# 一般化線形モデルを Excelで極め活用する

BioStat研究所（株）
高橋 行雄

# プロビット変換による$LD_{50}$

- 二値の用量反応の実験データからプロビット変換を使って50%致死量$LD_{50}$と95%信頼区間を求めています．反復重み付き回帰による最尤法で行なっています．
- ロジット変換の場合の計算方法を教えてもらえませんか？
- 低用量から反応が出る場合の$LD_{10}$の推定には，補2重対数変換がいいと言われています．どのように計算したらよいのですか？

# プロビット・ロジット・補2重対数

- $LD_{50}$の信頼区間の計算に，佐久間（1977），薬効評価－計画と解析 I，p335 に示されている Fiellerの式を用いて計算しています．
- ロジット変換を用いた場合にも計算できるようにしたいのです．どうしたらよいのでしょうか？
- 補２重対数変換（最小極値分布）による $LD_{10}$ の推定と95%信頼区間の場合も教えて頂けませんか？

# さて困った

- SAS/PROBITを使えば全て計算ができるのだが，持っていないと言う.
- ロジット変換の場合であれば，JMPのロジスティック回帰の逆推定を用いて計算はできるが，プロビット変換と補２重対数変換については，どうしたらよいのだろうか.
- 田中ら訳(2008)「一般化線形モデル入門」では，どのような説明があるのだろうか.

# 「一般化線形モデル入門」

- 反復重み付き回帰による最尤法が丁寧に書いてある．例示はポアソン回帰の場合だけで，プロビット・ロジット・補２重対数の例示はない．
- 反復に際しての重みの行列 $W$ の計算式，変換値 $z$ の計算式が示されているので，これを使えば何とかなりそうだ．

# Excel による反復重み付き回帰

- 吉村功編著(1987), 毒性・薬効データの統計解析:226-32, に示されている例でプロビット変換による反復重み付き回帰をExcel に実行してみよう.
- 残念ながら Excel には重み付き回帰の関数がないので, 行列関数を使うしかない.
- 重み$W$は行列の対角要素なので, 扱いにくいので, 縦ベクトル$w$とし, $X$の列ごとに$w$を掛けて $X^T W = (X * w)^T$ として計算しよう.

# プロビット変換での補正値z と重みw

プロビット変換

リンク関数 $$g(\mu_i)=\Phi^{-1}(\mu_i)+5=\boldsymbol{x}^T\boldsymbol{\beta}=\eta_i$$

補正値 $z_i$ $$z_i=\eta_i+(y_i-\mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right)=\eta_i+\frac{(y_i-\mu_i)}{\phi(\eta_i-5)}$$

重み $w_{ii}$ $$w_{ii}=\frac{1}{\mathrm{var}(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2=\frac{n_i\left[\phi(\eta_i-5)\right]^2}{\mu_i(1-\mu_i)}$$

# Excel による反復重み付き回帰

| | 用量 | 切片 | 対数 *dose* | 出現数 | 例数 | 出現率 | probit 変換 | | 回帰推定値 | 出現率推定値 | 補正 probit | 重み | 更新推定値 | 10) 差 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | *dose* | $x_0$ | $x_1$ | $r$ | $n$ | $y$ | $z^*$ | | 3) $\eta^{(1)}$ | $\mu$ | 4) $z$ | 5) $w$ | 9) $\eta^{(2)}$ | $\eta^{(1)}$-$\eta^{(2)}$ |
| 1 | 101 | 1 | 2.004 | 0 | 10 | 0.000 | - | | 3.522 | 0.070 | 3.001 | 2.762 | 3.253 | 0.270 |
| 2 | 136 | 1 | 2.134 | 2 | 10 | 0.200 | 4.158 | | 4.241 | 0.224 | 4.161 | 5.146 | 4.081 | 0.160 |
| 3 | 183 | 1 | 2.262 | 5 | 10 | 0.500 | 5.000 | | 4.957 | 0.483 | 5.000 | 6.362 | 4.907 | 0.050 |
| 4 | 247 | 1 | 2.393 | 8 | 10 | 0.800 | 5.842 | | 5.681 | 0.752 | 5.833 | 5.367 | 5.743 | -0.061 |
| 5 | 333 | 1 | 2.522 | 9 | 10 | 0.900 | 6.282 | | 6.403 | 0.920 | 6.271 | 3.011 | 6.574 | -0.172 |
| 6 | 450 | 1 | 2.653 | 10 | 10 | 1.000 | - | | 7.130 | 0.983 | 7.531 | 1.045 | 7.413 | -0.283 |
| | | | **X** | | | 1) Intercept()= | -7.621 | $\beta_0$^= | -7.621 | | | 8) 更新 $\beta_0$^ = | -9.598 | 2.1E-01 |
| | | | デザイン行列 | | | Slope()= | 5.559 | $\beta_1$^= | 5.559 | | | $\beta_1$^ = | 6.411 | 平方和 |
| | | | | | | 回帰係数 初期値 | | | 2) 係数・貼付 | | | $(X^TWX)^{-1}X^TWz$ | | |
| | | | | | 6) | 23.69 | 54.12 | 6) | 7.363 | -3.205 | 7) | 119.58 | | |
| | | | | | | 54.12 | 124.34 | | -3.205 | 1.403 | | 277.70 | | |
| | | | | | | $X^TWX$ | | | $(X^TWX)^{-1}$ | | | $X^TWz$ | | |

