

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 10 回 高橋セミナー

高橋 行雄

表紙裏

改訂の記録

2002年1月13日，新規作成

目 次

1.	はじめに.....	1
2.	プロビット変換, ロジット変換.....	3
2.1.	プロビット変換.....	3
2.2.	ロジット変換.....	6
2.3.	2 値反応に対する解析法.....	9
2.4.	回帰曲線の 95%信頼区間.....	10
2.5.	逆推定.....	11
2.6.	推定された有効量についての標準誤差の近似.....	13
2.7.	フィラーの定理を用いた有効用量の信頼区間.....	15
2.8.	ED50 の信頼区間の計算事例.....	18
3.	最尤法による計算の例.....	23
3.1.	吉村の例、JMP の行列計算言語による.....	23
3.2.	SAS の POBIT プロシジャによるプロビット法による計算.....	25
3.3.	JMP の行列言語によるロジスティック回帰.....	27

図 表 目 次

表 2.1	LD ₅₀ を求める数値例.....	4
表 2.2	プロビット法による解析結果.....	4
表 2.3	ロジスティック回帰モデルの結果.....	8
表 2.4	プロビット法とロジット法の推定値 p の差.....	8
表 2.5	ロジット法による逆推定.....	9
表 2.6	ロジット法による 95%信頼区間.....	11
図 2.1	プロビット曲線の当てはめ.....	5
図 2.2	プロビット曲線とロジット曲線の比較.....	7
図 2.3	ロジット曲線と 95%信頼区間の表示.....	11
図 2.4	ED ₅₀ とその信頼区間の近似計算、Excel による計算.....	19
図 2.5	ED ₅₀ とその信頼区間のフィラーの公式、Excel による計算.....	20
図 2.6	ED ₅₀ とその信頼区間の近似計算.....	20
図 2.7	ロジット法のフィラーの公式による 95%信頼区間.....	21
図 2.8	2 回反復のプロビット法のフィラーの公式による 95%信頼区間.....	22

第10回－高橋セミナーのご案内

2002年4月20日, 医薬安全研の午前中にセミナーを行います.

「第9回のテーマ: 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」をすべて終了することは不可能でした. 第10回目は, 第9回目で芳賀先生から指摘を受けた改定した「ロジット曲線とプロビット曲線の比較」も含め, 2値データの解析に用いられている最尤法, ニュートン・ラフソン法の復習を行い, 「5. 相対力値」以後をテーマとします.

1. プロビット変換, ロジット変換
2. プロビット法とロジスティック回帰によるシグモイド曲線の推定
3. 逆推定によるLD50の推定
4. LD50の95%信頼区間
5. 相対力値
6. 自然反応
7. 非線形ロジット回帰モデル
8. complementary log-log
9. ランダム効果を含む2つの用量反応データの比較

このテーマは, 題1回目の課題でも有りました. この時は, JMPで取り扱うことがどこまで可能なのか, 本当にできるのか, に集中した盛りだくさんの内容でした. この内容は, <http://www.yukms.com/biostat/>からダウンロードできます.

今回は, 実験研究者が引用できる文献を紹介しつつ, JMPでの解析の演習を行います. これにより, 実験研究者, プログラマー, 統計家が実務でJMPあるいは他の統計パッケージの見直しができるようになります. 参加希望の方はJMPがインストールされたノートPCを準備してください. 文献は, D. Collett (1991) Modeling Binary Data, Chapman and Hall/CRC の第4章 生物検定法といくつかの適用です. そして, ピンク本 5.4節 LD50の推定 です. 時間があれば, Overdispersionの問題にもふれたい. この問題は, 次のような実験データの解析に現れる.

6.1.1 Variation between the response probabilities

実験ユニットでバッチの数が, 同じ条件で得られたとき, 応答確率がバッチとバッチの間で, それにもかかわらず異なるかもしれない.

胎児の例, 催奇形性試験として知られている.

母体の遺伝的影響により同じ実験条件でも奇形出現率は異なる.

奇形出現率の分散は, それが定数の場合よりも大きい.

この変動を説明する変数を実験者は知ることができない。
実験条件が十分にコントロールされていない場にも起きる。

Example 6.2 Germination of Orobanche (岩波生物学辞典: シソ花類; ハマウツボ)

1:ageyp75				1:ageyp75			
1:Bean		2:Cucumber		1:Bean		2:Cucumber	
y	n	y	n	y	n	y	n
10	39	5	6	8	16	3	12
23	62	53	74	10	30	22	41
23	81	55	72	8	28	15	30
26	51	32	51	23	45	32	51
17	39	46	79	0	4	3	7
		10	13				

インゲン豆ときゅうりの引き抜いた根の中での2つのハマウツボ種類の発芽の割合を
考える。

1. はじめに

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，平行性検定について基礎にもどって講義と実習を行う。

このテーマは，第 1 回目セミナーの課題であった。この時は，JMP で生物検定法の代表的な課題に対してどこまで対応可能なのか，本当にできるのか，などについての解析結果を主体にした (<http://www.yukms.com/biostat/>からダウンロードできる)。今回は，実験研究者が，引用できる文献を紹介しつつ，JMP での解析の演習を行う。これにより，実験研究者，プログラマー，統計家が実務で JMP あるいは他の統計パッケージを利用した生物検定法の見直しができるように意図した。今回の講義と実習は，D. Collett (1991), *Modeling Binary Data* をベースにする。この本は，統計モデルを用いた豊富な薬理試験，毒性試験などの生物系の実験データの解析事例を用いつつ，2 値データの解析の理論的な背景が丁寧に解説されている。生物検定法は 4 章 Bioassay and some applications で取り上げられている。参加者になじみのある吉村編著 (1987) 毒性・薬効データの統計解析，5.4 節 LD50 の推定法，の事例を取り上げる。ロジスティック回帰については，丹後 (1996) ロジスティック回帰 - SAS を利用した統計解析の実際 - ，がロジスティック回帰全般について丁寧に述べられている。この本にも，3.9 節で LD50、ED50 の推定が取り上げられている。

生物検定法 (Biological Assay) とは

生物を用いて未知の化合物の生物活性を
既知の化合物の生物活性に対して
相対的に比較するために体系化された
応用統計学の一つの分野である
代表的な生物検定法 : 50 パーセント致死量の推定

生物検定法は回帰分析の応用

直線の当てはまりの欠如 (Lack of Fit)
非平行性 (Lack of Parallelism)
逆推定とその信頼区間
用量反応関係を論ずるために欠かせない
他の応用統計の分野では軽視

SAS と JMP による生物検定法

SAS : proc PROBIT

逆推定を取り扱えるのは 1 群の場合

プロビット変換よりむしろロジット変換

ロジスティック回帰分析として定式化

SAS : proc LOGISTIC、

proc CATMOD

proc GENMOD

JMP : “Fit Y by X”、

“Fit Model”

逆推定、 Y_0 となる X は、

生物検定法で常用される逆推定、95%信頼区間

SAS では標準的には求められない

JMP では、Inverse Prediction の問題として対応

生物検定法のための

統計パッケージとして JMP が適している

今回は次の課題を、基礎概念も含めて再度、取り上げ、非線形混合モデルの応用の可能性についても触れてみたい。

1. プロビット変換, ロジット変換
2. プロビット法とロジスティック回帰によるシグモイド曲線の推定
3. 逆推定による LD50 の推定
4. LD50 の 95%信頼区間
5. 相対力値
6. 自然反応
7. 非線形ロジット回帰モデル
8. 2重対数変換
9. ランダム効果を含む 2 つの用量反応データの比較

2. プロビット変換, ロジット変換

2.1. プロビット変換

毒性・薬効データの統計解析の 5.4 節 LD₅₀ の推定法生物検定法では、プロビット法による LD₅₀ の推定とその 95%信頼区間の計算法が示されている。

「LD₅₀ を毒性の目安とするときには、暗黙のうちに、用量と反応の関係が単調増加であることを前提にしている。その単調増加曲線としては、ほとんどの場合プロビット曲線を想定している。“プロビット曲線” probit curve とは、図 2.1 に示すように反応率が、用量の対数に対して、正規分布の累積確率、すなわち正規分布関数の関係をもつ曲線のことである。見方を変えれば、用量を横軸にとり、反応率を縦軸に取って、データを対数正規確率紙にプロットしたときに、関係が正の勾配を持った直線になることである。式で書けば用量 d と反応率 p (原文では y) の関係が

$$p = \int_{-\infty}^{\log_{10} d} \frac{\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}{\sqrt{2\pi\sigma^2}} dx \quad (2.1)$$

