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 11 日

第 11 章　デビアンス・逸脱度・テコ比・4 種の残差

一般化線形モデルによるポアソン回帰分析では，通常の回帰分析とは結果の表記法がかなり異なり，なかなか馴染めないのではないだろうか．デビアンス・逸脱度は，その代表的な例であろう．逸脱度を理解するためには対数尤度と最尤法の理解も必要となり最小 2 乗法に慣れ親しんだきた人たちは，茫然自失の状態になるかもしれない．さらに，スチューデント化デビアンス残差に関連して「テコ比」も登場する．「テコ比」は，通常の回帰分析にも登場することもあるが，マイナーな存在である．そこで，通常の回帰分析と対比しつつポアソン回帰で使われている統計用語とその意味づけについてこれまで取り上げてきた事例を用いて関連付けを行う．

第 11 章　目 次

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

第 9 回 続高橋セミナー
最尤法によるポアソン回帰分析入門

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

目 次

11. デビアンス・逸脱度・テコ比・4 種の残差

一般化線形モデルによるポアソン回帰分析では，通常の回帰分析とは結果の表記法がかなり異なり，なかなか馴染めないのではないだろうか．デビアンス・逸脱度は，その代表的な例であろう．逸脱度を理解するためには対数尤度と最尤法の理解も必要となり最小 2 乗法に慣れ親しんだきた人たちは，茫然自失の状態になるかもしれない．さらに，スチューデント化デビアンス残差に関連して「テコ比」も登場する．「テコ比」は，通常の回帰分析にも登場することもあるが，マイナーな存在である．そこで，通常の回帰分析と対比しつつポアソン回帰で使われている統計用語とその意味づけについてこれまで取り上げてきた事例を用いて関連付けを行う．

11.1. デビアンス

第 1.9 節では，久保（2012）で取り上げられている種子数のデータについて表 1.34 に Excel で計算した 3 種の対数尤度（縮小モデル，完全モデル，飽和モデル）を示し，デビアンスを飽和モデルの対数尤度 $\ln L_{飽和}$ から完全モデルの対数尤度 $\ln L_{完全}$ の差の 2 倍，84.9930 として示した．このデビアンス 84.9930 が，飽和モデルの自由度 100 に対し，完全モデルの自由度 2 との差の自由度 98 のカイ 2 乗分布に従うことから，上側確率が 0.8226 となり，100 個のパラメータで推定した飽和モデルに対し，2 個のパラメータで推定した完全モデルが統計的に遜色ないことを示した．

単にデビアンスと言う場合は，データに対して何らかの仮定をした 2 つのモデルから計算される対数尤度の差の 2 倍した統計量を意味する．ややこしいのは，統計ソフトを使うと「デビエンス残差」あるいは「スチューデント化デビアンス残差」なども出てくる．第 1.9 節で JMP によるポアソン回帰で出力される「スチューデント化デビアンス残差プロット」を図 1.9 に示したが，その計算方法を示さずに「ほとんどが（−2〜+2）の範囲に入っていることからあてはめの妥当性が示されている」との解説をしている．なぜ，そのような判断ができるのか，本章で Excel による計算を通して理解を深めたい．

第 9.5 節では，R による負の 2 項回帰（ガンマ・ポアソン回帰）による結果を文献から引用した際に「Deviance Residuals］と「Residual deviance」について Excel での計算例を示し，「デビアンス・逸脱度」について断片的な解説をした．しかし，他では意図的に使ってこなかった．それに代えてマイナス 2 倍の対数尤度，対数尤度の差の 2 倍，尤度比カイ 2 乗値を用い

てきた．第 1.9 節では，Pearson のカイ 2 乗値の計算方法を示したが，デビエンスについては深入りせず，Pearson のカイ 2 乗値を多用してきた．これは，デビアンス残差の計算で用いている「飽和モデルの対数尤度」の概念が通常の最小 2 乗法の世界で，対応する概念がないために，あえて言及しなかった．

本章では，通常の回帰分析で標準的に使われている「分散分析表」の概念と比較しながらデビアンスついて解説する．さらに，通常の回帰分析で行われている通常の残差に加え，スチューデント化残差（標準化残差）と対比して（Peason 残差，スチューデント化 Peason 残差，デビアンス残差，スチューデント化デビアンス残差）について，統計ソフトの結果と対比しつつ Excel を用いた計算法を提示し，それらの使い分けについて概説する．

Excel での各種の残差を計算する際に，厄介な問題に直面する．それは，テコ比 h_{ii} がハット行列 $\boldsymbol{H}$

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

の対角要素と定義されているため，Excel の行列関数で対角要素を取り出すことが容易ではない．そのために，デザイン行列 $\boldsymbol{X}$ の i 行目のベクトル $\boldsymbol{x}_i$ を使った

$$h_{ii} = \boldsymbol{x}_i(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i^T$$

計算により代替する．この式を使って $h_{1,1}$， $h_{2,2}$, . . . を列ベクトルとして得ることができる．このような，Excel による逐次的な計算手順を経験することにより，テコ比 h_{ii} の意味付けと活用方法について理解を深めてもらいたい．

テコ比 h_{ii} を使うことにより，通常の回帰分析でのスチューデント化残差（標準化残差）を計算することができるようになり，ポアソン回帰では，スチューデント化デビアンス残差の計算が可能となる．

11.2.　通常の回帰分析におけるスチューデント化残差

回帰パラメータの推定

第 4 章で示した Excel の行列関数を用いた回帰分析について要点をまとめる．用いるデータは，第 1.4 節のポアソン回帰のための人工データである［ドブソン（2008)］．表 11.1 に示すように，まず，説明変数 X の平均からの偏差を計算し，それらの平方から平方和 S_{XX} を計算する．次に，反応変数 Y の偏差と X の偏差の積和 S_{XY} を計算し，それらから，回帰パラメータ傾き $\hat{\beta}_1$ を

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{24.000}{4.8889} = 4.9091$$

により計算する．説明変数 X の平均 $\bar{X}$ と反応変数 Y の平均 $\bar{Y}$ と傾き $\hat{\beta}_1$ を用いて回帰パラメータの切片 $\hat{\beta}_0$ を，

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 8.0 - 4.9091 \times 0.1111 = 7.4545$$

として計算する．なお，表中に $S_T = S_{YY}$ があるが，回帰パラメータの推定には使われない．

表 11.1　回帰パラメータの Excel シート上での推定（表 4.1 再掲）

i	X	Y	X偏差	X偏差^2	Y偏差	Y偏差^2	XY偏差		
1	-1	2	-1.1111	1.2346	-6.0000	36.0000	6.6667	β_1^=	4.9091
2	-1	3	-1.1111	1.2346	-5.0000	25.0000	5.5556	β_0^=	7.4545
3	0	6	-0.1111	0.0123	-2.0000	4.0000	0.2222		
4	0	7	-0.1111	0.0123	-1.0000	1.0000	0.1111		
5	0	8	-0.1111	0.0123	0.0000	0.0000	0.0000		
6	0	9	-0.1111	0.0123	1.0000	1.0000	-0.1111		
7	1	10	0.8889	0.7901	2.0000	4.0000	1.7778		
8	1	12	0.8889	0.7901	4.0000	16.0000	3.5556		
9	1	15	0.8889	0.7901	7.0000	49.0000	6.2222		
	0.1111	8.0000	0.0000	4.8889	0.0000	136.0000	24.0000		
	平均	平均	合計	平方和	合計	平方和	平方和		
	$\bar{X}$	$\bar{Y}$		S_{XX}		S_T=S_{YY}	S_{XY}		

Excel での一般的な計算方法は，セルに対し計算式を与え，そのセルをプルダウンして参照セルを変化させつつ計算式をコピーする．これは，Excel の画期的な計算機能であるが，操作ミスがあっても発見しづらい．セルごとの計算に代え，範囲を使った計算を使うのが確実である．たとえば，8 行 1 列の「X 偏差」は，=(***X*** の範囲－X の平均）のように行列計算の要領で一括計算ができる．「X 偏差^2」は，=(***X*** 偏差の範囲)^2 によって 8 行 1 列分の計算を一括して行っている．「XY 偏差」は，＝(***X*** 偏差 * ***Y*** 偏差) として一括計算している．範囲を指定した場合の四則演算は，対応するセル同士の計算となり，片方がスカラーのような場合は，相手のサイズに合わせてくれる．

分散分析表

表 11.2 に示すように分散分析表に必要な平方和の計算を逐次的に行った結果を示す. なお, 各種の平方和の計算は, SumSq() 関数を使用するのが効率的である.

総平方和　$S_T = \sum_{i=1}^{9}(Y_i - \bar{Y})^2 = \mathrm{SumSq}(\boldsymbol{Y}\text{の範囲} - \bar{Y}) = 136.0000$　$df = 9-1 = 8$

推定値　$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \mathrm{Mmult}(\boldsymbol{X}\text{の範囲}, \hat{\boldsymbol{\beta}}\text{の範囲})$　$df = 2$

回帰の平方和　$S_R = \sum_{i=1}^{9}(\hat{Y}_i - \bar{Y})^2 = \mathrm{SumSq}(\hat{\boldsymbol{Y}}\text{の範囲} - \bar{Y}) = 117.8182$　$df = 2-1 = 1$

誤差平方和　$S_e = \sum_{i=1}^{9}(Y_i - \hat{Y}_i)^2 = \mathrm{SumSq}(\boldsymbol{Y}\text{の範囲} - \hat{\boldsymbol{Y}}\text{の範囲}) = 18.1818$　$df = 9-2 = 7$

回帰の平方和 S_R は,

$$S_R = S_T - S_e = 136.0000 - 18.1818 = 117.8182$$

としても計算できる.