**表1.　初期値を貼付け，更新パラメータの貼付けを繰り返す．5サイクル目で表示桁数で一致する．**

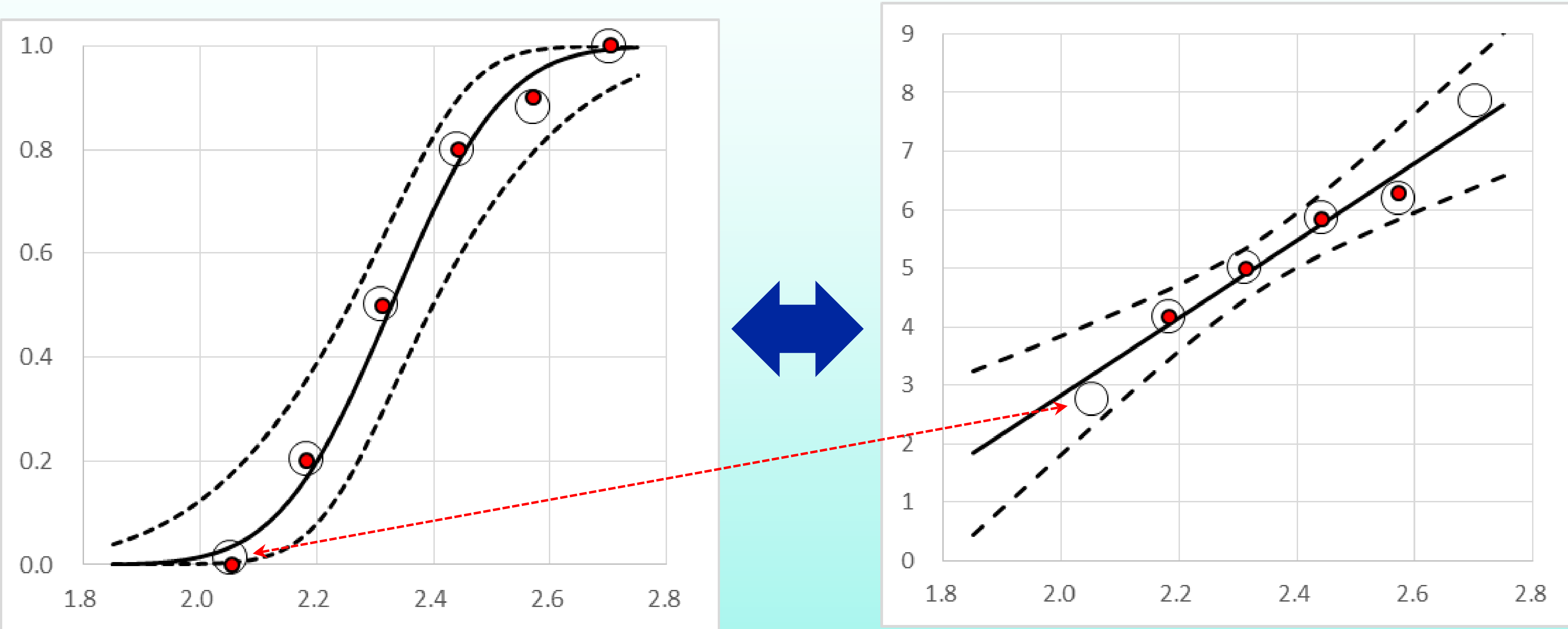
# 95%信頼区間

- プロビット変換に対する回帰パラメータが推定できたので，回帰直線の推定値の分散をデルタ法で求め，95%信頼区間を計算しグラフ上に描く.

| | | | | | | |
|---|---|---|---|---|---|---|
| var($x_i$) = | 1 | | 10.321 | -4.498 | 1 | $x_i$ |
| | $x_i$ | | -4.498 | 1.970 | | |
| | $\boldsymbol{x}^T$ | | $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ | | $\boldsymbol{x}$ | |

- プロビット変換の逆変換で，シグモイド曲線に対する95%信頼区間も描く.

# シグモイド曲線 vs プロビット変換




● は観測された死亡率. ○ はプロビット変換の補正値. 0%と100%はプロビット変換ができないので ● が表示されない. ● と ○ がわずかに乖離していることが見いだされる.