となるものである。このとき、 $\mu = \log_{10} \text{LD}_{50}$ である。“probit” は、probability unit の省略である。

以下に、プロビット曲線を前提にして、実験データから LD₅₀ を推定する手法を紹介するが、計算では次の式で定義するプロビット関数を使う。

$$\text{probit}(p) = \left\{ p = \int_{-\infty}^{y-5} \frac{\exp\left\{-\frac{x^2}{2}\right\}}{\sqrt{2\pi}} dx \text{ となる } y \text{ の値} \right\} \quad (2.2)$$

「 y は基準正規分布の累積確率が p となる場所の横軸 (正規偏差) に 5 を加えたものである。」

数値例として表 2.1 が示され、本的な LD₅₀ の求め方として重み付き最小 2 乗法の繰返し計算による最尤法が示されている。計算手順は煩雑なので、JMP の行列計算言語によるプログラミング例と結果を付録に示した。ここでは、得られた結果の解釈を通じてプロビット法の考え方を示す。式 (2.2) のプロビットは、標準正規分布の逆関数 $\Phi^{-1}(p)$ に 5 を加える。JMP では Normal Quartile 関数を用いることにより直接計算できる。死亡率が 0% あるいは 100% は、マイナス無限大とプラスの無限大となる。

表 2.1 LD₅₀を求める数値例

群 i	投与量 mg/kg (公比 1.35)	群の大きさ	死亡数	死亡率 p_i	プロビット probit(p_i)
1	110	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	∞

投与量の常用対数を取り、変換されたプロビットに、重み付き最小2乗法の繰り返しにより尤度を最大化するような次の回帰直線をあてはめる。3.1 節に示したJMPの行列計算プログラムにより、次の回帰係数が得られる。

$$\eta_i = \text{probit}(\hat{p}_i) = -10.0414 + 6.6030 \log_{10}(\text{dose}_i)$$

表 2.2 に、この式によって計算されたプロビットの推定値を示す。死亡率の推定値は、標準正規分布 $\Phi(\eta_i - 5)$ によって計算される。

表 2.2 プロビット法による解析結果

群	$\log_{10}(\text{dose}_i)$	$\text{probit}(\hat{p}_i)$	\hat{p}_i
1	2.0043	3.1931	0.035
2	2.1335	4.0463	0.170
3	2.2625	4.8975	0.459
4	2.3927	5.7575	0.776
5	2.5224	6.6142	0.947
6	2.6532	7.4777	0.993

死亡率 p が 0.50 となる、投与量は式 (2.1) の μ と σ は、

$$\eta_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

が、

$$\eta_i = \frac{x - (-(\hat{\beta}_0 - 5) / \hat{\beta}_1)}{1 / \hat{\beta}_1}$$

と書き換えられるので、

$$\hat{\mu} = \log_{10} \text{LD}_{50} = -((\hat{\beta}_0 - 5) / \hat{\beta}_1) = -(-10.0414 - 5) / 6.6030 = 2.278$$

となり、

$$LD_{50} = 10^{2.278} = 190.0 \text{ mg/kg}$$

が得られる。 σ は、

$$\hat{\sigma} = 1/\hat{\beta}_1 = 1/6.6030 = 0.1514$$

と計算される。これらの $\mu = 2.278$ 、 $\sigma = 0.1514$ となる正規分布、および2.4節で述べる方法でもとめた回帰直線の 95%信頼区間を逆プロビット変換して、当てはめたのが図 2.1 のシグモイド曲線である。

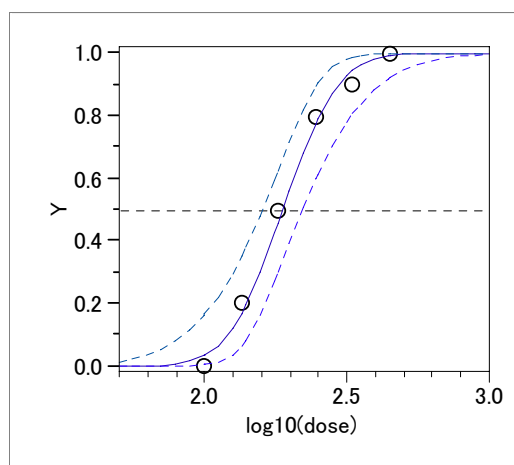


図 2.1 プロビット曲線の当てはめ

LD_{50} の 95%信頼区間は、図 2.1 のY軸の 0.5 を通る水平線上の 95%信頼区間を横切るときのX軸の目盛りを読むことにより得られる。解析的には、フィラーの式により、常用対数で (2.206, 2.347) となり、投与用量に変換して (161, 223) が得られる。フィラーの式は、2.7 節で詳しく述べるので、ここでは結果のみを示した。

ここで示した計算結果は、重みつき最小 2 乗法を 2 回繰り返した場合であり、数値計算的には収束していない。

プロビット法で十分に収束した結果を求めるためにSASのproc PROBITを用いた結果 (3.2 節に計算プログラムと結果) を示す。これは、JMPでは、プロビット法の計算はサポートされていないためである。なお、SASでは、proc LOGISTIC、proc GENMODでもプロビット法での計算は行えるが、 LD_{50} の 95%信頼区間の計算がサポートされていない。

SAS の proc PROBIT の反復計算は、初期の回帰係数 $\beta_0 = 0$ 、 $\beta_1 = 0$ からスタートして、6回のニュートン・ラフソン法による反復の結果 $\beta_0 = -15.0703$ 、 $\beta_1 = 6.6152$ が得られた。切片が2回の重み付き最小2乗法で得られた-10.0414に比べてかなり小さくなっているが、これは、プロビット変換時に定数5を加えたためであり、-5を加えると-15.0414となりほぼ同じである。LD50の95%信頼区間は、フィラーの式により、常用対数で(2.204, 2.348)と計算され、投与用量に変換して(160, 223)と、JMPの重み付最小2乗法の2回の反復と計算誤差範囲内で一致している。

2.2. ロジット変換

シグモイド曲線となる2値反応の出現率に、正規分布をあてはめることにより、さまざまな生物検定法における応用がされてきた。シグモイド曲線を得るために正規分布の数値計算は煩雑であることから、数値計算が簡単なロジスティック分布をシグモイド曲線に用いる方法が多用されるようになってきた。

$$f(x) = \frac{\exp\left(\frac{x-\mu}{\tau}\right)}{\tau \left\{1 + \exp\left(\frac{x-\mu}{\tau}\right)\right\}^2}, \quad -\infty < x < \infty \quad (2.3)$$

ここで、 $-\infty < \mu < \infty$ 、 $\tau < 0$ であり、平均と分散は、それぞれ μ と $\pi^2\tau^2/3$ である。確率密度関数 $f(x)$ は、正規分布に比べ簡潔とはいえないが、分布関数は、

$$p_i = \int_{-\infty}^{\log_{10} d_i} \frac{\exp\left(\frac{x-\mu}{\tau}\right)}{\tau \left\{1 + \exp\left(\frac{x-\mu}{\tau}\right)\right\}^2} dx = \frac{\exp\left(\frac{\log_{10} d_i - \mu}{\tau}\right)}{1 + \exp\left(\frac{\log_{10} d_i - \mu}{\tau}\right)} \quad (2.4)$$

と簡潔な指数関数となる。ここで、 $\beta_0 = -\mu/\tau$ 、 $\beta_1 = 1/\tau$ とおけば、ロジスティック分布関数は、

$$p_i = \frac{\exp(\beta_0 + \beta_1 \log_{10} d_i)}{1 + \exp(\beta_0 + \beta_1 \log_{10} d_i)} \quad (2.5)$$

となる。簡単な式の変形により、ロジット、

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 \log_{10} d_i$$

が得られる。式 (2.5) が正規分布を用いた場合とほぼ同様のシグモイド曲線を与える。このシグモイド曲線を用いて 50%致死量 LD_{50} を推定するのがロジスティック回帰モデルである。

