

第 9 回 続高橋セミナー
最尤法によるポアソン回帰分析入門
2020 年 6 月 6 日

第 5 章 反復重み付き最尤法によるポアソン回帰

ポアソン回帰を含む一般化線形モデルに対する最尤法には、2 種類の解法がある。伝統的には、重み付き回帰の反復によって最尤解を求める方法である。他方は、第 2 章で導入したニュートン・ラフソン法による最尤解を求める方法である。手計算の時代から反復重み付き回帰によって、2 値反応データから 50 パーセント致死量を求める方法とし、プロビット法が知られていた。コンピュータの性能向上により、2 階の偏微分行列を使うニュートン・ラフソン法による計算も可能となってきた。どちらでもパラメータの推定値は一致するが、パラメータの共分散行列には、若干の相違がある。反復重み付き最尤法は、パラメータ数が増えても対応が容易であり、Excel での計算には適している。

第 5 章 目 次

| | | |
|------|---|-----|
| 5 | 反復重み付き最尤法によるポアソン回帰 | 175 |
| 5.1. | 反復重み付き回帰 | 175 |
| 5.2. | 重み付き回帰の基礎 | 177 |
| | 正規方程式, 重みを含む行列計算 | |
| 5.3. | 恒等リンクの場合のポアソン回帰 | 180 |
| | 初期パラメータの推定, 重み付き回帰, 反復重み付き回帰(2), 反復重み付き回帰(3), 回帰パラメータについてのワルド検定, 尤度比検定 | |
| 5.4. | 対数リンクでのポアソン回帰 | 186 |
| | 対数リンク, 重み付き回帰(1), 重み付き回帰(2)および(3), 95%信頼区間, 2 次式のあてはめ | |

| | |
|---|-----|
| 5.5. 対数リンクでオフセットがある場合のポアソン回帰----- | 195 |
| オフセットを含めたポアソン回帰, 補正值, 反復計算, オフセットを用いた推定 | |
| 5.6. 2 項分布を仮定した(プロビット・補 2 重対数・ロジット)解析----- | 200 |
| プロビット, 補 2 重対数, ロジット, ポアソン・プロビット・補 2 重対数・ロジット, 上限があるシグモイド曲線のあてはめ | |
| 文献索引, 索引, 解析用ファイル一覧----- | 207 |

第 9 回 続高橋セミナー 最尤法によるポアソン回帰分析入門

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

目 次

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

5. 反復重み付き最尤法によるポアソン回帰

ポアソン回帰を含む一般化線形モデルに対する最尤法には、2種類の解法がある。伝統的には、重み付き回帰の反復によって最尤解を求める方法である。他方は、第2章で導入したニュートン・ラフソン法による最尤解を求める方法である。手計算の時代から反復重み付き回帰によって、2値反応データから50パーセント致死量を求める方法とし、プロビット法が知られていた。コンピュータの性能向上により、2階の偏微分行列を使うニュートン・ラフソン法による計算も可能となってきた。どちらでもパラメータの推定値は一致するが、パラメータの共分散行列には、若干の相違がある。反復重み付き最尤法は、パラメータ数が増えても対応が容易であり、Excelでの計算には適している。

5.1. 反復重み付きポアソン回帰

第1.4節の人工データでポアソン回帰のパラメータ推定に対し、説明なしに反復重み付き回帰による最尤法を用いた。ポアソン回帰は、説明変数 X_i が大きくなるにつれ推定値 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ の分散が推定値 \hat{Y}_i 比例して大きくなり、観測値 Y_i と回帰の推定値 \hat{Y}_i との偏差 $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ に等分散性を仮定することができなくなる場合に用いられる。

回帰の推定値に対して観測値が、ポアソン分布に従う場合には、最小2乗法による回帰分析ではなく、ポアソン分布の確率関数を用いた最尤法を適用する必要がある。第2章で導入したポアソン分布の確率関数を用いる最尤法は、計算機の性能が低い時代には、容易ではなかった。そのために、偏差 $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ の平方を求める際に \hat{Y}_i の分散で割って基準化した平方和を用いる重み付き回帰が考案された。

ただし、厄介な問題が内在する。これは、基準化のために、何らかの方法で得られた回帰の推定値 $\hat{Y}_i^{(0)}$ の分散を用い、重み付き回帰を実施して新たな推定値 $\hat{Y}_i^{(1)}$ と分散を得ると $\hat{Y}_i^{(0)}$ との間にわずかなブレが生じて一致しなくなる。そのために、推定値 $\hat{Y}_i^{(1)}$ の分散を用いて重み付き回帰を行い、推定値 $\hat{Y}_i^{(2)}$ 得る。といった反復を繰り返して、回帰パラメータのブレがなくなるまで、重み付き回帰を反復する必要が生ずる。

第 1.9 節で示した植物の種子数のデータの解析は、説明変数 X の増加に伴い反応 Y が指数関数的に増加する場合に対し、ポアソン分布の確率関数を用いたポアソン回帰を取り上げ、第 2.5 節で理論的側面を示した。このような現象に対し、反応 Y に対し対数変換によって直線化して通常の回帰分析を行えばとの誘惑にかられる。しかし、厄介な問題が発生する。元のカウント・データにゼロがあると対数変換できない。元のデータは、ある X に対してポアソン分布を仮定していたとしても、対数変換後の分布は、「対数ポアソン分布？」などの問題に直面する。

これらの問題を解決するための方法とし、元データを対数変換するのではなく、仮の対数推定値 $\ln(\hat{Y}_i)$ に対し、後述する式 (5.20) から、

$$Z_i = \ln(\hat{Y}_i) + \frac{\hat{Y}_i - \exp[\ln(\hat{Y}_i)]}{\exp[\ln(\hat{Y}_i)]}$$

のような巧妙な変換を施し、 Z_i について反復重み付き回帰を適用する。 $Y_i = 0$, $\ln(\hat{Y}_i) = -0.1$ の場合であれば、

$$\begin{aligned} Z_i &= \ln(\hat{Y}_i) + \frac{\hat{Y}_i - \exp[\ln(\hat{Y}_i)]}{\exp[\ln(\hat{Y}_i)]} \\ &= -0.1 + \frac{0 - 0.9048}{0.9048} = -1.1000 \end{aligned}$$

として重み付き回帰の対象とする「データ」とするのである。元のスケールでは

$$\exp(Z_i) = \exp(-1.10) = 0.3329$$

となり、元のゼロの値にげたをはかせ、対数変換したことになる。これは、ポアソン回帰だけではなく一般化線形モデルとしてのロジスティック回帰、プロビット回帰でも同様な変換が行なわれている。詳しくは、高橋 (2017)、「一般化線形モデルを Excel で極め活用するープロビット法・ロジット法・補 2 重対数法ー」を参照のこと

第 1.5 節、第 1.10 節、第 1.11 節では、ある地域で発生した死亡者数などのカウント・データの解析に際し、部分母集団の人数が得られる場合のデータを示した。年齢などで層別した場合にも、部分母集団の人数が得られるような場合に、その部分集団の人数に対し対数を取ってオフセットとしたポアソン回帰を示した。オフセットがある場合に対するポアソン回帰は、部分母集団の 1 人あたりのイベント発生数（発生率）をベースにしており、本章では、対数リンクでオフセットを考慮した反復重み付き回帰による最尤法による計算方法を示す。

5.2. 重み付き回帰の基礎

重み付き回帰では、偏差平方和の計算に際して分散 $Var(\hat{Y}_i)$ の逆数を掛けることで分散に関して基準化している。ポアソン分布は、平均と分散が等しいことから、得られた推定値 \hat{Y}_i と分散が $Var(\hat{Y}_i) = \hat{Y}_i$ と同じになる。さて、推定値 \hat{Y}_i を得るためには、何らかの回帰分析が必要となるのだが、最初は「重み」となる \hat{Y}_i が求まっていないので重み付き回帰は実施できない。そのために通常の回帰分析を行い、仮の推定値 $\hat{Y}_i^{(0)} = \hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} X_i$ を求め、 $\hat{Y}_i^{(0)}$ を用いて重み付き回帰を行う。

正規方程式

誤差がポアソン分布に従う回帰式を、

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad \varepsilon_i \sim \text{Poisson}(Y_i) \quad (5.1)$$

または、

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \text{Poisson}(\mathbf{Y}) \quad (5.2)$$

とする。推定されたパラメータを用いた $\hat{Y}_i^{(0)} = \hat{\beta}_0^{(0)} + \hat{\beta}_1^{(0)} X_i$ の推定値 $\hat{Y}_i^{(0)}$ に対して重みを $\hat{w}_i = 1/\hat{Y}_i^{(0)}$ としたときに、 $\hat{Y}_i^{(0)}$ からの偏差 ε_i の重み付き平方和は、

$$Q = \sum_{i=1}^n \hat{w}_i \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (5.3)$$

となる。式 (5.3) を β_0 と β_1 で偏微分すると次式が得られる。

$$\begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum_{i=1}^n \hat{w}_i (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum_{i=1}^n \hat{w}_i (Y_i - \beta_0 - \beta_1 X_i) X_i \end{aligned} \quad (5.4)$$

式 (5.4) を 0 と置くと、 β_0 と β_1 の推定値としての $\hat{\beta}_0$ と $\hat{\beta}_1$ が得られる。これらの偏微分した式を整理すると、

$$\begin{aligned} \sum_{i=1}^n \hat{w}_i Y_i - \hat{\beta}_0 \sum_{i=1}^n \hat{w}_i - \hat{\beta}_1 \sum_{i=1}^n \hat{w}_i X_i &= 0 \\ \sum_{i=1}^n \hat{w}_i X_i Y_i - \hat{\beta}_0 \sum_{i=1}^n \hat{w}_i X_i - \hat{\beta}_1 \sum_{i=1}^n \hat{w}_i X_i^2 &= 0 \end{aligned} \quad (5.5)$$

となり、 $\hat{\beta}_0$ と $\hat{\beta}_1$ が含まれない項を右辺に移して整理すると、次の正規方程式

$$\begin{aligned} \hat{\beta}_0 \sum_{i=1}^n \hat{w}_i + \hat{\beta}_1 \sum_{i=1}^n \hat{w}_i X_i &= \sum_{i=1}^n \hat{w}_i Y_i \\ \hat{\beta}_0 \sum_{i=1}^n \hat{w}_i X_i + \hat{\beta}_1 \sum_{i=1}^n \hat{w}_i X_i^2 &= \sum_{i=1}^n \hat{w}_i X_i Y_i \end{aligned} \quad (5.6)$$

が得られる。なお、これらの式について理解が不十分と思われた場合には、第 4 章を先に学習してもらいたい。

デザイン行列を用いた計算に際し、重み \hat{w}_i をベクトル化した場合だとサイズが合わないの
でデザイン行列の積 $\mathbf{X}^T \mathbf{X}$ に組み込めない．そこで、対角要素を $\hat{w}_{ii} = \hat{w}_i$ ，それ以外は $\mathbf{0}$ とす
る $n \times n$ の行列 \mathbf{W} を定義する．これを、 $\mathbf{X}^T \mathbf{X}$ の中に入れ込み

とすると、行列のそれぞれの内側のサイズが一致する。表 1.6 の人工データのデータ数は $n=9$ なので、 9×9 の行列として例示する。重み行列 \hat{W} は、 9×9 の行列であり、 $\mathbf{X}^T \mathbf{X}$ の中に入れると、内側のそれぞれの行列のサイズが一致する。

\mathbf{X}^T と $\hat{\mathbf{W}}$ の積は、 2×9 の行列となり、さらに \mathbf{X} との積は、 2×2 の行列となり、正規方程式 (5.6) の左辺と同様の形式となる。

2×2 の $\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}$ 行列に回帰の推定値 $\hat{\boldsymbol{\beta}}$ を掛けシグマで表記すると,

178

$\mathbf{X}^T \hat{\mathbf{W}}$ は既に計算しているので、 $\mathbf{X}^T \hat{\mathbf{W}}$ と \mathbf{Y} との積は、 2×1 のベクトルとなり、回帰パラメータの推定値正規方程式 (5.6) の右辺に等しくなる。

デザイン行列を用いた計算結果が，正規方程式 (5.6) の右辺と左辺に等しいことから，

| | | | | |
|-------------------------|---------------------------|------------------|-----|-----------------------------|
| $X^T W \wedge X$ | | β^\wedge | | $X^T W \wedge Y$ |
| Σw_i^\wedge | $\Sigma w_i^\wedge X_i$ | β_0^\wedge | $=$ | $\Sigma w_i^\wedge Y_i$ |
| $\Sigma w_j^\wedge X_i$ | $\Sigma w_i^\wedge X_i^2$ | β_1^\wedge | | $\Sigma w_j^\wedge X_i Y_i$ |

$$\hat{\beta} = (X^T \hat{W} X)^{-1} X^T \hat{W} Y \quad (5.9)$$
$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (5.10)$$

179

5.3. 恒等リンクの場合のポアソン回帰

初期パラメータの推定

第 1.4 節でドブソン (2008) の人工データについて詳細を示さずに反復重み付き回帰を用いた結果を示した。第 5.2 節で示した Excel で重み付き回帰のために必要な行列計算の方法を用いて、反復計算の詳細を示す。表 5.1 に示すように、重みなしの回帰パラメータ $\hat{\beta}^{(0)}$ を推定し、推定値 $\hat{Y}^{(0)}$ を求め、重みを計算する。これは、前節の表 4.2 と同様の計算結果で、重みの計算が追加されている。なお、第 1.4 節では、ここに示した回帰係数を初期値としては用いずに、ドブソン (2008) で示されている $\hat{\beta}^{(0)} = [7.0 \ 5.0]^T$ を使っている。