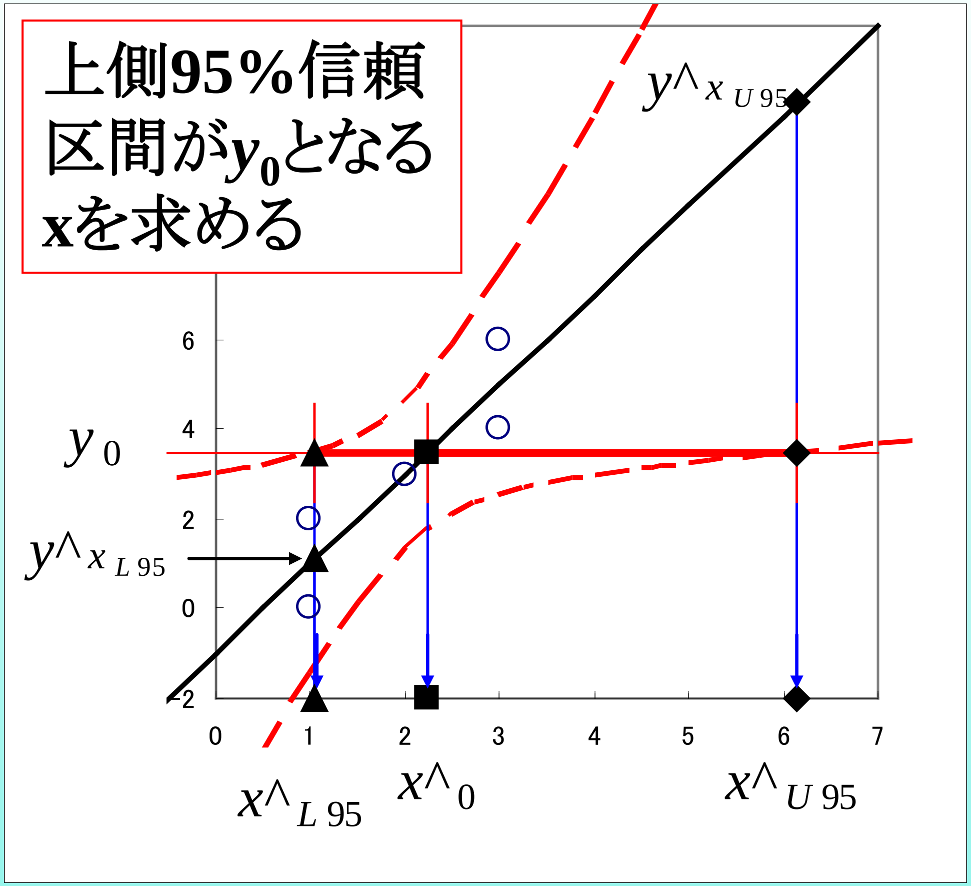
# $LD_{50}$の推定と近似95%信頼区間

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 1) Intercept()= | | | -7.621 | $\beta_0$^= | -10.070 | | 8) 更新 $\beta_0$ ^ = | | -10.070 | 8.2E-11 |
| | | Slope()= | 5.559 | $\beta_1$^= | 6.615 | | | $\beta_1$^ = | 6.615 | 平方和 |
| | | 回帰係数 初期値 | | 2) 係数・貼付 | | | | $(X^TWX)^{-1}X^TWz$ | | |
| | | | | | | | | | | |
| | 6) | 20.64 | 47.13 | 6) | 10.320 | -4.498 | 7) | 103.94 | | |
| | | 47.13 | 108.13 | | -4.498 | 1.970 | | 240.72 | | |
| | | $X^TWX$ | | | $(X^TWX)^{-1}$ | | | $X^TWz$ | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $(5-\beta_0\hat{})/\beta_1\hat{}$ | | 偏微分式 | $d$ | $d^T(X^TWX)^{-1}d$ | | | 対数 | *dose* |
| 対数$LD_{50}$= | | 2.278 | $d_0$= -1/$\beta_1$^= | | -0.151 | *Var* = | 0.0011 | L95%= | 2.213 | 163.3 |
| | $LD_{50}$= | 189.7 | -(5-$\beta_0$^)/$\beta_1$^$^2$= | | -0.344 | *SE* = | 0.0333 | U95%= | 2.343 | 220.5 |

**表2. 分散共分散行列$(X^TWX)^{-1}$を用いたデルタ法で計算したSEから95%信頼区間を計算する.**

# 正確な信頼区間 模式図




- 重み付き回帰でも逆推定値の正確な信頼区間は，単回帰分析の場合と同じである.
- 竹内啓（1963）数理統計学，312-315.
- 反応の推定値の95%信頼区間は推定反応に対して上下で対称だが，逆推定の場合は、用量についての推定なので95%信頼区間は左右で非対称になる.