ロジスティック法で推定された回帰係数は、プロビット法のそれとは異なる。プロビット法では、 $1/\beta_1$ が標準偏差となるが、ロジスティック法での標準偏差は $\sqrt{\pi^2\tau^2/3}$ であり、プロビット法で推定された推定された標準偏差を $\hat{\sigma}$ とすれば、

$$\hat{\tau} = \sqrt{\frac{3\hat{\sigma}^2}{\pi^2}} = \sqrt{\frac{3 \times 0.1514^2}{3.14^2}} = 0.0831$$

となる。

JMPによる LD_{50} の推定は、プロビット法ではなく“ロジット法”による推定となっている。プロビット法によって得られた $LD_{50} = 2.278$ と $\hat{\sigma} = 0.1514$ から推定されるプロビット曲線、 $LD_{50} = 2.278$ と $\hat{\tau} = 0.0831$ から推定されるロジスティック曲線を比較した結果を図 2.2 に示す。違いは、ロジスティック回帰が裾広がりとなるが、図にしてみるとごくわずかである。

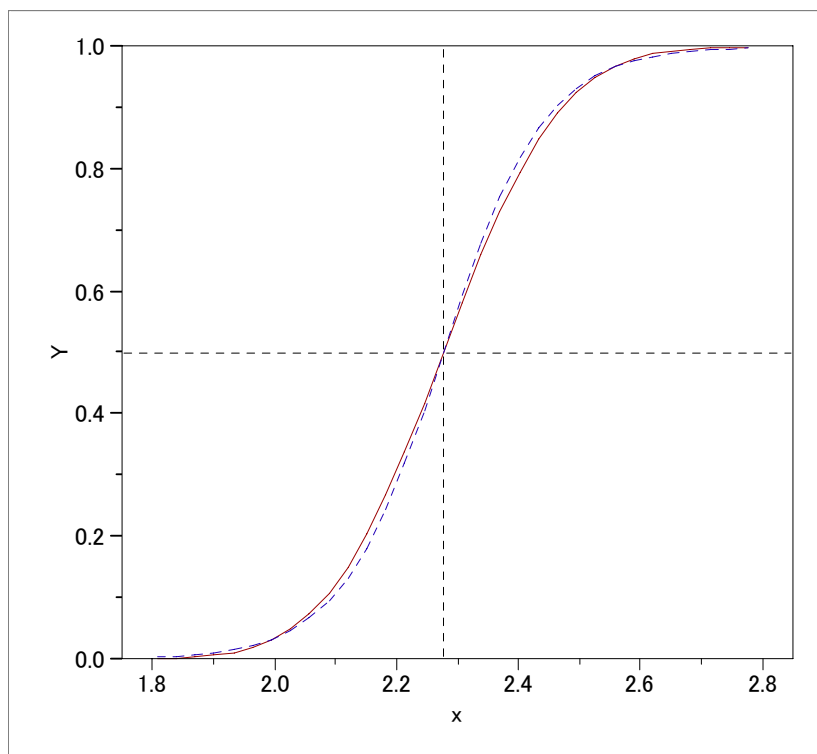


図 2.2 プロビット曲線とロジット曲線の比較
— プロビット、--- ロジット

JMPでロジスティック回帰を行った結果を表 2.3 に示す。回帰係数 $\beta_1 = 11.5229$ と勾配が急になり、裾広がり解消している。

表 2.3 ロジスティック回帰モデルの結果

パラメータ推定値				
項	推定値	標準誤差	カイ2乗	p値(Prob>ChiSq)
切片	-26.211489	6.350431	17.04	<.0001
log10(dose)	11.5229099	2.7804107	17.18	<.0001

50%致死量は、求められた回帰係数から、

$$\hat{\mu} = \log_{10} LD_{50} = -\hat{\beta}_0 / \hat{\beta}_1 = -(-26.2115) / 11.5229 = 2.275$$

となり、

$$LD_{50} = 10^{2.275} = 188.2 \text{ mg/kg}$$

と推定できる。プロビット法の 190.0mg/kg と比べて約 1%の違いである。分散は、

$$\hat{\sigma}^2 = \sqrt{\frac{\pi^2 \tau^2}{3}} = \sqrt{\frac{3.14^2 \times (1/11.52)^2}{3}} = 0.157$$

とプロビット法の 0.151 とほぼ同じである。

プロビット法とロジット法による推定値 p の差を表 2.4 に示す。その結果でもロジット法の結果は、裾広がりとなっているが、その差は 0.5%程度の違いとわずかである。

表 2.4 プロビット法とロジット法の推定値 p の差

$\log_{10}(dose_i)$	Probit	logit	logit-probit
1.809	0.001	0.005	0.004
1.903	0.007	0.014	0.007
2.028	0.049	0.055	0.006
2.153	0.205	0.197	-0.007
2.278	0.500	0.509	0.009
2.403	0.795	0.814	0.019
2.528	0.951	0.949	-0.002
2.653	0.993	0.987	-0.006
2.747	0.999	0.996	-0.003

JMPでは、逆推定の機能により、任意の死亡率について計算できるので、90%、50%、

および 10%のそれぞれについて推定した結果を 表 2.5 に示す。LD₅₀ の 95%信頼区間は、フィラーの式により、常用対数で (2.196, 2.351) と推定される。

表 2.5 ロジット法による逆推定

確率	予測値 log10(dose)	下限	上限	1-Alpha
0.90000000	2.46541141	2.38150573	2.66236231	0.9500
0.50000000	2.27472827	2.19554355	2.35091729	
0.10000000	2.08404514	1.87888963	2.17016401	

JMPのロジスティック回帰モデルの計算手順は、ニュートン・ラフソン法による反復計算による尤度が最大となるような解が求められている。どのような計算手順なのかを示すために 3.3節 JMPの行列計算プログラムと実行結を示す。反復計算は、初期の回帰係数 $\beta_0 = 0$ 、 $\beta_1 = 0$ からスタートして、6回の反復の結果 $\beta_0 = -26.212$ 、 $\beta_1 = 11.523$ が得られ、結果は一致する。

2.3. 2 値反応に対する解析法

これまでに多くの“法”を使ってきたので、これらの関連を整理してみよう。

2 値データの変換法の反復

プロビット変換 (5 を加える場合も有る)

ロジット変換

シグモイド曲線をあてはめて 50%推定値を求める方法の総称

プロビット変換による場合：プロビット法

ロジット変換による場合：ロジット法、または、ロジスティック回帰モデル

最尤解を求める反復計算の方法、どちらの方法でも解は一致する

重みつき最小 2 乗法

ニュートン・ラフソン法

方法の組合せ

プロビット変換・重み付き最小 2 乗法の反復

吉村らの本、SAS/LOGISTIC

プロビット変換・ニュートン・ラフソン法

SAS/PROBIT (逆推定)、SAS/LOGISTIC、SAS/GENMOD

ロジット変換・重み付き最小 2 乗法の反復

SAS/LOGISTIC

ロジット変換・ニュートン・ラフソン法

JMP (逆推定)、SAS/LOGISTIC、SAS/GENMOD

2.4. 回帰曲線の 95%信頼区間

プロビット曲線またはロジスティック曲線の 95%信頼区間を求めよう。 $\eta_i = \mathbf{x}_i \hat{\boldsymbol{\beta}}$ の推定値の分散は、分散共分散行列を Σ としたときに

$$\sigma_x^2 = \mathbf{x}_i' \Sigma \mathbf{x}_i$$

となる。単回帰分析の場合は、

$$\eta_i = 1 \cdot \hat{\beta}_0 + \log_{10}(\text{dose}_i) \cdot \hat{\beta}_1$$

なので、 $\hat{\beta}_0$ の分散を v_{00} 、 $\hat{\beta}_1$ の分散を v_{11} 、 $\hat{\beta}_0$ と $\hat{\beta}_1$ の共分散を v_{01} とすると、

$$\sigma_x^2 = \mathbf{x}_i' \Sigma \mathbf{x}_i = v_{00} + 2 \cdot \log_{10}(\text{dose}_i) \cdot v_{01} + \log_{10}(\text{dose}_i)^2 v_{11}$$