表 5.1 反復 (0) の回帰

| デザイン行列 | | | 回帰 | | | | | |
|--------|-------|-------|--------------------|----------------------|-------------------------|------------------|---------|---------------------------|
| X | | | 推定値 ⁽⁰⁾ | 重み | | | | |
| i | X_0 | X_1 | Y | Y^\wedge | $w^\wedge = 1/Y^\wedge$ | | | |
| 1 | 1 | -1 | 2 | 2.5455 | 0.3929 | $X^T X =$ | 9.0000 | 1.0000 |
| 2 | 1 | -1 | 3 | 2.5455 | 0.3929 | | 1.0000 | 5.0000 |
| 3 | 1 | 0 | 6 | 7.4545 | 0.1341 | $(X^T X)^{-1} =$ | | |
| 4 | 1 | 0 | 7 | 7.4545 | 0.1341 | | 0.1136 | -0.0227 |
| 5 | 1 | 0 | 8 | 7.4545 | 0.1341 | | -0.0227 | 0.2045 |
| 6 | 1 | 0 | 9 | 7.4545 | 0.1341 | $X^T Y =$ | | |
| 7 | 1 | 1 | 10 | 12.3636 | 0.0809 | | 72.0000 | $\beta_0^\wedge = 7.4545$ |
| 8 | 1 | 1 | 12 | 12.3636 | 0.0809 | | 32.0000 | $\beta_1^\wedge = 4.9091$ |
| 9 | 1 | 1 | 15 | 12.3636 | 0.0809 | | | $\beta^{(0)\wedge}$ |
| | | | | $X\beta^{(0)\wedge}$ | | | | |

推定値 $\hat{\beta}^{(0)}$ を次式で求め、

$$\begin{bmatrix} 7.4545 \\ 4.9091 \end{bmatrix} = \begin{bmatrix} 0.1136 & -0.0227 \\ -0.0227 & 0.2045 \end{bmatrix} \begin{bmatrix} 72.0000 \\ 32.0000 \end{bmatrix}$$

$$\beta^{(0)\wedge} = (X^T X)^{-1} X^T Y$$

推定値 $\hat{Y}^{(0)} = X\hat{\beta}^{(0)}$ を得る。さらに、重みを列ベクトル $\hat{w}^{(0)} = 1/\hat{Y}^{(0)}$ として計算する。重みは、 $x_{1,1} = -1$ の場合に $\hat{w}_1 = 1/2.5455 = 0.3929$ 、 $x_{1,3} = 0$ の場合に $\hat{w}_3 = 1/7.4545 = 0.1341$ 、 $x_{1,7} = 1$ の場合に $\hat{w}_7 = 1/12.3636 = 0.0809$ と小さくなっている。これは、偏差平方和の計算過程において、分散の逆数を掛けたことに相当し、基準化した偏差平方和を計算することになる。

重み付き回帰

表 5.1 では、重みを列ベクトル \hat{w} として計算しているので、これを対角要素とするマトリックス \hat{W} とする必要がある。ただし、Excel の行列関数に列ベクトルを対角行列とする行列関数がないので、シート上に 9×9 の行列の枠を作成し、その対角要素に重みを手作業で代入する必要がある。この方法は、データ数 n の変化に対し煩雑な操作を必要とするので、代替の計算手段が必要となる。

表 5.2 反復が(1)の重み付き回帰

| デザイン行列 | | | 回帰 | | リンク | 重み付回帰 | 推定値 | |
|--------|-------|---------------------|--------|----------------------|-----------------------|----------------------------|----------------------|-----------------------|
| X | | | 推定値 | 重み | 関数 | 推定値 | 差 | |
| i | X_0 | X_1 | Y | Y^\wedge | $w^\wedge=1/Y^\wedge$ | $Z=Y$ | Z^\wedge | $Y^\wedge - Z^\wedge$ |
| 1 | 1 | -1 | 2 | 2.5455 | 0.3929 | 2.00 | 2.5165 | 0.0289 |
| 2 | 1 | -1 | 3 | 2.5455 | 0.3929 | 3.00 | 2.5165 | 0.0289 |
| 3 | 1 | 0 | 6 | 7.4545 | 0.1341 | 6.00 | 7.4517 | 0.0029 |
| 4 | 1 | 0 | 7 | 7.4545 | 0.1341 | 7.00 | 7.4517 | 0.0029 |
| 5 | 1 | 0 | 8 | 7.4545 | 0.1341 | 8.00 | 7.4517 | 0.0029 |
| 6 | 1 | 0 | 9 | 7.4545 | 0.1341 | 9.00 | 7.4517 | 0.0029 |
| 7 | 1 | 1 | 10 | 12.3636 | 0.0809 | 10.00 | 12.3868 | -0.0232 |
| 8 | 1 | 1 | 12 | 12.3636 | 0.0809 | 12.00 | 12.3868 | -0.0232 |
| 9 | 1 | 1 | 15 | 12.3636 | 0.0809 | 15.00 | 12.3868 | -0.0232 |
| | | $\beta_0^\wedge=$ | 7.4545 | $X\beta^{(0)\wedge}$ | $\beta_0^\wedge=$ | 7.4517 | $Z\beta^{(1)\wedge}$ | 0.1389 |
| | | $\beta_1^\wedge=$ | 4.9091 | | $\beta_1^\wedge=$ | 4.9351 | | 絶対値の和 |
| | | $\beta^{(0)\wedge}$ | | | $\beta^{(1)\wedge}$ | | | |
| | | | | | | | | |
| | | | | 1.5649 | -0.5431 | 0.7824 | 0.4132 | 8.9813 |
| | | | | -0.5431 | 1.0284 | 0.4132 | 1.1906 | 1.0284 |
| | | | | $(X^*w^\wedge)^T X$ | | $[(X^*w^\wedge)^T X]^{-1}$ | | $(X^*w^\wedge)^T Z$ |

さらに、 $\hat{Z} = X\hat{\beta}^{(1)}$ を計算し、次いで推定値間の差 $\hat{Y} - \hat{Z}$ から、その絶対値の和を求め、乖離の度合を絶対偏差の和で評価すると、

0.1389
絶対値の和

が得られる。推定されたパラメータ間の差は、小数点以下2桁目で異なっている。

| | | | | |
|---------------------|---|---------------------|---|---------|
| 7.4517 | - | 7.4545 | = | -0.0029 |
| 4.9351 | - | 4.9091 | = | 0.0261 |
| $\beta^{(1)\wedge}$ | | $\beta^{(0)\wedge}$ | | 差 |

反復重み付き回帰 (2)

表 5.2 で示した通常の回帰分析に引き続き、重み付き回帰で得られたパラメータは、初期パラメータに対して差の絶対値の和は 0.1389 とかなり大きい。表 5.3 に示すように、第 2 回目の重み付き回帰を繰り返してみる。反復が(1)の重み付き回帰で推定されたパラメータをコピーし、 $\hat{\beta}^{(1)}$ には計算式が入っているので値のみをペースト

第 1 反復：

| | | | | |
|---------------------------|----------------------|---------------------------|----------------------|---------------|
| $\beta_0^\wedge = 7.4545$ | $X\beta^{(0)\wedge}$ | $\beta_0^\wedge = 7.4517$ | $Z\beta^{(1)\wedge}$ | 0.1389 |
| $\beta_1^\wedge = 4.9091$ | | $\beta_1^\wedge = 4.9351$ | | 絶対値の和 |
| $\beta^{(0)\wedge}$ | | $\beta^{(1)\wedge}$ | | |

すると、

第2 反復：

| | | | | | | |
|----------------------|---------------------|----------------------|----------------------|---------------------|----------------------|--------|
| $\beta_0^{\wedge} =$ | 7.4517 | $X\beta^{(1)\wedge}$ | $\beta_0^{\wedge} =$ | 7.4516 | $Z\beta^{(2)\wedge}$ | 0.0008 |
| $\beta_1^{\wedge} =$ | 4.9351 | | $\beta_1^{\wedge} =$ | 4.9353 | | 絶対値の和 |
| | $\beta^{(1)\wedge}$ | | | $\beta^{(2)\wedge}$ | | |

$\hat{\beta}^{(2)}$ が自動的に計算される. 表 5.3 に示すように絶対値の和は, 0.0008 とかなり小さくなり, パラメータも少数点以下 3 桁まで一致している.

表 5.3 反復が(2)の重み付き回帰

| デザイン行列 | | | 回帰 | | リンク | 重み付回帰 | 推定値 | |
|--------|----------------|-------------------------------|--------|------------------------------------|----------------------------------|--|--------------------|------------------------------------|
| X | | | 推定値 | 重み | 関数 | 推定値 | 差 | |
| i | X ₀ | X ₁ | Y | Y [^] | w [^] =1/Y [^] | Z=Y | Z [^] | Y [^] - Z [^] |
| 1 | 1 | -1 | 2 | 2.5165 | 0.3974 | 2.00 | 2.5163 | 0.0002 |
| 2 | 1 | -1 | 3 | 2.5165 | 0.3974 | 3.00 | 2.5163 | 0.0002 |
| 3 | 1 | 0 | 6 | 7.4517 | 0.1342 | 6.00 | 7.4516 | 0.0000 |
| : | | | | | | | | |
| 9 | 1 | 1 | 15 | 12.3868 | 0.0807 | 15.00 | 12.3869 | -0.0001 |
| | | β ₀ [^] = | 7.4517 | Xβ ^{(1)^} | β ₀ [^] = | 7.4516 | Zβ ^{(2)^} | 0.0008 |
| | | β ₁ [^] = | 4.9351 | | β ₁ [^] = | 4.9353 | | 絶対値の和 |
| | | β ^{(1)^} | | | β ^{(2)^} | | | |
| | | | | | | | | |
| | | | | 1.5737 | -0.5526 | 0.7817 | 0.4165 | 8.9999 |
| | | | | -0.5526 | 1.0369 | 0.4165 | 1.1863 | 1.0002 |
| | | | | (X* ^{w^}) ^T X | | [(X* ^{w^}) ^T X] ⁻¹ | | (X* ^{w^}) ^T Z |

反復重み付き回帰 (3)

表 5.4 に示すように, 反復が(3)の重み付き回帰の結果は, 絶対値の和が, 0.0000 となり,

表 5.4 反復が(3)の重み付き回帰

| デザイン行列 | | | 回帰 | | リンク | 重み付回帰 | 推定値 | |
|--------|----------------|-------------------------------|--------|-------------------------------|----------------------------------|---|--------------------|---------------------------------|
| X | | | 推定値 | 重み | 関数 | 推定値 | 差 | |
| i | X ₀ | X ₁ | Y | Y [^] | w [^] =1/Y [^] | Z=Y | Z [^] | Y [^] - Z [^] |
| 1 | 1 | -1 | 2 | 2.5163 | 0.3974 | 2.00 | 2.5163 | 0.0000 |
| 2 | 1 | -1 | 3 | 2.5163 | 0.3974 | 3.00 | 2.5163 | 0.0000 |
| 3 | 1 | 0 | 6 | 7.4516 | 0.1342 | 6.00 | 7.4516 | 0.0000 |
| : | | | | | | | | |
| 9 | 1 | 1 | 15 | 12.3869 | 0.0807 | 15.00 | 12.3869 | 0.0000 |
| | | β ₀ [^] = | 7.4516 | Xβ ^{(2)^} | β ₀ [^] = | 7.4516 | Zβ ^{(3)^} | 0.0000 |
| | | β ₁ [^] = | 4.9353 | | β ₁ [^] = | 4.9353 | | 絶対値の和 |
| | | β ^{(2)^} | | | β ^{(3)^} | | | |
| | | | | 1.5738 | -0.5526 | 0.7817 | 0.4166 | 9.0000 |
| | | | | -0.5526 | 1.0370 | 0.4166 | 1.1863 | 1.0000 |
| | | | | (X* <i>w</i>) ^T X | | [(X* <i>w</i>) ^T X] ⁻¹ | | (X* <i>w</i>) ^T Z |

$\hat{\beta}^{(2)}$ と $\hat{\beta}^{(3)}$ は, 少数点以下 4 桁まで等しくなったので, 解が求まったとみなす. 反復を繰り返せば更に精度の高い推定値を得ることができる.

なお、反復(0)として表 5.1 に示した回帰分析を行うことは、煩わしい。ここに示した手作業の場合には、最初の初期値として $\hat{\mathbf{Y}}$ が 0 にならないような初期値 $\hat{\boldsymbol{\beta}}^{(0)} = [2.0 \ 1.0]^T$ などを与えても、反復計算が 1 回ほど多くなるだけで表 5.4 と同様な解を求めることができる。

これまでに示した結果は、第 1.4 節に対する補足説明となっているので、結果の解釈、共分散行列を用いた 95%信頼区間の作図などを、合わせて見てほしい。なお、第 1.4 節では、 $\hat{\boldsymbol{\beta}}^{(0)} = [7.0 \ 5.0]^T$ と異なる初期値を与えたが、収束結果は当然ながら一致している。

回帰パラメータについてのワルド検定

反復重付き回帰によるポアソン回帰では、パラメータに関する共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ が、反復計算の過程で

$$[(X * \hat{\mathbf{w}})^T X]^{-1} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = \begin{array}{|c|c|} \hline 0.7817 & 0.4166 \\ \hline 0.4166 & 1.1863 \\ \hline \end{array} \quad [(X * \hat{\mathbf{w}})^T X]^{-1} \quad (5.11)$$

として求まっている。パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ の対角要素が、 $\hat{\beta}_0$ と $\hat{\beta}_1$ の分散なので、それぞれの分散の平方根で標準誤差 SE を計算し、

$$\text{Wald カイ 2 乗} = \left(\frac{\text{推定値}}{SE} \right)^2 \quad (5.12)$$

により Wald カイ 2 乗検定統計量を求められる。これが、自由度 1 のカイ 2 乗分布に従うことから、 p 値を求めることができる。表 5.5 に示したように、 $\hat{\beta}_1$ の分散は、1.1863 なので、

$$\chi^2_{\hat{\beta}_1} = \left(\frac{\hat{\beta}_1}{\sqrt{\text{Var}(\hat{\beta}_1)}} \right)^2 = \left(\frac{4.9353}{\sqrt{1.1863}} \right)^2 = 20.5320$$

となる。

表 5.5 反復重み付き回帰による Wald 検定

| 項 | 推定値 | 分散 | 標準誤差 | Waldカイ2乗 | p 値 |
|-------|--------|--------|--------|----------|--------|
| X_0 | 7.4516 | 0.7817 | 0.8841 | 71.0357 | 0.0000 |
| X_1 | 4.9353 | 1.1863 | 1.0892 | 20.5320 | 0.0000 |