# $LD_{10}$の正確な95%信頼区間

| | |
|---|---|
| 26.840 | -15.082 |
| -15.082 | 8.481 |

$I^{-1}$分散共分散

$$Var(\hat{\beta}_0) = 26.840$$

$$Var(\hat{\beta}_1) = 8.481$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = -15.082$$

$$t_\alpha = 1.96$$

2次式の
解の公式

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $p_0$= | 0.10 | $\eta_0$= | -2.197 | $LD_{p0}$= | 1.708 | $L95\%$= | 1.691 | | |
| $a$= | -864.6 | $b$= | 1013.9 | $c$= | -297.2 | $U95\%$= | 1.720 | | |

$$\hat{x}_{L95} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

$$\underbrace{\left[Var(\hat{\beta}_0) - \frac{(\eta_0 - \hat{\beta}_0)^2}{t_\alpha^2}\right]}_{a} + \underbrace{\left[2Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(\eta_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2}\right]}_{b} \hat{x}_{L95} + \underbrace{\left[Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_\alpha^2}\right]}_{c} \hat{x}_{L95}^2 = 0$$

***a*, *b*, *c* は，2次式の係数**

# 注）*LD*$_{50}$などの95%信頼区間

- ◆ *LD*$_{50}$など逆推定の95%信頼区間を求めるために，伝統的に「Fiellerの式」が使われている.
- ◆「Fiellerの式」は，*LD*$_{50}$が回帰係数の比によって推定されることから，比の正確な信頼区間を推定する式である.
- ◆ *LD*$_{50}$が回帰直線の95%信頼区間を横切るときの 用量 $x$ から推定しても，同じ式に帰着する.

# ロジット変換での重み付き回帰

## リンク関数から補正値 *z* と 重み *w* の計算式を求める

ロジット変換

リンク関数 $g(\mu_i) = \log[\mu_i / (1-\mu_i)] = \boldsymbol{x}^T \boldsymbol{\beta} = \eta_i$

補正値 $z_i$ $z_i = \eta_i + (y_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) = \eta_i + \frac{y_i - \mu_i}{\boxed{\mu_i(1-\mu_i)}}$

重み $w_{ii}$ $w_{ii} = \frac{1}{\mathrm{var}(Y_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \frac{\boxed{n_i[\mu_i(1-\mu_i)]^2}}{\mu_i(1-\mu_i)}$

□ がプロビットと異なる.

# 補2重対数変換での重み付き回帰

## リンク関数から補正値 *z* と 重み *w* の計算式を求める

リンク関数 $g(\mu)=\ln[-\ln(1-\mu)]=\beta_0+\beta_1 x=\eta$

補正値 $z_i$ $$z_i=\eta_i+\frac{y_i-\mu_i}{\boxed{-\ln(1-\mu_i)(1-\mu_i)}}$$

重み $w_{ii}$ $$w_{ii}=\frac{\boxed{n_i[-\ln(1-\mu_i)(1-\mu_i)]^2}}{\mu_i(1-\mu_i)}$$

□ がプロビット・ロジットと異なる.

# 情報行列を用いた最尤法

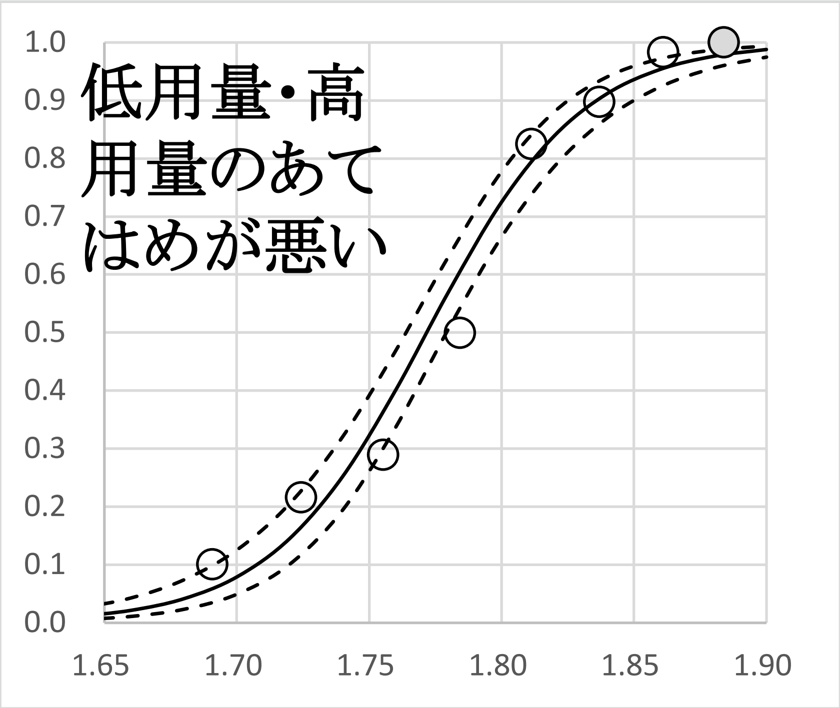
- 反復重み付き回帰は分かったが、他の計算方法はないのだろか.
- 最近は最尤法が一般に用いられているが，どのように実施すればいいのだろうか.
- 田中ら訳（2008），「一般化線形モデル入門」，第7.3節 用量反応モデル 表7.2 カブト虫の死亡データを例として，ロジット変換と補2重対数変換についてExcelで計算してみよう.

