第 6 回 続高橋セミナー

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

2017 年 8 月

高橋 行雄　　BioStat 研究所(株)

要約：　一般化線形モデル（GLIM）の解を求めるためには，反復重み付き回帰，あるいは，情報行列を用いた反復計算による最尤法による解析が必要である．しかし，最尤法の計算過程は煩雑であり，統計ソフトに計算を委ねるのが一般的である．その結果として多くの成書では，計算過程をブラックボックスとして扱い，統計ソフトの使い方と結果の解釈を主体とする記述となりがちである．そこで，解を求めるための煩雑な反復計算の過程を Excel により可視化することにより最尤法について理解を深め，統計ソフトが対応していない種々の実用上の問題に対して解決の糸口を提示する．成書で GLIM の理論を学習する際に，自ら Excel で計算を行い，計算結果を追試することは，理論を正しく理解できているかを確認するために有益である．統計ソフトが対応していない種々の応用上の課題を，統計ソフトの出力を Excel に取り込み再利用することよって実現できる．第 6 回目のセミナーとしては，二値反応に対する用量反応関係にシグモイド曲線をあてはめる方法に特化する．

目 次

1. はじめに ---------- 1
2. 反復重み付き回帰によるプロビット法 ---------- 3
3. 反復重み付き回帰によるロジット法 ---------- 23
4. 反復重み付き回帰による補 2 重対数法 ---------- 29
5. 情報行列を用いたロジット法によるシグモイド曲線の推定 ------- 37
6. 情報行列を用いた補 2 重対数法によるシグモイド曲線の推定 - 54
7. 情報行列を用いプロビット法によるシグモイド曲線の推定 ------- 64
8. 3 種のシグモイド曲線の比較 ---------- 69
9. JMP によるロジット･ロビット･補 2 重対数 ---------- 73
10 まとめ ---------- 84
参考文献 ---------- 86
索引 ---------- 87
Excel, JMP ファイル 一覧 ---------- 90

非売品，無断複製を禁ずる
連絡先:高橋 行雄, 〒105-0014 東京都 港区 芝 1-12－3
takahashi.stat@nifty.com ，FAX : 03-342-8035

目 次　　一般化線形モデルを EXCEL で極め活用する

1. はじめに　1

2. 反復重み付き回帰によるプロビット法　3

2.1 環境ホルモン実験におけるベンチマークドーズ　3
2.2 一般化線形モデルによるプロビット法　4
2.3 プロビット法　6
2.4 EXCLE による反復重み付き回帰によるプロビット法　8
2.5 50 パーセント致死量の推定　14
2.6 シグモイド曲線の 95%パーセント信頼区間　16
2.7 補遺　20

3. 反復重み付き回帰によるロジット法　23

3.1 ロジット変換での重み付き回帰　23
3.2 ロジット変換での 50%出現率の推定　26
3.3 補遺　27

4. 反復重み付き回帰による補 2 重対数法　29

4.1 最小極値分布とワイブル分布の関係　29
4.2 ワイブル分布を最小極値分布へ変換　30
4.3 最小極値分布をシグモイド曲線とする補 2 重対数法　32

5. 情報行列を用いたロジット法によるシグモイド曲線の推定　37

5.1 情報行列を用いたロジット法　37
5.2 情報行列を用いた EXCEL によるロジット法　40
5.3 シグモイド曲線の 95%信頼区間　45
5.4 出現率が 10 パーセントとなる用量とその近似 95%信頼区間　47
5.5 LD_{10} の正確な 95%信頼区間　49
5.6 古典的なフィラーの式を用いた 95%信頼区間　51

6. 情報行列を用いた補 2 重対数法によるシグモイド曲線の推定　54

6.1 EXCEL ソルバーを用いた最尤法　54
6.2 最小極値分布をシグモイド曲線とする補 2 重対数法　56

6.3 情報行列を用いた最尤法 58
6.4 分散共分散行列を用いたシグモイド曲線の 95%信頼区間 61

7. 情報行列を用いたプロビット法によるシグモイド曲線の推定 64

7.1 位置パラメータ μ と形状パラメータ σ の推定 64
7.2 情報行列を用いた最尤法 66

8. 3 種のシグモイド曲線の比較 69

8.1 位置パラメータと形状パラメータを用いた比較 69
8.2 3 種のシグモイド曲線の形状の比較 71

9. JMP によるロジット・プロビット・補 2 重対数 73

9.1 統計ソフト JMP による一般化線形モデル 73
9.2 JMP によるロジット法 74
9.3 JMP によるプロビット法 78
9.4 JMP による補 2 重対数法 80

10. まとめ 84

文献 86
索引 87
EXCEL，JMP ファイル 一覧 90

1. はじめに

2017 年 3 月の日本計量生物学会・年会で「一般化線形モデルを Excel で極め活用する」と題して一般演題で発表を行なった．この際の 6 ページ分の抄録と発表に用いたパワーポイントを元に第 6 回目のセミナーを行なう．

一般化線形モデル（GLIM，Genralized Linear Model）の解を求めるためには，反復重み付き回帰，あるいは，情報行列を用いた反復計算による最尤法による解析が必要である．しかし，最尤法の計算過程は煩雑であり，統計ソフトに計算を委ねるのが一般的である．その結果として多くの成書では，計算過程をブラックボックスとして扱い，統計ソフトの使い方と結果の解釈を主体とする記述となりがちである．

一般線形モデルの略号として古くから GLM が使われているので，一般化線形モデルを GLM と称すると区別がつかない．区別のために，普通の最小 2 乗法を OLS（Ordinary Least Square regression）が用いられているが，OLS を一般線形モデル（Genelal Linear Model）とは言い難い．一般線形モデルは，回帰分析，分散分析，共分散分析なども含め一般化した概念であり，普通の最小 2 乗回帰（OLS）とは異なる．

本セミナーは, GLIM の解を求めるための煩雑な反復計算の過程を Excel により可視化することにより最尤法について理解を深め，統計ソフトが対応していない種々の実用上の問題に対して解決の糸口を提示する．これにより，GLIM に限らず各種の最尤法による統計モデルの解析を自在に扱えるようになることを期待している．Excel には最小限の行列関数が備わっており，これらを用いて反復計算アルゴリズムを一枚の計算シート上に実現することにより「手計算」を可能とする．

成書で GLIM の理論を学習する際に，自ら Excel で計算を行い，計算結果を追試することは，理論を正しく理解できているかを確認するために有益である．理論の理解が，統計ソフトで対応していない種々の応用上の課題を，統計ソフトの出力を Excel に取り込み再利用することよって実現できる．

事例として，古典的な生物検定法である 50 パーセント致死量（LD_{50}），あるいは，50 パーセント有効濃度（EC_{50}）を推定するためのプロビット法を事例として取り上げ，GLIM の理論に基づいてロジット法などへの応用を例示する．プロビット法の事例は，(0, 1)反応の実験データにシグモイド状の用量反応曲線をあてはめ，反応が 50 パーセントとなる用量を逆推定し，その 95%信頼区間を推定する問題である．

本セミナーでは，GLIM で定式化された計算方法により LD_{50} を推定することを事例として取り上げる．歴史的にプロビット法は，計算時間の短縮のために反復重み付き回帰を適用する方法が，多くの成書で示されてきた．しかし，理論的な説明が省略されて計算方法を主体とした記述になりがちであり，種々の応用あるいは拡張を試みることは実務家にとって困難であった．

田中ら訳（2008）の「一般化線形モデル入門」は，GLIM の理論について詳細な記述があり，各種の応用を試みるためには，必読の成書である．残念ながら，反復重み付き回帰による最尤法については，ポアソン回帰の事例が示されているだけで，プロビット法に関する例示がない．そこで，GLIM の理論に則って，Excel を用いた反復重み付き回帰をプロビット法の反復計算の過程を詳細に示し，各種の応用法について例示する．

シグモイド曲線として累積正規分布に変え，累積ロジスティック分布を用いたロジット法とする場合の変更点を示し応用事例とする．プロビット法およびロジット法でのシグモイド曲線は点対称であり，低用量に裾を引く反応に対するシグモイド曲線として累積最小極値分布（補 2 重対数法）が望ましい性質を持っているので，GLIM のさらなる応用事例として例示する．

反復重み付き回帰による最尤法は、統計モデルをリンク関数の形式で線形化できる場合に限定された方法であり，寿命データなどで打ち切りのある統計モデルに対しては適用できない．情報行列をベースとしたニュートン・ラフソン法による最尤法は，GLIM のみならず多様な統計モデルに対して適用できるので，反復重み付き回帰による最尤法と対比して Excel での反復計算により可視化する．

統計ソフト JMP を用いて，本セミナーで取り上げた Excel での計算結果が正しいのかの検証も兼ねて統計ソフトの出力と対比する．これにより，統計ソフトの出力をブラックボックスとしてではなく，能動的に使いこなすスキルの向上を願っている．

最近の統計ソフトには，情報行列を計算するための 2 階の偏微分式を内部で解く機能を持っている．複雑な対数尤度の偏微分式を解くことは難儀なことで，プラス・マイナスをミスすることがしばしば起きる．そのときに，自ら解いた偏微分式による数値計算と統計ソフトで計算した偏微分式の計算結果とを対比することにより，Excel のシートに入力した偏微分式の正しさ検証できる．

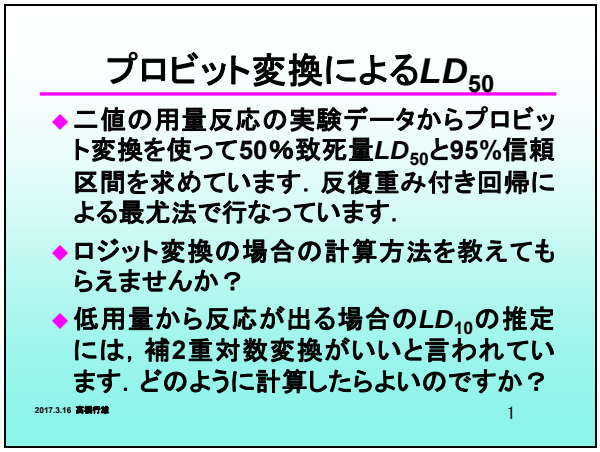
2.　反復重み付き回帰によるプロビット法

Finney（1971）により反復重み付き回帰を用いたプロビット法による LD_{50} の推定法が示され、佐久間（1977，2017），吉村編著（1987）でも踏襲され，一般的な方法として普及してきた．この方法は，重み付き平方和を用いた計算方法で，重みおよび *probit* 値を反復ごとに補正し，新に推定された推定値との差が近似的に 0 となった場合に収束したとみなす方法である．

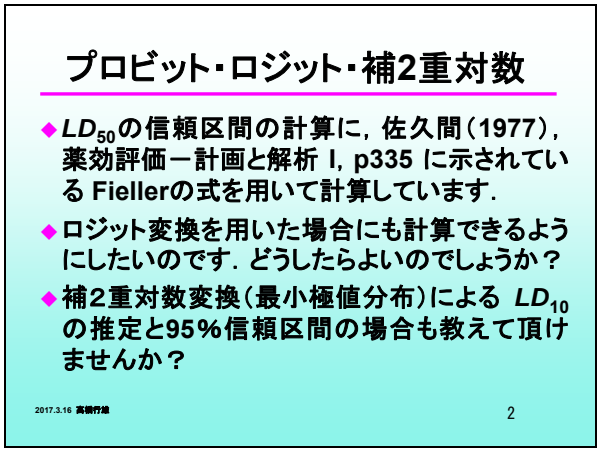
2.1　環境ホルモン実験におけるベンチマークドーズ

スライド 1 と 2 に示すように，ある実験研究者から相談があった．研究施設で昔から使われている Basic 言語で作成されたプログラムライブラリーを使用しているようだ．「低用量から反応が出る場合の LD_{10} の推定には，補 2 重対数法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？」との質問から環境ホルモンなどの実験研究で使われ始めたベンチマークドーズの要求に対応したいと思われた．

スライド 1



スライド 2



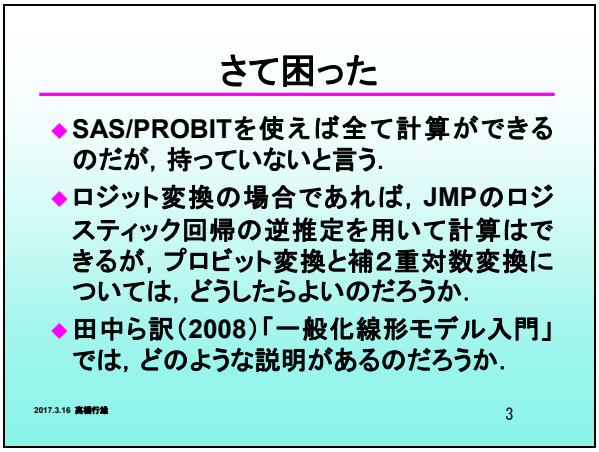
統計ソフト SAS が 1980 年代に提供された時代から PROBIT プロシジャが提供されていて，50 パーセント致死量の推定のみならず，10 パーセント致死量の推定も可能であり実務的には，解決済みの問題であった．近年では，統計ソフト JMP も，これらの逆推定について対応している．なお，JMP による逆推定については，第 9 章で詳しく示す

様々な統計相談を多くの方々から受けるようになり，スライド 3 に示すように SAS/PROBIT を使えばできるとの回答は躊躇せざるを得なくなった．質問に解答できるような文

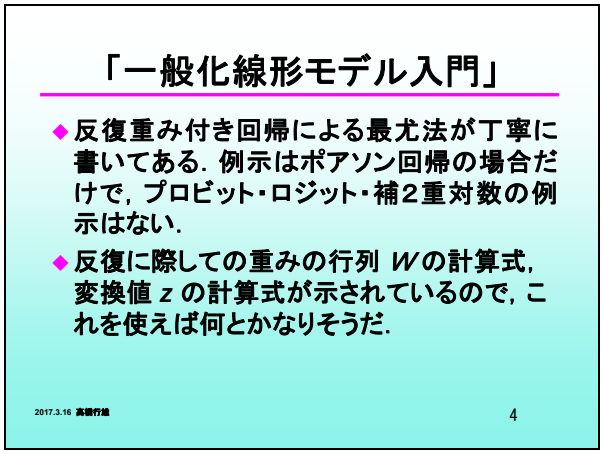
献を WEB で検索しても見当だせず，本当に困ってしまった．

スライド 4 に示すように，ロジット法の場合であれば，JMP の古いバージョンからサポートされているロジスティック回帰の逆推定を用いて計算は簡単にできるが，プロビット法と補 2 重対数法については，どうしたらよいのだろうか．田中ら訳（2008）では，どのような説明があるのだろうか．プロビット法，ロジット法，補 2 重対数法も一般化線形モデルで統一的に解決できることは承知してはいたが，実際の計算方法を自ら説明することは容易ではなかった．

スライド 3



スライド 4



2.2　一般化線形モデルによるプロビット法

一般化線形モデル（GLIM）では，期待値を $E(Y_i)=\mu_i$，リンク関数を $g(\mu_i)=\boldsymbol{x}^T\boldsymbol{\beta}=\eta_i$ としたときに，重み付き回帰式

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

の $\boldsymbol{W}$ と $\boldsymbol{z}$ を逐次的に変化させて $\hat{\boldsymbol{\beta}}$ を推定し，何らかの収束条件が満たされるまで反復を繰り返す．逐次的な変化は，田中ら訳(2008)の(4.23)式および(4.24)式より，重み w_{ii} および補正 *probit* 値 z_i は

$$w_{ii}=\frac{1}{Var(Y_i)}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2,\qquad z_i=\eta_i+(y_i-\mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right)$$

であり，これらを逐次更新し，収束するまで反復することでパラメータの推定を行う．出現率 y_i に対する用量反応関係にリンク関数として標準正規分布の逆累積分布 $\Phi^{-1}(\mu_i)$ に 5 を加えた *probit* 値が伝統的に用いられている．吉村編著（1987）の第 5 章のデータ(表 5-2)を用いて例示する．伝統的にはプロビット法の計算の例示には，計算過程を逐次的に示すために偏

差平方和を用いた計算式が使われてきた．現代の標準的な計算手段となっている Excel で偏差平方和計算するのは，煩雑であり行列計算でスマートに対応したい．そこで，一般線形モデル(GLM)で定式化されている行列関数を用い行列の積による偏差平方和を一括計算することにした．これにより，反復重み付き回帰によるプロビット法の計算を一枚の Excel シート上に展開することができ，煩雑な計算を簡素化することができる．

ここでは，出現率を y_i とし，推定値を μ_i としたが，出現数を y_i とし，出現率を $p_i = y_i / n_i$ とする表記もあり，統一的に記述すると，一般線化形モデル(GLIM)での標準的な表記法との異なる場合もあり，混在することに容赦を願いたい．

吉村編著(1987)で示されている「プロビット法による LD_{50} の推定」には，反復計算の過程が逐次的に示されているので，これを参考に GLIM の理論に従って Excel の行列計算で再現することにした．

スライド 5 に示すように吉村編著 : p226-32 の例でプロビット変換による反復重み付き回帰を Excel で実行し，計算法について理解を深めることにした．文献には，表 2.1 に示す数値例が示されており，重み付き回帰についても計算手順が丁寧に述べられている．

スライド 5

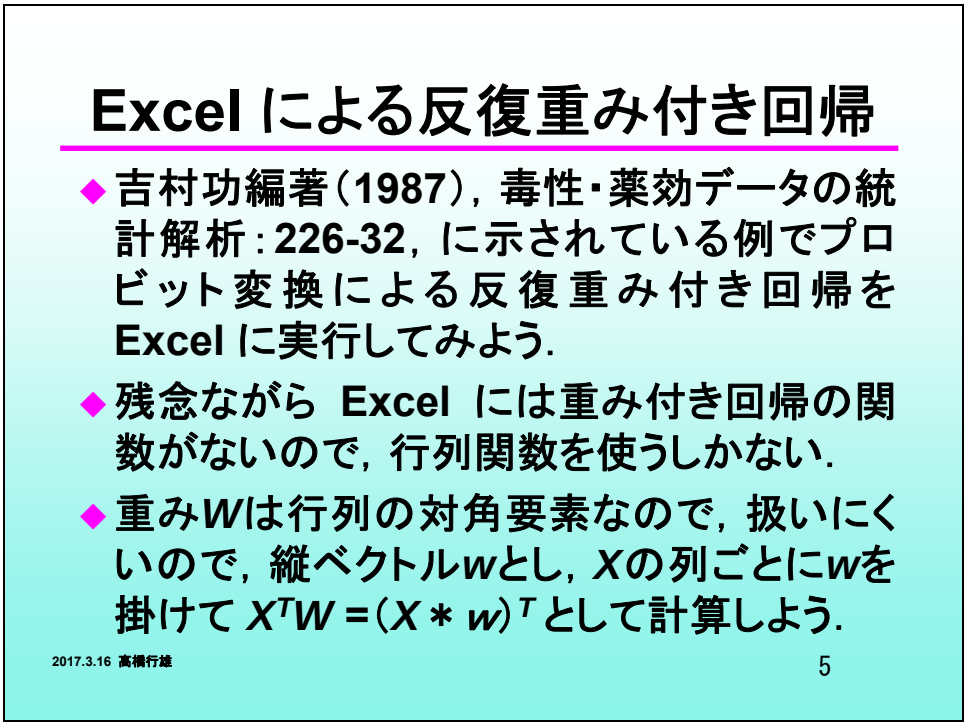


表 2.1　LD_{50} を求める数値例（吉村）

群 i	投与量 mg/kg (公比 1.35)	群の大きさ n_i	死亡数 y_i	死亡出現率 p_i	プロビット *probit*
1	101	10	0	0.000	$-\infty$
2	136	10	2	0.200	4.1584
3	183	10	5	0.500	5.0000
4	247	10	8	0.800	5.8416
5	333	10	9	0.900	6.2816
6	450	10	10	1.000	∞

実験結果は，公比 1.35 で設定された投与量に対する各群 10 匹中の動物の死亡数であり，死亡の出現率が示されている．対数用量に対して出現率の用量反応曲線は，シグモイド曲線状になることが経験的に知られている．図 2.1 に示すようにシグモイド曲線状とは，低い用量での出現率が 0 パーセントから緩やかに上昇し，ある用量から次第に立ち上がり，ある用量から次第に緩やかになり 100 パーセントに収束するような単調増加の曲線の総称で S 字曲線とも呼ばれている．

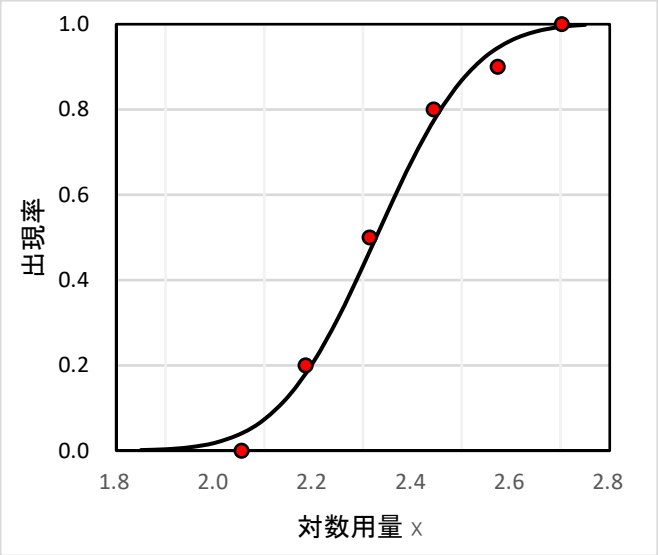


図 2.1 シグモイド曲線のあてはめ

2.3　プロビット法

用量反応関係がシグモイド曲線状となる関数として，累積正規分布が歴史的に用いられてきた．統計モデルとして考えるならば，出現率が 50 パーセントとなる用量を平均値とし，標準偏差を用量反応の形状から求める問題である．シグモイド曲線を直接あてはめることは，計算手段の乏しい時代には困難であった．そこで，ある用量 x_i に対する出現率 $p_i = y_i / n_i$ につ

いて累積正規分布の下側パーセント点 $\Phi^{-1}(p_i)$ に 5 を加えてプロビット変換

$$probit_i = z_i = \Phi^{-1}(p_i) + 5$$

し，その変換値 z_i に反復重み付き回帰によって図 2.2 に示すように直線をあてはめる．

プロビット変換は，出現率が 0%および 100%の場合には，マイナス無限大およびプラス無限大となり，そのままでは回帰分析でも用いることができない．そのために，例数 n_i に 1 を加え，反応 y_i に 0.5 を加えて

$$p_i^* = \frac{y_i + 0.5}{n_i + 1}$$

のように出現率を調整（経験プロビット）してプロビットを計算できるようにして，重み付き回帰を行う便宜的な対応も散見する．ただし，出現率が 0%の場合は，$n_i = 10$ としたとき出現率は 4.5%, 100%の場合は 95.9%となり, 直線の傾きが小さくなるバイアスが生ずる. 表 2.1 でも出現率が 0%と 100%となる場合に，出現率の調整を行うこともできるが，最尤法での反復計算に初期値として用いる必然性はないので割愛する．

プロビット法は，対数用量 x に対する回帰直線

$$z_i = \beta_0 + \beta_1 x$$

をあてはめる．あてはめられた直線から，

$$\pi_i = \Phi(z_i - 5) = \Phi(\beta_0 + \beta_1 x_i - 5)$$

によって累積正規分布の推定値 π が得られる．正規分布の平均 μ と標準偏差 σ を用いて

$$\pi_i = \Phi(\beta_0 + \beta_1 x_i - 5) = \Phi\left(\frac{x_i - \mu}{\sigma} - 5\right) = \Phi\left(-\frac{\mu}{\sigma} - 5 + \frac{1}{\sigma} x_i\right)$$

のように示すことができる．回帰式 $z_i = \beta_0 + \beta_1 x$ の切片は $\beta_0 = -\mu/\sigma - 5$，傾きは $\beta_1 = 1/\sigma$ であり，この関係から $\mu = -(\beta_0 - 5)/\beta_1$ と標準偏差 $\sigma = 1/\beta_1$ が導出できる．

表 2.1 の Excel シート上の *probit* は，出現率を累積正規分布の下側確率としたときのパーセント点 x に 5 を加えたものである．出現率 $p_2 = 0.200$ に対する *probit* は 4.1584 であり，これは，Excel の関数

$$probit = \Phi^{-1}(p_2) + 5 = \texttt{Norminv(0.200, 0, 1)+5=4.1584}$$

で計算されたものである． $p_1 = 0.0$ の場合は，

$$probit = \Phi^{-1}(p_1) + 5 = \texttt{Norminv(0.000, 0, 1)+5=-}\infty$$

となり，$p_6 = 1.0$ の場合も図に示ことができない．プロビット変換できる $i = 2,...,5$ の 4 ポイントを用いて目の子で回帰直線を引いたのが 図 2.2 である．

なぜ，プロビット変換するのだろうか．これは，シグモイド曲線状の累積正規分布を“線形化”するための方法として考案されたもので，5 を加えたのは，変換値がマイナスになりにくいようにするための計算上の工夫である．

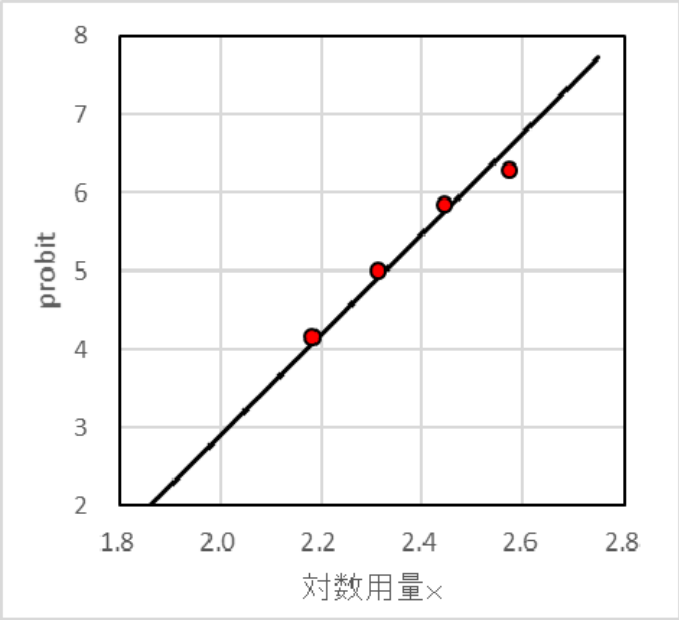


図 2.2 プロビット値に対する直線のあてはめ

シグモイド曲線の線形化に伴って，いやな問題が発生する．通常の回帰分析は，X の範囲内で回帰直線の推定値に対して Y は等分散であるとの暗黙の仮定が存在する．図 2.2 を見るかぎり等分散性が成り立っていると判断される．しかし，広い範囲の出現率 π に対して，2 項分布の分散は，$Var(\pi) = \pi(1-\pi)/n$ なので $\pi = 0.5$ で最大となり等分散ではない．プロビット変換した場合には，逆に $\pi = 0.5$ で最小になり，等分散を仮定することはできないので，分散の逆数を重みとする回帰分析が必要となる．なお，これらの問題については，第 2.6 節 補遺 で示す．

2.4 Excle による反復重み付き回帰によるプロビット法

Excel のシート上での計算過程をプログラミング言語的に示すことは困難なので，計算手順を逐次的に示すことにする．表 2.2 に出現率から計算した probit 値を用いた回帰分析の推定値の計算に引き続き，重み付き回帰の計算シートを示す．このシート上に示した 手順 1）から 手順 10）の番号順に計算の過程を示す．

手順 1）出現率 p_i が 0%と 100%の場合を除いたプロビット変換値 $\boldsymbol{z}^*$ と対数用量 $\boldsymbol{x}_1$ から，Excel の `Intercept` および `Slope` 関数を用いて回帰係数

$\hat{\beta}_0$=`Intecept(`x_1 の範囲，z*の範囲`)=-7.621`

$\hat{\beta}_1$=`Slope(`x_1 の範囲，z*の範囲`)=5.559`

を求める．

表 2.2　Excel によるプロビット法の初期値に対する重み付き回帰

	用量	切片	対数用量	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	y	n	p	z*		3) $\eta^{(1)}$	μ	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	β_0^=	-7.621		8) 更新 β_0^=		-9.598	2.1E-01
			デザイン行列		Slope()=		5.559	β_1^=	5.559			β_1^=	6.411	平方和
					1) 回帰係数 初期値			2) 係数・貼付				$(\boldsymbol{X}^T\boldsymbol{WX})^{-1}\boldsymbol{X}^T\boldsymbol{Wz}$		
					6)	23.69	54.12	6)	7.363	-3.205	7)	119.58		
						54.12	124.34		-3.205	1.403		277.70		
						$\boldsymbol{X}^T\boldsymbol{WX}$			$(\boldsymbol{X}^T\boldsymbol{WX})^{-1}$			$\boldsymbol{X}^T\boldsymbol{Wz}$		

手順 2）推定された切片 $\hat{\beta}_0 = -7.621$，傾き $\hat{\beta}_1 = 5.559$ をコピーして［係数・貼付］欄に数値のみを貼り付ける．

手順 3）貼り付けられた回帰係数から推定値

$$\eta_1^{(1)} = -7.621x_{0,1} + 5.559x_{1,1} = 3.522$$
$$\vdots$$
$$\eta_6^{(1)} = -7.621x_{0,1} + 5.559x_{1,6} = 7.130$$

が計算されている．計算に Excel の行列の積の関数 `Mmult` を用いて

$\boldsymbol{\eta}^{(1)} = \boldsymbol{X\beta}$ = `Mmult(` $\boldsymbol{X}$ の範囲，$\boldsymbol{\beta}$ の範囲`)`

として一括して計算することができる．

1	2.004	-7.621	=	3.522
1	2.134	5.559		4.241
1	2.262	$\boldsymbol{\beta}$		4.957
1	2.393			5.681
1	2.522			6.403
1	2.653			7.130
X				$\boldsymbol{\eta}$

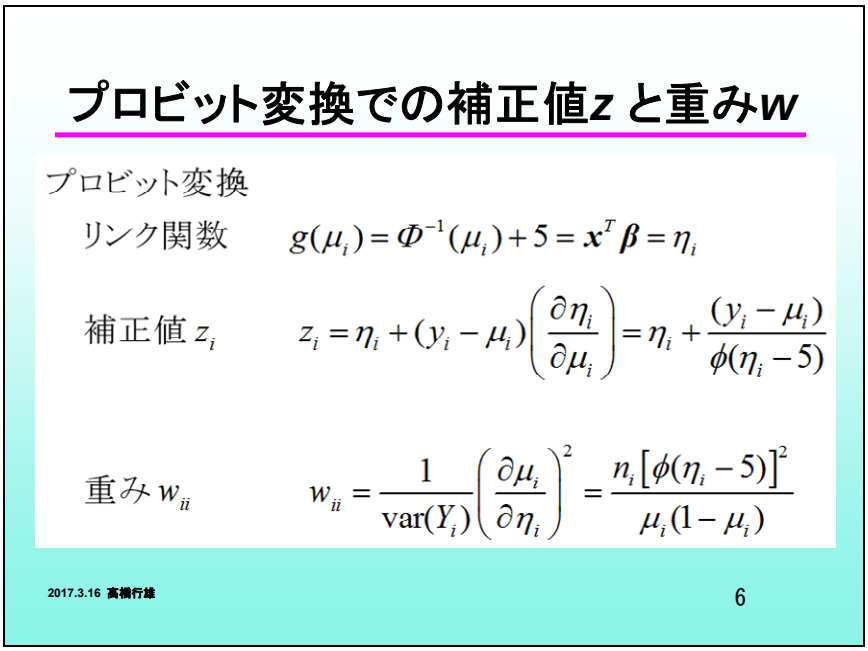
$\boldsymbol{\eta}^{(1)}$ は，*probit* の推定値なので，＝Normdist（$\eta_i^{(1)}$-5,0,1,true） によって出現率に戻す．

$$\mu_1 = \Phi(\eta_1 - 5) \quad = \texttt{Normdist(3.522-5,0,1, true)=0.070}$$

$$\vdots$$

$$\mu_6 = \Phi(\eta_6 - 5) \quad = \texttt{Normdist(7.130-5,0,1, true)=0.983}$$

スライド 6



スライド中の y_i は出現率なので，本文中では全体の整合性を図るために p_i とした．
ここで用いた微分については，第 2.6 節 補遺 で示す．

手順 4）スライド 6 に示すように補正 *probit* 値 z_i を次式によって計算する．

$$z_1 = \eta_1 + \frac{(p_1 - \mu_1)}{\phi(\eta_1 - 5)} \quad = \texttt{3.522+(0.000-0.070)/Normdist(3.522-5,0,1,false)}$$

$$= \texttt{3.001}$$

$$\vdots$$

$$z_6 = \eta_6 + \frac{(p_6 - \mu_6)}{\phi(\eta_6 - 5)} \quad = \texttt{7.130+(1.000-0983)/Normdist(7.130-5,0,1,false)}$$

$$= \texttt{7.531}$$

手順 5）スライド 6 に示すように重み w_i を次式で計算する．

$$w_1 = \frac{n_1[\phi(\eta_1 - 5)]^2}{\mu_1(1-\mu_1)} = \texttt{10*(Normdist(3.522-5,0,1,false))^2/(0.070*(1-0.070))}$$

$$= \texttt{2.762}$$

:

$$w_6 = \frac{n_6\left[\phi(\eta_6 - 5)\right]^2}{\mu_6(1-\mu_6)}$$ =10*(Normdist(7.130-5,0,1,false))^2/(0.983*(1-0.983))

