第 9 回 続高橋セミナー
最尤法によるポアソン回帰分析入門
2020 年 7 月 13 日

第 12 章　パラメータの共分散行列の活用

パラメータの共分散行列は，ポアソン回帰のみならず通常の回帰分析における種々の推定値に対する 95%信頼区間を求めるために不可欠な統計量であることを繰返し示してきた．一般的に共分散行列といった場合には，「データの共分散行列」を意味するので，「パラメータの共分散行列」のように明確に区別する必要がある．本章では，最初にデータの共分散行列について Excel による行列計算の入門として取り上げ，パラメータの共分散行列との違いを明確にする．次に，重回帰分析の伝統的な「偏差平方和行列をベース」について Excel の行列計算を用いてチャレンジすし，デザイン行列をベースにした重回帰分析につなげる．パラメータの共分散行列の活用事例として「2 次曲線の 95%信頼区間」を取り上げる．

第 12 章　　目 次

続く

連絡先:高橋 行雄, 〒105-0014 東京都 港区 芝 1―12－3－1005

takahashi.stat@nifty.com ， FAX : 03-342-8035

第 9 回 続高橋セミナー
最尤法によるポアソン回帰分析入門

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

目 次

12. パラメータの共分散行列の活用

パラメータの共分散行列は，ポアソン回帰のみならず通常の回帰分析における種々の推定値に対する 95%信頼区間を求めるために不可欠な統計量であることを繰返し示してきた．一般的に共分散行列といった場合には，「データの共分散行列」を意味するので，「パラメータの共分散行列」のように明確に区別する必要がある．本章では，最初にデータの共分散行列について Excel による行列計算の入門として取り上げ，パラメータの共分散行列との違いを明確にする．次に，重回帰分析の伝統的な「偏差平方和行列をベース」について Excel の行列計算を用いてチャレンジすし，デザイン行列をベースにした重回帰分析につなげる．パラメータの共分散行列の活用事例として「2 次曲線の 95%信頼区間」を取り上げる．

12.1. データの共分散行列・パラメータの共分散行列

共分散は，身近な統計量であり 2 つの変数の相関係数を算出する過程で使われており，多変量データの関連を概観するための相関行列を算出する際にも共分散行列が使われている．共分散分析も良く知られた統計的方法であるが，質的変数に対する 1 元配置分散分析に際し「共変量」についての回帰分析を併合した解析として知られていて，「共分散」を用いた解析ではない．共分散分析については，第 13 章で取り上げるので，ここでは取り扱わない．

第 4 章では，デザイン行列を用いた回帰分析で回帰パラメータの標準誤差を求める際，「パラメータの共分散行列」の対角要素が，パラメータの分散となるので，その平方根が標準誤差として求めることを示した．また，回帰直線の 95%信頼区間および予測区間（個別データの信頼区間）を求める際に，一般的にシグマが用いられている計算式に代え，「パラメータの共分散行列」を活用する方法も示してきた．

多変量ポアソン回帰を Excel によるニュートン・ラフソン法を用いた解析，および，反復重み付き回帰を用いた解析に際し，計算過程の中で「パラメータの共分散行列」を当然のごとく使い，各種の推定値の 95%信頼区間の計算に際しても，「パラメータの共分散行列」を使ってきた．「パラメータの共分散行列」は，多変量ポアソン回帰のみならず，通常の回帰でも重回帰でも共通の存在であるが，「パラメータの共分散行列」は，日陰の存在であり続けている．

以前より，JMP の「モデルあてはめ」による重回帰分析では，パラメータ間の相関行列を出力することができても，その元となるパラメータの共分散行列が出力できないない状態が現在でも続いている．他方,「モデルのあてはめ」でポアソン回帰を行う一般化線形モデルでは，パラメータの共分散行列および相関行列の出力がサポートされている．なぜなのだろうか．多くのユーザが必要としない機能に対し，統計ソフト・ベンダーが対応していない状況と思われる．

各種の分散分析モデルおよび重回帰分析は，ほとんどすべて一般線形モデルの応用によって解決でき, 長年にわたり統計ソフトのバージョンアップが行われてきた. その結果として，多くのユーザの必要性に答えてきたことが,「パラメータの共分散行列」を用いて解析したいと思うような事例をほとんど網羅してきたのだろうか．第 12.5 節で 2 次式の 95%信頼区間を描く事例を取り上げるが，計算方法の例示が全く見いだせない．「パラメータの共分散行列」を使って Excel で簡単に解決する方法を示す．