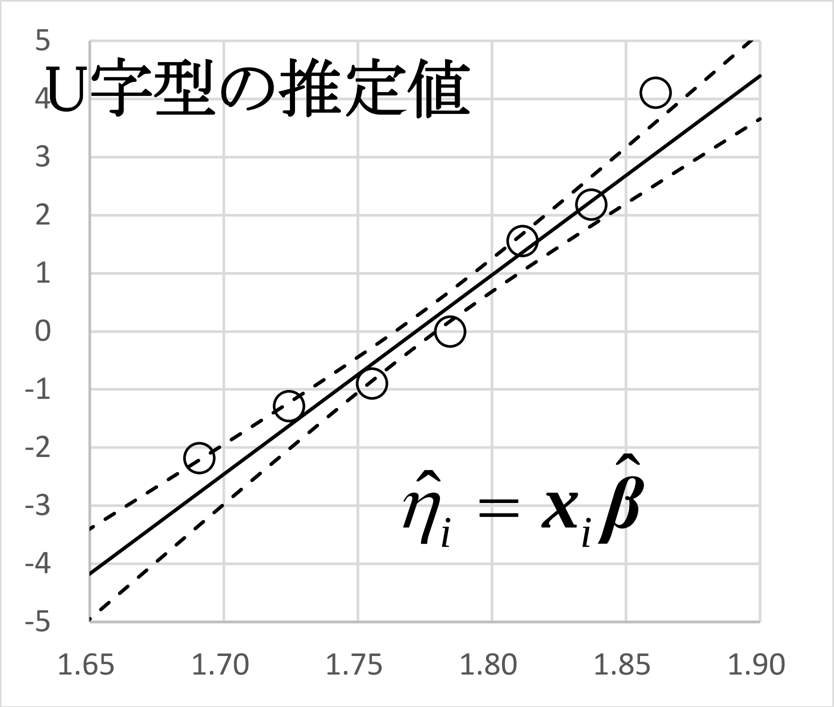
# 情報行列を用いたロジット法 初期値

| | 対数 | 死亡数 | 合計 | 出現率 | 推定値 | 対数尤度 | 行列 $\boldsymbol{U}$ の要素 | | 情報行列 $\boldsymbol{I}$ の要素 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 用量 $x_i$ | $y_i$ | $n_i$ | $p_i$ | $\pi_i$ | $\ln(L_i)$ | $U_{i1}$ | $U_{i2}$ | $I_{i11}$ | $I_{i12}$ | $I_{i21}$ | $I_{i22}$ |
| 1 | 1.6907 | 6 | 59 | 0.102 | 0.500 | -40.896 | -23.500 | -39.731 | 14.750 | 24.938 | 24.938 | 42.162 |
| 2 | 1.7242 | 13 | 60 | 0.217 | 0.500 | -41.589 | -17.000 | -29.311 | 15.000 | 25.863 | 25.863 | 44.593 |
| 3 | 1.7552 | 18 | 62 | 0.290 | 0.500 | -42.975 | -13.000 | -22.818 | 15.500 | 27.206 | 27.206 | 47.751 |
| 4 | 1.7842 | 28 | 56 | 0.500 | 0.500 | -38.816 | 0.000 | 0.000 | 14.000 | 24.979 | 24.979 | 44.567 |
| 5 | 1.8113 | 52 | 63 | 0.825 | 0.500 | -43.668 | 20.500 | 37.132 | 15.750 | 28.528 | 28.528 | 51.673 |
| 6 | 1.8369 | 53 | 59 | 0.898 | 0.500 | -40.896 | 23.500 | 43.167 | 14.750 | 27.094 | 27.094 | 49.769 |
| 7 | 1.8610 | 61 | 62 | 0.984 | 0.500 | -42.975 | 30.000 | 55.830 | 15.500 | 28.846 | 28.846 | 53.681 |
| 8 | 1.8839 | 60 | 60 | 1.000 | 0.500 | -41.589 | 30.000 | 56.517 | 15.000 | 28.259 | 28.259 | 53.236 |
| | | | | | $\ln L=$ | -333.40 | 50.500 | 100.785 | 120.25 | 215.71 | 215.71 | 387.43 |
| | | 切片 $\beta_0$^= | | 0.000 | -37.856 | -37.856 | | 50.500 | 120.25 | 215.71 | 6.743 | -3.754 |
| | | 傾き $\beta_1$^= | | 0.000 | 21.337 | 21.337 | | 100.785 | 215.71 | 387.43 | -3.754 | 2.093 |
| | | | | $\boldsymbol{\beta}^{(m-1)}+\left[\boldsymbol{I}^{(m-1)}\right]^{-1}\boldsymbol{U}^{(m-1)}=\boldsymbol{\beta}^{(m)}$ | | | | $\boldsymbol{U}^{(m-1)}$ | $\boldsymbol{I}^{(m-1)}$ | | $\boldsymbol{I}^{-1}$分散共分散 | |