表 11.2　回帰の平方和の計算（表 4.3 再掲）

i	$\boldsymbol{X}$		Y_i	$\bar{Y}$	$Y_i - \bar{Y}$	Y_i^	Y_i^ $- \bar{Y}$	$Y_i - Y_i$^		$\boldsymbol{\beta}$^
1	1	-1	2	8.00	-6.00	2.55	-5.45	-0.55	β_0^ =	7.4545
2	1	-1	3	8.00	-5.00	2.55	-5.45	0.45	β_1^ =	4.9091
3	1	0	6	8.00	-2.00	7.45	-0.55	-1.45		
4	1	0	7	8.00	-1.00	7.45	-0.55	-0.45		
5	1	0	8	8.00	0.00	7.45	-0.55	0.55		
6	1	0	9	8.00	1.00	7.45	-0.55	1.55		
7	1	1	10	8.00	2.00	12.36	4.36	-2.36		
8	1	1	12	8.00	4.00	12.36	4.36	-0.36		
9	1	1	15	8.00	7.00	12.36	4.36	2.64		
		5.00		8.0000	136.0000		117.8182	18.1818		
		ΣX^2		$\bar{Y}$	S_T		S_R	S_e		
自由度 df			9	1	9-1=8	2	2-1=1	9-2=7		2

計算結果を, 表 11.3 の分散分析表にまとめる. 分散分析表の自由度については, 各種の便宜的な説明が行なわれているが, 自由度の本質が把握しづらいので, ここに示したように偏差平方和をベースにした計算式に対応する自由度の計算法を示す.

表 11.3　回帰に対する分散分析表（表 4.4 再掲）

要因		平方和	自由度	平均平方	F 値	p 値
回帰	S_R	117.8182	2-1=1	117.8182	45.3600	0.0003
誤差	S_e	18.1818	9-2=7	2.5974		
全体	S_T	136.0000	9-1=8			

パラメータの共分散行列

分散分析表の誤差の平均平方（誤差分散 $\hat{\sigma}^2$）を用いて，回帰パラメータ $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$ は，正規方程式の解を用いて，式（4.23）および式（4.22）から

$$Var(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum X_i^2}{n\sum (X_i - \bar{X})^2}$$
$$= \frac{2.5974 \times 5.00}{9 \times 4.8889} = 0.2952$$

$$Var(\beta_1) = \frac{\hat{\sigma}^2}{\sum (X_i - \bar{X})^2}$$
$$= \frac{2.5974}{4.8889} = 0.5313$$

が得られる．また，式（4.25）から，それぞれの分散は，

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2 = \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} = \begin{bmatrix} \frac{\Sigma X_i^2}{n\Sigma (X_i - \bar{X})^2} & \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} & \frac{1}{\Sigma (X_i - \bar{X})^2} \end{bmatrix} \hat{\sigma}^2$$

パラメータの共分散行列の対角要素であることも示した．実際の計算は，

$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$=

0.1136	-0.0227
-0.0227	**0.2045**

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$

2.5974
σ^2

=

0.2952	-0.0590
-0.0590	**0.5313**

$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^2

となり，$Var(\hat{\beta}_0) = 0.2952$，$Var(\hat{\beta}_1) = 0.5313$ が得られる．それらの平方根から SE を計算しパラメータに関する t 検定が行なう．この行列計算の方法は，変数の数が増えても同じであり，偏差平方和をベースにした計算手順よりも簡潔である．詳しくは第 12.3 節で示す．

表 11.4　回帰パラメータの推定値（表 4.2 再掲）

項	推定値	分散	SE	t 値	p 値
β_0^	7.4545	0.2952	0.5433	13.72	0.0000
β_1^	4.9091	0.5313	0.7289	6.73	0.0003

スチューデント化残差

通常の回帰分析において，残差の検討の重要性は常に強調されている．代表的なのは，図 11.1 に示す予測値 $\hat{y}_i$ に対する残差 $\hat{\varepsilon}_i = y_i - \hat{y}_i$ のプロットおよびスチューデント化残差プロットである．この残差プロットは，予測値 $\hat{y}_i$ が大きくなるにつれ扇型に広がっている．したがって，推定された回帰直線に対して誤差が均一とはみなせないので，変数変換などで，誤差が均一になるような変換を検討することになる．

スチューデント化残差 $\hat{\varepsilon}_i'$ は，残差 $\hat{\varepsilon}_i$ の標準誤差で割って基準化したものであり，残差自体について統計的な考察ができる．図 11.1 右に示すように，扇型に広がっていはいるが残差 $\hat{\varepsilon}_i'$ は，(–2 ~ +2) の範囲に入っており，通常の回帰分析を適用しても差し支えないと判断される．

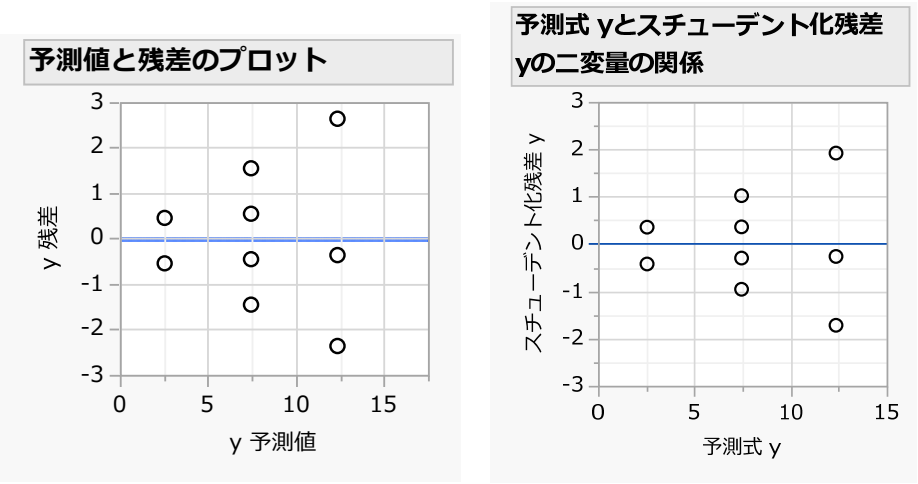


図 11.1　通常の回帰分析における残差およびスチューデント化残差プロット

テコ比・ハット行列

スチューデント化残差 $\hat{\varepsilon}_i'$ は，残差 $\hat{\varepsilon}_i$ の分散 $Var(\hat{\varepsilon}_i) = \hat{\sigma}^2(1-h_{ii})$ を考慮したものである．ここで h_{ii} は，通称テコ比ともいわれており，スチューデント化残差 $\hat{\varepsilon}_i'$ は，ハット行列の対角要素 h_{ii} を用いて基準化したもので

$$\hat{\varepsilon}_i' = \frac{\hat{\varepsilon}_i}{\sqrt{Var(\hat{\varepsilon}_i)}} = \frac{\hat{\varepsilon}_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}}$$

として計算されている．ハット行列 $\boldsymbol{H}$ は，$\boldsymbol{Y}$ の推定値を求める式 $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ に $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ を代入し，

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

最後の $\boldsymbol{Y}$ を除いた行列

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$$

で定義されている．実際に計算すると次のような結果となる．まず，$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ の結果を示す．次に，$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ を挟んで $\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$ の行列計算を行う．このようにハット行列 $\boldsymbol{H}$ の計算は，ストレスなく行えるのであるが，その対角要素を列ベクトル化する関数が Excel にないために，この式は使えない．

$\boldsymbol{X}^T$									$\boldsymbol{X}$			$\boldsymbol{X}^T\boldsymbol{X}$			$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$	
1	1	1	1	1	1	1	1	1	1	-1	=	**9**	1		**0.1136**	-0.0227
-1	-1	0	0	0	0	1	1	1	1	-1		1	**5**		-0.0227	**0.2045**
									1	0						
=Transpose($\boldsymbol{X}$の範囲)									1	0					=Minverse($\boldsymbol{X}^T\boldsymbol{X}$の範囲)	
									1	0						
									1	0						
									1	0		=Mmult($\boldsymbol{X}^T$の範囲, $\boldsymbol{X}$の範囲)				
									1	1						
									1	1						
									1	1						
			(2×9)						(9×2)			(2×2)			(2×2)	

$\boldsymbol{X}$		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		$\boldsymbol{X}^T$										$\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$								
1	-1	**0.1136**	-0.0227	1	1	1	1	1	1	1	1	1	=	**0.36**	0.36	0.14	0.14	0.14	0.14	-0.1	-0.1	-0.1
1	-1	-0.0227	**0.2045**	-1	-1	0	0	0	0	1	1	1		0.36	**0.36**	0.14	0.14	0.14	0.14	-0.1	-0.1	-0.1
1	0													0.14	0.14	**0.11**	0.11	0.11	0.11	0.09	0.09	0.09
1	0													0.14	0.14	0.11	**0.11**	0.11	0.11	0.09	0.09	0.09
1	0													0.14	0.14	0.11	0.11	**0.11**	0.11	0.09	0.09	0.09
1	0													0.14	0.14	0.11	0.11	0.11	**0.11**	0.09	0.09	0.09
1	1													-0.1	-0.1	0.09	0.09	0.09	0.09	**0.27**	0.27	0.27
1	1													-0.1	-0.1	0.09	0.09	0.09	0.09	0.27	**0.27**	0.27
1	1													-0.1	-0.1	0.09	0.09	0.09	0.09	0.27	0.27	**0.27**
(9×2)		(2×2)					(2×9)										(9×9)					
												=Mmult(Mmult($\boldsymbol{X}$の範囲, $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$の範囲),$\boldsymbol{X}^T$の範囲)								

ハット行列 $\boldsymbol{H}$ 対角要素を h_{ii} としたとき，デザイン行列 $\boldsymbol{X}$ の i 行目のベクトルを $\boldsymbol{x}_i$ を用いることにより，テコ比 h_{ii} は，

$$h_{ii} = \boldsymbol{x}_i(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i^T$$

として求めることができる．この式で，それぞれの h_{ii} を計算することにより，ハット行列 $\boldsymbol{H}$ を計算しなくとも次式で示すように，対角要素をベクトル化することができる．

x_1		$(X^TX)^{-1}$		x_1^T		h_{11}
1	-1	**0.1136**	-0.0227	1	=	**0.3636**
		-0.0227	**0.2045**	-1		
:						
x_8		$(X^TX)^{-1}$		x_8^T		h_{88}
1	1	**0.1136**	-0.0227	1	=	**0.2727**
		-0.0227	**0.2045**	1		