で計算される。

JMP のロジスティック回帰で求められた回帰係数は $\hat{\beta}_0 = -26.211$ 、 $\hat{\beta}_1 = 11.52$ 、JMP の行列計算言語で計算した分散共分散行列は、

$$\Sigma = \begin{bmatrix} 40.328 & -17.623 \\ -17.623 & 7.731 \end{bmatrix}$$

である。 η_i の場合について計算してみる。

$$\eta_i = 1 \cdot \hat{\beta}_0 + \log_{10}(\text{dose}_i) \cdot \hat{\beta}_1 = -26.211 + \log_{10}(101) \times 11.52 = -3.1159$$

$$\hat{p}_1 = \frac{e^{-3.115}}{1 + e^{-3.1159}}$$

$$\begin{aligned} \sigma_x^2 &= \mathbf{x}_i' \Sigma \mathbf{x}_i = v_{00} + 2 \cdot \log_{10}(\text{dose}_i) \cdot v_{01} + \log_{10}(\text{dose}_i)^2 v_{11} \\ &= 40.328 + 2 \times \log_{10}(101) \times (-17.623) + \log_{10}(101)^2 \times 7.731 = 0.7391 \end{aligned}$$

$$\eta_i \pm 1.96 \sigma_x = -3.1159 \pm 1.96 \sqrt{0.7391} = (-4.8010, -1.4310)$$

$$\hat{p}_1(L95) = \frac{e^{-4.8010}}{1 + e^{-4.8010}} = 0.0082, \quad \hat{p}_1(U95) = \frac{e^{-1.4310}}{1 + e^{-1.4310}} = 0.1930$$

他の用量についても同様に計算した結果を表 2.6 に示す。

表 2.6 ロジット法による 95%信頼区間

i	dose	n	Y	p	log10(dose)	eta	p^	Var(eta)	s.e.(eta)	L95%	U95%
1	101	10	0	0.0	2.0043	-3.1159	0.0420	0.7391	0.8597	0.0082	0.1930
2	136	10	2	0.2	2.1335	-1.6269	0.1640	0.3182	0.5641	0.0611	0.3725
3	183	10	5	0.5	2.2625	-0.1415	0.4650	0.1554	0.3942	0.2861	0.6528
4	247	10	8	0.8	2.3927	1.3593	0.7960	0.2519	0.5019	0.5928	0.9124
5	333	10	9	0.9	2.5224	2.8544	0.9460	0.6088	0.7803	0.7900	0.9877
6	450	10	10	1.0	2.6532	4.3612	0.9870	1.2319	1.1099	0.8990	0.9986

投与量を連続的に変えて推定値とその 95%信頼区間計算した結果を図 2.3 に示す。図の右は拡大した図であり、反応が 0.5 となるような 95%の投与量が逆推定されている。

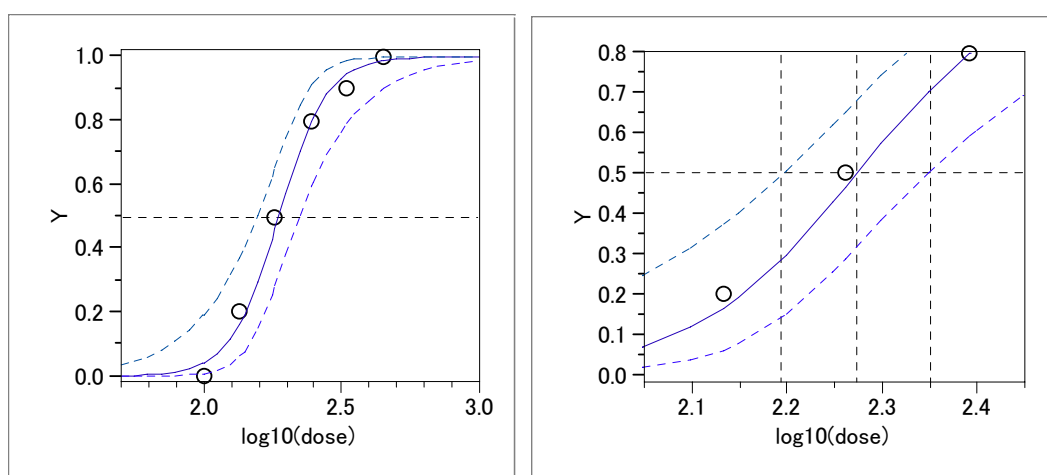


図 2.3 ロジット曲線と 95%信頼区間の表示

2.5. 逆推定

応答確率が、単一の説明変数 x の関数としてモデル化される場合、応答確率に対応する x の値を推定することにしばしば興味がある。例えば、生物検定法では、化学薬品に曝露された個体の 50%に反応すると期待される濃度に、しばしば興味をもたれる。この用量はメディアン用量と名付けられ、一般に $ED50$ 値と呼ばれる。試験の反応が死である場合、これはメディアン致死量、あるいは $LD50$ 値と言い換えられる。他の反応率を期待する濃度は、同様の方法出で表される。例えば、各個体の 90%に反応させる用量は $ED90$ 値などである。これらの量での要約は、分析しようとしている化学薬品の力価、および、その後の異なる物質の間の比較の基礎となる。

生物検定法により応答確率 p 、および用量 d 、の関連を記述するために、次に示す、

線形ロジスティクス・モデルを使用するとしよう。

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 d$$

$p = 0.5$ が $ED50$ 値である用量は、 $\text{logit}(0.5) = \log(1) = 0$ であるので、 $ED50$ 値は、式

$$\beta_0 + \beta_1 ED50 = 0$$

となり、その結果 $ED50 = -\beta_0 / \beta_1$ となる。線形ロジスティクス・モデルをあてはめた後に、未知のパラメータの推定値 $\hat{\beta}_0$ 、 $\hat{\beta}_1$ が得られ、 $ED50$ 値は、

$$ED50 = -\frac{\hat{\beta}_0}{\hat{\beta}_1}$$

によって推定される。 $ED90$ 値を得るために、 $p = 0.90$ を、線形ロジスティクス・モデル

$$\log\left(\frac{0.9}{0.1}\right) = \beta_0 + \beta_1 ED90$$

で用い、その結果、 $ED90$ 値は、 $(2.1972 - \beta_0) / \beta_1$ であるので、 $ED90 = (2.1972 - \hat{\beta}_0) / \hat{\beta}_1$ と推定される。

用量ではなく $\log(\text{dose})$ が、モデルの中で説明変数として使用される場合、これらの式は、修正を少し必要とする。モデル

$$\text{logit}(p) = \beta_0 + \beta_1 \log(d)$$

があてはめられる場合、 $ED50$ 値は、

$$\beta_0 + \beta_1 \log(ED50) = 0$$

となるので、 $ED50 = \exp(-\beta_0 / \beta_1)$ となり、 $ED50 = \exp(-\hat{\beta}_0 / \hat{\beta}_1)$ で推定される。同様に、 $ED90$ 値は、

$$ED90 = \exp\left\{\frac{2.1972 - \hat{\beta}_0}{\hat{\beta}_1}\right\}$$

によって推定される。他の対数の底が、モデルの中で使用される場合、この底は、この

式の e を代えなければならない。コンピューターによる解析では、出力によって、さらなる計算を正確に実行するために、対数の底 (10 または e) を知る必要がある。

$ED50$ と $ED90$ 値の推定は、プロビット回帰モデルのでも同様に得ることができる。例えば、 $\log(\text{dose})$ が説明変数として使用される場合、モデルは

$$\text{probit}(p) = \Phi^{-1}(p) = \beta_0 + \beta_1 \log(d)$$

となる。 $P = 0.5$ とした場合、 $\text{probit}(0.5) = 0$ であり、ロジスティックス・モデルと同様に、 $ED50$ 値は、 $ED50 = \exp(-\hat{\beta}_0/\hat{\beta}_1)$ によって推定される。ここで、 $\hat{\beta}_0$ 、 $\hat{\beta}_1$ はプロビット・モデルでのパラメータ推定値である。 $P = 0.9$ のとき、 $\text{probit}(0.9) = 1.2816$ であり、 $ED90$ 値は、 $\exp[(1.2816 - \hat{\beta}_0)/\hat{\beta}_1]$ によって推定される。