文献事例を使った探索的ポアソン回帰をこれまで示したきたのであるが，JMP の一般化線形モデルにおけるポアソン回帰の出力は，回帰パラメータの推定に限定されており，Excel を用いて追加の推定を行う必要があった．その際に，パラメータの共分散行列が，中心的な役割を果すことを実感した．JMP の「予測プロファイル」は，これまでの統計ソフトにはない，素朴ではあるが画期的な出力であり, Excel でパラメータの共分散行列を用いて再現することを通じ，これまで私も真剣に向き合わなかった各種の課題に対し，丁寧な解説を試みる切っ掛けとなった．

回帰分析のパラメータの標準誤差を求める際に，それらの分散を計算する場合にシグマを用いた計算式が示されることが普通であるが，それらをパラメータの共分散行列としてまとめて扱っている成書に遭遇することはまれである．実際，パラメータの共分散行列が得られたとしても，それらを活用した種々の推定を行うためには，行列計算を前提にする必要があるために，意図的に避けているとも思われる．もちろん私も自分自身の理解を深める為に統計ソフトに付随する行列計算言語による解析を行ってきたのであるが，それを推奨することが難儀であることを実感してきた．

共分散行列といえばデータの共分散行列がメジャーであり，パラメータの共分散行列といっても「データの共分散行列のこと？」と思われるに違いない．パラメータの共分散行列を理解し活用するためには，データの共分散行列について Excel の行列関数を使った計算方法を知ることも必要である．データの共分散行列が得られれば，相関行列も定義に従って，行列計算により簡単に求めることができる．

Excel の行列関数を使って統計計算を行う入門としては，2 変数の相関係数の算出は，統計の基礎知識でもあり，Excel の行列計算でなくともシグマ的な計算でも，相関係数を求める Correl() 関数でも容易に求めることができ，多変数の相関行列を作成もイメージしやすい．

多変量データについての相関行列の算出は，データの共分散行列をベースにしているので Excel の行列関数を使った計算方法の入門に適している．これらの行列計算は，偏差平方和ベースの重回帰分析の基礎であり，また，デザイン行列ベースでの重回帰分析の基礎でもある．さらに，反復重み付き回帰によるポアソン回帰への拡張に対しても必須の知識でもある．

第 4 章は，Excel によるデザイン行列を用いた回帰分析の入門としたが，相関行列の作成は，逆行列が含まれないので，回帰分析よりも行列計算の入門として適していると思われる．

12.2.　アイリスデータの共分散行列および相関行列

多変量データとして，表 12.1 に示すようにフィッシャーのアイリスデータからバーシカラー種の 50 個のデータを抜き出して用いる．このアイリスのデータは，多変量解析の代表的な事例であり，Web 上で沢山の解説記事が見いだされ，データを手軽にダウンロードすることができる．

どんな統計ソフトでも，多変量の共分散行列および相関行列の計算は標準的にサポートされているので，実用的には Excel で計算する必要性は全くない．しかし，理論を学習し応用力を養うためには，Excel の行列計算などにより，各種の統計計算を実際に行う経験を積むことが，理論を確実なものにすると期待される．多変量データの相関行列をいかにスマートに計算するかは，行列計算の最初の課題として適している．

Excel の行列関数を用いた相関行列の算出

表 12.1 に示すデータは，50 行 4 列データの矩形上の集まりであり，行列 $\boldsymbol{X}$ とする．行列 $\boldsymbol{X}$ の 1 列目を列ベクトル $\boldsymbol{X}_1$（50 行×1 列）とし，順次 $\boldsymbol{X}_2$，$\boldsymbol{X}_3$，$\boldsymbol{X}_4$ とし，行方向は，行ベクトル $\boldsymbol{x}_1$（1 行×4 列）とし，順次 $\boldsymbol{x}_2, \ldots, \boldsymbol{x}_{50}$ とする．

