95
3.1.	2×2 の分割表に対する尤度比検定の基礎	95
3.2.	一般化線形モデルで二項分布を仮定した2群間比較	100
3.3.	ポアソン回帰を用いた2群間の比較	104
3.4.	2×2 の要因配置モデルに対する各種のデザイン行列	108
3.5.	2本の回帰直線に対する各種のデザイン行列	119
3.6.	オフセットを含む対数リンクでの2本の2次曲線のあてはめ	125
	文献索引, 索引, 解析用ファイル 一覧	135

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解析 -----	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分析 -----	293
10. オフセットを含む探索的ポアソン回帰-----	321
11. デビアンس・逸脱度・残差・テコ比-----	359
12. パラメータの共分散行列の活用 -----	338
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

3. 尤度比検定のためのデザイン行列

ポアソン分布を仮定した 2 群間の尤度比検定については、第 1.6 節でポアソン回帰による方法を示した。尤度比検定は、身近な 2×2 の分割表に対する検定の一つとして良く知られおり、ポアソン分布を仮定した 2 群間の尤度比検定の考え方と対比する。ポアソン回帰を使用して様々な解析を行うためには、回帰分析の基礎となるデザイン行列について習熟し、目的とする比較を行うために適合するデザイン行列を自ら生成することが必要である。そのために、 2×2 の要因配置型のデータに対してポアソン回帰を行うために必要な各種のデザイン行列について基礎的な考え方に基づいた応用方法を示す。さらに、2 本のポアソン回帰直線をあてはめる場合に、切片が共通の場合、傾きが共通の場合、交互作用を検討する場合、別々の直線をあてはめる場合、などに必要なデザイン行列の型の選択方法について示す。

3.1. 2×2 の分割表に対する尤度比検定の基礎

尤度比検定は、 2×2 の分割表に対する Pearson のカイ 2 乗検定と同様に良く知られている。そこで、 2×2 の分割表に対するこれらの検定方法を復習し、ポアソン分布を仮定した 2 群間の尤度比検定の基礎となる考え方を示す。表 3.1 に示した 2×2 の分割表は、第 1.6 節の表 1.15 の 1 日当たりの犯罪件数データを犯罪の (0:なし, 1:あり) にまとめ直したものである [アルトマン (1999)]。

表 3.1 インド 3 地域の 5 年間の満月と新月の日に起きた犯罪の件数

	犯罪		計		犯罪		計
	0:なし	1:あり			0:なし	1:あり	
1:満月	40	143	183	1:満月	n_{11}	n_{12}	$n_{1\cdot}$
2:新月	114	72	186	2:新月	n_{21}	n_{22}	$n_{2\cdot}$
計	154	215	369	計	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot \cdot}$

分割表に対する 2 種類の検定

JMP の「二変量の関係」を用い、 2×2 の分割表としての解析結果を表 3.2 に示す。尤度比のカイ 2 乗値が 60.9396、Pearson のカイ 2 乗値が 58.9845 として出力されている。これらのカイ 2 乗値が、自由度 1 のカイ 2 乗分布に従うことから p 値が求められている。

表 3.2 満月と新月の日に起きた犯罪の有無に対する 2×2 の分割表に対する検定

		犯罪		
度数 期待値		0:なし	1:あり	合計
月	1:満月	40	143	183
		76.374	106.626	
日	2:新月	114	72	186
		77.626	108.374	
合計		154	215	369

	N	自由度	(-1)*対数尤度	R2乗(U)
	369	1	30.4698	0.1215

検定	カイ2乗	p値(Prob>ChiSq)
尤度比	60.9396	<.0001*
Pearson	58.9845	<.0001*

どちらの検定も、2×2 の分割表に対する期待度数 $\hat{\mu}_{ij}$ を用いて定式化されている。期待度数は、周辺の度数を用いて計算される。ここで、 $n_{i\cdot}$ は行方向の和、 $n_{\cdot j}$ は列方向の和、 $n_{\cdot\cdot}$ はセル全体の和である。期待度数 $\hat{\mu}_{ij}$ は、次のように周辺度数から求められた割合の積から期待割合を算出し、全体の数 $n_{\cdot\cdot}$ を掛けて期待度数を算出している。

$$\hat{\mu}_{ij} = \frac{n_{i\cdot}}{n_{\cdot\cdot}} \cdot \frac{n_{\cdot j}}{n_{\cdot\cdot}} \cdot n_{\cdot\cdot} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n_{\cdot\cdot}}, \quad i=1,2, \quad j=1,2 \quad (3.1)$$

Pearson のカイ 2 乗検定統計量は、各セルの実現値 n_{ij} と期待度数 $\hat{\mu}_{ij}$ の差の平方を期待度数 $\hat{\mu}_{ij}$ で割り、全てのセルの和として定義されている。

$$\begin{aligned} \chi_{Pearson}^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \\ &= \frac{(40 - 76.374)^2}{76.374} + \dots + \frac{(72 - 108.374)^2}{108.374} \\ &= 17.3235 + \dots + 12.2083 \\ &= 58.9845 \end{aligned} \quad (3.2)$$

簡便な計算公式として、セルの対角要素の積の差を平方し $n_{\cdot\cdot}$ を掛けて、全ての周辺の和で割った計算式

$$\begin{aligned} \chi_{Pearson}^2 &= \frac{(n_{11}n_{22} - n_{21}n_{12})^2 n_{\cdot\cdot}}{n_{1\cdot}n_{\cdot 1}n_{\cdot 2}n_{2\cdot}} \\ &= \frac{(40 \times 72 - 114 \times 143)^2 \times 369}{183 \times 186 \times 154 \times 215} \\ &= 58.9845 \end{aligned} \quad (3.3)$$

も良く知られている。

尤度比カイ 2 乗値は、各セルの実現値 n_{ij} を期待度数 $\hat{\mu}_{ij}$ で割り、対数を取り、実現値 n_{ij} を掛け、全てのセルの和とする計算手順が知られている。この手順は、多くの統計の教科書でも示されていて、もちろん統計ソフトでの計算手順の解説でもこの方法が示されている。

$$\begin{aligned}
\chi^2_{likelihood} &= 2 \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \ln \left(\frac{n_{ij}}{\hat{\mu}_{ij}} \right) \\
&= 2 \left[40 \times \ln \left(\frac{40}{76.3740} \right) + \dots + 72 \times \ln \left(\frac{72}{108.3740} \right) \right] \quad (3.4) \\
&= (-51.7410) + \dots + (-58.8848) \\
&= 60.9396
\end{aligned}$$

これらの式を用いて Excel によって計算した結果を表 3.3 に示す。この結果は、もちろん表 3.2 に示した JMP による計算結果と一致する。

表 3.3 Excel による期待度数から求めた Pearson および尤度比のカイ 2 乗値

	0:なし	1:あり	計	期待度数		Pearson		尤度比	
1:満月	40	143	183	76.3740	106.6260	17.3235	12.4085	-51.7410	83.9459
2:新月	114	72	186	77.6260	108.3740	17.0441	12.2083	87.6194	-58.8848
計	154	215	369			総和	58.9845	総和	60.9396

期待度数 $\hat{\mu}_{ij}$ を用いた計算手順は、簡潔ではあるが、表 1.19 で示したポアソン分布を仮定した 2 群間の尤度比検定の計算の考え方とは異なる。式 (3.4) は、簡便な計算公式として広く使われている。

出現確率を用いた尤度比検定

第 1.6 節の表 1.19 で計算したポアソン分布を仮定した尤度比検定と同様に、 2×2 の分割表に対しても出現確率を求めて、ポアソン分布と同様な方法を試みる。表 3.4 に示すように満月における 1 日当たりの犯罪が [1:あり] の確率は、 $P_{\text{満月},1} = 143/183 = 0.7814$ であり、143 日の確率は、 $P_{\text{満月},1} = 0.7814^{143}$ となる。犯罪が「0:なし」の 40 日の確率は、 $P_{\text{満月},0} = 0.2186^{40}$ である。満月の日の対数尤度の和 $\ln L_{\text{満月}}$ は、

$$\begin{aligned}
\ln L_{\text{満月}} &= \ln(0.2186^{40}) + \ln(0.7814^{143}) \\
&= -60.8243 - 35.2697 = -96.0940 \quad (3.5)
\end{aligned}$$

表 3.4 出現率に対する対数尤度

犯罪	満月			新月			満月 + 新月		
	$n_{\text{満月}}$	$P_{\text{満月}}$	$\ln L_{\text{満月}}$	$n_{\text{新月}}$	$P_{\text{新月}}$	$\ln L_{\text{新月}}$	$n_{\text{満+新}}$	$P_{\text{満+新}}$	$\ln L_{\text{満+新}}$
0:なし	40	0.2186	-60.8243	114	0.6129	-55.8085	154	0.4173	-134.5720
1:あり	143	0.7814	-35.2697	72	0.3871	-68.3338	215	0.5827	-116.1341
計	183		-96.0940	186		-124.1423	369		-250.7061
	$L_{\text{満月}}$		1.8489E-42	$L_{\text{新月}}$		1.2181E-54	$L_{\text{満+新}}$		1.3174E-109
	完全モデル $-96.0940 - 124.1423 = -220.2363$						縮小モデル		

である。ここでの対数尤度の計算は、1日ごとの犯罪が（0:なし，1:あり）の2値反応に対するベルヌーイ分布としての確率を用いている。同様に，新月の $\ln L_{\text{新月}}$ は，

$$\begin{aligned}\ln L_{\text{新月}} &= \ln(0.6129^{114}) + \ln(0.3871^{72}) \\ &= -55.8085 - 68.3338 = -124.1423\end{aligned}\quad (3.6)$$

であり，満月と新月を合わせた $\ln L_{\text{満月+新月}}$ は，

$$\begin{aligned}\ln L_{\text{満月+新月}} &= \ln(0.4173^{154}) + \ln(0.5827^{215}) \\ &= -134.5720 - 116.1341 = -250.7061\end{aligned}\quad (3.7)$$

となる。

尤度比検定は，満月と新月を合わせた全体の犯罪発生率の対数尤度（縮小モデル）と，満月と新月を別々にした犯罪発生率の対数尤度（完全モデル）の差の2倍を検定統計量としていいる。別々の満月と新月の対数尤度を加えた完全モデルは，

$$\begin{aligned}\ln L_{\text{満月, 新月}} &= \ln L_{\text{満月}} + \ln L_{\text{新月}} \\ &= -96.0940 - 124.1423 = -220.2363\end{aligned}\quad (3.8)$$

であり，縮小モデルの対数尤度 $\ln L_{\text{満月+新月}} = -250.7061$ との差の2倍の $\chi^2_{\text{likelihood}}$ は，

$$\begin{aligned}\chi^2_{\text{likelihood}} &= 2 \times [(-220.2363) - (-250.7061)] \\ &= 2 \times 30.4698 = 60.9396\end{aligned}\quad (3.9)$$

となる。このように，完全モデルと縮小モデルの確率の積の対数を基本にした尤度比検定の結果と，表3.3に示した 2×2 の分割表に対する尤度比検定の結果が一致することが確認される。

尤度比検定は「比」と言っているのに，実際は「差」としている。元々は，尤度の比について対数を取り2倍した統計量が，カイ2乗分布に従うことから，伝統的に「尤度比」と言われている。尤度比の形式では，

$$\begin{aligned}\chi^2_{\text{likelihood}} &= 2 \ln \left(\frac{\text{完全モデル}}{\text{縮小モデル}} \right) \\ &= 2 \ln \left(\frac{L_{\text{満月}} L_{\text{新月}}}{L_{\text{満月+新月}}} \right) \\ &= 2 \ln \left(\frac{1.8489 \times 10^{-42} \times 1.2181 \times 10^{-54}}{1.3174 \times 10^{-109}} \right) \\ &= 60.9396\end{aligned}\quad (3.10)$$

となり， 2×2 の分割表に対する式(3.4)に示した簡便公式の尤度比検定統計量と一致することが確認される。

分割表に対する簡便公式の尤度比検定統計量の誘導

これまで、出現確率を用いた尤度比検定の考え方を示してきたが、尤度比検定の定義に従った計算式は、

$$\chi^2_{likelihood} = 2 \ln \left[\frac{\prod_{i=1}^2 \prod_{j=1}^1 \left(\frac{n_{ij}}{n_{i\cdot}} \right)^{n_{ij}}}{\prod_{j=1}^2 \left(\frac{n_{\cdot j}}{n_{\cdot\cdot}} \right)^{n_{\cdot j}}} \right] \quad (3.11)$$

である。式を変形することにより、分割表の簡便公式による尤度比検定統計量が次のように導出できる。

$$\begin{aligned} \chi^2_{likelihood} &= 2 \ln \left[\frac{\prod_{i=1}^2 \prod_{j=1}^2 \left(\frac{n_{ij}}{n_{i\cdot}} \right)^{n_{ij}}}{\prod_{j=1}^2 \left(\frac{n_{\cdot j}}{n_{\cdot\cdot}} \right)^{n_{\cdot j} + n_{2j}}} \right] \\ &= 2 \ln \left[\prod_{i=1}^2 \prod_{j=1}^2 \left(\frac{n_{ij}}{n_{i\cdot} n_{\cdot j} / n_{\cdot\cdot}} \right)^{n_{ij}} \right] \\ &= 2 \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \ln \left(\frac{n_{ij}}{\hat{\mu}_{ij}} \right) \end{aligned} \quad (3.12)$$

式 (3.12) は、式 (3.4) に一致することから、ポアソン分布を仮定した 2 群間比較に対する尤度比検定の考え方は、 2×2 の分割表における期待度数 $\hat{\mu}_{ij}$ を用いた尤度比検定と同様な方法であることが確認される。

ポアソン分布を仮定した 2 群のカウント・データに対し、表 1.8 に示したようにポアソン回帰により尤度比検定の結果が得られ、表 1.19 に示したように Excel による尤度比検定を試み結果が一致した。同様の考え方でベルヌーイ分布を仮定した 2×2 の分割表に対する式 (3.11) で示した尤度比検定の結果が、いわゆる 2×2 の分割表に対する尤度比検定の結果に一致することを確認できた。これは、式 (3.11) が式 (3.4) に等しくなることを式 (3.12) によって確認した。

分割表に対する尤度比検定の式 (3.4) の導出は、尤度比検定の最も基礎的な課題であるとの認識で、参考になる文献を探して見たが、なかなか見出すことができなかった。唯一、Agresti, A. (2013), *Categorical Data Analysis* 3rd ed の 3.2.1 Pearson and Likelihood-Ratio Chi-Squared Test に分割表の一般式として同様の導出方法が示されている。

3.2. 一般化線形モデルで2項分布を仮定した2群間比較

「満月の日： $x=0$ 」, 「新月の日： $x=1$ 」, 「犯罪のなし： $y=0$ 」, 「犯罪のあり： $y=1$ 」として一般化線形モデルで、2項分布を仮定した2群間比較を行う。JMPの「一般化線形モデル」を使うために表3.5に示すように件数を縦方向に並べ直す。

表 3.5 一般化線形モデルを適用するためのデータリスト

切片	満新	犯罪	件数
μ	x	y	n
1	0	0	40
1	0	1	143
1	1	0	114
1	1	1	72
		計	369

リンク関数は、(ロジット, プロビット, 補2重対数)の3種類あるが、群間比較の場合は、完全フィットするので、尤度比カイ2乗値は、どれでも同じ結果となる。表3.6にリンク関数を「ロジット」とした場合についての尤度比検定の結果を示す。

表 3.6 犯罪の(なし, あり)に対する一般化線形モデル(ロジット・リンク)

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	30.4698	60.9396	1	<.0001*
完全	220.2363			
縮小	250.7061			