テコ比の活用

単純な誤差のプロットに加えて，テコ比を考慮したスチューデント化残差（標準化残差）による検討も有益である．なお，テコ比とハット行列の意味付けについては，野沢（1992），「テコ比とハット行列」が詳しい．

テコ比 h_{ii} は，回帰の 95%信頼区間を求めるための分散 $Var(\hat{y}_i)$ の計算にも関係している．式（4.36）から，

$$\begin{aligned} Var(\hat{y}_i) &= \boldsymbol{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \boldsymbol{x}_i^T \\ &= \boldsymbol{x}_i [(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2] \boldsymbol{x}_i^T \\ &= h_{ii}\hat{\sigma}^2 \end{aligned}$$

が導かれる．個々のデータの分散 $\hat{\sigma}^2$ に対しテコ比 h_{ii} は，回帰の推定値の分散 $Var(\hat{y}_i)$ を求めるための割引係数として解される．表 11.5 に Excel でテコ比 h_{ii} を計算し，スチューデント化残差 $\hat{\varepsilon}_i'$ および分散 $Var(\hat{y}_i)$ を計算した結果を示す．

表 11.5　テコ比を用いたスチューデント化残差

					残差	テコ比	ス.化残差	残差 比	分散
i	X		Y	Y^	$\varepsilon = Y - Y$^	h_{ii}	ε'	ε'/ε	$Var(y$^$)$
1	1	-1	2	2.5455	-0.5455	**0.3636**	**-0.4243**	0.7778	**0.9445**
2	1	-1	3	2.5455	0.4545	0.3636	0.3536	0.7778	0.9445
3	1	0	6	7.4545	-1.4545	0.1136	-0.9586	0.6591	0.2952
4	1	0	7	7.4545	-0.4545	0.1136	-0.2996	0.6591	0.2952
5	1	0	8	7.4545	0.5455	0.1136	0.3595	0.6591	0.2952
6	1	0	9	7.4545	1.5455	0.1136	1.0185	0.6591	0.2952
7	1	1	10	12.3636	-2.3636	0.2727	-1.7197	0.7276	0.7084
8	1	1	12	12.3636	-0.3636	0.2727	-0.2646	0.7276	0.7084
9	1	1	15	12.3636	2.6364	0.2727	1.9182	0.7276	0.7084
	9.00	1.00	0.1136	-0.0227	72.0000	β_0=	7.4545	$\varepsilon^T\varepsilon$=	18.1818
	1.00	5.00	-0.0227	0.2045	32.0000	β_1=	4.9091	σ^^{2}=	**2.5974**
	X^TX		$(X^TX)^{-1}$		X^TY	$(X^TX)^{-1}X^TY$			

実際の計算過程を $i=1$ の場合について示す．まずテコ比 h_{11} を計算し，

h_{11} =

1	-1	**0.1136**	-0.0227	1	=	**0.3636**
		-0.0227	**0.2045**	-1		
$\boldsymbol{x}_1$		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		$\boldsymbol{x}_1^T$		h_{11}

スチューデント化残差 $\hat{\varepsilon}_1'$ を求める．

$$\hat{\varepsilon}_1' = \frac{\hat{\varepsilon}_1}{\sqrt{\hat{\sigma}^2(1-h_{11})}} = \frac{-0.5455}{\sqrt{2.5974\times(1-0.3636)}} = -0.4243$$

テコ比 h_{11} を用いて回帰の推定値 $\hat{y}_1$ の分散 $Var(\hat{y}_1)$ の計算もできる．

$$Var(\hat{y}_1) = h_{11}\hat{\sigma}^2 = 0.3636\times 2.5974 = 0.9445$$

なお，図 11.1 右に示したスチューデント化残差プロットは，JMP のスチューデント化残差プロットに X 軸に予測値を指定できないので，スチューデント化残差をファイルに出力して，別途「二変量の関係」を使って作図したものである．

テコ比は，回帰直線の推定値の分散を誤差分散 $\hat{\sigma}^2$ に対して割引係数としても理解される．回帰直線の中心部は小さく，外側に向かって大きくなり，推定値の分散が大きくなり 95%信頼区間の幅の変化を示す統計量とも解される．

Excel の「標準残差」に対する使用上の注意

Excel の回帰分析の活用は，分散分析表および回帰パラメータの推定値に関してデザイン行列の計算の煩わしさを軽減するために有益であることを示したきた．さらに，Excel の回帰分析で「残差」に加えて「標準残差」を Excel シートに出力することができる．「標準残差」はスチューデント化残差（標準化残差）と紛らわしいが，テコ比を含まない計算であり，別物である．表 11.5 の場合では，全ての残差 $\hat{\varepsilon}_i$ に 0.6633 を掛けている．これは，分散 $\hat{\sigma}^2$ の平方根の逆数 0.6205 に近い値であるが，どのような計算なのかは不明である．いずれにしても，スチューデント化残差（標準化残差）ではないことに注意が必要である．

Excel の統計解析については，正確性が欠けるとの指摘があることは十分に承知しており，「標準残差」もその一例であろう．もっとひどい事例は，第 10.3 節で例示した，折れ線グラフの誤差範囲の設定のいい加減さであり，それを回避する方法を知る必要がある．なお，行列計算などの精度で問題になった経験はないが，計算過程の脆弱性を常に認識し，他のソフトでの検証を怠ってはならない．

11.3.　ポアソン回帰におけるデビアンス・逸脱度

対数尤度を用いた恒等リンクのポアソン回帰は，回帰式を

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i , \qquad \varepsilon_i \sim \text{ポアソン分布}$$

としたときに，対数尤度 $\ln L_{回帰}$ を最大にするような回帰パラメータ $\hat{\beta}_0$ と $\hat{\beta}_1$

$$\ln L_{回帰} = \sum_i \ln[\text{Poisson.dist}(y_i, \hat{\beta}_0 + \hat{\beta}_1 x_i, false)]$$

を推定する．$\ln L_{回帰}$ を最大化するために，第 1.4 節および第 5 章で反復重み付き回帰を用いる方法，第 2 章で対数尤度関数の 2 階の偏微分式行列を用いる方法，第 1.9 節で Excel のソルバーにより $\ln L_{回帰}$ が最大になるような $\hat{\beta}_0$ と $\hat{\beta}_1$ を直接求める方法を示してきた．ここでは，簡便な Excel のソルバーを用いる方法を使う．

デビアンス・カイ 2 乗

ポアソン回帰では，第 1.9 節で示したように飽和モデル，完全（最大）モデル，縮小モデル，切片のみの場合には（null モデル）など幾つかの「モデル」が登場する．そして，逸脱度/デビアンスは，飽和モデルと各モデルとのマイナス 2 倍の対数尤度の差で定義されている．

「飽和モデル」の概念が通常の回帰分析にはないので，理解に苦しむことになる．表 11.6 に示すように，飽和モデルの対数尤度 $\ln L_{飽和}$ は，y_i のポアソン分布の確率を求めるための推定値として $\hat{y}_i = y_i$ のように自分自身 y_i を用いて，

$$\text{飽和モデル：}\ \hat{y}_i = y_i ,\begin{cases} \ln L_{飽和} = \sum_i \ln(\text{Poisson.dist}(y_i, y_i, false)) \\ \quad = \ln(0.2707) + \ln(0.2240) + \cdots + \ln(0.1024) \\ \quad = -1.3069 - 1.4959 - \cdots - 2.2785 \\ \quad = -17.0566 \end{cases}$$

としてポアソン分布の確率を計算している．通常の回帰分析は，偏差平方和の計算で組み立てられているので，無理に飽和モデルを考えても，次のようにゼロなので

$$S_{飽和} = \sum_i (y_i - y_i)^2 = 0$$

何の役にも立たない．したがって，飽和モデルの概念がない．（完全 or 最大 or 回帰）モデルは，$\hat{\beta}_0$ と $\hat{\beta}_1$ を使った回帰モデルで，$\ln L_{回帰}$ に変えて $\ln L_{完全}$ とし，

$$\text{完全モデル，}\ \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \ :\begin{cases} \ln L_{完全} = \sum_i \ln(\text{Poisson.dist}(y_i, \hat{\beta}_0 + \hat{\beta}_1 x_i, false)) \\ \quad = \ln(0.2557) + \ln(0.2144) + \cdots + \ln(0.0791) \\ \quad = -1.3639 - 1.5397 - \cdots - 2.5366 \\ \quad = -18.0039 \end{cases}$$

となる．これは，表 11.7 に示すように通常の回帰分析の「誤差平方和 S_e」に該当する．

表 11.6　最尤法によるポアソン回帰

			縮小（null）モデル			完全（最大）モデル			飽和モデル		
			$y^=\beta_0^$			$y^=\beta_0^+\beta_1^x$			$y^=y$		
				$\beta_0^=$	**8.0000**		$\beta_0^=$	**7.4516**			
							$\beta_1^=$	**4.9353**			
				$\ln L_{縮小}=$	**-26.2669**		$\ln L_{完全}=$	**-18.0039**		$\ln L_{飽和}=$	**-17.0566**
			$\beta_0^$	確率	対数尤度	回帰	確率	対数尤度	y	確率	対数尤度
i	x	y	$y^$	P	$\ln L_i$	$y^$	P	$\ln L_i$	$y^$	P	$\ln L_i$
1	-1	2	**8.00**	0.0107	-4.5343	**2.52**	0.2557	-1.3639	**2**	0.2707	-1.3069
2	-1	3	**8.00**	0.0286	-3.5534	**2.52**	0.2144	-1.5397	**3**	0.2240	-1.4959
3	0	6	**8.00**	0.1221	-2.1026	**7.45**	0.1380	-1.9803	**6**	0.1606	-1.8287
4	0	7	**8.00**	0.1396	-1.9691	**7.45**	0.1469	-1.9178	**7**	0.1490	-1.9038
5	0	8	**8.00**	0.1396	-1.9691	**7.45**	0.1369	-1.9888	**8**	0.1396	-1.9691
6	0	9	**8.00**	0.1241	-2.0869	**7.45**	0.1133	-2.1776	**9**	0.1318	-2.0268
7	1	10	**8.00**	0.0993	-2.3100	**12.39**	0.0978	-2.3249	**10**	0.1251	-2.0786
8	1	12	**8.00**	0.0481	-3.0339	**12.39**	0.1137	-2.1744	**12**	0.1144	-2.1683
9	1	15	**8.00**	0.0090	-4.7076	**12.39**	0.0791	-2.5366	**15**	0.1024	-2.2785