第 2.4 節では、反復重付き回帰ではなく、ニュートン・ラフソンによる解析法を示し、計算の過程で求められている負のヘッセ行列の逆数 $(-\mathbf{H})^{-1}$ がパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ とな

ることを示した。第 2.4 節では、尤度比カイ 2 乗検定の結果を示したが、Wald カイ 2 乗検定を行うことも可能である。表 2.14 および表 2.15 に示したパラメータの共分散行列 $(-H)^{-1}$ を次に示す。

$$(-H)^{-1} = \Sigma(\hat{\beta}) = \begin{array}{c} \begin{array}{|c|} \hline \text{負の逆行列} \\ \hline \end{array} \\ \begin{array}{|c|} \hline (-H)^{-1} \\ \hline \end{array} \\ \begin{array}{|c|c|} \hline 0.7817 & 0.4160 \\ \hline 0.4160 & 1.1915 \\ \hline \end{array} \end{array} \quad (5.13)$$

共分散行列といっても式 (5.13) で示したのと式 (5.11) で示したのでは、若干の差異がある。その結果として、表 5.6 に示した共分散行列 $(-H)^{-1}$ の対角要素を分散とした Wald 検定結果も表 5.5 のカイ 2 乗値と若干の差異が生ずる。

表 5.6 負のヘッセ行列の逆行列 $(-H)^{-1}$ を用いた Wald 検定

| 項 | 推定値 | 分散 | 標準誤差 | Waldカイ2乗 | p 値 |
|-------|--------|--------|--------|----------|--------|
| X_0 | 7.4516 | 0.7817 | 0.8842 | 71.0300 | 0.0000 |
| X_1 | 4.9353 | 1.1915 | 1.0915 | 20.4428 | 0.0000 |

尤度比検定

第 2.4 節の表 2.17 で示した JMP によるポアソン回帰は、尤度比検定を標準的に出力するので、Wald 検定とは若干の差異が生ずる。表 2.17 を表 5.7 として再掲する。表 5.5 に示した反復重み付きによる $\hat{\beta}_1$ のカイ 2 乗値が 20.5320 であるのに対し、尤度比カイ 2 乗の結果は 16.5260 とかなり異なる。

表 5.7 JMP によるポアソン回帰の結果 (表 2.16 再掲)

| パラメータ推定値 | | | | | 推定値の共分散 | | |
|----------|--------|--------|---------|----------------|---------|--------|--------|
| 項 | 推定値 | 標準誤差 | 尤度比カイ2乗 | p値(Prob>ChiSq) | 共分散 | | |
| 切片 | 7.4516 | 0.8842 | 71.0299 | <.0001* | 切片 | 0.7817 | 0.4160 |
| x | 4.9353 | 1.0915 | 16.5260 | <.0001* | x | 0.4160 | 1.1915 |

正規分布を仮定した最小 2 乗法による回帰分析では、各種の統計ソフトによって結果が異なることはなくなったが、一般化線形モデルとしてのポアソン回帰では、SAS/GENMODE プロシジャの解析手法と JMP/一般線形モデルに対する最尤法の解析方法が異なるために、若干の差異が生ずる。従って、解析結果を公表する際には、使用した統計ソフトが用いている最尤法の計算方法、尤度比検定なのかワルド検定なのかを明示する必要がある。

5.4. 対数リンクでのポアソン回帰

第 1.5 節で取り上げた冠動脈心疾患の死亡者数は、オフセットがある対数リンクの事例であるが、オフセットを無視して、表 5.8 に示すようにオフセットなしの対数リンクの事例として取り上げる [ドブソン (2008)].

対数リンク

リンク関数が「恒等」の場合に対し「対数リンク」の場合は、ポアソン回帰の特徴的な反復重み付き回帰となる。死亡者数 y_i が、指数関数的に増加するとして、

$$Y_i = \exp(\beta_0 + \beta_1 X_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{Poisson}(\hat{y}_i) \quad (5.14)$$

をあてはめている。この式に対して両辺に対数を取ると誤差 ε_i が和の形で入っているために線形化できない。

$$\ln Y_i = \ln[\exp(\beta_0 + \beta_1 X_i) + \varepsilon_i] \quad (5.15)$$

回帰係数を推定値で表記することにより、

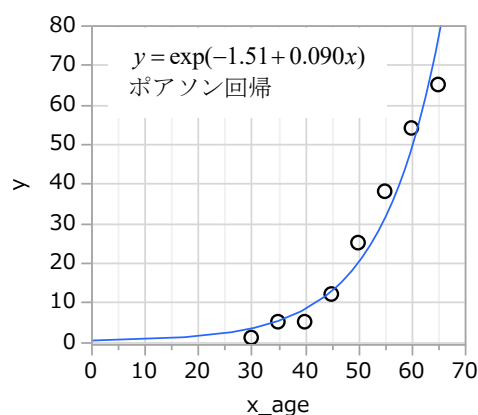
$$\ln \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (5.16)$$

と、線形化でき、反復重み付き回帰を行なう。このために指数ではなく、対数リンクと言われている。

さて、元のデータ y_i は、カウント・データなので $y_i = 0$ も発生する。ゼロについて対数はマイナスの無限大となり、計算不能となってデータに含められない。どうしても良いのだろうか。便宜的には、全データに 1 を加えて対数を取ること考えられるし、ゼロの場合に 0.5 に置き換えることも考えられる。

表 5.8 オーストラリアのある地方の冠動脈心疾患の死亡者数

| No. | 年齢層 X | 死亡者数 Y |
|-----|---------|----------|
| 1 | 30 | 1 |
| 2 | 35 | 5 |
| 3 | 40 | 5 |
| 4 | 45 | 12 |
| 5 | 50 | 25 |
| 6 | 55 | 38 |
| 7 | 60 | 54 |
| 8 | 65 | 65 |



一般化線形モデルは、ポアソン回帰の対数リンクのみならず、この様なリンク関数によっては、変換不能となるような場合でも、元の反応データのまま適切に対応できる計算アルゴリズムとなっている。年齢層が 30 の場合に死亡者数が $Y_1=0$ 人と仮定した場合に、表 5.9 に示すように初期値の計算で、便宜的に $Y_1=0.5$ 人として、回帰パラメータを推定すると表 5.9 に示すように

$$\ln Y_1 = \ln(0.5) = -0.6931$$

を得る。他のデータについても対数を取り、回帰係数を求めることは可能であるが、便宜的な対応と言わざるを得ない。

表 5.9 対数変換できない場合の便宜的な対応

| | デザイン行列 | | 死亡 | 回帰 | | | | |
|-----|--------|-------|-----|---------|------------------|---------|----------------------|----------------------------|
| | X | | 者数 | 推定値 | | | | |
| i | X_0 | X_1 | Y | $\ln Y$ | | | | |
| 1 | 1 | 30 | 0.5 | -0.6931 | $X^T X =$ | 8.0 | 380.0 | |
| 2 | 1 | 35 | 5 | 1.6094 | | 380.0 | 19100.0 | |
| 3 | 1 | 40 | 5 | 1.6094 | | | | |
| 4 | 1 | 45 | 12 | 2.4849 | $(X^T X)^{-1} =$ | 2.2738 | -0.0452 | |
| 5 | 1 | 50 | 25 | 3.2189 | | -0.0452 | 0.0010 | |
| 6 | 1 | 55 | 38 | 3.6376 | | | | $(X^T X)^{-1} X^T (\ln Y)$ |
| 7 | 1 | 60 | 54 | 3.9890 | $X^T \ln Y =$ | 20.0305 | $\beta_0^{\wedge} =$ | -3.4664 |
| 8 | 1 | 65 | 65 | 4.1744 | | 1083.42 | $\beta_1^{\wedge} =$ | 0.1257 |

一般化線形モデルでの反復重み付き回帰のアルゴリズムでは、何らかの最初の初期パラメータを得た後に、便宜的な対応ではなく、元のスケールのデータを変換する際に「補正式」を用いている。

対数リンクの場合、 $Y_i=0$ としたときの場合について例示する。補正式は、一般化線形モデルの公式により、

$$\hat{Z}_i = \ln \hat{Y}_i + \frac{Y_i - \exp(\ln \hat{Y}_i)}{\exp(\ln \hat{Y}_i)} \quad (5.17)$$

であり、何らかの初期パラメータによって、 $\hat{Y}_i > 0$ が推定された場合に、補正值 \hat{Z}_i は、 $Y_i = 0$ の場合でも計算不能とならない。表 5.9 で推定された $\hat{\beta}_0 = -3.4664$ 、 $\hat{\beta}_1 = 0.1257$ の場合に、推定値は、

$$\begin{aligned} \hat{Y}_1 &= -3.4664 + 0.1257 X_1 \\ &= -3.4664 + 0.1257 \times 30 = 0.3043 \end{aligned}$$

となり、

$$\begin{aligned}
\hat{Z}_i &= \ln \hat{Y}_i + \frac{Y_i - \exp(\ln \hat{Y}_i)}{\exp(\ln \hat{Y}_i)} \\
&= 0.3043 + \frac{0 - 0.3043}{0.3043} \\
&= -0.6957
\end{aligned}$$

これを元のスケールに直すと

$$Y'_i = \exp(-0.6957) = 0.4987$$

おおよそ、0.5 程度の値となる。他の \hat{Z}_i , $i=2,3,\dots,9$ についても同じ補正式で調整して、この調整値 \hat{z}_i について重み付き回帰分析を行う。

式 (5.17) は、リンク関数が「対数」の場合の一般化線形モデルの公式から与えられるとしたのであるが、他のリンク関数の場合を含む一般公式を用いて導出する。対数リンク式を

$$\eta_i = \ln \hat{Y}_i \quad (5.18)$$

とし、元の推定値を

$$\begin{aligned}
\mu_i &= \exp(\ln \hat{Y}_i) \\
\ln \mu_i &= \ln \hat{Y}_i
\end{aligned} \quad (5.19)$$

としたときに、一般線形モデルの公式により、補正した対数 Z_i は、

$$\begin{aligned}
Z_i &= \eta_i + (Y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \\
&= \eta_i + (Y_i - \mu_i) \left(\frac{\partial \ln(\mu_i)}{\partial \mu_i} \right) \\
&= \eta_i + \frac{(Y_i - \mu_i)}{\mu_i} \\
&= \ln \hat{Y}_i + \frac{Y_i - \exp(\ln \hat{Y}_i)}{\exp(\ln \hat{Y}_i)}
\end{aligned} \quad (5.20)$$

となる。重みは、

$$\begin{aligned}
w_i &= \frac{1}{\text{var}(\hat{Y}_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \\
&= \frac{1}{\text{var}(\hat{Y}_i)} \left(\frac{\partial \ln \mu_i}{\partial \mu_i} \right)^2 \\
&= \frac{1}{\frac{\mu_i}{\mu_i^2}} = \mu_i = \exp(\eta_i) = \exp(\ln \hat{Y}_i)
\end{aligned} \quad (5.21)$$

となる。この公式は、リンク関数が「対数」のみならず、「プロビット」、「ロジット」、「補 2 重対数」などでも使われる。

重み付き回帰 (1)

さて, $y_1 = 0.5$ を $y_1 = 1$ と元のデータに戻し, 回帰係数を推定し直すと初期値が,

$$\hat{\beta}^{(0)} = (X^T X)^{-1} X^T \ln(Y) = \begin{bmatrix} -2.8310 \\ 0.1141 \end{bmatrix}$$

として得られる. この推定値を用いて表 5.10 に示すように反復重み付き回帰を行う. 得られた初期値 $\hat{\beta}^{(0)}$ を $\hat{\beta}^{(m-1)}$ の欄に値のみペーストすると, 反復 1 の結果が $\hat{\beta}^{(m)}$ の欄に求められている.

表 5.10 <反復 1> 冠動脈心疾患の死亡者数 (対数リンク)

| | デザイン行列 | | 死亡 | 回帰 | | 補正リンク | 重付回帰 | 推定値 |
|---|---------|--------------------|------------------------|----------------------|-----------------------|-----------------------------|------------|---------------------------|
| | X | | 者数 | 推定値 | 重み | 関数 | 推定値 | 差 |
| i | X_0 | X_1 | Y | $\ln Y^\wedge$ | $w^\wedge = Y^\wedge$ | Z | Z^\wedge | $\ln Y^\wedge - Z^\wedge$ |
| 1 | 1 | 30 | 1 | 0.5931 | 1.8096 | 0.1457 | 1.2472 | -0.6541 |
| 2 | 1 | 35 | 5 | 1.1638 | 3.2019 | 1.7253 | 1.6903 | -0.5266 |
| 3 | 1 | 40 | 5 | 1.7344 | 5.6657 | 1.6169 | 2.1335 | -0.3991 |
| 4 | 1 | 45 | 12 | 2.3051 | 10.0253 | 2.5021 | 2.5767 | -0.2716 |
| 5 | 1 | 50 | 25 | 2.8758 | 17.7394 | 3.2851 | 3.0199 | -0.1441 |
| 6 | 1 | 55 | 38 | 3.4465 | 31.3893 | 3.6571 | 3.4631 | -0.0166 |
| 7 | 1 | 60 | 54 | 4.0171 | 55.5423 | 3.9894 | 3.9063 | 0.1109 |
| 8 | 1 | 65 | 65 | 4.5878 | 98.2800 | 4.2492 | 4.3495 | 0.2384 |
| $\beta_0^\wedge =$ | -2.8310 | $\beta_0^\wedge =$ | -2.8310 | | $\beta_0^\wedge =$ | -1.4120 | | 2.3613 |
| $\beta_1^\wedge =$ | 0.1141 | $\beta_1^\wedge =$ | 0.1141 | | $\beta_1^\wedge =$ | 0.0886 | 絶対値の和 | |
| $\hat{\beta}^{(0)} = (X^T X)^{-1} X^T \ln(Y)$ | | | $\beta^{(m-1)^\wedge}$ | | | $\beta^{(m)^\wedge}$ | | |
| | | | | 223.7 | 13178.2 | 0.2734 | -0.0046 | 852.29 |
| | | | | 13178.2 | 789404.0 | -0.0046 | 0.0001 | 51363.34 |
| | | | | $(X * w^\wedge)^T X$ | | $[(X * w^\wedge)^T X]^{-1}$ | | $(X * w^\wedge)^T Z$ |

Excel シートに埋め込まれている計算式は, 次の通りである.