「縮小」の行の「(-1)*対数尤度」列の250.7061が、表3.4に示した $\ln L_{\text{満月+新月}} = -250.7061$ に対応し、「完全」の220.2363が、 $\ln L_{\text{満月}} + \ln L_{\text{新月}} = -220.2363$ に対応する。「差分」の行の「尤度比カイ2乗」の60.9396が、式(3.9)の $\chi^2_{\text{likelihood}} = 2 \times 30.4698 = 60.9396$ に対応する。

表3.7に回帰パラメータについての推定値, 標準誤差, 尤度比カイ2乗の結果を示す。「項」の列の「x」の行の「尤度比カイ2乗」の結果も60.9396と表3.6の結果に一致する。

表 3.7 ロジット変換に対する回帰パラメータの推定

パラメータ推定値			
項	推定値	標準誤差	尤度比カイ2乗
切片	-1.2740	0.1789	61.5039
x	1.7335	0.2338	60.9396

推定値は、切片が、「満月の日： $x=0$ 」とした場合の「犯罪のなし： $y=0$ 」のロジット

$$\text{logit} = \ln\left(\frac{p_{\text{満月},y=0}}{1-p_{\text{満月},y=0}}\right) = \ln\left(\frac{0.2186}{1-0.2186}\right) = -1.2740 \quad (3.13)$$

に一致し、「新月の日： $x=1$ 」とした場合には、満月と新月のロジットの差

$$\begin{aligned} \text{logit} &= \ln\left(\frac{p_{\text{新月},y=0}}{1-p_{\text{新月},y=0}}\right) - \ln\left(\frac{p_{\text{満月},y=0}}{1-p_{\text{満月},y=0}}\right) \\ &= \ln\left(\frac{0.6129}{1-0.6129}\right) - (-1.2740) \\ &= 0.4595 + 1.2740 \\ &= 1.7335 \end{aligned} \quad (3.14)$$

となっている。

Excel ソルバーを用いたロジスティック回帰

一般化線形モデルで分布を「2項」、リンク関数を「ロジット」と設定し、ロジスティック回帰による 2×2 の分割表に対する尤度比検定を行い、結果の見方について示した。ポアソン回帰のみならず、ロジスティック回帰も Excel のソルバーを用いて簡単にできることを示す。

ロジスティック回帰は、表 3.5 に示すように、満月の日を $x=0$ 、新月の日を $x=1$ とし、2値反応 y に対して式 (3.13) および式 (3.14) に示したロジット変換 (*logit*) し、直線をあてはめる方法である。図 3.1 (左) に示すように2ポイントしかないのでそれらの点を通る直線となる。もちろん切片は $\beta_0 = -1.2740$ 、傾きは $\beta_1 = 1.7335$ である。

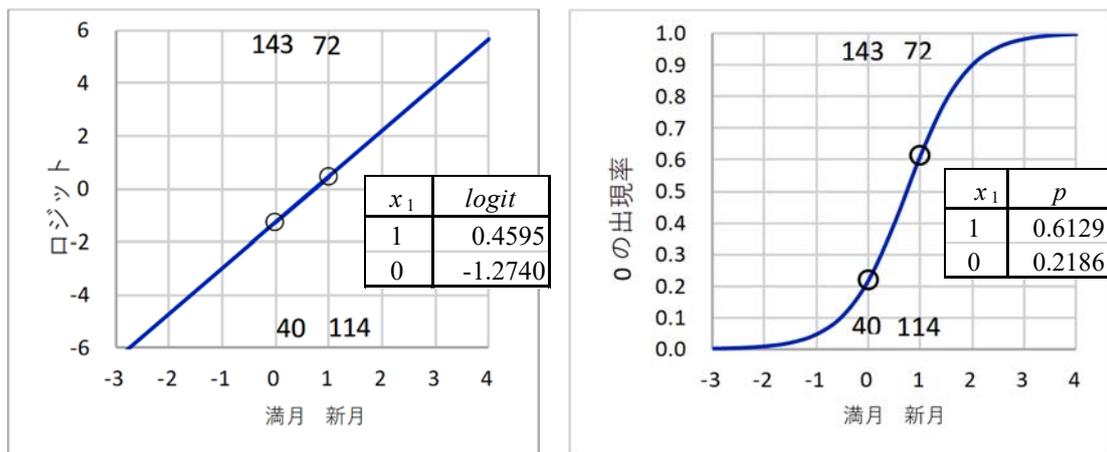


図 3.1 ロジスティック直線およびロジスティック曲線のあてはめ

シグモイド曲線として，ロジスティック分布の累積分布関数 $F(x)$

$$F(x) = \frac{\exp\left(\frac{x-\mu}{\sigma}\right)}{1 + \exp\left(\frac{x-\mu}{\sigma}\right)} = \frac{\exp\left(\frac{\mu}{\sigma} - \frac{1}{\sigma}x\right)}{1 + \exp\left(\frac{\mu}{\sigma} - \frac{1}{\sigma}x\right)} \quad (3.15)$$

$$= \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}$$

ただし， $\beta_0 = \mu/\sigma$ ， $\beta_1 = -1/\sigma$

を使う．表 3.7 で得られた回帰パラメータは， $\beta_0 = -1.2740$ ， $\beta_1 = 1.7335$ なので，図 3.1 (右) に示すように推定値は，

$$\text{満月 } x=0 : P_0 = F(x=0) = \frac{\exp(-1.2740 + 1.7335 \times 0)}{1 + \exp(-1.2740 + 1.7335 \times 0)} = 0.2186$$

$$\text{新月 } x=1 : P_1 = F(x=1) = \frac{\exp(-1.2740 + 1.7335 \times 1)}{1 + \exp(-1.2740 + 1.7335 \times 1)} = 0.6128$$

となる．

表 3.8 に Excel ソルバーを用いた最尤法によるロジスティックス回帰の解析法を示す．初期値を $\hat{\beta}_0 = 0.0$ ， $\hat{\beta}_1 = 0.0$ として

$$\text{logit}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (3.16)$$

表 3.8 Excel ソルバーによるロジスティック回帰

		初期値	β_0^{\wedge}	0.0000			
			β_1^{\wedge}	0.0000			
切片	満新	犯罪	件数	二項分布			
x_0	x_1	y	n	logit	π^{\wedge}	P^{\wedge}	$\ln L_i$
1	0	0	40	0.0000	0.5000	0.5000	-27.7259
1	0	1	143	0.0000	0.5000	0.5000	-99.1200
1	1	0	114	0.0000	0.5000	0.5000	-79.0188
1	1	1	72	0.0000	0.5000	0.5000	-49.9066
計			369			$\ln L =$	-255.7713
		最尤解	β_0^{\wedge}	-1.2740			
			β_1^{\wedge}	1.7335			
切片	満新	犯罪	件数	二項分布			
x_0	x_1	y	n	logit	π^{\wedge}	P^{\wedge}	$\ln L_i$
1	0	0	40	-1.2740	0.2186	0.2186	-60.8243
1	0	1	143	-1.2740	0.2186	0.7814	-35.2697
1	1	0	114	0.4595	0.6129	0.6129	-55.8086
1	1	1	72	0.4595	0.6129	0.3871	-68.3337
計			369			$\ln L =$	-220.2363

$$\hat{\pi}_i = \frac{\exp(\text{logit}_i)}{1 + \exp(\text{logit}_i)} \quad (3.17)$$

$$\hat{P}_i = \text{Binom.dist}(1 - y_i, 1, \hat{\pi}_i, \text{false}) \quad (3.18)$$

$$\ln(L_i) = n_i \ln(\hat{P}_i) \quad (3.19)$$

$$\ln L = \sum_i \ln(L_i) = -255.7713 \quad (3.20)$$

が計算されている。Excel ソルバーで、 $\ln L$ の最大化するために、 $\beta_0 = 0.0$ 、 $\beta_1 = 0.0$ を変化させると $\hat{\beta}_0 = -1.2740$ 、 $\hat{\beta}_1 = 1.7335$ が得られ、対数尤度 $\ln L = -220.2363$ は、表 3.4 の完全モデルに一致する。

切片に対する尤度比検定に対する補足

さて、表 3.7 に示したロジット変換に対する回帰パラメータ推定値で、切片項の推定値が -1.2740 に対し、尤度比カイ 2 乗値が、61.5039 と高度に有意となっている。どのような計算が行われているのだろうか。尤度比検定は、完全モデルの尤度と縮小モデルの尤度の比であるので、モデル式から切片項を除いたモデルであるが、実際にはどのように計算したらよいのであろうか。

切片がない場合のモデルは、「原点を通る直線」をあてはめることになる。ロジット変換した場合にゼロとなる出現率は、0.50 である。確認すると

$$\text{logit} = \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{0.50}{1-0.50}\right) = 0 \quad (3.21)$$

確かにゼロとなる。従って、「満月の日： $x=0$ 」, 「新月の日： $x=1$ 」, としたので、満月の日の犯罪がない率を 0.50 とし、新月の日の犯罪がない率 0.6129 とした場合が縮小モデルとなる。表 3.9 に「切片 x_0 抜き縮小モデル」の対数尤度が -250.9882 となり、完全モデルとの差の 2 倍が、61.5039 となり、表 3.7 の結果に一致する。

表 3.9 切片抜き縮小モデルの対数尤度

切片	満月新月		犯罪	件数	完全モデル		x_1 抜き縮小モデル		x_0 抜き縮小モデル	
	x_0	x_1			P^\wedge	$\ln L_i$	P^\wedge	$\ln L_i$	P^\wedge	$\ln L_i$
1	0	0	40	0.2186	-60.8243	0.4173	-34.9538	0.5000	-27.7259	
1	0	1	143	0.7814	-35.2697	0.5827	-77.2427	0.5000	-99.1200	
1	1	0	114	0.6129	-55.8085	0.4173	-99.6182	0.6129	-55.8085	
1	1	1	72	0.3871	-68.3338	0.5827	-38.8914	0.3871	-68.3338	
			369		-220.2363		-250.7061		-250.9882	
				完全モデルを基準とした対数尤度の差				30.4698		30.7519
				2倍の対数尤度				60.9396		61.5039

3.3. ポアソン回帰を用いた 2 群間の比較

第 1.6 節の表 1.19 で Excel によるポアソン分布を仮定した 2 群間の尤度比検定の結果を説明なしで示した。第 3.2 節で 2×2 の分割で尤度比検定の考え方を示したので、ポアソン分布の場合について拡張する。完全モデルの尤度は、満月および新月の日の犯罪件数に対し、それぞれについてポアソン分布を仮定した場合である。縮小モデルは、満月と新月を合わせた場合にポアソン分布を仮定する。尤度比検定は、それらの比として定義され、仮定する分布が 2 項分布でもポアソン分布でも尤度比検定の考え方は同じである。

$$\chi^2_{likelihood} = 2 \ln \left(\frac{\text{完全モデルでの尤度}}{\text{縮小モデルでの尤度}} \right) \quad (3.22)$$

満月の日の尤度、新月の日の尤度をそれぞれ $L_{\text{満月}}$ と $L_{\text{新月}}$ とし、満月と新月を合算した場合の尤度を $L_{\text{満月+新月}}$ とすると尤度比検定は、

$$\begin{aligned} \chi^2_{likelihood} &= 2 \ln \left(\frac{L_{\text{満月}} L_{\text{新月}}}{L_{\text{満月+新月}}} \right) \\ &= 2 (\ln L_{\text{満月}} + \ln L_{\text{新月}} - \ln L_{\text{満月+新月}}) \\ &= 2 \times [(-268.4776) + (-176.1202) - (-484.8865)] \\ &= 80.5774 \end{aligned} \quad (3.23)$$

となることを示した。この式は、 2×2 の分割の尤度比検定で用いた式 (3.10) と同じであるが、尤度の計算にポアソン分布を仮定した対数尤度を用いている。表 3.10 に JMP のポアソン回帰による縮小モデル、完全モデル、それらの差、および、マイナス 2 倍の対数尤度を示す。

表 3.10 ポアソン回帰による 2 群間比較

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	40.2887	80.5774	1	<.0001*
完全	444.5978			
縮小	484.8865			

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	1.3989	0.0874	256.0000	<.0001*
x	-0.8935	0.1018	80.5774	<.0001*

表 3.11 にポアソン回帰による 2 群間の比較の Excel の計算シートを示す。「満月の日： $x_1 = 0$ 」, 「新月の日： $x_1 = 1$ 」とし、犯罪発現件数を y , 日数を n とする。

完全モデルは,

$$\begin{aligned}\hat{y}_i &= \text{完全}\hat{\beta}_0 x_{0,i} + \text{完全}\hat{\beta}_1 x_{1,i} \\ &= 1.3989 - 0.8935 x_{1,i}\end{aligned}$$

$$\ln L_i = n_i \ln [\text{Poisson.dist}(y_i, \hat{y}_i, \text{false})] \quad (3.24)$$

$$\begin{aligned}\ln L &= -55.9562 - 68.0458 - \dots - 6.4132 \\ &= -444.5978\end{aligned}$$

と計算されている. 切片のみの縮小モデルは,

$$\begin{aligned}\hat{y}_i &= \text{縮小}\hat{\beta}_0 x_{0,i} \\ &= 0.9485\end{aligned} \quad (3.25)$$

$$\ln L_i = n_i \ln [\text{Poisson.dist}(y_i, \hat{y}_i, \text{false})] \quad (3.26)$$

$$\begin{aligned}\ln L &= -37.9404 - 64.0979 - \dots - 4.3380 \\ &= -484.8865\end{aligned} \quad (3.27)$$

と計算されている. 完全モデルの尤度 -444.5978 と縮小モデル -484.8865 との差の 2 倍で, 表 3.11 の最後の行に示すように, 80.5774 となり式 (3.23) と一致する.

傾きだけの縮小モデルは, 計算エラーとなり計算されていない. これは, 切片 β_0 が無いモデルとなり, $x_1 = 0$ の場合にポアソン分布のパラメータが, $\hat{\mu}_i = x_1 = 0$ となり, 1 以上のデータ

表 3.11 犯罪件数に対する (0, 1) 形式でのポアソン回帰

		完全 $\hat{\beta}_0 = 1.3989$		縮小 $\hat{\beta}_0 = 0.9485$		—		—		
		完全 $\hat{\beta}_1 = -0.8935$		—		縮小 $\hat{\beta}_1 = 0.5054$		—		
		完全モデル				x_1 縮小モデル		x_0 縮小モデル		
i	切片 x_0	満新 x_1	件数 y	日数 n	y^{\wedge}	$\ln L_i$	y^{\wedge}	$\ln L_i$	y^{\wedge}	$\ln L_i$
1	1	0	0	40	1.3989	-55.9562	0.9485	-37.9404	0.0000	0.0000
2	1	0	1	64	1.3989	-68.0458	0.9485	-64.0879	0.0000	#NUM!
3	1	0	2	56	1.3989	-79.5576	0.9485	-97.8535	0.0000	#NUM!
4	1	0	3	19	1.3989	-41.4883	0.9485	-55.0783	0.0000	#NUM!
5	1	0	4	1	1.3989	-3.2342	0.9485	-4.3380	0.0000	#NUM!
6	1	0	5	2	1.3989	-9.0159	0.9485	-12.0006	0.0000	#NUM!
7	1	0	9	1	1.3989	-11.1795	0.9485	-14.2261	0.0000	#NUM!
8	1	1	0	114	0.5054	-57.6128	0.9485	-108.1301	0.5054	-57.6129
9	1	1	1	56	0.5054	-66.5184	0.9485	-56.0769	0.5054	-66.5184
10	1	1	2	11	0.5054	-28.1977	0.9485	-19.2212	0.5054	-28.1977
11	1	1	3	4	0.5054	-17.3780	0.9485	-11.5954	0.5054	-17.3780
12	1	1	4	1	0.5054	-6.4132	0.9485	-4.3380	0.5054	-6.4132
					$\ln L =$	-444.5978	$\ln L =$	-484.8865	$\ln L =$	#NUM!
					対数尤度の差=		40.2887		#NUM!	
					2倍の差=		80.5774		#NUM!	