（縮小 or 切片 or Null）モデルは，

$$\text{縮小モデル，}\ \hat{y}_i=\hat{\beta}_0\ :\begin{cases}\ln L_{縮小}=\sum_i \ln(\text{Poisson.dist}(y_i,\hat{\beta}_0,false))\\ \quad=\ln(0.0107)+\ln(0.0286)+\cdots+\ln(0.0090)\\ \quad=-4.5343-3.5534-\cdots-4.7076\\ \quad=-26.2669\end{cases}$$

として計算される．通常の回帰分析の「総平方和 S_T」に該当する．

回帰の平方和 S_R は，$S_R=S_T-S_e$ で求められると同様に，傾き $\hat{\beta}_1$ に対する 2 倍の対数尤度は，差分

$$\text{差分}:\begin{cases}2\ln L_R=2(\ln L_{完全}-\ln L_{縮小})\\ \quad=2\times[-18.0039-(-26.2669)]\\ \quad=16.5260\end{cases}$$

として求められる．この 2 倍の対数尤度に対する検定統計量は，それぞれの自由度の差のカイ 2 乗分布に従うことにより有意差検定が行なえる．完全モデル（$\hat{y}_i=\hat{\beta}_0+\hat{\beta}_1 x_i$）と縮小モデル（$\hat{y}_i=\hat{\beta}_0$）の対数用尤度の差 2 倍なので，$\hat{\beta}_1$ に対する尤度比検定統計量となり，カイ 2 乗検定が行なえる．

通常の回帰分析では，S_R をその自由度 df_R で割った平均平方を，S_e をその自由度 df_e で割った平均平方（誤差分散）との分散比

$$F=\frac{S_R/df_R}{S_e/df_e}$$

が，分母の自由度 df_R，分子の自由度 df_e の F 分布に従うことから有意差検定を行なっている．

JMP によるポアソン回帰の結果を表 11.7 に示す．通常の回帰分析の「分散分析表」と対比して，（差分，完全，縮小）の意味を理解してもらいたい．その後に続く「適合度統計量」は通常の「分散分析表」にはない概念である．しいて言えば，仮定した正規分布からのからの乖離度の統計量であろうか．「適合度統計量」の欄の「デビアンス」の行のカイ 2 乗値＝1.8947 となっているのは，飽和モデルの対数尤度 $\ln L_{飽和} = -17.0566$ と $\ln L_{完全} = -18.0039$ との差の 2 倍で

$$\begin{aligned} デビアンス・カイ2乗 &= 2(\ln L_{飽和} - \ln L_{完全}) \\ &= 2\times[-17.0566-(-18.0039)] \\ &= 1.8947 \end{aligned}$$

と，デビアンスが計算されている．自由度は，$df_{飽和} = 9$ と $df_{完全} = 2$ の差から 7 となっている．このデビアンスが，第 9.5 節で示した R 言語の一般化線形モデルで出力されている Residual deviance に対応する．

表 11.7　通常の回帰とポアソン回帰の対比

通常の回帰分析		
要因		自由度
回帰	S_R	2-1=1
誤差	S_e	9-2=7
全体	S_T	9-1=8

モデル全体の検定

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値
差分	8.2630	16.5260	1	<.0001*
完全	18.0039			
縮小	26.2669			

適合度統計量	カイ2乗	自由度	p値
Pearson	1.8944	7	0.9655
デビアンス	1.8947	7	0.9654

AICc
42.0078

Pearson･カイ 2 乗

Pearson の適合度統計量は，反応変数 y_i と推定値 $\hat{y}_i$ の差の 2 乗をその分散 $Var(y_i) = \hat{y}_i$ で割って加えたもので

$$\begin{aligned} \text{Pearson}\cdot\text{カイ2乗} &= \sum_i \frac{(y_i-\hat{y}_i)^2}{Var(\hat{y}_i)} = \sum_i \frac{(y_i-\hat{y}_i)^2}{\hat{y}_i} \\ &= \frac{(2-2.52)^2}{2.52} + \frac{(3-2.52)^2}{2.52} + \cdots + \frac{(15-12.39)^2}{12.39} \\ &= 1.8944 \end{aligned}$$

として計算されている．これらのデビアンスおよび Peason のカイ 2 乗値を自由度で除したのが過分散の調整パラメータである．この例では，明らかにカイ 2 乗値が自由度 7 より小さいので，過分散が起きていないと判断される．

通常の回帰分析の分散分析表の「要因」の欄の（回帰，誤差，全体）との表記は，ポアソン回帰の「モデル」の表記（差分，完全，縮小）と全く異なるので，それらの対応関係を関連付けることによりによって理解を深めてもらいたい．ただし，「回帰」が「差分」に対応し，「誤差」が「完全」に，「全体」が「縮小」にそれぞれ対応している．このような用語の対応付けを行なったとしても，同義語としては全く認識されないので，それらの対数尤度の定義を理解した上での説明を加えながら注意深く使うことが必要と思われる．

AICc

AICc は，修正済み赤池の情報量量基準で，k をパラメータ数 2 とし

$$\begin{aligned}\mathrm{AICc} &= -2\ln L_{\text{完全}} + 2k + \frac{2k(k+1)}{n-k-1} \\ &= -2\times(-18.0039) + 2\times 2 + \frac{2\times 2\times(2+1)}{9-2-1} \\ &= 42.0078\end{aligned}$$

と計算されている．AICc は，パラメータ数が異なるモデルを比較する際に役に立つ．

表 11.8 のパラメータの関する尤度比検定の「項」の欄の「x」に対する尤度比カイ 2 乗は，16.5260 であり，モデル全体の検定の差分の尤度比カイ 2 乗 16.5260 に一致する．なお，標準誤差は，第 4.2 節で詳しく述べたように対数尤度関数の 2 階の偏微分行列 $\boldsymbol{H}$ の負の逆行列 $(-\boldsymbol{H})^{-1}$ がパラメータの共分散行列なので，その対角要素の平方根である．

表 11.8　パラメータに関する尤度比検定

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	7.4516	0.8842	71.0299	<.0001*
x	4.9353	1.0915	16.5260	<.0001*

11.4. ポアソン回帰における 4 種の残差

ポアソン回帰の場合は，恒等リンクの場合でも対数リンクの場合でも予測値の大きさに比例して分散が大きくなるので，単純な残差プロットによる残差の検討は，不適切である．そのために，ピアソン残差，スチューデント化ピアソン残差，デビアンス残差，あるいは，スチューデント化デビアンス残差などが JMP の一般化線形モデル，SAS の GENMOD プロシジャで提供されている．R では，デビアンス残差が使われている．

デビアンス残差

表 11.7 に示した JMP の「モデル全体の検定」での「適合度統計量」としてのデビアンスは，飽和モデルの $\ln L_{飽和} = -17.0566$ と完全モデル $\ln L_{完全} = -18.0039$ との差の 2 倍で

$$\begin{aligned} \text{デビアンス} &= 2(\ln L_{飽和} - \ln L_{完全}) \\ &= 2\times[-17.0566-(-18.0039)], \quad df = 7 \\ &= 1.8947 \end{aligned}$$

であることを示した．これが有意であれば，ポアソン回帰で取り上げた説明変数では説明しきれない誤差変動が残っていることを意味している．他に追加できる変数がなければ，誤差分布がポアソン分布に対し過分散となっていることを意味する．過分散となっている場合は，パラメータの推定値の標準誤差が小さくなり，結果を過大評価することになる．そのために，過分散パラメータで標準誤差を大きくする修正，あるいは，負の 2 項分布（ガンマ・ポアソン分布）などを仮定した解析が必要となる．

デビアンスは，飽和モデルと完全モデルの対数尤度の差の 2 倍であり，それぞれの対数尤度は，個々のデータの対数尤度 $\ln L_i$ の和であることを表 11.6 で示した．デビアンス残差 (Deviance Residuals) は，飽和モデルと完全モデルの個々の対数尤度の差の 2 倍の平方根として定義されている．ただし，符号が全てプラスなので残差とは言えないので，完全モデルの推定値 $\hat{y}_{完全,i}$ と飽和モデルの推定値 $\hat{y}_{飽和,i} = y_i$ の差の符号を付けてデビアンス残差 $\varepsilon_i^{(D)}$ とする．

$$d_i = 2[\ln(L_{飽和,i}) - \ln(L_{完全,i})] \tag{11.1}$$

$$\varepsilon_i^{(D)} = \mathrm{Sign}(\hat{y}_{飽和,i} - \hat{y}_{完全,i})\sqrt{d_i} \tag{11.2}$$

なお，個々のデビアンス d_i は，簡略化した計算公式（11.3）

$$\begin{aligned} d_i &= 2\left[\ln\left(\frac{\hat{y}_{飽和,i}^{y_i} e^{-\hat{y}_{飽和,i}}}{y_i!}\right) - \ln\left(\frac{\hat{y}_{完全,i}^{y_i} e^{-\hat{y}_{完全,i}}}{y_i!}\right)\right] \\ &= 2\left[y_i\ln(y_i) - y_i - \ln(y_i!) - y_i\ln(\hat{y}_{完全,i}) + \hat{y}_{完全,i} + \ln(y_i!)\right] \\ &= 2\left[y_i\ln\left(\frac{y_i}{\hat{y}_{完全,i}}\right) - (y_i - \hat{y}_{完全,i})\right] \end{aligned} \tag{11.3}$$