=1.531

手順 6）重み付き回帰係数 $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$ を計算するための重みは行列$\boldsymbol{W}$であるが，対角要素以外は 0 であり対角要素をベクトル$\boldsymbol{w}$として Excel で行列とベクトルの計算は，

$\boldsymbol{X} * \boldsymbol{w}$ =（$\boldsymbol{X}$ の範囲＊$\boldsymbol{\beta}$ の範囲）

とし Transpose 関数で

$(\boldsymbol{X} * \boldsymbol{w})^T$ =Transpse （$\boldsymbol{X} * \boldsymbol{w}$ の範囲）

を計算すると

$$\boldsymbol{X}^T\boldsymbol{W} = (\boldsymbol{X} * \boldsymbol{w})^T$$

と同値になることを活用し，計算シートの簡潔化を図る．

行列$\boldsymbol{W}$としての計算

1	1	1	1	1	1		2.76	0	0	0	0	0	=	2.76	5.15	6.36	5.37	3.01	1.05
2.00	2.13	2.26	2.39	2.52	2.65		0	5.15	0	0	0	0		5.54	10.98	14.39	12.84	7.60	2.77
		$\boldsymbol{X}^T$					0	0	6.36	0	0	0				$\boldsymbol{X}^T\boldsymbol{W}$			
							0	0	0	5.37	0	0							
							0	0	0	0	3.01	0							
							0	0	0	0	0	1.05							
									$\boldsymbol{W}$										

ベクトル$\boldsymbol{w}$としての計算

1	2.00	*	2.76	=	2.76	5.54		2.76	5.15	6.36	5.37	3.01	1.05
1	2.13		5.15		5.15	10.98		5.54	10.98	14.39	12.84	7.60	2.77
1	2.26		6.36		6.36	14.39				$(\boldsymbol{X}*\boldsymbol{w})^T$			
1	2.39		5.37		5.37	12.84							
1	2.52		3.01		3.01	7.60							
1	2.65		1.05		1.05	2.77							
$\boldsymbol{X}$			$\boldsymbol{w}$		$\boldsymbol{X}*\boldsymbol{w}$								

さらに積和行列を $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} = (\boldsymbol{X} * \boldsymbol{w})^T\boldsymbol{X}$ で計算し，Minverse 関数で分散共分散行列 $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$=Minverse（$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$ の範囲）を計算する．

2.76	5.15	6.36	5.37	3.01	1.05		1	2.00	=	23.69	54.12		7.363	-3.205
5.54	10.98	14.39	12.84	7.60	2.77		1	2.13		54.12	124.34		-3.205	1.403
		$(X*w)^T$					1	2.26		$X^T WX$			$(X^T WX)^{-1}$	
							1	2.39						
							1	2.52						
							1	2.65						
							X							

手順 7）同様に $X^T Wz = (X * w)^T z$ を計算する．

2.76	5.15	6.36	5.37	3.01	1.05		3.00	=	119.58
5.54	10.98	14.39	12.84	7.60	2.77		4.16		277.70
		$X^T W=(X*w)^T$					5.00		$X^T Wz$
							5.83		
							6.27		
							7.53		
							z		

手順 8）重み付き回帰係数を行列の積を $\hat{\boldsymbol{\beta}}^{\text{new}} = \left[(X^T WX)^{-1}\right]\left[X^T Wz\right]$ で計算し，その結果が［更新］した回帰係数 $[-9.598, 6.411]^T$ として求められている．

		X		1) Intercept()=	-7.621	β_0^=	-7.621			8) 更新 β_0^=	-9.598	2.1E-01	
	デザイン行列			Slope()=	5.559	β_1^=	5.559			β_1^ =	6.411	平方和	
				1) 回帰係数 初期値		2) 係数・貼付				$(X^T WX)^{-1} X^T Wz$			
				6)	23.69	54.12	6)	7.363	-3.205	7)	119.58		
					54.12	124.34		-3.205	1.403		277.70		
					$X^T WX$			$(X^T WX)^{-1}$			$X^T Wz$		

手順 9）［更新］$\hat{\boldsymbol{\beta}}^{\text{new}}$ を用いて推定値 $\eta^{(2)}$ を $\eta_i^{(2)} = \boldsymbol{x}_i \boldsymbol{\beta}^{\text{new}}$ で計算し，さらに推定値 $\eta^{(1)}$ との差 $\eta_i^{(1)} - \eta_i^{(2)}$ を計算した結果が示されている．ここまでが 1 サイクルである．

3.522	1	2.00	-9.598	=	3.253		0.270
4.241	1	2.13	6.411		4.081		0.160
4.957	1	2.26	更新$\boldsymbol{\beta}$		4.907		0.050
5.681	1	2.39			5.743		-0.061
6.403	1	2.52			6.574		-0.172
7.130	1	2.65			7.413		-0.283
$\eta^{(1)}$	X				$\eta^{(2)}$		$\eta^{(1)}-\eta^{(2)}$

スライド 7

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) $\eta^{(1)}$	μ	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	β_0^=	-7.621			8) 更新 β_0^=	-9.598	2.1E-01
		デザイン行列			Slope()=		5.559	β_1^=	5.559			β_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		

2017.3.16 高橋行雄

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．

7

スライド中の r_i は，本文中では全体の整合性を図るために y_i とし，また，y_i は p_i とする．

手順 10a) 差 $\eta_i^{(1)}-\eta_i^{(2)}$ が 0 とはみなせられず，平方和も大きいと判断されるので，［更新］$\hat{\boldsymbol{\beta}}^{\text{new}}$ をコピーして「係数・貼付」欄に値のみを張りつける．これが 2 サイクル目で，更新された［更新］$\boldsymbol{\beta}^{\text{new}}=[-10.041, 6.603]^T$ が得られる．

2 サイクル目

β_0^=	-9.598		8) 更新 β_0^=	-10.041	1.1E-02
β_1^=	6.411		β_1^=	6.603	平方和

手順 10b) 手順 10a) と同様の処理を行い 3 サイクル目で［更新］$\boldsymbol{\beta}^{\text{new}}=[-10.069, 6.615]^T$ が得られる．

3 サイクル目

β_0^=	-10.041		8) 更新 β_0^=	-10.069	4.2E-05
β_1^=	6.603		β_1^=	6.615	平方和

手順 10c) 同様に 4 サイクル目とすると，差 $\eta_i^{(1)}-\eta_i^{(2)}$ が小数点以下 3 桁目はすべて 0 で，平方和は 5.3×10^{-8} となる．

4 サイクル目	β_0^=	-10.069		8) 更新 β_0^=	-10.070	5.3E-08
	β_1^=	6.615		β_1^ =	6.615	平方和

手順 10d）係数が小数点以下 3 桁で異なるので，5 サイクル目を行い，係数が小数点以下 3 桁まで一致し，平方和は8.2×10^{-11} となるので繰返しを終了する．

表 2.3　反復 5 回目の重み付き回帰によって得られた推定値

	1) Intercept()=		-7.621	β_0^=	-10.070		8) 更新 β_0^=	-10.070	8.2E-11	
	Slope()=		5.559	β_1^=	6.615		β_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		

2.5　50 パーセント致死量の推定

実験で知りたいのは，50 パーセントの動物が死亡する用量である．この用量を直接推定する実験を行うことは困難なので，間接的に幾つかの用量に対して死亡数を実験的に求め，シグモイド曲線上で 50%となる用量を推定する．最尤法で得られたのは，*probit* 値に対する回帰係数なので，死亡が 50%出現する *probit* 値は，$\eta_0 = \Phi^{-1}(0.5)+5=5$ であり，$\eta_0 = 5 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ の関係から，$x_0 = (5-\hat{\beta}_0)/\hat{\beta}_1$ で計算できる．表 2.3 の回帰係数から図 2.3 に示すように 50%対数致死量は，

$$\text{対数}\, LD_{50} = \frac{5-\hat{\beta}_0}{\hat{\beta}} = \frac{5-(-7.621)}{5.559} = 2.278,$$

$$LD_{50} = 10^{2.278} = 189.7$$

として推定される．

LD_{50} の近似 95%信頼区間は，縦軸方向ではなく，横軸方向であるので，次節で示す縦軸方向の 95%信頼区間とは異なる．回帰直線の横軸方向の 95%信頼区間の推定方法は，デルタ法を用いた近似的なもの方法，回帰パラメータの比についての正確な信頼区間を直接求める方法（第 5.6 節），回帰直線の 95%信頼区間から逆推定する方法（第 5.5 節）があり，ここでは，近似的な方法を示す．

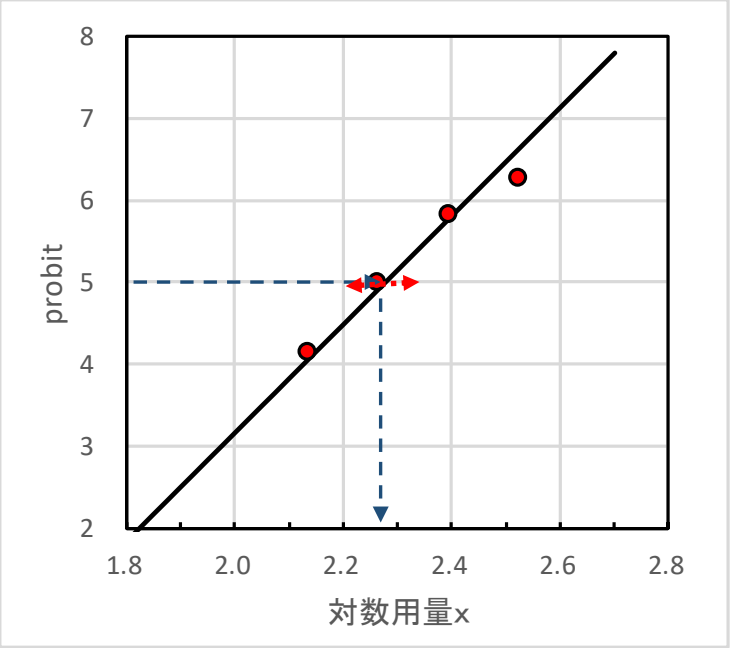


図 2.3 LD_{50} の推定

対数 LD_{50} は，$LD_{50}=(5-\hat{\beta}_0)/\hat{\beta}_1$ でとして推定できるが，分散 $Var(LD_{50})$ は未知であり，$Var(\hat{\beta}_0)$ と $Var(\hat{\beta}_1)$ から近似的（デルタ法）に推定する．$LD_{50}=(5-\hat{\beta}_0)/\hat{\beta}_1$ について $\hat{\beta}_0$ と $\hat{\beta}_1$ で偏微分すると

$$d_0=\frac{\partial LD_{50}}{\partial\hat{\beta}_0}=\frac{\partial\left[(5-\hat{\beta}_0)/\hat{\beta}_1\right]}{\partial\hat{\beta}_0}=-\frac{1}{\hat{\beta}_1}=-0.151$$

$$d_1=\frac{\partial LD_{50}}{\partial\hat{\beta}_1}=\frac{\partial\left[(5-\hat{\beta}_0)/\hat{\beta}_1\right]}{\partial\hat{\beta}_1}=-\frac{(5-\hat{\beta}_0)}{\hat{\beta}_1^2}=-0.344$$

が得られる．これらを縦ベクトル $\boldsymbol{d}$ としたときに LD_{50} 分散は，分散共分散行列 $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ を挟んだ 2 次形式，

-0.151	-0.344		10.320	-4.498		-0.151	=	0.0011
$\boldsymbol{d}^T$			-4.498	1.970		-0.344		
			$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$			$\boldsymbol{d}$		

$$Var(LD_{50})=\boldsymbol{d}^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{d}=0.0011$$

=Mmult(Mmult(Transpose($\boldsymbol{d}$ の範囲) , $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ の範囲) , $\boldsymbol{d}$ の範囲)

として計算される．この分散を元にして，表 2.4 に示すように LD_{50} の近似 95%信頼区間が

L95%： $L95\%=2.278-1.96\sqrt{Var(LD_{50})}=2.278-1.96\times0.0333=2.213$

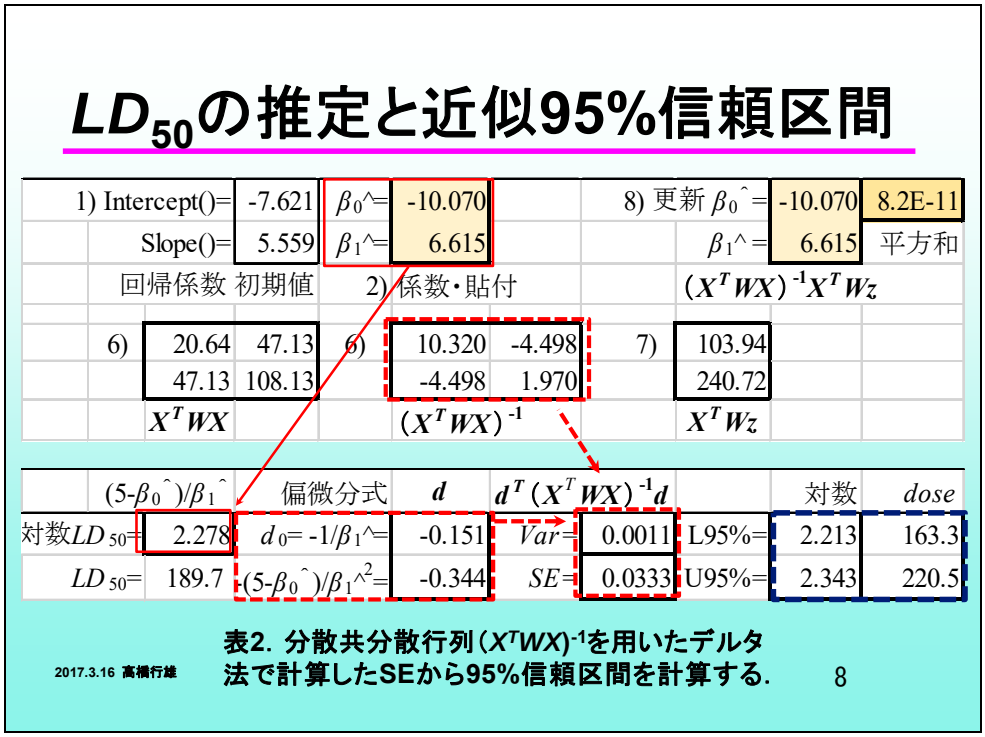
U95%： $L95\%=2.278+1.96\sqrt{Var(LD_{50})}=2.278-1.96\times0.0333=2.343$

であり，元の用量で（163.3，220.5）として計算できる．

表 2.4　LD_{50} および近似 95%信頼区間

	$(5-\beta_0\hat{})/\beta_1\hat{}$	偏微分式	d	$d^T(X^TWX)^{-1}d$			対数	$dose$
対数LD_{50}=	2.278	d_0= -1/β_1^=	-0.151	Var=	0.0011	L95%=	2.213	163.3
LD_{50}=	189.7	-(5-$\beta_0\hat{}$)/β_1^2=	-0.344	SE=	0.0333	U95%=	2.343	220.5

スライド 8



2.6　シグモイド曲線の 95%パーセント信頼区間

回帰分析をした場合には，その結果を図示する際に回帰直線の 95%信頼区間も合わせて示すことが多い．ところが，シグモイド曲線の信頼区間を表示する統計ソフトは少ない．SAS/PROBIT では，図 2.4 に示すように出力できるとの例示がある．しかし，グラフィカルな出力が充実している JMP（Ver13）では，標準出力にはないが，予測式と 95%信頼区間を JMP テーブルに出力できるので，重ね合わせプロットを使って作成することは可能である．

多くの統計ソフトで，パラメータに関する分散共分散行列が出力されるので，これを用いて Excel でシグモイド曲線とその 95%信頼区間を描くことは難しくはない．シグモイド曲線の 95%信頼区間は，プロビット変換した $\hat{\eta}_i$ についての分散を計算する．$Var(\hat{\eta}_i)$ は，

$$Var(\hat{\eta}_i) = Var(\hat{\beta}_0 x_{0,i} + \hat{\beta}_1 x_{1,i})$$

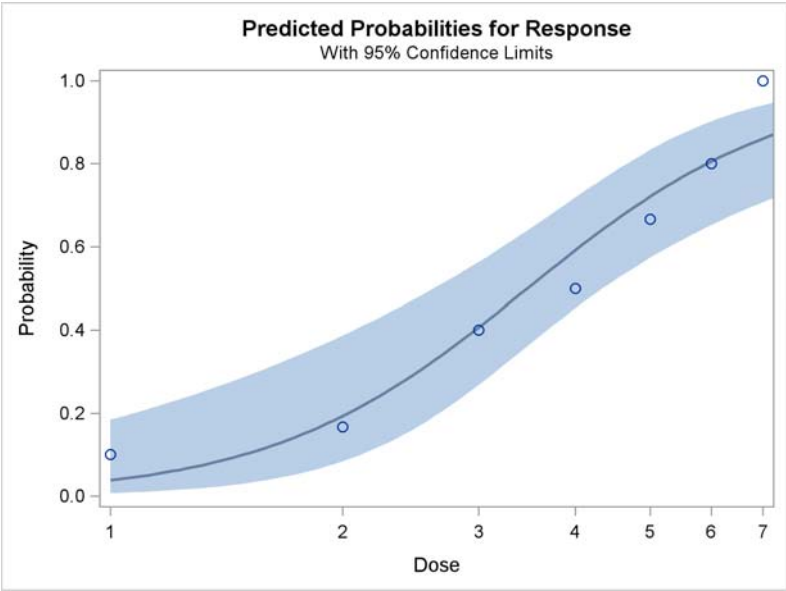


図 2.4　SAS/PROBIT による 95%信頼区間の表示例（SAS 9.3 の User's Guide）

であり，$\hat{\eta}_i = \hat{\beta}_0 x_{0,i} + \hat{\beta}_1 x_{1,i}$ は，パラメータに関する線形式なので，切片 $\hat{\beta}_0$ の分散 $Var(\hat{\beta}_0)$，傾き $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$，それらの共分散 $Cov(\hat{\beta}_1, \hat{\beta}_1)$ から，

$$Var(\hat{\eta}_i) = Var(\hat{\beta}_0 x_{0,i} + \hat{\beta}_1 x_{1,i}) = Var(\hat{\beta}_0) x_{0,i}{}^2 + 2Cov(\hat{\beta}_0, \hat{\beta}_1) x_{0,i} x_{1,i} + Var(\hat{\beta}_1) x_{1,i}{}^2$$

として計算することができる．ところで，分散共分散行列は，

$$(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} = \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} = \begin{bmatrix} 10.320 & -4.498 \\ -4.498 & 1.970 \end{bmatrix}$$

として計算されているので，行列のままで計算したい．そこで，$x_{0,i} = 1$ として，横ベクトル $\boldsymbol{x}_i = (x_{0,i}, x_{1,i})$ に関して分散共分散行列を挟んだ 2 次形式での計算が簡便である．

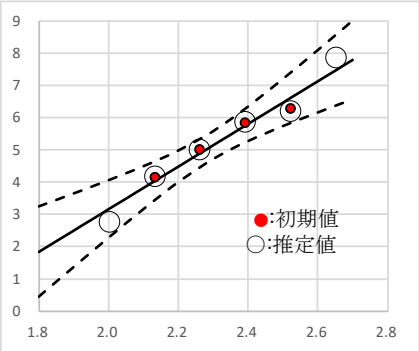
$$Var(\hat{\eta}_1) = \boldsymbol{x}_1 (\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1} \boldsymbol{x}_1{}^T = 0.202$$

$Var(\eta_1)$=	1	2.004		10.320	-4.498	1	=	0.202
	$\boldsymbol{x}_1$			-4.498	1.970	2.004		
				$(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X})^{-1}$		$\boldsymbol{x}_1{}^T$		

プロビット変換値 $\hat{\eta}_1$ に対する直線の 95%信頼区間は，

$$\mathrm{L95\%}(\hat{\eta}_1) = \hat{\eta}_1 - 1.96\sqrt{Var(\hat{\eta}_1)}, \quad \mathrm{L95\%}(\hat{\eta}_1) = 3.189 - 1.96\sqrt{0.202} = 2.307$$
$$\mathrm{U95\%}(\hat{\eta}_1) = \hat{\eta}_1 + 1.96\sqrt{Var(\hat{\eta}_1)}, \quad \mathrm{U95\%}(\hat{\eta}_1) = 3.189 + 1.96\sqrt{0.202} = 4.070$$

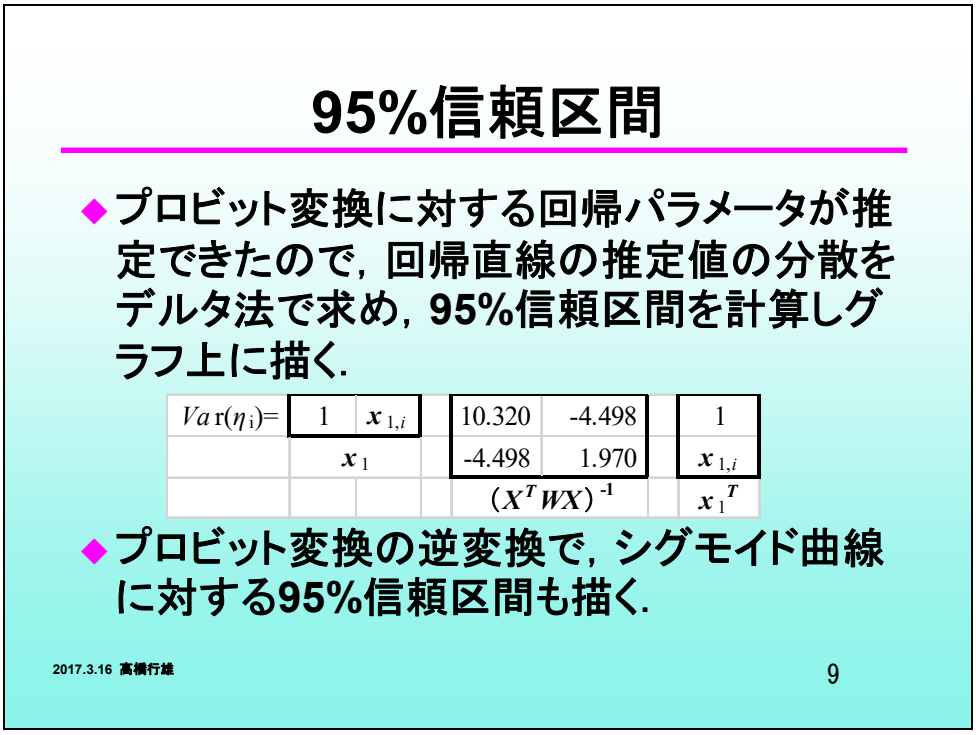
として計算される．他のベクトル $\boldsymbol{x}_i = (1, x_{1,i})$ についても計算し，その結果を図 2.5 に示す．



probit 変換	補正 *probit*	更新 推定値	分散	95%信頼区間	
z^*	4) z	9) $\eta^{(2)}$	$Var(\eta)$	L95%	U95%
-	2.736	3.189	0.202	2.307	4.070
4.158	4.165	4.043	0.093	3.446	4.641
5.000	5.000	4.896	0.049	4.461	5.332
5.842	5.839	5.758	0.072	5.232	6.283
6.282	6.182	6.616	0.161	5.830	7.402
-	7.838	7.481	0.317	6.377	8.585

図 2.5 プロビット変換値に対する初回の重み付き回帰直線と 95%信頼区間

スライド 9



プロビット変換（正規分布のパーセント点）に対する 95%信頼区間を用いて，95%信頼区間の下限値と上限値を逆プロビット変換（累積正規分布）して，出現率に対するシグモイド曲線とその 95%信頼区間の計算は，

$\hat{\mu}_1$ =Normdist(3.189-5,0,1,true)= 0.012

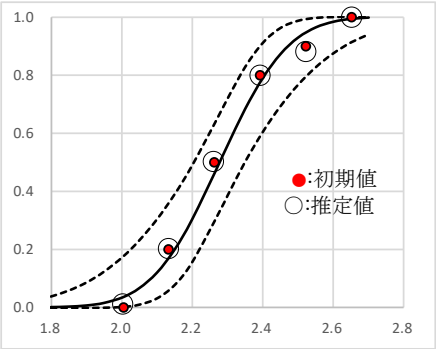
L95%($\hat{\mu}_1$) =Normdist(2.307-5,0,1,true)= 0.004

U95%($\hat{\mu}_1$) =Normdist(3.189-5,0,1,true)= 0.176

として求める．他の点について計算した結果を図 2.6 に示す．

シグモイド曲線を実験結果にあてはめた場合に，実験結果に対して推定されたシグモイド

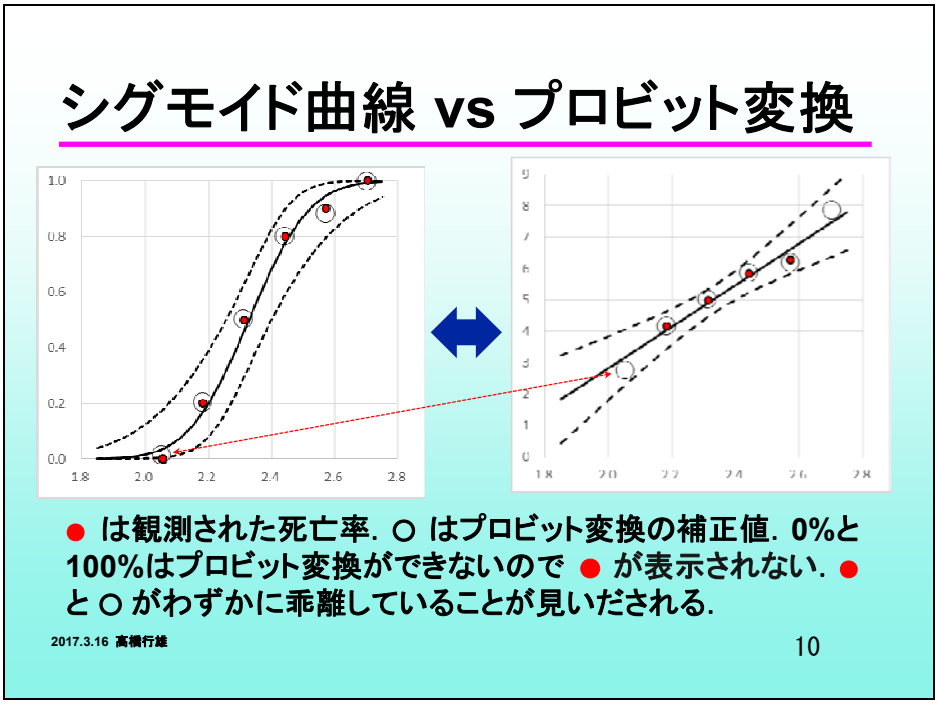
曲線を図 2.6 に示すように，描き，その 95%信頼区間を示すことが，実験結果の適切な理解となる．



出現率	補正出現率	推定出現率	95%信頼区間	
y	p^*	μ^	L95%	U95%
0.000	0.012	0.035	0.004	0.176
0.200	0.202	0.169	0.060	0.360
0.500	0.500	0.459	0.295	0.630
0.800	0.799	0.776	0.592	0.900
0.900	0.881	0.947	0.797	0.992
1.000	0.998	0.993	0.916	1.000

図 2.6 出現率に対するシグモイド曲線と 95%信頼区間

スライド 10



多くの一般化線形モデルに対する統計ソフトの出力は，回帰パラメータの推定と SE の算出が主たる結果であり，各種の推定問題については，対応はまちまちである．また，プロビット法の結果の解釈に際して，回帰係数の統計的意義はあるものの実験結果の解釈の為には全く役に立たない．

統計ソフトの出力に共分散分析行列があれば，この行列を Excel にコピーして実験データの解釈の為の統計量の算出とその 95%信頼区間などの推定を自在にできる．統計ソフトから出力された共分散分析行列の活用については，第 9 章で示す．

2.7　補遺

1）出現率π_iの分散およびプロビット変換したη_iの分散

第 2.3 節で，プロビット変換した場合の分散について，次のように述べた．

広い範囲の出現率πに対して，2 項分布の分散は，$Var(\pi)=\pi(1-\pi)/n$ なので$\pi=0.5$で最大となり等分散ではない．プロビット変換した場合には，逆に $\pi=0.5$ で最小になり，等分散を仮定することはできないので，分散の逆数を重みとする回帰分析が必要となる．

実際に計算してみれば，$n=10$としたときに，

$$\pi_1=0.01 \qquad Var(\pi_1)=0.01\times(1-0.01)/10=0.00099$$

$$\pi_2=0.1 \qquad Var(\pi_2)=0.1\times(1-0.1)/10=0.009$$

$$\pi_3=0.5 \qquad Var(\pi_3)=0.5\times(1-0.5)/10=0.025$$

のように明らかに異なる．プロビット変換は，正規分布のパーセント点に 5 を加えたものなので，

$\pi_1=0.01$　　η_1=Normsinv(0.01)+5= 2.6737

$\pi_2=0.1$　　η_2=Normsinv(0.10)+5= 4.7184

$\pi_3=0.5$　　η_3=Normsinv(0.50)+5= 5.0000

のように計算できる．

出現率 π_iの分散は，$Var(\pi_i)=\pi_i(1-\pi_i)/n_i$ で，プロビット変換は，$\eta_i=\Phi^{-1}(\pi_i)+5$ であり，その分散 $Var(\eta_i)$ を求めたい．合成分散の公式から $Var(\eta_i)$ は，変換式 $\left[\Phi^{-1}(\pi_i)+5\right]$を$\pi_i$ で偏微分し，それを $Var(\pi_i)$ の両側から掛けて求めることができる．

$$Var(\eta_i)=\left[\frac{\partial\Phi^{-1}(\pi_i)}{\partial\pi_i}\right]Var(\pi_i)\left[\frac{\partial\Phi^{-1}(\pi_i)}{\partial\pi_i}\right]$$

となる．正規分布のパーセント点 $\Phi^{-1}(\pi_i)$ の微分は，正規分布の分布関数 $\Phi(\eta_i)$ を微分した場合に図 2.8 に示すように確率密度関数 $\phi(\eta_i)$ となる関係から，その逆数となる．

$$\frac{\partial\Phi^{-1}(\pi_i)}{\partial\pi_i}=\frac{1}{\phi[\eta_i-5]}$$ この関係を用いて，

$$Var(\eta_i)=\left[\frac{\partial\Phi^{-1}(\pi_i)}{\partial\pi_i}\right]^2 Var(\pi_i)=\left[\frac{1}{\phi(\eta_i-5)}\right]^2\frac{\pi_i(1-\pi_i)}{n_i}$$

となる．

図 2.7 に$n_i=10$と固定し，π_iを変化させた場合の$Var(\pi_i)$と$Var(\eta_i)$を示す．$Var(\pi_i)$は，$\pi_i=0$，$\pi_i=1$で 0 となり，$\pi_i=0.5$で最大になる．他方，$Var(\eta_i)$は，0.1～0.9 の範囲でフラットで分散が異なるとは言い切れないが，0 または 1 に近づくにつれ急激に上昇するので，広範なπ_iに対し