- 1) $\ln \hat{Y} = \text{Mmult}(X \text{ の範囲}, \hat{\beta}^{(m-1=0)} \text{ の範囲})$
- 2) $\hat{w} = \exp(\ln \hat{Y} \text{ の範囲}) = \hat{Y}$
- 3) $Z = \ln \hat{Y} + (Y - \exp(\ln \hat{Y})) / \exp(\ln \hat{Y})$: 以下「の範囲」は省略
- 4) $(X * \hat{w})^T X = \text{Mmult}(\text{Transpose}(X * \hat{w}), X)$
- 5) $[(X * \hat{w})^T X]^{-1} = \text{Minvers}((X * \hat{w})^T X)$
- 6) $(X * \hat{w})^T Z = \text{Mmult}(\text{Transpose}(X * \hat{w}), Z)$
- 7) $\hat{\beta}^{(1)} = \text{Mmult}([(X * \hat{w})^T X]^{-1}, (X * \hat{w})^T Z)$
- 8) $\hat{Z} = \text{Mmult}(Z, \hat{\beta}^{(m=1)})$
- 9) $\ln \hat{Y} - \hat{Z} = \ln \hat{Y} \text{ の範囲} - \hat{Z} \text{ の範囲}$
- 10) 絶対値の和 = $\text{Sum}(\text{Abs}(\ln \hat{Y} - \hat{Z}))$

重み付き回帰 (2) および (3)

第1反復では、「 $\ln \hat{Y} - \hat{Z}$ の絶対値の和」が、2.3613と大きいので、第2反復を行う。第1反復で得られたパラメータの推定値

$$\hat{\beta}^{(1)} = \begin{matrix} -1.4120 \\ 0.0886 \\ \beta^{(m)\wedge} \end{matrix} \text{ を } \begin{matrix} -2.8310 \\ 0.1141 \\ \beta^{(m-1)\wedge} \end{matrix} \text{ に値のみをペースト}$$

すると第2反復の結果は

$$(m=2): \begin{matrix} -1.5053 \\ 0.0898 \\ \beta^{(m)\wedge} \end{matrix}$$

が得られるが、「 $\ln \hat{Y} - \hat{Z}$ の絶対値の和」が、0.2866と、まだ大きいので、この結果を、 $\hat{\beta}^{(m-1)}$ に値のみをペーストして第3反復を行った結果を表5.11に示す。

表 5.11 <反復3>冠動脈心疾患の死亡者数 (対数リンク)

| | デザイン行列 | | 死亡 | 回帰 | | 補正リンク | 重付回帰 | 推定値 |
|---|---------|-----------------------|---------|----------------------|-----------------------|-----------------------------|------------|---------------------------|
| | X | | 者数 | 推定値 | 重み | 関数 | 推定値 | 差 |
| i | X_0 | X_1 | Y | $\ln Y^\wedge$ | $w^\wedge = Y^\wedge$ | Z | Z^\wedge | $\ln Y^\wedge - Z^\wedge$ |
| 1 | 1 | 30 | 1 | 1.1901 | 3.2875 | 0.4943 | 1.1888 | 0.0014 |
| 2 | 1 | 35 | 5 | 1.6394 | 5.1520 | 1.6099 | 1.6382 | 0.0012 |
| 3 | 1 | 40 | 5 | 2.0886 | 8.0737 | 1.7079 | 2.0876 | 0.0010 |
| : | | | | | | | | |
| 8 | 1 | 65 | 65 | 4.3348 | 76.3105 | 4.1866 | 4.3348 | 0.0001 |
| $\beta_0^\wedge =$ | -2.8310 | $\beta_0^\wedge =$ | -1.5053 | | $\beta_0^\wedge =$ | -1.5078 | | 0.0058 |
| $\beta_1^\wedge =$ | 0.1141 | $\beta_1^\wedge =$ | 0.0898 | | $\beta_1^\wedge =$ | 0.0899 | | 絶対値の和 |
| $\hat{\beta}^{(0)} = (X^T X)^{-1} X^T \ln(Y)$ | | $\beta^{(m-1)\wedge}$ | | | | $\beta^{(m)\wedge}$ | | |
| | | | | 205.1 | 11753.5 | 0.2176 | -0.0037 | 747.26 |
| | | | | 11753.5 | 689087.2 | -0.0037 | 0.0001 | 44216.90 |
| | | | | $(X * w^\wedge)^T X$ | | $[(X * w^\wedge)^T X]^{-1}$ | | $(X * w^\wedge)^T Z$ |

$\ln \hat{Y} - \hat{Z}$ の絶対値の和が、0.0058とかなり小さくなる。第3反復は、まだ収束はしていないが、次の反復で収束する直前の結果である。したがって、

$$(m=3): \begin{matrix} \beta_0^\wedge = -1.5078 \\ \beta_1^\wedge = 0.0899 \end{matrix}$$

が、オフセットなしの対数リンクで推定されたパラメータになる。パラメータの共分散行列 $\Sigma(\beta^\wedge)$ は、

$$\text{共分散行列: } \Sigma(\beta^\wedge) = \begin{matrix} \begin{matrix} 0.2176 & -0.0037 \\ -0.0037 & 0.0001 \end{matrix} \\ \begin{matrix} [(X * w^\wedge)^T X]^{-1} \end{matrix} \end{matrix}$$

となる。

95%信頼区間

$\Sigma(\hat{\beta})$ を用いて、回帰直線の95%信頼区間を求める。それぞれの*i*ごとに、 $\mathbf{x}_i = [x_{0,i} \ x_{1,i}]$ としたとき \hat{Z}_i の分散は、次の2次形式で

$$Var(\hat{Z}_i) = \mathbf{x}_i \Sigma(\hat{\beta}) \mathbf{x}_i^T$$

求められる。 \hat{Z}_i の95%信頼区間は、

$$\hat{Z}_i \pm 1.96 \sqrt{Var(\hat{Z}_i)} = (L95\%, U95\%)$$

であり、表 5.12 に計算した結果を示す。元のスケールは、対数の95%信頼区間について指数を取って計算したものである。

表 5.12 対数リンクの場合の95%信頼区間

| \mathbf{X} | | | 対数 | | | | | 元のスケール | | | |
|--------------|----------------|----------------|--------------|---------------------|--------------------------|--------------|--------------|--------------|---------------------|--------------|--------------|
| <i>i</i> | \mathbf{X}_0 | \mathbf{X}_1 | \mathbf{Z} | \mathbf{Z}^\wedge | $Var(\mathbf{Z}^\wedge)$ | <i>L</i> 95% | <i>U</i> 95% | \mathbf{Y} | \mathbf{Y}^\wedge | <i>L</i> 95% | <i>U</i> 95% |
| 1 | 1 | 30 | 0.4943 | 1.1888 | 0.0532 | 0.7367 | 1.6408 | 1 | 3.283 | 2.089 | 5.159 |
| 2 | 1 | 35 | 1.6099 | 1.6382 | 0.0371 | 1.2606 | 2.0158 | 5 | 5.146 | 3.527 | 7.507 |
| 3 | 1 | 40 | 1.7079 | 2.0876 | 0.0243 | 1.7821 | 2.3931 | 5 | 8.066 | 5.943 | 10.947 |
| : | | | | | | | | | | | |
| 8 | 1 | 65 | 4.1866 | 4.3348 | 0.0087 | 4.1519 | 4.5176 | 65 | 76.306 | 63.556 | 91.614 |

パラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた計算の実例を、*i*=1の場合について示す。まず、対数についての $Var(\hat{Z}_1)$ は、2次形式の計算法で、0.0532 が得られる。

$$Var(\mathbf{Z}_1^\wedge) = \begin{bmatrix} 1 & 30 \end{bmatrix} \begin{bmatrix} 0.2178 & -0.0037 \\ -0.0037 & 0.0001 \end{bmatrix} \begin{bmatrix} 1 \\ 30 \end{bmatrix} = 0.0532$$

$$\Sigma(\hat{\beta}) = [(\mathbf{X}^* \mathbf{w}^\wedge)^T \mathbf{X}]^{-1} \quad \mathbf{x}_1^T$$

信頼区間の計算は、

$$\hat{Z}_1 \pm 1.96 \sqrt{Var(\hat{Z}_1)} = 1.1888 \pm 1.96 \sqrt{0.0532} = (0.7365, 1.6408)$$

で得られる。指数を取って元のスケールでは、

$$\hat{Y}_1 = \exp(\hat{Z}_1) = \exp(1.1888) = 3.283$$

$$L95\% = \exp(0.7365) = 2.089$$

$$U95\% = \exp(1.6408) = 5.160$$

となる。

これらの計算結果より，図 5.1 に対数軸と元のスケールでの 95%信頼区間を示す．なお，この図は Excel ではなく，JMP に対数リンクでのポアソン回帰を行い，推定値および 95%信頼区間を JMP ファイルに書き出して Excel の結果と一致することを確認の上，「重ね合わせプロット」で作成したものである．

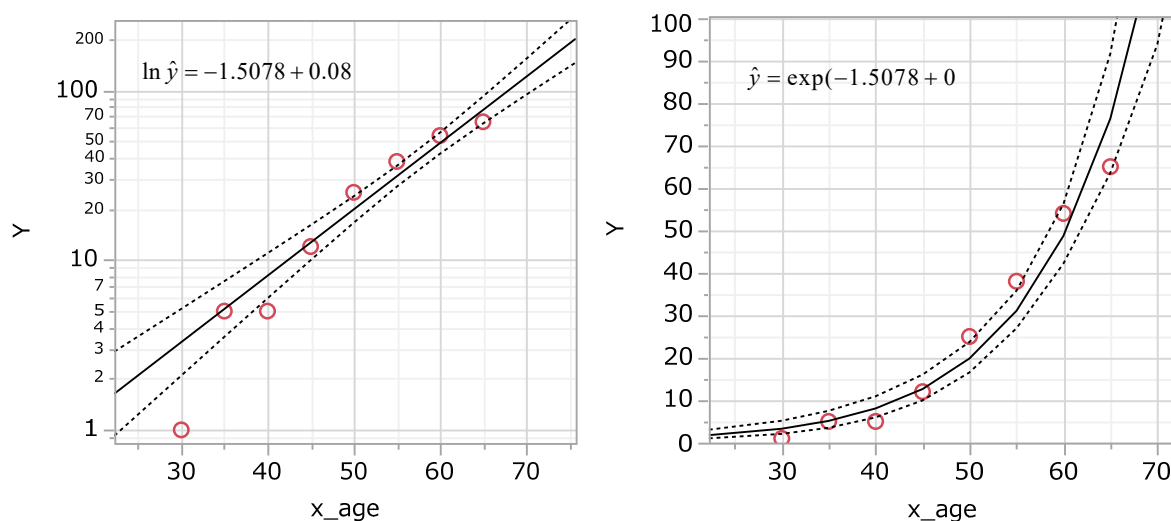


図 5.1 冠動脈心疾患の死亡者数に対するポアソン回帰曲線と 95%信頼区間

2 次式のあてはめ

図 5.1 左の対数リンクによる回帰直線に対して，元データのプロットは上に凸の曲線状となり，直線のあてはめには疑問が残る．これは，年齢が高くなれば，人口が減り冠動脈心疾患による死亡者数も相対的に減るであろうし，10 万人当たりの死亡数も全死亡の一部でもあり，上限があり頭打ちになることも考えられる．そこで，65 歳までのデータから頭打ちが確認できるのか検討する．表 5.13 に示すように年齢について 2 乗の項を追加し，直線のあてはめが妥当かを検討する．

反復重み付き回帰の利便性は，説明変数の数を増やしても，行列のサイズが大きくなるだけで，Excel の計算シートの本体に大きな変更がない．表 5.13 に示すように，デザイン行列 X のサイズを 8×2 から 8×3 に変更し，パラメータ β のサイズを 2×1 から 3×1 に変更するが，最初の $\beta^{(0)}$ の計算式 $\beta^{(0)} = (X^T X)^{-1} X^T \ln(Y)$ に変更はない．ただし，Excel では，行列の範囲をフィルハンドルによって拡大することができる．他の計算式も同様に行列の範囲を変更しなければならないので，煩雑さはいないが機械的な作業レベルで対応できる．

表 5.13 冠動脈心疾患の死亡者数に対する 2 次式のあてはめ (対数リンク)

| | デザイン行列 | | | 死亡 | 回帰 | | 補正リンク | 重付回帰 | 推定値 |
|---|---------|-------|----------------------|-----------------------|----------------|--|---------------------|------------|---------------------------|
| | X | | | 者数 | 推定値 | 重み | 関数 | 推定値 | 差 |
| i | X_0 | X_1 | $X_2 = X_1^2$ | Y | $\ln Y^\wedge$ | $w^\wedge = Y^\wedge$ | Z | Z^\wedge | $\ln Y^\wedge - Z^\wedge$ |
| 1 | 1 | 30 | 900 | 1 | 0.2151 | 1.2400 | 0.0216 | 0.2151 | 0.0000 |
| 2 | 1 | 35 | 1225 | 5 | 1.1259 | 3.0829 | 1.7477 | 1.1259 | 0.0000 |
| 3 | 1 | 40 | 1600 | 5 | 1.9224 | 6.8372 | 1.6537 | 1.9224 | 0.0000 |
| 4 | 1 | 45 | 2025 | 12 | 2.6046 | 13.5262 | 2.4918 | 2.6046 | 0.0000 |
| 5 | 1 | 50 | 2500 | 25 | 3.1726 | 23.8702 | 3.2200 | 3.1726 | 0.0000 |
| 6 | 1 | 55 | 3025 | 38 | 3.6264 | 37.5769 | 3.6376 | 3.6264 | 0.0000 |
| 7 | 1 | 60 | 3600 | 54 | 3.9659 | 52.7675 | 3.9893 | 3.9659 | 0.0000 |
| 8 | 1 | 65 | 4225 | 65 | 4.1912 | 66.0991 | 4.1745 | 4.1912 | 0.0000 |
| $\beta_0^\wedge =$ | -7.6076 | | $\beta_0^\wedge =$ | -7.6486 | | $\beta_0^\wedge =$ | -7.6486 | | 0.0000 |
| $\beta_1^\wedge =$ | 0.3277 | | $\beta_1^\wedge =$ | 0.3307 | | $\beta_1^\wedge =$ | 0.3307 | | 絶対値の和 |
| $\beta_2^\wedge =$ | -0.0022 | | $\beta_2^\wedge =$ | -0.0023 | | $\beta_2^\wedge =$ | -0.0023 | | |
| $\hat{\beta}^{(0)} = (X^T X)^{-1} X^T \ln(Y)$ | | | | $\beta^{(m-1)\wedge}$ | | | $\beta^{(m)\wedge}$ | | |
| | | | 2.05E+02 | 1.18E+04 | 6.86E+05 | 6.9415 | -0.2619 | 0.0024 | 7.50E+02 |
| | | | 1.18E+04 | 6.86E+05 | 4.06E+07 | -0.2619 | 0.0100 | -0.0001 | 4.41E+04 |
| | | | 6.86E+05 | 4.06E+07 | 2.44E+09 | 0.0024 | -0.0001 | 8.73E-07 | 2.62E+06 |
| | | | $(X^* w^\wedge)^T X$ | | | $\Sigma(\beta^\wedge) = [(X^* w^\wedge)^T X]^{-1}$ | | | $(X^* w^\wedge)^T Z$ |

