2004. 2. 10

最小2乗法，最尤法
線形モデル，非線形モデル

芳賀　敏郎

目 次

0　まえがき　1

0.1　**テキストの目的**　1

0.2　**非線形関係と非線形回帰分析**　2

1　線形モデルの最小2乗法　4

1.1　**平均値**　4

(1)　**平均値は最小2乗推定**　4

(2)　**平均値の標準誤差**　5

1.2　**単回帰式**　7

(1)　**正規方程式の導出**　7

(2)　**正規方程式の解法**　8

(3)　LINEST **関数による解法**　10

(4)　Excel **ソルバーによる回帰係数の推定**　11

(5)　**回帰係数の標準誤差の推定**　12

1.3　**共通の切片を持つ2本の回帰直線**　16

(1)　**データと単純な解析**　16

(2)　**共通の切片を持つ回帰式の推定**　16

(3)　**別のモデルによる解析**　18

1.4　**重みつき最小2乗法（未完）**　19

2 非線形モデルの最小２乗法 **20**

2.1 **回帰式による逆推定** · · · · · 20

(1) **従来の方法** · · · · · 20

(2) **非線形回帰式の当てはめ** · · · · · 22

(3) c **の標準誤差** · · · · · 24

(4) c **の区間推定** · · · · · 25

2.2 **傾斜の比の推定** · · · · · 26

2.3 **ロジスティック曲線の当てはめ** · · · · · 28

(1) **例題** · · · · · 28

(2) Excel **ソルバーによる解析** · · · · · 29

(3) **推定値** c **の標準誤差** · · · · · 29

(4) **特殊な場合の注意** · · · · · 31

3 最尤法 **33**

3.1 **確率と尤度** · · · · · 33

(1) **２項分布** · · · · · 33

(2) χ **2乗分布** · · · · · 34

3.2 **２項分布の** π **の推定** · · · · · 35

(1) **点推定** · · · · · 35

(2) **区間推定** · · · · · 36

(3) **補足: 最小**2**乗推定と最尤推定** · · · · · 37

3.3 **ロジスティック回帰分析** · · · · · 38

(1) **データとモデル** · · · · · 38

(2) **最尤法によるロジスティック回帰分析** · · · · · 39

(3) c **の信頼区間** · · · · · 40

(4) D50 **の差の推定と検定** · · · · · 41

0　まえがき

0.1　テキストの目的

ほとんどの統計のテキストには「回帰分析」が説明されている．回帰係数の推定には「最小2乗法」が用いられる．

しかし，最小2乗法はもっと簡単な場合，すなわち，平均値も最小2乗法で導かれる．

平均値や回帰係数の標準誤差を求める式はよく知られているが，その意味をべつの面から理解することは，通常の回帰分析（線形回帰分析）を非線形回帰分析に発展させることに役立つであろう．

ソルバーを使って非線形回帰式のパラメータを推定し，さらに，その標準誤差や信頼区間を求める手順を説明する．

最小2乗法による推定は分かっても，最尤法による推定を十分に理解して使っている人は少ない．

最尤法を理解するためには，まず，確率 と 尤度 の違いを把握しなければならない．

最尤法の考え方と手順を，簡単な例から始め，最後にロジスティック回帰分析を説明する．

従来の統計のテキストは，数式の羅列か，または，解析プログラムの使い方マニュアルに近いものが少なくない．

このテキストでは，数式は最小限に抑え，簡単な数値を使って，解析の考え方を説明し，結果をグラフ化して示すように努めた．

このテキストは，読むだけでなく，添付されるExcelファイルを開き，計算の過程を一つ一つ追いかけることにより，理解を深めることができるであろう．

Excelシートの左上の「名前ボックス」をクリックすると，プルダウンメニューに`fig1_1`，...などが表示される．

テキストの表示番号に対応する項目をクリックすると，テキストの表示を見ることができる．

Excel ファイルの中にはテキストには省略した表やグラフが含まれている．

0.2　非線形関係と非線形回帰分析

通常の重回帰分析（線形回帰分析）は，目的変数 y と説明変数 $x_1,\ x_2,\ \ldots$ について

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon$$

というモデルを仮定して，解析するものである．

ここで，説明変数 x は単純な変数である必要はなく，

$$y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

のように，多項式であったり，積の項を含んでもかまわない．また，

$$y = \beta_0 + \beta_1 \ln(x_1) + \beta_2 \exp(x_2)$$

のような関数であってもよい．

これらは，説明変数と目的変数は曲線関係（非線形関係）である．

しかし，$x_1^2, x_1 x_2,\ \ln(x_1), \exp(x_1)$ などを一つの変数としてあつかえば，回帰分析によってモデルのパラメータを推定することができる．

そのための条件は，「パラメータ β に関して線形（一次式）である」ことである．

数学的には，y を一つのパラメータ β で偏微分するとすべてのパラメータが消えることに対応する．この性質は本文で再度取り上げる．

それに対して，

$$y = \alpha_0 \alpha_1^{x_1} \alpha_2^{x_2}$$
$$y = \alpha_0 x_1^{\alpha_1} x_2^{\alpha_2}$$

は，パラメータ α に関して線形ではない．したがってこのモデルのパラメータは通常の回帰分析では解けない．

このような問題は，パラメータに関して線形になるように，モデル式を変形して解析することが多い．上の例では，両辺の対数をとることにより，

$$\ln(y) = \ln(\alpha_0) + \ln(\alpha_1) x_1 + \ln(\alpha_2) x_2 = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$
$$\ln(y) = \ln(\alpha_0) + \alpha_1 \ln(x_1) + \alpha_2 \ln(x_2) = \beta_0 + \alpha_1 X_1 + \alpha_2 X_2$$

となり，パラメータに関して線形のモデルに変換し，通常の回帰分析が適用される．

この場合は，元のモデルで等分散が成立しても，対数変換後には等分散が成立しない場合があり，注意が必要である．逆に，対数変換することにより，等分散も成立するという場合も少なくない．

ロジスティック回帰分析のモデルは，

$$\pi = \frac{1}{1+\exp(-(\beta_0+\beta_1 x))} \tag{0.1}$$

である．これをロジット変換すると

$$z = \ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x \tag{0.2}$$

となり，パラメータに関して線形式が導かれる．従来はこの関係を使って，ロジステック回帰分析が実行されていた．

しかし，式(0.2) で z を計算するとき，π の代わりのその推定値である $p = r/n$ を用いると，$r = 0$ またはn のとき，$p = 0$ または1 となり，z が $\pm\infty$ になるために，経験ロジット

$$z = \ln\left(\frac{r+0.5}{n-r+0.5}\right)$$

を用いるという技巧をこらす必要があった．

また，$\pi = 0.5$ となる x（これを $x_{0.5}$ で表わす）を知りたいときは，$\beta_0,\ \beta_1$ の推定値 $b_0,\ b_1$ を使って

$$x_{0.5} = -b_0/b_1$$

として推定される．

ここで，$\beta_0,\ \beta_1$ の標準誤差から $\widehat{x}_{0.5}$ の標準誤差を求めるためには特別の技巧を必要とする．

非線形回帰分析の計算方法が進歩し，それらを備えた統計解析プログラムが普及した現在，無理やり線形化して解析する必要は少なくなったといえるであろう．

ここでは，母数を $\beta,\ \varepsilon$ のギリシャ文字で，推定値を b のアルファベットで区別したが，次章以降では，一部を除き，あえて区別しないで，アルファベットを用いる．

1　線形モデルの最小２乗法

1.1　平均値

(1)　平均値は最小２乗推定

観測値 x_i の代表値 a として,

$$Q = \sum_{i=1}^{n} (x_i - a)^2 \quad \Rightarrow \quad \min$$

が最小となる値を用いる（最小2 乗法）.

横軸に a を，縦軸に Q を取って，グラフを描くと放物線（2次曲線）となる.

この曲線は,

$$\begin{aligned} Q &= \sum_{i=0}^{n} (x_i - a)^2 = \sum_{i=0}^{n} \{(x_i - \overline{x}) - (\overline{x} - a)\}^2 \\ &= \sum_{i=0}^{n} (x_i - \overline{x})^2 - 2 \underbrace{\sum_{i=0}^{n} (x_i - \overline{x})}_{=0} (\overline{x} - a) + \sum_{i=0}^{n} (\overline{x} - a)^2 \\ &= \sum_{i=0}^{n} (x_i - \overline{x})^2 + n(\overline{x} - a)^2 \end{aligned} \tag{1.1}$$

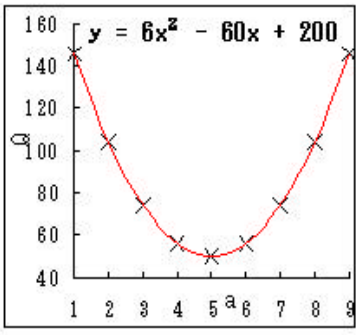
で表される.

2 次曲線の極小点は平均値 $\overline{x}$ で，そこでの Q の値は残差平方和 $S = \sum_{i=0}^{n} (x_i - \overline{x})^2$ となることが分かる.

以上を簡単な数値（$n = 6,\ \overline{x} = 5$) で確認する.

表示1.1: a と Q の関係

	a								
x	1	2	3	4	5	6	7	8	9
1	0	−1	−2	−3	−4	−5	−6	−7	−8
3	2	1	0	−1	−2	−3	−4	−5	−6
4	3	2	1	0	−1	−2	−3	−4	−5
5	4	3	2	1	0	−1	−2	−3	−4
7	6	5	4	3	2	1	0	−1	−2
10	9	8	7	6	5	4	3	2	1
Q	146	104	74	56	50	56	74	104	146



Q **は** $a = 5(= \overline{x})$ **で極小となり，その値は** $S_e = 50$ **である．**

a **と** Q **のグラフで求められた**2**次式は**

$$Q = 200 - 60a + 6a^2 = 50 + 6(25 - 10a + a^2) = 50 + 6(a-5)^2$$

と変形される．この関係は式(1.1) **に対応し，**2 **次項の係数の** 6 **は** n **に対応していることが分かる．**

(2) 平均値の標準誤差

平均値 $\overline{x}$ **の標準誤差**(Standard Error) $se(\overline{x})$ **は，**

$$se(\overline{x}) = \frac{\sigma}{\sqrt{n}} \tag{1.2}$$

である．

分子の σ **は，残差平方和** S_e **を自由度** $f_e = n - 1$ **で割った平均平方** V_e **の平方根で推定される．推定値を** s **で表わすと，**

$$s = \sqrt{V_e} = \sqrt{\frac{S_e}{n-1}} = \sqrt{\frac{50}{5}} = \sqrt{10}$$

となる．式(1.2) **の分子の** σ **をその推定値で置き換えると，**

$$se(\overline{x}) = \sqrt{\frac{V_e}{n}} = \sqrt{\frac{10}{6}} = 1.291$$

が得られる．

$a = \overline{x} \pm se(\overline{x}) = 5 \pm 1.291$ **に対する** Q **は**

$$Q = S_e + n(a - \overline{x})^2 = S_e + n \times se(\overline{x})^2 = S_e + V_e = 50 + 10 = 60$$