ては，等分散性を仮定することはできない．

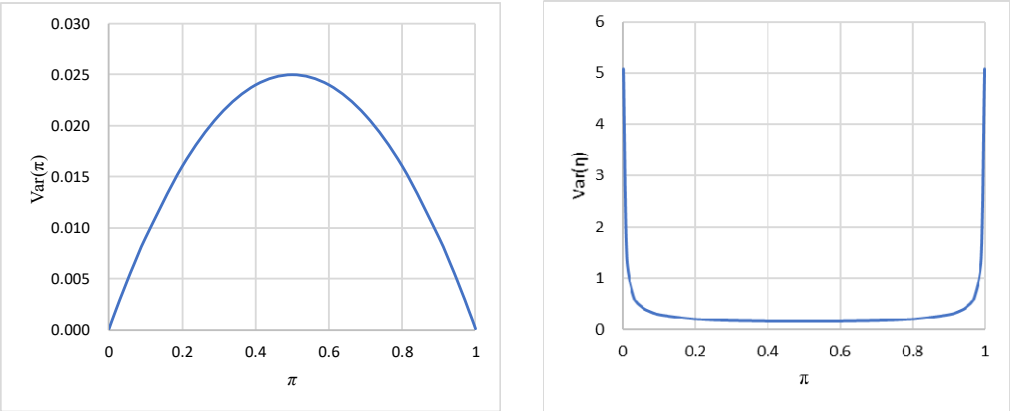


図 2.7　出現率 π_i の分散およびプロビット変換した η_i の分散

2）反復重み付き回帰で用いる補正値 z と重み w

スライド 5 で，プロビット変換でのリンク関数として次式を示した．

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

補正値 z_i については，

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

と天下り的に示した．表記法を統一すると，リンク関数は，

$$g(\pi_i) = \Phi^{-1}(\pi_i) + 5 = \boldsymbol{x}_i^T \boldsymbol{\beta} = \eta_i$$

補正値は，

$$z_i = \eta_i + (p_i - \pi_i)\left(\frac{\partial \eta_i}{\partial \pi_i}\right) = \eta_i + \frac{(p_i - \pi_i)}{\phi(\eta_i - 5)}$$

に対応する．

プロビット変換式 $\eta_i = \Phi^{-1}(\mu_i + 5)$ を μ_i での微分した場合に，$1/\phi(\eta_i - 5)$ との結果を示している．この関係は，図 2.8 の正規分布の分布関数を転置した μ の関数として表すと理解しやすい．正規分布の分布関数 $\Phi(\eta_i)$ の微分は，図中に示した曲線の接線の傾きである．分布関数を転置した曲線の同じ点の傾きは，逆数となることは自明である．分布関数 $\Phi(\eta_i)$ の微分は確率密度関数 $\phi(\eta_i)$ なので，転置した曲線の微分式は，

$$\frac{\partial \eta_i}{\partial \mu_i} = \frac{1}{\phi\left[\Phi^{-1}(\mu_i)\right]} = \frac{1}{\phi(\eta_i - 5)}$$

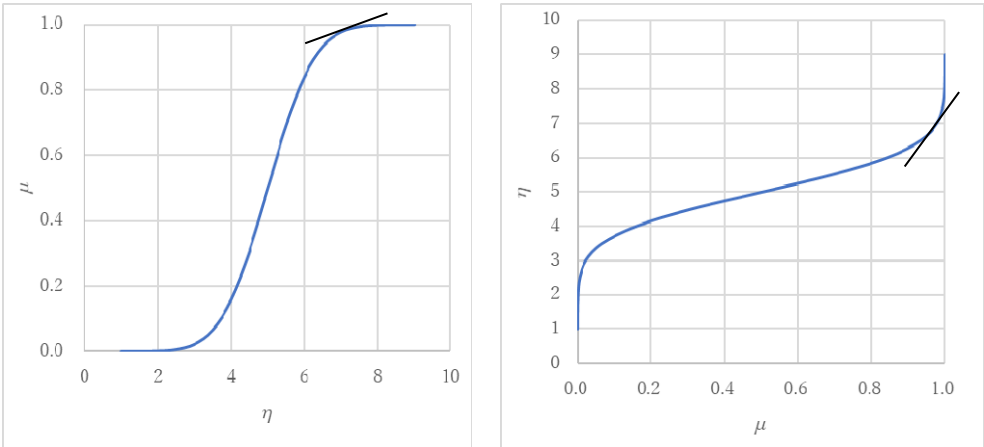


図 2.8　正規分布の分布関数の転置した正規分布のパーセント点 μ の関数

となる．

重み w_i は

$$w_i = \frac{1}{Var(\mu_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \frac{n_i\left[\phi(\eta_i - 5)\right]^2}{\mu_i(1-\mu_i)}$$

で与えた．ところで，重み w_i は，分散の逆数なので，次式で，これまでの微分式を用いて，次のようになる．

$$Var(\eta_i) = \frac{1}{w_i} = Var(\mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2 = Var(\mu_i)\left(\frac{\partial \Phi^{-1}(\mu_i)}{\partial \mu_i}\right)^2 = \frac{\mu_i(1-\mu_i)}{n_i\left[\phi(\eta_i - 5)\right]^2}$$

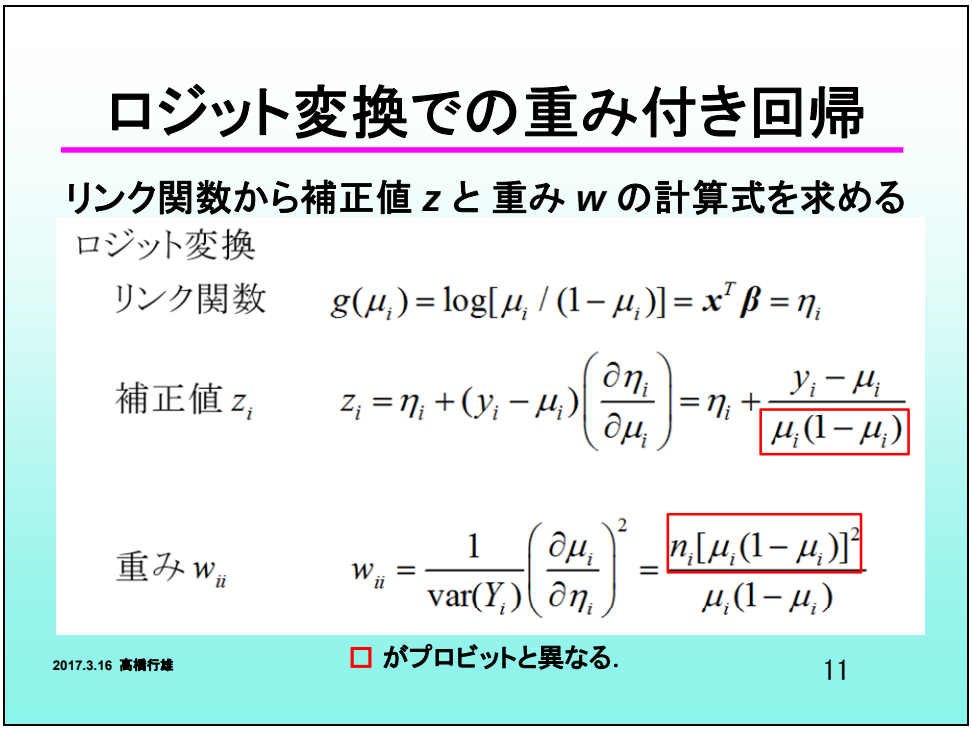
この式は，1）でのプロビット変換した場合の分散と一致する．

3.　反復重み付き回帰によるロジット法

3.1　ロジット変換での重み付き回帰

死亡率に対する用量反応曲線に正規分布の積分として定義されている累積分布を用いるプロビット法に代わり，累積分布が関数として定義されるロジスティック分布を用いたロジット法がプロビット法に代わり使われるようになってきた．スライド 11 に示すように，ロジット法を用いた LD_{50} とその信頼区間の推定するためのリンク関数 $g(\mu_i)$ および補正値 z_i は，田中ら訳（2008）に示されている一般式(4.24)および式(4.23)から容易に導出できる．詳しくは，第 3.3 節 補遺を参照のこと．

スライド 11



スライド中の y_i は，本文中では全体の整合性を図るために p_i とし，Y_i は μ_i とした．

表 2.2 のプロビット法の Excel シートを表 3.1 に示すようにロジット法による反復重み付き回帰となるように変更する．

手順 1）プロビット変換の z^* をロジット変換式で置き換える．ここで，ロジットに+5 を加えたのは，伝統的なプロビット法との結果の比較のためである．

$$z_2{}^* = \log\left(\frac{p_2}{1-p_2}\right) + 5 = \mathtt{Ln(0.200/(1-0.200))+5=3.614}$$

:

$$z_2{}^* = \log\left(\frac{p_2}{1-p_2}\right) + 5 = \mathtt{Ln(0.900/(1-0.900))+5=7.197}$$

すると，Intercept 及び Slope 関数により，回帰係数が計算されるので，隣の $\hat{\beta}_0$ と $\hat{\beta}_1$ の欄に値のみを張りつける．この結果として，回帰係数の推定値 η が更新される．

Intercept()=	-16.232	β_0^=	-16.232	
Slope()=	9.357	β_1^=	9.357	
回帰係数 初期値			係数 貼付	

手順 2）回帰係数の推定値 η を逆ロジット変換で出現率の推定値 μ を計算する．

$$\mu_1 = \frac{\exp(\eta_1 - 5)}{1+\exp(\eta_1 - 5)} = \mathtt{Exp(2.523-5)/(1+Exp(2.523-5))=0.077}$$

:

$$\mu_6 = \frac{\exp(\eta_6 - 5)}{1+\exp(\eta_6 - 5)} = \mathtt{Exp(8.594-5)/(1+Exp(8.594-5))=0.973}$$

表 3.1　ロジット法の初期値に対する重み付き回帰

	dose	切片	対数用量	出現数	例数	出現率	ロジット変換		回帰推定値	出現率推定値	補正 *logit*	重み	更新推定値	差
i	*dose*	x_0	x_1	*y*	*n*	*p*	1) z^*		1) $\eta^{(1)}$	2) μ	3) *z*	3) *w*	4) $\eta^{(2)}$	$\eta^{(1)}$-$\eta^{(2)}$
1	101	1	2.004	0	10	0.000	-		2.523	0.077	1.439	0.715	1.997	0.526
2	136	1	2.134	2	10	0.200	3.614		3.732	0.220	3.618	1.714	3.434	0.297
3	183	1	2.262	5	10	0.500	5.000		4.938	0.485	5.000	2.498	4.868	0.070
4	247	1	2.393	8	10	0.800	6.386		6.157	0.761	6.372	1.820	6.317	-0.160
5	333	1	2.522	9	10	0.900	7.197		7.371	0.915	7.184	0.781	7.760	-0.390
6	450	1	2.653	10	10	1.000	-		8.594	0.973	9.622	0.260	9.215	-0.621
		X			Intercept()=		-16.232	β_0^=	-16.232		4) 更新 β_0^=		-20.298	9.3E-01
	デザイン行列				Slope()=		9.357	β_1^=	9.357			β_1^=	11.123	平方和
					回帰係数 初期値			1)	係数 貼付			$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$		

手順 3）ロジット法でのリンク関数は

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{x}^T\boldsymbol{\beta} = \eta_i$$

なので，補正値 z_i および重み w_{ii} は，

$$z_i = \eta_i + (p_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) = \eta_i + \frac{p_i - \mu_i}{\mu_i(1-\mu_i)}$$

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

と導出される．プロビット法の Excel の計算式を置き換える．

$z_1 = \eta_1 + \frac{p_1 - \mu_1}{\mu_1(1-\mu_1)}$ =2.523+(0.000-0.007)/(0.007*(1-0.007))

=1.439

:

$z_6 = \eta_6 + \frac{p_6 - \mu_6}{\mu_6(1-\mu_1)}$ =8.594+(1.000-0.973)/(0.973*(1-0.973))

=9.622

$w_{11} = \frac{n_i[\mu_i(1-\mu_i)]^2}{\mu_i(1-\mu_i)}$ =10*(0.077*(1-0.077))^2/(0.077*(1-0.077))

=0.715

:

$w_{66} = \frac{n_6[\mu_6(1-\mu_6)]^2}{\mu_6(1-\mu_6)}$ =10*(0.973*(1-0.973))^2/(0.973*(1-0.973))

=0.260

手順 4）積和行列を $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} = (\boldsymbol{X} * \boldsymbol{w})^T \boldsymbol{X}$ で計算し，Minverse 関数で分散共分散行列 $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ および $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z} = (\boldsymbol{X} * \boldsymbol{w})^T \boldsymbol{z}$ を計算する．回帰係数を

$$\hat{\boldsymbol{\beta}}^{\text{new}} = \left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\right]\left[\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}\right]$$

で計算し，その結果が［更新］した回帰係数として求める．

Intercept()=	-16.232	β_0^=	-16.232		4) 更新 β_0^=	-20.298	9.3E-01	
Slope()=	9.357	β_1^=	9.357		β_1^=	11.123	平方和	
回帰係数 初期値		1)	係数 貼付		$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$			
7.79	17.76		26.498	-11.566	39.43			
17.76	40.68		-11.566	5.073	92.10			
$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$			$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$		$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$			

手順 5a)第 1 サイクル目で得られた回帰パラメータを, 1)係数貼付欄に値のみを張りつける．自動的に第 2 サイクルの重み付き回帰が実行される．同様に繰返すと第 5 サイクルで，更新された係数に変化が無くなるので，解が求まったとする．

第 2 サイクル目：

β_0^=	-16.232		4) 更新 β_0^=	-20.298
β_1^=	9.357		β_1^=	11.123

第 3 サイクル目：	β_0^=	-20.298		4) 更新 β_0^=	-21.177
	β_1^=	11.123		β_1^=	11.508
第 4 サイクル目：	β_0^=	-21.177		4) 更新 β_0^=	-21.211
	β_1^=	11.508		β_1^=	11.523
第 5 サイクル目：	β_0^=	-21.211		4) 更新 β_0^=	-21.212
	β_1^=	11.523		β_1^=	11.523

表 3.2　ロジット法の 6 サイクル目での収束結果

	dose	切片	対数用量	出現数	例数	出現率	ロジット変換		回帰推定値	出現率推定値	補正 *logit*	重み	更新推定値	差
i	*dose*	x_0	x_1	y	n	p	1) z^*		1) $\eta^{(1)}$	2) μ	3) z	3) w	4) $\eta^{(2)}$	$\eta^{(1)}$-$\eta^{(2)}$
1	101	1	2.004	0	10	0.000	-		1.884	0.042	0.840	0.407	1.884	0.000
2	136	1	2.134	2	10	0.200	3.614		3.373	0.164	3.633	1.373	3.373	0.000
3	183	1	2.262	5	10	0.500	5.000		4.859	0.465	5.000	2.488	4.859	0.000
4	247	1	2.393	8	10	0.800	6.386		6.359	0.796	6.386	1.626	6.359	0.000
5	333	1	2.522	9	10	0.900	7.197		7.854	0.946	6.970	0.515	7.854	0.000
6	450	1	2.653	10	10	1.000	-		9.361	0.987	10.374	0.124	9.361	0.000
			X		Intercept()=		-16.232	β_0^=	-21.212		4) 更新 β_0^=		-21.212	6.8E-22
			デザイン行列		Slope()=		9.357	β_1^=	11.523			β_1^=	11.523	平方和
					回帰係数 初期値			1)	係数 貼付			$(X^TWX)^{-1}X^TWz$		
						6.53	14.89		40.328	-17.623		33.03		
						14.89	34.08		-17.623	7.731		76.79		
						X^TWX			$(X^TWX)^{-1}$			X^TWz		

3.2　ロジット変換での 50%出現率の推定

ロジット法での回帰係数は，プロビット法の場合と全く異なるが，表 3.3 に示すように LD_{50} =188.2，95%信頼区間は (161.5, 219.4) とほぼ同じ結果となる．このことから，LD_{50} の算出に際しては，どちらを選択しても本質は同じである．

表 3.3　ロジット法による LD_{50} の推定値と近似 95%信頼区間

	$(5-\beta_0\hat{})/\beta_1\hat{}$	偏微分式	d	$d^T(X^TWX)^{-1}d$		対数	*dose*
対数LD_{50}=	2.275	d_0= -1/β_1^=	-0.087	*Var*=	0.0012	*L*95%=	2.208	161.49
LD_{50}=	188.2	-(5-$\beta_0\hat{})/\beta_1$^2=	-0.197	*SE*=	0.0340	*U*95%=	2.341	219.44

計算の見通しの観点から，プロピット法は，正規分布の確率密度関数および積分で定義されている累積分布関数を用いているために，微分などの計算では煩雑となり見通しが悪くなる．他方，ロジット法では，確率密度関数および累積分布関数対数が，対数と指数の関数であり，微分などの計算が容易である．

回帰係数の推定値およびその SE を 表 3.4 に示す. 統計的にはこれらの推定値の 95%信頼区間などで，各種の考察をしたいのであるが，これらの係数はロジスティック曲線を推定するための計算上の便宜のために使われたもので，解釈不能である．無理やり説明すると，「$\hat{\beta}_0 = -21.212$ は，常用対数が 0.0（元の用量で 1.0 mg/kg）でのロジットの切片であり出現率は逆ロジットで6.134×10^{-8}%となる．$\hat{\beta}_1 = 11.523$は，単位対数用量あたりのロジットの増分である.」とのように，実験結果の要約統計量としては，全く意味をなさない．

表 3.4　ロジット法による回帰係数の推定値と *SE*

係数	推定値	分散	SE
β_0^	-21.212	40.328	6.350
β_1^	11.523	7.731	2.780

ロジット法についての各種の推定については，プロビット法で示した計算方法と同じであり，第 5 章の情報行列を用いた最尤法で詳細に示すので，ここでは割愛する．

3.3　補遺

ロジット法の補正値 z_i および重み w_{ii} は，

$$z_i = \eta_i + (p_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) = \eta_i + \frac{p_i - \mu_i}{\mu_i(1-\mu_i)}$$

$$w_{ii} = \frac{1}{Var(\pi_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \frac{n_i[\mu_i(1-\mu_i)]^2}{\mu_i(1-\mu_i)}$$

と導出される．ロジット法でのリンク関数は

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{x}^T\boldsymbol{\beta} = \eta_i$$

なので，微分法の公式を用いて

$$
\begin{aligned}
\frac{\partial \eta_i}{\partial \mu_i} &= \frac{\partial \log\left(\mu_i / (1-\mu_i)\right)}{\partial \mu_i} \\
&= \frac{1/(1-\mu) + \mu_i / (1-\mu_i)^2}{\mu_i / (1-\mu_i)} \\
&= \frac{1}{\mu_i (1-\mu_i)}
\end{aligned}
$$

と計算される．重み w_i は，分散の逆数なので，次式で，これまでの微分式を用いて，次のようになる．

$$
Var(\eta_i) = Var(\pi_i)\left[\frac{1}{\mu_i(1-\mu_i)}\right]^2 = \frac{\mu_i(1-\mu_i)}{n_i[\mu_i(1-\mu_i)]^2}
$$

この式から，分散の逆数 w_i は，

$$
w_i = \frac{1}{Var(\eta_i)} = \frac{1}{Var(\pi_i)}\left[\mu_i(1-\mu_i)\right]^2 = \frac{n_i[\mu_i(1-\mu_i)]^2}{\mu_i(1-\mu_i)}
$$

となる．

4. 反復重み付き回帰による補 2 重対数法

4.1 最小極値分布とワイブル分布の関係

低用量の出現率が裾を引く場合には，プロビット法およびロジット法では，推定値が低め（過小評価）となりがちである．そのために，左に裾を引く分布をシグモイド曲線のための関数が必要となる．最小値の分布関数として知られている最小極値分布がこの性質を持っている．

最小極値分布は，ワイブル分布と密接な関係がある．対数正規分布するような変数 x について，対数を取ると正規分布になると同様に，ワイブル分布に従う変数 x について，対数を取ると最小極値分布となる．したがって，最小極値分布は，対数ワイブル分布ともいえる．最小極値分布は，正規分布と同様に一山型の確立密度関数を持ち，累積分布関数は累積正規分布と同様にシグモイド曲線として用いられている．

ワイブル分布は，寿命データの分布として知られているが，最小極値分布については，対数を 2 重に取る（補 2 重対数変換）と線形化できることが知られていて，ワイブル確率プロットで活用されているが，最小極値分布は黒子の存在となっている．

寿命時間解析で用いられているワイブル分布の時間 t の対数を取り，対数時間とした場合に最小極値分布となる．ワイブル分布は，寿命時間解析では知られているが，最小極値分布の認知度は低いので，ワイブル分布と最小極値分布の関係について示す．

ワイブル分布の累積分布関数 $F(t)$ は，t を寿命時間とし，

$$F(t) = 1 - \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$$

で与えられる，尺度（scale）パラメータを α，形状（shape）パラメータを β といい，共に正である．確率密度関数 $f(t)$ は，累積分布関数 $F(t)$ を t で微分して

$$f(t) = \frac{\beta}{\alpha^{\beta}} t^{\beta-1} \exp\left\{-\left(\frac{t}{\alpha}\right)^{\beta}\right\}$$

が得られる．累積分布関数 $F(t)$ に $t = \alpha$ を代入すると β にかかわらず，

$$F(t = \alpha) = 1 - e^{-1} = 0.632$$

と計算される．したがって，ワイブル分布の尺度パラメータ α は，下側確率 0.632 となり，おおよその “平均” を示す．

図 4.1 に尺度パラメータを $\alpha=10$ と固定し，形状パラメータを $\beta=0.5, 1, 2, 4, 6$ と変化させたときのワイブル分布の確率密度関数（左）と累積分布関数（右）を Excel の Weibull 関数を用いて計算し，Excel により作成した線グラフを示す．

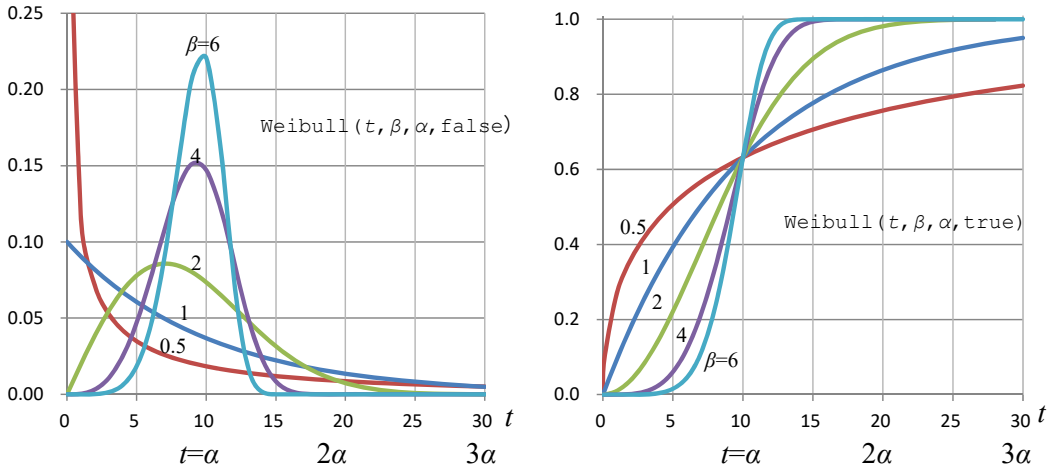


図 4.1　ワイブル分布の確率密度関数（左）と累積分布関数（右）

ワイブル分布の尺度パラメータを $\alpha=10$ としたので，累積分布関数では $t=10$ で形状パラメータが $\beta=0.5, 1, 2, 4, 6$ と異なっても同じ下側確率 0.632 を通ることが図（右）から確認される．確率密度関数の場合は，$\beta\le 1$ 以下の場合に単調減少の形状となり，$\beta>1$ の場合には一山型の形状で，山のピークは右に移動する．形状パラメータが $\beta=4$ の場合には正規分布に近く，$\beta>4$ の場合には，確率密度関数の山がさらに高く，左に裾を引くようになり，幅も狭まる．

4.2　ワイブル分布を最小極値分布へ変換

ワイブル分布の時間 t について $y=\ln(t)$ と自然対数を取り，Weibull α も自然対数 $\mu_{SEV}=\ln(\alpha)$ とし，Weibull β の逆数を $\sigma_{\mathrm{SEV}}=\delta=1/\beta$ とすることにより，正規分布と同様の統計的な性質を持つ最小極値分布（SEV : Smallest Extreme Value Distribution）に置き換えることができる．これらをワイブル分布の累積分布関数 $F(t)$ に $\alpha=\exp(\mu_{SEV})$，$\beta=1/\sigma_{SEV}$ を代入する．最小極値分布のパラメータは，μ_{SEV}，σ_{SEV} とするのが適切だが，文脈の中で識別できると思われる場合には，“SEV” は除くことにする．

$$
\begin{aligned}
F(t) &= 1-\exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right] \\
&= 1-\exp\left[-\left(\frac{t}{\exp(\mu)}\right)^{1/\sigma}\right] \\
&= 1-\exp\left\{-\exp\left[\ln\left(\frac{t}{\exp(\mu)}\right)^{1/\sigma}\right]\right\} \\
&= 1-\exp\left[-\exp\left(\frac{\ln(t)-\mu}{\sigma}\right)\right]
\end{aligned}
$$

ここで，$x=\ln(t)$ として，次の最小極値分布

$$
F(y) = 1-\exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right] \quad , \quad -\infty < y < \infty
$$

を得る．

ワイブル分布と最小極値分布の関係は，正規分布と対数正規分布の関係と同様である．このような対数変換を行うことにより，ワイブル分布を正規分布と同様に，“平均値”として μ_{SEV} が，“標準偏差”として σ_{SEV} を用いて表すことができる．

最小極値分布の確率密度関数 $f(y)$ は，累積分布関数 $F(x)$ を x で微分することにより，

$$
f(x) = \frac{1}{\sigma}\cdot\exp\left(\frac{x-\mu}{\sigma}\right)\cdot\exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\}, \quad -\infty < y < \infty
$$

が得られる．パラメータ $\mu_{SEV}=2.0$ としてパラメータ σ_{SEV} を（1.5，1.0，0.8，0.4）と変えたときの確率密度関数および累積分布関数を 図 4.2 に示す．

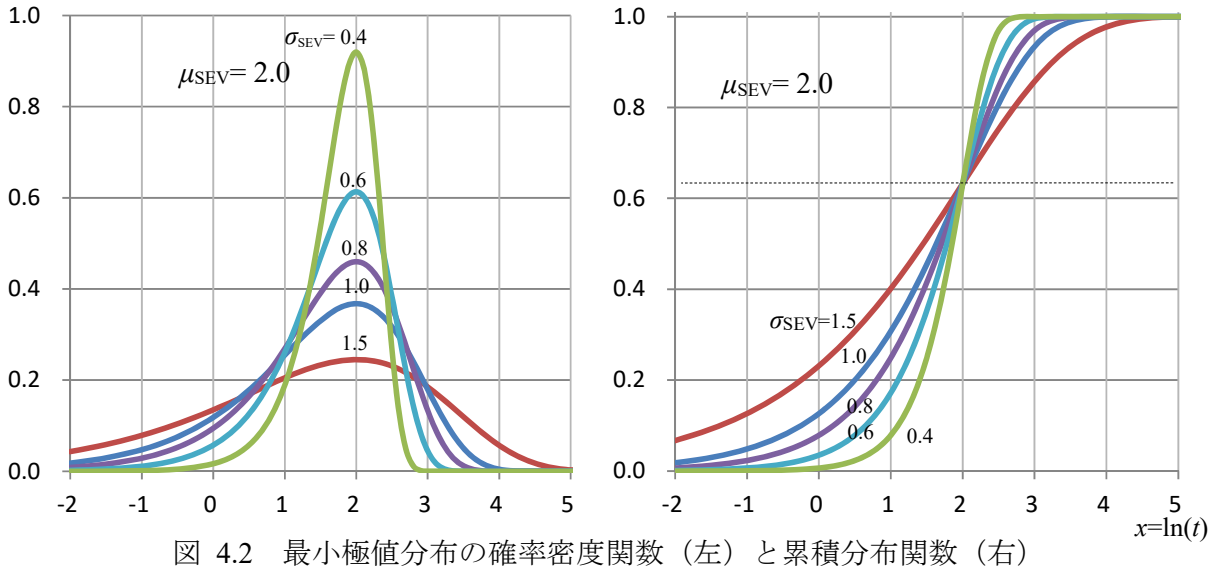


図 4.2　最小極値分布の確率密度関数（左）と累積分布関数（右）

ワイブル分布の確率密度関数の形状は，形状パラメータ $\beta \le 1$ の場合に単調減少，$\beta > 1.0$ の場合は右に裾をひく分布で一山型ではあるがその位置が一定ではなかった．時間について自然対数を取ることにより $\sigma_{SEV} = 1/\beta$ の大きさにかかわらず確率密度関数の最大値は，$\mu_{\mathrm{SEV}} = 2.0$ の位置となり，$\sigma_{SEV} = 1.0$ の場合には，左に裾は引くものの正規分布の確率密度関数の形状に近い．

累積分布関数は，ワイブル分布の場合には，図 4.1 に示したように Weibull $\beta \le 1$ の場合に上に凸の単調増加，$\beta > 1.0$ の場合に S 字型であったが，最小極値分布では，$\sigma_{SEV} = 1/\beta$ の大きさにかかわらず S 字型となっている．また，$y = \mu_{SEV} = 2.0$ における下側確率は，

$$F(x=\mu) = 1 - \exp\left[-\exp\left(\frac{\mu-\mu}{\sigma}\right)\right] = 1 - e^{-1} = 0.6321$$

と一定であることは，ワイブル分布の $t = \alpha$ 場合に 0.632 になることと同じである．

このように最小極値分布は，パラメータ μ_{SEV} および σ_{SEV} に関して正規分布と同様の性質を持つために，寿命データの解析に際してワイブル分布のままでは推定が困難な各種の信頼区間の推定に際して黒子的な役割を果している．

最小極値分布

$$F(x) = 1 - \exp\left[-\exp\left(\frac{x-\mu}{\sigma}\right)\right]$$

の 1 を左辺に移行しマイナス 1 を掛け，両辺に対数を取ると

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

となり，さらに両辺に対数を取ると，線形化することができる．

$$\ln\left[-\ln\left(1-\pi\right)\right] = \beta_0 + \beta_1 x$$

$$\text{ただし，}\ \pi = F(x),\ \ \beta_0 = -\frac{\mu}{\sigma}\ \ ,\ \ \beta_1 = \frac{1}{\sigma}$$

最小極値分布は，2 重に対数を取ることにより線形化できるので，プロビット法あるいはロジット法に対応して補 2 重対数法とも称されている．

4.3　最小極値分布をシグモイド曲線とする補 2 重対数法

用量反応曲線のための累積最小極値分布は，

$$F_{SEV}(\mu) = 1 - \exp[-\exp(\beta_0 + \beta_1 x)]$$

であり，スライド 12 に示すように 2 重に対数を取ることにより，リンク関数は，

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

となり，補正値 z_i および重み w_{ii} を，

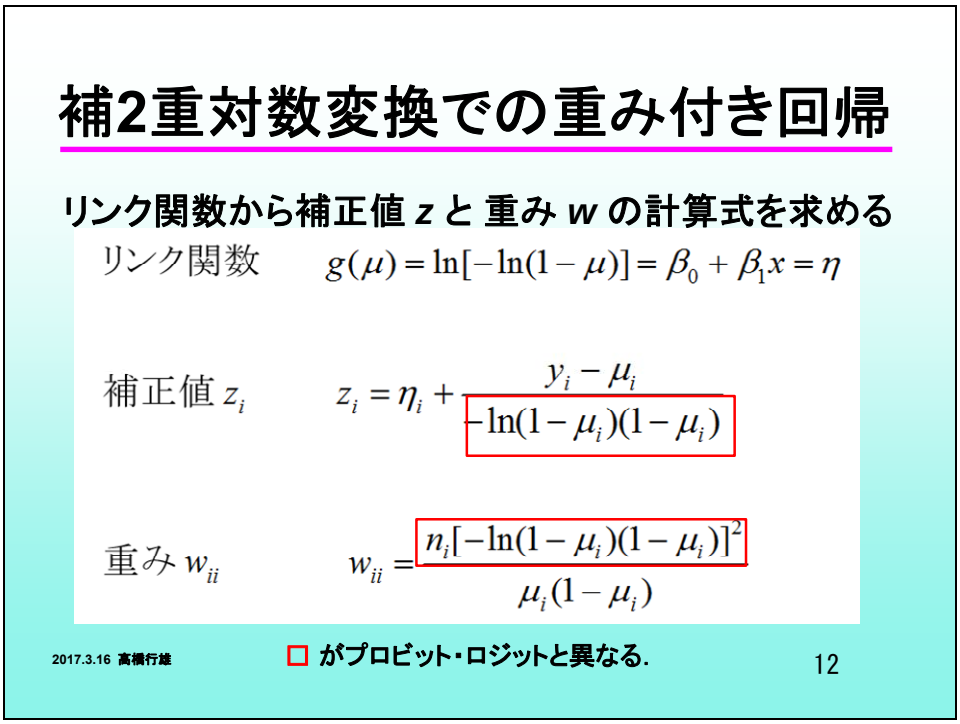
$$z_i = \eta_i + (p_i - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) = \eta_i + \frac{p_i - \mu_i}{-\ln(1-\mu_i)(1-\mu_i)}$$

$$w_{ii} = \frac{1}{Var(\pi_i)}\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 = \frac{n_i[-\ln(1-\mu_i)(1-\mu_i)]^2}{\mu_i(1-\mu_i)}$$