表 12.1　アイリスデータのバーシカラー種の相関行列の計算

	種類: versicolor								
i	がくの長さ x_1	がくの幅 x_2	花弁の長さ x_3	花弁の幅 x_4		がくの長さ x_1	がくの幅 x_2	花弁の長さ x_3	花弁の幅 x_4
1	7.0	3.2	4.7	1.4	平均 $\boldsymbol{x}^{-}$=	5.9360	2.7700	4.2600	1.3260
2	6.4	3.2	4.5	1.5					
3	6.9	3.1	4.9	1.5	共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$=	**0.2664**	0.0852	0.1829	0.0558
4	5.5	2.3	4.0	1.3		0.0852	**0.0985**	0.0827	0.0412
5	6.5	2.8	4.6	1.5		0.1829	0.0827	**0.2208**	0.0731
6	5.7	2.8	4.5	1.3		0.0558	0.0412	0.0731	**0.0391**
7	6.3	3.3	4.7	1.6					
8	4.9	2.4	3.3	1.0	分散 $\boldsymbol{\sigma}^2$=	0.2664	0.0985	0.2208	0.0391
9	6.6	2.9	4.6	1.3					
10	5.2	2.7	3.9	1.4	相関行列 $\boldsymbol{R}(\boldsymbol{x})$=	**1**	0.5259	0.7540	0.5465
11	5.0	2.0	3.5	1.0		0.5259	**1**	0.5605	0.6640
12	5.9	3.0	4.2	1.5		0.7540	0.5605	**1**	0.7867
13	6.0	2.2	4.0	1.0		0.5465	0.6640	0.7867	**1**
:									
48	6.2	2.9	4.3	1.3					
49	5.1	2.5	3.0	1.1					
50	5.7	2.8	4.1	1.3					

手順 1） $\boldsymbol{X}_1$ の平均を $\bar{x}_1$ = Average($\boldsymbol{X}_1$の範囲）により計算し，右方向にフィルハンドルで計算式をコピーし，4 個の平均をベクトル $\bar{\boldsymbol{x}}$ する．

	がくの長さ x_1	がくの幅 x_2	花弁の長さ x_3	花弁の幅 x_4
平均 $\bar{\boldsymbol{x}}$ =	5.9360	2.7700	4.2600	1.3260

手順 2）偏差は［($\boldsymbol{X}$の範囲)－($\bar{\boldsymbol{x}}$の範囲)］の行列の引き算として計算する．行列計算では，50×4 の行列と 1×4 のベクトルとの差は．50×4 の行列となり，平均からの偏差が計算される．行列の積 Mmult() 関数で一気に 4×4 のデータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$ を作成する．なお，49 は，自由度である

$\boldsymbol{\Sigma}(\boldsymbol{x})$=Mmult(Transpose(($\boldsymbol{X}$の範囲)－($\bar{\boldsymbol{x}}$の範囲))，(($\boldsymbol{X}$の範囲)－($\bar{\boldsymbol{x}}$の範囲)))/49

	Trannspose(($\boldsymbol{X}$の範囲) - ($\boldsymbol{x}^-$の範囲))						(($\boldsymbol{X}$の範囲) - ($\boldsymbol{x}^-$の範囲))			
	1	2	3	…	50		x_1	x_2	x_3	x_4	
x_1	1.0640	0.4640	0.9640	…	-0.2360		1.0640	0.4300	0.4400	0.0740	1
x_2	0.4300	0.4300	0.3300		0.0300		0.4640	0.4300	0.2400	0.1740	2
x_3	0.4400	0.2400	0.6400		-0.1600		0.9640	0.3300	0.6400	0.1740	3
x_4	0.0740	0.1740	0.1740		-0.0260		-0.4360	-0.4700	-0.2600	-0.0260	4
							:				:
							-0.2360	0.0300	-0.1600	-0.0260	50

共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$=				
	0.2664	0.0852	0.1829	0.0558
	0.0852	**0.0985**	0.0827	0.0412
	0.1829	0.0827	**0.2208**	0.0731
	0.0558	0.0412	0.0731	**0.0391**