となる．これは，表示1.1 **の** a **に** 3.709, 6.291 **を入力することで確かめられる．**

これを逆に考えると，$Q = S_e + V_e$ **となる** a **が分かれば，推定値** $\overline{x}$ **の標準誤差が求められることになる．**

これを拡張すると，母平均 μ **の信頼率** 0.05 **の信頼区間**

$$\overline{x} - t(f_e, 0.05)se(\overline{x}) < \mu < \overline{x} + t(f_e, 0.05)se(\overline{x})$$

$$5 - 2.571 \times 1.291 < \mu < 5 + 2.571 \times 1.291$$

$$1.681 < \mu < 8.319$$

は,

$$\begin{aligned} Q &= S_e + F(1, f_e;\ 0.05)V \\ &= 50 + 6.608 \times 10 = 116.08 \end{aligned}$$

となる a に対応することが導かれる．ここで，$t(f_e, 0.05)^2 = F(1, f_e;\ 0.05)$ の性質を利用している．

以上の手順により，推定値（ここでは，平均値）の標準誤差を，推定値 a と Q の関係から導くことができた．この関係は，次節以降様々に展開されていく．

1.2 単回帰式

最小２乗法を，単回帰分析に拡張する．
平均値の場合と同様に，下に示す簡単な数値を使って説明する．

表示1.2: **回帰分析用データ**

	データ						平均	平方和
x	1	3	4	5	7	10	5.0	50.0
y	5	5	7	6	9	10	7.0	22.0

まず，統計の標準的な教科書に書かれている，単回帰分析の手順を説明し，ついで，Excel**の**LINEST**関数による解法とソルバーによる解法を説明する．**

(1) 正規方程式の導出

線形回帰式

$$y = a + bx$$

の係数 a, b **は**

$$Q = \sum_{i=1}^{n} (y_i - (a + bx_i))^2$$

が最小になるように決められる．Q **を** $a,\ b$ **で偏微分すると，**

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} 2(y_i - (a + bx_i)) \frac{\partial}{\partial a}(y_i - (a + bx_i)) = \sum_{i=1}^{n} 2(y_i - (a + bx_i))(-1)$$
$$\frac{\partial Q}{\partial b} = \sum_{i=1}^{n} 2(y_i - (a + bx_i)) \frac{\partial}{\partial b}(y_i - (a + bx_i)) = \sum_{i=1}^{n} 2(y_i - (a + bx_i))(-x_i) \tag{1.3}$$

となる．これらの式に $=0$ **を追加すると，**$a,\ b$ **を未知数とする連立１次方程式**

$$\begin{aligned} na &+ \sum_{i=1}^{n} x_i b &= \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i a &+ \sum_{i=1}^{n} x_i^2 b &= \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{1.4}$$

が導かれる．これを，正規方程式 と呼ぶ．

Q **を偏微分する過程で，**$\dfrac{\partial}{\partial a}(y_i - (a + bx_i)),\ \dfrac{\partial}{\partial b}(y_i - (a + bx_i))$ **が** $-1,\ -x_i$ **となり，**$a,\ b$ **が消え，**$a,\ b$ **の１次式になる．これが，線形最小２乗法の基本条件である．**

このデータに対する正規方程式は

$$\begin{aligned} 6a &+ 30b &= 42 \\ 30a &+ 200b &= 241 \end{aligned}$$

である.

(2)　正規方程式の解法

正規方程式を解くと，$a,\ b$ が求められる．その標準的な方法は消去法（掃き出し計算）である.

計算の過程を表示1.3に示す.

連立方程式の係数行列の下に$\sum y_i,\ \sum x_i y_i,\ \sum y_i^2$ を追加し，さらに，右に単位行列を追加して，表示1.3 の上の3行（A）を準備する.

表示1.3: 正規方程式の解法

		a	b	1	a	b
A	$(I)_A$	**6**	30	42	1	0
	$(II)_A$	30	200	241	0	1
	$(III)_A$	42	241	316	0	0
B	$(I)_B = (I)_A/6$	1	**5**	**7**	0.167	0
	$(II)_B = (II)_A - 30(I)_B$	0	**50**	**31**	–5.000	1
	$(III)_B = (III)_A - 42(I)_B$	0	31	**22**	–7.000	0
C	$(I)_C = (I)_B - 5(II)_C$	1	0	**3.90**	**0.667**	**–0.100**
	$(II)_C = (II)_B/50$	0	1	**0.62**	**–0.100**	**0.020**
	$(III)_C = (III)_B - 31(II)_C$	0	0	**2.78**	–3.900	–0.620

A の **6** をピボットとして掃き出し，B を求める．$(I)_B$ の列には $\overline{x} = 5,\ \overline{y} = 7$ が，$(II)_B,\ (III)_B$ には平方和と積和 $S_{xx} = 50,\ S_{xy} = 31,\ S_{yy} = 22$ が得られている.

B の**50** をピボットとして掃き出し，C を求める．C の 1 の列に，係数 $a = 3.90,\ b = 0.62$ と残差平方和 $S_e = 2.78$ が得られている.

C の右の $\begin{pmatrix} 0.667 & -0.100 \\ -0.100 & 0.020 \end{pmatrix}$ は正規方程式の係数行列 $\begin{pmatrix} 6 & 30 \\ 30 & 200 \end{pmatrix}$ の逆行列である.

データ行列から正規方程式の係数行列を求め，逆行列を経て，解と残差平方和を求める過程は，Excel の行列関数を使うと極めて簡潔になる.

表示1.4に示す.

表示1.4: **行列演算による解法**

	A	B	C	D	E	F	G	H	I	J
47		1	x	y						
48		1	1	5						
49		1	3	5						
50		1	4	7						
51		1	5	6						
52		1	7	9						
53		1	10	10						
54										
55		1	x	y						
56	1	6	30	42	B56:D58	=MMULT(TRANSPOSE(B48:D53),B48:D53)				
57	x	30	200	241						
58	y	42	241	316						
59										
60	1	0.67	-0.10	3.90	B60:C61	=MINVERSE(B56:C57)				
61	x	-0.10	0.02	0.62	D60:D61	=MMULT(B60:C61,D56:D57)				
62	y			2.78	D62	=D58-MMULT(B58:C58,D60:D61)				

表示1.4**の上に示すように，データ** $x,\ y$ **の左に** 1 **の列を追加する．この行列を** $\boldsymbol{XY}$ **で表わす．**

$\boldsymbol{XY}$ **の転置行列** $\boldsymbol{XY}^T$ **と** $\boldsymbol{XY}$ **の積を「行列の積を求める関数」**MMULT **で求める．黒枠で囲った領域を反転して，表示**1.4 **の右に示す式を入力したのち，**Ctrl **キーと** Shift **キーを押しながら** Enter **キーを押す**[1]**．**

得られた行列を4 **つの行列に分解する．**

$$\boldsymbol{XY}^T\boldsymbol{XY} = \begin{pmatrix} \boldsymbol{X}^T\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{Y} \\ \boldsymbol{Y}^T\boldsymbol{X} & \boldsymbol{Y}^T\boldsymbol{Y} \end{pmatrix}$$

$\boldsymbol{X}^T\boldsymbol{X}$ **の逆行列** $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ **を，「行列の逆行列を求める関数」**MINVERSE **で求める．**

係数 $a,\ b$ **のベクトル** $\boldsymbol{B}$ **は，** $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ **で求められる．残差平方和** S_e **は** $\boldsymbol{Y}^T\boldsymbol{Y}\boldsymbol{B}$ **で求められる．**

当然のことながら，表示1.3 **の結果と一致する．**

残差の自由度はデータ数 n **から推定したパラメータの個数（ここでは** a,b **の**2**個）を引**

[1] MMULT **関数のように，複数のセルが同時に求めるときは，出力領域を反転してから，数式を入力し，**Ctrl **キーと** Shift **キーを押しながら** Enter **キーを押す．以下の** MINVERSE **関数，次項の** LINEST **関数も同様である．**
このような場合は，セルの内容を見ると，入力した式の前後に{ } **が付加されている．**

いた $f_e = n - 2 = 6 - 2 = 4$ である．残差平均平方は $V_e = S_e/f_e = 2.78/4 = 0.695$ として求められる．

係数 a, b の分散は，V_e と逆行列の対角要素の積となる．従って，それらの標準誤差は

$$
\begin{aligned}
se(a) &= \sqrt{0.695 \times 0.667} = 0.6807 \\
se(b) &= \sqrt{0.695 \times 0.020} = 0.1179
\end{aligned}
\tag{1.5}
$$

となる．

以上をまとめると，推定された回帰式は

$$
\begin{array}{rccc}
y = & 3.900 & + 0.620x & \\
 & (0.681) & (0.118) &
\end{array}
$$

となる．係数の下に標準誤差を示す．

(3)　LINEST 関数による解法

以上が，回帰分析の標準的な手順であるが，Excel の`LINEST`関数を用いると，一度に解を求めることができる．

5行2列の出力領域を反転してから，

`=LINEST(y` の範囲，`x` の範囲，，`TRUE)`

を入力し，**Ctrl** キーと **Shift** キーを押しながら Enter キーを押す．

表示1.5 の結果が得られる（黒枠の内側が出力で，周囲の文字は追加したものである）．

表示 1.5: LINEST 関数の結果

	x	const	
b	0.620	3.900	
$se(b)$	0.118	0.681	
R^2	0.874	0.834	s_e
F	27.655	4	f_e
S_R	19.220	2.780	S_e

回帰式の係数とその標準誤差が出力の1,2行目に求められている．

残差平方和 $S_e = 2.780$ が出力の右下に，その上に自由度が求められている．

実務データの解析に，LINEST 関数や分析ツールを用いるのはかまわないが，その裏ではここに述べたような計算が実行されていることを承知していることが望ましい．

(4) Excel ソルバーによる回帰係数の推定

前節では $a,\ b$ を代数的に計算した.

さらに，最小2乗法の基本に立ち返って，回帰式を推定する方法を考える.

表示1.6の左のように，$x,\ y$ を入力し，C列に $\widehat{y}$ の列を準備する.

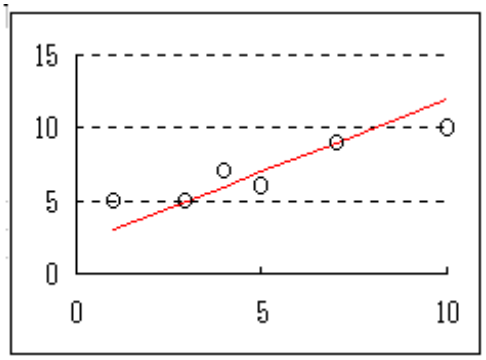
C2，C3 のセルに $a,\ b$ の近似値（たとえば，2.0, 1.0）を入力する.

C4 のセルに `=$C$2 +$C$3 * A4` を入力して，$\widehat{y}_1$ を求める．このセルを下にコピーする.

C10 のセルに $Q = \sum(y_i - \widehat{y}_i)^2$ を求めるために，`=SUMSQ($B$4:$B$9-C4:C9)` を入力し，Ctrlキーと Shift キーを押しながら Enter キーを押す[2].