が，一般的に計算公式として用いられている．ただし，この計算式からでは何を意味しているのか，推測しがたい．元の対数尤度での定義式による理解を勧める．

実際に $i=1$ の場合の計算を次に示し，全ての i についての結果を表 11.9 に示す．

$$d_1 = 2[(-1.3069)-(-1.3639)]$$
$$= 2\times 0.0570 = 0.1140$$
$$\mathrm{Sign}(\hat{y}_{飽和,1}-\hat{y}_{完全,1}) = \mathrm{Sign}(2.0-2.5163) = マイナス$$
$$\varepsilon_1^{(D)} = \mathrm{Sign}(\hat{y}_{飽和,1}-\hat{y}_{完全,1})\sqrt{d_i}$$
$$= -\sqrt{0.1140} = -0.3377$$

表 11.9　デビアンス残差

				完全(最大)モデル			飽和モデル			デビアンス残差		
					$\hat{\beta}_0=$	**7.4516**						
					$\hat{\beta}_1=$	**4.9353**					平方和 =	**1.8947**
					$\ln L_{完全}=$	**-18.0039**		$\ln L_{飽和}=$	**-17.0566**	尤度の差	平方根	残差
i	X		y	$y_{完全}$^	$P_{完全}$	$\ln L_{完全,i}$	$y_{飽和}$^	$P_{飽和}$	$\ln L_{飽和,i}$	d	$\sqrt{\ }(d)$	$\varepsilon^{(D)}$
1	1	-1	2	2.5163	0.2557	**-1.3639**	2	0.2707	**-1.3069**	**0.1140**	**0.3377**	**-0.3377**
2	1	-1	3	2.5163	0.2144	-1.5397	3	0.2240	-1.4959	0.0875	0.2958	0.2958
3	1	0	6	7.4516	0.1380	-1.9803	6	0.1606	-1.8287	0.3032	0.5506	-0.5506
4	1	0	7	7.4516	0.1469	-1.9178	7	0.1490	-1.9038	0.0279	0.1672	-0.1672
5	1	0	8	7.4516	0.1369	-1.9888	8	0.1396	-1.9691	0.0394	0.1985	0.1985
6	1	0	9	7.4516	0.1133	-2.1776	9	0.1318	-2.0268	0.3015	0.5491	0.5491
7	1	1	10	12.3869	0.0978	-2.3249	10	0.1251	-2.0786	0.4927	0.7019	-0.7019
8	1	1	12	12.3869	0.1137	-2.1744	12	0.1144	-2.1683	0.0122	0.1105	-0.1105
9	1	1	15	12.3869	0.0791	-2.5366	15	0.1024	-2.2785	0.5161	0.7184	0.7184

デビアンス残差 $\varepsilon_i^{(D)}$ の平方和は，

$$\text{デビアンス}=\sum_{i=1}^{9}\left[\varepsilon_i^{(D)}\right]^2$$
$$= (-0.3377)^2 + 0.2958^2 + \cdots + 0.7184^2 = 1.8947$$

となり，デビアンス・カイ 2 乗値に

$$\text{デビアンス・カイ2乗} = 2(\ln L_{飽和} - \ln L_{完全})$$
$$= 2\times[-17.0566-(-18.0039)] = 1.8947$$

一致することが確認される．もちろん，計算公式によっても，次のように

$$d_1 = 2\left[y_1 \ln\left(\frac{y_1}{\hat{y}_{完全,1}}\right) - (y_1 - \hat{y}_{完全,1})\right]$$
$$= 2\times\left[2\times\ln\left(\frac{2}{2.5163}\right) - (2-2.5163)\right] = 0.1140$$

求め，一致することが確認できる．

スチューデント化デビアンス残差

スチューデント化デビアンス残差は，反復重み付き回帰での重みを加味したハット行列 $\boldsymbol{H}'$ の対角要素テコ比 h_{ii}' を用いる．恒等リンクの場合の重みは，推定値 $\hat{y}_i$ の逆数であるので，対角要素にそれぞれの重を持つ行列を $\hat{\boldsymbol{W}}$ とする．推定値 $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ のパラメータ推定値 $\hat{\boldsymbol{\beta}}$ を重み付き回帰の計算式 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{Y}$ を代入すると，

$$\begin{aligned}\hat{\boldsymbol{Y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{Y}\end{aligned} \tag{11.4}$$

となる．ハット行列 $\boldsymbol{H}'$ は，回帰の推定値を求める際に $\hat{\boldsymbol{Y}}$ の式から最後の $\boldsymbol{Y}$ を除いた行列

$$\boldsymbol{H}' = \boldsymbol{X}(\boldsymbol{X}^T\hat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{W}} \tag{11.5}$$

で定義されている．重み行列 $\hat{\boldsymbol{W}}$ の対角要素からなるベクトルを $\hat{\boldsymbol{w}}$ とし，デザイン行列 $\boldsymbol{X}$ の i 行目のベクトルを $\boldsymbol{x}_i$ とすれば，テコ比 h_{ii}' は，

$$h_{ii}' = \boldsymbol{x}_i[(\boldsymbol{X}*\hat{\boldsymbol{w}})^T\boldsymbol{X}]^{-1}\boldsymbol{x}_i^T\hat{w}_i \tag{11.6}$$

として求めることができる．スチューデント化デビアンス残差 $\varepsilon_i'^{(D)}$ は，

$$\hat{\varepsilon}_i'^{(D)} = \frac{\hat{\varepsilon}_i^{(D)}}{\sqrt{1-h_{ii}'}} \tag{11.7}$$

として求められる．表 11.10 にスチューデント化デビアンス残差の計算結果を示す．デビアンス残差は，表 11.9 に示した結果を用いる．重み $\hat{w}_i$ を推定値 $\hat{y}_i$ の逆数とし，テコ比を求め，スチューデント化デビアンス残差 $\hat{\varepsilon}_i'^{(D)}$ が計算されている．

表 11.10　スチューデント化デビアンス残差

				完全モデル		飽和モデル						
				$\hat{\beta}_0$ =	**7.4516**			デビアンス残差		スチューデント化		
				$\hat{\beta}_1$ =	**4.9353**			平方和 =	**1.8947**	デビアンス残差		
				$\ln L_{完全}$ =	**-18.0039**	$\ln L_{飽和}$ =	**-17.0566**	尤度の差	残差	重み w	テコ比	残差
i	$\boldsymbol{X}$		y	$\hat{y}_{完全}$	$\ln L_{完全,i}$	$\hat{y}_{飽和}$	$\ln L_{飽和,i}$	d	$\varepsilon^{(D)}$	$1/\hat{y}_{完全}$	h_{ii}'	$\varepsilon'^{(D)}$
1	1	-1	2	**2.5163**	-1.3639	2.0	-1.3069	0.1140	**-0.3377**	**0.3974**	**0.4510**	**-0.4558**
2	1	-1	3	2.5163	-1.5397	3.0	-1.4959	0.0875	0.2958	0.3974	0.4510	0.3993
3	1	0	6	7.4516	-1.9803	6.0	-1.8287	0.3032	-0.5506	0.1342	0.1049	-0.5820
4	1	0	7	7.4516	-1.9178	7.0	-1.9038	0.0279	-0.1672	0.1342	0.1049	-0.1767
5	1	0	8	7.4516	-1.9888	8.0	-1.9691	0.0394	0.1985	0.1342	0.1049	0.2098
6	1	0	9	7.4516	-2.1776	9.0	-2.0268	0.3015	0.5491	0.1342	0.1049	0.5804
7	1	1	10	12.3869	-2.3249	10.0	-2.0786	0.4927	-0.7019	0.0807	0.2261	-0.7979
8	1	1	12	12.3869	-2.1744	12.0	-2.1683	0.0122	-0.1105	0.0807	0.2261	-0.1256
9	1	1	15	12.3869	-2.5366	15.0	-2.2785	0.5161	0.7184	0.0807	0.2261	0.8167
										0.7817	0.4166	
										0.4166	**1.1863**	
										$[(\boldsymbol{X}*\boldsymbol{w})^T\boldsymbol{X}]^{-1}$		

実際の計算を$i=1$の場合について示す．スチューデント化デビアンス残差は，

$$\hat{w}_1 = 1/\hat{y}_{完全,1}$$
$$= 1/2.5163 = 0.3974$$

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = [(\boldsymbol{X} * \hat{\boldsymbol{w}})^T \boldsymbol{X}]^{-1}$$
$$= \text{Minverse(Mmult(Transpose(}\boldsymbol{X}\text{の範囲} * \hat{\boldsymbol{w}}\text{の範囲)},}\boldsymbol{X}\text{の範囲)}$$

$h_{11}' =$

1	-1	**0.7817**	0.4166	1	×	0.3974	=	0.4510
		0.4166	**1.1863**	-1				
$\boldsymbol{x}_1$		$[(\boldsymbol{X} * \boldsymbol{w})^T \boldsymbol{X}]^{-1}$		$\boldsymbol{x}_1^T$		w_1		h_{ii}'

$$\hat{\varepsilon}_i'^{(D)} = \frac{\hat{\varepsilon}_i^{(D)}}{\sqrt{1-h_{ii}'}}$$
$$= \frac{-0.3377}{\sqrt{1-0.4510}} = -0.4558$$

で求められる．

デビアンス残差およびスチューデント化デビアンス残差について，JMP のポアソン回帰で作成した残差プロットを図 11.2 に示す．デビアンス残差に対してスチューデント化デビアンス残差の方が，残差の絶対値が大きめになっていることが確認できる．ただし，実際の解析で複数の残差を用いることは非現実的であり，第 11.5 節を参考にして，選択してほしい．

自ら計算する場合には，手軽に計算できるデビアンス残差であるが，統計的には，スチューデント化デビアンス残差が望ましい思われる．JMP のディフォルトの残差プロットは，スチューデント化デビアンス残差が使用されているが，SAS/GENMOD プロシジャでは，ユーザの選択に任されている．