ただし，$\dfrac{\partial \eta_i}{\partial \mu_i} = \dfrac{1}{-\ln(1-\mu_i)} \cdot \dfrac{1}{1-\mu_i} = \dfrac{1}{-\ln(1-\mu_i)(1-\mu_i)}$

として，ロジット法と同様に計算できる．

スライド 12



スライド中の y_i は，本文中では全体の整合性を図るために p_i とし，Y_i は π_i とした．

表 3.1 のロジット法の Excel シートを 表 4.1 に示すように補 2 重対数法による反復重み付き回帰となるように変更する．

手順 1）ロジット変換の $z*$ を補 2 重対数変換式で置き換える．ここで，補 2 重対数に+5 を加えたのは，伝統的なプロビット法との結果の比較のためである．

$z_2* = 5 + \log[-\log(1-p_2)]$ =5+Ln(-Ln(1-0.200))=3.500

：

$$z_5^* = 5 + \log\left[-\log(1-p_5)\right] = 5+\mathrm{Ln}(-\mathrm{Ln}(1-0.900)) = 5.834$$

すると，Intercept 及び Slope 関数により，回帰係数が計算されるので，隣の $\hat{\beta}_0$ と $\hat{\beta}_1$ の欄に値のみを張りつける．この結果として，回帰係数の推定値 η が更新される．

Intercept()=	-9.214	β_0^=	-9.214	
Slope()=	6.046	β_1^=	6.046	
回帰係数 初期値			係数 貼付	

手順 2） 回帰係数の推定値 η を逆補 2 重対数変換（累積最小極値分布）で出現率の推定値 μ を計算する．

$$\mu_1 = 1 - \exp\left[-\exp(\eta_1 - 5)\right] = 1-\mathrm{Exp}(-\mathrm{Exp}(2.523-5)) = 0.116$$

$$\vdots$$

$$\mu_6 = 1 - \exp\left[-\exp(\eta_6 - 5)\right] = 1-\mathrm{Exp}(-\mathrm{Exp}(6.829-5)) = 0.998$$

表 4.1 補 2 重対数法の初期値に対する重み付き回帰の結果

	dose	切片	対数用量	出現数	例数	出現率	補2重変換		回帰推定値	出現率推定値	補正補2重	重み	更新推定値	差
i	$dose$	x_0	x_1	r	n	y	z^*		$\eta^{(1)}$	μ	z	w	$\eta^{(2)}$	$\eta^{(1)}-\eta^{(2)}$
1	101	1	2.004	0	10	0.000	-		2.905	0.116	1.841	1.157	2.643	0.262
2	136	1	2.134	2	10	0.200	3.500		3.686	0.236	3.512	2.343	3.510	0.177
3	183	1	2.262	5	10	0.500	4.633		4.466	0.444	4.639	4.311	4.373	0.092
4	247	1	2.393	8	10	0.800	5.476		5.253	0.724	5.467	6.319	5.246	0.007
5	333	1	2.522	9	10	0.900	5.834		6.038	0.941	5.796	5.035	6.116	-0.078
6	450	1	2.653	10	10	1.000	-		6.829	0.998	6.989	0.769	6.992	-0.164
		$\boldsymbol{X}$			Intercept()=		-9.214	β_0^=	-9.214			更新 β_0^=	-10.790	1.4E-01
	デザイン行列				Slope()=		6.046	β_1^=	6.046			β_1^=	6.702	平方和
					回帰係数 初期値				係数 貼付			$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$		

手順 3） 補 2 重対数法でのリンク関数は

$$g(\mu_i) = \log\left[-\log(1-\mu_i)\right] = \boldsymbol{x}^T\boldsymbol{\beta} = \eta_i$$

なので，補正値 z_i および重み w_{ii} は，

$$z_1 = \eta_1 + \frac{y_i - \mu_i}{-\log(1-\mu_i)(1-\mu_i)}$$

$$=2.523+(0.000-0.116)/(-\mathrm{Ln}(1-0.116)*(1-0.116)) = 1.841$$

$$\vdots$$

$$w_{11} = \frac{-n_i\left[-\log(1-\mu_i)(1-\mu_i)\right]^2}{\mu_i(1-\mu_i)}$$

$$=-10*(-\mathrm{Ln}(1-0.116)*(1-0.116))^2/(0.116*(1-0.116)) = 0.715$$

$$\vdots$$

手順 4）積和行列を $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} = (\boldsymbol{X} * \boldsymbol{w})^T \boldsymbol{X}$ で計算し，Minverse 関数で分散共分散行列 $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ および $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z} = (\boldsymbol{X} * \boldsymbol{w})^T \boldsymbol{z}$ を計算する．回帰係数を

$$\hat{\boldsymbol{\beta}}^{\text{new}} = \left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\right]\left[\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}\right]$$

で計算し，その結果が［更新］した回帰係数として求められている．

Intercept()=	-9.214	β_0^=	-9.214		更新 β_0^=	-10.790	1.4E-01	
Slope()=	6.046	β_1^=	6.046		β_1^=	6.702	平方和	
回帰係数 初期値			係数 貼付		$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$			
19.93	46.93		10.864	-4.593	99.45			
46.93	111.00		-4.593	1.951	237.58			
$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$			$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$		$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$			

手順 5a）第 1 サイクル目で得られた回帰パラメータを， 1）係数貼付欄に値のみを張りつける．自動的に第 2 サイクルの重み付き回帰が実行される．同様に繰返すと第 5 サイクルで，更新された係数に変化が無くなるので，解が求まったとする．

第 2 サイクル：	β_0^=	-10.790		更新 β_0^=	-10.893
	β_1^=	6.702		β_1^=	6.743

第 3 サイクル：	β_0^=	-10.893		更新 β_0^=	-10.892
	β_1^=	6.743		β_1^=	6.743

第 4 サイクル：	β_0^=	-10.892		更新 β_0^=	-10.893
	β_1^=	6.743		β_1^=	6.743

表 4.2 に示すように補 2 重対数法での回帰係数は，ロジット法の場合と全く異なるが，$LD_{50} = 227.5$，95%信頼区間は (194.4, 266.2) とやや高めに推定される．これは，投与量の低い場合にはロジット法に比べ裾が長い分，高用量では裾が短くなるためである．なお，ロジット法との比較検討は，情報行列を用いた最尤法の場合について第 8 章で示す．

表 4.2　補 2 重対数法による LD_{50} の推定値と近似 95%信頼区間

	dose	切片	対数用量	出現数	例数	出現率	SEV変換		回帰推定値	出現率推定値	補正SEV	重み	更新推定値	差
i	$dose$	x_0	x_1	r	n	y	z^*		$\eta^{(1)}$	μ	z	w	$\eta^{(2)}$	$\eta^{(1)}-\eta^{(2)}$
1	101	1	2.004	0	10	0.000	-		2.623	0.089	1.575	0.886	2.623	0.000
2	136	1	2.134	2	10	0.200	3.500		3.494	0.199	3.500	1.981	3.494	0.000
3	183	1	2.262	5	10	0.500	4.633		4.363	0.411	4.649	4.013	4.363	0.000
4	247	1	2.393	8	10	0.800	5.476		5.241	0.720	5.466	6.302	5.241	0.000
5	333	1	2.522	9	10	0.900	5.834		6.116	0.953	5.750	4.618	6.116	0.000
6	450	1	2.653	10	10	1.000	-		6.998	0.999	7.134	0.341	6.998	0.000
		$\boldsymbol{X}$			Intercept()=		-9.214	β_0^=	-10.893			更新 β_0^=	-10.893	2.5E-10
		デザイン行列			Slope()=		6.046	β_1^=	6.743			β_1^=	6.743	平方和
					回帰係数		初期値		係数	貼付		$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$		
						18.14	42.71		13.604	-5.755		90.42		
						42.71	100.98		-5.755	2.444		215.65		
						$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$			$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$			$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{z}$		
					$(5-\beta_0\hat{})/\beta_1\hat{}$		偏微分式		$\boldsymbol{d}$	$\boldsymbol{d}^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{d}$			対数	$dose$
				対数LD_{50}=		2.357	d_0= -1/β_1^=		-0.148	Var=	0.0012	L95%=	2.289	194.37
				LD_{50}=		227.5	$(5-\beta_0\hat{})/\beta_1$^2=		-0.350	SE=	0.0348	U95%=	2.425	266.16

5. 情報行列を用いたロジット法によるシグモイド曲線の推定

5.1　情報行列を用いたロジット法

一般化線形モデルに対する最尤法にパラメータの推定には，反復重み付き回帰による方法と情報行列を用いた 2 つの方法がある．これまで示してきた反復重み付き回帰による方法は，各種の累積分布関数に対応するリンク関数の微分を Excel のシートの計算式に組み込むことにより，反復重み付き回帰のアルゴリズムを変えることなく実現できた．

反復重み付き回帰による最尤法は，リンク関数を用いてシグモイド曲線を直線化し，重み付き回帰分析により回帰係数を求め，その推定値からシグモイド曲線を間接的に推定している．情報行列を用いた最尤法は，回帰係数を用いてシグモイド曲線としての累積分布関数を定義するが，回帰直線のあてはめではなく，シグモイド曲線のあてはめを直接行う方法である．

田中ら訳（2008）では，ロジット法の例示に反復重み付き回帰による最尤法ではなく，情報行列を用いた最尤法が示されている．そこで，情報行列を用いた最尤法を Excel シート上で実現することにする．

スライド 13

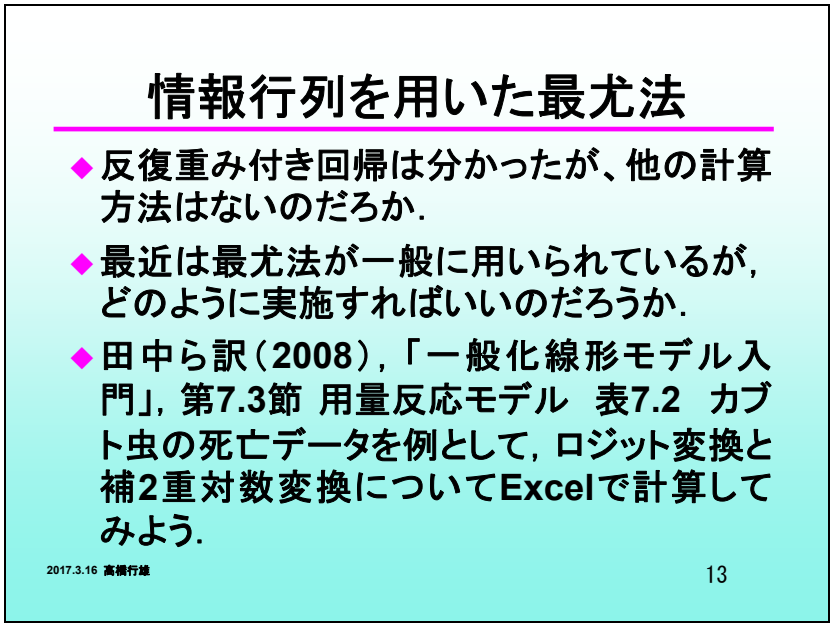


表 5.1（田中ら訳：表 7.2）に二硫化炭酸ガスに 5 時間曝露されたカブトムシの死亡数とガス濃度との関係をまとめたものである. 低濃度 1.6907 の場合には 59 匹中 6 匹の死亡であり，濃度が高くなるにつれて死亡数は増加し，高濃度 1.8839 の場合には，60 匹中 60 匹が死亡している．シグモイド曲線として累積ロジスティック分布をあてはめる．対数濃度を x_i としたときに死亡率 π_i は，次式で与えられる．

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}, \quad 1 - \pi_i = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

両辺に対数をとり整理して次式が得られる．

$$\begin{aligned} \ln \pi_i &= \beta_1 + \beta_2 x_i - \ln[1 + \exp(\beta_1 + \beta_2 x_i)] \\ &= \beta_1 + \beta_2 x_i + \ln(1 - \pi_i) \end{aligned}$$

さらに整理すると反復重み付きでのリンク関数が得られる．

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_1 + \beta_2 x_i$$

表 5.1 カブトムシの死亡データ（田中）

群	対数	反応		合計	出現率
i	濃度 x_i	0	1 死亡数	n_i	p_i
1	1.6907	53	6	59	0.102
2	1.7242	47	13	60	0.217
3	1.7552	44	18	62	0.290
4	1.7842	28	28	56	0.500
5	1.8113	11	52	63	0.825
6	1.8369	6	53	59	0.898
7	1.8610	1	61	62	0.984
8	1.8839	0	60	60	1.000
	（対数は常用対数で単位は，$CS_2 mgl^{-1}$）				

対数尤度 $\ln L_i$ は，累積ロジスティック分布の推定出現率（死亡率）に対して，それぞれの濃度 x_i に対するデータから得られる出現率が，2 項分布に従うとして定義される．対数用量 x_i での出現率を π_i としたときに，n_i 回中 y_i 回の反応が起きる確率は 2 項分布

$$f(y_i; n_i, \pi_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{(n_i - y_i)}$$

に従うとすると，それぞれの π_i に対する対数尤度 $\ln L_i$ は，

$$\ln L_i = \ln\left[\binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{(n_i - y_i)}\right] = \ln\binom{n_i}{y_i} + y_i \ln \pi_i + (n_i - y_i)\ln(1 - \pi_i)$$

であり，対数尤度 $\ln L_i$ を β_0 と β_1 について整理し，次式を得る．

$$
\begin{aligned}
\ln L_i &= \ln\binom{n_i}{y_i} + y_i \ln \pi_i + (n_i - y_i)\ln(1-\pi_i) \\
&= \ln\binom{n_i}{y_i} + y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) + n_i \ln(1-\pi_i) \\
&= \ln\binom{n_i}{y_i} + y_i(\beta_0 + \beta_1 x_i) - n_i \ln\left[1+\exp(\beta_0+\beta_1 x_i)\right]
\end{aligned}
$$

それぞれの対数尤度の和が $\ln L$ を最大にするように，β_0，β_1 をニュートン・ラフソン法で最大化する．それぞれの対数尤度 $\ln L_i$ を用量群 N 全体の対数尤度は，

$$
\ln L = \sum_{i=1}^{N} \ln L_i = \sum_{i=1}^{N} \ln\left[\binom{n_i}{y_i} \hat{\pi}_i^{\,y_i} (1-\hat{\pi}_i)^{n_i - y_i}\right]
$$

であり，この式は，

$$
\ln L = \sum_{i=1}^{N}\left[\ln\binom{n_i}{y_i} + y_i(\beta_0+\beta_1 x_i) - n_i \ln\left[1+\exp(\beta_0+\beta_1 x_i)\right]\right]
$$

と，回帰係数を用いた式として示される．ニュートン・ラフソン法は，対数尤度 $\ln L$ を回帰係数 β_0, β_1 で偏微分した式，さらにそれらの偏微分式を β_0, β_1 で偏微分した式を必用とする．

まず，$\ln L_i$ を β_0 と β_1 で偏微分して次式を得る．

$$
u_{i1} = \frac{\partial \ln L_i}{\partial \beta_0} = y_i - n_i \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} = y_i - n_i \pi_i
$$

$$
u_{i2} = \frac{\partial \ln L_i}{\partial \beta_1} = \left[y_i - n_i \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right] x_i = (y_i - n_i \pi_i) x_i
$$

これらの 1 階の偏微分式を $\boldsymbol{U}$ （スコアベクトル）として

$$
\boldsymbol{U} = \begin{bmatrix} \sum_{i=1}^{N} u_{i1} \\ \sum_{i=1}^{N} u_{i2} \end{bmatrix}
$$

として，まとめ，u_{i1} と u_{i2} をさらに β_0 と β_1 で偏微分する．

$$
h_{i11} = \frac{\partial^2 \ln L_i}{\partial \beta_0 \beta_0} = -n_i \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)} \cdot \left[1 - \frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right] = -n_i \pi_i (1-\pi_i)
$$

$$
h_{i12} = \frac{\partial^2 \ln L_i}{\partial \beta_0 \beta_1} = -n_i \pi_i (1-\pi_i) x_i = h_{i11} x_i
$$

$$
h_{i21} = \frac{\partial^2 \ln L_i}{\partial \beta_1 \beta_0} = h_{i12} = -n_i \pi_i (1-\pi_i) x_i = h_{i12}
$$

$$
h_{i22} = \frac{\partial^2 \ln L_i}{\partial \beta_1 \beta_1} = -n_i \pi_i (1-\pi_i) x_i^2 = h_{i11} x_i^2
$$

これらを，2×2 の 2 階の偏微分行列 $\boldsymbol{H}$（ヘッセ行列 $\boldsymbol{H}$，情報行列 $\boldsymbol{I} = -\boldsymbol{H}$）としてまとめる．負のヘッセ行列（情報行列 $\boldsymbol{I}$）の逆行列が分散共分散行列 $\boldsymbol{\Sigma}$ となる．

$$\boldsymbol{H}=\begin{bmatrix}\sum_{i=1}^{N}h_{i11} & \sum_{i=1}^{N}h_{i12}\\ \sum_{i=1}^{N}h_{i21} & \sum_{i=1}^{N}h_{i22}\end{bmatrix},\quad \boldsymbol{I}=\begin{bmatrix}-\sum_{i=1}^{N}h_{i11} & -\sum_{i=1}^{N}h_{i12}\\ -\sum_{i=1}^{N}h_{i21} & -\sum_{i=1}^{N}h_{i22}\end{bmatrix}$$

$$\boldsymbol{\Sigma}=(-\boldsymbol{H})^{-1}=\boldsymbol{I}^{-1}=\begin{bmatrix}v_{11} & v_{12}\\ v_{21} & v_{22}\end{bmatrix}=\begin{bmatrix}Var(\beta_0) & Cov(\beta_0,\beta_1)\\ Cov(\beta_0,\beta_1) & Var(\beta_1)\end{bmatrix}$$

5.2　情報行列を用いた Excel によるロジット法

表 5.2 にスコアベクトル $\boldsymbol{U}$ および情報行列 $\boldsymbol{I}$ を用いて Excel によるニュートン・ラフソン法の計算シートを示す.

表 5.2 情報行列を用いたロジット法による初期値での計算結果

	対数	死亡数	合計	出現率	推定値	2) 対数尤度	3) 1階の偏微分		4) 2階の偏微分（ヘッセ行 $\boldsymbol{H}$）			
i	用量 x_i	y_i	n_i	p_i	π_i	$\ln(L_i)$	u_{i1}	u_{i2}	h_{i11}	h_{i12}	h_{i21}	h_{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
	1) 切片 β_0^=			0.000	-37.856	-37.856		50.500	120.25	215.71	6.743	-3.754
	傾き β_1^=			0.000	21.337	21.337		100.785	215.71	387.43	-3.754	2.093
			8) $\boldsymbol{\beta}^{(m-1)}+\left[\boldsymbol{I}^{(m-1)}\right]^{-1}\boldsymbol{U}^{(m-1)}=\boldsymbol{\beta}^{(m)}$				5) $\boldsymbol{U}^{(m-1)}$		6) $\boldsymbol{I}^{(m-1)}$		7) $\boldsymbol{I}^{-1}$分散共分散	

手順 1）用量 x_i に対して合計 n_i に対する死亡数 y_i の出現率 p_i（死亡率）が計算されている．回帰係数の初期値として $\hat{\beta}_0=0$，$\hat{\beta}_1=0$ を与え，推定値 $\hat{\pi}_1$ が，$x_1=1.6907$ に対して

$$\pi_1=\frac{\exp(\beta_0+\beta_1x_1)}{1+\exp(\beta_0+\beta_1x_1)}=\texttt{EXP((0+0*1.6907)/(1+EXP(0+0*1.6907))=0.5}$$

として計算されている．他の $\hat{\pi}_i$ も指数項が 1 なので，すべて 0.5 となっている．

手順 2）Excel には 2 項分布 `Binomdist` 関数があるので，対数尤度 $\ln L_1$ は，

$$\ln L_i=\ln\left[\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{(n_i-y_i)}\right]=\texttt{Ln(Binomdist(6,59,0.5,false))}$$

$$=\texttt{-23.272}$$

で計算できる．対数尤度 $\ln L$ を最大化する際に π_i を変化させるが，π_i が変化しても第 1 項の 2 項係数は定数なので，最大化する際になんら寄与しないので，省略しても差し支えない．推定値 $\hat{\pi}_1 = 0.5$ の場合の対数尤度は，

$$\ln L_1' = y_1 \ln \pi_1 + (n_1 - y_1)\ln(1-\pi_1) = 6*\mathrm{Ln}(0.5)+(59-6)*\mathrm{Ln}(1-0.5)$$

$$=-40.896$$

$$\vdots$$

となり，それぞれの対数尤度 $\ln L_i'$ を加えて全体の対数尤度 $\ln L'$

$$\ln L' = \sum_{i=1}^{N} \ln L_i' = \mathrm{Sum}(L_i' \text{の範囲}) = -333.40$$

が下段に計算されている．

手順 3）対数尤度の回帰係数による 1 階の偏微分式は，推定値 $\hat{\pi}_1 = 0.5$ に対して

$$u_{1,1} = \frac{\partial \ln L_1}{\partial \beta_0} = y_1 - n_1\pi_1 = 6-59*0.5=-23.500$$

$$u_{1,2} = \frac{\partial \ln L_1}{\partial \beta_1} = (y_1 - n_1\pi_1)x_1 = (6-59*0.5)*1.6907=-39.731$$

$$\vdots$$

が計算されている．それぞれの u_{i1} と u_{i2} の $i = 1,\ldots,,N$ について加えてた結果

$$\sum_{i=1}^{N} u_{i1} = =\mathrm{Sum}(u_{i1} \text{の範囲}) = 50.500$$

$$\sum_{i=1}^{N} u_{i2} = =\mathrm{Sum}(u_{i2} \text{の範囲}) = 100.785$$

が下段に示されている．

手順 4）対数尤度の回帰係数による 2 階の偏微分式は，推定値 $\hat{\pi}_1 = 0.5$ に対して

$$h_{111} = \frac{\partial^2 \ln L_1}{\partial \beta_0 \beta_0} = -n_1\pi_1(1-\pi_1) = -59*0.5*(1-0.5)=-14.750$$

$$h_{112} = \frac{\partial^2 \ln L_1}{\partial \beta_0 \beta_1} = h_{111}x_1 = -14.750*1.6907=-24.938$$

$$h_{121} = \frac{\partial^2 \ln L_1}{\partial \beta_1 \beta_0} = h_{112} = -24.938$$

$$h_{i22} = \frac{\partial^2 \ln L_i}{\partial \beta_1 \beta_1} = h_{i11}x_i^2 = -14.750*1.6907\text{^}2=-42.162$$

$$\vdots$$

が計算されている．それぞれの $h_{i11}, h_{i11}, h_{i11}, h_{i11}$ を加えた結果が，

$\sum_{i=1}^{N} h_{i11}$ =Sum（h_{i11} の範囲）=-120.25

$\sum_{i=1}^{N} h_{i12} = \sum_{i=1}^{N} h_{i21}$ =Sum（h_{i12} の範囲）=-215.71

$\sum_{i=1}^{N} h_{i22}$ =Sum（h_{i22} の範囲）= -387.43

が下段に示されている．

手順 5）手順 3）で計算された 1 階の微分 u_{i1} と u_{i2} の i についての合計を縦ベクトル化しスコアベクトル $\boldsymbol{U}$ とする．

$$\boldsymbol{U} = \begin{bmatrix} \sum_{i=1}^{N} u_{i1} \\ \sum_{i=1}^{N} u_{i2} \end{bmatrix} = \begin{array}{|c|} \hline 50.500 \\ \hline 100.785 \\ \hline \end{array}$$

手順 6）手順 4）で計算された 2 階の微分の合計の負を取り 2×2 の行列にまとめて情報行列 $\boldsymbol{I}$ とする．

$$\boldsymbol{I} = \begin{bmatrix} -\sum_{i=1}^{N} h_{i11} & -\sum_{i=1}^{N} h_{i12} \\ -\sum_{i=1}^{N} h_{i21} & -\sum_{i=1}^{N} h_{i22} \end{bmatrix} = \begin{array}{|c|c|} \hline 120.25 & 215.71 \\ \hline 215.71 & 387.43 \\ \hline \end{array}$$

手順 7）情報行列 $\boldsymbol{I}$ の逆行列を計算し，回帰係数の分散共分散行列を計算する．

120.25	215.71
215.71	387.43

$\boldsymbol{I}$ 情報行列

6.743	-3.754
-3.754	2.093

$\boldsymbol{I}^{-1}$ 分散共分散行列

=MINVERSE（$\boldsymbol{I}$ の範囲）

手順 8）ニュートン・ラフソン法の第 1 回目の計算で回帰係数を更新する．

$$\begin{array}{|c|} \hline -37.856 \\ 21.337 \\ \hline \end{array} = \begin{array}{|c|} \hline 0.000 \\ 0.000 \\ \hline \end{array} + \begin{array}{|cc|} \hline 6.743 & -3.754 \\ -3.754 & 2.093 \\ \hline \end{array} \begin{array}{|c|} \hline 50.500 \\ 100.785 \\ \hline \end{array}$$

$$\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m-1)} + \left[\boldsymbol{I}^{(m-1)}\right]^{-1} \boldsymbol{U}^{(m-1)}$$

$$\begin{array}{|c|} \hline -37.856 \\ 21.337 \\ \hline \end{array}$$

$\boldsymbol{\beta}^{(1)}$ の範囲 = $\boldsymbol{\beta}^{(0)}$ の範囲 + Mmult（$\boldsymbol{I}^{-1}$ の範囲，$\boldsymbol{U}$ の範囲）

スライド 14 に，手順 1）から手順 8）について模式化した結果を示す．

スライド 14

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	π_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 β_0=		0.000	-37.856	-37.856		50.500	120.25	215.71	6.743	-3.754
		傾き β_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	

2017.3.16 高橋行雄

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．

14

1 階の偏微分を「行列 $\boldsymbol{U}$ の要素」， 2 階の偏微分の負を「情報行列 $\boldsymbol{I}$の要素」とした

手順 9a）更新された $\boldsymbol{\beta}^{(m)}$ の範囲データをコピーして $\boldsymbol{\beta}^{(m-1)}$ のデータの範囲に値のみを貼り付けると，手順 1）から手順 8）までが自動的に更新され，次の結果を得る．

$$\begin{bmatrix} -53.853 \\ 30.384 \end{bmatrix} = \begin{bmatrix} -37.856 \\ 21.337 \end{bmatrix} + \begin{bmatrix} -15.997 \\ 9.046 \end{bmatrix}$$

$$\boldsymbol{\beta}^{(2)} = \boldsymbol{\beta}^{(1)} + \left[\boldsymbol{I}^{(1)}\right]^{-1}\boldsymbol{U}^{(1)}$$

手順 9b）手順 9a）と同様に更新された $\boldsymbol{\beta}^{(m)}$ の範囲データをコピーして $\boldsymbol{\beta}^{(m-1)}$ のデータの範囲に値のみを貼り付けると，手順 1）から手順 8）までが自動的に更新され，次の結果を得る．このようにして反復計算を継続する．第 5 回目の $\boldsymbol{\beta}^{(m-1)}$ に対する $\boldsymbol{\beta}^{(m)}$ が同じ値となり収束したと判定する．

表 5.3 反復計算によって得られたパラメータ

		m=0	m=1	m=2	m=3	m=4	m=5	m=6
切片 β_0^=		0.000	-37.856	-53.853	-59.965	-60.708	-60.717	-60.717
傾き β_1^=		0.000	21.337	30.384	33.844	34.265	34.270	34.270

表 5.4 情報行列を用いたロジット法による最終的な計算結果

	対数	死亡数	合計	出現率	推定値	2) 対数尤度	3) 1階の偏微分		4) 2階の偏微分（ヘッセ行 $\boldsymbol{H}$）			
i	用量 x_i	y_i	n_i	p_i	π_i	$\ln(L_i)$	u_{i1}	u_{i2}	h_{i11}	h_{i12}	h_{i21}	h_{i22}
1	1.6907	6	59	0.102	0.059	-20.223	2.543	4.299	-3.255	-5.503	-5.503	-9.304
2	1.7242	13	60	0.217	0.164	-31.921	3.158	5.446	-8.227	-14.186	-14.186	-24.459
3	1.7552	18	62	0.290	0.362	-38.067	-4.451	-7.813	-14.321	-25.137	-25.137	-44.120
4	1.7842	28	56	0.500	0.605	-40.087	-5.898	-10.523	-13.379	-23.871	-23.871	-42.590
5	1.8113	52	63	0.825	0.795	-29.360	1.904	3.449	-10.261	-18.586	-18.586	-33.664
6	1.8369	53	59	0.898	0.903	-19.407	-0.291	-0.534	-5.157	-9.472	-9.472	-17.400
7	1.8610	61	62	0.984	0.955	-5.902	1.778	3.309	-2.653	-4.938	-4.938	-9.190
8	1.8839	60	60	1.000	0.979	-1.270	1.257	2.368	-1.231	-2.319	-2.319	-4.368
					$\ln L$=	-186.24	0.000	0.000	-58.48	-104.01	-104.01	-185.09
	1) 切片 β_0^=			-60.717	0.000	-60.717		0.000	58.48	104.01	26.840	-15.082
	傾き β_1^=			34.270	0.000	34.270		0.000	104.01	185.09	-15.082	8.481
			8) $\boldsymbol{\beta}^{(m-1)}+\left[\boldsymbol{I}^{(m-1)}\right]^{-1}\boldsymbol{U}^{(m-1)}=\boldsymbol{\beta}^{(m)}$				5) $\boldsymbol{U}^{(m-1)}$		6) $\boldsymbol{I}^{(m-1)}$		7) $\boldsymbol{I}^{-1}$分散共分散	

推定された回帰係数とその標準誤差は，

$$\hat{\beta}_0 = -60.717\,, \qquad SE(\hat{\beta}_0) = \sqrt{26.840} = 5.181$$

$$\hat{\beta}_1 = 34.270\,, \qquad SE(\hat{\beta}_1) = \sqrt{8.418} = 2.912$$

と推定される．

出現率に対するロジスティック曲線のあてはめた結果を 図 5.1 に示す．低濃度では推定曲線が観測値の下側，中ほどでは上側，高濃度ではやや下側となっていて，系統的な誤差となっていて，あてはめは良いとは言えない．

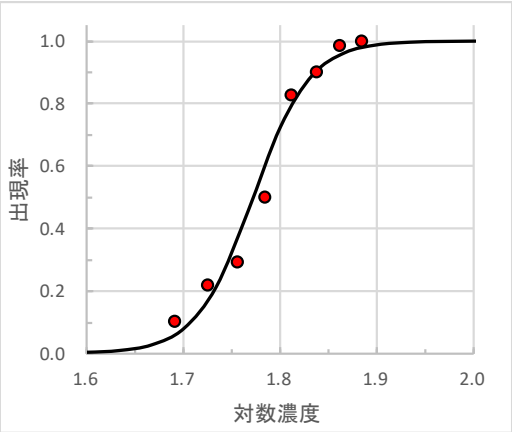


図 5.1　出現率に対するロジスティック曲線の直接あてはめ

5.3　シグモイド曲線の 95%信頼区間

シグモイド曲線の 95%信頼区間の推定方法は，2.5 節 の反復重み付き回帰の場合と同様な考え方で推測できる．反復重み付き回帰の場合の分散共分散行列は，重み付き回帰係数を求める計算式