が全くないことになり，ポアソン分布が定義できないために

$$\begin{aligned} (x_0) \text{ 縮小} \quad \ln L_i &= n_i \ln[\text{Poisson.dist}(y_i = 0, \hat{\mu}_i = 0, \text{false})] \\ &= n_i \ln[1] = 0 \end{aligned} \quad (3.28)$$

$$\begin{aligned} (x_0) \text{ 縮小} \quad \ln L_i &= n_i \ln[\text{Poisson.dist}(y_i > 0, \hat{\mu}_i = 0, \text{false})] \\ &= n_i \ln[0] = \text{\#NUM!} \end{aligned} \quad (3.29)$$

計算不能（#NUM!）となるためである．JMP および SAS での計算結果は，表 1.18 および表 1.20 に示したように，尤度比カイ 2 乗が共に，256.00 と出力されている．これは，共分散行列から計算された SE を用いたワルド統計量を便宜的に使っているためと思われる．

計算不能となったのは，満月の日： $x_1 = 0$ ，「新月の日： $x_1 = 1$ 」としたために起きた現象で，表 3.12 に示すように，「満月の日： $x_1 = 0.5$ 」，「新月の日： $x_1 = 1.5$ 」のようなダミー変数とすれば，計算可能となる．

表 3.12 犯罪件数に対する (0.5, 1.5) 形式でのポアソン回帰

		完全 $\beta_0^{\wedge} = 1.8457$		縮小 $\beta_0^{\wedge} = 0.9485$		—		—		
		完全 $\beta_1^{\wedge} = -0.8935$		—		縮小 $\beta_1^{\wedge} = 0.9447$		—		
		完全モデル		x_1 縮小モデル		x_0 縮小モデル				
i	切片 x_0	満新 x_1	件数 y	日数 n	y^{\wedge}	$\ln L_i$	y^{\wedge}	$\ln L_i$	y^{\wedge}	$\ln L_i$
1	1	0.5	0	40	1.3989	-55.9563	0.9485	-37.9404	0.4723	-18.8934
2	1	0.5	1	64	1.3989	-68.0458	0.9485	-64.0879	0.4723	-78.2337
3	1	0.5	2	56	1.3989	-79.5576	0.9485	-97.8535	0.4723	-149.2745
4	1	0.5	3	19	1.3989	-41.4883	0.9485	-55.0783	0.4723	-85.7716
5	1	0.5	4	1	1.3989	-3.2342	0.9485	-4.3380	0.4723	-6.6507
6	1	0.5	5	2	1.3989	-9.0159	0.9485	-12.0006	0.4723	-18.0203
7	1	0.5	9	1	1.3989	-11.1795	0.9485	-14.2261	0.4723	-20.0248
8	1	1.5	0	114	0.5054	-57.6128	0.9485	-108.1301	1.4170	-161.5385
9	1	1.5	1	56	0.5054	-66.5184	0.9485	-56.0769	1.4170	-59.8337
10	1	1.5	2	11	0.5054	-28.1977	0.9485	-19.2212	1.4170	-15.5437
11	1	1.5	3	4	0.5054	-17.3780	0.9485	-11.5954	1.4170	-8.6525
12	1	1.5	4	1	0.5054	-6.4132	0.9485	-4.3380	1.4170	-3.2009
					$\ln L =$	-444.5978	$\ln L =$	-484.8865	$\ln L =$	-625.6383
					対数尤度の差=		40.2887		181.0406	
					2倍の差=		80.5774		362.0811	

この結果が正しいのか，JMP での結果を表 3.13 の最後の行に示すように，切片の尤度比カイ 2 乗は，362.0811 となり，Excel での結果と一致する．この様にポアソン回帰の場合には，名義尺度の与える数値により，結果が異なるので，切片に対する「尤度比カイ 2 乗」は，全く意味をなさないことに注意が必要である．

表 3.13 JMP による犯罪件数に対する (0.5, 1.5) 形式でのポアソン回帰

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	40.2887	80.5774	1	<.0001*
完全	444.5978			
縮小	484.8865			

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	1.8457	0.1337	362.0811	<.0001*
x3	-0.8935	0.1018	80.5774	<.0001*

名義尺度にどのような数値（ダミー変数）を与えるかは、自己責任で自由に設定できる。ただし、切片に対する尤度比カイ2乗値は、計算の原理からは計算不能となる場合もあるが、JMP および SAS では、便宜的な対応をしていることが推測される。

3.4. 2×2の要因配置モデルに対する各種のデザイン行列

デザイン行列に与える変数（ダミー変数）

ポアソン分布を仮定した解析を行うためには、一般化線形モデルで分布をポアソン分布、リンク関数を恒等と設定してポアソン回帰を実施する方法をこれまで示してきた。基本的なポアソン回帰の場合は、

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{ポアソン分布} \quad (3.30)$$

のように、通常の回帰分析と同様な回帰式を用いてきた。この式で β_0 の推定値は、 $x=0$ における y の値なのでY切片となり、 β_1 の推定値は、回帰直線の傾きで x のプラス1増加した場合における y の増分である。なぜ、そのように言えるのか、数式を使って説明する。回帰パラメータの推定値 β_0 と $\hat{\beta}_1$ が得られ、回帰式で

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (3.31)$$

と推定値 \hat{y}_i が得られた場合に、 $x=0$ とした場合に

$$\begin{aligned} \hat{y}_{x=0} &= \hat{\beta}_0 + \hat{\beta}_1 x \\ &= \hat{\beta}_0 + \hat{\beta}_1 \times 0 = \hat{\beta}_0 \end{aligned} \quad (3.32)$$

推定値は、 $\hat{y}_{x=0} = \hat{\beta}_0$ で、いわゆるY切片である。次に x_i に適当な値として $x=2$ を代入した場合、

$$\begin{aligned} \hat{y}_{x=2} &= \hat{\beta}_0 + \hat{\beta}_1 \times 2 \\ &= \hat{y}_{x=0} + 2\hat{\beta}_1 \end{aligned} \quad (3.33)$$

となり、 $\hat{\beta}_1$ について解くと

$$\hat{\beta}_1 = \frac{\hat{y}_{x=2} - \hat{y}_{x=0}}{2} \quad (3.34)$$

が得られる。これは、 x が2の場合の $\hat{y}_{x=2}$ から、 x が0の場合の $\hat{y}_{x=0}$ を差し引いて2で割っており、 $\hat{\beta}_1$ は、 x が1増加した場合の \hat{y} の増分であり、回帰分析における「傾き」である。

第1.6節の満月と新月の犯罪件数では、2群間の比較をする際に、名義尺度に対して満月の場合に $x=0$ 、新月の場合に $x=1$ としてポアソン回帰を行った。もちろん次式で β_0 と β_1 を推定するのであるが、

$$\begin{aligned} x=0 \text{ の場合} &: \hat{\beta}_0 = \hat{y}_{\text{満月}} \\ x=1 \text{ の場合} &: \hat{\beta}_1 = \hat{y}_{x=1} - \hat{y}_{x=0} = \hat{y}_{\text{新月}} - \hat{y}_{\text{満月}} \end{aligned} \quad (3.35)$$

となり、(切片・傾き)とは、言い難くなる。

第 1.4 節で反復重み付き回帰式を，事前の説明なしに，

$$\hat{\beta}^{(m)} = \left[(X^T \hat{W} X)^{(m-1)} \right]^{-1} (X^T \hat{W} \hat{Z})^{(m-1)} \quad (3.36)$$

として示した．この式の中の X 行列は，計画行列またはデザイン行列と言われており，ポアソン回帰による各種の推定を行う際に中心的な役割を果たす．なお，重み \hat{W} の対角要素が全て 1 の場合には， $X^T \hat{W} = X^T$ なので，通常の線形回帰式

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3.37)$$

となる．

第 1.4 節の表 1.8 では，デザイン行列 X を， 9×2 の矩形データとし，Excel の行列関数を用いて，回帰パラメータを推定した．

$$X = \begin{array}{|c|c|c|} \hline & x_0 & x_1 \\ \hline & 1 & -1 \\ \hline & 1 & -1 \\ \hline & 1 & 0 \\ \hline & 1 & 1 \\ \hline & 1 & 1 \\ \hline & 1 & 1 \\ \hline \end{array}$$

一般的な回帰分析では，切片を含めない変数のみを設定するが，デザイン行列を用いた回帰式を用いる場合は，

$$y_i = \beta_0 x_{0,i} + \beta_1 x_{1,i} + \varepsilon_i \quad \varepsilon_i \sim Normal(0, \sigma), \quad i = 1, 2, \dots, n \quad (3.38)$$

のように切片を説明変数として $x_{0,i} = 1$ を含めた式とする必要がある．

2×2 の要因配置実験

第 1.7 節の表 1.22 に示した Ames 試験におけるコロニー数については， 2×2 の要因配置デザインと見なすこともでき，表 3.14 に結果のサマリーを示す．各セルの平均は異なるが，それぞれの分散の比が 1 に近いことから，全体としてポアソン分布に従うと判断される [吉村ら (1992)]．

この実験を 2 因子実験と見なすならば，主効果を (A : 溶媒, B : 代謝活性化)，交互作用として (A×B) を含めて解析することができる．因子 B の効果は，統計的に云々する必要がないくらい明らかに B_1 に比べて B_2 が小さいが，因子 A の効果は， B_1 内では A_1 に対し A_2 は減少傾向であり， B_2 内では A_1 に対し A_2 は増加傾向で統計的な判断が必要となる．

表 3.14 A:溶媒と B:代謝活性化の組合せ結果 (各セル, n=50)

A:溶媒	B:代謝活性化			
	1:なし		2:あり	
	平均	分散	平均	分散
1:蒸留水	14.54	17.23	7.54	6.34
2:DMOS	12.48	11.64	8.28	6.80

表 3.14 に示した 2×2 の実験データについてポアソン回帰による 2 元配置型の解析を行なう。JMP のモデル効果の構成は、通常の二元配置分散分析に準じて設定する。表 3.15 にパラメータの推定結果を示す。

表 3.15 JMP による対比型のポアソン回帰の結果

モデル効果の構成

追加 A:溶媒
交差 B:活性化
A:溶媒*B:活性化

手法: 一般化線形モデル
分布: Poisson
リンク関数 恒等

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	78.8560	157.7120	3	<.0001*
完全	507.2925			
縮小	586.1485			

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	10.7100	0.2314	2142.0000	<.0001*
A:溶媒[A1]	0.3300	0.2314	2.0339	0.1538
B:活性化[B1]	2.8000	0.2314	148.1204	<.0001*
A:溶媒[A1]*B:活性化[B1]	0.7000	0.2314	9.1569	0.0025*

ここで、切片=10.7100 は、何を意味しているのだろうか。A:溶媒[A1]=0.3300 は何を意味しているのだろうか。表 3.16 には、2×2 表について A:溶媒についての平均、B:代謝活性化に

表 3.16 セル平均の平均と総平均からの差

	B:代謝活性化		全体	
	1:なし	2:あり		
A:溶媒	平均	平均	平均の平均	効果(差)
1:蒸留水	14.54	7.54	11.04	0.33
2:DMOS	12.48	8.28	10.38	-0.33
平均の平均	13.51	7.91	10.71	
効果(差)	2.80	-2.80		

ついでに平均、さらに全体の平均 10.71、全体平均から差を効果（差）の計算結果が示されている。これから、切片= 10.71 は、総平均であると推測でき、

$$A:\text{溶媒}[A1] = 0.33$$

は、 $A_1=11.04$ から全体の平均 10.71 との差が 0.33 であり、 A_1 の平均と A_2 の平均の差の 2 分の 1 と一致する。

$$\text{因子 } A_1 \text{ の効果} : \frac{(14.54+7.54)/2-(12.48+8.28)/2}{2} = \frac{0.66}{2} = 0.33$$

同様に、

$$B:\text{活性化}[B1] = 2.80$$

は、 B_1 の平均 13.51 と総平均 10.71 の差であり、 B_1 の平均と B_2 の平均の差の 2 分の 1 と一致する。

$$\text{因子 } B_1 \text{ の効果} : \frac{(14.54+12.48)/2-(7.54+8.28)/2}{2} = \frac{5.60}{2} = 2.80$$

では、交互作用

$$A:\text{溶媒 } [A1]*B:\text{活性化}[B1] = 0.70$$

は、何を意味しているのであろうか。試行錯誤的に検討すると、たすき掛けで加えた平均の差の平均

$$\text{交互作用 } 1 : \frac{(14.54+8.28)/2-(12.48+7.54)/2}{2} = \frac{1.40}{2} = 0.70$$

に一致する。更に、表 3.17 に示すように A_1 内の B_1 と B_2 の差の平均、 A_2 内の B_1 と B_2 の差の平均、それらの差の平均の平均で

$$\text{交互作用 } 2 : \frac{(14.54-7.54)/2-(12.48-8.28)/2}{2} = \frac{1.40}{2} = 0.70$$

同じ結果が得られるが、式を変形すれば、互いに同じ式になることが確認できる。

表 3.17 セル平均の差の差

	B:代謝活性化		全体 差/2
	1:なし 平均	2:あり 平均	
A:溶媒 1:蒸留水	14.54	7.54	3.50
2:DMOS	12.48	8.28	2.10
差/2	1.03	-0.37	1.40
	差の差の効果		0.70

交互作用の効果としてプラスの効果 0.70 が得られたが、効果としてはマイナスの効果がないとバランスしない。たすき掛けの $(A_1B_1 \times A_2B_2) - (A_1B_2 \times A_2B_1)$ 差を第 1 項と第 2 項を

入れ替えると

$$\text{交互作用 3 : } \frac{(12.48+7.54)/2-(14.54+8.28)/2}{2} = \frac{-1.40}{2} = -0.70$$

マイナスの効果が出てくる。これらから、表 3.18 に示すように交互作用の効果としては、

表 3.18 交互作用の効果

A:溶媒	B:代謝活性化		和
	1:なし	2:あり	
1:蒸留水	0.70	-0.70	0.00
2:DMOS	-0.70	0.70	0.00
和	0.00	0.00	0.00

これらの効果を用いて 2×2 のセル平均を求めてみると、次のようにセル平均を求めることができる。

$$A_1B_1 = \text{総平均} + A_1 \text{効果} + B_1 \text{効果} + \text{交互作用}_{11} = 10.71 + 0.33 + 2.80 + 0.70 = 14.54$$

$$A_1B_2 = \text{総平均} + A_1 \text{効果} + B_2 \text{効果} + \text{交互作用}_{12} = 10.71 + 0.33 - 2.80 - 0.70 = 7.45$$

$$A_2B_1 = \text{総平均} + A_2 \text{効果} + B_1 \text{効果} + \text{交互作用}_{21} = 10.71 - 0.33 + 2.80 - 0.70 = 12.48$$

$$A_2B_2 = \text{総平均} + A_2 \text{効果} + B_2 \text{効果} + \text{交互作用}_{22} = 10.71 - 0.33 - 2.80 + 0.70 = 8.28$$

このような探索的な見当は、統計ソフトの出力を理解し解釈するために欠かせない。ここでの結果は、JMP 一般線形モデルの出力に対する結果であるが、SAS の GENMOD プロシジャでは、第 13.4 節で示すように全く異なる。統計ソフトの結果の解釈には、統計ソフトの名義尺度に対する数値の与え方を確認することが必須である。