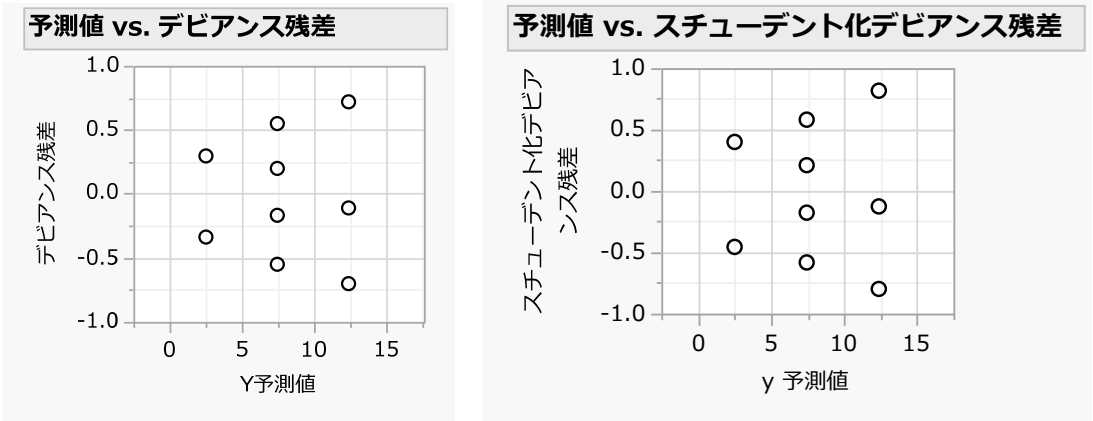


図 11.2　デビアンス残差とスチューデント化デビアンス残差の比較

スチューデント化 Pearson 残差

第 11.2 節で通常の回帰分析の場合に残差 $\hat{\varepsilon}_i = y_i - \hat{y}_i$ に対してテコ比 h_{ii} を用いてスチューデント化残差 $\hat{\varepsilon}_i'$ を求めた．ポアソン回帰の場合，単純な残差 $\hat{\varepsilon}_i$ は，推定値 $\hat{y}_i$ に比例して大きくなるので残差の検討に使えない，そこで，$\hat{y}_i$ の標準誤差で基準化した Pearson 残差 $\hat{\varepsilon}_i^{(P)}$ が使われている．Pearson 残差 $\hat{\varepsilon}_i^{(P)}$ は，

$$\begin{aligned}\hat{\varepsilon}_i^{(P)} &= \frac{y_i - \hat{y}_i}{\sqrt{Var(\hat{y}_i)}} \\ &= \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}\end{aligned} \tag{11.8}$$

として計算される．Pearson 残差 $\hat{\varepsilon}_i^{(P)}$ は過小評価になりがちなので，テコ比 h_{ii} により調整したスチューデント化 Pearson 残差も使われている．重みは，$\hat{w}_i = 1/\hat{y}_i$ であり，テコ比は，スチューデント化デビエンス残差で計算した表 11.10 と同じ式であり，スチューデント化 Pearson 残差 $\hat{\varepsilon}_i'^{(P)}$ は，スチューデント化デビアンス残差 $\varepsilon_i'^{(D)}$ の場合と同様に

$$\hat{\varepsilon}_1'^{(P)} = \frac{\hat{\varepsilon}_1^{(P)}}{\sqrt{1-h_{11}}} \tag{11.9}$$

となる．表 11.11 に Pearson 残差およびスチューデント化 Pearson 残差についての計算結果を示す．

表 11.11　Pearson 残差およびスチューデント化 Pearson 残差

					Pearson	スチューデント化Pearson残差		
				推定値	残差	重み	テコ比	残差
i	$\boldsymbol{X}$		y	y^	$\varepsilon^{(P)}$	$w=1/y$	h_{ii}'	$\varepsilon'^{(P)}$
1	1	-1	2	**2.5163**	**-0.3255**	**0.3974**	0.4510	**-0.4393**
2	1	-1	3	2.5163	0.3049	0.3974	0.4510	0.4115
3	1	0	6	7.4516	-0.5318	0.1342	0.1049	-0.5621
4	1	0	7	7.4516	-0.1654	0.1342	0.1049	-0.1749
5	1	0	8	7.4516	0.2009	0.1342	0.1049	0.2123
6	1	0	9	7.4516	0.5672	0.1342	0.1049	0.5995
7	1	1	10	12.3869	-0.6782	0.0807	0.2261	-0.7709
8	1	1	12	12.3869	-0.1099	0.0807	0.2261	-0.1250
9	1	1	15	12.3869	0.7425	0.0807	0.2261	0.8440

推定値は，表 11.10 の完全モデルの推定値，
テコ比は，スチューデント化デビアンス残差でのテコ比に等しい．

実際の計算を $i=1$ の場合について示す．Pearson 残差は，

$$\begin{aligned}\varepsilon_1^{(P)} &= \frac{y_1 - \hat{y}_1}{\sqrt{\hat{y}_1}} \\ &= \frac{2 - 2.5163}{\sqrt{2.5163}} \\ &= -0.3255\end{aligned}$$

スチューデント化デ Pearson 残差は,

$$\hat{w}_1 = 1/\hat{y}_1$$
$$= 1/2.5163 = 0.3974$$

$h_{11}' =$

1	-1	**0.7817**	0.4166	1	×	0.3974	=	0.4510
		0.4166	**1.1863**	-1				
$\boldsymbol{x}_1$		$[(\boldsymbol{X}*\boldsymbol{w})^T \boldsymbol{X}]^{-1}$		$\boldsymbol{x}_1^T$		w_1		h_{ii}'

$$\hat{\varepsilon}_1^{(P)'} = \frac{\hat{\varepsilon}_i^{(P)}}{\sqrt{1-h_{11}'}}$$
$$= \frac{-0.3255}{\sqrt{1-0.4510}} = -0.4393$$

で求められる.

SAS/GENMOD による各種の残差

SAS/GENMOD を用いて，スチューデント化 Pearson 残差を含め，これまでに示したポアソン回帰での 4 種の残差,（デビアンス残差，スチューデント化デビアンス残差，Pearson 残差，スチューデント化 Pearson 残差）を計算し，これまでの結果を検証する．SAS/GENMOD のオプションで「residual」が各種の残差の一括出力指示で,「plots=stdresdev(xbeta)」が，スチューデント化デビアンス残差を推定値 $\hat{y}_i$ に対する残差プロットの作成を指示している.

＜＜SAS/GENMOD によるポアソン回帰＞＞

```
Title "デビアンス残差_a01.sas   2020/01/20  Y.Takahashi" ;
data d01 ;
  input  x y @@ ;
datalines ;
-1  2    -1  3    0  6    0  7    0  8    0  9    1  10    1  12    1  15
;
proc genmod  data=d01  plots=reschi(xbeta)  plots=stdreschi(xbeta)
                       plots=resdev(xbeta)  plots=stdresdev(xbeta) ;
    model   y = X / dist=poisson link= identity  residual ;
    run;
```

表 11.12 に示した SAS の残差の出力で,「標準化」となっているのが「スチューデント化」の意味である「未加工残差」は，$\hat{\varepsilon}_i = y_i - \hat{y}_i$ であるが，右端の「尤度残差」は，スチューデント化デビアンス残差とスチューデント化 Pearson 残差の両方を使って，それらの中間的な残差である[SAS Institute (2016), The GENMOD Procedure, The GENMOD Procedure :3164-3165]. SAS の残差の出力結果と，Excel で計算した 4 種の残差を照合し，一致することが確認される.

このように沢山の残差があり，どれを使うか選択に窮するのである．最も簡便なのが Pearson 残差であるが，スチューデント化 Pearson 残差に比べて小さめに出るので，残差の大きさを少々過小評価となる．同様にデビアンス残差もスチューデント化デビアンス残差に比べて小さめになっている．

スチューデント化 Pearson 残差か，スチューデント化デビアンス残差かの選択については，この事例ではどちらとも言い切れない．元々の計算過程で，対数尤度を使っているので，飽和モデルと完全モデルの対数尤度の差を用いるデビアンス残差，あるいは，スチューデント化デビアンス残差を使うのが自然の流れのように思われる．

表 11.12　SAS/GENMOD の出力：ポアソン回帰の各種の残差統計量

観測値の統計量						
オブザベーション	未加工残差	Pearson 残差	デビアンス残差	標準化デビアンス残差	標準化 Pearson 残差	尤度残差
1	−0.5163	−0.3255	−0.3377	−0.4558	−0.4393	−0.4484
2	0.4837	0.3049	0.2958	0.3993	0.4115	0.4048
3	−1.4516	−0.5318	−0.5506	−0.5820	−0.5621	−0.5799
:						
9	2.6131	0.7425	0.7184	0.8167	0.8440	0.8229