表 5.13 で求めたパラメータ $\boldsymbol{\beta}^{(m)}$ ，共分散行列の対角要素から $\boldsymbol{\beta}^{(m)}$ の分散を表 5.14 に取り出し，検定統計量の計算を追加する．

表 5.14 冠動脈心疾患の死亡者数に対する Wald 検定の結果

| 項 | 推定値 | 分散 | 標準誤差 | Waldカイ2乗 | p 値 |
|---------------|---------|----------|----------|----------|--------|
| X_0 | -7.6486 | 6.9415 | 2.6347 | 8.42771 | 0.0037 |
| X_1 | 0.3307 | 0.0100 | 0.1000 | 10.93412 | 0.0009 |
| $X_2 = X_1^2$ | -0.0023 | 8.73E-07 | 9.34E-04 | 5.98016 | 0.0145 |

Wald 検定の結果から、2 次の項が $p=0.0145$ と有意であることから、65 歳までのデータで頭打ちが統計的に確認されたことになる。そこで、表 5.12 と同様に表 5.15 に 95%信頼区間の計算を行い、図 5.2 にポアソン回帰の 2 次曲線を示す。図 5.1 に比べて、あてはまりは良くなり、対数死亡者数目盛りの場合に場合に年齢が増えるにつれ徐々に頭打ちとなることが観察される。実目盛りでは、70 歳がピークとなり、減少に転じている。

対数目盛上での 2 次式のあてはめは便宜的な方法であり，得られた 2 次式のパラメータについて意味づけすることは生産的でないが，少なくとも直線のあてはめには難点があると理解できることに意義がある．

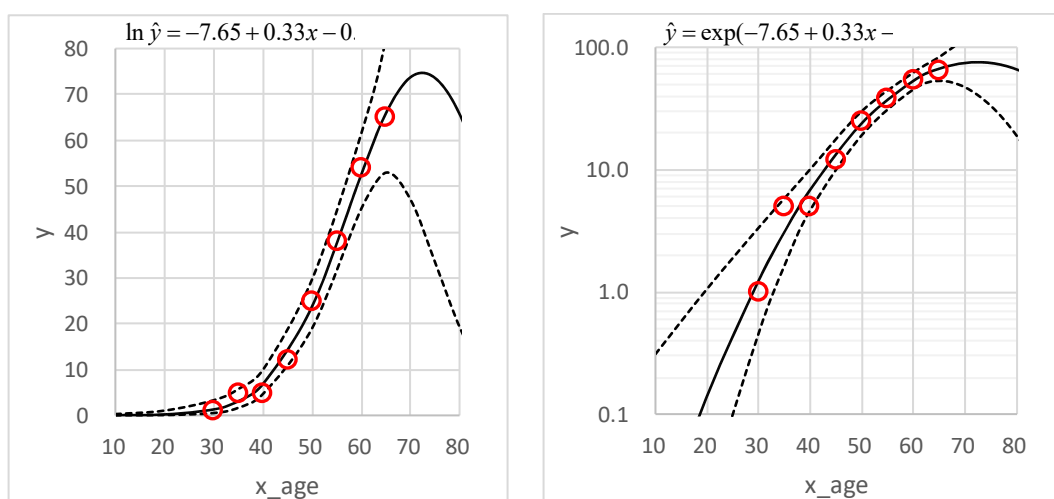


図 5.2 冠動脈心疾患の死亡者数に対するポアソン 2 次回帰と 95%信頼区間

表 5.15 2 次式の 95%信頼区間

| i | X | | | 対数 | | | | 元のスケール | | |
|-----|-------|-------|-------|------------|-----------------|---------|---------|------------|--------|---------|
| | X_0 | X_1 | X_2 | Z^\wedge | $Var(Z^\wedge)$ | L95% | U95% | Y^\wedge | L95% | U95% |
| 1 | 1 | 10 | 100 | -4.5704 | 3.0094 | -7.9705 | -1.1702 | 0.010 | 0.000 | 0.310 |
| 2 | 1 | 20 | 400 | -1.9491 | 1.0481 | -3.9557 | 0.0575 | 0.142 | 0.019 | 1.059 |
| 3 | 1 | 30 | 900 | 0.2151 | 0.2553 | -0.7752 | 1.2054 | 1.240 | 0.461 | 3.338 |
| 4 | 1 | 35 | 1225 | 1.1259 | 0.1036 | 0.4951 | 1.7567 | 3.083 | 1.641 | 5.793 |
| 5 | 1 | 40 | 1600 | 1.9224 | 0.0380 | 1.5402 | 2.3045 | 6.837 | 4.666 | 10.020 |
| 6 | 1 | 50 | 2500 | 3.1726 | 0.0130 | 2.9495 | 3.3958 | 23.870 | 19.096 | 29.838 |
| 7 | 1 | 60 | 3600 | 3.9659 | 0.0063 | 3.8101 | 4.1217 | 52.768 | 45.154 | 61.664 |
| 8 | 1 | 65 | 4225 | 4.1912 | 0.0129 | 3.9687 | 4.4136 | 66.099 | 52.914 | 82.570 |
| 9 | 1 | 70 | 4900 | 4.3022 | 0.0538 | 3.8474 | 4.7569 | 73.860 | 46.873 | 116.384 |
| 10 | 1 | 75 | 5625 | 4.2989 | 0.1663 | 3.4995 | 5.0983 | 73.621 | 33.100 | 163.748 |
| 11 | 1 | 80 | 6400 | 4.1814 | 0.4008 | 2.9407 | 5.4222 | 65.460 | 18.928 | 226.382 |
| 12 | 1 | 85 | 7225 | 3.9497 | 0.8204 | 2.1744 | 5.7250 | 51.920 | 8.797 | 306.446 |

対数リンクの 2 次式の分散も表 5.12 の 1 次式の場合と同様に，パラメータの共分散行列 $\Sigma(\hat{\beta})$ を用いた 2 次形式で求めることができる．計算の実例を， $i=1$ の場合について示す．まず，対数についての $Var(\hat{Z}_1)$ は，2 次形式の計算法で，3.0094 が得られる．この分散を用いて 95%信頼区間の計算を行える．

| | | | | | | | | | |
|---|---|-------|-----|---------|---------|----------|---------|---|--------|
| $Var(Z_1^\wedge) =$ | 1 | 10 | 100 | 6.9415 | -0.2619 | 0.0024 | 1 | = | 3.0094 |
| | | x_1 | | -0.2619 | 0.0100 | -0.0001 | 10 | | |
| | | | | 0.0024 | -0.0001 | 8.73E-07 | 100 | | |
| $\Sigma(\hat{\beta}^\wedge) = [(X^*w^\wedge)^T X]^{-1}$ | | | | | | | x_1^T | | |

5.5. 対数リンクでオフセットがある場合のポアソン回帰

第 5.4 節では、冠動脈心疾患の死亡者数についてオフセットを無視して対数リンクによるポアソン回帰を例示した。元のデータである第 1.5 節データには、部分母集団の人数が得られているので、表 5.16 にように死亡率を求めることができる [ドブソン (2008)]。このデータについては、第 2.6 節で、ポアソン確率を用いたニュートン・ラフソン法による最尤法の解析法でも取り上げており、ここでは、反復重み付き回帰による解析法について示す。

表 5.16 冠動脈心疾患の死亡率 (表 1.11 再掲)

| | 年齢層 | 死亡者数 | 母集団 | 死亡率 | 1 万比 |
|-----|-----|------|--------|--------|-------|
| No. | X | Y | 人数 n | % | 人 |
| 1 | 30 | 1 | 17,742 | 0.0056 | 0.56 |
| 2 | 35 | 5 | 16,554 | 0.0302 | 3.02 |
| 3 | 40 | 5 | 16,059 | 0.0311 | 3.11 |
| 4 | 45 | 12 | 13,083 | 0.0917 | 9.17 |
| 5 | 50 | 25 | 10,784 | 0.2318 | 23.18 |
| 6 | 55 | 38 | 9,645 | 0.3940 | 39.40 |
| 7 | 60 | 54 | 10,706 | 0.5044 | 50.44 |
| 8 | 65 | 65 | 9,933 | 0.6544 | 65.44 |

この様に死亡率が求められるデータに対しては、一般化線形モデルで、分布を 2 項分布とし、リンク関数を (ロジット or プロビット or 補 2 重対数) とした解析が適しているようにも思われる。しかし、死亡率が最も高い年齢 65 歳階層でも 1.0 パーセントにも届かないデータに、0~100 パーセントの範囲を規定するシグモイド曲線のあてはめるような統計モデルを適用することは、違和感が付きまとう。

全ての死亡を対象にした場合には、0~100 パーセントの死亡率を想定することは理にかなっているように思われる。しかし、冠動脈心疾患での高年齢での死亡率はおおよそ 15%程度 (厚労省・2018 年, 人口動態統計, 第 6 表 性別にみた死因順位 (第 10 位まで) 別 死亡数・死亡率 (人口 10 万対)・構成割合) であるので、0~100 パーセントを仮定した方法の適用は、違和感が付きまとう。リンク関数を対数としたポアソン回帰が望ましいように思われるが、分母があるカウント・データに対して適用できるのだろうか。

オフセットを含めたポアソン回帰

死亡人数 Y_i に対して部分母集団の人数 n_i を加えた指数曲線をあてはめるモデル

$$Y_i = n_i \exp(\beta_0 + \beta_1 X_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{Poisson}(Y_i) \quad (5.22)$$

について、誤差項を含まない推定式として両辺に対数を取ると、

$$\ln \hat{Y}_i = \ln n_i + \hat{\beta}_0 + \hat{\beta}_0 X_i \quad (5.23)$$

対数リンクの形になり、右辺にオフセット $\ln n_i$ を含む式が得られる。オフセット $\ln n_i$ を左辺に移動して整理すると

$$\ln \frac{\hat{Y}_i}{n_i} = \hat{\beta}_0 + \hat{\beta}_0 X_i \quad (5.24)$$

が得られる。これは、死亡率の対数について、回帰直線をあてはめることに相当する。もちろん、単純に死亡率の対数に対し、通常の回帰分析を行なう問題ではない。誤差の分布はポアソン分布を仮定しなければならない。式 (5.23) で考えれば、 $\ln n_i + \beta_0$ がそれぞれの年齢層 i についての切片となっている。このことから、傾きは共通で、切片に年齢層別にオフセットを持つ回帰直線を求める問題と解される。ただし、推定する切片は、対数死亡率に対する切片 β_0 であり、それぞれの年齢層 i について推定しているわけではなく、単にオフセットしているだけである。なお、オフセットとは、ある基準からのズレの意味で、 $\hat{\beta}_0$ を基準として $\ln n_i$ 分ずらし切片とすることを意味する。なお、オフセットを考慮した回帰直線のあてはめについては、第 2.6 節の図 2.3 を参照してもらいたい。

補正值

反復重み付き回帰で、オフセットがある対数リンク式を

$$\eta_i = \ln \left(\frac{\hat{Y}_i}{n_i} \right)$$

とし、元の推定値を

$$\mu_i = n_i \exp[\ln(\hat{Y}_i / n_i)]$$

$$\ln(\mu_i / n_i) = \ln(\hat{Y}_i / n_i)$$

としたときに、一般線形モデルの公式により、対する補正值 Z_i は

$$\begin{aligned} Z_i &= \eta_i + (Y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \\ &= \eta_i + (Y_i - \mu_i) \left(\frac{\partial \ln(\mu_i)}{\partial \mu_i} \right) \\ &= \eta_i + \frac{(Y_i - \mu_i)}{\mu_i} \\ &= \ln \left(\frac{\hat{Y}_i}{n_i} \right) + \frac{Y_i - n_i \exp[\ln(\hat{Y}_i / n_i)]}{n_i \exp[\ln(\hat{Y}_i / n_i)]} \end{aligned} \quad (5.25)$$

となる。式 (5.22) の誤差分布の分散は、 $\text{Var}(Y_i) = \mu_i$ なので、計算に用いる重み w_i は、

$$\begin{aligned}
w_i &= \frac{1}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \\
&= \frac{1}{\mu_i \left(\frac{\partial \ln \mu_i}{\partial \mu_i} \right)^2} \\
&= \frac{1}{\frac{\mu_i}{\mu_i^2}} = \mu_i = n_i \exp[\ln(\hat{Y}_i / n_i)]
\end{aligned} \tag{5.26}$$

となる.

反復計算

さて、元のデータに戻って、初期値を、表 5.17 から

$$\hat{\boldsymbol{\beta}}^{(0)} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \ln\left(\frac{\mathbf{Y}}{\mathbf{n}}\right) = \begin{bmatrix} -13.1819 \\ 0.1331 \end{bmatrix}$$

ただし、 $\ln\left(\frac{\mathbf{Y}}{\mathbf{n}}\right)$ は、セル同士の除算結果に対し対数を取ったベクトル

として求められる.