対比型デザイン行列

前項では、JMP のポアソン回帰の解析で出力されるパラメータの推定値に対して探索的な見当により意味づけを行った。この計算プロセスを行列で整理してみよう。セル平均を表 3.19 に示す順番で縦方向に展開し、行と列の入れ替えをする転置記号 $[\dots]^T$ を用いてセル平均 y_i を列ベクトル

$$\mathbf{y} = [y_1 \ y_2 \ y_3 \ y_4]^T = [14.54 \ 7.54 \ 12.48 \ 8.28]^T$$

とする。デザイン行列は、前項でセル平均の計算でプラスとなっている場合を 1 とし、マイナスとなっている場合を -1 と置き換えたものをデザイン行列 \mathbf{X} としている。

デザイン行列の列ベクトル \mathbf{x}_0 は、全て 1 で「切片」ともいわれるが、その役割は他のデザイン行列の設定によって微妙に異なる。列ベクトル \mathbf{x}_1 は、因子 A の水準 A_1 に対して +1, A_2

に対して -1 のように足して 0 となる数値のセットを対比型という。列ベクトル x_2 は、因子 B の水準 B_1 に対して +1, B_2 に対して -1 と対比型で設定する。列ベクトル x_3 は、因子 A と因子 B の交互作用で、 $x_3 = x_1 \cdot x_2$ のように積で計算されている。

表 3.19 対比型 (1, -1) のデザイン行列

A:溶媒	B:代謝 活性化	平均	y	X				β	ε	β^{\wedge}
				x_0	x_1	x_2	x_3			
1:蒸留水	1:なし	14.54	y_1	1	1	1	1	β_0	ε_1	10.71
	2:あり	7.54	y_2	1	1	-1	-1	β_1	ε_2	0.33
2:DMOS	1:なし	12.48	y_3	1	-1	1	-1	β_2	ε_3	2.8
	2:あり	8.28	y_4	1	-1	-1	1	β_3	ε_4	0.7

推定値 $\hat{\beta}$ は、[総平均 A₁効果 B₁効果 交互作用効果] で、
で

$$\hat{\beta} = [\hat{\beta}_0 \ \hat{\beta}_1 \ \hat{\beta}_2 \ \hat{\beta}_3]^T = [10.71 \ 0.33 \ 2.80 \ 0.70]^T$$

のように推定されている。セル平均 \hat{y}_i は、 $\hat{y} = X\hat{\beta}$ で計算されていて、以下に示すように展開することができる。

$$\hat{y}_1 = +10.71 + 0.33 + 2.80 + 0.70 = 14.54$$

$$\hat{y}_2 = +10.71 + 0.33 - 2.80 - 0.70 = 7.45$$

$$\hat{y}_3 = +10.71 - 0.33 + 2.80 - 0.70 = 12.48$$

$$\hat{y}_4 = +10.71 - 0.33 - 2.80 + 0.70 = 8.28$$

この行列 X を一般的には、デザイン行列または計画行列とも言われている。行列計算になれていない場合は、先に第 4 章で統計計算に必要な行列計算の基礎を学習してもらいたい。

デザイン行列の行方向は、反応変数の並びである。デザイン行列の列方向は、説明変数ベクトル x_j で表すと j 列目は、 x_{ij} となり、以下のようにも表すことができる。

$$\begin{aligned} \hat{y}_1 &= 10.71 x_{0,1} + 0.33 x_{1,1} + 2.80 x_{2,1} + 0.70 x_{3,1} = 14.54 \\ & \vdots \\ \hat{y}_4 &= 10.71 x_{0,4} + 0.33 x_{1,4} + 2.80 x_{2,4} + 0.70 x_{3,4} = 8.28 \end{aligned}$$

パラメータ β は、デザイン行列 X が明示的に示されれば、 $\hat{y} = X\hat{\beta}$ となるような推定値 $\hat{\beta}$ を推定することができる。

表 1.22 に示した Ames 試験におけるコロニー数の表を並び替えて整理した結果を表 3.20 に示す。

表 3.20 対比型のデザイン行列を用いたポアソン回帰

G	A	B	y	n	— デザイン行列 X —				y [^]	P _i	ln L _i	最大化
					x ₀	x ₁	x ₂	x ₃				
1	A1	B1	4	1	1	1	1	1	14.54	0.001	-7.01	ln L = -507.2925
1	A1	B1	7	2	1	1	1	1	14.54	0.013	-8.65	β [^] ₀ = 10.7100
1	A1	B1	8	1	1	1	1	1	14.54	0.024	-3.73	β [^] ₁ = 0.3300
1	A1	B1	9	2	1	1	1	1	14.54	0.039	-6.50	β [^] ₂ = 2.8000
1	A1	B1	10	3	1	1	1	1	14.54	0.056	-8.63	β [^] ₃ = 0.7000
1	A1	B1	11	5	1	1	1	1	14.54	0.075	-12.98	
1	A1	B1	12	3	1	1	1	1	14.54	0.090	-7.21	
1	A1	B1	13	2	1	1	1	1	14.54	0.101	-4.58	初期値
1	A1	B1	14	3	1	1	1	1	14.54	0.105	-6.76	ln L = -549.4206
1	A1	B1	15	3	1	1	1	1	14.54	0.102	-6.86	β [^] ₀ = 10.0000
1	A1	B1	16	6	1	1	1	1	14.54	0.092	-14.29	β [^] ₁ = 1.0000
1	A1	B1	17	6	1	1	1	1	14.54	0.079	-15.23	β [^] ₂ = 1.0000
1	A1	B1	18	4	1	1	1	1	14.54	0.064	-11.00	β [^] ₃ = 1.0000
1	A1	B1	19	5	1	1	1	1	14.54	0.049	-15.09	
1	A1	B1	20	1	1	1	1	1	14.54	0.036	-3.34	
1	A1	B1	21	2	1	1	1	1	14.54	0.025	-7.41	
1	A1	B1	22	1	1	1	1	1	14.54	0.016	-4.12	
2	A1	B2	3	1	1	1	-1	-1	7.54	0.038	-3.27	
2	A1	B2	4	5	1	1	-1	-1	7.54	0.072	-13.19	
2	A1	B2	5	4	1	1	-1	-1	7.54	0.108	-8.91	
2	A1	B2	6	10	1	1	-1	-1	7.54	0.136	-19.98	
2	A1	B2	7	7	1	1	-1	-1	7.54	0.146	-13.47	
2	A1	B2	8	6	1	1	-1	-1	7.54	0.138	-11.90	
2	A1	B2	9	5	1	1	-1	-1	7.54	0.115	-10.80	
2	A1	B2	10	6	1	1	-1	-1	7.54	0.087	-14.65	
2	A1	B2	11	3	1	1	-1	-1	7.54	0.060	-8.46	
2	A1	B2	12	1	1	1	-1	-1	7.54	0.037	-3.28	
2	A1	B2	13	1	1	1	-1	-1	7.54	0.022	-3.83	
2	A1	B2	14	1	1	1	-1	-1	7.54	0.012	-4.45	
3	A2	B1	5	1	1	-1	1	-1	12.48	0.010	-4.65	
3	A2	B1	6	1	1	-1	1	-1	12.48	0.020	-3.91	
3	A2	B1	7	2	1	-1	1	-1	12.48	0.036	-6.67	
3	A2	B1	8	2	1	-1	1	-1	12.48	0.055	-5.78	
3	A2	B1	9	3	1	-1	1	-1	12.48	0.077	-7.69	
3	A2	B1	10	4	1	-1	1	-1	12.48	0.096	-9.37	
3	A2	B1	11	7	1	-1	1	-1	12.48	0.109	-15.52	
3	A2	B1	12	7	1	-1	1	-1	12.48	0.113	-15.24	
3	A2	B1	13	5	1	-1	1	-1	12.48	0.109	-11.09	
3	A2	B1	14	5	1	-1	1	-1	12.48	0.097	-11.67	
3	A2	B1	15	1	1	-1	1	-1	12.48	0.081	-2.52	
3	A2	B1	16	6	1	-1	1	-1	12.48	0.063	-16.59	
3	A2	B1	17	2	1	-1	1	-1	12.48	0.046	-6.15	
3	A2	B1	18	2	1	-1	1	-1	12.48	0.032	-6.88	
3	A2	B1	19	1	1	-1	1	-1	12.48	0.021	-3.86	
3	A2	B1	20	1	1	-1	1	-1	12.48	0.013	-4.33	
4	A2	B2	4	2	1	-1	-1	1	8.28	0.050	-6.01	
4	A2	B2	5	3	1	-1	-1	1	8.28	0.082	-7.49	
4	A2	B2	6	5	1	-1	-1	1	8.28	0.113	-10.88	
4	A2	B2	7	12	1	-1	-1	1	8.28	0.134	-24.10	
4	A2	B2	8	10	1	-1	-1	1	8.28	0.139	-19.74	
4	A2	B2	9	4	1	-1	-1	1	8.28	0.128	-8.23	
4	A2	B2	10	6	1	-1	-1	1	8.28	0.106	-13.48	
4	A2	B2	11	2	1	-1	-1	1	8.28	0.080	-5.06	
4	A2	B2	12	2	1	-1	-1	1	8.28	0.055	-5.80	
4	A2	B2	13	1	1	-1	-1	1	8.28	0.035	-3.35	
4	A2	B2	14	3	1	-1	-1	1	8.28	0.021	-11.63	

それぞれの y_i に対し Excel の関数で、

$$\begin{aligned} x_{0,i} &= 1 \\ x_{1,i} &= \text{if}((A_i = "A1"), 1, -1) \\ x_{2,i} &= \text{if}((B_i = "B1"), 1, -1) \\ x_{3,i} &= x_{2,i} * x_{3,i} \quad \text{交互作用は, 主効果の積} \\ \hat{y}_i &= \text{Mmult}(x_i \text{の範囲}, \hat{\beta} \text{の範囲}) \\ P_i &= \text{Poisson.dist}(y_i, \hat{y}_i, \text{false}) \\ \ln L_i &= n_i * P_i \\ \ln L &= \sum_i \ln L_i \end{aligned}$$

のように計算した。初期値として $\hat{\beta} = [10 \ 1 \ 1 \ 1]^T$ とし、ソルバーで $\ln L$ を最大化するように $\hat{\beta}$ の範囲を設定した。結果は、「最大化」で示した。最大化された対数尤度とパラメータの推定値は、

$$\begin{aligned} \ln L &= \sum_i \ln L_i = -507.292 \\ \hat{\beta} &= [10.710 \ 0.330 \ 2.800 \ 0.700]^T \end{aligned}$$

と表 3.15 に示した JMP でのポアソン回帰の出力結果に一致する。

(0, 1)型デザイン行列

デザイン行列 \mathbf{X} は、回帰パラメータとして何を推定したいのかによって自由に設定することができる。表 3.19 に示したように総平均からの効果を求めたいためには (1, -1) の対比型のデザイン行列 \mathbf{X} が適している。

さて、(1, -1) 型のデザイン行列 \mathbf{X} ではなく、表 3.21 に示すように (0, 1) 型のデザイン行列 \mathbf{X} とすることは可能なのだろうか。実はどのような数値のデザイン行列 \mathbf{X} を定義するかは、全く制約がない。(0, 1) 型ではなく、(1, 0) 型としても、(-1, 1) 型としても、(1, 2) 型としても何ら問題はない。ただし、その結果として推定される回帰パラメータ $\hat{\beta}$ がどのような意味を持つのかが変わる。それは、推定された回帰パラメータ $\hat{\beta}$ をどのように解釈するかが、全て自己責任となるので、何を推定したいのかの明確確認し、その目的が達成されるようなデザイン行列を設定する必要がある。さらに、推定された回帰パラメータが、目的通りに推定されているかの検証も欠かしてはならない。

さて、(1, -1) 対比型ではなく表 3.21 に示すような (0, 1) 型のデザイン行列 \mathbf{X} とした場合に、回帰パラメータ $\hat{\beta}$ は何を推定することになるのだろうか。結果は、

$$\hat{\beta} = [\hat{\beta}_0 \ \hat{\beta}_1 \ \hat{\beta}_2 \ \hat{\beta}_3]^T = [14.54 \ -0.66 \ -5.60 \ -1.40]^T$$

となる。表 3.19 の場合は、

$$\hat{\beta} = [10.71 \quad 0.33 \quad 2.80 \quad 0.70]^T$$

と全く異なっていることに、注意してもらいたい。

表 3.21 (0, 1)型のデザイン行列

A:溶媒	B:代謝 活性化	平均	y	X				β	ε	$\hat{\beta}$
				x_0	x_1	x_2	x_3			
1:蒸留水	1:なし	14.54	y_1	1	0	0	0	β_0	ε_1	14.54
	2:あり	7.54	y_2	1	0	1	1	β_1	ε_2	-0.66
2:DMOS	1:なし	12.48	y_3	1	1	0	1	β_2	ε_3	-5.6
	2:あり	8.28	y_4	1	1	1	0	β_3	ε_4	-1.4

表 3.22 に示すように、回帰パラメータ $\hat{\beta}_0 = 14.54$ は、 A_1B_1 のセル平均であり、 $\hat{\beta}_1 = -0.66$ は、 A_1 水準のセル平均の平均からの A_2 水準の差であり、 $\hat{\beta}_2 = -5.60$ は、 B_1 水準のセル平均の平均からの B_2 水準の差である。さて、 $\hat{\beta}_3 = -1.40$ は、たすき掛の和の平均の差

$$\hat{\beta}_3 : \text{交互作用} : \frac{12.48 + 7.54}{2} - \frac{14.54 + 8.28}{2} = -1.40$$

となっていて、表 3.19 の交互作用 $\hat{\beta}_3 = 0.70$ のマイナス 2 倍となっている。交互作用の意味するところは、主効果 A と主効果 B の効果のずれを計量化した結果である。表 3.17 および表 3.18 を参照しつつ、意味するところを感じ取ってもらいたい。

表 3.22 (0, 1)型でのポアソン回帰の結果

	B:代謝活性化		全体	
	1:なし	2:あり	平均の平均	差
A:溶媒	平均	平均		
1:蒸留水	14.54	7.54	11.04	基準
2:DMOS	12.48	8.28	10.38	-0.66
平均の平均	13.51	7.91	10.71	
差	基準	-5.60		

厳密には、与えられたデザイン行列 X に対する $y = X\beta$ を展開し

$$y_1 = \beta_0 \times 1 + \beta_1 \times 0 + \beta_2 \times 0 + \beta_3 \times 0 \quad (1)$$

$$y_2 = \beta_0 \times 1 + \beta_1 \times 0 + \beta_2 \times 1 + \beta_3 \times 1 \quad (2)$$

$$y_3 = \beta_0 \times 1 + \beta_1 \times 1 + \beta_2 \times 0 + \beta_3 \times 1 \quad (3)$$

$$y_4 = \beta_0 \times 1 + \beta_1 \times 1 + \beta_2 \times 1 + \beta_3 \times 0 \quad (4)$$

これらの式を $\beta_0, \beta_1, \beta_2, \beta_3$ について解くことが必要である。(1) 式から β_0 は、

$$y_1 = \beta_0 \times 1$$

$$\hat{\beta}_0 = y_1 = 14.54$$

が得られ、これは A_1B_1 のセル平均となっている。 β_1 は、[(3)+(4)] - [(1)+(2)] によって

$$(y_3 + y_4) - (y_1 + y_2) = 2\beta_1$$

$$\hat{\beta}_1 = \frac{y_3 + y_4}{2} - \frac{y_1 + y_2}{2} = \frac{12.48 + 8.28}{2} - \frac{14.45 + 7.54}{2} = -0.66$$

が得られ A の水準の差となっている。 β_2 は、[(2)+(4)] - [(1)+(3)] によって

$$(y_2 + y_4) - (y_1 + y_3) = 2\beta_2$$

$$\hat{\beta}_2 = \frac{y_2 + y_4}{2} - \frac{y_1 + y_3}{2} = \frac{7.54 + 8.28}{2} - \frac{14.54 + 12.54}{2} = -5.60$$