表示 1.6: ソルバーによる解析（初期画面）

	A	B	C	D
1				
2			2.00	a
3	x	y	1.00	b
4	1	5	3.00	
5	3	5	5.00	
6	4	7	6.00	
7	5	6	7.00	
8	7	9	9.00	
9	10	10	12.00	
10			10.00	Q



横軸に x，縦軸に y と $\widehat{y}$ をとって散布図を描き，推定値については線で結ぶと，表示1.6右のグラフが得られる.

a, b の値を変化すると，$\widehat{y}, Q$ が再計算され，グラフも更新される．Q を最小とする $a,\ b$ を求めるのが最小2乗法である.

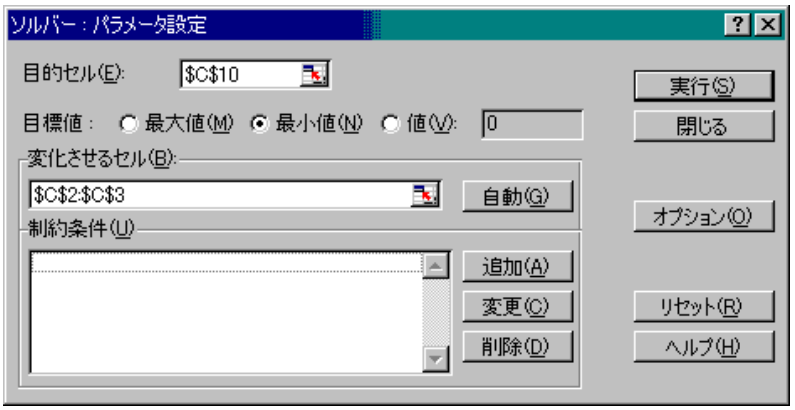
これを手作業による試行錯誤で求める代わりに，実行してくれるのがExcel のソルバーである.

トップメニューから「ツール」,「ソルバー」を選択すると，表示1.7 に示すようなパラメータ設定画面が現われる.

---

[2] 普通は，$\widehat{y}$ の右に 残差 $= y_i - \widehat{y}_i$ の列を作り，SUMSQ(残差) とする．ここでは，後の計算のために，残差の列を作らないで，SUMSQ 関数の中で残差を計算している.
出力セルが 1 個であっても，関数の中で複数のセルの演算が含まれる場合は，MMULT や LINEST 関数と同様に，Ctrlキーと Shift キーを押しながら Enter キーを押す必要がある.

表示1.7: ソルバーのパラメータ設定



目的セルに Q のセルを，目標値に「最小値」を，変化させるセルに $b,\ c$ のセルを指定する．

実行をクリックすると，$a = 3.90,\ b = 0.62,\ Q = 2.78$ が得られる．

(5)　回帰係数の標準誤差の推定

$a,\ b$ を 推定値−標準誤差，推定値，　推定値+標準誤差 の3段階に変えて，9つの組合せについての Q の値を計算した結果を表示1.8に示す．

表示1.8: $a,\ b$ と Q の関係

	F	G	H	I
2	$b \backslash a$	3.219	3.900	4.581
3	0.502	13.155	5.560	3.525
4	0.620	5.560	2.780	5.560
5	0.738	3.525	5.560	13.155

ここで，Q を計算するために，F3 のセルに

```
=SUMSQ($B$4:$B$9-(G$2+$F3*$A$4:$A$9)
```

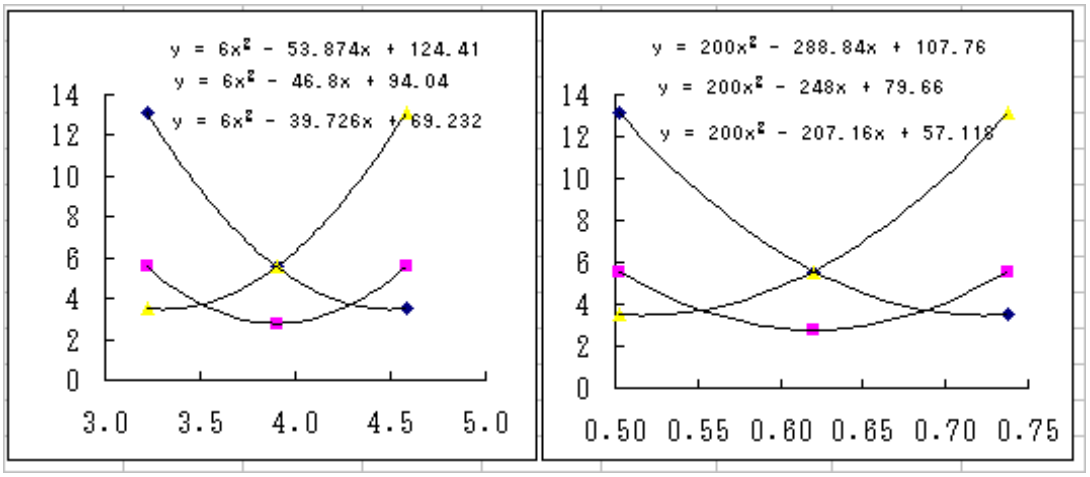
を入力して，Ctrl キーと Shift キーを押しながら，Enter キーを押す．このセルを下と右にコピーすると表示1.8が得られる．

$a,\ b$ の推定値から，a または b の何れか一方をその標準誤差だけずらしたとき $Q = 5.560$

となる．この値は $S_e + V_e = 2.780 + 0.695 = 3.475$ とは大きく異なる．$a + se(a),\ b - se(b)$ としたときの $Q = 3.525$ に近いが，一致はしない．

横または縦に並んだ３つの点に2次式を当てはめた結果が表示1.9のグラフである．

表示1.9: δ と Q の関係



3本の曲線の2次項の係数は同じで，a については 6，b については 200 である．

これらの値は，正規方程式の左辺の対角線項の係数に対応している．

6つのQに対して，$a,\ b$ の2次式をあてはめ，等高線を描いたグラフを表示1.10に示す(JMP による)．

表示1.9のグラフは，表示1.10の等高線を，水平線または垂直線で切ったときの断面図を表わすものである．

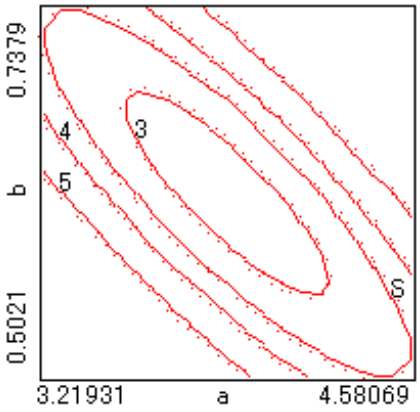
$Q = 3, 4, 5$ の等高線に $Q = 3.475$ の等高線（内から2つ目）を追加したものである．これを見ると，$Q = 3.475$ の等高線 は四辺形に内接していることが分かる．

これから，$b + se(b)$ に固定し，Q を最小になる a と，そのときの Q を求めると $S_e + V_e$ となることが分かる．

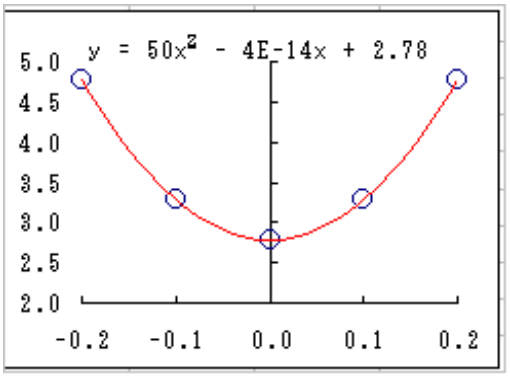
これは，標準誤差 $se(b)$ が分かっていたから可能であるが，分からないときにはどうしたら良いであろうか．

b を推定値の前後に $\delta(= \pm 1,\ \pm 2$ だけ変化して，Q を最小にする a をソルバーを使って求める．

表示1.10: Q の等高線



表示1.11: b と Q ののの関係



計算表を表示1.12に示す.

ここでは，b を変化する個数は3個で十分であるが，次章の非線形回帰に拡張するために，５段階に変化させることにした.

表示1.12のように，上の行に δ を入力し，b の行に，$b-\delta$ の値を入力する．a の行には推定値を入力する．Q の列には，表示1.8の計算で用いたのと同様の式を入力する.

表示1.12: δ と Q の関係

δ	−0.20	−0.10	0.00	0.10	0.20
b	0.82	0.72	0.62	0.52	0.42
a	2.90	3.40	3.90	4.40	4.90
Q	4.780	3.280	2.780	3.280	4.780

ここで，a の値毎にソルバーを使って b を推定するのは面倒であるので，複数のパラメータについてソルバーによる解析をまとめて実行する VBA マクロ Solv-Min を準備した.

表示1.12 で黒枠で囲まれた部分が，ソルバーとやり取りする情報の範囲である．このマクロは，一番下の行が 最小化するセル，その上のセル（ここでは1つであるが，複数でも可）は変化させるセル であると判断して解を求める．横には何個並んでいても構わない.

黒枠で囲まれた範囲を反転したのち，マクロ「Solv-Min」を実行する．表示1.12 はソルバーで解いたあとの結果である.

ソルバーを直接使うときは，最小化するセルと変化させるセルの位置はどこでも構わないが，マクロかするための標準化として，上に示した順にこれらのセルを配置しなければならない．

横軸に δ **を，縦軸に** Q **を取った散布図が表示**1.11 **である．この点に**2 **次式を当てはめる．**

$$Q = 2.78 + 50\delta^2 \tag{1.6}$$

が得られた．５つの点は2 **次曲線にぴったり乗っている．**

これから， $se(b)$ **は，**

$$Q = S_e + V_e = 2.78 + 0.695 = 3.475$$

となる δ **であるから，**

$$\begin{aligned}\delta &= \sqrt{\frac{Q - 2.78}{50}} \\ &= se(b) = \sqrt{\frac{V_e}{\text{2 次の係数}}} = \sqrt{\frac{0.695}{50}} = 0.118\end{aligned} \tag{1.7}$$

が得られる．この値は 前に得られた結果と一致している．

式(1.6) **の定数項** 2.78 **は** S_e **である．** δ^2 **の係数** 50 **は，式**(1.7) **と式**(1.5) **を対応させると，逆行列の対角要素** 0.020 **の逆数であることが分かる．**

補足　表示1.10 **の等高線は楕円である．** $\overline{x} = 0$ **のとき，円となる．** $\overline{x}/s_x$ **が大きくなると，楕円の扁平度が大きくなる（楕円の長軸と短軸の比が大きくなる）．**

楕円が極端に扁平であるとき，楕円の中心である $a,\ b$ **を無造作に四捨五入すると具合の悪いことが起こる．**

a,b **が共に正のとき，** $a,\ b$ **を共に切上げ，または，切捨てると，中心から右上，または，左下に移動するため，** Q **が大きくなる．逆に，** $a,\ b$ **の一方を切上げ，他方を切捨てると，右下，または，左上に移動するので** Q **の変化は少ない．**

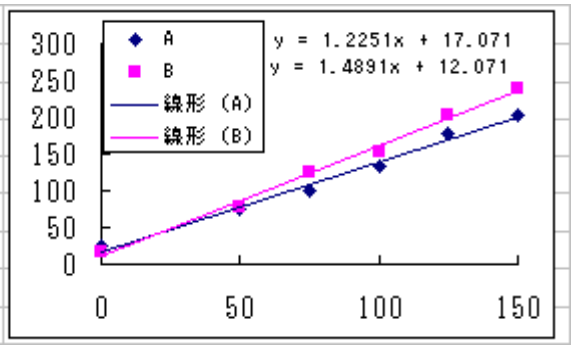
1.3　共通の切片を持つ2本の回帰直線

(1)　データと単純な解析

2種類の添加物　A, B の効果を比較するために，添加量 x を変化させて，製品の特性 y を測定したところ表示1.13が得られた.