$ED50$ のような推定量を得たときに、推定精度を判断することができるように、標準誤差、あるいは信頼区間を与えることが一般に望まれる。パラメータ推定関数の標準誤差、対応する信頼区間を得るため用いられる式の展開は、重要であり単純ではないので、ここで詳しく示す。推定された $ED50$ 値の近似の標準誤差は、最初にセクション2.6 で誘導され、この標準誤差に基づいた真の $ED50$ 値の近似の信頼区間が与えられる。真の $ED50$ 値のより正確な区間は、セクション2.7節で記述されるアプローチ、フィラーの定理を使用して求められる。

2.6. 推定された有効量についての標準誤差の近似

$ED50$ が例えば 2 つのパラメータの関数 $g = (\hat{\beta}_0, \hat{\beta}_1)$ であるばあい、関数の分散の近似の標準的な解法、テイラー級数近似法、または、デルタ法により、推定値の標準誤差を得るために使用できる。この結果は、 $g = (\hat{\beta}_0, \hat{\beta}_1)$ の分散がほぼ式 (2.6) であることを示している。

$$\left(\frac{\partial g}{\partial \hat{\beta}_0} \right)^2 \text{Var}(\hat{\beta}_0) + \left(\frac{\partial g}{\partial \hat{\beta}_1} \right)^2 \text{Var}(\hat{\beta}_1) + 2 \left(\frac{\partial g}{\partial \hat{\beta}_0} \frac{\partial g}{\partial \hat{\beta}_1} \right) \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \quad (2.6)$$

モデルが説明変数として用量を含んでいる場合、 $g = (\hat{\beta}_0, \hat{\beta}_1)$ は $-\hat{\beta}_0/\hat{\beta}_1$ であることから得られる。 $\hat{\rho} = \hat{\beta}_0/\hat{\beta}_1$ 、 $v_{00} = \text{Var}(\hat{\beta}_0)$ 、 $v_{11} = \text{Var}(\hat{\beta}_1)$ および $v_{01} = \text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$ とおくと、式 (2.6) は

$$\text{Var}(E\hat{D}50) \approx \frac{v_{00} - 2\hat{\rho}v_{01} + 2\hat{\rho}^2 v_{11}}{\hat{\beta}_1^2}$$

となる。実際には、この式での分散および共分散は、正確には、わからないのであるが、 \hat{v}_{00} 、 \hat{v}_{11} および \hat{v}_{01} の近似値が、式の中で用いられるであろう。推定された $ED50$ 値と $\text{s.e.}(E\hat{D}50)$ の標準誤差は、その後、式 (2.7) によって与えられる。

$$\text{s.e.}(E\hat{D}50) \approx \left\{ \frac{v_{00} - 2\hat{\rho}v_{01} + 2\hat{\rho}^2 v_{11}}{\hat{\beta}_1^2} \right\}^{1/2} \quad (2.7)$$

このような方法による推定値の標準誤差は、対応するパラメータについての近似的な信頼区間を得ることである。例えば、 $E\hat{D}50 \pm 1.96 \text{s.e.}(E\hat{D}50)$ を使用して、 $ED50$ 値の 95% の信頼区間を得ることができる。この方法で計算された信頼区間の低い限界は、負になることが起きるかもしれない。ゼロと単にこれを取り替えることができるかもしれないが、この問題を回避する 1 つの方法はすべて $ED50$ 値の対数を使うことである。 $\log(E\hat{D}50)$ の標準誤差は、標準的な計算法により、

$$\text{s.e.}\{\log(E\hat{D}50)\} \approx (E\hat{D}50)^{-1} \text{s.e.}(E\hat{D}50)$$

を得る。

$\log(E\hat{D}50)$ 値の近似的な 95% 信頼区間は、 $\log(E\hat{D}50) \pm 1.96 \text{s.e.}\{\log(E\hat{D}50)\}$ であり、 $ED50$ 値の信頼区間は、これらの 2 つの範囲の指数をとることにより得られる。しかしながら、この方法は、信頼区間のより低い限界が負でないことを保証するが、一般的には勧められない。これは推定された $\log(ED50)$ 値は、対称的に $ED50$ 値自体が分布しにくいからであり、従って、信頼区間を構築するために用いられる正規分布の仮定は合理的とは思われない。

ロジスティックスのモデルの中で用いられたとき、説明変数が $\log(\text{dose})$ であるとき、 $ED50$ 値は $\exp(-\hat{\beta}_0 / \hat{\beta}_1)$ によって推定され、そして、式 (2.7) の関係から、 $\log(E\hat{D}50)$ の標準誤差は、式(4.3)によって与えられる。

$$\text{s.e.}\{\log(E\hat{D}50)\} \approx \left\{ \frac{\hat{v}_{00} - 2\hat{\rho}\hat{v}_{01} + \hat{\rho}^2\hat{v}_{11}}{\hat{\beta}_1^2} \right\}^{1/2} \quad (2.8)$$

$E\hat{D}50$ 自体の標準誤差は、式 (2.9) を用いて得られる。

$$\text{s.e.}(ED50) \approx ED50 \text{ s.e.}\{\log(ED50)\} \quad (2.9)$$

$\log(\text{dose})$ が説明変数として使用される場合、 $ED50$ 値の対数の限界の指数をとって $ED50$ 値の信頼区間を計算するほうがよく、式 (2.9) に基づいた対称な間隔を使用するのではなく、式 (2.8) を使用することが望ましい。

同様の手順は、 $ED90$ 値のような他の多量の興味がある量の標準誤差および対応する信頼区間を得るために用いられる。

2.7. フィラーの定理を用いた有効用量の信頼区間

フィラーの定理は 2 つの正規分布の確率変数の比率の信頼区間によって得ることができる一般的な計算結果である。この結果は、 $ED50$ 値の信頼区間の構成を適用する前の一般的な用語で最初に与えられる。 $\rho = \beta_0 / \beta_1$ を考えよう。ここで β_0 と β_1 は、 $\hat{\beta}_0$ と $\hat{\beta}_1$ によって推定され、その平均が β_0 と β_1 、分散が v_{00} と v_{11} 、共分散が v_{01} の正規分布となると仮定される関数 $\psi = \hat{\beta}_0 - \rho\hat{\beta}_1$ を考えよう。

そのとき、 $\hat{\beta}_0$ と $\hat{\beta}_1$ が β_0 と β_1 の不偏推定量であるので、 $E(\psi) = \beta_0 - \rho\beta_1 = 0$ となり、 ψ の分散は、

$$V = \text{Var}(\psi) = v_{00} + \rho^2 v_{11} - 2\rho v_{01} \quad (2.10)$$

で与えられる。 $\hat{\beta}_0$ 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

$$\frac{\hat{\beta}_0 - \rho\hat{\beta}_1}{\sqrt{V}}$$

は、標準正規分布となる。

従って、 $z_{\alpha/2}$ が、標準正規分布の上側 $\alpha/2$ 点であるとしたときに、 ρ の 100 (1- α) % 信頼区間は、

$$|\hat{\beta}_0 - \rho\hat{\beta}_1| \leq z_{\alpha/2} \sqrt{V}$$

値のセットとなる。

両辺を2乗し、等式とし、

$$\hat{\beta}_0^2 + \rho^2 \hat{\beta}_1^2 - 2\rho \hat{\beta}_0 \hat{\beta}_1 - z_{\alpha/2}^2 V = 0$$

を与える。式 (2.10) により V を代入したのちに、式の整理と、次のように ρ に関する二次方程式が得られる。

$$(\hat{\beta}_1^2 - z_{\alpha/2}^2 \hat{v}_{11})\rho^2 + (2\hat{v}_{01}z_{\alpha/2}^2 - 2\hat{\beta}_0\hat{\beta}_1)\rho + (\hat{\beta}_0^2 - \hat{v}_{00}z_{\alpha/2}^2) = 0 \quad (2.11)$$

この二次方程式の2つの根は、 ρ のための信頼限界を構成する。これが、フィラーの結果である。この結果を $ED50 = -\hat{\beta}_0 / \hat{\beta}_1$ の信頼区間を得るために用いるために、式 (2.11) の ρ を $-ED50$ と置き換える。

さらに、パラメータの推定が、線形のロジスティクス・モデルで得られたものならば、大きなサンプルについてのみの近似であるので、分散共分散 \hat{v}_{00} 、 \hat{v}_{11} および \hat{v}_{01} の近似を v_{00} 、 v_{11} および v_{01} の代わりに使用しなければならない。