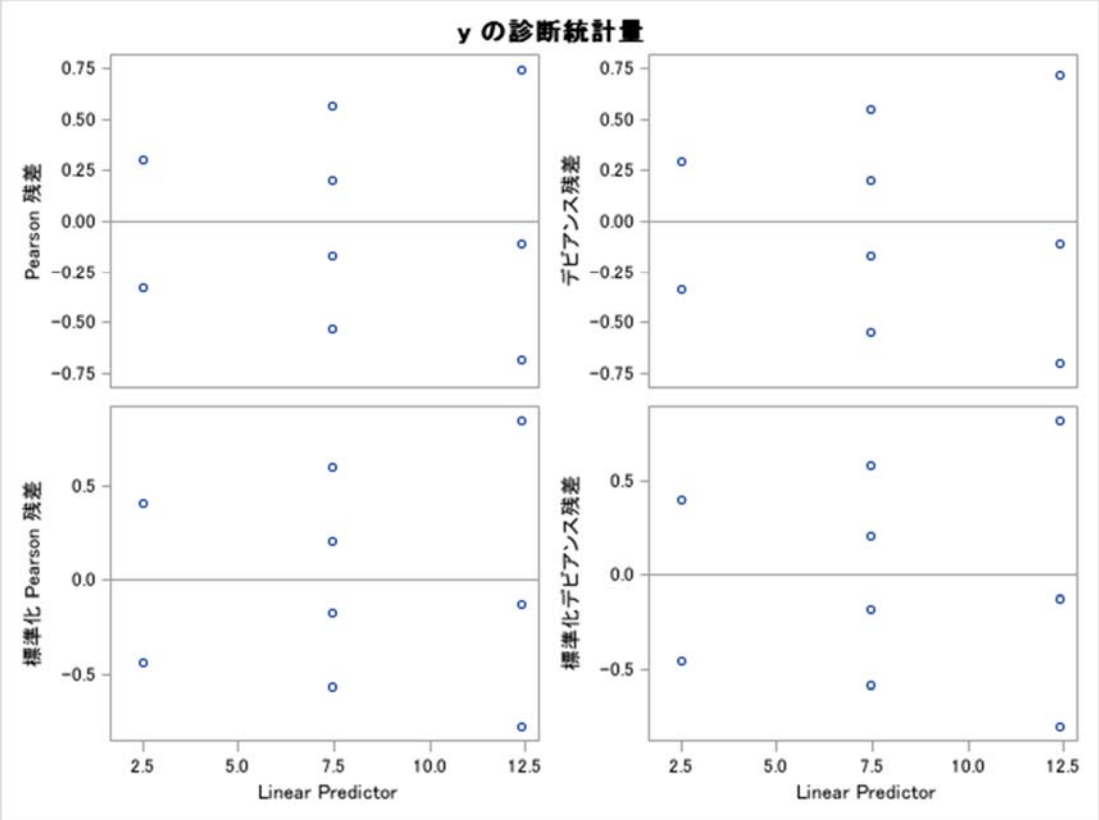


図 11.3　SAS/GENMOD による 4 種の残差プロット

11.5.　カブトガニの事例における 4 種の残差

アグレスティ（2003）のカブトガニのデータについて第 1.13 節で概要を示し，第 7.2 節で探索的な解析の事例として用いた．ここでは，ポアソン回帰で用いられている 4 種の残差を比較検討するために用いる．このデータは，雌のカブトガニに連結する雄のサテライト数（Satellite 数）について 173 匹について，名義尺度（甲羅の色，後体部の棘の状態）の 2 変数，連続尺度（甲羅の幅，体重）の 2 変数，反応変数としてサテライト数が含まれている．

JMP による 4 種の残差の計算

第 1.13 節では，雌のカブトガニに連結する雄のサテライト数を反応変数とし，雌の甲羅の幅を説明変数とした対数リンクでのポアソン回帰を行い，散布図に回帰曲線および 95%信頼区間と予測区間（個別データの 95%信頼区間）を示し，多くのデータが 95%予測区間の外にあることを示した．残差プロットには，JMP の過分散の調整に使われている Pearson 残差について，「予測値 vs. Pearson 残差」を例示した．

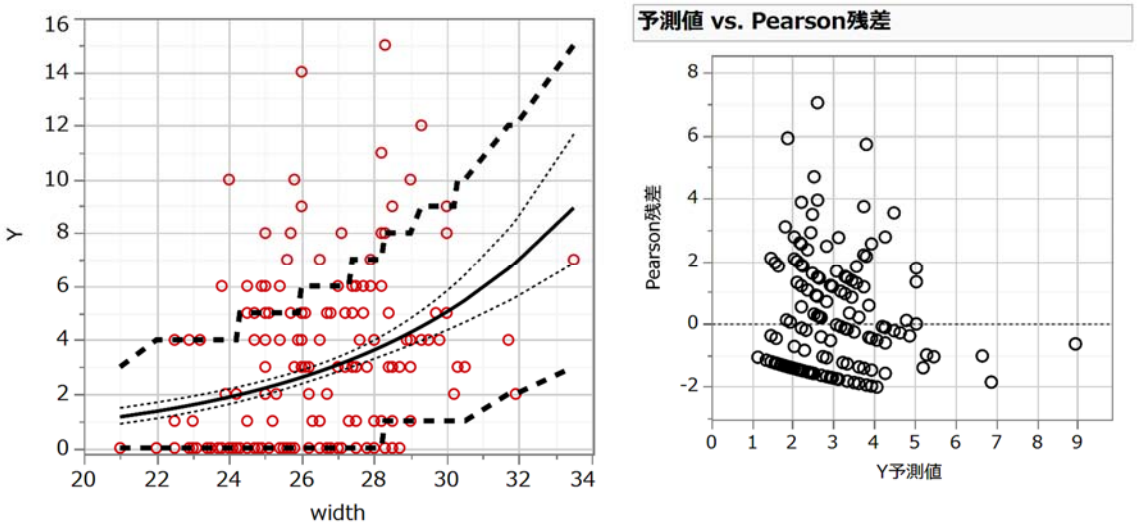
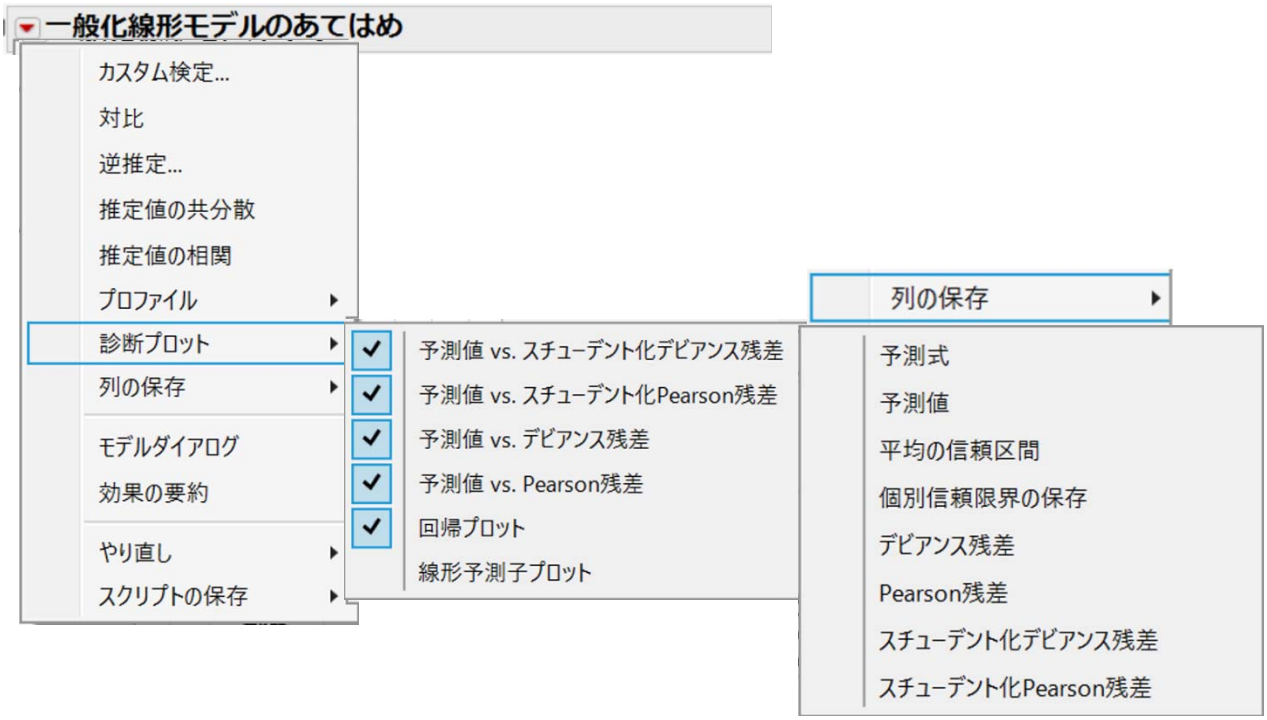


図 11.4　ポアソン回帰に対する 95%信頼区間および pearson 残差（図 1.15 再掲）

Pearson 残差プロットは，プラス側に大きく歪んでいる．これは，推定値に対してポアソン分布がプラス側に裾を引くことによる必然的な現象であること，さらに，ゼロを含む場合にプラス側に大きく裾を引くことも影響している．このような必然的に起きるバイアスを少しでも解消するためにデビエンス残差の使用が望ましい．

JMP で過分散なしの対数リンクでのポアソン回帰分析を行い，4 種の診断プロットを選択し，さらに，4 種の残差を JMP ファイルとして出力して，相互の比較を行う．