表 5.17 オフセットを考慮した初期値の計算

| | デザイン行列 | | 死亡 | 母集団 | 死亡率の | | | | |
|-----|----------------|----------------|--------------|--------------|------------------------------|---|----------|----------------------|---|
| | \mathbf{X} | | 者数 | 人数 | 対数 | | | | |
| i | \mathbf{X}_0 | \mathbf{X}_1 | \mathbf{Y} | \mathbf{n} | $\ln(\mathbf{Y}/\mathbf{n})$ | $\mathbf{X}^T \mathbf{X} =$ | 8.00 | 380.00 | |
| 1 | 1 | 30 | 1 | 17,742 | -9.7837 | | 380.00 | 19100.00 | |
| 2 | 1 | 35 | 5 | 16,554 | -8.1049 | | | | |
| 3 | 1 | 40 | 5 | 16,059 | -8.0746 | $(\mathbf{X}^T \mathbf{X})^{-1} =$ | 2.2738 | -0.0452 | |
| 4 | 1 | 45 | 12 | 13,083 | -6.9942 | | -0.0452 | 0.0010 | |
| 5 | 1 | 50 | 25 | 10,784 | -6.0669 | | | | $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\ln \mathbf{Y} / \mathbf{n})$ |
| 6 | 1 | 55 | 38 | 9,645 | -5.5366 | $\mathbf{X}^T \ln(\mathbf{Y}/\mathbf{n}) =$ | -54.8797 | $\beta_0^{\wedge} =$ | -13.1819 |
| 7 | 1 | 60 | 54 | 10,706 | -5.2896 | | -2467.04 | $\beta_1^{\wedge} =$ | 0.1331 |
| 8 | 1 | 65 | 65 | 9,933 | -5.0292 | | | | $\boldsymbol{\beta}^{(0)}$ |

この初期値を用いて表 5.18 に示すように反復重み付き回帰を行う. この結果は、3 反復目の結果で、まだ収束はしていないが、次の反復で収束する直前の結果である. なお、計算の手順は、第 5.4 節の対数リンクの場合の $\ln(\mathbf{Y})$ を $\ln(\mathbf{Y} / \mathbf{n})$ に変更しただけで、他の手順は全く同じである.

表 5.18 <第3 反復>冠動脈心疾患の死亡者数（対数リンク，オフセットあり）

| | デザイン行列 | | 死亡者数 | 母集団人数 | 死亡率の対数 | 回帰推定値 | | リンク関数 | 重付回帰推定値 | 推定値差 |
|-----|--------|---|----------|--------|-----------------------|----------------|----------------|---------------------|------------|---------------------------|
| | X | | | | | | | | | |
| i | X_0 | X_1 | Y | n | $\ln(Y/n)$ | $\ln Y^\wedge$ | $w = Y^\wedge$ | Z | Z^\wedge | $\ln Y^\wedge - Z^\wedge$ |
| 1 | 1 | 30 | 1 | 17,742 | -9.7837 | -8.4919 | 3.6393 | -9.2171 | -8.4947 | 0.0028 |
| 2 | 1 | 35 | 5 | 16,554 | -8.1049 | -7.9701 | 5.7220 | -8.0962 | -7.9725 | 0.0024 |
| 3 | 1 | 40 | 5 | 16,059 | -8.0746 | -7.4483 | 9.3537 | -7.9137 | -7.4503 | 0.0020 |
| 4 | 1 | 45 | 12 | 13,083 | -6.9942 | -6.9264 | 12.8408 | -6.9919 | -6.9281 | 0.0017 |
| 5 | 1 | 50 | 25 | 10,784 | -6.0669 | -6.4046 | 17.8356 | -6.0029 | -6.4059 | 0.0013 |
| 6 | 1 | 55 | 38 | 9,645 | -5.5366 | -5.8828 | 26.8802 | -5.4691 | -5.8837 | 0.0009 |
| 7 | 1 | 60 | 54 | 10,706 | -5.2896 | -5.3610 | 50.2783 | -5.2870 | -5.3615 | 0.0006 |
| 8 | 1 | 65 | 65 | 9,933 | -5.0292 | -4.8392 | 78.6061 | -5.0123 | -4.8394 | 0.0002 |
| | | $\beta_0^\wedge =$ | -13.1819 | | $\beta_0^\wedge =$ | -11.6228 | | $\beta_0^\wedge =$ | -11.6278 | 0.0119 |
| | | $\beta_1^\wedge =$ | 0.1331 | | $\beta_1^\wedge =$ | 0.1044 | | $\beta_1^\wedge =$ | 0.1044 | 絶対値の和 |
| | | $\hat{\beta}^{(0)} = (X^T X)^{-1} X^T \ln(Y/n)$ | | | $\beta^{(m-1)\wedge}$ | | | $\beta^{(m)\wedge}$ | | |
| | | | | | | 205.2 | 11757.7 | 0.2049 | -0.0035 | -1157.6 |
| | | | | | | 11757.7 | 690267.6 | -0.0035 | 0.0001 | -64626.5 |
| | | | | | | $(X^*w)^T X$ | | $[(X^*w)^T X]^{-1}$ | | $(X^*w)^T Z$ |

オフセットを用いた推定

この推定結果から，発現率に対する回帰式

$$\ln \frac{\hat{Y}_i}{n_i} = -11.6278 + 0.1044 X_{1,i}$$

が得られる．そこで，表 5.19 に示すように，年齢階層別の人数の推定値に換算しよう．年齢が 30 の場合の死亡率の推定値は，

$$\hat{Z}_1 = -11.6278 + 0.1044 \times 30 = -8.4947$$

となる．これは，死亡率 $p_1 = 1/17,742 = 0.000056$ に対する推定値 $\pi_1 = \exp(-8.4974) = 0.00020$ である．年齢階層別の人数の推定値は，

$$\begin{aligned} \hat{Y}_1 &= n_1 \exp(\hat{\beta}_0 + \hat{\beta}_1 X_{1,1}) \\ &= 17,742 \exp(-8.4947) = 3.63 \end{aligned}$$

となる．それぞれの階層での分母が異なるので，オフセットされた推定値が適切であるが，全体として解釈するためには適さない．そのために，分母を区切りの良い人数に揃えて結果を示すことにする．分母の人数は，9,647～17,742 人なので，10,000 人あたりに換算する．年齢階層 30 歳の得られた死亡数は，1 人なので，

$$Y'_1 = \frac{10,000}{n_i} Y_1 = \frac{10,000}{17,742} \times 1 = 0.56$$

0.56 人となり，10,000 人あたりの推定値は，

$$\hat{Y}'_1 = 10,000 \times \exp(-8.4947) = 2.05$$

と，2.05 人となる．この様に換算することにより，人口 10,000 人あたりの死亡数となり，指

数曲線での図が作成可能となる。オフセット付きの人数では、各層間の直接的な比較は不向きであり、またグラフ表示にもなじまない

表 5.19 一万人当たりの冠動脈心疾患の死亡者数の推定

| | デザイン行列 | | 死亡者数 | 母集団人数 | 回帰推定値 | 人数推定値 | 10,000人当たりの人数 | 10,000人当たりの推定人数 |
|-----|--------|-------|------|--------|--------------|--------------|---------------|-----------------|
| | X | | Y | n | Z^{\wedge} | Y^{\wedge} | | |
| i | X_0 | X_1 | | | | | | |
| 1 | 1 | 30 | 1 | 17,742 | -8.4947 | 3.63 | 0.56 | 2.05 |
| 2 | 1 | 35 | 5 | 16,554 | -7.9725 | 5.71 | 3.02 | 3.45 |
| 3 | 1 | 40 | 5 | 16,059 | -7.4503 | 9.33 | 3.11 | 5.81 |
| 4 | 1 | 45 | 12 | 13,083 | -6.9281 | 12.82 | 9.17 | 9.80 |
| 5 | 1 | 50 | 25 | 10,784 | -6.4059 | 17.81 | 23.18 | 16.52 |
| 6 | 1 | 55 | 38 | 9,645 | -5.8837 | 26.86 | 39.40 | 27.84 |
| 7 | 1 | 60 | 54 | 10,706 | -5.3615 | 50.25 | 50.44 | 46.94 |
| 8 | 1 | 65 | 65 | 9,933 | -4.8394 | 78.59 | 65.44 | 79.12 |

図 5.3 に 1 万人比の死亡者数に対して、対数目盛と実目盛りでのポアソン回帰の結果を示す。前節と同様に、各年齢のプロット点が推定値に対して上に凸であり、前節と同様に 2 次式のあてはめが必要となる。

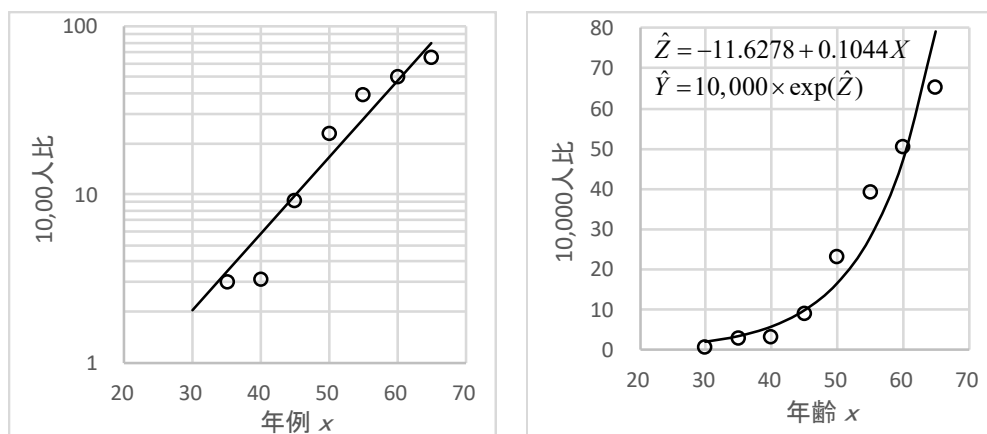


図 5.3 冠動脈心疾患の人口 10,000 人当たりの死亡者数にポアソン回帰

5.6. 2 項分布を仮定した(プロビット・補 2 重対数・ロジット)解析

オフセットがある対数リンクのポアソン回帰については、前節で解析事例を示した。オフセットがある場合には、観測されたカウント・データに対して部分母集団のサイズが分かっている場合は、ポアソン分布をしなくても、2 項分布を仮定した (0, 1) 反応に対するロジスティック回帰などの適用が妥当とも思われる。前節では、「死亡率が最も高い年齢 65 歳階層でも 1.0 パーセントにも届かないデータに、0~100 パーセントの範囲を仮定するシグモイド曲線のあてはめを行う統計モデルを適用することは、不適切である。」と述べた。

ポアソン回帰も一般化線形モデルの枠組みであるので、分布を 2 項分布とし、リンク関数を (プロビット or 補 2 重対数 or ロジット) による一般化線形モデル解析を適用した場合に結果をどのように解釈するのか、ポアソン回帰の結果と対比して示す。一般化線形モデルでのリンク関数をロジットとした場合には、独立した統計手法として確立しているロジスティック回帰と同じ結果を得るので、「プロビット」の場合を最初に解説する。

プロビット

2 項分布に従う出現率 $P_i = r_i / n_i$ について、説明変数を X_i とした場合に、標準正規分布の分布関数をシグモイド曲線として

$$P_i = \Phi(\beta_0 + \beta_1 X_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{Binomial}(r_i, n_i; \pi_i) \quad (5.27)$$

ただし、 Φ は、標準正規分布

をあてはめたい。ここで、

$$\beta_0 = \frac{-\mu}{\sigma}, \quad \beta_1 = \frac{1}{\sigma}$$

で置き換えると、

$$\begin{aligned} \Phi(\beta_0 + \beta_1 X_i) &= \Phi\left(\frac{-\mu}{\sigma} + \frac{1}{\sigma} X_i\right) \\ &= \Phi\left(\frac{X_i - \mu}{\sigma}\right) \end{aligned}$$

となり、線形式のパラメータから、標準正規分布への変換パラメータは、

$$\mu = \frac{-\beta_0}{\beta_1}, \quad \sigma = \frac{1}{\beta_1}$$

として与えられる。

リンク関数の「プロビット」は、標準正規分布の逆関数 Φ^{-1} を用いて

$$Z_i = \Phi^{-1}(\hat{P}_i) = \hat{\beta}_0 + \hat{\beta}_1 X_i \quad (5.28)$$

ただし、 Φ^{-1} ：逆標準正規分布

で線形化し、反復重み付き回帰で解くので、線形化した場合の分布を気にする必要がなくなる。この方法は、生物統計の分野で 50%致死量 LD_{50} を求めるための方法として Finney (1971), Probit Analysis 3rd ed. および Finney (1978), Statistical Method in Biological Assay 3rd ed. により知られている方法でもあり、詳細は、高橋 (2017) を参照のこと。

ここでは、プロビット変換した式 (5.28) ではなく、式 (5.27) によりシグモイド曲線の直接あてはめを行う。表 5.20 に示すように冠動脈心疾患の死亡率に対して、年齢が 30 歳 ($X_{1,1}=30$) 正規分布の下側確率を

$$\begin{aligned} \hat{\pi}_1 &= \Phi(\hat{\beta}_0 + \hat{\beta}_1 X_{1,1}) \\ &= \Phi(-4.6119 + 0.0337 \times 30) \\ &= \text{Norm.dist}(-3.5999, 0, 1, \text{true}) = 0.000159 \end{aligned}$$

で求めている。この予測死亡率 π_1 に対し、2 項分布の確率を「尤度」として

$$\begin{aligned} L_1 &= \text{Binom.dist}(Y_1, n_1, \hat{\pi}_1, \text{false}) \\ &= \text{Binom.dist}(1, 17742, 0.000159, \text{false}) \\ &= 0.1676 \end{aligned}$$

で求める。その対数尤度の和は、

$$\begin{aligned} \ln L &= \sum_{i=1}^8 \ln L_i \\ &= \ln(0.1676) + \ln(0.1755) + \dots + \ln(0.0176) \\ &= -1.7861 - 1.7402 - \dots - 4.0383 \\ &= -23.2990 \end{aligned}$$