$ED50$ による二次方程式を書き換えると、

$$(\hat{\beta}_1^2 - z_{\alpha/2}^2 \hat{v}_{11})ED50^2 - (2\hat{v}_{01}z_{\alpha/2}^2 - 2\hat{\beta}_0\hat{\beta}_1)ED50 + (\hat{\beta}_0^2 - \hat{v}_{00}z_{\alpha/2}^2) = 0$$

がえられ、この2方程式を標準的な手順により解き、 $ED50$ 値の $100(1-\alpha)\%$ の信頼限界のために次の式を得る。

$$ED50 = \frac{-\left(\hat{\rho} - g \frac{\hat{v}_{01}}{\hat{v}_{11}}\right) \pm \frac{z_{\alpha/2}}{\hat{\beta}} \left\{ \hat{v}_{00} - 2\hat{\rho}\hat{v}_{01} + \hat{\rho}^2\hat{v}_{11} - g \left(\hat{v}_{00} - \frac{\hat{v}_{01}^2}{\hat{v}_{11}} \right) \right\}^{1/2}}{1 - g} \quad (2.12)$$

ここで、 $\hat{\rho} = \hat{\beta}_0 / \hat{\beta}_1$ 、 $g = z_{\alpha/2}^2 \hat{v}_{11} / \hat{\beta}_1^2$ である。

強い用量反応関係があるとき、 $\hat{\beta}_1$ は0にたいして高度に有意となり、また、 $\hat{\beta}_1 / \sqrt{\hat{v}_{11}}$ は、 $z_{\alpha/2}$ より極めて大きくなるであろう。これらの場合に g は、小さくなる、すなわち、より有意となるような関連の場合、 g はより無視できるようになる。

g が式 (2.12) でゼロである場合、 $ED50$ 値の信頼限界は、式 (2.7) の中で与えられた $E50$ 値の標準誤差の近似に基づくものと一致する。

$\log(\text{dose})$ が説明変数として使用されている場合、 $ED50$ 値の信頼区間は、フィラーの定理を用い $\log(ED50) = -\beta_0 / \beta_1$ について信頼限界を最初に得ることにより計算でき、次に、その値についての間隔の推定限界の結果の指数をとればよい。

2.8. ED50 の信頼区間の計算事例

表 2.1 の SAS/PROBIT によるプロビット法の 6 回のニュートン・ラフソン法の反復の結果 $\hat{\beta}_0 = -15.0703$ 、 $\hat{\beta}_1 = 6.6152$ が得られ、50%致死量は $\log(ED50) = 2.27814$ 、その 95%信頼区間は (2.204, 2.348) が出力されている。2.6節 および 2.7節 で示したデルタ法による近似、フィラーの公式による 95%信頼区間を計算例として示す。

信頼区間の計算のためには、 $\hat{\beta}_0$ および $\hat{\beta}_1$ の分散と共分散が必要であり、SAS/PROBIT の model ステートメントのオプションで covb を指定しておく。

SAS の proc PROBIT の反復計算は、初期の回帰係数 $\beta_0 = 0$ 、 $\beta_1 = 0$ からスタートして、回帰係数として、

$$\beta_0 = -15.0703, \beta_1 = 6.6152$$

分散共分散は、

$$\Sigma = \begin{bmatrix} 10.53242 & -4.58163 \\ -4.58163 & 2.00220 \end{bmatrix}$$

が出力される。デルタ法 (近似法) による信頼区間の計算を試みよう。

$$\log(ED50) = -\frac{\hat{\beta}_0}{\hat{\beta}_1} = \frac{-15.0703}{6.6152} = 2.2781$$

$$\hat{\rho} = \frac{\hat{\beta}_0}{\hat{\beta}_1} = \frac{-15.0703}{6.6152} = -2.2781$$

$$\begin{aligned} \text{s.e.}\{\log(ED50)\} &\approx \left\{ \frac{\hat{v}_{00} - 2\hat{\rho}\hat{v}_{01} + \hat{\rho}^2\hat{v}_{11}}{\hat{\beta}_1^2} \right\}^{1/2} \\ &= \left\{ \frac{10.53242 - 2 \cdot (-2.2781) \cdot (-4.58163) + (-2.2781)^2 \cdot 2.00220}{6.6152^2} \right\}^{1/2} \\ &= 0.033289 \end{aligned}$$

$$\log(ED50) \pm 1.96 \text{ s.e.}(\log(ED50)) = 2.2781 \pm 0.033289 = (2.213, 2.343)$$

これらの数値計算を行うためには 図 2.4 に示したように Excel での使用が便利である。この結果は、フィラーの式での結果 (2.204, 2.348) に対して小さめに推定されている。

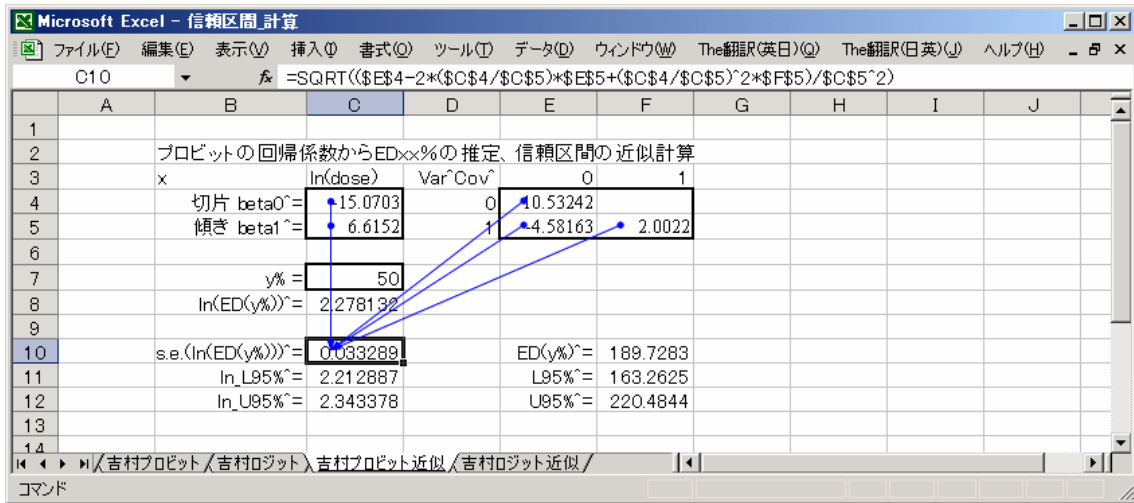


図 2.4 ED50 とその信頼区間の近似計算、Excel による計算

次に、フィラーの式での計算を試みよう。対数の ED50 について 95% 信頼区間は、次の 2 次式の根を求めることにより得られる。

$$(\hat{\beta}_1^2 - z_{\alpha/2}^2 \hat{v}_{11}) \log(ED50)^2 - (2\hat{v}_{01} z_{\alpha/2}^2 - 2\hat{\beta}_0 \hat{\beta}_1) \log(ED50) + (\hat{\beta}_0^2 - \hat{v}_{00} z_{\alpha/2}^2) = 0$$

それぞれの係数は、

$$a = \hat{\beta}_1^2 - z_{\alpha/2}^2 \hat{v}_{11} = 6.6152^2 - 1.96^2 \cdot 2.0022 = 36.80616$$

$$b = -(2\hat{v}_{01} z_{\alpha/2}^2 - 2\hat{\beta}_0 \hat{\beta}_1) = -(2 \cdot (-4.03832) \cdot 1.96^2 - 2 \cdot (-15.0414) \cdot 6.603) = -164.185$$

$$c = \hat{\beta}_0^2 - \hat{v}_{00} z_{\alpha/2}^2 = (-15.0414)^2 - 9.26724 \cdot 1.96^2 = 186.6526$$

となる。これを 2 次式の公式に代入すると 95% 信頼区間が得られる。

$$\log(ED50) \pm 1.96 \text{ s.e.}(\log(ED50)) = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = (2.204, 2.348)$$