が得られ B_2 水準の差となっている。 β_3 は、[(2)+(3)] - [(1)+(4)] によって

$$(y_2 + y_3) - (y_1 + y_4) = 2\beta_3$$

$$\hat{\beta}_3 = \frac{y_2 + y_3}{2} - \frac{y_1 + y_4}{2} = \frac{7.54 + 12.48}{2} - \frac{14.54 + 8.28}{2} = -1.40$$

が得られ、たすき掛の和の平均の差であるが、 A_1 内の B_1 と B_2 の差の 2 分の 1、 A_2 内の B_1 と B_2 の差の 2 分の 1 を求め、それらの差とも解釈される。これらは、交互作用といわれているのであるが、どのように解釈するかについて、試行錯誤を重ねてもらいたい。

$$\hat{\beta}_3 = \frac{y_3 - y_4}{2} - \frac{y_1 - y_2}{2} = \frac{12.48 - 8.28}{2} - \frac{14.54 - 7.54}{2} = 2.10 - 3.50 = -1.40$$

・推定された回帰パラメータの意味することは、与えられたデザイン行列 X に対し回帰パラメータ β を掛けて、それぞれの β_i について解き、その内容を吟味して後付的に得られる。

基準との差 (標示型 or Indicator 型デザイン行列)

さて、表 3.23 に示したデザイン行列 X は、 2×2 の要因配置ではなく、4 水準の 1 元配置型と見なして、最初的水準を基準として、他の水準との差の検定を行いたい場合などに使わ

表 3.23 基準との差の推定 (標示型 または Indicator 型)

A:溶媒	B:代謝		平均	y	X				β	ε	β^{\wedge}
	活性化				x_0	x_1	x_2	x_3			
1:蒸留水	1:なし		14.54	y_1	1	0	0	0	$\beta_0^{\wedge} + \varepsilon_1$		14.54
	2:あり		7.54	y_2	1	1	0	0	$\beta_1^{\wedge} + \varepsilon_2$		-7.00
2:DMOS	1:なし		12.48	y_3	1	0	1	0	$\beta_2^{\wedge} + \varepsilon_3$		-2.06
	2:あり		8.28	y_4	1	0	0	1	$\beta_3^{\wedge} + \varepsilon_4$		-6.26

れる。基準を最後の水準として設定することもできる。SAS/GENMOD のデフォルトのデザイン行列として伝統的に使われている。

さて、推定される回帰パラメータの意味を考えてみよう。切片 β_0 は、 $\beta_0 = y_1$ であることは自明であろう。 β_1 は、 $\beta_1 = y_2 - y_1$ となることも容易に求められるであろう。同様に、 $\beta_2 = y_3 - y_1$ 、 $\beta_3 = y_4 - y_1$ が得られる。このように基準となる組合せ水準からの差について尤度比検定がまとめてポアソン回帰により行える。

表 3.20 で示した Excel による最尤法のデータを JMP ファイルとし、標示型のデザイン行列にしてポアソン回帰を行った結果を表 3.34 に示す。尤度比カイ 2 乗検定の結果は、全て有意な差であることが確認される。

表 3.24 JMP による基準との差の推定 (標示型 または Indicator 型)

パラメータ推定値						
項	推定値	標準誤差	尤度比カイ2乗	p値	下側信頼限界	上側信頼限界
切片	14.5400	0.5393	727.0000	<.0001*	13.5085	15.6227
x1	-7.0000	0.6645	112.8979	<.0001*	-8.3116	-5.7046
x2	-2.0600	0.7351	7.8603	0.0051*	-3.5038	-0.6201
x3	-6.2600	0.6756	86.9731	<.0001*	-7.5921	-4.9419

交互作用の吟味

さて、B: 代謝活性化が (2:あり) の場合に、溶媒を (1:蒸留水) から (2:DMOS) にした場合にコロニー数がわずかに増加しているが、統計的にはどうであろうか。この検討のために、(0, 1) 型のデザイン行列を設定することもできるが、簡便的には、因子 B 別に因子 A 内の比較を (1, -1) 型で行うこともできる。表 3.25 に示すように、 $p=0.1882$ と有意ではない。推定値は、切片が A1 と A2 の平均で、A:溶媒[A1] は、A1 と A2 の差の 2 分の 1

$$A:溶媒[A1] = (7.54 - 8.28) / 2 = -0.74 / 2 = -0.37$$

となっている。

表 3.25 代謝活性がある場合の A1:蒸留水と A2: DMOS との差

パラメータ推定値						
項	推定値	標準誤差	尤度比カイ2乗	p値	下側信頼限界	上側信頼限界
切片	7.9100	0.2812	791.0000	<.0001*	7.3715	8.4741
A:溶媒[A1]	-0.3700	0.2812	1.7314	0.1882	-0.9225	0.1813

3.5. 2本の回帰直線に対する各種のデザイン行列

前節では 2×2 の要因配置型のデータに対して解析のための幾つかのデザイン行列を例示し、それぞれ異なるパラメータの推定値が得られることを示した。その結果の解釈を行うための考え方も詳細に述べてきた。通常、回帰分析については、第4章で詳しく述べるが、ここでは、2本の回帰直線をさまざまな観点から同時あてはめを行うためのデザイン行列に焦点をあてる。ここでは、2本のポアソン回帰直線を扱っているが、最小2乗法による2本の回帰直線の解析でもここに示したデザイン行列の考え方は、全く同じである。

切片を共通とする場合

第1.8節では、2本の回帰直線が傾きは異なるが共通の切片を持つ場合を示した。解析に用いたデザイン行列は、SASのDATAステップで内部的に作成し、詳しい説明を避けた。どのようなデザイン行列が作成されたのであろうか。実際のデータを用いた例示はサイズが大きくなりすぎるので、最小限のデータを用いて説明する。

表3.26に示したのは、第1.8節の表1.27で示した細菌を用いた用量反応試験の結果を要約したものである[富山ら(2004)]。薬剤の濃度を(0, 50, 100)の3段階にし、反応を変異コロニー数の平均を整数化し $n=6$ の小さなデータとして、さまざまなデザイン行列をコンパクトにし、比較が容易に行えるようにした。

T薬とS薬に共通の切片を持ち、それぞれの薬剤に別々の傾きを持つようなデザイン行列を設定し、ポアソン分布による対数尤度 $\ln L$ を設定し、Excelのソルバーにより、 $\ln L$ を最大化するように $\hat{\beta}$ を変化させた結果である。推定された2本の回帰式は、

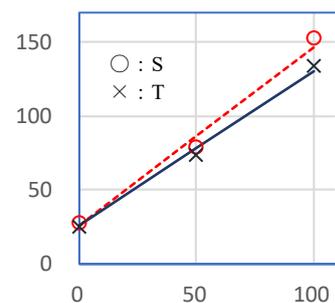
$$\text{S薬} : y^{(1)} = 25.626 + 1.212x$$

$$\text{T薬} : y^{(2)} = 25.626 + 1.049x$$

のように同じ切片を持つが、異なる傾きを持つ回帰式である。

表 3.26 切片を共通にする T薬と S薬のポアソン回帰直線

薬 剤	濃 度	y	y^{\wedge}	デザイン行列 X			$\hat{\beta}$	$\ln L_i$	対数尤度
				x_0	x_1	x_2			
S	0	28	25.6263	1	0	0	25.6263	β_0^{\wedge}	-2.6947
	50	79	86.2337	1	50	0	1.2121	β_1^{\wedge}	-3.4170
	100	153	146.8411	1	100	0	1.0495	β_2^{\wedge}	-3.5621
T	0	25	25.6263	1	0	0			-2.5394
	50	74	78.0996	1	0	50			-3.1816
	100	134	130.5729	1	0	100			-3.4131
$\ln L =$									-18.8079



デザイン行列の x_0 は、切片を求めるために全て 1 とし、 x_1 には、薬剤 S に対応した行のみに濃度が設定され、 x_2 には、薬剤 T に対応した行のみに濃度が設定されている。推定値 \hat{y}_i は、デザイン行列 X と回帰パラメータのベクトル $\hat{\beta}$ の積で、次のように計算される。

$$\begin{aligned}\hat{y}_1 &= \hat{\beta}_0 x_{0,1} + \hat{\beta}_1 x_{1,1} + \hat{\beta}_2 x_{2,1} = 25.63 \times 1 + 1.21 \times 0 + 1.05 \times 0 = 25.6 \\ \hat{y}_2 &= \hat{\beta}_0 x_{0,2} + \hat{\beta}_1 x_{1,2} + \hat{\beta}_2 x_{2,2} = 25.63 \times 1 + 1.21 \times 50 + 1.05 \times 0 = 86.2 \\ &\vdots \\ \hat{y}_6 &= \hat{\beta}_0 x_{0,6} + \hat{\beta}_1 x_{1,6} + \hat{\beta}_2 x_{2,6} = 25.63 \times 1 + 1.21 \times 0 + 1.05 \times 100 = 130.6\end{aligned}$$

対数尤度 $\ln L_i$ は、平均を \hat{y}_i とするポアソン分布の確率の対数で、次のように計算される

$$\begin{aligned}\ln L_1 &= \ln[\text{Poisson.dist}(y_1, \hat{y}_1, \text{false})] = \ln[\text{Poisson.dist}(28, 25.6, \text{false})] = \ln[0.0676] = -2.695 \\ \ln L_2 &= \ln[\text{Poisson.dist}(y_2, \hat{y}_2, \text{false})] = \ln[\text{Poisson.dist}(79, 86.2, \text{false})] = \ln[0.0328] = -3.417 \\ &\vdots \\ \ln L_6 &= \ln[\text{Poisson.dist}(y_6, \hat{y}_6, \text{false})] = \ln[\text{Poisson.dist}(134, 130.6, \text{false})] = \ln[0.0329] = -3.413\end{aligned}$$

対数尤度 $\ln L$ は、それぞれの $\ln L_i$ の合計であり、ソルバーにより $\hat{\beta}$ を変化させて最大化した結果である。S 薬について \hat{y}_i の推定値は、

$$\begin{aligned}\hat{y}_1 &= \hat{\beta}_0 + \hat{\beta}_1 x_{1,1} = 25.63 + 1.21 \times 0 = 25.6 \\ \hat{y}_2 &= \hat{\beta}_0 + \hat{\beta}_1 x_{1,2} = 25.63 + 1.21 \times 50 = 86.2 \\ \hat{y}_3 &= \hat{\beta}_0 + \hat{\beta}_1 x_{1,3} = 25.63 + 1.21 \times 100 = 136.8\end{aligned}$$

であり、 $\hat{y}_1 = \hat{\beta}_0 = 25.6$ から、 $x_1 = 0$ の場合の推定値であり、「Y 切片」となり、 $\hat{y}_2 = 86.2$ は、切片に濃度 50 に $\hat{\beta}_1 = 1.21$ を掛けて足したものである。したがって、 $\hat{\beta}_1$ は濃度が 1 増加した時の S 薬の反応の増加分である。濃度 100 の場合の推定値は $\hat{y}_3 = 146.8$ となる。

T 薬については、

$$\begin{aligned}\hat{y}_4 &= \hat{\beta}_0 + \hat{\beta}_2 x_{2,4} = 25.63 + 1.05 \times 0 = 25.6 \\ \hat{y}_5 &= \hat{\beta}_0 + \hat{\beta}_2 x_{2,5} = 25.63 + 1.05 \times 50 = 78.1 \\ \hat{y}_6 &= \hat{\beta}_0 + \hat{\beta}_2 x_{2,6} = 25.63 + 1.05 \times 100 = 130.6\end{aligned}$$

であり、 $\hat{y}_4 = \hat{\beta}_0 = 25.6$ から、 $x_2 = 0$ の場合の推定値であり、S 薬と共通の Y 切片となり、 $\hat{y}_5 = 78.1$ は、切片に濃度 50 に $\hat{\beta}_2 = 1.05$ を掛けて足したものである。したがって、 $\hat{\beta}_2$ は濃度が 1 増加した時の T 薬の反応の増加分となり、T 薬の場合の傾きである。

傾きを共通とする平行線(0, 1)型

表 3.27 は、S 薬と T 薬の傾きを共通とする回帰直線を得るためのデザイン行列である。S 薬と T 薬の濃度が同じ変数 x_2 となっていることから回帰パラメータ $\hat{\beta}_2$ は共通となることが推論される。推定された 2 本の回帰式は、

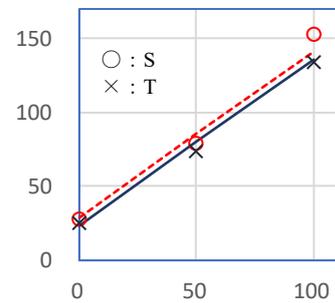
$$y^{(1)} = 28.41 + 1.13x_2$$

$$\begin{aligned} y^{(2)} &= 28.41 - 5.33 + 1.13x_2 \\ &= 23.08 + 1.13x_2 \end{aligned}$$

のように同じ傾きであるが、異なる切片を持つ回帰式である。

表 3.27 傾きを共通とする S 薬と T 薬の回帰直線

薬 剤	濃 度	y	\hat{y}	デザイン行列 X			$\hat{\beta}$		対数尤度 $\ln L_i$
				x_0	x_1	x_2			
S	0	28	28.4085	1	0	0	28.4085	β_0^{\wedge}	-2.5910
	50	79	84.8326	1	0	50	-5.3319	β_1^{\wedge}	-3.3100
	100	153	141.2568	1	0	100	1.1285	β_2^{\wedge}	-3.9098
T	0	25	23.0765	1	1	0			-2.6097
	50	74	79.5007	1	1	50			-3.2669
	100	134	135.9248	1	1	100			-3.3822
$\ln L =$									-19.0696



では $\hat{\beta}_0$ と $\hat{\beta}_1$ は、何を推定してるのであろうか。デザイン行列の x_1 は、S 薬の場合には 0、T 薬の場合には 1 となっているので、 \hat{y}_1 と \hat{y}_4 の推定式は、

$$\text{S 薬} : \hat{y}_1 = \hat{\beta}_0 = 28.41$$

$$\text{T 薬} : \hat{y}_4 = \hat{\beta}_0 + \hat{\beta}_1 = 23.08$$

となる。これらから、S 薬の切片が、 $\hat{\beta}_0 = 28.41$ であり、T 薬の切片が、 $\hat{\beta}_0 + \hat{\beta}_1 = 23.08$ となる。このことから、

$$\hat{\beta}_1 = 23.08 - \hat{\beta}_0 = 23.08 - 28.41 = -5.33$$

となり、 $\hat{\beta}_1 = -5.33$ は、S 薬の切片からの T 薬の切片との減少分であることがわかる。

傾きを共通とする平行線(標示型)

表 3.28 は、表 3.27 と同じ回帰直線となるが、デザイン行列が異なる。変数 x_0 は、S 薬の時にのみ 1 で、T 薬では 0 となっている。変数 x_1 は、逆に S 薬の時に 0 で、T 薬では 1 となっている。この様なデザイン行列にすることにより、T 薬と S 薬の切片を $\hat{\beta}_1 = 28.4$ と直接推定