表示1.13: データと2本の回帰直線

x	A	B
0	25	18
50	74	79
75	100	126
100	134	153
125	178	202
150	204	239



このデータに A, B 毎に直線を当てはめると,

$$y = 17.071 + 1.2251x, \quad (A)$$
$$= 12.071 + 1.4891x, \quad (B)$$

となり，切片の値 a が異なる．a は添加率が0の場合の特性値であるから，両者は一致するはずである．また，添加率 0 の値 A, B の区別がないから，添加率が 0 の観測値（太線で囲まれた2個）を A, B に分けることには意味がない.

(2)　共通の切片を持つ回帰式の推定

そこで，添加率が 0 の2個の値を区別しないで，共通の切片を持つ２つの回帰式

$$y = a + \begin{pmatrix} b_A x & (A) \\ b_B x & (B) \end{pmatrix} \tag{1.8}$$

を推定する方法を考える.

式(1.8)を

$$y = a + b_A x_A + b_B x_B \tag{1.9}$$

と変形する．ここに，x_A は．A のとき x，B のとき 0，x_B は．A のとき 0，B のとき x という変数である．

データの行列を表示1.14の左に示す．

表示1.14: LINEST 関数のためのデータと解

	x	xA	xB	y		xB	xA	1
A	0	0	0	25		1.4669	1.2474	14.5714
A	50	50	0	74		0.0475	0.0475	4.1000
A	75	75	0	100		0.9915	7.2317	#N/A
A	100	100	0	134		526.6	9	#N/A
A	125	125	0	178		55076	471	#N/A
A	150	150	0	204				
B	0	0	0	18		xA	x	1
B	50	0	50	79		-0.2196	1.4669	14.5714
B	75	0	75	126		0.0431	0.0475	4.1000
B	100	0	100	153		0.9915	7.2317	#N/A
B	125	0	125	202		526.6	9	#N/A
B	150	0	150	239		55076	471	#N/A

x_A, x_B を説明変数とする回帰式をLINEST で求めた結果を表示1.14の右上にしめす．

$$\begin{array}{rcrcrcr} y & = & 14.5714 & + & 1.2474x_1 & + & 1.4669x_2 \\ & & & & (0.0475) & & (0.0475) \end{array}$$

係数の下の括弧内には，推定値の標準誤差を示す．

A と B の傾斜の差は 0.2196 である．この値の信頼区間を求めたい．b_1, b_2 の標準誤差は LINEST 関数の出力の2行目に求められている．２つの推定値 b_1, b_2 が独立（無相関）であれば，差の標準誤差は２つの標準誤差から

$$SE(b_B - b_A) = \sqrt{SE(b_B)^2 + SE(b_A)^2} = 0.0672$$

として計算されるが，２つの推定値は添加率 0 の値を共通に用いているので独立ではないこの結果は正しくない．

正しい標準誤差を求めるためには，正規方程式の逆行列が必要である．表示1.15 に行列演算で逆行列を求めた結果を示す．

表示 1.15: 行列演算による解

12	500	500	1532	
500	56250	0	77450	
500	0	56250	89800	
1532	77450	89800	251132	
0.3214	-0.0029	-0.0029	14.5714	
-0.0029	4.317E-05	2.540E-05	1.2474	
-0.0029	2.540E-05	4.317E-05	1.4669	
			470.673	Se
			52.297	Ve

b_A, b_B の分散は，逆行列の対角要素（共に 4.317E–05）と残差分散 V_e (52.297) の積 2.258E-03 として求められる．この平方根は 0.0475 で，上に求めて標準誤差に一致する．

$b_B - b_A$ の分散を求めるためには，b_A と b_B の共分散が必要となる．これは，逆行列の非対角要素 (2.54E–05) と V_e の積 (1.328E–03) として求められる．

$b_B - b_A$ の分散は，b_B, b_A の分散の和から，共分散の 2 倍を引いて求められる．標準誤差は，分散の平方根を取り，

$$se(b_B - b_A) = \sqrt{2.54E-05 + 2.54E-05 - 2 \times 1.328E-03}$$
$$= \sqrt{1.859E-03} = 0.0431$$

となる．

(3)　別のモデルによる解析

式(1.8)を

$$y = a + bx + (b_A - b_B)x_B$$
$$= a + bx + cx_B \tag{1.10}$$
$$c = b_A - b_B \tag{1.11}$$

のように変形する．

x, x_A を説明変数として，LINEST 関数を使って解いた結果を表示 1.14 の右下に示す．

$$y = 14.5714 + 1.4669x - 0.2196x_A$$
$$(0.0457) \quad (0.0431)$$

が得られる．

この方法を用いると，傾斜の差 c の標準誤差が 0.0431 と求められる．この値は上に求めた正しい値と一致する．この例が示すように，モデル式を工夫することにより，推定値の標準誤差が簡単に求められる．

c の信頼区間は次のように計算される．

$$-2.196 - t(0.05, 9) \times 0.0431 < c < -2.196 + t(0.05, 9) \times 0.0431$$
$$-0.3171 < c < -0.1220$$

２つの傾斜間に有意差があるかどうかは，

$$t = \frac{-2.196}{0.0431} = -5.092$$

を計算し，自由度9 のｔ分布で $|t|$ が5.092 以上となる確率 p 値を求める．ｐ値は 0.0007 となり，傾斜間には高度に有意な差が認められる．

この節では，2本の傾斜の差の検定と推定を実行したが，傾斜の比が２つの添加剤の効果を比較するためには有効である．比の検定と推定については次の章で再び取り上げる．

1.4　重みつき最小２乗法（未完）

回帰分析のモデル

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

で，誤差 ε_i が正規分布 $\mathrm{N}(0,\ \sigma_i^2)$ に従う，すなわち，i によって分散の大きさが異なる（等分散性が成立しない）ときは，通常の最小2乗法ではなく，重みつき最小2乗法を用いなければならない．

重みつき最小2乗法による推定方法を説明した後に，残差平方和の分布，誤差の推定方法，得られた推定値の標準誤差の求め方などを分かりやすく説明する必要がある．

2　非線形モデルの最小２乗法

2.1　回帰式による逆推定

前章で得られた回帰式

$$y = a + bx = 3.900 + 0.620x$$

で，$y = 8$ となる x は

$$x = \frac{y - a}{b} = \frac{8 - 3.900}{0.620} = 6.613$$

として推定される．これを 逆推定 という．

この推定誤差を知り，x の信頼区間を求めたい．

(1)　従来の方法

回帰分析の教科書には，推定値 $\widehat{y}$ の標準誤差は

$$se(\widehat{y}) = \sqrt{\left(\frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}\right)\sigma^2}$$

であると書かれている．

ある x に対する y の期待値 μ の信頼区間は，σ^2 の代わりに V_e を用い，$\widehat{y}$ に $se(\widehat{y})$ の $t(f_e,\ 0.05) = 2.78$ 倍を加減して求められる．

$$\mu \sim a + bx \pm t(f_e,\ 0.05)\sqrt{\left(\frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}\right)V_e} \tag{2.1}$$

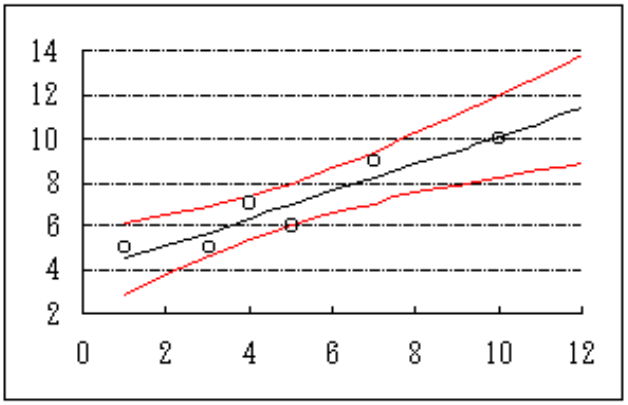
表示2.1: y の区間推定

x	$\widehat{y}$	y_L	y_U
2	5.14	3.78	6.50
4	6.38	5.38	7.38
6	7.62	6.62	8.62
8	8.86	7.50	10.22
10	10.10	8.21	11.99
6.613	**8.00**	6.92	9.08
5.088	7.05	6.11	**8.00**
9.383	9.72	**8.00**	11.44

x を変化させて信頼限界を計算すると，表示2.1 の上半分が得られる．

これをグラフ化すると表示2.2が得られる．

表示2.2: y **の区間推定**



$y = 8$ **とする** x **の信頼限界は式**(2.1) **で，** $\mu = 8$ **として解くと求められるが，２次方程式を解かねばならない．**

近似値は，表示2.2 **のグラフで，** $y = 8$ **を横切る点の** x **座標の値を読み取ることで得られる．**

より正確な値は，式(2.1) **で信頼区間を計算する表で，信頼区間が** 8 **になる** x **を** Excel **のゴールシークを使って求めたのが，表示**2.1 **の下の**3 **行である．**

y **を** 6, 7, 8, 9 **となる** x **の区間推定を求めると，表示**2.3 **が得られる．**

表示2.3: x **の区間推定**

y	x_L	$\widehat{x}$	x_U	$\widehat{x} - x_L$	$x_U - \widehat{x}$
6	0.616	3.387	4.912	2.772	1.525
7	3.206	5.000	6.795	1.794	1.795
8	5.088	6.613	9.383	1.525	2.770
9	6.507	8.226	12.438	1.719	4.213

表示2.2 **のグラフで，信頼限界の曲線は，回帰直線から上下に同じ幅で引かれているが，** x **の信頼区間は 左右が対称ではない．**

(2)　非線形回帰式の当てはめ

回帰式を

$$y = 8 + b(x - c) \tag{2.2}$$

と書き直す．ここで，$x - c$ が 0 のとき $y = 8$ になるから，c は 求めたい $y = 8$ となる x の値である．

式(2.2)のモデル式について，§1.2 単回帰式 と同様の考え方で正規方程式を導出して見よう．

$$Q = \sum_{i=1}^{n} (y_i - (8 + b(x_i - c)))^2$$

が最小になるように $b,\ c$ を決めるために，Q を $b,\ c$ で微分する．

$$\begin{aligned}
\frac{\partial Q}{\partial a} &= \sum_{i=1}^{n} 2(y_i - (8 + bx_i - cx_i)) \frac{\partial}{\partial a}(y_i - (8 + bx_i - cx_i)) \\
&= \sum_{i=1}^{n} 2(y_i - (a + bx_i - cx_i))(-x_i + c) \\
\frac{\partial Q}{\partial b} &= \sum_{i=1}^{n} 2(y_i - (8 + bx_i - cx_i)) \frac{\partial}{\partial b}(y_i - (a + bx_i)) \\
&= \sum_{i=1}^{n} 2(y_i - (8 + bx_i - cx_i))(-b)
\end{aligned}$$