初期値は(0, 0)で，6回の反復で収束．パラメータの変化は，参考文献の表7.3 と一致．

# ロジット変換での95%信頼区間




$$\pi_i = 1/[1+\exp(-\eta_i)]$$

$$Var(\eta_i) = \boldsymbol{x}_i \boldsymbol{I}^{-1} \boldsymbol{x}_i^T$$

**ロジット変換後の推定値が回帰直線に対してU型に分布しており，当てはめが良いとはいえない**

# 補2重対数変換でのあてはめ

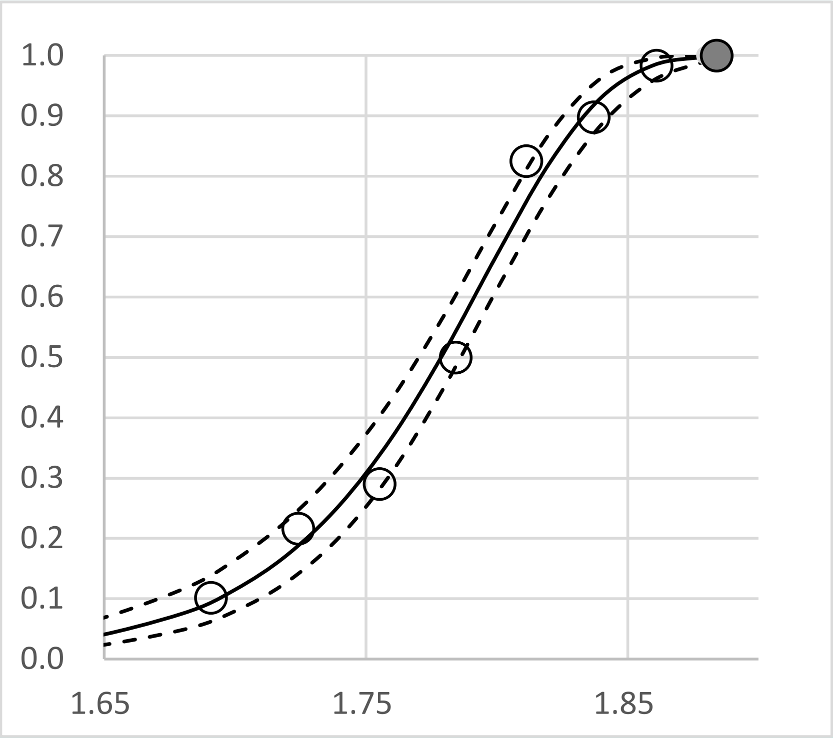
| | 対数 | 死亡数 | 合計 | 出現率 | 推定値 | 対数尤度 | | 行列 $\boldsymbol{U}$ の要素 | | 行列 $\boldsymbol{I}$ の要素 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $i$ | 用量 $x$ | $y_i$ | $n_i$ | $p_i$ | $\pi_i$^ | $\ln(L_i)$ | $z_i$ | $U_{i1}$ | $U_{i2}$ | $I_{i11}$ | $I_{i12}$ | $I_{i21}$ | $I_{i22}$ |
| 1 | 1.6907 | 6 | 59 | 0.102 | 0.095 | -19.415 | 0.100 | 0.431 | 0.729 | 5.564 | 9.407 | 9.407 | 15.904 |
| 2 | 1.7242 | 13 | 60 | 0.217 | 0.188 | -31.515 | 0.208 | 1.904 | 3.284 | 11.049 | 19.050 | 19.050 | 32.846 |
| 3 | 1.7552 | 18 | 62 | 0.290 | 0.338 | -37.674 | 0.412 | -3.605 | -6.328 | 21.352 | 37.478 | 37.478 | 65.781 |
| 4 | 1.7842 | 28 | 56 | 0.500 | 0.542 | -39.017 | 0.782 | -3.415 | -6.093 | 30.032 | 53.583 | 53.583 | 95.603 |
| 5 | 1.8113 | 52 | 63 | 0.825 | 0.758 | -30.007 | 1.420 | 7.910 | 14.328 | 36.164 | 65.504 | 65.504 | 118.648 |
| 6 | 1.8369 | 53 | 59 | 0.898 | 0.918 | -19.536 | 2.497 | -3.109 | -5.711 | 35.417 | 65.057 | 65.057 | 119.503 |
| 7 | 1.8610 | 61 | 62 | 0.984 | 0.986 | -5.126 | 4.247 | -0.488 | -0.908 | 16.687 | 31.054 | 31.054 | 57.791 |
| 8 | 1.8839 | 60 | 60 | 1.000 | 0.999 | -0.053 | 7.036 | 0.372 | 0.700 | 2.246 | 4.231 | 4.231 | 7.970 |
| | | | | | 和 $\ln L$= | -182.343 | | 0.000 | 0.000 | 158.51 | 285.36 | 285.36 | 514.05 |
| | | 切片 $\beta_0$^= | | -39.572 | 0.000 | -39.572 | | | 0.000 | 158.51 | 285.36 | 10.427 | -5.788 |
| | | 傾き $\beta_1$^= | | 22.041 | 0.000 | 22.041 | | | 0.000 | 285.36 | 514.05 | -5.788 | 3.215 |
| | | | | $\boldsymbol{\beta}^{(m-1)}$ | $\left[\boldsymbol{H}^{(m-1)}\right]^{-1}$ $\boldsymbol{U}^{(m-1)}$ | $\boldsymbol{\beta}^{(m)}$ | | | $\boldsymbol{U}$ | $\boldsymbol{I}$ | | $\boldsymbol{I}^{-1}$分散共分散 | |