表 5.20 冠動脈心疾患の死亡率に対する正規分布曲線のあてはめ

| | デザイン行列 | | 死亡者数 | 母集団人数 | 死亡率 | | シグモイド | | | 組合せ |
|-----|--------|-------|------|--------|----------------------|-------------------|----------------|--------------|-----------|------------|
| | X | | | | Y/n | $X\beta^{\wedge}$ | 標準正規 | 二項分布 | | 無 |
| i | X_0 | X_1 | Y | n | P | Z^{\wedge} | π^{\wedge} | 尤度 L_i | $\ln L_i$ | $\ln L_i'$ |
| 1 | 1 | 30 | 1 | 17,742 | 0.0001 | -3.5999 | 0.000159 | 0.1676 | -1.7861 | -11.57 |
| 2 | 1 | 35 | 5 | 16,554 | 0.0003 | -3.4312 | 0.000300 | 0.1755 | -1.7402 | -45.52 |
| 3 | 1 | 40 | 5 | 16,059 | 0.0003 | -3.2625 | 0.000552 | 0.0644 | -2.7426 | -46.37 |
| 4 | 1 | 45 | 12 | 13,083 | 0.0009 | -3.0939 | 0.000988 | 0.1106 | -2.2017 | -95.96 |
| 5 | 1 | 50 | 25 | 10,784 | 0.0023 | -2.9252 | 0.001721 | 0.0290 | -3.5392 | -177.65 |
| 6 | 1 | 55 | 38 | 9,645 | 0.0039 | -2.7566 | 0.002921 | 0.0137 | -4.2875 | -249.87 |
| 7 | 1 | 60 | 54 | 10,706 | 0.0050 | -2.5879 | 0.004828 | 0.0516 | -2.9635 | -339.55 |
| 8 | 1 | 65 | 65 | 9,933 | 0.0065 | -2.4192 | 0.007777 | 0.0176 | -4.0383 | -392.72 |
| | | | | | $\beta_0^{\wedge} =$ | -4.6119 | | | -23.2990 | -1,359.22 |
| | | | | | $\beta_1^{\wedge} =$ | 0.0337 | | 対数尤度 $\ln L$ | | $\ln L'$ |

となっている。これは、ソルバーで $\ln L$ が最大になるように $\hat{\beta}_0$ と $\hat{\beta}_1$ を変化させた結果である。ここに示した対数尤度は、2 項分布の確率の対数の和であるが、統計ソフトでは、2 項確率の組み合わせ計算を除いている場合があるので、 $\ln L'$

$$\begin{aligned}\ln L' &= \sum_{i=1}^8 [y_i \ln \hat{\pi}_i + (n_i - y_i) \ln(1 - \hat{\pi}_i)] \\ &= -11.57 - 45.52 - \dots - 392.72 \\ &= -1,359.22\end{aligned}$$

の計算も付け足してある。この $\ln L' = -1,359.22$ は、表 5.21 に示す JMP によるプロビット解析の「モデル」の欄の「完全」、「(-1)*対数尤度」の欄の「1359.2183」に一致する。

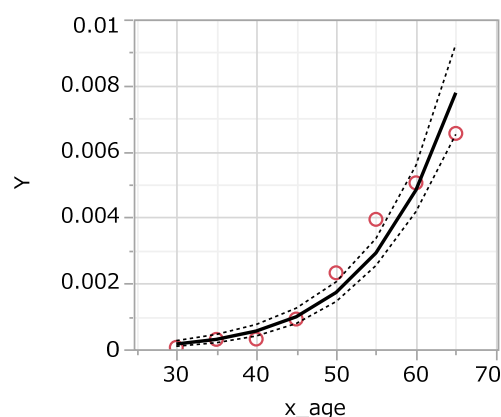
JMP によるプロビット解析は、データを死亡の (0 : あり, 1 : なし) を行方向に展開し、それぞれの人数を設定し、一般線形モデルで次のように分布とリンク関数を設定する。

| | |
|-------|----------|
| 手法: | 一般化線形モデル |
| 分布: | 二項 |
| リンク関数 | プロビット |

表 5.21 に結果を示す。Excel で求めた結果と一致することが確かめられる。95%信頼区間の図は、推定値を JMP ファイルに別途書き出して、重ね合わせプロットで作成したものである。

表 5.21 JMP によるプロビット解析

| モデル | (-1)*対数尤度 | 尤度比カイ2乗 | |
|-------|-----------|----------|-----------|
| 差分 | 123.5484 | 247.0968 | |
| 完全 | 1359.2183 | | |
| 縮小 | 1482.7667 | | |
| 項 | 推定値 | 標準誤差 | 尤度比カイ2乗 |
| 切片 | -4.6119 | 0.1402 | 2580.1056 |
| x_age | 0.0337 | 0.0025 | 247.0968 |



標準正規分布に変換するためのパラメータは、

$$\hat{\mu} = \frac{-\hat{\beta}_0}{\hat{\beta}_1} = \frac{-(-4.6119)}{0.0337} = 136.72, \quad \hat{\sigma} = \frac{1}{\hat{\beta}_1} = \frac{1}{0.0337} = 29.64$$

なので、形式的に平均が 136.72 歳、標準偏差が 29.64 歳とした正規分布をあてはめていることになる。

補 2 重対数

2 項分布に従う出現率 $P_i = r_i / n_i$ について，説明変数を X_i とした場合に，最小極値分布をシグモイド曲線として

$$P_i = 1 - \exp[-\exp(\beta_0 + \beta_1 X_i)] + \varepsilon_i, \quad \varepsilon_i \sim \text{Binomial}(r_i, n_i; \pi_i) \quad (5.29)$$

用いる．式 (5.29) の推定値にした回帰式について解くと

$$\begin{aligned} \exp[-\exp(\hat{\beta}_0 + \hat{\beta}_1 X_i)] &= 1 - \hat{P}_i \\ \exp(\hat{\beta}_0 + \hat{\beta}_1 X_i) &= -\ln(1 - \hat{P}_i) \\ Z_i = \ln[-\ln(1 - \hat{P}_i)] &= \hat{\beta}_0 + \hat{\beta}_1 X_i \end{aligned} \quad (5.30)$$

となり，標準的な関数で変換式が表せることから一般的に使われている．プロビット法と同様に，リンク関数が「補 2 重対数」の形式になっている．

ここでは，補 2 重対数変換した式 (5.30) ではなく，式 (5.29) によるシグモイド曲線の直接あてはめを行う．表 5.22 に対数尤度 $\ln L_i$ が計算できるような適当な $\hat{\beta}_0$ と $\hat{\beta}_1$ を設定し，Excel のソルバーで，対数尤度 $\ln L$ を最大化するように $\hat{\beta}_0$ と $\hat{\beta}_1$ を変化させた結果を示す．

表 5.22 冠動脈心疾患の死亡率に対する最小極値分布のあてはめ

| | デザイン行列 | | 死亡 | 母集団 | 死亡率 | | シグモイド | | | 組合せ |
|-----|--------|-------|-----|--------|----------------------|-------------------|----------------|--------------|-----------|------------|
| | X | | 者数 | 人数 | Y/n | $X\beta^{\wedge}$ | 最小極値 | 二項分布 | | 無 |
| i | X_0 | X_1 | Y | n | P | Z^{\wedge} | π^{\wedge} | 尤度 L_i | $\ln L_i$ | $\ln L_i'$ |
| 1 | 1 | 30 | 1 | 17,742 | 0.0001 | -8.4956 | 0.000204 | 0.0966 | -2.3376 | -12.12 |
| 2 | 1 | 35 | 5 | 16,554 | 0.0003 | -7.9729 | 0.000345 | 0.1677 | -1.7857 | -45.57 |
| 3 | 1 | 40 | 5 | 16,059 | 0.0003 | -7.4501 | 0.000581 | 0.0522 | -2.9536 | -46.59 |
| 4 | 1 | 45 | 12 | 13,083 | 0.0009 | -6.9273 | 0.000980 | 0.1114 | -2.1949 | -95.95 |
| 5 | 1 | 50 | 25 | 10,784 | 0.0023 | -6.4045 | 0.001653 | 0.0220 | -3.8157 | -177.93 |
| 6 | 1 | 55 | 38 | 9,645 | 0.0039 | -5.8817 | 0.002786 | 0.0084 | -4.7837 | -250.36 |
| 7 | 1 | 60 | 54 | 10,706 | 0.0050 | -5.3590 | 0.004695 | 0.0474 | -3.0488 | -339.64 |
| 8 | 1 | 65 | 65 | 9,933 | 0.0065 | -4.8362 | 0.007906 | 0.0142 | -4.2522 | -392.93 |
| | | | | | $\beta_0^{\wedge} =$ | -11.6323 | | | -25.1721 | -1,361.09 |
| | | | | | $\beta_1^{\wedge} =$ | 0.1046 | | 対数尤度 $\ln L$ | | $\ln L'$ |

プロビットの場合とほぼ同様なパラメータの推定値

$$\hat{\beta}_0 = -11.6323, \quad \hat{\beta}_1 = 0.1046$$

が得られている．

$$\hat{\mu}_{SEV} = \frac{-\hat{\beta}_0}{\hat{\beta}_1} = \frac{-(-11.6323)}{0.1046} = 111.25, \quad \hat{\sigma}_{SEV} = \frac{1}{\hat{\beta}_1} = \frac{1}{0.1046} = 9.56$$

従って，位置パラメータが 111.25 歳，形状パラメータが 9.56 歳の最小極値分布をあてはめたことになる．

JMP による補 2 重対数による解析結果を表 5.23 に示す。Excel での結果に一致することが確認される。

表 5.23 JMP による補 2 重対数解析

| モデル | (-1)*対数尤度 | 尤度比カイ2乗 | 項 | 推定値 | 標準誤差 | 尤度比カイ2乗 |
|-----|-----------|----------|-------|----------|--------|-----------|
| 差分 | 121.6753 | 243.3506 | 切片 | -11.6323 | 0.4531 | 1501.6613 |
| 完全 | 1361.0914 | | x_age | 0.1046 | 0.0078 | 243.3506 |
| 縮小 | 1482.7667 | | | | | |

ロジット

2 項分布に従う出現率 $P_i = r_i / n_i$ について、説明変数を X_i とした場合に、ロジスティック分布をシグモイド曲線として

$$P_i = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} + \varepsilon_i, \quad \varepsilon_i \sim \text{Binomial}(r_i, n_i; \pi_i) \quad (5.31)$$

用いる。リンク関数の「ロジット」は、式 (5.31) を推定式にした回帰式について解くと

$$\begin{aligned} \hat{P}_i + \hat{P}_i \exp(\hat{\beta}_0 + \hat{\beta}_1 X_i) &= \exp(\hat{\beta}_0 + \hat{\beta}_1 X_i) \\ \hat{P}_i &= (1 - \hat{P}_i) \exp(\hat{\beta}_0 + \hat{\beta}_1 X) \\ Z_i = \ln \left(\frac{\hat{P}_i}{1 - \hat{P}_i} \right) &= \hat{\beta}_0 + \hat{\beta}_1 X_i \end{aligned} \quad (5.32)$$

となり、標準的な関数で表わされることから広く使われている。

表 5.24 冠動脈心疾患の死亡率に対するロジスティック曲線のあてはめ

| | デザイン行列 | | 死亡 | 母集団 | 死亡率 | ロジスティック | | | 組合せ | |
|-----|--------|-------|-----|--------|----------------------|----------|----------------|----------|--------------|------------|
| | X | | 者数 | 人数 | Y/n | $X\beta$ | 曲線 | 二項分布 | 無 | |
| i | X_0 | X_1 | Y | n | p | z | π^{\wedge} | 尤度 L_i | $\ln L_i$ | $\ln L_i'$ |
| 1 | 1 | 30 | 1 | 17,742 | 0.0001 | -8.4978 | 0.000204 | 0.0971 | -2.3317 | -12.12 |
| 2 | 1 | 35 | 5 | 16,554 | 0.0003 | -7.9741 | 0.000344 | 0.1679 | -1.7846 | -45.57 |
| 3 | 1 | 40 | 5 | 16,059 | 0.0003 | -7.4505 | 0.000581 | 0.0523 | -2.9505 | -46.58 |
| 4 | 1 | 45 | 12 | 13,083 | 0.0009 | -6.9269 | 0.000980 | 0.1114 | -2.1949 | -95.95 |
| 5 | 1 | 50 | 25 | 10,784 | 0.0023 | -6.4032 | 0.001653 | 0.0221 | -3.8124 | -177.93 |
| 6 | 1 | 55 | 38 | 9,645 | 0.0039 | -5.8796 | 0.002788 | 0.0084 | -4.7755 | -250.35 |
| 7 | 1 | 60 | 54 | 10,706 | 0.0050 | -5.3560 | 0.004698 | 0.0475 | -3.0464 | -339.63 |
| 8 | 1 | 65 | 65 | 9,933 | 0.0065 | -4.8323 | 0.007905 | 0.0143 | -4.2504 | -392.93 |
| | | | | | $\beta_0^{\wedge} =$ | -11.6395 | | | -25.1465 | -1,361.07 |
| | | | | | $\beta_1^{\wedge} =$ | 0.1047 | | | 対数尤度 $\ln L$ | $\ln L'$ |

ここでは、ロジット変換した 式 (5.32) ではなく、式 (5.31) によるシグモイド曲線の直接あてはめを行う。表 5.22 に対数尤度 $\ln L_i$ が計算できるような適当な $\hat{\beta}_0$ と $\hat{\beta}_1$ を設定し、Excel のソルバーで、対数尤度 $\ln L$ を最大化するように $\hat{\beta}_0$ と $\hat{\beta}_1$ を変化させた結果を示す。

補 2 重対数の場合と同様なパラメータの推定値

$$\hat{\beta}_0 = -11.6393, \quad \hat{\beta}_1 = 0.1047$$

が得られている。

$$\hat{\mu}_{LGT} = \frac{-\hat{\beta}_0}{\hat{\beta}_1} = \frac{-(-11.6323)}{0.1046} = 111.14, \quad \hat{\sigma}_{LGT} = \frac{1}{\hat{\beta}_1} = \frac{1}{0.1047} = 9.55$$