これらの式に $=0$ を追加して，連立方程式を導くと，§1 の線形最小2 乗法の場合と異なり，$b,\ c$ に関して連立１次方程式とはならない．

これは，y を b または c で偏微分すると，

$$\frac{\partial y}{\partial b} = x - c, \qquad \frac{\partial y}{\partial c} = -b$$

となり，パラメータが残るためである．．したがって，通常の回帰分析のように正規方程式を立てて解いたり，LINEST関数を使って解くことはできない．

一般には「非線形回帰分析」の特殊プログラムを使って解析するが，Excel のソルバーを使えば線形回帰分析の場合と全く同様の手順で解析することができる．

表示2.4の左に示すように，$x,\ y$ の表を作成し，その右に y の推定値 $\widehat{y}$ の列を準備する．$b,\ c$ のセルに適当な初期値（たとえば，$b = 0.6,\ c = 6$) を入力する．

表示2.4: **ソルバーによる解析**

				初期値	**結果**
	K	L	M	N	
1			b	0.6	0.620
2			c	6	6.613
3		x	y	$\widehat{y}$	$\widehat{y}$
4	1	1	5	5.00	4.52
5	2	3	5	6.20	5.76
6	3	4	7	6.80	6.38
7	4	5	6	7.40	7.00
8	5	7	9	8.60	8.24
9	6	10	10	10.40	10.10
10			Q	3.760	2.780

$\widehat{y}$ の最初のセルに

```
=8+N$1*($L4-N$2)
```

の関数を入力し，下にコピーする．

Q のセル に $\sum e_i^2$ を計算する．3.760 **となる．**

Q を最少にする $b,\ c$ を求めるために，ソルバーを実行すると，表示2.4 **の右の列の解が得られる．解は，$b = 0.620,\ c = 6.613$ で，前に求めた値に一致する．**

$$y = 8 + 0.620(x - 6.613) = 3.900 + 0.620x$$

この問題をJMP **の「非線形回帰分析」を使って解いた結果を表示**2.5 **に示す．**

表示2.5: JMP **によるの解**

解

SSE	DFE	MSE	RMSE	
2.78	4	0.695	0.8336666	
パラメータ	推定値	近似標準誤差	下側信頼限界	上側信頼限界
b	0.62	0.11789826	0.34121491	0.89878509
c	6.6129032258	0.62881217	5.30751838	8.73579603

(3)　c の標準誤差

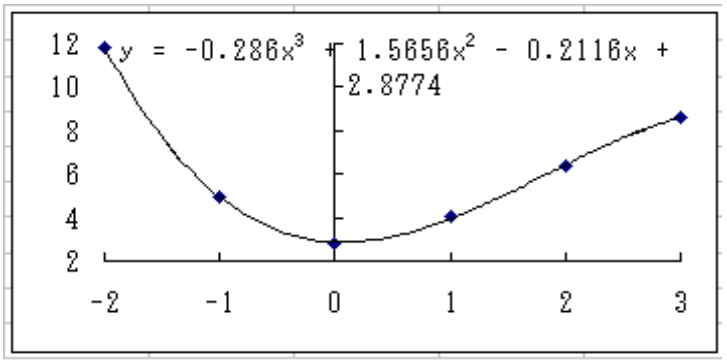
JMP の解には b, c の近似標準誤差と信頼区間が求められている．c の近似標準誤差は 0.6288 である．これを，回帰係数の標準誤差を求めたときと同様に，ソルバーを使って求める．

最小2乗推定値 c に δ を加えて固定し，Q を最小とする b を マクロ Solv-min を使って求める．

表示2.6 の上の計算表に示すように，1行目に δ を入力して，2行目の c を求める．3行目に変化させる変数 b の初期値として最小2乗推定値を入力する．4行目に Q を計算する式を入力する．

表示2.6: c と Q の関係

δ	−2	−1	0	1	2	3
c	4.613	5.613	6.613	7.613	8.613	9.613
b	0.563	0.664	0.620	0.513	0.411	0.330
Q	11.843	4.987	2.780	4.048	6.375	8.622



入力が完了したら，b と Q の行を反転させて，マクロSolv-min を実行すると，表示2.6 の上半分の計算表が得られる．

計算表の下には，横軸に δ を縦軸に Q を取った散布図に，3 次の近似曲線を当てはめたグラフを示す．

線形回帰分析の場合と異なり，左右対称の 2 次曲線ではない．3次式は，

$$Q = 2.8774 - 0.2116\delta + 1.5656\delta^2 - 0.286\delta^3 \tag{2.3}$$

となる.

§1.4 **の場合と同様に，式**(2.3) **で,**

$$Q = S_e + V_e = 2.780 + \frac{2.780}{4} = 2.780 + 0.695 = 3.475$$

となる δ **が** c **の標準誤差** $se(c)$ **となる.**

これを，ゴールシークを使って求めると

$$\delta = -0.531,\ 0.748$$

となる．2つの $|\delta|$ **の算術平均または幾何平均を計算すると，**c **の標準誤差の近似値として,**

$$se(c) = \frac{0.531 + 0.748}{2} = 0.640, \qquad = \sqrt{0.531 \times 0.748} = 0.630$$

が得られる．後者はJMP **の結果** 0.628 **に近い値である.**

もう一つの近似標準誤差は，上の3 **次式の**2**次項の係数** 1.5656 **から**

$$se(c) = \sqrt{\frac{V_e}{1.5656}} = 0.666$$

という近似値が得られる.

(4)　**c **の区間推定

式(2.3) **で,**

$$Q = S_e + F(1, f_e;\ 0.05)V_e = 2.780 + 7.709 \times 0.695 = 8.138$$

となる δ **に** $c = 6.613$ **を加えると** c **の信頼区間が得られる．ここで，**$t(f_e,\ 0.05)^2 = F(1, f_e;\ 0.05)$ **の関係を用いている.**

ゴールシークを使って求めると,

$$\delta = -1.565, \quad 2.724$$

$$c \sim (5.048,\ 9.337)$$

という信頼区間が得られる.

この値を逆推定で求めた (5.088, 9.383) **と比較すると，概ね一致する.**

JMP **の信頼区間は** (5.308, 8.736) **は上に求めた2つの区間に比べてかなり狭い.**

2.2　傾斜の比の推定

§1.3 で切片を共通とする 2 本の回帰式を求め，傾斜の差の推定と検定をした．

2 本の回帰直線を表わす式

$$y = a + \begin{pmatrix} b_A \\ b_B \end{pmatrix} x, \qquad \begin{matrix} (A) \\ (B) \end{matrix} \tag{2.4}$$

であった．

このような場合は，2 種類の添加剤の効果は，差ではなく，傾斜の比（有効率の比率）で評価されることが多い．

傾斜の比 c は，それぞれの傾斜 b から簡単に求められる．§1.3 の例では

$$c = \frac{b_A}{b_B} = \frac{1.2474}{1.4669} = 0.8503$$

である．この比率の区間推定をするためには，式 (2.4) を次のように変形する．

$$y = a + \begin{pmatrix} b_B c \\ b_B \end{pmatrix} x = a + \begin{pmatrix} b_B \times c \\ b_B \times 1 \end{pmatrix} x, \qquad \begin{matrix} (A) \\ (B) \end{matrix} \tag{2.5}$$

式 (2.5) はパラメータ $b,\ c$ の積が含まれるため，パラメータに関して線形ではない．したがって，非線形最小 2 乗法を用いる必要がある．

そのためには，工夫が必要である．

A のとき c，B のとき 1 になるような変数を生成するために，A のときのみ 1, B のときのみ 1 で，他は 0 となる変数を準備する．表示 2.7 の左では，このための変数を $c,\ 1-c$ で表わす．

$\widehat{y}$ の列を準備し，その上にパラメータ $a,\ b,\ c$ の初期値を入力する．$\widehat{y}$ の列の J6 のセルには

```
=J$2+J$3*(J$4*$G6+$H6)*$F6
```

と入力し，下にコピーする．その下の Q のセルには

```
=SUMSQ($I6:$I17-J6:J17)
```

を入力する．

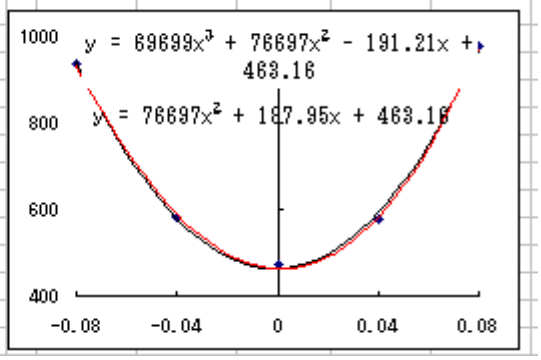
Q を最少にする $a,\ b,\ c$ をソルバーで求めた結果が表示 2.7 の左半分である．

これから，

$$y = 1.57 + \begin{pmatrix} 1.467 \times 0.850 \\ 1.467 \end{pmatrix} x, \qquad \begin{matrix} (A) \\ (B) \end{matrix}$$

表示2.7: 切片を共通とする2本の回帰式

	E	F	G	H	I	J	K	L	M	N	O	P	Q
1					δ			δ	-0.08	-0.04	0	0.04	0.08
2					a	14.57		c	0.770	0.810	0.850	0.890	0.930
3					b	1.467		a	17.17	15.62	14.57	14.00	14.00
4					c	0.850		b	1.498	1.485	1.467	1.443	1.443
5		x	c	1-c	y	y-hat		Q	935	584	471	578	976
6	A	0	1	0	25	14.6							
7	A	50	1	0	74	76.9							
8	A	75	1	0	100	108.1							
9	A	100	1	0	134	139.3							
10	A	125	1	0	178	170.5							
11	A	150	1	0	204	201.7							
12	B	0	0	1	18	14.6							
13	B	50	0	1	79	87.9							
14	B	75	0	1	126	124.6							
15	B	100	0	1	153	161.3							
16	B	125	0	1	202	197.9							
17	B	150	0	1	239	234.6							
18					Q	471							



が得られる．当然のことながら，得られた式は一致し，残差平方和 Q も同じである．

残差の平方和 Q_e は 471 で，その自由度は観測値の個数 12 から推定したパラメータの個数 3 を引いた 9 であるから，残差の平均平方 V は 52.3 となる．