することができる。推定された2本の回帰式は、

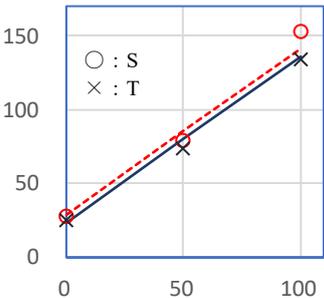
$$\text{S薬} : y^{(1)} = 28.41 + 1.13x$$

$$\text{T薬} : y^{(2)} = 23.08 + 1.13x$$

のように同じ傾きであるが、異なる切片を持つ回帰式である。

表 3.28 傾きを共通とする T薬と S薬の切片の直接推定

薬 剤	濃 度	y	y [^]	デザイン行列 X			$\hat{\beta}$		対数尤度 ln L _i
				x ₀	x ₁	x ₂			
S	0	28	28.4085	1	0	0	28.4085	β_0^{\wedge}	-2.5910
	50	79	84.8326	1	0	50	23.0765	β_1^{\wedge}	-3.3100
	100	153	141.2567	1	0	100	1.1285	β_2^{\wedge}	-3.9098
T	0	25	23.0765	0	1	0			-2.6097
	50	74	79.5006	0	1	50			-3.2669
	100	134	135.9247	0	1	100			-3.3822
ln L =									-19.0696



交互作用

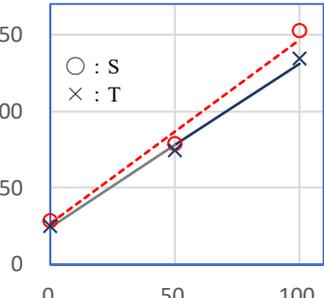
共通の傾きを持つ回帰直線のあてはめが、統計的に支持されるのかを検査するために、薬剤に対する変数 x_1 、濃度に関する変数 x_2 の積を変数 $x_3 = x_1x_2$ とする。これが薬剤と濃度の交互作用となり、尤度比検定が行える。

表 3.29 に交互作用列を追加した 6×4 のデザイン行列に対するポアソン回帰の解析結果を示す。パラメータ $\hat{\beta}_0 = 26.814$ は、S薬の切片であり、 $\hat{\beta}_1 = -2.390$ は S薬の切片と T薬の切片の差であり、T薬の切片は、

$$\hat{\beta}_0 + \hat{\beta}_1 = 26.814 - 2.390 = 24.423$$

表 3.29 交互作用の検討

薬 剤	濃 度	y	y [^]	デザイン行列 X				$\hat{\beta}$		対数尤度 ln L _i
				x ₀	x ₁	x ₂	x ₃			
S	0	28	26.814	1	0	0	0	26.8140	β_0^{\wedge}	-2.6139
	50	79	86.667	1	0	50	0	-2.3905	β_1^{\wedge}	-3.4543
	100	153	146.519	1	0	100	0	1.1971	β_2^{\wedge}	-3.5760
T	0	25	24.424	1	1	0	0	-0.1322	β_3^{\wedge}	-2.5385
	50	74	77.667	1	1	50	50			-3.1600
	100	134	130.910	1	1	100	100			-3.4047
ln L =									-18.7473	



となる。推定された2本の回帰式は、

$$\hat{y}^{(S)} = 26.814 + 1.197x$$

$$\begin{aligned}\hat{y}^{(T)} &= 26.814 - 2.390 + (1.197 - 0.132)x \\ &= 24.423 + 1.065x\end{aligned}$$

のように異なる切片および異なる傾きを持つ回帰式である。

傾きに関する $\hat{\beta}_2 = 1.197$ と $\hat{\beta}_3 = 0.132$ はどのように解釈したらよいのであろうか。濃度が100の場合の推定値 \hat{y}_3 , \hat{y}_6 の推定式は、

$$\hat{y}_3 = \hat{\beta}_0 + 100\hat{\beta}_2 = 26.814 + 100\hat{\beta}_2 = 146.519$$

$$\hat{\beta}_2 = \frac{146.519 - 26.814}{100} = 1.197$$

$$\begin{aligned}\hat{y}_6 &= \hat{\beta}_0 + \hat{\beta}_1 + 100\hat{\beta}_2 + 100\hat{\beta}_3 \\ &= 26.814 - 2.390 + 100 \times 1.197 + 100\hat{\beta}_3 \\ &= 144.124 + 100\hat{\beta}_3 = 130.910\end{aligned}$$

$$\hat{\beta}_3 = \frac{130.910 - 144.124}{100} = -0.132$$

となり、 $\hat{\beta}_2 = 1.197$ は、S薬の傾きであり、 $\hat{\beta}_3 = -0.132$ は、S薬とT薬の傾きの差であると解釈される。この交互作用モデルの対数尤度は、 $\ln L^{(\text{交互作用})} = -18.7473$ であり、表3.28の主効果のみのモデルの対数尤度 $\ln L^{(\text{主効果})} = -19.0696$ との差の2倍

$$\chi_{x_3}^2 = 2(\ln L^{(\text{交互作用})} - \ln L^{(\text{主効果})}) = 2 \times (-18.7473 + 19.0696) = 0.6446, \quad p = 0.4220 \text{ N.S.}$$

が自由度1のカイ2乗分布に従うことで検定ができる。表3.30にJMPによるポアソン回帰による交互作用モデルの結果を示した。交互作用に関する尤度比カイ2乗値は、0.6447であり、四捨五入による誤差の範囲内で一致していることが確認できる。

表 3.30 JMP によるポアソン回帰での交互作用に対する尤度比検定

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	項	推定値	標準誤差	尤度比カイ2乗	p値
差分	89.2672	178.5344	3	切片	26.8140	4.9468	29.3810	<.0001*
完全	18.7473			x1	-2.3905	6.8659	0.1212	0.7277
縮小	108.0145			x2	1.1971	0.1191	94.9808	<.0001*
				x3	-0.1322	0.1646	0.6447	0.4220

別々の回帰直線

交互作用の検討では、薬剤と濃度についての主効果モデルに、交互作用列を加えたモデルによって行った。別々の回帰直線のパラメータの推定を直接求めることもデザイン行列の設定次第で自由にできる。表 3.31 に示したデザイン行列は、S 薬と T 薬に別々の切片と傾きを同時推定することができる。推定された 2 本の回帰式は、

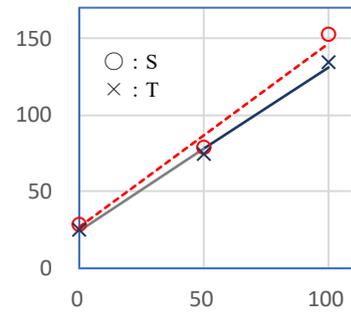
$$y^{(S)} = 26.814 + 1.197x$$

$$y^{(T)} = 24.423 + 1.065x$$

のように異なる切片および傾きを持つ回帰式である。

表 3.31 別々の回帰直線のパラメータ推定

薬 剤	濃 度	y	y^{\wedge}	デザイン行列 X				β^{\wedge}		対数尤度 $\ln L_i$
				x_0	x_1	x_2	x_3			
S	0	28	26.814	1	0	0	0	26.8140	β_0^{\wedge}	-2.6139
	50	79	86.667	1	0	50	0	24.4235	β_1^{\wedge}	-3.4543
	100	153	146.519	1	0	100	0	1.1971	β_2^{\wedge}	-3.5760
T	0	25	24.423	0	1	0	0	1.0649	β_3^{\wedge}	-2.5385
	50	74	77.667	0	1	0	50			-3.1600
	100	134	130.910	0	1	0	100			-3.4047
$\ln L =$										-18.7473



推定結果を見ると、S 薬の切片： $\hat{\beta}_0 = 26.814$ ，傾き： $\hat{\beta}_2 = 1.197$ ，T 薬の切片： $\hat{\beta}_1 = 24.423$ ，傾き： $\hat{\beta}_3 = 1.065$ でありとなる。交互作用モデルと比較すると、推定されたパラメータは異なるが、 $\hat{\beta}_1$ は完全に一致している。もちろん対数尤度も $\ln L = -18.7473$ と一致している。

3. 6. オフセットを含む対数リンクでの 2 本の 2 次曲線のあてはめ

第 1.11 節で取り上げた「喫煙習慣と冠動脈疾患による死亡」データを表 3.32 に再掲する [ドブソン (2008)]. 第 1.11 節では, オフセットを含む 2 群の対数リンクによるポアソン回帰の導入のために, 80 歳のデータを除いた結果を示した. これは, 80 歳での 10 万人比での死亡数が, 非喫煙者と喫煙者で逆転し, 非喫煙者群の死亡数が多くなり, 1 次式ではなく 2 次のポアソン回帰を検討する必要性があった.

表 3.32 年齢階層別の喫煙習慣による冠動脈心疾患による死亡数 (表 1.38 再掲)

年齢		非喫煙者 ($x = 0$)			喫煙者 ($x = 1$)		
範囲	歳	死亡	人年	10万人比	死亡	人年	10万人比
35-44	40	2	18,790	10.6	32	52,407	61.1
45-54	50	12	10,673	112.4	104	43,248	240.5
55-64	60	28	5,710	490.4	206	28,612	720.0
65-74	70	28	2,585	1083.2	186	12,663	1468.8
75-84	80	31	1,462	2120.4	102	5,317	1918.4

各年代での部分母集団の大きさが, 35-44 歳代に比べ 75-84 歳代は 10 分の 1 以下となるので, 10 万人比による散布図を図 3.2 に示す. 左は実目盛りであり, 死亡数が年齢と共に指数関数的に増加している. 40 歳から 70 歳までは, 喫煙者の死亡数が非喫煙者よりも多いが 80 歳で逆転している. 部分母集団の数が, それぞれ (5,317, 1,462) と小さいので統計的な変動の範囲内なのかも知れない.

図 3.2 右の対数目盛りでの散布図から, 年齢が 40 歳の場合の対数死亡数が喫煙者と非喫煙者の間で最も開いていて, 年齢が上がるにつれ差が縮まり 80 歳ではほとんど同じとなっている. また, 年齢が上がるにつれ対数死亡数の増加は減弱している. これらのことから, 単純

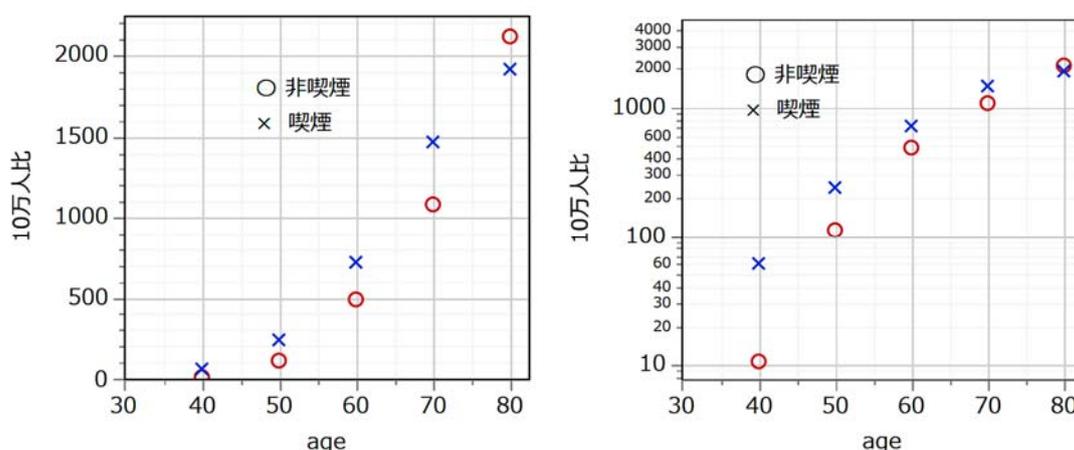


図 3.2 非喫煙・喫煙別の 10 万人比での死亡数

な対数リンクによる 2 本の直線のあてはめではなく、年齢に 2 次の項を追加する必要性が感じられる。さらに喫煙習慣と年齢の交互作用も推察される。

一足飛びに 2 次式のあてはめを行う前に、最も簡単なモデルから複雑なモデルへと段階的に進めることにより、ポアソン回帰の応用について理解を深めてもらいたい。

(非喫煙・喫煙) の 2 群間比較

年齢区分を無視して、(0:非喫煙・1:喫煙) で各 5 組の死亡数 y_i とその部分母集団数 n_i が得られたとしよう。表 3.33 の Excel シートは、年齢区分を除いた 10×2 のデザイン行列でのポアソン回帰を行う。推定値を $\hat{y}_i = n_i \cdot \exp(\hat{\beta}_0 x_{0i} + \hat{\beta}_1 x_{1i})$ とし、ポアソン確率を $P_i = \text{Poisson.dist}(y_i, \hat{y}_i, \text{false})$ 、対数尤度を $\ln L_i = \ln(P_i)$ とし、その合計を $\ln L$ とし、 $\ln L$ を最大化するように $(\hat{\beta}_0, \hat{\beta}_1)$ をソルバーで変化させた結果である。推定されたパラメータから、10 万人比での非喫煙者の死亡数が 257.5 人、喫煙者の場合は 385.6 人と推定される。

$$\text{非喫煙者} : \hat{y}_0 = 100,000 \times \exp(-5.9618 + 0.5422 \times 0) = 257.5 \text{ 人}$$

$$\text{喫煙者} \quad \hat{y}_1 = 100,000 \times \exp(-5.9618 + 0.5422 \times 1) = 385.6 \text{ 人}$$

表 3.33 喫煙習慣による冠動脈心疾患による死亡数

切片	喫煙	死亡	人年	10万人比	推定値	確率	対数尤度			
x_0	x_1	y	n	y'	\hat{y}	P	$\ln L_i$		最尤解	
1	0	2	18,790	11	48.4	1.13E-18	-41.3229	$\hat{\beta}_0 =$	-5.9618	
1	0	12	10,673	112	27.5	4.49E-04	-7.7087	$\hat{\beta}_1 =$	0.5422	
1	0	28	5,710	490	14.7	6.58E-04	-7.3260			
1	0	28	2,585	1,083	6.7	4.75E-10	-21.4682			
1	0	31	1,462	2,120	3.8	1.99E-18	-40.7593			
1	1	32	52,407	61	232.1	3.02E-61	-139.3535			
1	1	104	43,248	240	191.5	1.44E-12	-27.2691			
1	1	206	28,612	720	126.7	2.54E-11	-24.3980			
1	1	186	12,663	1,469	56.1	1.10E-42	-96.6127			
1	1	102	5,317	1,918	23.5	5.37E-33	-74.3037			
							$\ln L =$	-480.5221		

$$y'_i = (y_i / n_i) \cdot 100,000, \quad \hat{y}_i = n_i \cdot \exp(\hat{\beta}_0 x_{0i} + \hat{\beta}_1 x_{1i}), \quad P = \text{Poisson.dist}(y_i, \hat{y}_i, \text{false}), \quad \ln L_i = \ln(P_i)$$

2 群間の尤度比カイ 2 乗値は、切片のみをデザイン行列 (10×1) とした縮小モデルでの対数尤度を

縮小モデル:	$\hat{\beta}_0 =$	-5.5144	$\ln L =$	-495.0676
--------	-------------------	----------------	-----------	------------------

を求め、完全モデル $\ln L_{\text{完全}} = -480.5221$ と縮小モデル $\ln L_{\text{縮小}} = -495.0676$ のマイナス 1 倍が、表 3.34 に示す JMP の結果に一致する。これらの差の -2 倍は、「差分」の行で示され 29.0911 となり、有意な差であると判断される。

表 3.34 JMP による喫煙習慣に関する尤度比カイ 2 乗

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	14.5456	29.0911	1	<.0001*
完全	480.5221			
縮小	495.0676			

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	-5.9618	0.0995	77033.712	<.0001*
x1:smoke	0.5422	0.1072	29.0911	<.0001*

この推定結果を図 3.3 に示す. 実目盛りでは若年層での母集団数が多いことが反映され, 10 万人比目盛り上で低い位置に推定結果が表示されているが, 対数目盛り上では, 中ほどに推定結果が示されている.