これらの数値計算は煩雑であるので、Excel による計算シートを 図 2.5 に示す。

プロビット、回帰係数からEDxx%の推定、信頼区間の計算				
ln(dose)	Var [^] Cov [^]		0	1
切片 beta0 [^] =	-15.0703	0	10.53242	
傾き beta1 [^] =	6.6152	1	-4.58163	2.0022
y% =	50			
ln(ED(y%) [^])=	2.278132			ED(y%) [^] = 9.758436
Q_a=	36.06922	ln_L95% [^] =	2.204065	L95% [^] = 159.9798
Q_b=	-164.185	ln_U95% [^] =	2.347863	U95% [^] = 222.7735
Q_c=	186.6526			
計算はフィラーの式、2次式の根の公式により計算				

図 2.5 ED50 とその信頼区間のフィラーの公式、Excel による計算

ED50 とその信頼区間の計算公式は、プロビット法の場合のみならずロジット法の場合もそのまま適用できる。JMP のロジスティック回帰係数は、

$$\hat{\beta}_0 = -26.2115、\hat{\beta}_1 = 11.5229$$

分散共分散は、JMP の行列計算により

$$\Sigma = \begin{bmatrix} 40.328 & -17.623 \\ -17.623 & 7.731 \end{bmatrix}$$

が計算されているので、デルタ法 (近似法) による信頼区間の計算は、Excel の計算シートに数値を入れ替えることにより

$$\log(ED50) \pm 1.96 \text{ s.e.}(\log(ED50)) = 2.2748 \pm 0.0343 = (2.2075, 2.3420)$$

求められる。

ロジットの回帰係数からEDxx%の推定、信頼区間の計算				
ln(dose)	Var [^] Cov [^]		0	1
切片 beta0 [^] =	-26.212	0	40.328	
傾き beta1 [^] =	11.523	1	-17.623	7.731
y% =	50			
ln(ED(y%) [^])=	2.274755			
s.e.(ln(ED(y%))) [^] =	0.03429		ED(y%) [^] =	188.2586
ln_L95% [^] =	2.207546		L95% [^] =	161.2672
ln_U95% [^] =	2.341964		U95% [^] =	219.7676

図 2.6 ED50 とその信頼区間の近似計算

フィラーの公式も同様にロジスティック回帰の場合に適用できる。JMP による信頼区間の計算は、フィラーの公式によって計算されている。Excel の計算シートに数値を入

れ替えた結果を 図 2.7 に示す。

ロジットの回帰係数からEDxx%の推定、信頼区間の計算				
ln(dose)		Var [^] Cov [^]	0	1
切片 beta0 [^] =	-26.212	0	40.328	
傾き beta1 [^] =	11.523	1	-17.623	7.731
y% =	50			
ln(ED(y%)) [^] =	2.274755			ED(y%) [^] = 9.725534
Q_a=	103.0801		ln_L95% [^] = 2.19709	L95% [^] = 8.998785
Q_b=	-468.681		ln_U95% [^] = 2.349672	U95% [^] = 10.48213
Q_c=	532.1449			

図 2.7 ロジット法のフィラーの公式による 95%信頼区間

注) 計算精度が悪いので、分散の有効数字を増やす

吉村の例では、投与量の常用対数を取り、変換されたプロビットに、重み付き最小 2 乗法を 2 回繰り返しにより、次の回帰係数が得られた。

$$\eta_i = \text{probit}(\hat{p}_i) = -10.0414 + 6.6030 \log_{10}(\text{dose}_i)$$

死亡率の推定値は、

$$\hat{\mu} = \log_{10} \text{LD}_{50} = -((\hat{\beta}_0 - 5) / \hat{\beta}_1) = -(-10.0414 - 5) / 6.6030 = 2.278$$

と標準正規分布 $\Phi(\eta_i - 5)$ によって計算された。JMP の行列計算により、分散共分散は、

$$\hat{\Sigma} = \begin{bmatrix} 9.2672 & -4.0383 \\ -4.0383 & 1.7684 \end{bmatrix}$$

と推定されるので、これらを Excel のフィラーの計算シートから

切るときの X 軸の目盛りを読むことにより得られる。解析的には、フィラーの式により、常用対数で (2.206, 2.347) となり、投与用量に変換して (161, 223) が得られる。フィラーの式は、2.7 節で詳しく述べるので、ここでは結果のみを示した。

プロビット、回帰係数からEDxx%の推定、信頼区間の計算					
ln(dose)		Var [^] Cov [^]		0	1
切片 beta0 [^] =	-15.0414	0	9.26724		
傾き beta1 [^] =	6.603	1	-4.03832	1.7684	
y% =	50				
ln(ED(y%) [^])=	2.277965				ED(y%) [^] = 9.756801
Q_a=	36.80612		ln_L95% [^] = 2.208091	L95% [^] =	161.4696
Q_b=	-167.61		ln_U95% [^] = 2.345758	U95% [^] =	221.6959
Q_c=	190.6427				
計算はフィラーの式、2次式の根の公式により計算					

図 2.8 2 回反復のプロビット法のフィラーの公式による 95%信頼区間
 小数点 4 桁目が吉村のテキストと一致しない。計算誤差の問題であろう

3. 最尤法による計算の例

3.1. 吉村の例、JMP の行列計算言語による

プロビット変換、重み付最小 2 乗法の繰り返し 2 回

吉村本の表 5-13 の結果に合わせて計算結果を出しているので、対応を取ることで、最小 2 乗法の計算原理を学習してもらいたい。分散共分散行列の計算は、信頼区間の計算のために追加している。

```

/* Probit, yoshimura(1988), p226, Weighted linear Regression */

d = [101, 136, 183, 247, 333, 450] ;
x = log10(d) ;
y = [ 0,  2,  5,  8,  9, 10] ;
n = [ 10, 10, 10, 10, 10, 10] ;

p  = y ./ n ;
X  = J(nrow(X), 1) || X;
beta = [-7.6208, 5.5594] ;          show(round(beta, 4)) ;
eta  = X*beta ;                   show(round(eta0, 4)) ;
A    = d || x || y || n || p || eta ; show(round(A, 4)) ;

/** step 1***/
pi  = Normal Distribution(eta - 5) ; show(round(pi, 4)) ;
z   = Normal Density(eta - 5) ;    show(round(z, 4)) ;
w   = z^2 ./ (pi .* (1-pi)) ;      show(round(w, 4)) ;
nw  = n .* w ;                    show(round(nw, 4)) ;
y   = eta + (p - pi) ./ z ;        show(round(y, 4)) ;
W   = diag(w) ;
Beta = Inv(X`*W*X)*X`*W*y ;        show(round(beta, 4)) ;
etal = X*beta ;                   show(round(etal, 4)) ;
diff = abs(eta - etal) ;           show(round(diff, 4)) ;
B    = X || pi || z || nw || y || eta || diff ; show(round(B, 4)) ;
eta=etal ;

/** step2 ***/
pi  = Normal Distribution(eta - 5) ; show(round(pi, 4)) ;
z   = Normal Density(eta - 5) ;    show(round(z, 4)) ;
w   = z^2 ./ (pi .* (1-pi)) ;      show(round(w, 4)) ;
nw  = n .* w ;                    show(round(nw, 4)) ;
y   = eta + (p - pi) ./ z ;        show(round(y, 4)) ;
W   = diag(w) ;
Beta = Inv(X`*W*X)*X`*W*y ;        show(round(beta, 4)) ;
etal = X*beta ;                   show(round(etal, 4)) ;
diff = abs(eta - etal) ;           show(round(diff, 4)) ;
B    = X || pi || z || nw || y || eta || diff ; show(round(B, 4)) ;

res=(y-eta)`*W*(y-eta) ;
covb = Inv(X`*W*X)*res ;

```

Round(beta, 4) : [-7.6208, 5.5594]

Round(A, 4) :

```
/*      d      x      y      n      p      eta      */
[      101 1 2.0043 0 10 0 3.522,
      136 1 2.1335 2 10 0.2 4.2404,
      183 1 2.2625 5 10 0.5 4.9571,
      247 1 2.3927 8 10 0.8 5.6812,
      333 1 2.5224 9 10 0.9 6.4025,
      450 1 2.6532 10 10 1 7.1295]
```

Round(beta, 4) : [-9.5978, 6.4113]