これから，c の信頼率 0.05 の信頼区間は，

$$Q = S_e + F(1, f_e;\ 0.05)V_e = 471 + 5.117 \times 52.3 = 471 + 268 = 738$$

となる c として求められる．そこで，c の信頼区間を求める．

表示2.7の右半分に示すように，c の推定値の上下に0.04 刻みで c を設定し，Q を最小とする a, b を求める．

横軸に c の変化量 δ を，縦軸に Q を取って散布図を描き，2次式と3次式を当てはめると表示2.7 の右下のグラフが求められる．

ほぼ2 次式が当てはまる．

2.3　ロジスティック曲線の当てはめ

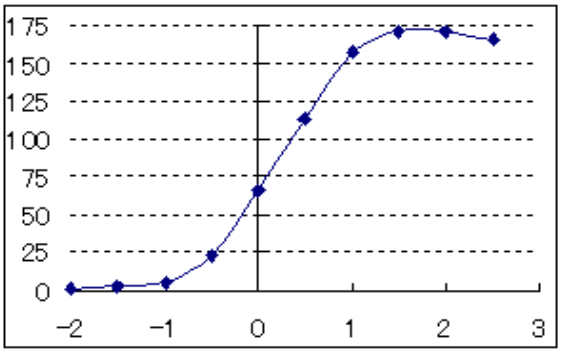
(1)　例題

ヒスタミンの投与量による平滑筋の収縮量の変化を観測した．

ヒスタミンの濃度（単位 μM）は，公比が $\sqrt{10}$ の等比級数となるように変化させた．結果は平滑筋の収縮量 (mm) である．

表示2.8: ヒスタミンによる平滑筋の収縮

ヒスタミン濃度	収縮量
0.0100	1
0.0316	3
0.100	5
0.316	23
1.00	66
3.16	113
10.0	158
31.6	171
100	171
316	165



表示2.8の右は，横軸に濃度の常用対数，縦軸に収縮量をプロットし，Excel で点を滑らかな曲線で結んだものである．

濃度の常用対数 x と収縮量 y の間にロジスティック曲線を当てはめる．

ロジスティック曲線は通常

$$y = y_{min} + \frac{y_{max} - y_{min}}{1 + e^{-(a+bx)}} \tag{2.6}$$

で表わされる．

a は $x = 0$ のときの $\ln\{\frac{y-y_{min}}{y_{max}-y}\}$ である．

この実験データでは $y_{min} = 0$ というモデルを当てはめるのが適当であろう．

このデータから，y_{max} と $y_{max}/2$ となる x を推定したい．

$y_{max}/2$ となる x を推定するためには，式(2.6)を

$$y = y_{max}\frac{1}{1 + e^{-b(x-c)}} \tag{2.7}$$

と書き換えると，c が知りたい x となる[3]．

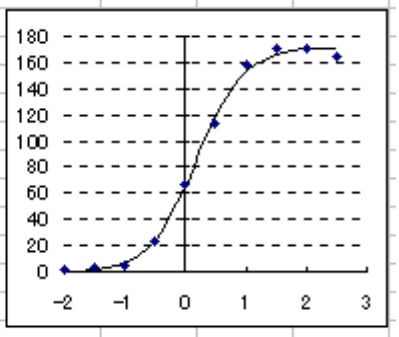
[3] $x = c$ のとき，$e^{-b(x-c)} = 1$，分母が 2 となる．

(2) Excel ソルバーによる解析

Excel のソルバーによる解析の過程を表示2.9 に示す.

表示2.9: ロジスティック曲線の当てはめ

	A	B	C	D	E
1					
2			ymax	175.00	171.58
3			b	2.00	2.689
4			c	0.300	0.201
5	x	log(x)	y	y-hat	y-hat
6	0.0100	-2.000	1	1.74	0.46
7	0.0316	-1.500	3	4.65	1.75
8	0.100	-1.000	5	12.10	6.54
9	0.316	-0.500	23	29.38	22.62
10	1.00	0.000	66	62.01	63.19
11	3.16	0.500	113	104.74	118.53
12	10.0	1.000	158	140.38	153.67
13	31.6	1.500	171	160.44	166.52
14	100	2.000	171	169.35	170.23
15	316	2.500	165	172.88	171.23
16			Se	665.22	121.10



A 列に濃度を，B 列には濃度の常用対数 x を，C 列に y を入力する．グラフから，$y_{max} \simeq 175,\ c \simeq 0.3$ 前後と予想される．b の初期値は 2 とする.

これらの値を初期値として D2:D4 に入力する.

$\widehat{y}_1$ の D6 には =D$2/(1+EXP(-D$3*($B6-D$4))) が,

S_e のD16 には =SUMSQ($C$6:$C$15-D6:D15) が入力されている.

Q を最小とする $y_{max},\ b,\ c$ をソルバーで求めた結果が E 列に示されている.

$y = y_{max}/2$ となる濃度は $10^c = 10^{0.201} = 1.59$ となる.

(3) 推定値 c の標準誤差

JMP で解析した結果を表示2.10 に示す.

当然のことながら，ソルバーと同じ推定値が得られている.

c の近似標準誤差と信頼区間が求められている.

これを，ソルバーで求めて見よう.

表示2.10: JMP による解析結果

解

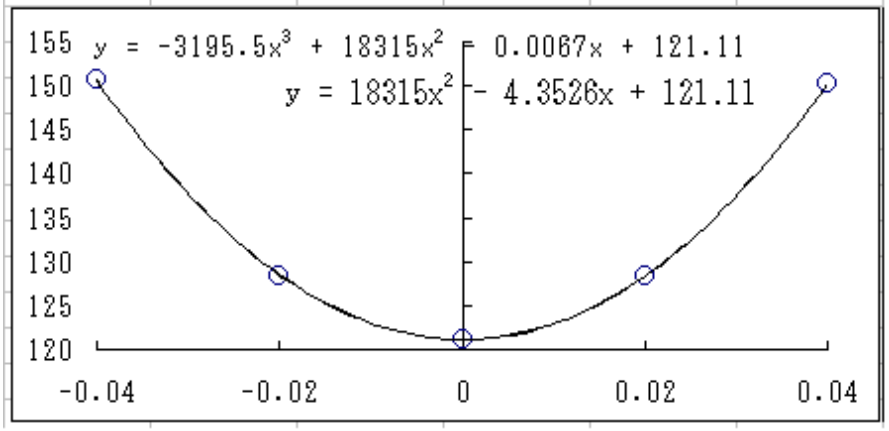
SSE	DFE	MSE	RMSE
121.39418906	7	17.342027	4.1643759

パラメータ	推定値	近似標準誤差	下側信頼限界	上側信頼限界
ymax	171.58076776	2.68479989	165.512476	177.961259
b	−2.688601247	0.18931033	−3.1554709	−2.3076803
c	0.2008256102	0.03101341	0.12840759	0.2741942

前節と同様に，1行目に c の変化量 δ を入力し，2行目に c を求める．3,4行目に変化させるパラメータ y_{max}, b の初期値を入力する．5行目に残差平方和 Q を求める式を入力する．

表示 2.11: ソルバーによる標準誤差の推定

δ	-0.04	-0.02	0.00	0.02	0.04
c	0.161	0.181	0.201	0.221	0.241
ymax	169.76	170.66	171.58	172.52	173.50
b	2.737	2.715	2.689	2.660	2.628
Q	150.62	128.46	121.10	128.41	150.21



Q を最小とする y_{max}, b をマクロ Solv-min で求める．

横軸に δ を，縦軸に Q を取って散布図を描き，「近似曲線の追加」で，2次式と3次式を当てはめる．

2 次式と 3 次式とは 2 本の曲線はかなり接近している．

3 次式

$$Q = 121.11 - 0.0067\delta + 18315\delta^2 - 3195.5\delta^3$$

で，$Q = S_e + V_e = 121.10 + 17.30 = 138.40$ となる δ をゴールシークで求めると，0.0308，−0.0306 となる．両者の絶対値の平均値 0.0307 は，JMP で得られた近似標準誤差 0.03097 に近い値である．

同様に，$Q = S_e + F(1, 7; 0.05)V_e = 217.84$ となる δ は，–0.0722, 0.0731 で，c の 95%信頼区間は (0.1288, 0.2741) となる．この結果は JMP の解（0.1284, 0.2742) と良く一致する．

別の方法として，2 次項の係数 18315 から，

$$se(c) = \sqrt{V_e/\text{係数}} = \sqrt{17.30/18315} = 0.0307$$

$$c \sim 0.201t(7, 0.05) \times 0.0307 = (0.128, 0.278)$$

という近似値を求めることができる．

(4)　特殊な場合の注意

水準の幅が広いとき，2 つの水準の間で y が急激に変化する場合がある．たとえば，表示 2.12 の左に示すようなデータが得られたとする．

このデータに，上に述べた手順で $y_{max} = 100$ のロジスティック曲線を当てはめると，表示 2.12 の結果が得られる．データと推定された曲線をグラフ化すると，表示 2.12 の右のように，まずまずの結果が得られる．

このグラフから，$y = 50$ を与える x, $D50$ は，3 番目と 4 番目の水準の間にあると予想される．

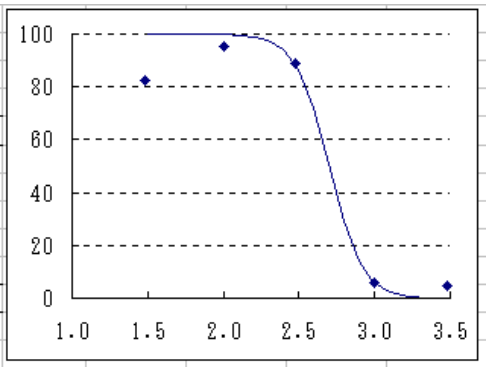
ここで，D50 の信頼区間を求めることを考える．

D50 を推定値の前後に変化させて，Q の値の変化を調べる．

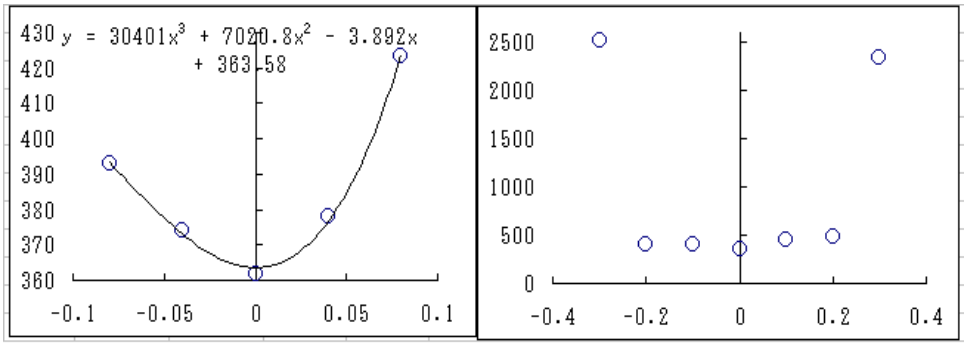
変化の範囲を狭く取ると，表示 2.13 の左のようなグラフが得られる．左右が完全に対称であるとは言えないが，3 次式で十分に近似できる曲線である．多くの統計解析プログラムは，この曲線の曲率から D50 の推定値の標準誤差を求めて，D50 の信頼区間を求めている．

表示2.12: 特殊な例と最小2 乗解

		ymax	100.00
		b	-9.19
		c	2.704
濃度	x	y	y-hat
30	1.477	82.2	100.00
100	2.000	95.2	99.85
300	2.477	89.1	88.92
1000	3.000	5.9	6.15
3000	3.477	4.9	0.08
		Q	361.69