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

の式の $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ が，共分散分析行列

$$(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1} = \begin{bmatrix} Var(\beta_0) & Cov(\beta_0,\beta_1) \\ Cov(\beta_0,\beta_1) & Var(\beta_1) \end{bmatrix}$$

となることを活用して，ベクトル $\boldsymbol{x}_i = (1, x_{1,i})$ を挟んだ 2 次形式で分散を

$$Var(\hat{\eta}_i) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i}) = \boldsymbol{x}_i(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{x}_i^T$$

で推定した．

情報行列を用いた最尤法の場合は，シグモイド曲線に対する対数尤度関数の 2 階の偏微分行列（ヘッセ行列 $\boldsymbol{H}$）を計算し，情報行列 $\boldsymbol{I} = -\boldsymbol{H}$ の逆行列が分散共分散行列 $\boldsymbol{\Sigma}$ となることをすでに示してきた．

表 5.4 の $\pi_1 = 0.059$ の 95%信頼区間を求めるためには，分散共分散行列は次の逆ロジット式

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

の回帰係数 β_0，β_1 に対応しているので，間接的に求めることになる．この式を回帰式について整理するとのリンク関数（ロジット変換）は，次式

$$\hat{\eta}_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \hat{\beta}_1 + \hat{\beta}_2 x_i$$

で与えられるので，推定値 $\hat{\eta}_i$ についての分散は，ベクトル $\boldsymbol{x}_i = (1, x_{1,i})$ を挟んだ

$$Var(\eta_i) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i}) = \boldsymbol{x}_i(-\boldsymbol{H})^{-1}\boldsymbol{x}_i^T$$

2 次形式で分散を求めることができる．線形式 $\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i}$ についての 95%信頼区間を，

$$\mathrm{L95\%}(\eta_i) = \eta_i - 1.96\sqrt{Var(\eta_i)}$$
$$\mathrm{U95\%}(\eta_i) = \eta_i + 1.96\sqrt{Var(\eta_i)}$$

で計算し，回帰直線上の 95%信頼区間を逆ロジット変換

$$\mathrm{L95\%}(\pi_i) = \frac{\exp\left[\mathrm{L95\%}(\eta_i)\right]}{1 + \exp\left[\mathrm{L95\%}(\eta_i)\right]}$$

$$\mathrm{U95\%}(\pi_i) = \frac{\exp\left[\mathrm{U95\%}(\eta_i)\right]}{1+\exp\left[\mathrm{U95\%}(\eta_i)\right]}$$

で求める．煩雑な計算とも思われるが，Excel を用いれば容易である．推定値 $\hat{\eta}_i$ は，

1	1.6907	-60.717	=	-2.777
1	1.7242	34.270		-1.629
1	1.7552			-0.566
1	1.7842			0.428
1	1.8113			1.356
1	1.8369			2.234
1	1.8610			3.060
1	1.8839			3.844
$\boldsymbol{X}$		$\boldsymbol{\beta}$		$\boldsymbol{\eta}$

のように，デザイン行列 $\boldsymbol{X}$ と回帰係数縦ベクトル $\boldsymbol{\beta}$ の関で求めることができる．$\hat{\eta}_1$ の分散 $Var(\hat{\eta}_1)$ は，分散共分散行列は，負のヘッセ行列（情報行列）の逆数を挟み込んだ，$\hat{\eta}_1$ に対応するデザイン行列のベクトルついての 2 次形式でも求めることができる．

$$Var(\eta_1) = \begin{bmatrix} 1 & 1.6907 \end{bmatrix} \begin{bmatrix} 26.840 & -15.082 \\ -15.082 & 8.481 \end{bmatrix} \begin{bmatrix} 1 & 1.6907 \end{bmatrix}^T = 0.082$$

$$\boldsymbol{x} \qquad (-\boldsymbol{H})^{-1} \qquad \boldsymbol{x}^T$$

この分散を用いて $\hat{\eta}_1$ の 95%信頼区間は，

$$\mathrm{L95\%}(\eta_1) = -2.777 - 1.96\sqrt{0.0.082} = -3.339$$

$$\mathrm{U95\%}(\eta_1) = -2.777 + 1.96\sqrt{0.0.082} = -2..214$$

として計算できる．出現率に対する 95%信頼区間は，逆ロジットで

$$\mathrm{L95\%}(\pi_1) = \frac{\exp\left[\mathrm{L95\%}(\eta_1)\right]}{1+\exp\left[\mathrm{L95\%}(\eta_1)\right]} = \frac{\exp(-3.339)}{1+\exp(-3.339)} = 0.034$$

$$\mathrm{U95\%}(\pi_1) = \frac{\exp\left[\mathrm{U95\%}(\eta_1)\right]}{1+\exp\left[\mathrm{U95\%}(\eta_1)\right]} = \frac{\exp(-2.214)}{1+\exp(-2.214)} = 0.098$$

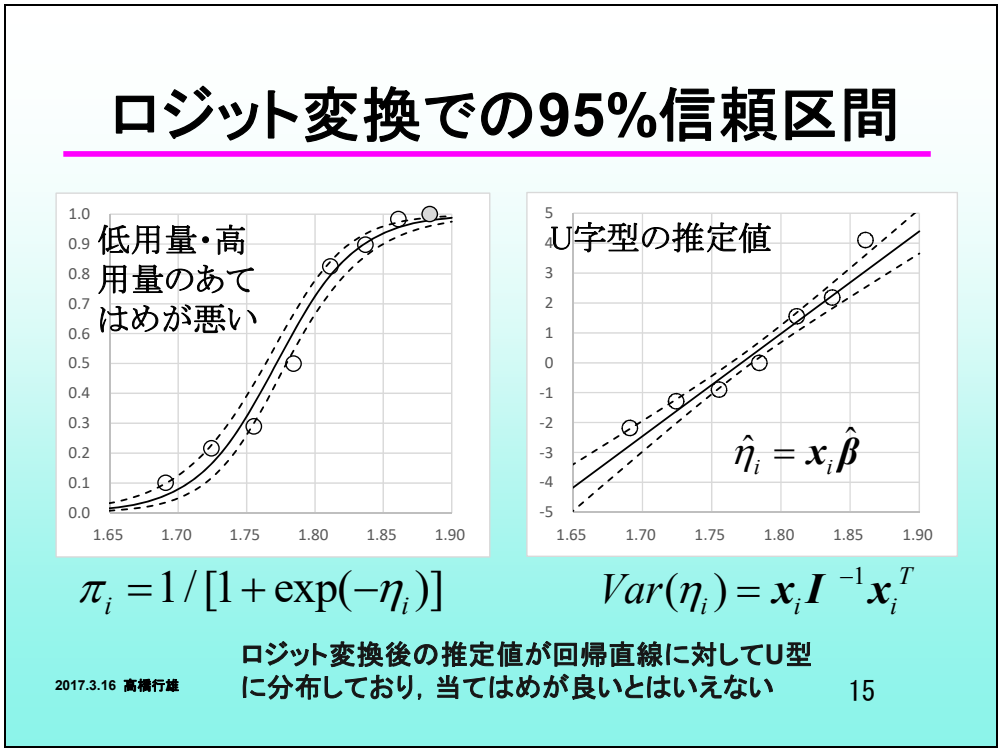
として求めることができる．この計算を全ての i について繰り返して，表 5.5 を得る．

推定値および 95%信頼区間を対数用量 x の適当な区分で計算し，それらの点を結んだ結果をスライド 15 に示す．

表 5.5 推定されたシグモイド曲線の 95%信頼区間

		対数濃度	死亡数	合計	出現率	推定値	回帰直線について				πについて	
i	x_0	x_1	y_i	n_i	p_i	π_i	η_i	分散	L95%	U95%	L95%	U95%
1	1	1.6907	6	59	0.102	0.059	-2.777	0.082	-3.339	-2.214	0.034	0.098
2	1	1.7242	13	60	0.217	0.164	-1.629	0.042	-2.030	-1.227	0.116	0.227
3	1	1.7552	18	62	0.290	0.362	-0.566	0.022	-0.855	-0.278	0.298	0.431
4	1	1.7842	28	56	0.500	0.605	0.428	0.017	0.169	0.686	0.542	0.665
5	1	1.8113	52	63	0.825	0.795	1.356	0.026	1.039	1.674	0.739	0.842
6	1	1.8369	53	59	0.898	0.903	2.234	0.046	1.813	2.654	0.860	0.934
7	1	1.8610	61	62	0.984	0.955	3.060	0.075	2.523	3.596	0.926	0.973
8	1	1.8839	60	60	1.000	0.979	3.844	0.111	3.190	4.499	0.960	0.989

スライド 15



5.4 出現率が 10 パーセントとなる用量とその近似 95%信頼区間

情報行列を用いた反復計算で求められた結果から LD_{50} とその近似信頼区間を求める方法は，反復重み付き回帰の場合と同様に，LD_{50} の分散 $\boldsymbol{d}^T(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{d}$ に代えて $\boldsymbol{d}^T\boldsymbol{I}^{-1}\boldsymbol{d}$ として表 2.4 と同様に求めることができる．一般的に出現率が p_0 となるような対数用量 x_{p_0} は，

$$p_0 = \frac{1}{1+\exp\left[-(\hat{\beta}_0 + \hat{\beta}_1 x_{p_0})\right]}$$

で与えられる．この式を x_{p_0} について解き p_0 パーセント出現率となる用量は，

$$x_{p_0} = \frac{-\hat{\beta}_0 + \ln(p_0/(1-p_0))}{\hat{\beta}_1}$$

となる．この分散は，パラメータによる偏微分を

$$d_0 = \frac{\partial x_{p_0}}{\partial \hat{\beta}_0} = -\frac{1}{\hat{\beta}_1}$$

$$d_1 = \frac{\partial x_{p_0}}{\partial \hat{\beta}_1} = \frac{\hat{\beta}_0 - \ln(p_0/(1-p_0))}{\hat{\beta}_1^2}$$

として，これら用いた 2 次形式により求められる．表 5.6 に第 5 回目の反復で得られた結果は，$\hat{\beta}_0 = -60.717$，$\hat{\beta}_1 = 34.270$ なので，10 パーセント出現率は，

$$LD_{10} = x_{p=0.1} = \frac{-\hat{\beta}_0 + \ln\left[0.1/(1-0.1)\right]}{\hat{\beta}_1} = \frac{-60.717 - 2.197}{34.270} = 1.707$$

となる．パラメータに関する偏微分は，

$$d_0 = -\frac{1}{\hat{\beta}_1} = -\frac{1}{34.270} = -0.029$$

$$d_1 = \frac{\hat{\beta}_0 - \ln(p_0/(1-p_0))}{\hat{\beta}_1^2} = \frac{-60.717 - 1.707}{34.270^2} = -0.050$$

となるので，分散 $Var(x_{p=0.10})$ は，情報行列を挟みこんだ次式で

-0.029	-0.050	26.840	-15.082	-0.029	=	0.00005079
		-15.082	8.481	-0.050		
$\boldsymbol{d}^T$		$\boldsymbol{I}^{-1}$		$\boldsymbol{d}$		

$$Var(x_{p=0.10}) = 0.00005079$$

で求めることができる．表 5.6 に示すように $SE(x_{p=0.10}) = 0.00713$ をもとに，$LD_{10} = 1.707$ の 95%信頼区間は，（1.694，1.722）と計算される．

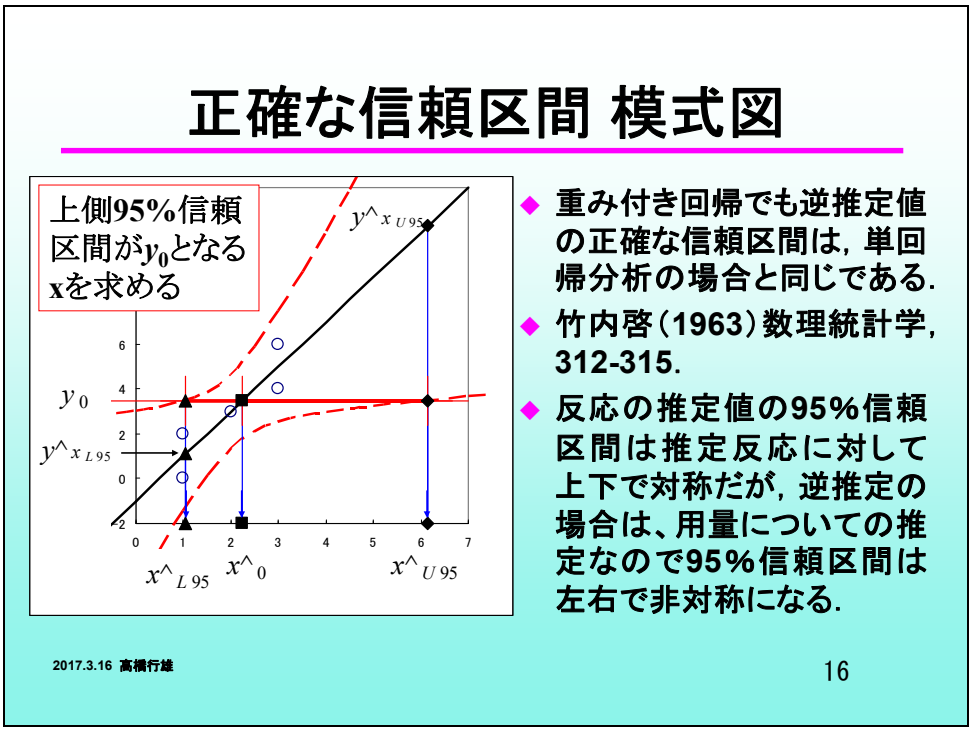
表 5.6　出現率が 10 パーセントとなる用量の推定

切片 β_0^=	-60.717	0.000	-60.717	-186.24		0.000	58.48	104.01	26.840	-15.082
傾き β_1^=	34.270	0.000	34.270	ln L		0.000	104.01	185.09	-15.082	8.481
$\boldsymbol{\beta}^{(m-1)} + \left[\boldsymbol{I}^{(m-1)}\right]^{-1} \boldsymbol{U}^{(m-1)} = \boldsymbol{\beta}^{(m)}$						$\boldsymbol{U}^{(m-1)}$	$\boldsymbol{I}^{(m-1)}$		$\boldsymbol{I}^{-1}$分散共分散	
	P=	0.10		$\boldsymbol{d}$		$\boldsymbol{d}^T(\boldsymbol{I})^{-1}\boldsymbol{d}$		x	$dose$	
	対数LD_p=	1.708	d_0=	-0.029	Var=	5.1E-05	L95%=	1.694	49.39	
	LD_p=	51.004	d_1=	-0.050	SE=	7.1E-03	U95%=	1.722	52.67	

5.5　LD_{10} の正確な 95%信頼区間

ある出現率 p_0 となる用量 x_0 の近似的な信頼区間は，前節に示したように分散共分散行列を挟み込んだ 2 次形式で算出できることを示した．スライド 15 の推定された回帰直線と上下の 95%信頼区間までの距離は，Y 軸方向で同じであるが，ある出現率 p_0 となる用量 x_0 の 95%信頼区間は，X 軸方向に上下の 95%信頼区間までの距離となる．スライド 16 に示すように，ある出現率 p_0 に対する $\eta_0 = \ln\left[p_0 / (1-p_0)\right]$ を通る X 軸に平行な直線が，95%信頼区間の下限と上限を横切る点までの対数用量 x_0 からの距離は，等しくならない．

スライド 16



スライドでは y_0 としているが η_0 と読み替える．

LD_{10} について正確な 95%信頼区間は，回帰直線 $\eta = \beta_0 + \beta_1 x$ の 95%信頼区間を基に求めることができる．竹内（1979）によれば，95%信頼区間の下限の推定値を $\hat{x}_{L95}$ としたときに，$\hat{x}_{L95}$ における回帰直線上の推定値は，$\hat{\eta}_{\hat{x}_{L95}} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}$ で与えられる．その推定値 $\hat{\eta}_{\hat{x}_{L95}}$ の 95%信頼区間は，$\hat{x}_{L95}$ における

$$95\%\text{信頼区間} = \hat{\eta}_{\hat{x}_{L95}} \pm t_\alpha \sqrt{Var(\hat{\eta}_{\hat{x}_{L95}})}$$

で与えられる．その上側の 95%信頼曲線が，x_0 の逆推定値 η_0 を通る X 軸に平行な直線との交点は，p_0 を LD_{p_0} としたときに，

$$\eta_0 = \ln\left[p_0 / (1-p_0)\right]$$

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} = \eta_0$$

で与えられる．この式を $\hat{x}_{L95}$ について解くことにより，$\hat{x}_0$ の下側の 95%信頼区間の式が求められる．

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

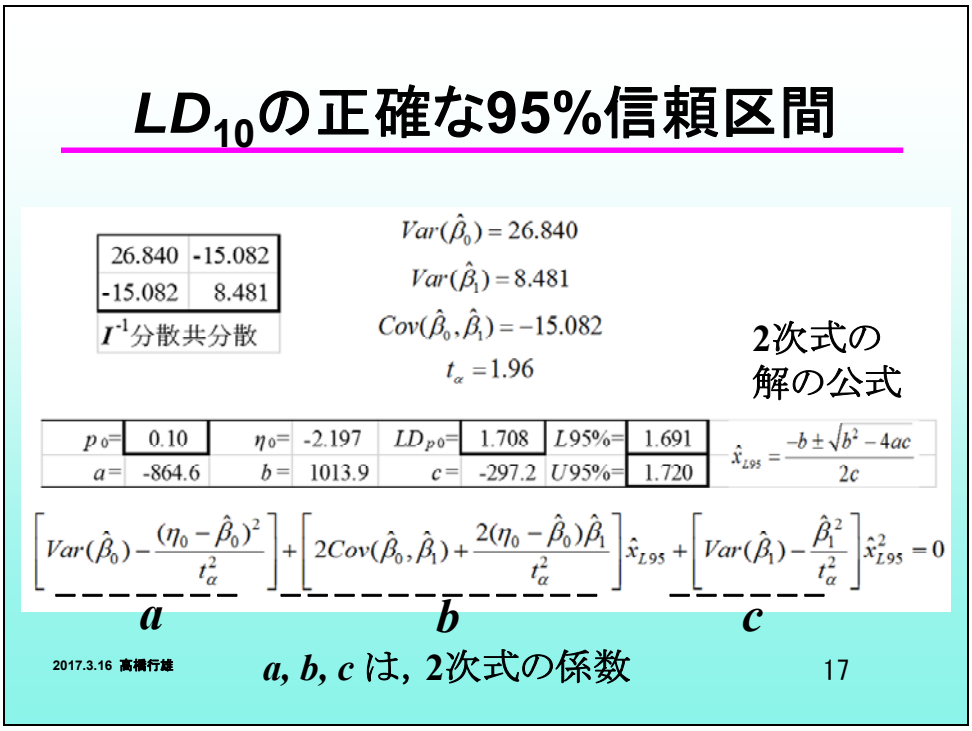
式は複雑であるが，$\hat{x}_{L95}$ に関して 2 次式

$$a+b\hat{x}_{L95}+c\hat{x}_{L95}^2=0$$

となる．従って，2 次式の解の公式により x_{L95} を求めることができる．解は 2 つあるが，小さい方が $\hat{x}_{L95}$ となり，大きい方が $\hat{x}_{U95}$ となる．

表 5.6 から $\hat{\beta}_0=-60.717$，$\hat{\beta}_1=34.270$，分散共分散行列 $\boldsymbol{I}^{-1}$ より，$Var(\hat{\beta}_0)=26.840$，$Var(\hat{\beta}_1)=8.481$，$Cov(\hat{\beta}_0,\hat{\beta}_1)=-15.082$ を用い，$t_\alpha=1.96$ として a，b，c を次式によりスライド 17 に示すように計算する．

スライド 17



$$\eta_0=\ln\left[p_0/(1-p_0)\right]=\ln\left[0.1/(1-0.1)\right]=-2.197$$

$$a=Var(\hat{\beta}_0)-\frac{(\eta_0-\hat{\beta}_0)^2}{t_\alpha^2}=26.840-\frac{(-2.197-60.171)^2}{1.96^2}=-864.6$$

$$b=2Cov(\hat{\beta}_0,\hat{\beta}_1)+\frac{2(\eta_0-\hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2}=2\times(-15.082)+\frac{2\times(-2.197-60.717)\times 34.270}{1.96^2}=1013.9$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_\alpha^2} = 8.481 - \frac{8.489^2}{1.96^2} = -297.2$$

これらを 2 次式の解の公式に代入して

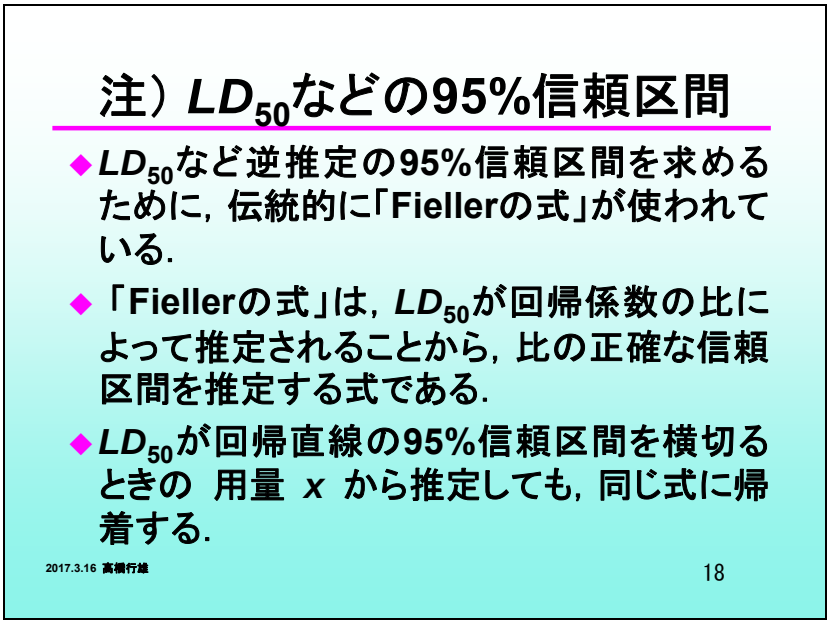
$$\hat{x}_{L95} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} = (1.691,\ 1.720)$$

を得る．

5.6　古典的なフィラーの式を用いた 95%信頼区間

プロビット法は，50 パーセント致死量 LD_{50} を正確に求める方法として知られている．さらに，LD_{50} の正確な 95%信頼区間も求めるための方法として“フィラーの式”により定式化された．フィラーの式(Fieller の式)は，正規分布に従うと仮定する回帰係数の切片 β_0 と傾き β_1 の比

スライド 18



$\rho = \beta_0 / \beta_1$ を使って 95%信頼区間を正確に計算する方法である．Collett（2003）は，分散共分散行列を用いてフィラーの式の導出を行っている．

推定された回帰係数を $\hat{\beta}_0$ および $\hat{\beta}_1$ とし，分散を $Var(\hat{\beta}_0)$ と $Var(\hat{\beta}_1)$ とし，共分散を $Cov(\hat{\beta}_0, \hat{\beta}_1)$ とする．次の関数 $\psi = \hat{\beta}_0 - \rho\hat{\beta}_1$ を考える．$\hat{\beta}_0$ と $\hat{\beta}_1$ が β_0 と β_1 の不偏推定量となるので，ψ の期待値は，

$$E(\psi)=\beta_0-\rho\beta_1=0$$

となり，ψ の分散は，

$$Var(\psi)=Var(\beta_0)-2\rho Cov(\beta_0,\beta_1)+\rho^2 Var(\beta_1)$$

で与えられる．そのとき，$\hat{\beta}_0$ と $\hat{\beta}_1$ は，正規分布に従うと仮定されるので，ψ は，同様に正規分布に従い

$$\frac{\hat{\beta}_0-\rho\hat{\beta}_1}{\sqrt{Var(\psi)}}$$

は，標準正規分布となる．従って，t_α が，標準正規分布の両側 α 点であるとしたときに，ρ の $100\cdot(1-\alpha)\%$ 信頼区間は，

$$\left|\hat{\beta}_0-\rho\hat{\beta}_1\right|\le t_\alpha\sqrt{Var(\psi)}$$

の式から 2 つ値となる．両辺を 2 乗し，等式として，

$$\hat{\beta}_0{}^2-2\rho\hat{\beta}_0\hat{\beta}_1+\rho^2\hat{\beta}_1{}^2-t_\alpha^2 Var(\psi)=0$$

となる．$Var(\psi)$ に $Var(\hat{\psi})$ を代入すると

$$\hat{\beta}_0{}^2-2\rho\hat{\beta}_0\hat{\beta}_1+\rho^2\hat{\beta}_1{}^2-t_\alpha^2\left[Var(\hat{\beta}_0)-2\rho Cov(\hat{\beta}_0,\hat{\beta}_1)+\rho^2 Var(\hat{\beta}_1)\right]=0$$

となり，式を整理と，次のように ρ に関する二次式が得られる

$$\left[\hat{\beta}_0{}^2-Var(\hat{\beta}_0)t_\alpha^2\right]+\left[2Cov(\hat{\beta}_0,\hat{\beta}_1)t_\alpha^2-2\hat{\beta}_0\hat{\beta}_1\right]\rho+\left[\hat{\beta}_1{}^2-Var(\hat{\beta}_1)t_\alpha^2\right]\rho^2=0$$

この 2 次方程式の 2 つの根は，ρ のための 95%信頼区間となる．

2 次式 $a+b\rho+c\rho^2=0$ の 係数 a，b，c は，それぞれ，

$$a=\hat{\beta}_0{}^2-Var(\hat{\beta}_0)t_\alpha^2$$
$$b=2Cov(\hat{\beta}_0,\hat{\beta}_1)t_\alpha^2-2\hat{\beta}_0\hat{\beta}$$
$$c=\hat{\beta}_1{}^2-Var(\hat{\beta}_1)t_\alpha^2$$

となり，式は複雑であるが，$\hat{x}_{L95}$ に関して 2 次式

$$a+b\hat{\rho}+c\hat{\rho}^2=0$$

となる．従って，2 次式の解の公式により LD_{50} の 95%信頼区間を求めることができる．解は 2 つあるが，小さい方が $\hat{x}_{L95}$ となり，大きい方が $\hat{x}_{U95}$ となる．

表 5.6 から $\hat{\beta}_0=-60.717$，$\hat{\beta}_1=34.270$，分散共分散行列 $\boldsymbol{I}^{-1}$ より，$Var(\hat{\beta}_0)=26.840$，$Var(\hat{\beta}_1)=8.481$，$Cov(\hat{\beta}_0,\hat{\beta}_1)=-15.082$ を用い，$t_\alpha=1.96$ として a，b，c は

$$a=\hat{\beta}_0{}^2-Var(\hat{\beta}_0)t_\alpha^2=(-60.717)^2-26.840\times1.96^2=3583.5$$

$$b=2Cov(\hat{\beta}_0,\hat{\beta}_1)t_\alpha^2-2\hat{\beta}_0\hat{\beta}=2\times(-15.082)-2\times(-60.717)\times34.270=4045.7$$

$$c = \hat{\beta}_1^2 - Var(\hat{\beta}_1)t_\alpha^2 = 34.270^2 - 8.481 \times 1.96^2 = 1141.9$$

のように計算される．これらを 2 次式の解の公式に代入して

$$\hat{\rho} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} = (-1.764,\ -1.779)$$

となるので，表 5.7 の Excel の計算シートに示すように LD_{50} の正確な 95%信頼区間 $LD_{50} = (1.764,\ 1.779)$ を得る．

表 5.7　出現率が 50 パーセントとなる用量の正確な 95%信頼区間

回帰の95%信頼区間からの導出								
	p_0=	0.50	η_0=	0.000	LD_{p0}=	1.772	L95%=	1.764
	a=	-932.8	b=	1053.1	c=	-297.2	U95%=	1.779
フィラーの式による導出								
	p_0=	0.50	η_0=	0.000	LD_{p0}=	1.772	L95%=	1.764
	a=	3583.5	b=	4045.7	c=	1141.9	U95%=	1.779

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

$$-\rho = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$$

正確な LD_{50} の 95%信頼区間の導出法を伝統的なフィラーの式によって示したが，前節での回帰の 95%信頼区間からの導出結果と一致することを確認した．ここでは，2 次式の係数を基に解の公式から求めたが，“フィラーの式”としては，2 次式の解

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

$$= \frac{-\left[2Cov(\hat{\beta}_0, \hat{\beta}_1)t_\alpha^2 - 2\hat{\beta}_0\hat{\beta}_1\right] \pm \sqrt{\left[2Cov(\hat{\beta}_0, \hat{\beta}_1)t_\alpha^2 - 2\hat{\beta}_0\hat{\beta}_1\right]^2 - 4\left[\hat{\beta}_0^2 - Var(\hat{\beta}_0)t_\alpha^2\right]}}{2\left[\hat{\beta}_1^2 - Var(\hat{\beta}_1)t_\alpha^2\right]}$$

を更に整理した式として文献では示されている．この式は煩雑であるので，元の 2 次式の係数を計算し，その後に 2 次式の解を求めるといった計算法を勧める．

フィラーの式は，元々は 2 つの正規分布に従う変数の比の正確な信頼区間を求めるための一般的な考え方であり，LD_{50} に応用したものである．

回帰の 95%信頼区間からの導出は，LD_{50} のみならず LD_{10} に対しても，スライド 16 に示したように，視覚的に説明可能であるので，推奨したい．

6. 情報行列を用いた補 2 重対数法によるシグモイド曲線の推定

6.1 Excel ソルバーを用いた最尤法

反復重み付き回帰を使う最尤法については，累積正規分布を用いるプロビット法に引き続き，ロジット法，および，補 2 重対数法を示してきた．情報行列を用いる最尤法では，プロビット法ではなく累積ロジスティック分布をシグモイド曲線として用いるロジット法について前章で示した．本章では，シグモイド曲線に累積最小極値分布を用いる補 2 重対数法について示す．

反復重み付き回帰を使う最尤法では，（プロビット法，ロジット法，補 2 重対数法）のどの方法でも，リンク関数に対応した補正値 z_i の計算，および 重み w_i の計算式が異なるだけで，計算のための方法は同一であることを示した．

情報行列を用いる最尤法の場合は，累積分布関数を $\pi = F(x)$ としたときに，それぞれの対数濃度 x_i に対して n_i 回中 y_i 回の反応が起きる確率が 2 項分布に従うとして，対数尤度を定義する．それぞれの対数用量 x_i での出現率を π_i としたときの，それぞれの尤度は，2 項分布の確率

$$f(y_i;n_i,\pi_i) = \binom{n_i}{y_i} \pi_i^{y_i} (1-\pi_i)^{(n_i-y_i)}$$

で定義される．最大化するための対数尤度 $\ln L$ は，

$$\ln L = \sum_{i=1}^{N} \ln L_i = \sum_{i=1}^{N} \ln\left[\binom{n_i}{y_i} \hat{\pi}_i^{y_i} (1-\hat{\pi}_i)^{n_i-y_i}\right]$$

である．シグモイド曲線としての累積分布が異なっても対数尤度の定義は，同じである．情報行列を用いた最尤法は，対数尤度 $\ln L$ を累積分布のパラメータで偏微分した式をスコア関数とし，さらにパラメータで 2 階偏微分した式の負を取った情報行列としてニュートン・ラフソン法で対数尤度 $\ln L$ を最大化した．

理解を深める為に必要だとしても，これらの式を正確に Excel に入力し，解を求めることは容易ではない．理解を深める第 1 歩として，Excel のソルバーを使って最尤解を求める方法を推奨する．これまで，Excel のソルバーでの最尤法を示してこなかったのは，統計ソフトでは必ずしも対応していないシグモイド曲線の 95%信頼区間，LD_{50} などの 95%信頼区間の実用上の課題を解くために，情報行列の逆行列である分散共分散行列を必要としたからである．残念ながら，Excel のソルバーでは，応用のために必須な分散共分散行列が得られない．

Excel のソルバーで最尤解を求めることは，統計モデルの対数尤度を正確に定義する必要もあり，自らニュートン・ラフソン法にチャレンジする前段階としての学習のための道具でもある．