Round(B, 4) :

```
/*      X      pi      z      nw      y(1)      eta(1)      diff(1)      */
[      1 2.0043 0.0697 0.1338 2.7621 3.0012 3.522 0.2695,
      1 2.1335 0.2237 0.299 5.1461 4.161 4.2404 0.1594,
      1 2.2625 0.4829 0.3986 6.3619 5 4.9571 0.0496,
      1 2.3927 0.7521 0.3163 5.3676 5.8325 5.6812 0.0614,
      1 2.5224 0.9196 0.1492 3.0116 6.271 6.4025 0.1719,
      1 2.6532 0.9834 0.0413 1.0457 7.5313 7.1295 0.2833]
```

Round(beta, 4) : [-10.0414, 6.603]

< ===== 回帰係数

Round(B, 4) :

```
/*      X      pi(2)      z(2)      nw(2)      y(2)      eta(2)      diff(2)      */
[      1 2.0043 0.0403 0.0867 1.9427 2.7877 3.2525 0.0594,
      1 2.1335 0.179 0.2615 4.653 4.1611 4.081 0.0347,
      1 2.2625 0.4631 0.3972 6.3464 5.0003 4.9075 0.01,
      1 2.3927 0.7711 0.3028 5.1957 5.8379 5.7425 0.015,
      1 2.5224 0.9423 0.1155 2.4547 6.2082 6.5744 0.0399,
      1 2.6532 0.9921 0.0217 0.6006 7.7773 7.4128 0.0649]
```

covb = Inv(X`*W*X)*res

< ===== 分散共分散行列

```
[      9.26723972492964 -4.03832252490738,
      -4.03832252490738 1.76839949392387]
```

3.2. SAS の POBIT プロシジャによるプロビット法による計算

プロビット変換、ニュートン・ラフソン法

```
Title 'ANZ09a.sas 2002-1-15 Y.Takahashi ' ;
```

```
data d01 ;
  input i dose n y p eta p_hat ;
datalines ;
1 101 10 0 0 3.1931 0.035
2 136 10 2 0.2 4.0463 0.170
3 183 10 5 0.5 4.8975 0.459
4 247 10 8 0.8 5.7575 0.776
5 333 10 9 0.9 6.6142 0.947
6 450 10 10 1 7.4777 0.993
;
proc probit data=d01 log10 inversecl ;
  model y/n = dose / dist=normal itprint covb ;
  output out=out01 p=p std=std xbeta=xbeta ;
run ;
proc print data=out01 ;
run ;
```

Probit Procedure

Iteration History for Parameter Estimates				
Iter	Ridge	Loglikelihood	Intercept	Log10(dose)
0	0	-41.588831	0	0
1	0	-22.989246	-9.338521413	4.0829799764
:				
6	0	-20.872686	-15.07025874	6.6151614919

Model Information	
Number of Observations	6
Number of Events	34
Number of Trials	60
Name of Distribution	Normal
Log Likelihood	-20.87268602

==== 回帰係数 =====

Analysis of Parameter Estimates							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-15.0703	3.2454	-21.4311	-8.7095	21.56	<.0001
Log10(dose)	1	6.6152	1.4150	3.8418	9.3885	21.86	<.0001

==== 分散共分散行列 =====

Estimated Covariance Matrix		
	Intercept	Log10(dose)
Intercept	10.532418	-4.581626
Log10(dose)	-4.581626	2.002200

Probit Model in Terms of Tolerance Distribution	
MU	SIGMA
2.2781392	0.15116789

Probit Analysis on Log10(dose)

Probability	Log10(dose)	95% Fiducial Limits	
0.10	2.08441	1.91484	2.16702
0.50	2.27814	2.20406	2.34788
0.90	2.47187	2.39327	2.62875

Probit Analysis on dose

Probability	dose	95% Fiducial Limits	
0.10	121.45342	82.19450	146.90043
0.50	189.73140	159.97966	222.78290
0.90	296.39348	247.32889	425.35599

OBS	i	dose	n	y	p	eta	p_hat	p2	xbeta	std
1	1	101	10	0	0.0	3.1931	0.035	0.03504	-1.81135	0.45800
2	2	136	10	2	0.2	4.0463	0.170	0.16940	-0.95655	0.31025
3	3	183	10	5	0.5	4.8975	0.459	0.45867	-0.10378	0.22280
4	4	247	10	8	0.8	5.7575	0.776	0.77572	0.75782	0.26481
5	5	333	10	9	0.9	6.6142	0.947	0.94697	1.61612	0.39758
6	6	450	10	10	1.0	7.4777	0.993	0.99345	2.48117	0.56118

3.3. JMP の行列言語によるロジスティック回帰

ロジット変換、ニュートン・ラフソン法

```

/* The linear logistic model, Newton-Raphson 2002-1-13 Y.Takahashi */

d = [101, 136, 183, 247, 333, 450] ;
x = log10(d) ;
y = [ 0, 2, 5, 8, 9, 10] ;
n = [ 10, 10, 10, 10, 10, 10] ;

X = J(nrow(x),1) || x ;
p = y :/ n ;
pi = (n:/2) :/ n ;
A = X || y || n || p || pi ; show(round(A, 3)) ;

beta=[0, 0] ;
L0 = [-1E10] ;
epsilon=1 ;

/** step << i >> ** */
for (i=1, epsilon>1E-8, i++,
  show("iteration"); show(round(i, 0)) ;
  u = X` * (y - n :* pi) ; show(round(u, 3)) ;
  v = pi :* (1-pi) :* n ; show(round(v, 3)) ;
  V = diag(v) ;
  Im = X` * V * X ; show(round(Im, 3)) ;
  beta= beta + inv(Im) * u ; show(round(beta, 3)) ;
  inv_Im=inv(Im) ; show(round(inv_Im, 3)) ;
  eta = X * beta ;
  pi = exp(eta) :/ (1 + exp(eta)) ;
  B = X || eta || pi ; show(round(B, 3)) ;

  L1 = y` * ln(pi) + (n - y)` * ln (1-pi) ;
  epsilon=abs(( L1-L0)/L0) ;
  C =L1 || L0 || epsilon ; show(C) ;

  L0=L1 ;
) ;

```

Round (A, 3) :

```
[ 1 2.004 0 10 0 0.5,
  1 2.134 2 10 0.2 0.5,
  1 2.262 5 10 0.5 0.5,
  1 2.393 8 10 0.8 0.5,
  1 2.522 9 10 0.9 0.5,
  1 2.653 10 10 1 0.5]
```

"iteration"

Round (i, 0) :1

Round (u, 3) : [4, 14.112]

Round (v, 3) : [2.5, 2.5, 2.5, 2.5, 2.5, 2.5]

Round (Im, 3) :

```
[ 15 34.922,
  34.922 82.038]
```

Round (beta, 3) : [-14.902, 6.515]

Round (inv_Im, 3) :

```
[ 7.425 -3.161,
 -3.161 1.358]
```

Round (B, 3) :

```
[ 1 2.004 -1.843 0.137,
  1 2.134 -1.001 0.269,
  1 2.262 -0.161 0.46,
  1 2.393 0.687 0.665,
  1 2.522 1.533 0.822,
  1 2.653 2.385 0.916]
```

G: [-23.3831353866953 -10000000000 0.99999997661687]

"iteration" Round (i, 0) :2 Round (beta, 3) : [-21.999, 9.653]

"iteration" Round (i, 0) :3 Round (beta, 3) : [-25.499, 11.206]

"iteration" Round (i, 0) :4 Round (beta, 3) : [-26.189, 11.513]

"iteration" Round (i, 0) :5 Round (beta, 3) : [-26.211, 11.523]

"iteration" Round (i, 0) :6

Round (u, 3) : [0, 0]

Round (v, 3) : [0.407, 1.373, 2.488, 1.626, 0.515, 0.124]

Round (Im, 3) :

```
[ 6.532 14.891,
  14.891 34.075]
```

Round (beta, 3) : [-26.212, 11.523] < ===== 回帰係数

Round (inv_Im, 3) :

```
[ 40.328 -17.623,
 -17.623 7.731]
```

< ===== 分散共分散行列

Round (B, 3) :

```
[ 1 2.004 -3.116 0.042,
  1 2.134 -1.627 0.164,
  1 2.262 -0.141 0.465,
  1 2.393 1.359 0.796,
  1 2.522 2.854 0.946,
  1 2.653 4.361 0.987]
```

G: [-20.9842048618161 -20.9842048618224 3.00345622228392e-13]