表示2.13: D50 を変化させたときの Q の変化



しかし，変化の範囲を広く取ると，表示2.13 の右のグラフが得られる．ここではもはや，多項式では近似できない曲線となっている．これは，D50 が第3 水準と第4 水準の外側にあるということは極めて不自然であることを表わしている．

したがって，このような場合は，数理統計学で取り扱われる範囲外であって，既存の統計解析プログラムを使うととんでもない結果が得られることになる．

3 最尤法

3.1 確率と尤度

(1) 2項分布

2項分布の確率は，次の式で表される．

$$prob(r) = {}_nC_r\pi^r(1-\pi)^{n-r} \tag{3.1}$$

$n = 10$ について，横に π を，縦に r を取った表に式(3.1)の計算値を求めると表示3.1が得られる．

表示3.1: 2項分布の確率と尤度

n	10																		
	π																		
r	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
0	.60	.35	.20	.11	.06	.03	.01	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
1	.32	.39	.35	.27	.19	.12	.07	.04	.02	.01	.00	.00	.00	.00	.00	.00	.00	.00	.00
2	.07	.19	.28	.30	.28	.23	.18	.12	.08	.04	.02	.01	.00	.00	.00	.00	.00	.00	.00
3	.01	.06	.13	.20	.25	.27	.25	.21	.17	.12	.07	.04	.02	.01	.00	.00	.00	.00	.00
4	.00	.01	.04	.09	.15	.20	.24	.25	.24	.21	.16	.11	.07	.04	.02	.01	.00	.00	.00
5	.00	.00	.01	.03	.06	.10	.15	.20	.23	.25	.23	.20	.15	.10	.06	.03	.01	.00	.00
6	.00	.00	.00	.01	.02	.04	.07	.11	.16	.21	.24	.25	.24	.20	.15	.09	.04	.01	.00
7	.00	.00	.00	.00	.00	.01	.02	.04	.07	.12	.17	.21	.25	.27	.25	.20	.13	.06	.01
8	.00	.00	.00	.00	.00	.00	.00	.01	.02	.04	.08	.12	.18	.23	.28	.30	.28	.19	.07
9	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.02	.04	.07	.12	.19	.27	.35	.39	.32
10	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.01	.03	.06	.11	.20	.35	.60

式(3.1)は，π, n, r の関数である．n, π を固定して r の値によってどのように変化するかを考え，左辺を $prob(r)$ とした．すなわち，r が得られる確率を表すものと考えた．

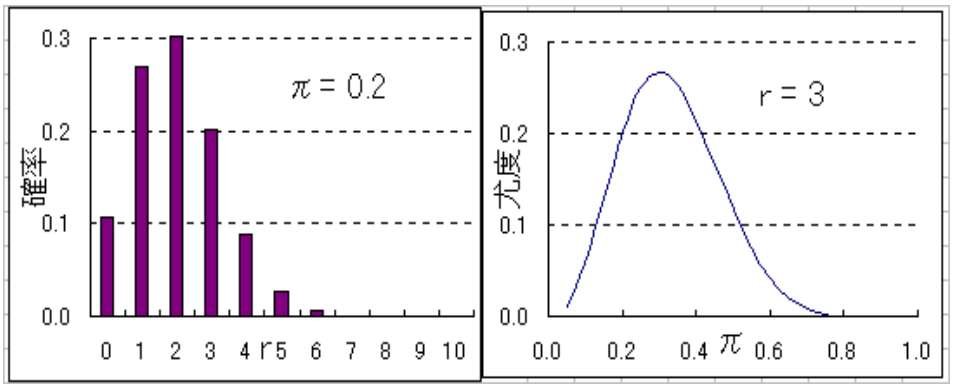
表示3.1を縦に見ると，2項分布の確率である．$\pi = 0.2$ のときのグラフを表示3.2の左に示す．

表示3.1を横に見て，横軸に π を取ってグラフ化したのが，表示3.2の右である．

n, r を固定して π によってどのように変化するかを見たものと考えると，式(3.1)の左辺は $L(\pi)$ と書くことができる．

$$L(\pi) = {}_nC_r\pi^r(1-\pi)^{n-r} \tag{3.2}$$

表示3.2: 確率と尤度グラフ



式の右辺は式(3.1)と同じであるが，左辺の括弧の中が r ではなく，π になっている．r という値が得られる確率が π によってどう変化するかを表している．

このとき，π である尤もらしさという意味で，尤度 Likelihood という．通常は，自然対数を取って，対数尤度と呼ぶ．

表示3.2で，確率の値は不連続な値 r によって変化するので，左のように，棒グラフで表され，尤度の値は連続な値 π によって変化するので，右のように，滑らかな曲線で表される．

(2)　χ 2乗分布

母分散が σ^2 の母集団から得られた $n = 11$ のサンプルの平方和を S とすると，

$$\chi^2 = \frac{S}{\sigma^2}$$

は自由度が $f = n - 1 = 10$ のχ 2乗分布に従う．

$\sigma^2 = 1.6,\ 1.8,\ 2.0,\ 2.2$ の場合の S の分布を表示3.3の左に示す．

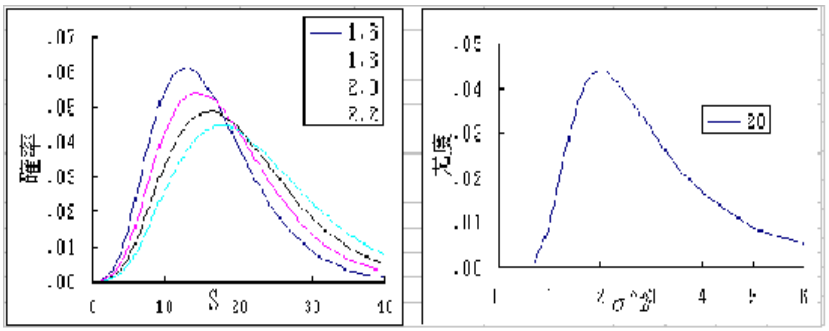
S の分布は，χ 2乗分布の，横軸を σ^2 倍に，縦軸を $1/\sigma^2$ 倍にしたものである．

いま，$S = 20$ が得られたとき，$S = 20$ に縦線を入れ，4本の曲線との交点を読むと，これが，$S = 20$ に対する尤度である．

横軸にσ^2 を取り，縦軸に尤度を取ると，表示3.3の右のグラフが得られる．

表示3.2は不連続分布について，表示3.3は連続分布について，確率と尤度の関係を表わすグラフである．

表示3.3: S の確率分布と尤度



3.2　2項分布の π の推定

(1)　点推定

$n,\ r$ から π を推定するためには，通常，不偏推定量として，$\hat{\pi} = r/n$ が用いられる．この推定値の期待値が π になることを利用した推定である．

対数尤度

$$\ln L = \ln({}_nC_r) + r\ln\pi + (n-r)\ln(1-\pi)$$

が最大になるπ の値を π の推定値とする方法を，最尤推定 という．上の式をπ で偏微分して0 と置くと，

$$\frac{\partial \ln L}{\partial \pi} = \frac{r}{\pi} - \frac{n-r}{1-\pi} = \frac{(1-\pi)r - \pi(n-r)}{\pi(1-\pi)} = \frac{r-\pi n}{\pi(1-\pi)} = 0$$

$$\pi = \frac{r}{n}$$

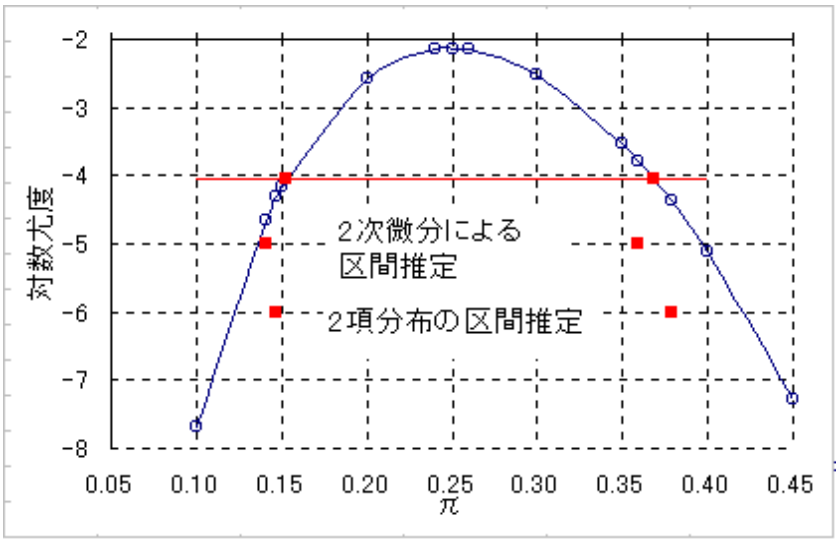
となり，この場合は，不偏推定値に一致する．

表示3.2の右の曲線を見ると，$\pi = 0.3$ で最大になることが分かる．

$n = 60,\ r = 15$ の場合について，$\pi = 0.1 \sim 045$ についての対数尤度を計算して，グラフを描くと表示3.4が得られる．

尤度は確率の見方を変えたものであるから，二項分布の確率を計算する関数を使って簡単に計算することができる．

表示3.4: π と対数尤度との関係




```
=LN(BINOMDIST($C$5,$B$5,D5,FALSE))
```

とすれば良い．BINOMDIST 関数の中のパラメータは左から，r, n, $\hat{\pi}$ で最後の FALSE は確率を計算することを指定する．

最尤推定値を頂点として2次曲線に近い形を取るが，厳密には左右対称ではない．

(2)　区間推定

最尤推定値 $\hat{\pi}$ の対数尤度と帰無仮説 π の対数尤度との差の2倍は，近似的に，自由度1のカイ2乗分布に従う．この性質を使って，π の信頼区間を求めることができる．

表示3.4のグラフで，曲線が対数尤度の最大値から3.841(自由度1のカイ2乗分布の上側5%点 $= 1.96^2$ 標準正規分布の両側5%点の2乗）の 半分 だけ低い位置（-4.056，横線で示す）を横切るπ が信頼区間となり，

$$0.152 \leq \pi \leq 0.369$$

が得られる．

このような曲線から信頼区間を求める代わりに，最尤推定値付近での曲線の曲率を求め，曲線が2次曲線であるという近似の下で，信頼区間を求めることができる．

2次の曲率は， $\pi = 0.24,\ 0.25,\ 0.26$ **に対する対数尤度，** —2.151, –2.135, –2.151 **から**

$$\frac{-2.135 - (-2.151 - 2.151)/2}{(0.25 - 0.24)^2} = 160$$

として求められる．これから， π **の信頼区間は**

$$\pi = \widehat{\pi} \pm \sqrt{3.841/(2 \times 160)} = \widehat{\pi} \pm 1.96/\sqrt{2 \times 160}$$

$$0.140 \leq \pi \leq 0.360$$

となる．この信頼区間は，最尤推定値の両側に等距離となっている．また，2次の曲率の逆数は最尤推定値の分散の近似値である．

2項分布の分布関数を使って， π **の正確な信頼区間を求めると，**

$$0.147 \leq \pi \leq 0.379$$

となる．

これらの3つの信頼区間は表示3.4の中に記されている．

(3) 補足: 最小2乗推定と最尤推定

未完

誤差が正規分布に従うとき，最尤推定は最小2乗推定になることを説明する．

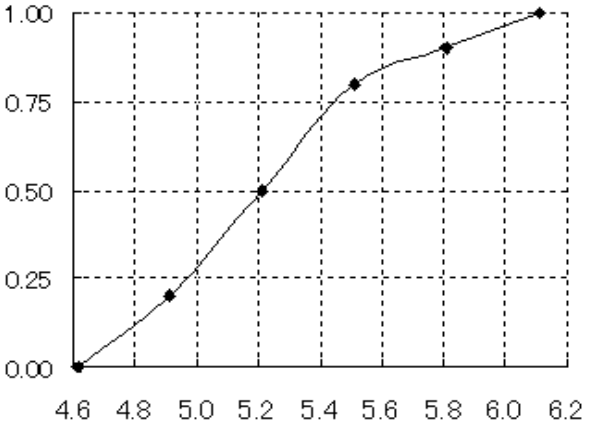
3.3　ロジスティック回帰分析

(1)　データとモデル

用量を6段階に変えて，10匹の実験動物に投与した結果，表示3.5 のデータ[4]が得られた.