スライド 19

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	π_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 β_0^=		-39.572	0.000	-39.572			0.000	158.51	285.36	10.427	-5.788
		傾き β_1^=		22.041	0.000	22.041			0.000	285.36	514.05	-5.788	3.215
				$\boldsymbol{\beta}^{(m-1)}$	$\lceil \boldsymbol{H}^{(m-1)} \rceil^{-1}$	$\boldsymbol{U}^{(m-1)}$	$\boldsymbol{\beta}^{(m)}$		$\boldsymbol{U}$	$\boldsymbol{I}$			$\boldsymbol{I}^{-1}$分散共分散

p_0=	0.10	η_0=	-2.250	LD_{p0}=	1.693	$L95\%$=	1.672
a=	-352.2	b=	416.7	c=	-123.2	$U95\%$=	1.709

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

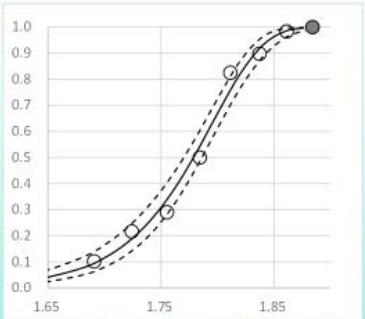
2017.3.16 高橋行雄

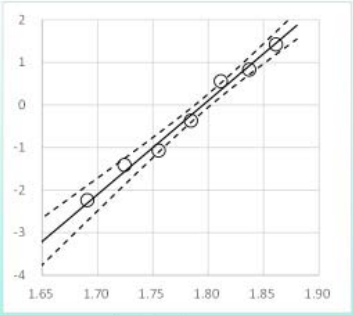
スコア行列U および 情報行列 I の計算式を補2重対数の式にかえるだけで，他はプロビット・ロジットの場合と全く同じである．

19

スライド 20

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．

2017.3.16 高橋行雄

20

6.2　最小極値分布をシグモイド曲線とする補 2 重対数法

最小極値分布の累積分布は，対数濃度 x に対して，位置パラメータを μ_{SEV}，形状パラメータを σ_{SEV} として次のように定義される．

$$\pi = F_{SEV}(x) = 1 - \exp\left[-\exp\left(\frac{x-\mu_{SEV}}{\sigma_{SEV}}\right)\right]$$

この式を

$$\pi = 1 - \exp\left[-\exp\left(-\frac{\mu_{SEV}}{\sigma_{SEV}} + \frac{1}{\sigma_{SEV}}x\right)\right]$$

と線形式に変換することにより，

$$\beta_0 = -\frac{\mu_{SEV}}{\sigma_{SEV}}, \quad \beta_1 = \frac{1}{\sigma_{SEV}}$$

として

$$\pi = 1 - \exp\left[-\exp\left(\beta_0 + \beta_1 x\right)\right]$$

に変換することができる．2 重に対数を取ることで

$$\ln\left[-\ln(1-\pi)\right] = \beta_0 + \beta_1 x$$

線形化することができ，反復重み付き回帰によって最尤解を求めることは第 4 章で示した．このような線形化は，情報行列を求めるための偏微分式を簡単化するための基本的な操作である．　Excel には 2 項分布の確率関数があるので，

$$\ln L_i = \ln\left[\binom{n_i}{y_i}\hat{\pi}_i^{y_i}(1-\hat{\pi}_i)^{n_i-y_i}\right] = \texttt{Ln(Binimdist(y}_\texttt{i}\texttt{,n}_\texttt{i}\texttt{,}\pi_\texttt{i}\texttt{,false))}$$

によって対数尤度を定義すればよい．多くの統計ソフトは，2 項分布の確率関数の対数を取るのではなく，2 項分布の組み合わせ数を含めない．前節のロジット法では，文献で示された対数尤度となるように，組み合わせ数を含めない対数尤度とした．

Excel のソルバーを用いた最尤法の計算結果を 表 6.1 に示す．これまで示したきた情報行列などを全く必要としない．また，反復計算などの仕組みも全くない．Excel のソルバーにより，対数尤度 $\ln L$ を最大にするように，回帰係数 β_0 と β_1 をソルバーで変化させただけである．

田中ら訳（2008）の表 7.4 に（ロジスティックモデル，プロビットモデル，極値モデル）の 3 種類の用量反応モデルの比較表があり，極値モデルの適合度がもっともよいと述べられている．ただし，回帰係数が示されていないので Excel のソルバーで求めた回帰係数が正しいか判断ができない．なお、Dobson, Barntt（2008）の Table7.4 には，3 種のモデルに対する回帰係数が示されており，Excel のソルバーで求めた回帰係数が正しいことが確認された．

表 6.1　カブト虫データに Excel ソルバーを用いた補 2 重対数法での最尤解

					補2重対数			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	$\ln L_i$
i	用量 x	y_i	n_i	p_i	2) π_i	3) $\ln L_i$		π_i	$\ln L_i$	定数項無
1	1.6907	6	59	0.102	0.632	-38.129		0.095	-1.792	-19.415
2	1.7242	13	60	0.217	0.632	-23.689		0.188	-2.242	-31.515
3	1.7552	18	62	0.290	0.632	-17.103		0.338	-2.521	-37.674
4	1.7842	28	56	0.500	0.632	-4.270		0.542	-2.444	-39.017
5	1.8113	52	63	0.825	0.632	-7.705		0.758	-2.860	-30.007
6	1.8369	53	59	0.898	0.632	-12.686		0.918	-1.912	-19.536
7	1.8610	61	62	0.984	0.632	-24.852		0.986	-0.999	-5.126
8	1.8839	60	60	1.000	0.632	-27.521		0.999	-0.053	-0.053
				切片 β_0^=	0.000	-155.954		-39.572	-14.822	-182.343
				傾き β_1^=	0.000	4) 和 $\ln L$		22.041	和 $\ln L$	
					1) 初期値			最尤解		

手順 1）初期値として $\hat{\beta}_0 = 0$ と $\hat{\beta}_1 = 0$ とする．

手順 2）推定値として $\pi_i = 1 - \exp\left[-\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]$ を各セルの計算する．

手順 3）対数尤度 $\ln L_i$ を 2 項分布関数を使って計算する．

$\ln L_i$ =Ln(Binimdist(y_i,n_i,π_i,false))

手順 4）$\ln L_i$ の合計 $\ln L$ を計算する．

$\ln L$ =Sum($\ln L_i$ の範囲)

手順 5) Excel のソルバーを ファイル＞オプション＞アドイン＞ソルバーアドイン でアクティブにする．データ＞ソルバー で起動する．

手順 6）ソルバーのパラメータ設定で，目的セル ← $\ln L$ セル，目標値：最大値を選択，変数セルの選択 ← β_0 と β_1 の範囲　をセットし，解決 する．

以上の操作で得られた結果が表の右に示されている．求められた対数尤度には，定数項である組み合わせ数も含まれていることに注意が必要である．定数項無の対数尤度は，

$$\ln L == \sum_{i=1}^{N}\left[y_i \ln(\hat{\pi}_i) + (n_i - y_i)\ln(1-\hat{\pi}_i)\right]$$

で計算したものである．

Excel のソルバーの良さは，一般化線形モデルの形式にとらわれず，表 6.2 に示すように位置パラメータを μ_{SEV} および 形状パラメータを σ_{SEV} を用いた非線形のままでも最尤解を求めることができる．

表 6.2　位置パラメータを μ_{SEV} および 形状パラメータを σ_{SEV} を用いた最尤法

					補2重対数		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
i	用量 x	y_i	n_i	p_i	π_i	$\ln L_i$	π_i	$\ln(L_i)$
1	1.6907	6	59	0.102	0.095	-1.792	0.095	-1.792
2	1.7242	13	60	0.217	0.188	-2.242	0.188	-2.242
3	1.7552	18	62	0.290	0.338	-2.521	0.338	-2.521
4	1.7842	28	56	0.500	0.542	-2.444	0.542	-2.444
5	1.8113	52	63	0.825	0.758	-2.860	0.758	-2.860
6	1.8369	53	59	0.898	0.918	-1.912	0.918	-1.912
7	1.8610	61	62	0.984	0.986	-0.999	0.986	-0.999
8	1.8839	60	60	1.000	0.999	-0.053	0.999	-0.053
				切片 β_0^=	-39.572	-14.822	1.795	-14.822
				傾き β_1^=	22.041	和 $\ln L$	0.045	和 $\ln L$ =
					最尤解			

線形パラメータの場合は，初期値として $\hat{\beta}_0 = 0$ と $\hat{\beta}_1 = 0$ などのようにとんでもない初期値を入れても対数尤度の最大化できるのであるが，非線形パラメータを μ_{SEV} と σ_{SEV} の場合には，初期値の設定には注意が必要である．

位置パラメータ μ_{SEV} には，出現率が 50%となるような対数用量 $\hat{\mu}_{SEV} = 1.8$ を与え，形状パラメータ σ_{SEV} には，対数用量の範囲の 1/4 程度 $\sigma_{SEV} = 0.05$ を与え，手順 6）と同様にソルバーで対数尤度を最大化すると，$\hat{\mu}_{SEV} = 1.795$ および $\sigma_{SEV} = 0.045$ を得る．この結果から回帰係数は，

$$\beta_0 = -\frac{\mu_{SEV}}{\sigma_{SEV}} = -\frac{1.795}{0.045} = -39.57, \qquad \beta_1 = \frac{1}{\sigma_{SEV}} = 22.04$$

のように計算でき，回帰係数を用いた最尤法の結果と一致することが確認できる．

6.3　情報行列を用いた最尤法

前章までは，Excel の基本機能のみで手作業によって反復計算を行ってきたが，Excel のソルバーを使って手作業なしに最尤解が求められることを表 6.1 で示した．

反復重み付き回帰によりシグモイド曲線に累積最小極値分布を用いる補 2 重対数法については，すでに示した．各用量の対数尤度 $\ln L_i$ は，

$$\ln L_i = \ln\binom{n_i}{y_i} + y_i \ln(\pi_i) + (n_i - y_i)\ln(1 - \pi_i)$$

であり，

$$\pi_i = 1 - \exp\left[-\exp\left(\beta_0 + \beta_1 x_i\right)\right], \quad 1 - \pi_i = \exp\left[-\exp\left(\beta_0 + \beta_1 x_i\right)\right]$$

を代入し整理すると，

$$\ln L_i = \ln\binom{n_i}{y_i} + y_i \ln\left\{1 - \exp\left[-\exp\left(\beta_0 + \beta_1 x_i\right)\right]\right\} - (n_i - y_i)\exp\left(\beta_0 + \beta_1 x_i\right)$$

が得られる．β_0 と β_1 で偏微分してスコア関数は，

$$U_{i1} = \frac{\partial \ln L_i}{\partial \beta_0} = \frac{y_i(1-\pi_i)z_i}{\pi_i} - (n_i - y_i)z_i$$

$$U_{i2} = \frac{\partial \ln L_i}{\partial \beta_1} = U_{i1} x_i$$

$$\text{ただし } z_i = \exp\left(\beta_0 + \beta_1 x_i\right)$$

となる．さらに β_0 と β_1 で偏微分して，２階の偏微分の負を取って情報行列の要素を計算する

$$I_{i11} = -\frac{\partial^2 \ln L_i}{\partial \beta_0 \beta_0} = -\left[\frac{y_i(1-\pi_i)(\pi_i - z_i)z_i}{\pi_i^2} - (n_i - y_i)z_i\right]$$

$$I_{i12} = I_{i11} x_i$$

$$I_{i21} = I_{i12}$$

$$I_{i22} = I_{i11} x_i^2$$

表 6.2 で対数尤度を最大化した計算シートに 表 6.3 に情報行列の各行の計算式を入れて，分散共分散行列を計算した結果を示す. Excel ソルバーを用いずに繰り返し計算を示してきたのは，自ら対数尤度の最大化の過程を体得し実感することが応用力をつけることになると判断したからである.

表 6.3　Excl ソルバーで対数尤度の最大化後の情報行列の計算

		対数	死亡数	合計	出現率	推定値	対数尤度		行列 $\boldsymbol{I}$ の要素			
i	切片	用量 x	y_i	n_i	p_i	π_i	$\ln(L_i)$	z_i	I_{i11}	I_{i12}	I_{i21}	I_{i22}
1	1	1.6907	6	59	0.102	0.095	-1.792	0.100	5.564	9.407	9.407	15.904
2	1	1.7242	13	60	0.217	0.188	-2.242	0.208	11.049	19.050	19.050	32.847
3	1	1.7552	18	62	0.290	0.338	-2.521	0.412	21.353	37.478	37.478	65.781
4	1	1.7842	28	56	0.500	0.542	-2.444	0.782	30.032	53.583	53.583	95.604
5	1	1.8113	52	63	0.825	0.758	-2.860	1.420	36.164	65.504	65.504	118.648
6	1	1.8369	53	59	0.898	0.918	-1.912	2.497	35.417	65.057	65.057	119.502
7	1	1.8610	61	62	0.984	0.986	-0.999	4.247	16.687	31.054	31.054	57.792
8	1	1.8839	60	60	1.000	0.999	-0.053	7.036	2.246	4.231	4.231	7.971
					切片 β_0^=	-39.572	-14.822		158.51	285.37	285.37	514.05
					傾き β_1^=	22.041	和 $\ln L$		158.51	285.37	10.426	-5.788
									285.37	514.05	-5.788	3.215
									$\boldsymbol{I}$		$\boldsymbol{I}^{-1}$分散共分散	

複雑な対数尤度関数を正しく偏微分することは，なかなか難儀でもあるし，符号の取り違えもしばしば経験する．スコアベクトルおよび情報行列を用いたニュートン・ラフソン法で尤度を最大化できれば，偏微分式が正しいことの確証が得られる．

手順 1）初期値として，切片 $\hat{\beta}_0 = 0$ と傾き $\hat{\beta}_1 = 0$ を初期値とする．

手順 2）出現率の推定値として

$$\pi_1 = 1 - \exp\left[-\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_1\right)\right]$$

=1-exp(-exp(0+0*1.6907))

=1-exp(-1)= 0.632

として各セルの計算する．

手順 3）対数尤度 $\ln L_1$ を 2 項分布関数を使って計算する．

$\ln L_1$ =Ln(Binimdist(y_1,n_1,π_1,false))

= Ln(Binimdist(6,59,0.632,false))

=-38.129

手順 4） $\ln L_i$ の合計 $\ln L$ を計算する．

$\ln L$ =Sum($\ln L_i$ の範囲)=-155.954

手順 5）ソルバーのパラメータ設定で，目的セル ← $\ln L$ セル，目標値：最大値を選択，変数セルの選択 ← β_0 と β_1 の範囲　をセットし，解決 する．

手順 6）切片 $\hat{\beta}_0 = -39.572$ と傾き $\hat{\beta}_1 = 22.041$， $\ln L = -14.822$ が得られる．

手順 7）情報行列の $i = 1$ の要素を

$$I_{1,1,1} = -\left[\frac{y_1(1-\pi_1)(\pi_1 - z_1)z_1}{\pi_1^2} - (n_1 - y_1)z_1\right]$$

=-(6*(1-0.095)*(0.095-0.1)/0.095^2)-(59-6)*0.1)

=5.564

ただし，z_1=exp(-39.572+22.041*1.6907)=0.100

$I_{1,1,2} = I_{1,1,1}x_1$=5.564*1.6907=9.407

$I_{1,2,1} = I_{1,1,1}x_1$=5.564*1.6907=9.407

$I_{1,2,2} = I_{1,1,1}x_1^2$=5.564*1.6907^2=15.904

として計算する．

手順 8）情報行列の $\boldsymbol{I}_{\cdot 11}$ の要素を計算する．

$\boldsymbol{I}_{\cdot 11}$ =Sum（$\boldsymbol{I}_{i,1,1}$ の範囲）=158.51

：

手順 9）情報行列の $\boldsymbol{I}$ を (2×2) として整形する．

$$\boldsymbol{I}=\begin{array}{|c|c|}\hline 158.51 & 285.37 \\ \hline 285.37 & 514.05 \\ \hline\end{array}$$

手順 10）分散共分散行列を $\boldsymbol{I}^{-1}$ で計算する．

$\boldsymbol{I}^{-1}$ =Minverse（$\boldsymbol{I}$ の範囲）=

10.426	-5.788
-5.788	3.215

6.4　分散共分散行列を用いたシグモイド曲線の 95%信頼区間

推定された回帰パラメータにより出現率の推定値 $\hat{\pi}_i$ は，

$$\hat{\pi}_i = 1-\exp\left[-\exp\left(\hat{\beta}_0+\hat{\beta}_1 x_i\right)\right]$$

と計算されている．回帰直線の推定値は，補 2 重対数変換

$$\hat{\eta}_i = \ln\left[-\ln(1-\pi_i)\right]=\hat{\beta}_0+\hat{\beta}_1 x$$

であり，この推定値の 95%信頼区間は，5.3 節と同様な考え方で推測できる分散共分散行列は，

$$\boldsymbol{I}^{-1}=\begin{bmatrix} Var(\beta_0) & Cov(\beta_0,\beta_1) \\ Cov(\beta_0,\beta_1) & Var(\beta_1) \end{bmatrix}$$

となることを活用して，ベクトル $\boldsymbol{x}_i=(1, x_{1,i})$ を挟んだ 2 次形式で分散を

$$Var(\hat{\eta}_i)=Var(\hat{\beta}_0+\hat{\beta}_1 x_{1,i})=\boldsymbol{x}_i\boldsymbol{I}^{-1}\boldsymbol{x}_i^T$$

で推定する．

$$\mathrm{L95\%}(\hat{\eta}_i)=\hat{\eta}_i-1.96\sqrt{Var(\hat{\eta}_i)}$$
$$\mathrm{U95\%}(\hat{\eta}_i)=\hat{\eta}_i+1.96\sqrt{Var(\hat{\eta}_i)}$$

で計算し，回帰直線上の 95%信頼区間を逆補 2 重対数変換

$$\mathrm{L95\%}(\hat{\pi}_i)=1-\exp\left\{-\exp\left[\mathrm{L95\%}(\hat{\eta}_i)\right]\right\}$$
$$\mathrm{U95\%}(\hat{\pi}_i)=1-\exp\left\{-\exp\left[\mathrm{U95\%}(\hat{\eta}_i)\right]\right\}$$

表 6.4 推定されたシグモイド曲線の 95%信頼区間

		対数	死亡数	合計	出現率	推定値	回帰直線について					πについて	
i	切片	用量 x	y_i	n_i	p_i	π_i	η	η^	分散	L95%	U95%	L95%	U95%
1	1	1.6907	6	59	0.102	0.095	-2.233	-2.307	0.045	-2.723	-1.892	0.064	0.140
2	1	1.7242	13	60	0.217	0.188	-1.410	-1.569	0.025	-1.878	-1.259	0.142	0.247
3	1	1.7552	18	62	0.290	0.338	-1.070	-0.886	0.013	-1.108	-0.664	0.281	0.403
4	1	1.7842	28	56	0.500	0.542	-0.367	-0.246	0.007	-0.412	-0.081	0.484	0.602
5	1	1.8113	52	63	0.825	0.758	0.557	0.351	0.007	0.190	0.511	0.702	0.811
6	1	1.8369	53	59	0.898	0.918	0.827	0.915	0.011	0.713	1.117	0.870	0.953
7	1	1.8610	61	62	0.984	0.986	1.418	1.446	0.018	1.182	1.710	0.962	0.996
8	1	1.8839	60	60	1.000	0.999	-	1.951	0.029	1.618	2.284	0.994	1.000

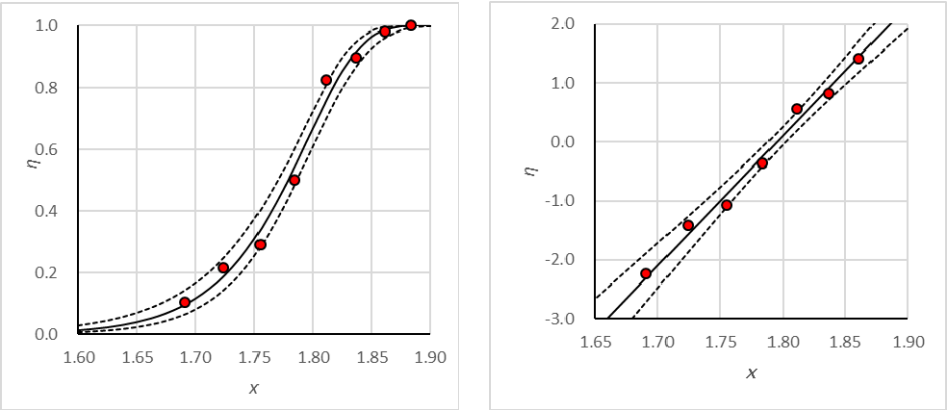


図 6.1　補 2 重対数変換に対する直線および出現率に対する最小極値分布のあてはめ

手順 1）推定値π_1は，推定された回帰係数 $\hat{\beta}_0 = -39.572$ と $\hat{\beta}_1 = 22.041$ を用いて次のように計算される．

$$\hat{\pi}_i = 1 - \exp\left[-\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right]$$

=1-exp(-exp(-39.572+22.041*1.6907))

=1-exp(-0.100)= 0.095

手順 2）出現率 p_1 についての補 2 重対数変換値η_1，推定値$\hat{\eta}_1$は，推定された回帰係数 $\hat{\beta}_0 = -39.572$ と $\hat{\beta}_1 = 22.041$ を用いて次のように計算される．

$$\eta_1 = \ln\left[-\ln(1-p_1)\right]$$ =Ln(-Ln(1-0.095))=-2.233

$$\hat{\eta}_1 = \ln\left[-\ln(1-\pi_1)\right] = \beta_0 + \beta_1 x$$ =-39.572+22.041*1.6907=-2.307

手順 3）出現率の推定値$\hat{\eta}_1$の分散$Var(\hat{\eta}_1)$は，ベクトル $\boldsymbol{x}_1 = (1, x_{1,1})$ を挟んだ 2 次形式

$$Var(\hat{\eta}_1) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_{1,1}) = \boldsymbol{x}_1 \boldsymbol{I}^{-1} \boldsymbol{x}_1^T$$

で推定する．

$$Var(\hat{\eta}_1) = \begin{bmatrix} 1 & 1.6907 \end{bmatrix} \begin{bmatrix} 10.426 & -5.788 \\ -5.788 & 3.215 \end{bmatrix} \begin{bmatrix} 1 \\ 1.6907 \end{bmatrix} = 0.045$$

手順 4）推定値 $\hat{\eta}_1$ の 95%信頼区間は，次のように計算される．

$$\mathrm{L95\%}(\hat{\eta}_1) = -2.233-1.96*\mathrm{sqrt}(0.045) = -2.723$$

$$\mathrm{U95\%}(\hat{\eta}_1) = -2.233+1.96*\mathrm{sqrt}(0.045) = -1.892$$

手順 5）推定値 π_1 は，次式で計算される．

$$\mathrm{L95\%}(\hat{\pi}_1) = 1-\mathrm{Exp}(-\mathrm{Exp}(-2.723)) = 0.064$$

$$\mathrm{U95\%}(\hat{\pi}_i) = 1-\mathrm{Exp}(-\mathrm{Exp}(-1.892)) = 0.140$$

手順 6）適当な初期用量，増分，最終用量を Excel シート上に展開し，手順 1）から手順 5）までの推定値を別途計算し，Excel の散布図を用いて図 6.1 に示すように作図する．

7. 情報行列を用いたプロビット法によるシグモイド曲線の推定

7.1　位置パラメータ μ と形状パラメータ σ の推定

反復重み付き回帰を使うプロビット法については，第 2 章で詳しく示した．情報行列を用いるプロビット法については，ロジット法とほぼ同様の結果となるので，これまで示してこなかった．

対数濃度 x_i，$i=1,2,...,n$ に対する出現率 π_i とする累積正規分布のシグモイド曲線は，累積標準正規分布を Φ としたときに, 位置パラメータを μ , 形状パラメータを σ とした場合に, 次のように定義される．

$$\pi_i = \Phi\left(\frac{x_i - \mu}{\sigma}\right)$$

この式を,

$$\pi_i = \Phi\left(-\frac{\mu}{\sigma} + \frac{1}{\sigma}x_i\right)$$

と線形式に変換することにより,

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

として

$$\pi_i = \Phi\left(\beta_0 + \beta_1 x_i\right)$$

のように線形式とすることができる．線形化することにより，μ と σ を用いるよりも対数尤度のパラメータに関する偏微分式が簡素化されるメリットがある．対数尤度 L は，π_i についての 2 項誤差の対数を L_i としたときに

$$\ln L = \sum_{i=1}^{n} \ln L_i = \sum_{i=1}^{n} \ln\left[\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}\right]$$

で定義される．Excel には 2 項分布の確率関数があるので,

$$\ln L_i = \ln\left[\binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i}\right]$$ =Ln(Binimdist(yi,ni,πi,false))

によって計算すことができる．Excel のソルバーを用いて，$\ln L$ を最大化するようなパラメータ β_0 および β_1 を求める．表 7.1 に示す Excel シートに沿った手順を示す．

手順 1）初期値として $\hat{\beta}_0 = 0$ と $\hat{\beta}_1 = 0$ をセットする．

手順 2）推定値として $\pi_i = \Phi(\beta_0 + \beta_1 x_i)$ をセル毎に計算する．

$\pi_i =$ Normdist($\beta_0 + \beta_1 x_i$,0,1,true)

手順 3）対数尤度 $\ln L_i$ を 2 項分布関数を使って計算する．

$\ln L_i$ =Ln(Binimdist(y_i,n_i,π_i,false))

手順 4） $\ln L_i$ の合計 $\ln L$ を計算する．

$\ln L$ =Sum($\ln L_i$ の範囲)

手順 5）ソルバーのパラメータ設定で，目的セル ← $\ln L$ セル，目標値：最大値を選択，変数セルの選択 ← β_0 と β_1 の範囲をセットし，解決 する．

表 7.1　カブト虫データに Excel ソルバーを用いたプロビット法での最尤解

					プロビット		プロビット			プロビット	
	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_i	2) π_i	3) $\ln L_i$	π_i	$\ln L_i$		π_i	$\ln(L_i)$
1	1.6907	6	59	0.102	0.500	-23.272	0.057	-2.680		0.057	-2.680
2	1.7242	13	60	0.217	0.500	-12.316	0.179	-2.366		0.179	-2.366
3	1.7552	18	62	0.290	0.500	-7.822	0.379	-3.267		0.379	-3.267
4	1.7842	28	56	0.500	0.500	-2.243	0.604	-3.478		0.604	-3.478
5	1.8113	52	63	0.825	0.500	-16.522	0.788	-2.312		0.788	-2.312
6	1.8369	53	59	0.898	0.500	-23.272	0.904	-1.785		0.904	-1.785
7	1.8610	61	62	0.984	0.500	-38.848	0.962	-1.494		0.962	-1.494
8	1.8839	60	60	1.000	0.500	-41.589	0.987	-0.777		0.987	-0.777
				切片 β_0^=	0.000	-165.884	-34.935	-18.159	μ ^=	1.7709	-18.159
				傾き β_1^=	0.000	4) 和 $\ln L$	19.728	和 $\ln L$	σ ^=	0.0507	和 $\ln L$ =
				1) 初期値		5)	最尤解			6) 別解	

手順 6）別解：正規分布の位置パラメータを $\mu = 1.8$，形状パラメータを $\sigma = 0.1$ を，初期値としてソルバーによって対数尤度を最大化する．ただし，π_i =Normdist(x_i, μ, σ,true)．推定された $\hat{\mu} = 1.7709$，$\hat{\sigma} = 0.0507$ より，回帰パラメータ

$$\hat{\beta}_0 = -\frac{\hat{\mu}}{\hat{\sigma}} = -\frac{1.7709}{0.0507} = -34.93, \quad \hat{\beta}_1 = \frac{1}{\hat{\sigma}} = \frac{1}{0.0507} = 19.72$$

が得られる．なお，求められた対数尤度には，定数項である組み合わせ数も含まれていることに注意が必要で，統計ソフトと異なる場合がある．

Excel のソルバーの良さは，一般化線形モデルの形式にとらわれず，表 7.1 に示すように位置パラメータを μ および 形状パラメータ σ を用いた非線形のままでも最尤解を求めることができる．ただし，パラメータの比が含まれているので，初期値の設定をいい加減にすると発散してしまい，解がもとまらないこともあるので，手作業で対数尤度が小さくなるよう

に初期値を探索する．初期値としては，位置パラメータ μ は，出現率 $p=0.5$ となる対数用量を，尺度パラメータ σ は，正規分布の $\mu-\sigma$ または $\mu+\sigma$ となるような出現率 0.16 または 0.841 の対数用量から μ との距離を用いる．

「最尤法とは，どのような方法でです」かとの質問にたいして，「尤度を最大化する方法です」との答えは，禅問答のごとくであり，解答になっていない．ここで示した方法に即していえば，「ある関数による推定値が得られた時，それらに対応する測定値に対して何らかの分布関数を仮定して確率(密度)とし，それらの積が最大となるような関数のパラメータを得る方法を最尤法と言う」，と一般化できる．また，「尤度とは，何ですか」との質問には，「確率(密度)の積です」との答えでどうだろうか．

7.2　情報行列を用いた最尤法

表 7.1 で示，Excel のソルバーを使って簡単に最尤解が求められることを示した．しかし，各種の 95%信頼区間の推定に必要な分散共分散行列を得ることができない．そこで，対数尤度について 2 階の偏微分式を Excel シート上に埋め込み，分散共分散行列を求める．

各用量の対数尤度 $\ln L_i$ は，

$$\ln L_i = \ln\binom{n_i}{y_i} + y_i \ln(\pi_i) + (n_i - y_i)\ln(1-\pi_i)$$

なので，

$$\pi_i = \Phi\left(\beta_0 + \beta_1 x_i\right)$$

を代入し整理すると，

$$\ln L_i = \ln\binom{n_i}{y_i} + y_i \ln\left[\Phi\left(\beta_0 + \beta_1 x_i\right)\right] - (n_i - y_i)\ln\left[1-\Phi\left(\beta_0 + \beta_1 x_i\right)\right]$$

が得られる．β_0 と β_1 で偏微分してスコア関数は，

$$\begin{aligned}
U_{i1} &= \frac{\partial \ln L_i}{\partial \beta_0} \\
&= y_i \frac{\phi\left(\beta_0 + \beta_1 x_i\right)}{\Phi\left(\beta_0 + \beta_1 x_i\right)} - (n_i - y_i)\frac{\phi\left(\beta_0 + \beta_1 x_i\right)}{1-\Phi\left(\beta_0 + \beta_1 x_i\right)} \\
&= \phi\left(z_i\right)\left[\frac{y_i}{\Phi\left(z_i\right)} - \frac{n_i - y_i}{1-\Phi\left(z_i\right)}\right]
\end{aligned}$$

$$U_{i2} = \frac{\partial \ln L_i}{\partial \beta_1} = U_{i1} x_i$$

ただし，$z_i = \beta_0 + \beta_1 x_i$

となる．さらにβ_0とβ_1による 2 階の偏微分の負を取って情報行列の要素を計算する．

$$
\begin{aligned}
I_{i11} &= -\frac{\partial^2 \ln L_i}{\partial\beta_0\beta_0} \\
&= -\phi(\beta_0+\beta_1 x_i)\left\{\begin{array}{l} -(\beta_0+\beta_1 x_i)\left[\dfrac{y_i}{\Phi(\beta_0+\beta_1 x_i)}-\dfrac{n_i-y_i}{1-\Phi(\beta_0+\beta_1 x_i)}\right] \\ -\dfrac{y_i\phi(\beta_0+\beta_1 x_i)}{\Phi(\beta_0+\beta_1 x_i)^2}-\dfrac{(n_i-y_i)\phi(\beta_0+\beta_1 x_i)}{\left[1-\Phi(\beta_0+\beta_1 x_i)\right]^2} \end{array}\right\} \\
&= \phi(z_i)\left\{z_i\left[\frac{y_i}{\Phi(z_i)}-\frac{n_i-y_i}{1-\Phi(z_i)}\right]+\phi(z_i)\left[\frac{y_i}{\Phi(z_i)^2}+\frac{(n_i-y_i)}{\left[1-\Phi(z_i)\right]^2}\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
I_{i12} &= I_{i11}x_i \\
I_{i21} &= I_{i12} \\
I_{i22} &= I_{i11}x_i^2
\end{aligned}
$$