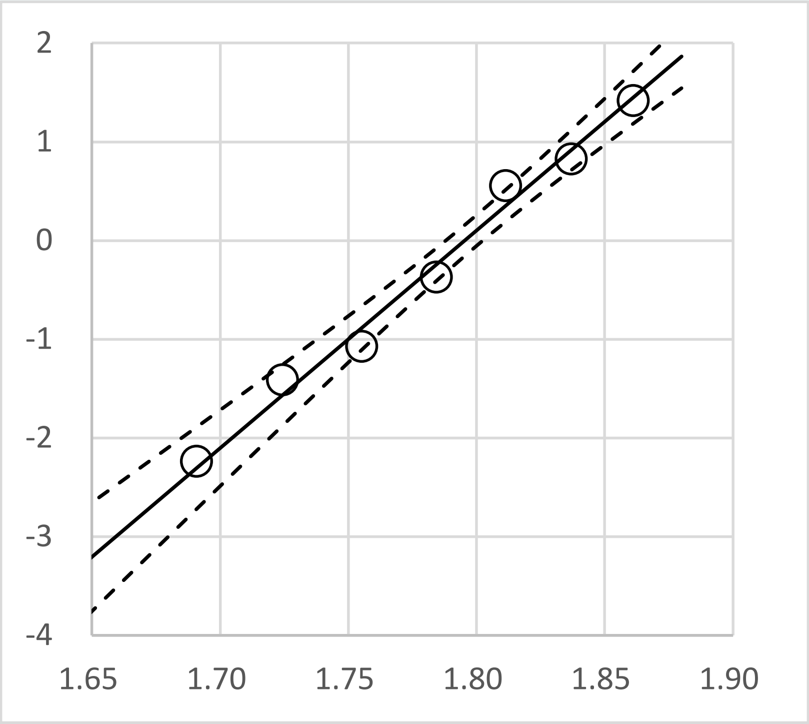
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $p_0$= | 0.10 | $\eta_0$= | -2.250 | $LD_{p0}$= | 1.693 | $L95\%$= | 1.672 |
| $a$ = | -352.2 | $b$ = | 416.7 | $c$ = | -123.2 | $U95\%$= | 1.709 |

$$\hat{x}_{L95} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

**スコア行列*U* および 情報行列 *I* の計算式を補2重対数の式にかえるだけで，他はプロビット・ロジットの場合と全く同じである．**

# 補2重対数変換 95%信頼区間




推定シグモイド曲線および直線の
上下に推定値がバランスよく分布している.