表示3.5: ロジスティック回帰分析用データ

用量	個体数	死亡	死亡率
101	10	0	0.0
136	10	2	0.2
183	10	5	0.5
247	10	8	0.8
333	10	9	0.9
450	10	10	1.0



表示3.5の右には，横軸に用量の自然対数を，縦軸に死亡率を取って実験値をプロットし，滑らかな線で結んだものである.

用量の自然対数を x，個体数を n，死亡数を r，死亡率を $p = r/n$ で表わす.

p は §2.2 ロジスティック曲線 と同様にS字状に変化する.

式(2.7)で，$y_{max} = 1$ とすると，

$$p = \frac{1}{1 + e^{-b(x-c)}} \tag{3.3}$$

となる.

§2.2 では，式 (2.7) に正規分布に従う誤差が加わったものと仮定され，最小2乗法により解析した.

しかし，p は2項分布に従うので，最尤法を用いて解析する[5].

---

[4] ピンク本のp.230 に記載されている．ただし，そこでは，プロビット分析のみが適用されている.

[5] 2項分布の誤差を仮定して，重みつき最小2乗法を反復して解く方法もある.

(2) 最尤法によるロジスティック回帰分析

§3.2 では**1つの** p **から未知のパラメータ** π **を最尤法で推定した.**

今度は, 6 **個の** p_i **から未知のパラメータ** $b,\ c$ **を推定する.**

まず, パラメータ $b,\ c$ **の初期値を与え,** $\widehat{p}_i$ **を計算する. それぞれについて対数尤度を求め, その合計が最大になる** $b,\ c$ **を求める.**

これをExcel **のソルバーで求めるためには, 表示**3.6 **のような計算表を準備する.**

表示3.6: Excel **によるロジスティック回帰分析**

	A	B	C	D	E	F	G
2	**用量**	x	r	n	p	$\widehat{p}$	L
3	101	4.615	0	10	0.00	0.042	–0.434
4	136	4.913	2	10	0.20	0.164	–1.241
5	183	5.209	5	10	0.50	0.465	–1.427
6	247	5.509	8	10	0.80	0.796	–1.198
7	333	5.808	9	10	0.90	0.946	–1.112
8	450	6.109	10	10	1.00	0.987	–0.127
9					b	5.004	
10					c	5.238	
11					$\sum L$		–5.539

`F9:F10` **に** $b,\ c$ **の初期値を入力する.**

`F3` **に**

```
=1/(1+exp(-F$9*($B3-F$10))
```

を入力して, $\widehat{p}_1$ **を求め, 下にコピーする.**

`G3` **に**

```
=LN(BINOMDIST($C3,$D3,F3,FALSE))
```

を入力して, p_1 **に対する対数尤度** L_1 **を求め, 下にコピーする.**

`G11` **に対数尤度の合計を求める.**

ソルバーを起動し,「目的セル」を対数尤度の合計に,「目的」を「最小値」に,「変化させるセル」を $b,\ c$ **に指定して, 実行する.**

その結果が表示3.6 **の右の部分である.**

これから, $p = 0.5$ **となる** x **の値** c **は** 5.238 **で, 用量に戻すと** $e^{5.238} = 188$ **が得られる.**

(3)　c の信頼区間

前項で得られた c の95% 信頼区間を求める.

c の推定値 5.238 を中心とし，$\pm 0.1,\ \pm 0.2$ で5段階に変化させて設定し，対数尤度の合計 L を最大にする b を求める．計算の過程を表示3.7に示す.

表示3.7: c と対数尤度の関係

	I	J	K	L	M	N
2	**δ**	−0.200	−0.100	0.000	0.100	0.200
3		0.156	0.079	0.042	0.030	0.030
4		0.378	0.258	0.164	0.115	0.098
5		0.665	0.583	0.465	0.351	0.276
6		0.868	0.851	0.796	0.695	0.575
7		0.956	0.959	0.946	0.906	0.827
8		0.986	0.990	0.987	0.976	0.945
9	**b**	3.987	4.695	5.004	4.808	4.229
19	c	5.038	5.138	5.238	5.338	5.438
11	L	–8.35	–6.32	–5.54	–6.34	–8.55

表示3.7は次の手順で作られた.

- **表示3.6のF列（$\widehat{p}$ と b, c）をコピーする.**
- **一番上の行にδ を入力し，c の行には，最尤法で求めた値に δ を加えた値を求める.**
- **対数尤度 L の行には，$\widehat{p}$ から直接 対数尤度の合計 を求める関数**

```
=SUM(LN(BINOMDIST($C$3:$C$8,$D$3:$D$8,J3:J8,FALSE)))
```

を入力し，Ctrl キーと Shift キーを押しながらEnter キーを押す（ここに，J は$\delta = -0.200$ の列である）.

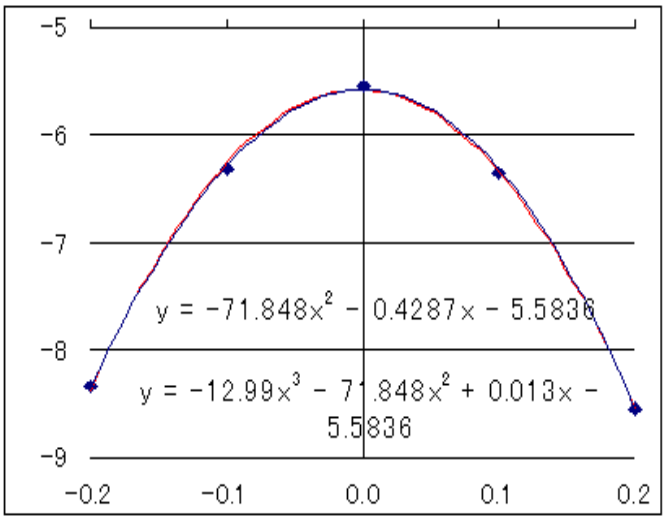
- **1列ずつ,「目的セル」に L を,「変化させるセル」に b （c を除くことに注意）を指定して，実行する.**

表示3.8に尤度の変化のグラフを示す.

「近似曲線の追加」を使ってこの点に2次式または3次式を当てはめた.

$$L = -5.5836 - 0.4287\delta - 71.848\delta^2$$
$$= -5.5836 + 0.0130\delta - 71.848\delta^2 - 12.99\delta^3$$

表示3.8: c **と対数尤度の関係**



3 **次式で，**L **が** $-5.539 - \chi^2(1, 0.05)/2 = -7.460$ **となる**δ **を，ゴールシークを使って求めると，表示**3.9 **が得られる．**

表示3.9: c **の信頼区間**

係数	−5.584	0.013	−71.85	-12.99	
c	1	δ	δ^2	δ^3	L
5.074	1	**−0.164**	0.0269	−0.004	**−7.460**
5.397	1	**0.159**	0.0254	0.004	**−7.460**

これから，

$$5.074 \leq c \leq 5.397$$

が得られる．

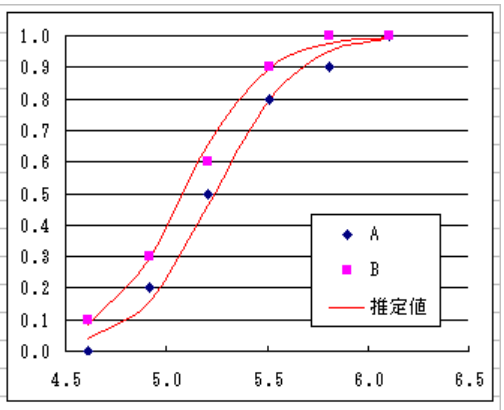
(4) D50 の差の推定と検定

§3.3 (1) **のような実験を** 2 **つの薬剤について実施し，表示**3.10 **の左に示すようなデータが得られた．**

このデータに，

表示3.10: 2つの薬剤についてのデータと解析結果

dose	Gr	x	r	n	p	p-hat	p-hat
101	0	4.62	0	10	0.0	0.040	0.062
136	0	4.91	2	10	0.2	0.159	0.225
183	0	5.21	5	10	0.5	0.464	0.560
247	0	5.51	8	10	0.8	0.800	0.850
333	0	5.81	9	10	0.9	0.949	0.962
450	0	6.11	10	10	1.0	0.989	0.991
101	1	4.62	1	10	0.1	0.083	0.062
136	1	4.91	3	10	0.3	0.293	0.225
183	1	5.21	6	10	0.6	0.655	0.560
247	1	5.51	9	10	0.9	0.898	0.850
333	1	5.81	10	10	1.0	0.976	0.962
450	1	6.11	10	10	1.0	0.995	0.991
					b	5.119	4.989
					c	5.238	5.161
					d	-0.154	
					L	-10.52	-11.50



$$p = \begin{array}{l} \frac{1}{1+e^{b(x-c)}} \quad (A) \\ \frac{1}{1+e^{b(x-(c+d))}} \quad (B) \end{array} \tag{3.4}$$

というモデルを当てはめる.

ここに，A と B の $D50$ は c, $c+d$ である．すなわち，d は $D50$ の差に相当する.

d を含むモデルの $\widehat{p}$ を求める式は,

```
=1/(1+EXP(-G$15*($C3-(G$16+G$17*B3))))
```

である．B 列には，薬剤が A のとき 0，B のとき 1 となる変数が準備されている．このような変数は ダミー変数 と呼ばれる.

このモデルの対数尤度 L は -10.52 である.

同じデータに，帰無仮説：D50 が等しい というモデルを当てはめると，対数尤度は -11.50 となる.

両者の差の2倍 $2\times(-10.52-(-11.50))=1.94$ は，帰無仮説が正しいとき，自由度が1のカイ2乗分布に従う.

1.94 以上の確率（p値）は 0.1634 となり，この2本のロジスティック曲線の間には有意な差が認められない.