表 6.2 で対数尤度を最大化した計算シートに 表 6.3 に情報行列の各行の計算式を入れて，分散共分散行列を計算した結果を示す．Excel ソルバーを用いずに繰り返し計算を示してきたのは，自ら対数尤度の最大化の過程を体得し実感することが応用力をつけることになると期待している．

複雑な対数尤度関数を正しく偏微分することは，なかなか難儀でもあるし，符号の取り違えもしばしば経験する．計算したスコアベクトルおよび情報行列をもちいた，ニュートン・ラフソン法で尤度を最大化できれば，偏微分式が正しいことの確証が得られる．

表 7.2　Excel ソルバーで対数尤度の最大化後のプロビット情報行列の計算

		対数	死亡数	合計	出現率	推定値	対数尤度		行列 $\boldsymbol{I}$ の要素			
i	切片	用量 x	y_i	n_i	p_i	π_i	$\ln(L_i)$	z_i	I_{i11}	I_{i12}	I_{i21}	I_{i22}
1	1	1.6907	6	59	0.102	0.057	-2.680	-1.581	16.076	27.179	27.179	45.952
2	1	1.7242	13	60	0.217	0.179	-2.366	-0.920	28.798	49.654	49.654	85.613
3	1	1.7552	18	62	0.290	0.379	-3.267	-0.309	37.385	65.618	65.618	115.173
4	1	1.7842	28	56	0.500	0.604	-3.477	0.263	35.430	63.214	63.214	112.786
5	1	1.8113	52	63	0.825	0.788	-2.313	0.798	30.879	55.931	55.931	101.307
6	1	1.8369	53	59	0.898	0.904	-1.785	1.303	19.934	36.616	36.616	67.260
7	1	1.8610	61	62	0.984	0.962	-1.494	1.778	10.566	19.663	19.663	36.594
8	1	1.8839	60	60	1.000	0.987	-0.777	2.230	4.566	8.602	8.602	16.205
					切片 β_0^=	-34.935	-18.159		183.63	326.48	326.48	580.89
					傾き β_1^=	19.728	和 $\ln L$		183.63	326.48	6.967	-3.916
									326.48	580.89	-3.916	2.202
									$\boldsymbol{I}$		$\boldsymbol{I}^{-1}$分散共分散	

手順 1）から 手順 6）までは，表 7.1 での手順と同じなので省略する．

手順 6）切片 $\hat{\beta}_0 = -34.935$ と傾き $\hat{\beta}_1 = 19.728$， $\ln L = -18.195$ が得られる．

手順 7）情報行列の $i=1$ の要素を

$$I_{i11} = \phi(z_1)\left\{ z_1\left[\frac{y_1}{\Phi(z_1)} - \frac{(n_1 - y_1)}{1-\Phi(z_1)}\right] + \phi(z_1)\left[\frac{y_1}{\Phi(z_1)^2} + \frac{(n_i - y_1)}{\left[1-\Phi(z_1)\right]^2}\right]\right\}$$

```
=0.1143*( -1.581*(6/0.0569-(59-6)/(1-0.0569))
          +0.1143*(6/0.0569^2)+(59-6)/(1-0.0569)^2)
= 16.076
```

ただし，z_1=-34.935+19.728*1.6907=-1.581

$\phi(z_1)$=Normdist(-1.581,0,1,fals)=0.1143

$\Phi(z_1)$=Normdist(-1.581,0,1,true)=0.0569

$I_{1,1,2} = I_{1,1,1}x_1$=16.076*1.6907=27.179

$I_{1,2,1} = I_{1,1,1}x_1$=16.076*1.6907=27.179

$I_{1,2,2} = I_{1,1,1}x_1^2$=16.076*1.6907^2=45.952

として計算する．

手順 8）情報行列の $\boldsymbol{I}_{\cdot 11}$ の要素を計算する．

$\boldsymbol{I}_{\cdot 11}$=Sum（$\boldsymbol{I}_{i,1,1}$ の範囲）= 183.63

：

手順 9）情報行列の $\boldsymbol{I}$ を (2×2) として整形する．

$\boldsymbol{I}=$

183.63	326.48
326.48	580.89

手順 10）分散共分散行列を $\boldsymbol{I}^{-1}$ で計算する．

$\boldsymbol{I}^{-1}$=Minverse（$\boldsymbol{I}$ の範囲）=

6.967	-3.916
-3.916	2.202

分散共分散行列が計算できれば，これまで繰り返し示してきたシグモイド曲線の 95%信頼区間の計算，50%出現率の 95%信頼区間などを計算することができる．

8. 3 種のシグモイド曲線の比較

8.1 位置パラメータと形状パラメータを用いた比較

これまでは，用量反応曲線にロジスティック分布，正規分布，最小極値分布のあてはめを，反復重み付き回帰による最尤法，情報行列を用いた最尤法それぞれについて行ってきた．用量反応曲線に正規分布を使うプロビット法とロジスティック分布を使うロジット法は，供に点対称であり，ほとんど同様なシグモイド曲線となる．最小極値分布を用いる補 2 重対数法は，低濃度に長く裾を引くシグモイド曲線となることを示してきた．

そこで，Excel のソルバーによって対数尤度を最大化する方法で，これら 3 種のシグモイド曲線を重ね書きして形状の違いを観察する．スライド 21 のロジットの欄では，線形式 $z_i = \beta_0 + \beta_1 x_i$ を用いた推定値が示されている．プロビットと補 2 重対数の欄では，$\hat{\mu}$ と $\hat{\sigma}$ を推定値としている．線形式を用いると 2 階の偏微分式が関係になるのであるが，$\hat{\mu}$ と $\hat{\sigma}$ を使う場合には非線形となり偏微分式が複雑になるので避けてきた．

スライド 21

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	π_i^	$\ln(L_i)$	π_i^	$\ln(L_i)$	π_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 β_0^=	-60.716	μ ^=	1.771	μ ^=	1.795
					傾き β_1^=	34.270	σ ^=	0.051	σ ^=	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

2017.3.16 高橋行雄　　21

線形式を用いた場合は，推定されたパラメータから，シグモイド曲線の形状を想定することは困難である．プロビット法の場合，対数濃度x_iに対する出現率π_iのシグモイド曲線は，累積標準正規分布をΦとしたときに，位置パラメータをμ_{NOR}，形状パラメータをσ_{NOR}とした場合に，次のように定義される．

$$\pi_i = \Phi\left(\frac{x_i - \mu_{NOR}}{\sigma_{NOR}}\right)$$

この式を，

$$\pi_i = \Phi\left(-\frac{\mu_{NOR}}{\sigma_{NOR}} + \frac{1}{\sigma_{NOR}} x_i\right)$$

のように線形式に変換することにより，

$$\beta_0 = -\frac{\mu_{NOR}}{\sigma_{NOR}}, \quad \beta_1 = \frac{1}{\sigma_{NOR}}$$

とすることにより，

$$\pi_i = \Phi\left(\beta_0 + \beta_1 x_i\right)$$

に線形式にすることができ，この式から対数尤度を定義し，パラメータβ_0とβ_1について偏微分した．

元の位置パラメータをμ_{NOR}，形状パラメータをσ_{NOR}としても，これらを用いて偏微分することは，可能ではあるが煩雑となる．Excle のソルバーを用いる場合には，内部で偏微分が行われるので，（μ_{NOR}，σ_{NOR}）を用いても煩雑さはなく，パラメータの意味づけは容易である．

ロジット法は，シグモイド曲線にロジスティック分布を用いていて，伝統的に次式

$$\pi_i = F_{LOG}(x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

を用いてきた．プロビット法で用いた位置パラメータをμ_{LOG}，形状パラメータをσ_{LOG}とした場合に，

$$\beta_0 = -\frac{\mu_{LOG}}{\sigma_{LOG}}, \quad \beta_1 = \frac{1}{\sigma_{LOG}}$$

で置きなおし整理すると

$$\pi_i = F_{LOG}(x_i) = \frac{\exp\left(\dfrac{x_i - \mu_{LOG}}{\sigma_{LOG}}\right)}{1 + \exp\left(\dfrac{x_i - \mu_{LOG}}{\sigma_{LOG}}\right)}$$

となり，推定されたパラメータのμ_{LOG}は，50%出現率の推定用量となる．もちろん，

$$\mu_{LOG} = -\frac{\beta_1}{\beta_0}, \quad \sigma_{LOG} = \frac{1}{\beta_1}$$

と置きなおせば，元の線形式になる．

補 2 重対数法は，シグモイド曲線に最小極値分布を，対数濃度 x_i に対して，位置パラメータを μ_{SEV}，形状パラメータを σ_{SEV} として次のように定義される．

$$\pi = F_{SEV}(x) = 1 - \exp\left[-\exp\left(\frac{x_i - \mu_{SEV}}{\sigma_{SEV}}\right)\right]$$

最小極値分布の場合の μ_{SEV} は，50%出現率の推定値ではなく，

$$F(x \mid \mu_{SEV}) = 1 - \exp\left[-\exp\left(\frac{x_i - \mu_{SEV}}{\sigma_{SEV}}\right)\right] = 1 - \exp\left[-\exp(0)\right] = 0.632$$

なので，63.2%出現率となる用量 $x_{p=0.632}$ の推定値になっている．

8.2　3 種のシグモイド曲線の形状の比較

スライド 21 の線形式を用いたロジット法に加えて，（μ_{LOG}，σ_{LOG}）を用いたロジット法，同様にプロビット法，補 2 重対数法で求めた最尤解表 8.1 を示す．

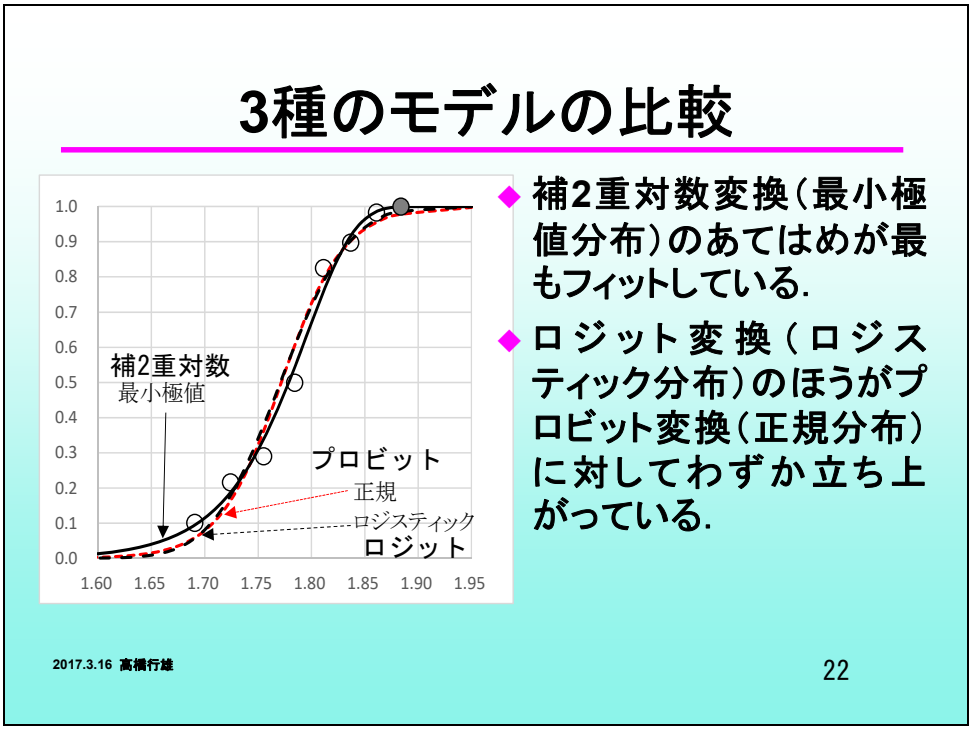
表 8.1　3 種のシグモイド曲線のパラメータ推定

					ロジット		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	推定値	対数尤度
i	用量 x	y_i	n_i	p	π_i	$\ln(L_i)$	π_i	$\ln(L_i)$	π_i	$\ln(L_i)$	π_i	$\ln(L_i)$
1	1.6907	6	59	0.102	0.059	-20.222	0.059	-20.223	0.057	-20.303	0.095	-19.415
2	1.7242	13	60	0.217	0.164	-31.921	0.164	-31.921	0.179	-31.639	0.188	-31.515
3	1.7552	18	62	0.290	0.362	-38.067	0.362	-38.067	0.379	-38.421	0.338	-37.674
4	1.7842	28	56	0.500	0.605	-40.087	0.605	-40.087	0.604	-40.051	0.542	-39.018
5	1.8113	52	63	0.825	0.795	-29.360	0.795	-29.360	0.788	-29.459	0.758	-30.006
6	1.8369	53	59	0.898	0.903	-19.407	0.903	-19.407	0.904	-19.408	0.918	-19.536
7	1.8610	61	62	0.984	0.955	-5.902	0.955	-5.902	0.962	-5.621	0.986	-5.126
8	1.8839	60	60	1.000	0.979	-1.270	0.979	-1.270	0.987	-0.777	0.999	-0.053
					和 $\ln L=$	-186.235	和 $\ln L=$	-186.235	和 $\ln L=$	-185.679	和 $\ln L=$	-182.343
					切片 β_0^=	-60.716	μ_{LOG}^=	1.772	μ_{NOR}^=	1.771	μ_{SEV}^=	1.795
					傾き β_1^=	34.270	σ_{LOG}^=	0.029	σ_{NOR}^=	0.051	σ_{SEV}^=	0.045

推定値 $\hat{\pi}_i$ の計算方法は，それぞれ異なるが，対数尤度 $\ln L_1$ の計算は，2 項分布の定数項を除いた式を用いている．対数尤度は，補 2 重対数法で -182.343 とロジット法の -186.235 に比べ 3.843 と大きく統計的は補 2 重対数法のあてはめが支持される．プロビット法がロジット法に比べ 0.556 大きいが，統計的にはほぼ同等の範囲内である．

Excel シート上で対数用量 x について細かな刻みでそれぞれの推定値 π を計算し Excel の散布図を用いてスライド 22 を作成した．

スライド 22



補 2 重対数法（最小極値分布）のあてはめが，全ての用量範囲で実験結果によく追従していることが読み取れる．ロジット法（ロジスティック分布）の方がプロビット法（正規分布）に対してわずか立ち上がっているが，ごくわずかである．

9.　JMP によるロジット・プロビット・補 2 重対数

9.1　統計ソフト JMP による一般化線形モデル

統計ソフト JMP（Ver13）を用いて，ロジット法，プロビット法，補 2 重対数法による解析を行う．JMP でのファイル形式は，Excel の場合とは異なり，表 9.1 に示すように反応 Y を名義尺度（0，1）とし，それぞれの出現数 n_i を度数が必要となる．目的とする反応ありを 0 とし反応なしを 1 とする．

表 9.1　カブト虫データの JMP でのデータ形式

	i	濃度 x	T	y	ni	p
1	1	1.6907	59	1	53	0.10
2	1	1.6907	59	0	6	0.10
3	2	1.7242	60	1	47	0.22
4	2	1.7242	60	0	13	0.22
5	3	1.7552	62	1	44	0.29
6	3	1.7552	62	0	18	0.29
7	4	1.7842	56	1	28	0.50
8	4	1.7842	56	0	28	0.50
9	5	1.8113	63	1	11	0.83
10	5	1.8113	63	0	52	0.83
11	6	1.8369	59	1	6	0.90
12	6	1.8369	59	0	53	0.90
13	7	1.8610	62	1	1	0.98
14	7	1.8610	62	0	61	0.98
15	8	1.8839	60	1	0	1.00
16	8	1.8839	60	0	60	1.00

JMP の「二変量の関係」では，目的変数に名義尺度の Y を，説明変数に濃度 x，度数に n_i と設定することにより，自動的にロジット法（ロジスティック回帰）が行われる．「二変量の関係」では，標準的なロジスティック回帰の出力に加えて，パラメータの推定で付加的に分散共分散行列が推定でき，任意のパーセント値に対する逆推定値と正確な 95%信頼区間のみならず 90%，あるいは，99%パーセントなど任意の信頼水準を設定することができる．さらに，ROC 曲線を出力することもできる．

JMP の「モデルのあてはめ」では，役割変数で名義尺度の Y を，度数に n_i をセットすると自動的に「名義ロジスティック」が選択され，モデル効果の構成で濃度 x を設定し実行する．出力内容は，「二変量の関係」と同様の出力結果である．

JMP の「モデルのあてはめ」で「名義ロジスティック」が自動的に選択された後に，プルダウン・メニューを展開し,「一般化線形モデル」を選択すると, 表 9.2 に示すように「分布」, および,「リンク関数」が現れる.「リンク関数」をプルダウンすると，ロジット，プロビット, 補対数・対数 (2 重補対数) などが現れる.「分布」をプルダウンすると, 正規, 2 項, Poisson, 指数が選択できる.

表 9.2　JMP による「一般化線形モデル」の選択

9.2　JMP によるロジット法

一般化線形モデルによるロジット法を例示する．手法で一般化線形モデル，分布で 2 項, リンク関数でロジットを選択する.

表 9.3　「一般化線形モデル」でのロジット法の設定

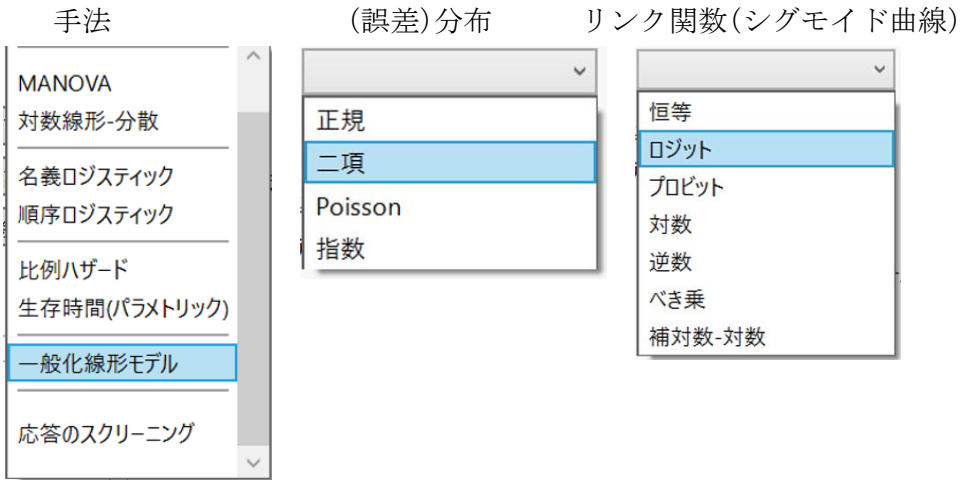
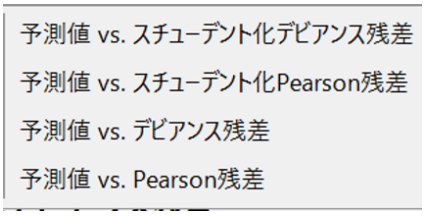


表 9.4　「一般化線形モデル」での各種のオプションの選択

オプションの選択

残差プロット



列の保存

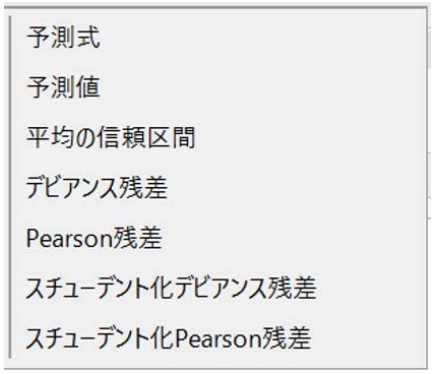


表 9.5　「一般化線形モデル」でのロジット法による標準出力

一般化線形モデルのあてはめ

度数: ni
応答: y
モデル化P(y=0)
分布: 二項
リンク関数: ロジット
推定法: 最尤法
オブザベーション(または重みの合計) = 481

モデル全体の検定

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
差分	136.485109	272.9702	1	<.0001*
完全	186.235403			
縮小	322.720512			

適合度統計量	カイ2乗	自由度	p値(Prob>ChiSq)
Pearson	436.2958	479	0.9194
デビアンス	372.4708	479	0.9999

AICc
376.4959

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(Prob>ChiSq)	下側信頼限界	上側信頼限界
切片	-60.71745	5.1807114	267.87418	<.0001*	-71.44235	-51.07882
濃度 x	34.270325	2.91214	272.97022	<.0001*	28.853907	40.300527

得られた結果を，表 5.4 の結果と比較する．まず，モデル全体の検定の「完全」の(-1)*対数尤度は，186.235403 で Excel の $\ln L = -186.24$ と一致する．パラメータの推定値の切片の推定値は，$\hat{\beta}_0 = -60.717$，濃度 x の推定値は，$\hat{\beta}_1 = 34.270$ に対応し一致する．

「列の保存」のプルダウン・メニューから「予測値」,「平均の信頼区間」を選択し，JMP テーブルに出力し，シート上で残差を計算する.

表 9.6　ロジット法でのシグモイド曲線の推定値および 95%信頼区間の保存

	i	濃度 x	T	y	ni	p	P(y) 計算式	残差	平均 yの下側95%	平均 yの上側95%
1	1	1.6907	59	1	53	0.10	0.059	0.043	0.034	0.098
2	1	1.6907	59	0	6	0.10	0.059	0.043	0.034	0.098
3	2	1.7242	60	1	47	0.22	0.164	0.053	0.116	0.227
4	2	1.7242	60	0	13	0.22	0.164	0.053	0.116	0.227
5	3	1.7552	62	1	44	0.29	0.362	-0.072	0.298	0.431
6	3	1.7552	62	0	18	0.29	0.362	-0.072	0.298	0.431
7	4	1.7842	56	1	28	0.50	0.605	-0.105	0.542	0.665
8	4	1.7842	56	0	28	0.50	0.605	-0.105	0.542	0.665
9	5	1.8113	63	1	11	0.83	0.795	0.030	0.739	0.842
10	5	1.8113	63	0	52	0.83	0.795	0.030	0.739	0.842
11	6	1.8369	59	1	6	0.90	0.903	-0.005	0.860	0.934
12	6	1.8369	59	0	53	0.90	0.903	-0.005	0.860	0.934
13	7	1.8610	62	1	1	0.98	0.955	0.029	0.926	0.973
14	7	1.8610	62	0	61	0.98	0.955	0.029	0.926	0.973
15	8	1.8839	60	1	0	1.00	0.979	0.021	0.960	0.989
16	8	1.8839	60	0	60	1.00	0.979	0.021	0.960	0.989

JMP テーブルに出力された推定式，および，95%信頼区間を「重ね合わせプロット」使って，グラフ化した結果を 図 9.1 に示す.

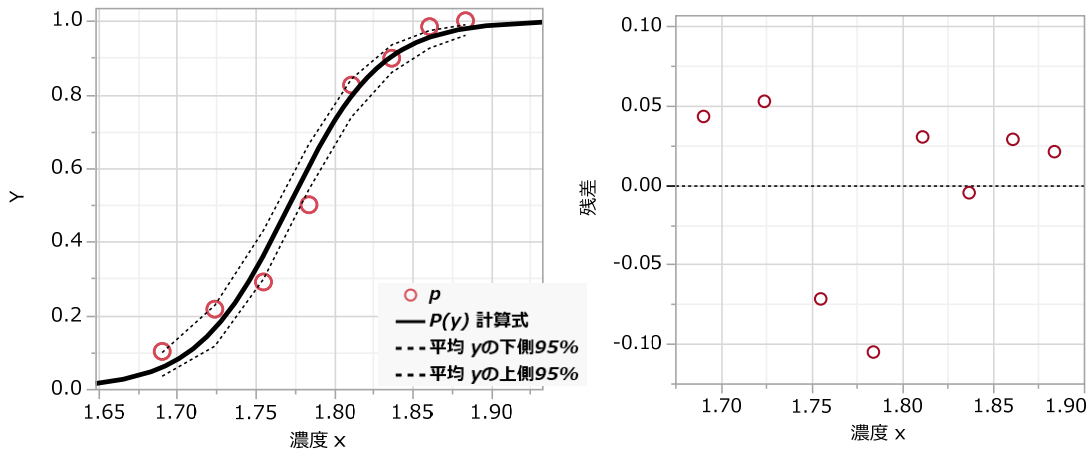


図 9.1　JMP の重ね合わせプロットでの作図（ロジット）

推定されたロジスティック曲線は，計算式が含まれているので，関数プロットを用いているために滑らかな曲線となっているが，95%信頼区間は，計算式が含まれていないので折れ

線になっている．残差プロットから残差が濃度 x に対して U 字型になっており，ロジット曲線のあてはめには，難点があることが示唆されている．

JMP の優れているのは，逆推定の機能が充実していることである．図 9.2 に幾つかのパーセント点を設定すると，その結果がきれいなグラフとして出力される．

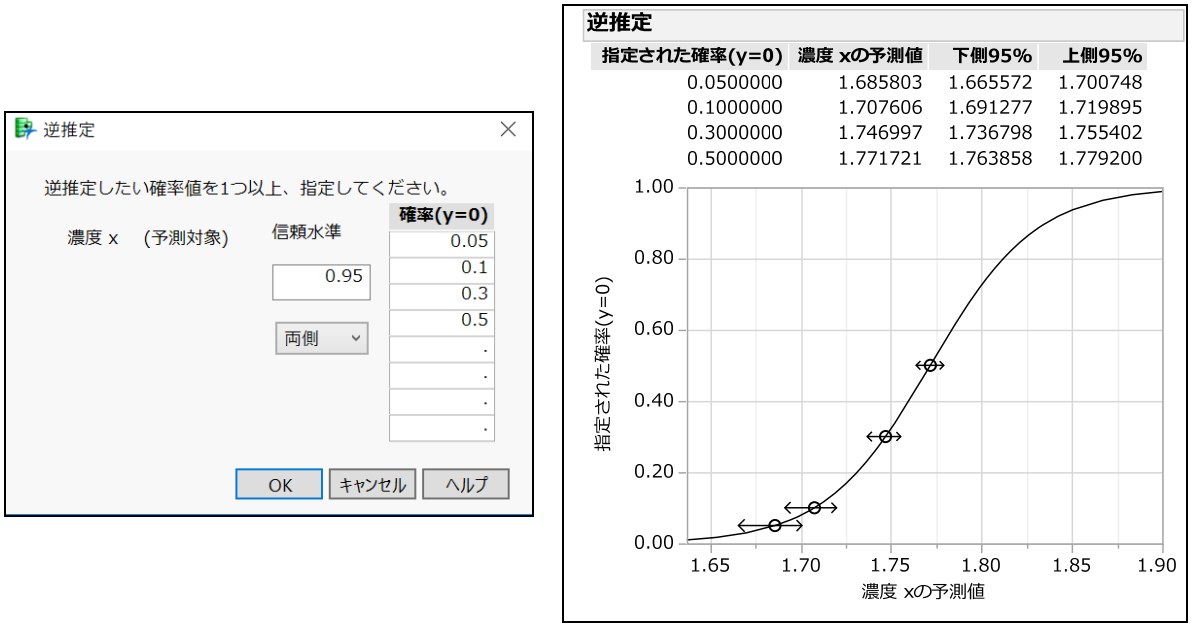


図 9.2　逆推定値の入力画面と出力結果（ロジット）

JMP の結果は，第 5 章で示した Excel での計算結果と一致する．統計ソフトだけでは，計算方法がブラックボックス的になりがちである．このように Excel を用いた計算を行うことで，統計ソフトの計算ロジックを理解することにより，統計ソフトが対応していない各種の応用が可能となる．

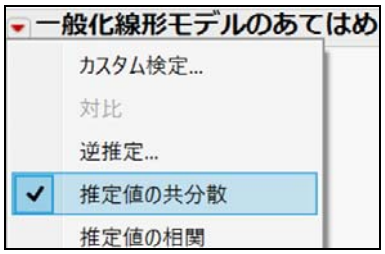
表 9.7　第 5 章で示した Excel での逆推定の 95%信頼区間の計算結果

回帰の95%信頼区間からの導出								
	p_0=	0.10	η_0=	-2.197	LD_{p0}=	1.708	L95%=	1.691
	a=	-864.6	b=	1013.9	c=	-297.2	U95%=	1.720
フィラーの式による導出								
	p_0=	0.50	η_0=	0.000	LD_{p0}=	1.772	L95%=	1.764
	a=	3583.5	b=	4045.7	c=	1141.9	U95%=	1.779

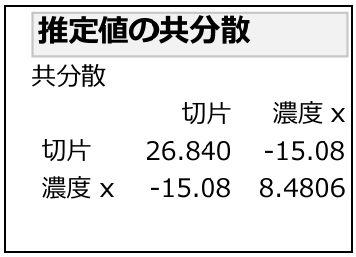
応用のためには，パラメータの推定値に加えて，表 9.8 に示すように分散共分散行列を Excel シートに取り込むことである．推定値および共分散を用いることにより，各種の推定問題が解けることを，これまで繰り返し示したきた．

表 9.8　共分散の出力（ロジット）

出力の指定



JMP の結果

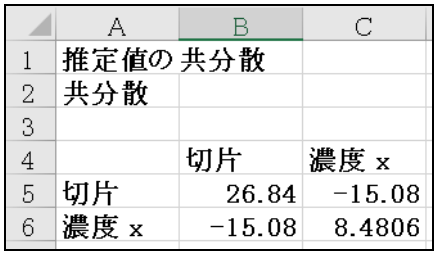


推定値の共分散

共分散

	切片	濃度 x
切片	26.840	-15.08
濃度 x	-15.08	8.4806

Excel シートへのペースト



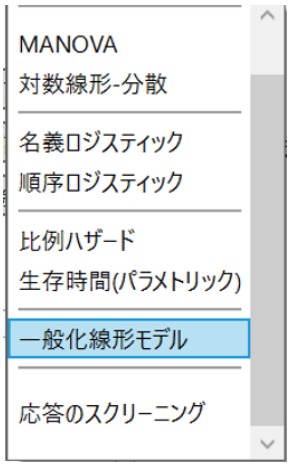
	A	B	C
1	推定値の共分散		
2	共分散		
3			
4		切片	濃度 x
5	切片	26.84	−15.08
6	濃度 x	−15.08	8.4806

9.3　JMP によるプロビット法

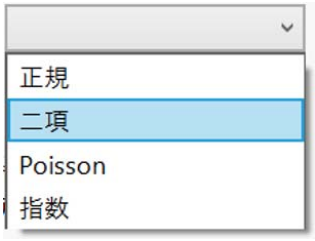
JMP でのプロビット法は，一般化線形モデルのリンク関数で「プロビット」を選択する．他の操作は，ロジットの場合と同様である．

表 9.9　「一般化線形モデル」でのプロビット法の設定

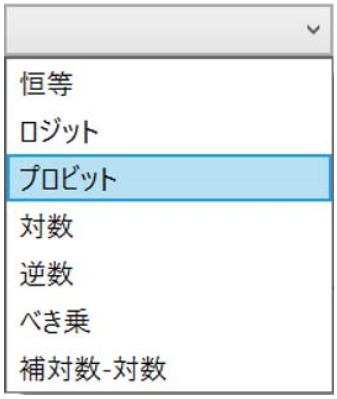
手法



(誤差)分布



リンク関数(シグモイド曲線)



出力結果を 表 9.10 に示す．得られた結果を，表 7.1 の結果と比較する．まず，パラメータの推定値の切片の推定値は，$\hat{\beta}_0 = -34.935$，濃度 x の推定値は，$\hat{\beta}_1 = 19.728$ に対応し一致する．モデル全体の検定の「完全」の(-1)*対数尤度は，185.679167 で Excel の $\ln L = -18.159$ とは一致しない．これは，JMP での対数尤度の計算式では，2 項分布の定数項を除いているた

めである．表 8.1 では，Excel での計算で JMP と同様の尤度の計算を行っており，$\ln L = -185.679$ と一致する．

表 9.10 「一般化線形モデル」でのプロビット法による出力

一般化線形モデルのあてはめ

重み: ni
応答: y
モデル化P(y=0)
分布: 二項
リンク関数: プロビット
推定法: 最尤法
オブザベーション(または重みの合計) = 481

モデル全体の検定

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
差分	137.041346	274.0827	1	<.0001*
完全	185.679167			
縮小	322.720512			