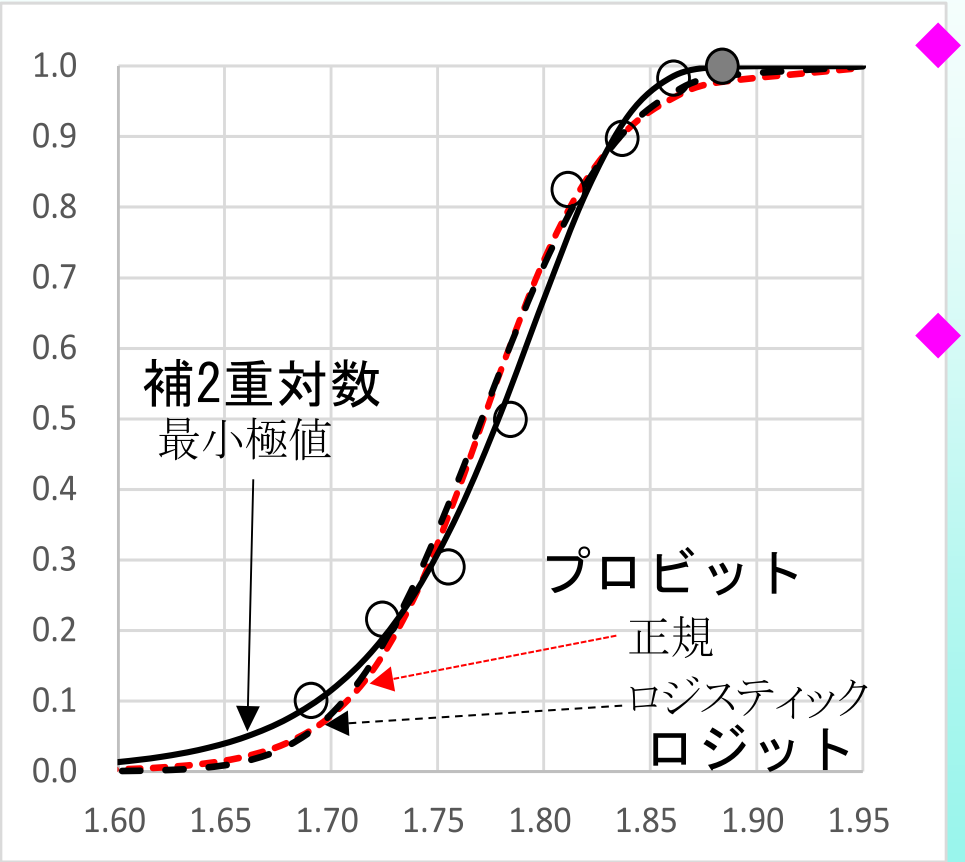
# ロジット・プロビット・補2重対数

| | | | | | ロジット | | プロビット | | 補2重対数 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 対数 | 死亡数 | 合計 | 出現率 | 推定値 | 対数尤度 | 推定値 | 対数尤度 | 推定値 | 対数尤度 |
| $i$ | 用量 $x$ | $y_i$ | $n_i$ | $p$ | $\pi_i$^ | $\ln(L_i)$ | $\pi_i$^ | $\ln(L_i)$ | $\pi_i$^ | $\ln(L_i)$ |
| 1 | 1.6907 | 6 | 59 | 0.102 | 0.059 | -20.222 | 0.057 | -20.303 | 0.095 | -19.415 |
| 2 | 1.7242 | 13 | 60 | 0.217 | 0.164 | -31.921 | 0.179 | -31.639 | 0.188 | -31.515 |
| 3 | 1.7552 | 18 | 62 | 0.290 | 0.362 | -38.067 | 0.379 | -38.421 | 0.338 | -37.674 |
| 4 | 1.7842 | 28 | 56 | 0.500 | 0.605 | -40.087 | 0.604 | -40.051 | 0.542 | -39.018 |
| 5 | 1.8113 | 52 | 63 | 0.825 | 0.795 | -29.360 | 0.788 | -29.459 | 0.758 | -30.006 |
| 6 | 1.8369 | 53 | 59 | 0.898 | 0.903 | -19.407 | 0.904 | -19.408 | 0.918 | -19.536 |
| 7 | 1.8610 | 61 | 62 | 0.984 | 0.955 | -5.902 | 0.962 | -5.621 | 0.986 | -5.126 |
| 8 | 1.8839 | 60 | 60 | 1.000 | 0.979 | -1.270 | 0.987 | -0.777 | 0.999 | -0.053 |
| | | | | | 和 $\ln L$ = | -186.235 | 和 $\ln L$ = | -185.679 | 和 $\ln L$ = | -182.343 |
| | | | | | 切片 $\beta_0$^= | -60.716 | $\mu$ ^= | 1.771 | $\mu$ ^= | 1.795 |
| | | | | | 傾き $\beta_1$^= | 34.270 | $\sigma$ ^= | 0.051 | $\sigma$ ^= | 0.045 |

**Excelソルバーにより，対数尤度を最大化した．**

**補2重対数変換（最小極値分布）のあてはめが最もよいことは，参考文献の表7.4に示されている**

# 3種のモデルの比較




- 補2重対数変換（最小極値分布）のあてはめが最もフィットしている.
- ロジット変換（ロジスティック分布）のほうがプロビット変換（正規分布）に対してわずか立ち上がっている.

# まとめ

- 一般線形モデル関連の教科書は統計ソフトの使い方のマニュアル的になっており，$LD_{50}$のような実用上の諸問題については言及されていない．
- 一般化線形モデル（GLIM）で用いられる反復重み付き回帰および情報行列を用いた最尤法の計算過程を，Excelで実施できるようにした．
- プロビット，ロジット，補2重対数など種々の変換に対応でき，応用力の向上が期待される．