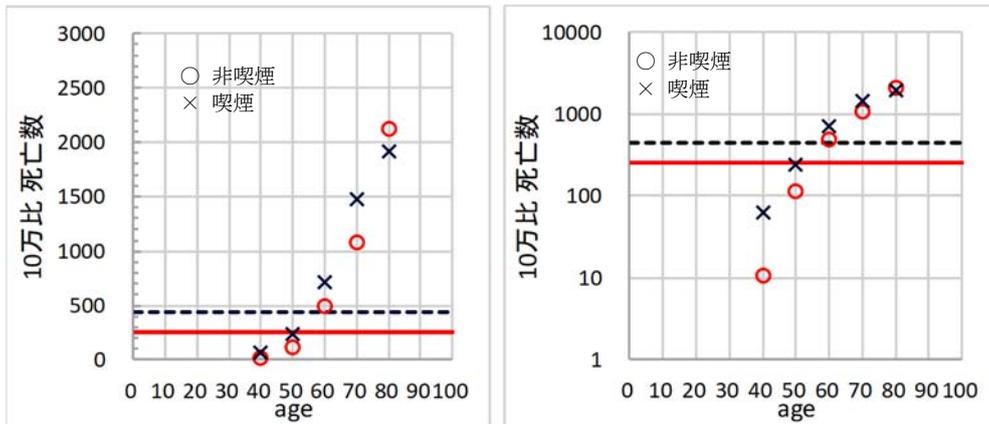


図 3.3 (非喫煙・喫煙) に対するポアソン回帰

喫煙習慣に年齢を加えた対数リンクでの平行な 2 本の直線

喫煙習慣に年齢を加えたモデルでは, 対数死亡数に対して 2 本の直線をあてはめることを考える. その際, 共通の切片を持つのか, 共通の傾きを持つ平行線をあてはめるのか, 別々の直線をあてはめるのかの判断が必要である. 図 3.3 から, 別々の直線のあてはめが適しているように思われるが, 喫煙習慣により死亡数に差あるかを検討したいので, 平行な直線をあてはめて, 年齢に関わらず喫煙習慣の差について平均的な尤度比カイ 2 乗値で検討したい.

表 3.35 に示すように, 喫煙習慣と年齢を加えた 10×3 のデザイン行列では, 非喫煙群の回帰直線の切片が $\hat{\beta}_0$ となり, 喫煙群の切片は $\hat{\beta}_0 + \hat{\beta}_1$ となる. 傾きは $\hat{\beta}_2$ で共通で, ように, 年齢

が 60 歳の場合の 10 万人比での推定値は,

$$\text{非喫煙者: } \hat{y}_{0,age=60} = 100,000 \times \exp(-10.6260 + 0.4064 \times 0 + 0.0836 \times 60) = 365.8 \text{ 人}$$

$$\text{喫煙者: } \hat{y}_{1,age=60} = 100,000 \times \exp(-10.6260 + 0.4064 \times 1 + 0.0836 \times 60) = 549.2 \text{ 人}$$

となる.

表 3.35 喫煙習慣別の年齢による対数リンクでの 2 本の直線のあてはめ

切片	喫煙	年齢	死亡	人年	10万人比	推定値	10万比	確率	対数尤度		
x_0	x_1	$x_2:age$	y	n	y'	\hat{y}	y''	P	$\ln Li$		最尤解
1	0	40	2	18,790	11	12.9	68.7	0.0002	-8.4925	$\hat{\beta}_0 =$	-10.6260
1	0	50	12	10,673	112	16.9	158.6	0.0515	-2.9664	$\hat{\beta}_1 =$	0.4064
1	0	60	28	5,710	490	20.9	365.8	0.0252	-3.6815	$\hat{\beta}_2 =$	0.0836
1	0	70	28	2,585	1,083	21.8	843.8	0.0336	-3.3930		
1	0	80	31	1,462	2,120	28.5	1946.4	0.0640	-2.7491		
1	1	40	32	52,407	61	54.1	103.2	0.0004	-7.9464		
1	1	50	104	43,248	240	103.0	238.1	0.0389	-3.2471		
1	1	60	206	28,612	720	157.1	549.2	0.0000	-10.4973		
1	1	70	186	12,663	1,469	160.4	1266.9	0.0042	-5.4708		
1	1	80	102	5,317	1,918	155.4	2922.4	0.0000	-13.6810		
									$\ln L =$	-62.1250	

$$y'_i = (y_i / n_i) \cdot 100,000, \quad \hat{y}_i = n_i \cdot \exp(\hat{\beta}_0 x_{0i} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}),$$

$$P_i = \text{Poisson.dist}(y_i, \hat{y}_i, \text{false}), \quad \ln L_i = \ln(P_i)$$

年齢をモデルに加えることにより, $\ln L = -62.1250$ と大きく変化する. 図 3.4 に示すように年齢を加味しても喫煙習慣による死亡数に違いがあるようだが, 尤度比カイ 2 乗値を求めるためには, 喫煙習慣を含めない年齢のみの縮小モデルでのポアソン回帰での対数尤度が必要となる. 計算のための Excel 表は省略するが, 次の結果から,

縮小モデル年齢:	$\hat{\beta}_0 =$	-10.2988		$\ln L =$	-70.0397
	$\hat{\beta}_2 =$	0.0838			

対数尤度は, $\ln L = -70.0397$ となる.

各モデルでの対数尤度, マイナス 2 倍の対数尤度, 縮小モデルとの差をまとめると,

モデル	対数尤度	マイナス 2 倍	縮小モデルとの差
切片 (縮小)	-495.0676	990.1353	
切片+煙習慣	-480.5221	961.0441	29.0911
切片+年齢 (縮小)	-70.0397	140.0795	
切片+喫煙習慣+年齢	-62.1250	124.2500	15.8294

となる. 年齢を加味しない場合の喫煙歴の差についての尤度比カイ 2 乗値=29.0911 に対し,

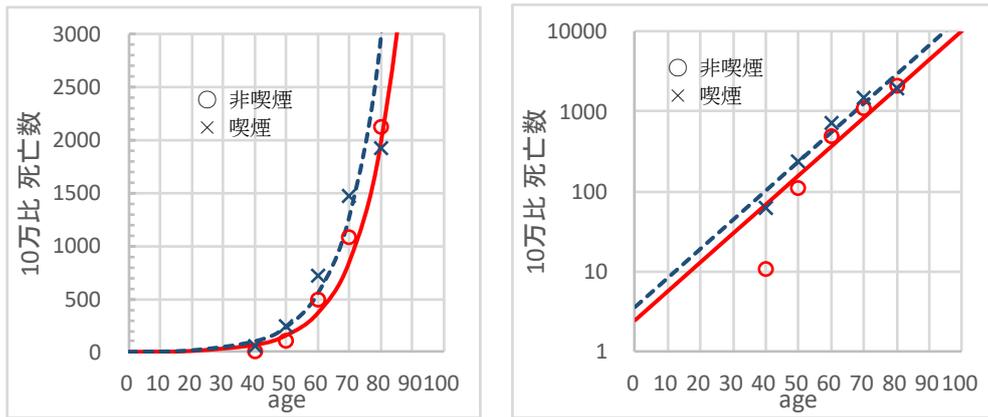


図 3.4 喫煙歴別の対数リンクでの 2 本のポアソン回帰直線

年齢を加味した場合の尤度比カイ 2 乗値=15.8294 と若干減少するが、自由度 1 のカイ 2 乗分布の 5%上側確率点は、3.84 なので、はるかに大きいカイ 2 乗値となっている。切片のみの縮小モデルに対し、年齢を加味した場合のマイナス 2 倍の対数尤度の差は、

$$\text{尤度比カイ 2 乗} = 990.1353 - 140.0795 = 850.0558$$

となり、喫煙習慣に対してはるかに大きい。また、(切片+喫煙習慣) と (切片+喫煙習慣+年齢) の差は、

$$\text{尤度比カイ 2 乗} = 961.0441 - 124.2500 = 836.7941$$

となる、喫煙習慣が含まれているモデルについて、年齢を加味した場合のマイナス 2 倍の対数尤度の差は、少し減ってはいるが、極めて大ききく、加齢とともに死亡数が劇的に増えることを意味している。

(切片+喫煙習慣+年齢) モデルに対し、表 3.36 に示す JMP のポアソン回帰の出力と対比する。「モデル」の「縮小」の (-1)*対数尤度 495.0676 は、切片のみのモデルに対応し、「完全」は、(切片+喫煙習慣+年齢) に対応し 62.1250 となる。「差分」の 865.8853 は、切片モデルに (喫煙習慣+年齢) を加えた場合の 2 倍の差である。

表 3.36 (切片+喫煙習慣+年齢) ポアソン回帰

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	432.9426	865.8853	2	<.0001*
完全	62.1250			
縮小	495.0676			

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	-10.6258	0.2097	3777.5415	<.0001*
x1:smoke	0.4064	0.1072	15.8294	<.0001*
x2:age	0.0836	0.0029	836.7941	<.0001*

「パラメータ推定値」の推定値は、表 3.35 の Excel での計算結果に一致している。「標準誤差」は、Excel の計算過程で共分散行列の計算をせずにソルバーで最尤解を求めたために対応できない。「x1:smoke」の「尤度比カイ 2 乗」は、15.8294 であり、すでに求めた（切片＋年齢）と（切片＋喫煙習慣＋年齢）モデルのマイナス 2 倍の対数尤度の差となっている。「x1:age」の 836.7941 は、（切片＋煙習慣）と（切片＋喫煙習慣＋年齢）の差であることもすでに示した。

切片のみが異なる 2 本の 2 次曲線のあてはめ

図 3.4 に示した散布図から、対数目盛上では、非喫煙群の 40 歳では、回帰直線からの乖離が大きく、全体的には各点が上に凸であり、喫煙群でも同様に上に凸となっている。統計的に意味があるかを検討するためには、2 次曲線のあてはめの必要性が示唆される。ただし、2 次曲線のあてはめ、統計的に有意となっても、直線のあてはめには問題があると指摘するだけで、2 次曲線の回帰パラメータに対する意味づけは困難である。

複数の直線のあてはめの場合でも、切片が同じで傾きが異なるモデル、切片は異なるが傾きが同じモデル、別々のモデル、3 通りのモデルを第 3.5 節で示してきた。1 本の 2 次曲線の場合についてならば、迷うことはないのであるが、2 本の場合については、年齢に依存せず喫煙習慣によらず死亡数が異なるかを年齢の 2 乗を追加して検討する。

表 3.37 年齢について 2 乗を追加

切片	喫煙	年齢	年齢 ²	死亡	人年	10万人比	推定値	10万人比	対数尤度		
x_0	x_1	$x_2:age$	$x_3:age^2$	y	n	y'	y^{\wedge}	$y^{\wedge'}$	$\ln Li$		
1	0	40	16	2	18,790	11	6.7	35.9	-3.6200	$\beta_0^{\wedge} =$	-17.8583
1	0	50	25	12	10,673	112	17.4	162.6	-3.0947	$\beta_1^{\wedge} =$	0.3547
1	0	60	36	28	5,710	490	28.5	499.4	-2.5927	$\beta_2^{\wedge} =$	0.3258
1	0	70	49	28	2,585	1,083	26.9	1040.4	-2.6105	$\beta_3^{\wedge} =$	-0.1942
1	0	80	64	31	1,462	2,120	21.5	1469.9	-4.4873		
1	1	40	16	32	52,407	61	26.8	51.2	-3.1253		
1	1	50	25	104	43,248	240	100.2	231.8	-3.3112		
1	1	60	36	206	28,612	720	203.7	712.1	-3.5959		
1	1	70	49	186	12,663	1,469	187.8	1483.4	-3.5413		
1	1	80	64	102	5,317	1,918	111.4	2095.7	-3.6430		
			$(age/10)^2$						$\ln L =$		-33.6218

縮小モデル（切片＋喫煙習慣＋年齢）の場合のマイナス 2 倍の対数尤度は、124.2500 であるのに対し、完全モデル（切片＋喫煙習慣＋年齢＋年齢²）の場合は、67.2435 と大幅に減少し、その差は 57.0065 であり、年齢²を加えることは統計的に支持される。

モデル	対数尤度	マイナス 2 倍	との差
切片+喫煙習慣+年齢 (縮小)	-62.1250	124.2500	
切片+喫煙習慣+年齢+年齢 ²	-33.6218	67.2435	57.0065

図 3.5 に 2 本の 2 次曲線を描いた結果を示す。実目盛り上では、非喫煙者の 80 歳が上方に大きく外れているが、母集団の数が小さいための誤差的な変動と解釈される。しかし、図 3.5 右の対数目盛のあてはまりを見ると 40 歳代の非喫煙群でのあてはまりは悪くなっている。

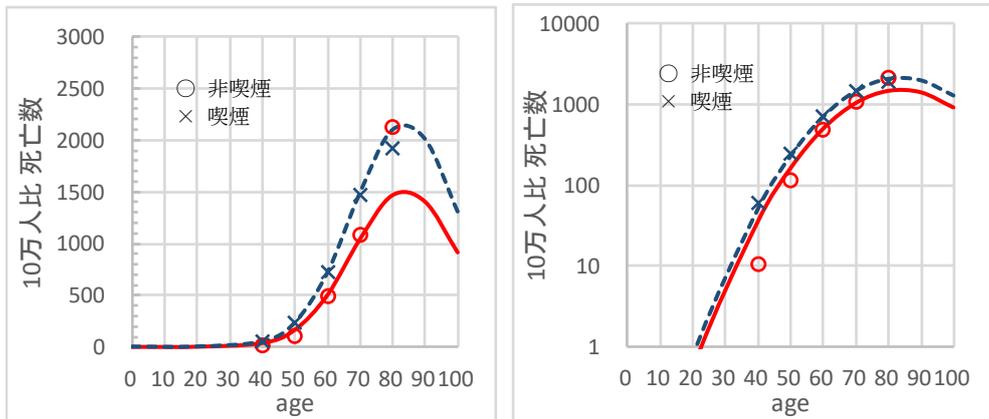


図 3.5 喫煙歴別の対数リンクでの 2 本の 2 次曲線

年齢が 60 歳の場合の 10 万人比の推定死亡数は、

$$\text{非喫煙者} : \hat{y}_{0, \text{age}=60} = 100,000 \times \exp(-17.8583 + 0.3547 \times 0 + 0.3258 \times 60 - 0.1942 \times 6^2) = 499.4 \text{ 人}$$

$$\text{喫煙者} \quad \hat{y}_{1, \text{age}=60} = 100,000 \times \exp(-17.8583 + 0.3547 \times 1 + 0.3258 \times 60 - 0.1942 \times 6^2) = 712.1 \text{ 人}$$

と推定される。

その次に、(切片+年齢+年齢²)を縮小モデルとして(切片+喫煙習慣+年齢+年齢²)との差により喫煙習慣の統計的性能を評価することになる。

縮小モデル:				
切片	$\hat{\beta}_0 =$	-17.7907	ln L =	-39.5501
	$\hat{\beta}_1 =$	-		
年齢	$\hat{\beta}_2 =$	0.3332		
年齢 ²	$\hat{\beta}_3 =$	-0.2000		

縮小モデルとの差の尤度比カイ 2 乗値は、11.8566 と統計的に有意な差となっている。

モデル	対数尤度	マイナス 2 倍	縮小モデルとの差
切片+年齢+年齢 ² (縮小)	-39.5501	67.2435	
切片+喫煙習慣+年齢+年齢 ²	-33.6218	79.1002	11.8566

喫煙習慣と年齢の交互作用を含む 2 本の 2 次曲線のあてはめ

図 3.5 で、喫煙習慣によって切片が異なるモデルによって 2 次曲線のあてはめを仔細にみると、良くあてはまっているとは、言い難い。そのために、喫煙習慣と年齢の交互作用を含めたモデルにチャレンジしたくなる。ただし、厄介なのは、喫煙習慣の違いによって異なる 2 次式をあてはめる場合には、交互作用として (喫煙習慣×年齢) を加え、さらに (喫煙習慣×年齢²) を加える必要があるかである。