適合度統計量	カイ2乗	自由度	p値(Prob>ChiSq)
Pearson	435.4817	14	<.0001*
デビアンス	371.3583	14	<.0001*

AICc
376.2814

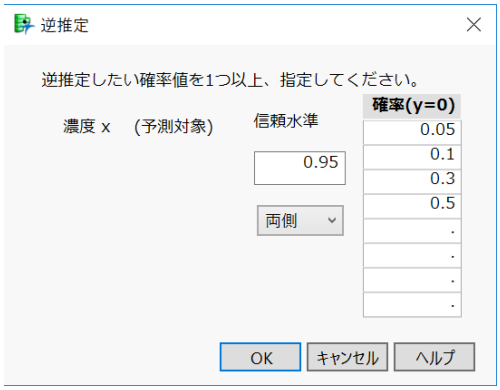
効果の検定

要因	自由度	尤度比カイ2乗	p値(Prob>ChiSq)
濃度 x	1	274.08269	<.0001*

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(Prob>ChiSq)	下側信頼限界	上側信頼限界
切片	-34.93526	2.6395037	268.94431	<.0001*	-40.28918	-29.92927
濃度 x	19.727934	1.4840583	274.08269	<.0001*	16.914808	22.739736

逆推定の結果を図 9.3 に示す．出現率が 10%の場合の推定値は，$probit\ LD_{10} = 1.706$，95%信頼区間は（1.691，1.717）となる．ロジット法の場合は，$logit\ LD_{10} = 1.708$，（1.691，1.720）とほぼ同様の推定値となっている．



逆推定

指定された確率(y=0)	濃度 xの予測値	下側95%	上側95%
0.0500000	1.687475	1.669956	1.700977
0.1000000	1.705891	1.691152	1.717431
0.3000000	1.744271	1.734546	1.752502
0.5000000	1.770852	1.763209	1.778184

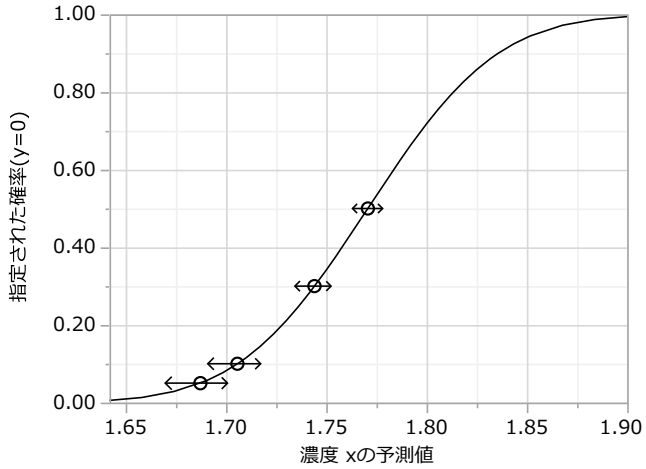


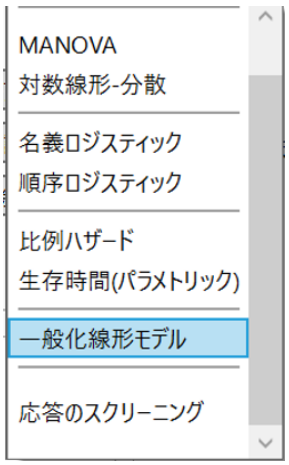
図 9.3　逆推定値の入力画面と出力結果（プロビット）

9.4　JMP による補 2 重対数法

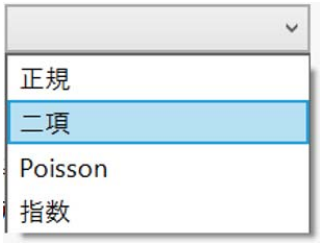
一般化線形モデルによる補 2 重対数法を例示する．手法で一般化線形モデル，分布で 2 項，リンク関数でロジットを選択する．

表 9.11　「一般化線形モデル」での補 2 重対数法の設定

手法



(誤差)分布



リンク関数(シグモイド曲線)

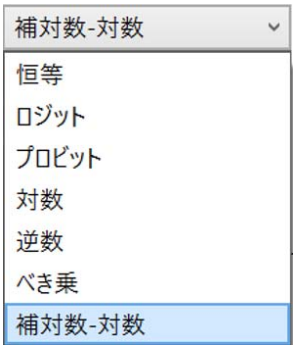


表 9.12　「一般化線形モデル」での補 2 重対数法による出力

一般化線形モデルのあてはめ

重み: ni
応答: y
モデル化P(y=0)
分布: 二項
リンク関数: 補対数-対数
推定法: 最尤法
オブザベーション(または重みの合計) = 481

モデル全体の検定

モデル	(-1)*対数尤度	尤度比カイ2乗	自由度	p値(Prob>ChiSq)
差分	140.378005	280.7560	1	<.0001*
完全	182.342507			
縮小	322.720512			

適合度統計量	カイ2乗	自由度	p値(Prob>ChiSq)
Pearson	437.0237	14	<.0001*
デビアンス	364.6850	14	<.0001*

AICc
369.6081

効果の検定

要因	自由度	尤度比カイ2乗	p値(Prob>ChiSq)
濃度 x	1	280.75601	<.0001*

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(Prob>ChiSq)	下側信頼限界	上側信頼限界
切片	-39.57231	3.229046	281.77577	<.0001*	-46.20362	-33.53859
濃度 x	22.04117	1.793088	280.75601	<.0001*	18.690245	25.722461

得られた結果を，表 6.1 の結果と比較する．まず，モデル全体の検定の「完全」の(-1)*対数尤度は，182.342507 で Excel の $\ln L = -182.343$ と一致する．パラメータの推定値の切片の推定値は，$\hat{\beta}_0 = -39.572$，濃度 x の推定値は，$\hat{\beta}_1 = 22.041$ に対応し一致する．

「列の保存」のプルダウン・メニューから「予測値」，「平均の信頼区間」を選択し，JMP テーブルに出力し，シート上で残差を計算する．

JMP テーブルに出力された推定式，および，95%信頼区間を「重ね合わせプロット」使って，グラフ化した結果を 図 9.4 に示す．残差プロットは，図 9.1 に示したロジット法の場合と比較すると全体的に 0 ラインに近づき，U 字型の残差構造が解消されている．そのために，それぞれの濃度で 2 項分布をあてはめた時の確率密度が全体的に大きくなり，それらの対数の和であらわされる対数尤度が，$\ln L = -182.343$ とロジット法での $logit \ln L = -186.235$ より 3.89 大きくなっていることからも，補 2 重対数法によるあてはめが統計的にもすぐれていることになる．

表 9.13　補 2 重対数法でのシグモイド曲線の推定値および 95%信頼区間の保存

	i	濃度 x	T	y	ni	p	P(y) 計算式	残差	平均 yの下側95%	平均 yの上側95%
1	1	1.6907	59	1	53	0.10	0.095	0.007	0.064	0.140
2	1	1.6907	59	0	6	0.10	0.095	0.007	0.064	0.140
3	2	1.7242	60	1	47	0.22	0.188	0.029	0.142	0.247
4	2	1.7242	60	0	13	0.22	0.188	0.029	0.142	0.247
5	3	1.7552	62	1	44	0.29	0.338	-0.048	0.281	0.403
6	3	1.7552	62	0	18	0.29	0.338	-0.048	0.281	0.403
7	4	1.7842	56	1	28	0.50	0.542	-0.042	0.484	0.602
8	4	1.7842	56	0	28	0.50	0.542	-0.042	0.484	0.602
9	5	1.8113	63	1	11	0.83	0.758	0.067	0.702	0.811
10	5	1.8113	63	0	52	0.83	0.758	0.067	0.702	0.811
11	6	1.8369	59	1	6	0.90	0.918	-0.019	0.870	0.953
12	6	1.8369	59	0	53	0.90	0.918	-0.019	0.870	0.953
13	7	1.8610	62	1	1	0.98	0.986	-0.002	0.962	0.996
14	7	1.8610	62	0	61	0.98	0.986	-0.002	0.962	0.996
15	8	1.8839	60	1	0	1.00	0.999	0.001	0.994	1.000
16	8	1.8839	60	0	60	1.00	0.999	0.001	0.994	1.000

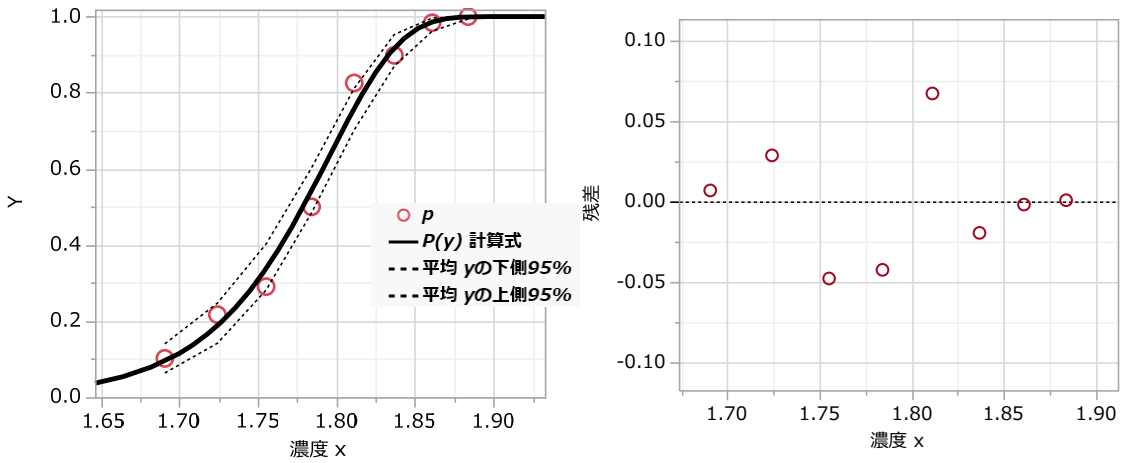
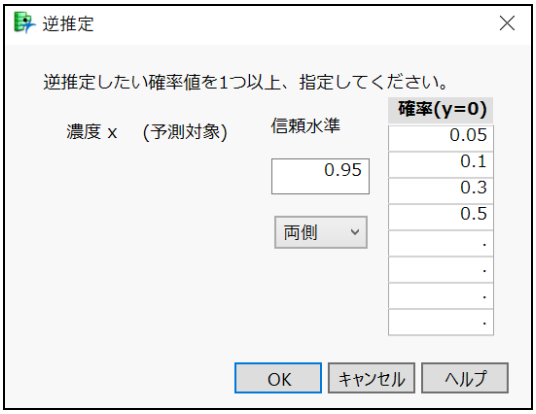


図 9.4　JMP の重ね合わせプロットでの作図（補 2 重対数）

補 2 重対数法による逆推定の結果を図 9.5 に示す．ロジット法，プロビット法の場合に比べて，低濃度域での信頼区間の幅が大きくなっている．

ロジット法と補 2 重対数法での逆推定の結果を出現率が 10%の場合について比較した結果を 表 9.14 に示す．元の濃度に戻した場合にロジット法の場合が，補 2 重対数法に比べて，大きめに推定され，ロジット法での予測値 51.004 は，補 2 重対数法の 95%上限 51.216 と同程度となっている．逆に，補 2 重対数法での予測値 49.350 は，ロジット法の 95%下限 49.122 と同程度となっている．



逆推定

指定された確率(y=0)	濃度 xの予測値	下側95%	上側95%
0.0500000	1.660625	1.633039	1.680925
0.1000000	1.693283	1.671580	1.709404
0.3000000	1.748609	1.736185	1.758336
0.5000000	1.778753	1.770217	1.786166

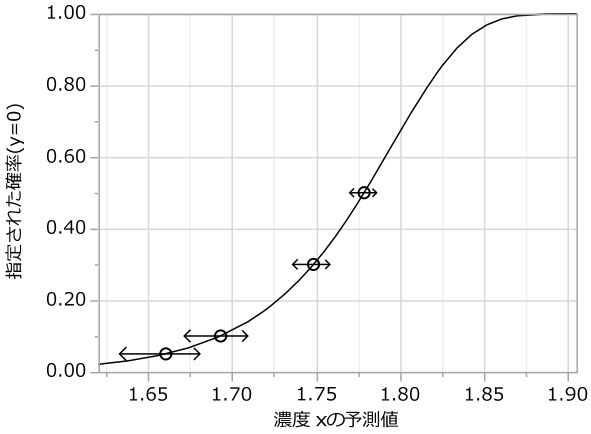


図 9.5　逆推定値の入力画面と出力結果（補 2 重対数）

表 9.14　ロジット法と補 2 重対数法で出現率が 10%の場合の推定値の比較

用量 x	出現率	推定値	ロジット	補2重対数	差
対数濃度	0.10	予測値	1.708	1.693	0.014
		L95%	1.691	1.672	0.020
		U95%	1.720	1.709	0.010
濃度	0.10	予測値	51.004	49.350	1.655
		L95%	49.122	46.944	2.178
		U95%	52.468	51.216	1.252
			単位 ($CS_2 mgl^{-1}$)		

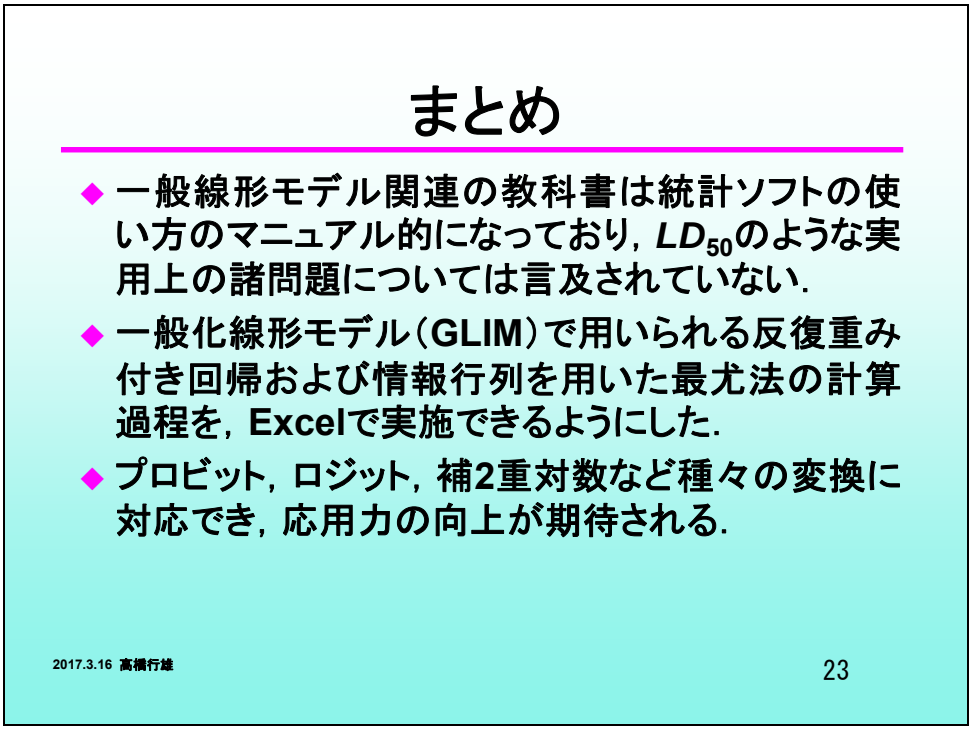
10. まとめ

プロビット法による 50%致死量 LD_{50} の推定は，Finney（1971）により示され，現在に至っているものの，各種の平方和を用いた反復重み付き回帰の計算方法である．さらに，LD_{50} の正確な 95%信頼区間の推定は，Filler の式による計算式が示されている．そのために，プロビット法をロジット法に変えた場合，あるいは，最小極値分布を用いた LD_{10} の推定などについて定式化するのが困難であった．

一般化線形モデル GLIM により，プロビット法，ロジット法あるいは最小極値分布について統一的な扱いが可能となり，Excel の行列関数を用いて反復重み付き回帰および情報行列を用いた反復計算を 1 枚の Excel シート上に展開することができた．

同一シート上で推定された回帰パラメータと分散共分散行列を用いて LD_{50} および LD_{10} の近似 95%信頼区間および正確な 95%信頼区間の計算結果を示した．なお，この計算方法は，プロビット法でもロジット法でもすべて共通であり，汎用的に用いることができる．

スライド 23



Excel による情報行列を用いた最尤法は，GLIM の枠に留まらず非線形の問題，右側，左側および区間打ち切りがある場合の回帰分析などにも適用することが可能であり，統計ソフト

が対応していない問題について容易に計算することができる．

続高橋セミナーの第 4 回目は，「寿命試験データの統計解析」で，すでに公開している．寿命試験データの解析は，右側打ち切りを含む統計解析で，誤差分布をワイブル分布とした回帰分析が主体であり，本セミナーと同様に，Excel および統計ソフト JMP を用いた最尤法を主体にしている．目次に含まれているように，合成分散の一般式（デルタ法），JMP による対数尤度関数の偏微分，について詳細に示してあるので，参考にしてもらいたい．

目 次

1. はじめに ---------- 1
2. 寿命試験データの例示 ---------- 4
3. ワイブル分布のあてはめ ---------- 13
4. 最尤法による加速試験データの統計解析 ---------- 35
5. JMP の寿命の二変量を用いた加速試験の解析 ---------- 48
6. 寿命時間の対数変換による最小極値分布の活用 ---------- 61
7. NEWTON-RAPHSON 法による尤度の最大化 ---------- 77
8. 回帰分析の拡張 ---------- 85
9. 多因子による寿命試験の事例 ---------- 95
10. 合成分散の一般式(デルタ法) ---------- 102
11. プロファイル尤度を用いた 95%信頼区間 ---------- 109
12. JMP による対数尤度関数の偏微分 ---------- 121

文献 ---------- 126

索引 ---------- 127

Excel, JMP ファイル 一覧 ---------- 132

最尤法は，統計モデルを使った統計解析の必須の方法論であり，基礎の基礎から丁寧な解説を行い，様々な応用事例も含めた入門書を作成する努力を続けている．本セミナーは，その中の一部の話題に関連して，実用上の諸問題を含めてまとめた．

本セミナーは，続高橋セミナー以前の高橋セミナー，「第 10 回 生物検定法の故郷、2 値データの用量反応に対する逆推定，平行性検定」，「第 18 回 LD50 を求めるための最尤法入門」，「第 26 回 低用量から発現する薬効・毒性の推定，極値分布およびゴンペルツ曲線の応用」，などを参考にして，新たに習得した「Excel による逐次計算法」を用いて書き直した．

参 考 文 献

Collett D.（2003）, Modelling Binary Data 2nd ed., 106-11,Chapman & Hall/CRC.

Dobson A.J.著（2002）田中豊, 森川敏彦, 中山竹春ら訳（2008）, 一般化線形モデル入門, 73-6, 139-48, 原著第 2 版, 共立出版.

Dobson A.J.,Barnett A.G.(2008）,An Introduction to Genaralozed Linear Models 3rd ed.,124-31,CRC Press.

Finney D.J (1971), Probit Analysis 3rd ed.,50-80, Cambridge University Press.

佐久間昭（1977）, 薬効評価－計画と解析 I, 330-9, 東大出版会.

佐久間昭著, 五所正明ら編（2017）, 新版薬効評価, 286-90, 東大出版会.

吉村功編著（1987）, 毒性・薬効データの統計解析, 226-32, サイエンティスト社.

竹内啓（1979）, 数理統計学, 312-5, 東洋経済新報社.

高橋行雄（2002）, 生物検定法の故郷, 2 値データの用量反応に対する逆推定, 平行性検定, http://www.yukms.com/biostat/takahasi/rec/010.htm

高橋行雄（2004）, LD50 を求めるための最尤法入門, http://www.yukms.com/biostat/takahasi/rec/018.htm

高橋行雄（2007）, 低用量から発現する薬効・毒性の推定, 極値分布およびゴンペルツ曲線の応用 http://www.yukms.com/biostat/takahasi/rec/026.htm

高橋行雄（2015）, 寿命試験データの統計解析, http://www.yukms.com/biostat/takahasi2/rec/004.htm

索 引

Excel	Binimdist	2項分布	40
Excel	Binimdist		56
Excel	Binimdist		65
Excel	Exp	指数	24
Excel	Intercept	切片	9
Excel	Ln	自然対数	24
Excel	Minverse	逆行列	11
Excel	Mmult	行列の積	9
Excel	Normdist	正規分布	10
Excel	Normdist		65
Excel	Norminv	逆正規	7
Excel	Slope	傾き	9
Excel	Transpose	行列の転置	11
Excel	Weibull	ワイブル分布	30
Excel	ソルバー		54
Excel	ソルバー		65
書籍	Collett(2003)		51
書籍	Dobson, Barntt(2008)		56
書籍	Finney(1971)		3
書籍	Finney(1971)		84
書籍	吉村編著(1987)		3
書籍	佐久間(1977)		3
書籍	佐久間(2017)		3
書籍	竹内(1979)		49
書籍	田中ら訳(2008)		2
書籍	田中ら訳(2008)		4
書籍	田中ら訳(2008)		37
書籍	田中ら訳(2008)		56
	1階の偏微分式	スコアベクトル	39
	2階の偏微分行列	ヘッセ行列	39
	2項分布	出現率	38
	2次形式	デルタ法	15
	2次形式	分散	46
	2次形式	分散共分散行列	17
	2次式の解	フィラーの式	53
	2次式の解の公式		50
	50 パーセント致死量		14
	50%出現率	ロジット変換	26
	95%信頼区間		45
	95%信頼区間	LD50	14
	95%信頼区間	LD50	52
	95%信頼区間	SAS/PROBIT	17
	95%信頼区間	逆推定	49
	95%信頼区間	シグモイド曲線	16
	95%信頼区間	シグモイド曲線	19
	95%信頼区間	正確	49
	95%信頼区間	比	51
	95%信頼区間	プロビット変換値	18
	95%信頼区間	下限・上限	49
	95%信頼区間	近似	36

95%信頼区間	分散共分散行列	61
Excel	逆推定	77
Excel	逐次計算法	85
Fiellerの式	フィラーの式	51
Fillerの式		84
GLIM	Genralized Linear Model	1
GLIM	一般化線形モデル	1
GLM	Genral Linear Model	1
JMP	カブト虫	73
JMP	プロビット法	78
JMP	モデルのあてはめ	73
JMP	モデル全体の検定	75
JMP	リンク関数	74
JMP	ロジット法	73
JMP	一般化線形モデル	73
JMP	関数プロット	76
JMP	逆推定	77
JMP	逆推定	83
JMP	共分散	78
JMP	重ね合わせプロット	76
JMP	重ね合わせプロット	82
JMP	二変量の関係	73
JMP	分布(誤差)	74
JMP	平均の信頼区間	81
JMP	補対数・対数	80
JMP	補対数・対数	74
JMP	予測値	81
JMP	列の保存	76
LD10	10パーセント用量	47
LD50	95%信頼区間	14
LD50	95%信頼区間	52
LD50	デルタ法	14
LD50	プロビット法	5
LD50	補2重対数法	36
OLS	Ordinary least Squares	1
probit	逆正規分布	7
PROBIT プロシジャ		3
S 字曲線	シグモイド曲線	6
SAS/PROBIT	95%信頼区間	17
SAS/PROBIT		3
scaleパラメータ	尺度パラメータ	29
SEV	最小極値分布	30
shapeパラメータ	形状パラメータ	29
X	デザイン行列	46
上に凸		32
カブトムシ	死亡数	38
カブト虫	JMP	73
カブト虫	ソルバー	56
逆推定	95%信頼区間	49
逆推定	JMP	3

最尤法		ソルバー	54
最尤法		情報行列	66
シグモイド曲線		3種	69
シグモイド曲線		3種	71
シグモイド曲線		95%信頼区間	16
シグモイド曲線		95%信頼区間	19
シグモイド曲線		95%信頼区間	45
シグモイド曲線		95%信頼区間	62
シグモイド曲線		パラメータ推定	71
シグモイド曲線		プロビット法	64
シグモイド曲線		形状の比較	71
シグモイド曲線		最小極値分布	56
シグモイド曲線		用量反応曲線	6
シグモイド曲線		累積正規分布	6
出現率		2項分布	38
スコアベクトル		1階の偏微分式	39
スコア関数		補2重対数法	59
線形化		プロビット変換	7
ソルバー		カブト虫	56
ソルバー		最尤法	54
ソルバー		位置パラメータ	57
ソルバー		形状パラメータ	57
デザイン行列		X	46
デルタ法		2 次形式	15
デルタ法		LD50	14
デルタ法		合成分散の一般式	85
デルタ法		偏微分	14
ニュートン・ラフソン法		プロビット法	67
ニュートン・ラフソン法		ロジット法	39
ニュートン・ラフソン法		情報行列	40
ニュートン・ラフソン法		補2重対数法	54
パラメータ		偏微分	48
比		95%信頼区間	51
比		二次式	52
フィラーの式		2次式の解	53
フィラーの式		95%信頼区間	51
フィラーの式		Fiellerの式	51
プロビット		リンク関数	78
プロビット法		補2重対数法	69
プロビット変換		線形化	8
プロビット変換		逆	18
プロビット変換		微分	21
プロビット変換		累積正規分布	7
プロビット変換値		95%信頼区間	18
プロビット法		JMP	78
プロビット法		LD50	5
プロビット法		シグモイド曲線	64
プロビット法		リンク関数	4
プロビット法		一般化線形モデル	79
プロビット法		重み付き回帰	4
プロビット法		情報行列	64
ヘッセ行列		2階の偏微分行列	39
ヘッセ行列		ロジット法	46
ヘッセ行列		情報行列	39
ベンチマークドーズ			3

右側打ち切り		寿命試験データ	85
モデルのあてはめ		JMP	73
モデル全体の検定		JMP	75
リンク関数		JMP	74
リンク関数		プロビット法	78
リンク関数		プロビット法	4
リンク関数		ロジット法	25
リンク関数		補2重対数法	34
ロジスティック曲線		出現率	44
ロジスティック分布		ロジット法	23
ロジット変換		50%出現率	26
ロジット変換		逆	24
ロジット変換		重み付き回帰	23
ロジット法		JMP	73
ロジット法		プロビット法	69
ロジット法		リンク関数	25
ロジット法		ロジスティック分布	23
ロジット法		回帰係数	27
ロジット法		情報行列	37
ロジット法		情報行列	40
ロジット法		推定値の比較	83
ロジット法		補2重対数法	69
ワイブル確率プロット			29
ワイブル分布		確率密度関数	30
ワイブル分布		誤差分布	85
ワイブル分布		最小極値分布	29
ワイブル分布		寿命時間解析	29
ワイブル分布		対数	29
ワイブル分布		累積分布関数	30
位置パラメータ		ソルバー	57
位置パラメータ		正規分布	64
位置パラメータ		正規分布	70
一般化線形モデル		GLIM	1
一般化線形モデル		GLIM	4
一般化線形モデル		JMP	73
一般化線形モデル		プロビット法	79
一般線形モデル		GLM	1
一般線形モデル		GLM	5
回帰係数		ロジット法	27
確率（密度）の積		最尤法	66
確率（密度）の積		尤度	66
環境ホルモン			3
関数プロット		JMP	76
逆推定		Excel	77
逆推定		JMP	77
逆推定		JMP	83
逆推定		出現率が10％	79
共分散		JMP	78
形状パラメータ		shapeパラメータ	29
形状パラメータ		ソルバー	57
形状パラメータ		正規分布	64
形状パラメータ		正規分布	70
経験プロビット			7
誤差分布		ワイブル分布	85
高橋セミナー		LD50・最尤法入門	85

高橋セミナー		寿命試験データ	85
高橋セミナー		生物検定法の故郷	85
高橋セミナー		低用量から発現	85
合成分散			20
合成分散の一般式		デルタ法	85
左に裾を引く分布		最小極値分布	29
最小2乗法		普通の	1
最小2乗法(普通の)		OLS	1
最小極値分布		シグモイド曲線	56
最小極値分布		ワイブル分布	29
最小極値分布		確率密度関数	31
最小極値分布		左に裾を引く分布	29
最小極値分布		出現率	62
最小極値分布		対数ワイブル分布	29
最小極値分布		微分	31
最小極値分布		補2重対数法	56
最小極値分布		累積分布関数	31
最尤法		打ち切り	84
最尤法		非線形の問題	84
最尤法とは		確率(密度)の積	66
最尤法とは		尤度を最大化	66
残差プロット		U 字型	77
尺度パラメータ		scaleパラメータ	29
寿命試験データ		右側打ち切り	85
寿命時間解析		ワイブル分布	29
重ね合わせプロット		JMP	76
重ね合わせプロット		JMP	82
重み		ベクトルw	11
重み		行列W	11
重み		分散の逆数	21
重み付き回帰		係数	12
重み付き回帰		プロビット法	4
重み付き回帰		ロジット変換	23
出現率		ロジスティック曲線	44
情報行列		最尤法	66
情報行列		プロビット法	64
情報行列		ヘッセ行列	39
情報行列		ロジット法	37
情報行列		ロジット法	40
情報行列		分散共分散行	45
情報行列		補2重対数法	58
推定値の比較		ロジット法	83
推定値の比較		補2重対数法	83
正規分布		位置パラメータ	64
正規分布		位置パラメータ	70
正規分布		形状パラメータ	64
正規分布		形状パラメータ	70
正規分布		微分	20
正規分布のパーセント点		微分	20
積和行列			11
積和行列			25
積和行列			35
打ち切り		最尤法	84
対数尤度		定数項なし	56
対数尤度		偏微分	39
対数尤度			38
逐次計算法		Excel	85
等分散でない			8
二変量の関係		JMP	73
二硫化炭酸ガス			38
反復重み付き回帰		分散共分散行列	45
反復重み付き回帰			8
反復重み付き回帰			13
比		正確な信頼区間	53
微分		プロビット変換	21
微分		最小極値分布	31
微分		正規分布	20
微分		正規分布のパーセント点	20
分散		2次形式	46
分散		プロビット変換	20
分散		プロビット変換	21
分散		合成	20
分散		出現率	20
分散		出現率	21
分散の逆数		重み	8
分散の逆数		重み	22
分散共分散行		情報行列	45
分散共分散行列		2 次形式	17
分散共分散行列		95%信頼区間	61
分散共分散行列		逆行列	39
分散共分散行列		反復重み付き回帰	45
分散共分散行列			15
分布(誤差)		JMP	74
平均の信頼区間		JMP	81
偏微分		パラメータ	48
偏微分		対数尤度	39
補2重対数変換		逆	34
補2重対数変換		逆	61
補2重対数変換		最小極値分布	29
補2重対数変換		直線	62
補2重対数法		JMP	80
補2重対数法		LD50	36
補2重対数法		スコア関数	59
補2重対数法		プロビット法	69
補2重対数法		プロビット法	69
補2重対数法		リンク関数	34
補2重対数法		ロジット法	69
補2重対数法		最小極値分布	56
補2重対数法		情報行列	58
補2重対数法		推定値の比較	83
補2重対数法			32
補対数・対数		JMP	80
補対数・対数		JMP	74
尤度とは		確率(密度)の積	66
予測値		JMP	81
用量反応曲線		シグモイド曲線	6
累積正規分布		プロビット変換	7
列の保存		JMP	76

Excel, JMP ファイル一覧

GLIM_第2章_01_反復重み付き_吉村プロビット	72 KB
GLIM_第2章_02_シグモイド95%信頼区間	35 KB
GLIM_第2章_03_表2_4_2次形式	28 KB
GLIM_第2章_04_図2_7_プロビット変換の分散の変化	26 KB
GLIM_第2章_05_図2_8_正規分布パーセント点の微分	24 KB
GLIM_第3章_06_ロジット	26 KB
GLIM_第4章_07_図41_図4_2_ワイブル_最小極値_分布	61 KB
GLIM_第4章_08_表4_1_表4_2_補2重対数	26 KB
GLIM_第5章_09_表5.1_カブト虫	150 KB
GLIM_第6章_10_補2重対数_極値	48 KB
GLIM_第7章_10_プロビット	28 KB
GLIM_第8章_11_3種のシグモイド曲線	30 KB
GLIM_第9章_12_dobson_カブト虫 JMP	9 KB

非売品，無断複製を禁ずる

第 6 回 続高橋セミナー

一般化線形モデルを Excel で極め活用する

－プロビット法・ロジット法・補2重対数法－

BioStat 研究所（株）

〒105-0014 東京都 港区 芝 1-12－3