従って、位置パラメータが 111.14 歳、形状パラメータが 9.55 歳のロジスティック分布をあてはめたことになる。

JMP によるロジットによる解析結果を表 5.26 に示す。Excel での結果に一致することが確認される。

表 5.25 JMP によるロジット解析

| モデル | (-1)*対数尤度 | 尤度比カイ2乗 | 項 | 推定値 | 標準誤差 | 尤度比カイ2乗 |
|-----|-----------|----------|-------|----------|--------|-----------|
| 差分 | 121.7009 | 243.4019 | 切片 | -11.6395 | 0.4538 | 1499.0600 |
| 完全 | 1361.0658 | | x_age | 0.1047 | 0.0078 | 243.4019 |
| 縮小 | 1482.7667 | | | | | |

ポアソン・プロビット・補 2 重対数・ロジット

対数リンクでオフセットがある場合のポアソン回帰に対比して、2 値データに対する誤差分布を 2 項分布、リンク関数を（プロビット・補 2 重対数・ロジット）とした解析結果を示してきた。いずれの方法でも同様の結果が得られるのであるが、推定されたパラメータを相互に比較する。表 5.26 に比較結果を示すが、プロビットを除いて、他の方法はほとんど同様な値となっていることをあらためて認識させられる。

（プロビット・補 2 重対数・ロジット）は、どれも 0～1 の範囲のシグモイド曲線としての分布関数のあてはめであり、単調増加を前提にしている。第 5.4 節で取り上げたように、反応に頭打ちがあるような場合に、2 次式に対するポアソン回帰を行った。シグモイド曲線のあてはめを前提にする解析に対して、シグモイド曲線を否定するような 2 次式の適用は受け入れがたい。したがって、稀な現象であるが母集団のサイズが分かっているような場合については、オフセット付きの対数リンクのポアソン回帰が標準的な方法として薦められる。

なお、補2重対数の最小極値分布は、左に長く裾を引く分布であり、その位置パラメータは、累積分布の0.623点となり、正規分布とロジスティック分布の0.50点と異なるが、ここでの事例は、発現率が1%未満であり、3種の方法にほとんど違いがない。

表 5.26 JMP によるポアソン・プロビット・補2重対数・ロジット解析

| | ポアソン | プロビット | 補2重対数 | ロジット |
|------------------|----------|---------|----------|----------|
| β_0^\wedge | -11.6278 | -4.6119 | -11.6323 | -11.6395 |
| β_1^\wedge | 0.1044 | 0.0337 | 0.1046 | 0.1047 |

ポアソン回帰のみならず、(0, 1) データに対する一般化線形モデルに対して、統計ソフトの結果と Excel のソルバーを用いた結果を確認することは、統計モデルに対する知識を確実なものにすることが期待される。

上限があるシグモイド曲線のあてはめ

シグモイド曲線のあてはめの基本は、反応が0%から100%の範囲である。ただし、現実の用量反応関係では、低用量の場合に自然反応などがあり0%に収束しない場合も良く知られている。また、特定の疾患の死亡率では、本質的に100%になりえない場合もあり、シグモイド曲線の上限を得られたデータから推測することが現実的である。第2.6節では、「死亡率の上限を新たな変数とするロジスティック回帰」について例示したので参考にしてもらいたい。同様な方法で、上限をパラメータとして加えたプロビット曲線を図5.4に示す。このようなシグモイド曲線が求められる条件は、低年齢層のデータおよび高年齢層の死亡率に頭打ち現象がみられ、全体としてシグモイド曲線らしき形状となっていることが必要である。

$$\hat{\pi} = 0.008966 \times \Phi(-5.2153 + 0.0901x)$$

$$\hat{\mu} = \frac{-\hat{\beta}_0}{\hat{\beta}_1} = \frac{-(-5.2153)}{0.0901} = 57.85, \quad \hat{\sigma} = \frac{1}{\hat{\beta}_1} = \frac{1}{0.0901} = 11.09$$

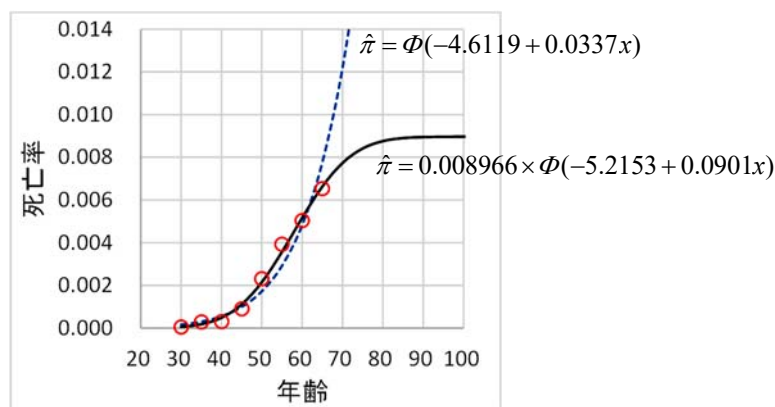


図 5.4 上限を持つプロビット曲線のあてはめ

第5章 文 献 索 引

| | |
|---|-----|
| 高橋(2017) - 一般化線形モデルをExcelで極め活用するープロビット法・ロジット法・補2重対数法ー | 176 |
| Finney(1971) - Probit Analysis 3rd ed. | 201 |
| Finney(1978) - Statistical Method in Biological Assay 3rd ed. | 201 |

第5章 索 引

| | | | | | |
|---|--------------------------------|-----|---|--------------------------------|----------|
| あ | 頭打ち現象 - 高年齢層 | 206 | さ | 死亡率 - 冠動脈心疾患 | 195 |
| | 10,000人あたり - オフセット | 198 | | - シグモイド曲線 | 195 |
| | 1万人比 - オフセット | 199 | | - 対数 | 196 |
| | 一般線形モデル - 重み | 188 | | - 2項分布 | 195 |
| | Excel - セル同士の積「*」 | 181 | | - 正規分布曲線 | 201 |
| | オフセット - 10,000人あたり | 198 | | JMP - プロビット解析 | 202 |
| | - 1万人比 | 199 | | - 補2重対数 | 204 |
| | - 重み | 196 | | - ロジット | 205 |
| | - 基準からのズレ | 196 | | JMP/一般線形モデル - SAS/GENMODE | 185 |
| | - 推定 | 198 | | JMPファイル - 重ね合わせプロット | 192 |
| | - 対数リンク | 195 | | 上限 - シグモイド曲線 | 206 |
| | - ドブソン(2008) | 195 | | 上限をパラメータ - プロビット曲線 | 206 |
| | - 反復計算 | 197 | | 初期パラメータ - 重みなしの回帰 | 180 |
| | - ポアソン回帰 | 195 | | 初期値 - ドブソン(2008) | 180 |
| | - 補正值 | 196 | | - 反復 | 180 |
| | 重み - 一般線形モデル | 188 | | 正規分布曲線 - 死亡率 | 201 |
| | - オフセット | 196 | | 正規方程式 - 重み付き回帰 | 177 |
| | - 行列計算 | 181 | | 生物統計 - 50%致死量 | 201 |
| | 重みなしの回帰 - 初期パラメータ | 180 | | 切片 - 年齢層別オフセット | 196 |
| | 重みの行列 - デザイン行列 | 178 | | セル同士の積「*」 - Excel | 181 |
| | 重み付き回帰 - 正規方程式 | 177 | | ゼロの値 - げた | 176 |
| | - 反復 | 182 | | ソルバー - 対数尤度 | 202 |
| | - ブレ | 175 | た | 対数 - 死亡率 | 196 |
| | - 厄介な問題 | 175 | | 対数ポアソン分布? - ポアソン分布 | 176 |
| | 重み付き平方和 - 偏微分 | 177 | | 対数リンク - オフセット | 195 |
| か | 回帰パラメータ - ワールド検定 | 184 | | - 95%信頼区間 | 191 |
| | 回帰曲線 - 95%信頼区間 | 192 | | - 2次式 | 192 |
| | 重ね合わせプロット - JMPファイル | 192 | | - ポアソン回帰 | 186 |
| | 冠動脈心疾患 - 死亡率 | 195 | | 対数尤度 - ソルバー | 202 |
| | - ドブソン(2008) | 186 | | 高橋(2017) - プロビット法 | 176, 201 |
| | 基準からのズレ - オフセット | 196 | | 調整 - 補正式 | 188 |
| | 逆標準正規分布 - 標準正規分布 | 201 | | 直接あてはめ - シグモイド曲線 | 201 |
| | 95%信頼区間 - 対数リンク | 191 | | デザイン行列 - 重みの行列 | 178 |
| | - 2次回帰 | 194 | | データ変換 - 補正式 | 187 |
| | 共分散行列 - 2次形式 | 191 | | ドブソン(2008) - オフセット | 195 |
| | - ヘッセ行列 | 184 | | - 冠動脈心疾患 | 186 |
| | 共分散行列 $\Sigma(\beta')$ - パラメータ | 184 | | - 初期値 | 180 |
| | 行列計算 - 重み | 181 | な | 2項分布 - 死亡率 | 195 |
| | げた - ゼロの値 | 176 | | - プロビット | 200 |
| | 恒等 - リンク関数 | 181 | | - 補2重対数 | 200, 203 |
| | 恒等リンク - ポアソン回帰 | 177 | | - ロジット | 195, 200 |
| | 高年齢層 - 頭打ち現象 | 206 | | 2項分布の確率 - 尤度 | 201 |
| | 50%致死量 - 生物統計 | 201 | | 2次回帰 - 95%信頼区間 | 194 |
| さ | 最小極値分布 - シグモイド曲線 | 203 | | - ポアソン回帰 | 194 |
| | SAS/GENMODE - JMP/一般線形モデル | 185 | | 2次形式 - 共分散行列 | 191 |
| | シグモイド曲線 - 最小極値分布 | 203 | | 2次式 - 対数リンク | 192 |
| | - 死亡率 | 195 | | 2次式のあてはめ - 便宜的な方法 | 193 |
| | - 上限 | 206 | | ニュートン・ラフソン法 - 反復重み付き | 175 |
| | - 直接あてはめ | 201 | | 年齢層別オフセット - 切片 | 196 |
| | - 標準正規分布 | 200 | は | パラメータ - 共分散行列 $\Sigma(\beta')$ | 184 |

| | |
|---------------------------|----------|
| は 反復 - 重み付き回帰 | 182 |
| - 初期値 | 180 |
| 反復計算 - オフセット | 197 |
| 反復重み付き - ニュートン・ラフソン法 | 175 |
| 反復重み付き回帰 - 利便性 | 192 |
| 標準正規分布 - 逆標準正規分布 | 201 |
| - シグモイド曲線 | 200 |
| Finney (1971) - プロビット法 | 201 |
| Finney (1978) - プロビット法 | 201 |
| ブレ - 重み付き回帰 | 175 |
| プロビット - 2項分布 | 200 |
| プロビット法 - 高橋 (2017) | 176, 201 |
| - Finney (1971) | 201 |
| - Finney (1978) | 201 |
| プロビット解析 - JMP | 202 |
| プロビット曲線 - 上限をパラメータ | 206 |
| ベクトル w^* - マトリックス W^* | 180 |
| ヘッセ行列 - 共分散行列 | 184 |
| 偏微分 - 重み付き平方和 | 177 |
| 変換不能 - リンク関数 | 187 |
| 便宜的な方法 - 2次式のあてはめ | 193 |
| ポアソン回帰 - オフセット | 195 |
| - 恒等リンク | 177 |
| - 対数リンク | 186 |
| - 2次回帰 | 194 |
| ポアソン分布 - 対数ポアソン分布？ | 176 |
| 補正式 - 調整 | 188 |
| - データ変換 | 187 |
| 補正值 - オフセット | 196 |
| 補2重対数 - JMP | 204 |
| - 2項分布 | 200, 203 |
| - リンク関数 | 203 |
| や 厄介な問題 - 重み付き回帰 | 175 |
| 尤度 - 2項分布の確率 | 201 |
| 尤度比検定 - Wald検定 | 185 |
| ら 利便性 - 反復重み付き回帰 | 192 |
| リンク関数 - 恒等 | 181 |
| - 変換不能 | 187 |
| - 補2重対数 | 203 |
| - ロジット | 204 |
| ロジスティック分布 - シグモイド曲線 | 204 |
| ロジット - JMP | 205 |
| - 2項分布 | 195, 200 |
| - リンク関数 | 204 |
| わ Wald検定 - 尤度比検定 | 185 |
| ワルド検定 - 回帰パラメータ | 184 |

第5章 解析用ファイル一覧

| | | | |
|---|-------|----------------|------------------------|
|  | 22 KB | 第5章01_対数リンク | Microsoft Excel ワークシート |
|  | 18 KB | 第5章02_重み_計算式 | Microsoft Excel ワークシート |
|  | 32 KB | 第5章03_重み_恒等 | Microsoft Excel ワークシート |
|  | 6 KB | 第5章04_重み_対数リンク | JMP Data Table |
|  | 55 KB | 第5章04_重み_対数リンク | Microsoft Excel ワークシート |
|  | 7 KB | 第5章04b_重み_2次式 | JMP Data Table |
|  | 42 KB | 第5章04b_重み_2次式 | Microsoft Excel ワークシート |
|  | 6 KB | 第5章05_オフセット | JMP Data Table |
|  | 43 KB | 第5章05_オフセット | Microsoft Excel ワークシート |
|  | 9 KB | 第5章05b_オフセット2次 | JMP Data Table |
|  | 6 KB | 第5章06_プロビット | JMP Data Table |
|  | 36 KB | 第5章06_プロビット | Microsoft Excel ワークシート |

非売品, 無断複製を禁ずる

第 9 回 続高橋セミナー

最尤法によるポアソン回帰分析入門 <<第 5 章>>

第 5 章 反復重み付き最尤法によるポアソン回帰

BioStat 研究所(株)

〒105-0014 東京都 港区 芝 1-12-3 の 1005

2020 年 6 月 6 日 高橋 行雄