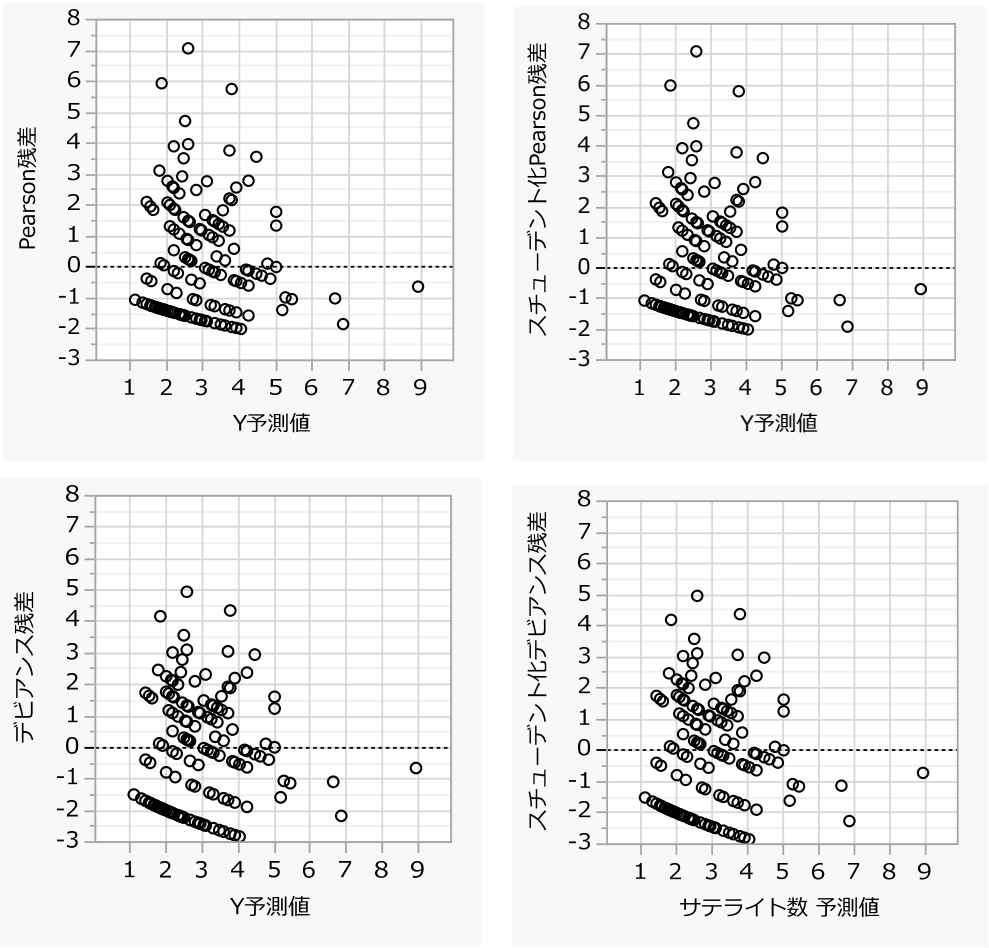


図 11.5　JMP/ポアソン回帰による 4 種の残差プロット

4 種の残差の比較

（Pearson 残差 vs. デビアンス残差）では，正の残差については，デビアンス残差を全体的に小さい方向へシフトし，負の残差の場合は，負の大きい方向にシフトしている．（スチューデント化 Pearson 残差 vs. スチューデント化デビアンス残差）も同様である．

（Pearson 残差 vs. スチューデント化 Pearson 残差）および（デビアンス残差 vs. スチューデント化デビアンス残差）については，式（11.7）および式（11.9）を用いたスチューデント化に用いているテコ比 h_{ii} が 1 以下であることから，必然的に残差ゼロを基準に残差の絶対値を大きい方に広げることになる．言い換えれば，プラスの残差はプラス方向に引き伸ばし，マイナスの残差は，マイナス方向に引き伸ばす．

どの程度の差が実際に起きるのかを実感するために，表 11.13 に JMP ファイルに出力された 4 種の残差を Excel に取り込みデビアンス残差の大きい順に並べて抜粋した結果を示す．スチューデント化した場合の残差の差について計算した結果は，大きいもので小数点以下 2 桁目での差であり，目立った違いではない．

Pearson 残差とデビアンス残差を比較すると，最初の行の No.15 では，Pearson 残差が 7.0448 であるのに対し，デビアンス残差は，4.9221 と明らかにデビアンス残差に縮小効果が表れている．また最後の No. 94 の場合は，-2.0171➔-2.8526 とマイナス方向への明らかな引き延ばしが確認される．

表 11.13　ポアソン回帰の各種の残差統計量

No	甲羅の色	後体部の棘	体重	甲羅の幅	サテライト数	予測式サテライト数	Pearson 残差	スチューデント化 Pearson 残差	Pearson 残差の差	デビアンス残差	スチューデント化デビアンス残差	デビアンス残差の差
15	2	1	2.30	26.0	14	2.6128	7.0448	7.0673	**0.0225**	4.9221	4.9379	**0.0157**
56	2	3	3.00	28.3	15	3.8103	5.7324	5.7608	**0.0284**	4.3279	4.3494	**0.0215**
13	2	3	3.05	28.2	11	3.7483	3.7456	3.7632	**0.0176**	3.0301	3.0443	**0.0142**
165	2	3	2.75	26.5	7	2.8361	2.4725	2.4799	**0.0074**	2.0787	2.0849	**0.0062**
28	2	1	2.70	26.8	5	2.9792	1.1708	1.1743	**0.0035**	1.0659	1.0692	**0.0032**
91	2	1	3.85	29.7	5	4.7941	0.0940	0.0951	**0.0011**	0.0934	0.0945	**0.0011**
124	2	3	1.65	24.2	2	1.9448	0.0396	0.0398	**0.0002**	0.0394	0.0396	**0.0002**
44	2	1	3.30	30.0	5	5.0359	-0.0160	-0.0162	**-0.0002**	-0.0160	-0.0162	**-0.0002**
63	3	1	2.45	27.0	3	3.0786	-0.0448	-0.0449	**-0.0001**	-0.0450	-0.0451	**-0.0001**
50	2	1	3.60	30.3	3	5.2900	-0.9956	-1.0122	**-0.0166**	-1.0848	-1.1028	**-0.0181**
81	3	2	2.25	24.5	0	2.0429	-1.4293	-1.4361	**-0.0068**	-2.0213	-2.0309	**-0.0096**
94	2	1	3.20	28.7	0	4.0688	-2.0171	-2.0297	**-0.0126**	-2.8526	-2.8705	**-0.0178**

全体で 173 サンプルをデビアンス残差の大きい順に並べ，最大値と最小値を残し，デビアンス残差がおおむね等間隔になるように 12 サンプルを抽出した．

図 11.6 に全 173 匹についてスチューデント化した場合の残差の変化について JMP の「対応のあるペア」によって比較した結果を示す．スチューデント化した場合，残差がゼロを起点にプラス方向とマイナス方向に引き伸ばされていることが確認される．

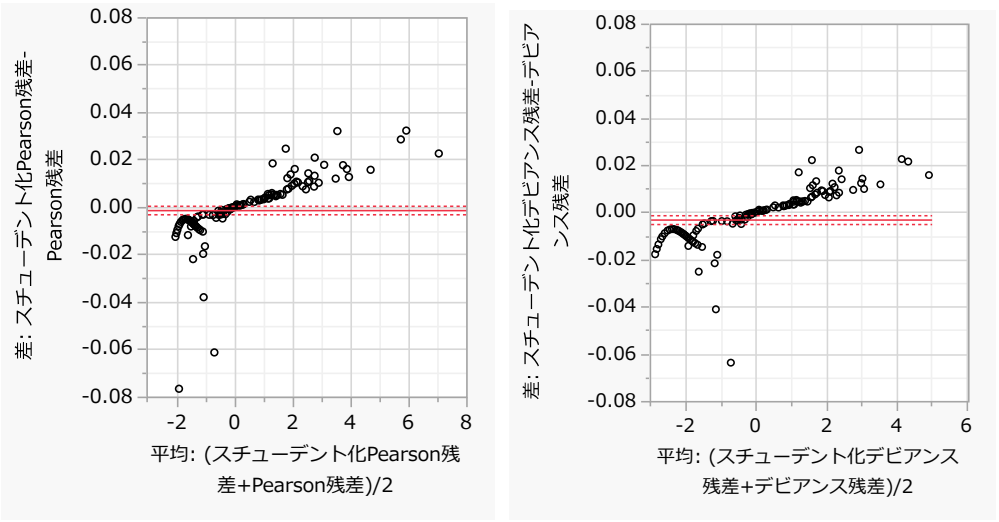


図 11.6　スチューデント化による残差の大きさの比効

図 11.7 に, スチューデント化 Pearson 残差とスチューデント化デビアンス残差の比較を JMP の「二変量の関係」によって作図した結果を示す．デビアンスにした場合に，残差がプラスの場合は圧縮され，マイナスの場合は，引き伸ばされていることが確認される．対数リンクの場合には，観測値が大いい方に裾を引くので，スチューデント化 Pearson 残差は，大きい方に引っ張られる．ビアンス残差あるいはスチューデント化デビアンス残差による補正が行われていることが確認さるれる．したがって，対数リンクの場合には，ビアンス残差あるいはスチューデント化デビアンス残差の使用が望ましい．

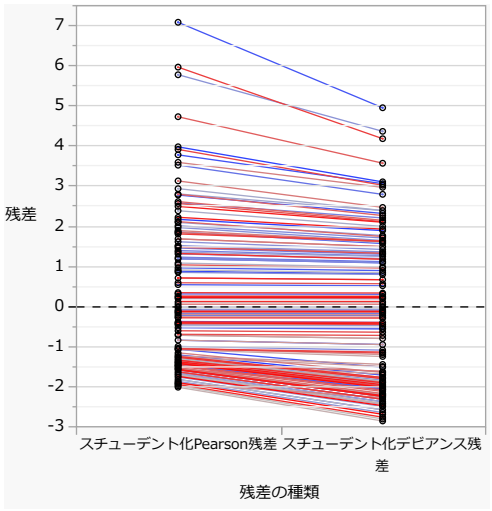


図 11.7　スチューデント化 Pearson 残差とスチューデント化デビアンス残差の比較

第 11 章　文 献 索　引

アグレスティ著, 渡邉・菅波・吉田・角野・寒水・松永 訳(2003) - カテゴリカルデー解析入門 379
SAS Institute(2016) - SAS/STAT® 14.2 User's Guide, The GENMOD Procedure 377
ドブソン著, 田中・森川・山中・富田 訳(2008) - 一般化線形モデル入門, 原著 第2版 361
野沢昌弘(1992) - テコ比とハット行列 366

第 11 章　索　引

第 11 章　Excel, JMP, SAS ファイル一覧

4 KB	第11章02_人工データ_線形	JMP Data Table
37 KB	第11章02_人工データ_線形	Microsoft Excel ワークシート
57 KB	第11章03_ポアソン回帰_デビアンス残差	Microsoft Excel ワークシート
49 KB	第11章04_人工データ_4種の残差	Microsoft Excel ワークシート
1 KB	第11章04_人工データ_4種の残差_SAS	テキスト ドキュメント
27 KB	第11章05_カブトガニ_4種の残差	JMP Data Table
28 KB	第11章05_カブトガニ_4種残差	Microsoft Excel ワークシート
72 KB	第11章05_カブトガニ_転置_グラフ作成	JMP Data Table

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 11 章＞

第 11 章　デビアンス・逸脱度・テコ比・4 種の残差

BioStat 研究所（株）

〒105-0014 東京都 港区 芝 1-12－3 の 1005

2020 年 7 月 11 日　　高橋 行雄