厄介なのは、喫煙習慣との交互作用を含めた場合には、統計的に有意となった場合には、年齢ごとの喫煙習慣による死亡者数の比較が必要となり、これまで示してきたような、年齢にかかわらず、喫煙習慣による死亡者を論ずることができなくなってしまう。

交互作用	モデル
なし(縮小)	切片+喫煙習慣+年齢+年齢 ²
1 次のみ	切片+喫煙習慣+年齢+年齢 ² + (喫煙習慣×年齢)
2 次まで	切片+喫煙習慣+年齢+年齢 ² + (喫煙習慣×年齢) + (喫煙習慣×年齢 ²)

このような変数が多くなると、対数尤度 $\ln L$ を最大化するための適当な初期値の設定が困難になる。与えた初期値では Excel のソルバーによって $\ln L$ を最大化できなくなることもしばしば起きる。また、パラメータ共分散行列を求めるために行列計算を行うことも極めて煩雑になるので、統計ソフト JMP でパラメータを推定し、Excel で整理し、さらにグラフ化する。表 3.38 に解析用の JMP データセットを示す。

一般化線形モデルで分布をポアソン、リンク関数を対数、オフセットを $\ln(n)$ とする。変数は、切片モデル、切片+喫煙習慣、..., 切片+喫煙習慣+...+ (喫煙習慣×年齢²) のように逐次変数を増やして解析し、得られたパラメータの推定値を順次表 3.39 に示す Excel シートにコピーする。それぞれの (-1)*対数尤度もコピーし、2 倍した後に差分を計算する。この差分が追加された変数に対する尤度比カイ 2 乗値となるので、自由度 1 のカイ 2 乗分布の上側確率を $\text{Chisq.dist.RT}()$ 関数で p 値を求めている。

表 3.38 喫煙習慣と年齢の交互作用解析のための JMP データセット

	x0	x1:smoke	x2:age	x3:age^2	x4:S*age	x5:S*a^2	y	n	ln n
1	1	0	40	16	0	0	2	18,790	9.8411
2	1	0	50	25	0	0	12	10,673	9.2755
3	1	0	60	36	0	0	28	5,710	8.6500
4	1	0	70	49	0	0	28	2,585	7.8575
5	1	0	80	64	0	0	31	1,462	7.2876
6	1	1	40	16	40	16	32	52,407	10.8668
7	1	1	50	25	50	25	104	43,248	10.6747
8	1	1	60	36	60	36	206	28,612	10.2616
9	1	1	70	49	70	49	186	12,663	9.4464
10	1	1	80	64	80	64	102	5,317	8.5787

表 3.39 変数の逐次増加に対する推定値および尤度比カイ 2 乗値

変数	β_0^{\wedge}	$\beta_0^{\wedge} \sim \beta_1^{\wedge}$	$\beta_0^{\wedge} \sim \beta_2^{\wedge}$	$\beta_0^{\wedge} \sim \beta_3^{\wedge}$	$\beta_0^{\wedge} \sim \beta_4^{\wedge}$	$\beta_0^{\wedge} \sim \beta_5^{\wedge}$
x_0 :	-5.5144	-5.9618	-10.6258	-17.8675	-19.7003	-21.4905
x_1 :smoke		0.5422	0.4064	0.3545	2.3636	4.4053
x_2 :age			0.0836	0.3261	0.3563	0.4136
x_3 :age ²				-0.1944	-0.1977	-0.2421
x_4 :s×age					-0.0308	-0.0964
x_5 :s×age ²						0.0511
	x_0	x_1 :smoke	x_2 :age	x_3 :age ²	x_4 :s×age	x_5 :s×age ²
(-1)対数尤度	495.0676	480.5221	62.1250	33.6217	28.3517	28.1568
2倍	990.1353	961.0441	124.2500	67.2435	56.7033	56.3136
差分		29.0911	836.7941	57.0065	10.5402	0.3898
p 値		0.0000	0.0000	0.0000	0.0012	0.5324

「差分」に対する p 値から、(喫煙習慣×年齢²) の追加は、統計的に支持されない。最終的なモデルとして

$$\text{切片} + \text{喫煙習慣} + \text{年齢} + \text{年齢}^2 + (\text{喫煙習慣} \times \text{年齢})$$

が、ポアソン回帰モデルとして有効である。ただし、前にも述べたように、喫煙習慣の統計的な検討には難点がある。

推定されたパラメータを用いて、喫煙習慣と年齢の交互作用を加えた推定曲線を図 3.6 および図 3.7 に示す。図 3.6 は、統計的には望ましいのであるが、曲線を 80 歳以上に外挿したときに動きが大きいくことに難点がある。

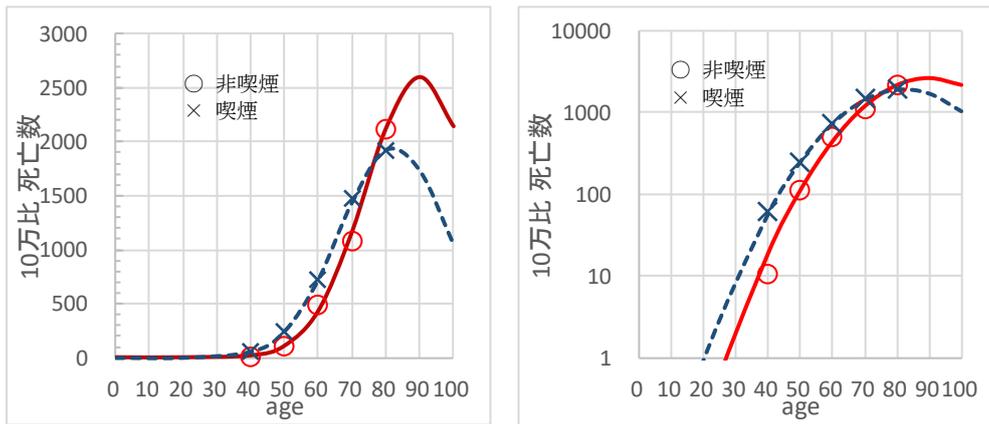


図 3.6 切片+喫煙習慣+年齢+年齢²+ (喫煙習慣×年齢)
ベストモデル

図 3.7 は、喫煙習慣と年齢の 2 乗の交互作用項を入れたモデルで、喫煙習慣について別々に 2 次曲線をあてはめた場合に相当する。2 次曲線のピークが、80 歳から 90 歳の間に含まれること、それらのピークがほぼ重なり合うことから、70 歳代までは喫煙群は非喫煙群に対して死亡数が多いが、80 歳以後には死亡数の差がほとんどなくなることを意味している。これは、喫煙者でも 80 歳まで生きてたとの条件下では、それ以後に死亡確率は非喫煙者と同等であることを意味している。

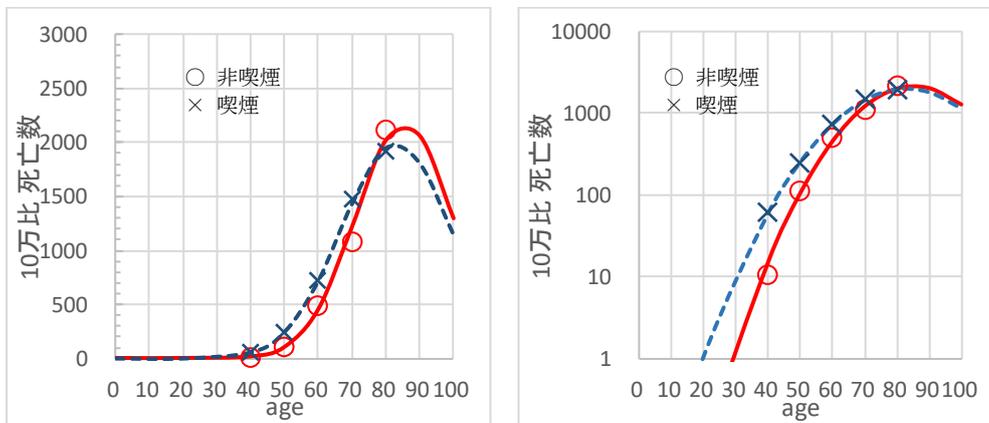


図 3.7 切片+喫煙習慣+年齢+年齢²+ (喫煙習慣×年齢) + (喫煙習慣×年齢²)
パラメータ過剰モデル

なお、第 12.6 節では、パラメータの共分散行列を用いて 95%信頼区間の計算およびグラフ表示を詳細に示す。

第3章 文献索引

Agresti.A. (2013) - Categorical Data Analysis 3rd ed

99

第3章 索引

あ	Agresti (2013) - 分割表	99	- 交互作用	123
	- 尤度比検定	99	- データセット	133
	アルトマン (1999) - 犯罪件数	95	- 2×2の分割表	96
	(1, -1) - 対比型	113	- 標示型	118
	一般化線形モデル - JMP	100	10万人比 - オフセット	125
	- 2項分布	100	縮小モデル - 完全モデル	126
	Indicator型 - 標示型	117	- 尤度比検定	98
	Ames試験 - コロニー数	109	出現確率 - 尤度比検定	97
	- 吉村ら (1992)	109	切片がない - 尤度比検定	103
	Excel - Chisq.dist.RT() 関数	132	切片を共通 - 2本の回帰直線	119
	- 計算不能	106	(0, 1) 型 - デザイン行列	115
	- 尤度比のカイ2乗値	96	ソルバー - ロジスティック回帰	102
	Excelソルバー - ロジスティック回帰	101	た 対数リンク - オフセット	125
	オフセット - 10万人比	125	- 2次曲線	125
	- 対数リンク	125	- 2本の直線	127
	- 部分母集団	125	対比型 - (1, -1)	113
か	Chisq.dist.RT() 関数 - Excel	132	- デザイン行列	112
	カイ2乗検定統計量 - Pearson	96	- ポアソン回帰	110
	回帰パラメータ - ロジット変換	100	たすき掛け - 交互作用	111
	回帰直線 - 2本	119	ダミー変数 - 1.5	106
	- 別々	124	- 0.5	106
	過剰モデル - ベストモデル	133	- 自己責任	107
	傾きを共通 - 平行線 (0, 1) 型	121	- デザイン行列	108
	冠動脈疾患 - ドブソン (2008)	125	逐次増加 - ポアソン回帰	133
	冠動脈心疾患 - 喫煙習慣	125	T - 転置記号	112
	完全モデル - 縮小モデル	126	デザイン行列 - (1, 0) 型	115
	- 尤度比検定	98	- (1, 2) 型	115
	簡便公式 - 分割表	99	- 矩形データ	109
	基準との差 - 標示型	117	- 計画行列	95
	(非喫煙・喫煙) - 2群間比較	126	- (0, 1) 型	115
	喫煙習慣 - 冠動脈心疾患	125	- 対比型	112
	喫煙習慣と年齢 - 交互作用	131	- ダミー変数	108
	共分散行列 - ワールド統計量	106	- 2×2	108
	吟味 - 交互作用	118	- 2本の回帰直線	119
	矩形データ - デザイン行列	109	- (-1, 1) 型	115
	計画行列 - デザイン行列	95	- 尤度比検定	95
	- 尤度比検定	95	データセット - JMP	133
	計算不能 - Excel	106	転置記号 - T	112
	交互作用 - 喫煙習慣と年齢	131	ドブソン (2008) - 冠動脈疾患	125
	- 吟味	118	富山ら (2004) - 用量反応試験	119
	- JMP	123	な 2×2 - デザイン行列	108
	- たすき掛け	111	- 要因配置実験	108
	- 2本の2次曲線	132	2×2の分割表 - JMP	96
	- ポアソン回帰	123	- ピアソンのカイ2乗検定	95
	- 薬剤と濃度	122	- 尤度比検定	95
	- 尤度比検定	123	2群間の比較 - ポアソン回帰	104
	コロニー数 - Ames試験	109	2群間比較 - (非喫煙・喫煙)	126
さ	差分 - 尤度比カイ2乗	100	- 2項分布	100
	シングモイド曲線 - ロジスティック分布	102	2元配置型 - ポアソン回帰	110
	自己責任 - ダミー変数	107	2項分布 - 一般化線形モデル	100
	JMP - 一般化線形モデル	100	- 2群間比較	100

2次曲線 - 対数リンク	125	ロジスティック分布 - シグモイド曲線	102
- 2本	125	ロジット - リンク関数	100
	130	ロジット変換 - 回帰パラメータ	100
- 年齢	130	わ ワールド統計量 - 共分散行列	106
2種類の検定 - 分割表	95		
2値反応 - ベルヌーイ分布	98		
2本 - 回帰直線	119		
- 2次曲線	125		
	130		
2本の2次曲線 - 交互作用	132		
2本の回帰直線 - 切片を共通	119		
- デザイン行列	119		
2本の直線 - 対数リンク	127		
年齢 - 2次曲線	130		
は 犯罪の有無 - 満月と新月	95		
犯罪件数 - アルトマン (1999)	95		
- ポアソン回帰	105		
Pearson - カイ2乗検定統計量	96		
ピアソンのカイ2乗検定 - 2×2の分割表	95		
標示型 - Indicator型	117		
- 基準との差	117		
- JMP	118		
部分母集団 - オフセット	125		
分割表 - Agresti (2013)	99		
- 簡便公式	99		
- 2種類の検定	95		
平行線 (0, 1) 型 - 傾きを共通	121		
ベストモデル - 過剰モデル	133		
別々 - 回帰直線	124		
ベルヌーイ分布 - 2値反応	98		
ポアソン回帰 - 対比型	110		
- 逐次増加	133		
- 2群間の比較	104		
- 2元配置型	110		
- 犯罪件数	105		
ポアソン分布 - 尤度比検定	104		
ま 満月と新月 - 犯罪の有無	95		
モデル - 完全モデル	98		
- 縮小モデル	98		
や 薬剤と濃度 - 交互作用	122		
尤度比カイ2乗 - 差分	100		
尤度比のカイ2乗値 - Excel	96		
尤度比検定 - Agresti (2013)	99		
- 完全モデル	98		
- 計画行列	95		
- 交互作用	123		
- 縮小モデル	98		
- 出現確率	97		
- 切片がない	103		
- デザイン行列	95		
- 2×2の分割表	95		
- ポアソン分布	104		
用量反応試験 - 富山ら (2004)	119		
要因配置実験 - 2×2	108		
吉村ら (1992) - Ames試験	109		
ら リンク関数 - ロジット	100		
ロジスティック回帰 - Excelソルバー	101		
- ソルバー	102		

第3章 解析用ファイル一覧

	2 KB	第3章01_満月新月_01	JMP Data Table
	17 KB	第3章01_満月新月_01	Microsoft Excel ワークシート
	3 KB	第3章02_満月新月_01ロジット	JMP Data Table
	31 KB	第3章02_満月新月_01ロジット	Microsoft Excel ワークシート
	6 KB	第3章03_満月新月_ポアソン	JMP Data Table
	19 KB	第3章03_満月新月_ポアソン	Microsoft Excel ワークシート
	11 KB	第3章04_細菌_01型	JMP Data Table
	27 KB	第3章04_細菌_2x2	Microsoft Excel ワークシート
	30 KB	第3章05_Ames_用量反	Microsoft Excel ワークシート
	3 KB	第3章05_Ames_用量反応	JMP Data Table
	7 KB	第3章06_タバコと冠動脈疾患	JMP Data Table
	101 KB	第3章06_タバコと冠動脈疾患	Microsoft Excel ワークシート
	4 KB	第3章06_タバコと冠動脈疾患b	JMP Data Table

非売品, 無断複製を禁ずる

第9回 続高橋セミナー

最尤法によるポアソン回帰分析入門

第3章 尤度比検定のためのデザイン行列

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3 の1005

2020年5月8日 高橋 行雄