手順 3）データの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$ の対角要素である変数 x_1 の分散を別途

$$Var(x_1) = \text{SumSq}((\boldsymbol{X}_1\text{の範囲}) - \bar{x}_1) / 49$$

あるいは， $Var(x_1) = \text{Var.S}(\boldsymbol{X}_1\text{の範囲})$

で計算し，計算式をフィルハンドルでコピーし，4 個の分散を計算する．

分散 $\boldsymbol{\sigma}^2$=	0.2664	0.0985	0.2208	0.0391

手順 4）相関行列 $\boldsymbol{R}(\boldsymbol{x})$ を計算する

$\boldsymbol{R}(\boldsymbol{x})$ ＝（$\boldsymbol{\Sigma}(\boldsymbol{x})$の範囲） / Sqrt($\boldsymbol{\sigma}^2$ の範囲) / Sqrt(Transpose($\boldsymbol{\sigma}^2$ の範囲))

	データの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$				SD		データの相関行列 $\boldsymbol{R}(\boldsymbol{x})$			
相関行列 $\boldsymbol{R}(\boldsymbol{x})$=	**0.2664**	0.0852	0.1829	0.0558	0.5162	=	**1**	0.5259	0.7540	0.5465
	0.0852	**0.0985**	0.0827	0.0412	0.3138		0.5259	**1**	0.5605	0.6640
	0.1829	0.0827	**0.2208**	0.0731	0.4699		0.7540	0.5605	**1**	0.7867
	0.0558	0.0412	0.0731	**0.0391**	0.1978		0.5465	0.6640	0.7867	**1**
				対応する行で除す						
SD=	0.5162	0.3138	0.4699	0.1978						
	対応する列で除す									

JMP の「多変量の相関」で計算した結果を表 12.2 に示す．Excel での計算結果と一致することが確認される．

表 12.2　JMP の「多変量の相関」によるデータの共分散行列および相関行列

共分散行列

	がくの長さ	がくの幅	花弁の長さ	花弁の幅
がくの長さ	0.2664	0.0852	0.1829	0.0558
がくの幅	0.0852	0.0985	0.0827	0.0412
花弁の長さ	0.1829	0.0827	0.2208	0.0731
花弁の幅	0.0558	0.0412	0.0731	0.0391

相関

	がくの長さ	がくの幅	花弁の長さ	花弁の幅
がくの長さ	1.0000	0.5259	0.7540	0.5465
がくの幅	0.5259	1.0000	0.5605	0.6640
花弁の長さ	0.7540	0.5605	1.0000	0.7867
花弁の幅	0.5465	0.6640	0.7867	1.0000

相関はリストワイズ法によって推定されました。

データの相関行列と標準偏差 *SD* を用いてデータの共分散行列を逆に求めることもできる．

手順 5）相関行列 $\boldsymbol{R}(\boldsymbol{x})$ と分散ベクトルから共分散行列を計算する．

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = (\boldsymbol{R}(\boldsymbol{x})\text{の範囲}) * \text{Sqrt}(\boldsymbol{\sigma}^2\text{ の範囲}) * \text{Sqrt}(\text{Transpose}(\boldsymbol{\sigma}^2\text{ の範囲}))$$

	データの相関行列 $\boldsymbol{R}(\boldsymbol{x})$				*SD*		データの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$			
共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$=	**1**	0.5259	0.7540	0.5465	0.5162	=	**0.2664**	0.0852	0.1829	0.0558
	0.5259	**1**	0.5605	0.6640	0.3138		0.0852	**0.0985**	0.0827	0.0412
	0.7540	0.5605	**1**	0.7867	0.4699		0.1829	0.0827	**0.2208**	0.0731
	0.5465	0.6640	0.7867	**1**	0.1978		0.0558	0.0412	0.0731	**0.0391**
				対応する行で掛ける						
SD=	0.5162	0.3138	0.4699	0.1978						
	対応する行で掛ける									

こに示した行列計算は，中間的な計算結果を示していないので，Excel の行列計算に不慣れな場合には，Ecxcel シート上に中間結果を書き出すことを薦める．「手順 2）偏差は $(\boldsymbol{X}$の範囲$)-(\bar{\boldsymbol{x}}$の範囲$)$で計算する」などは，練習用に小さな行列を用いることから始めるとよい．なお，行列計算の詳細は，第 4 章で丁寧に説明しているので，参考にしてもらいたい．

逆行列の練習には，相関行列から偏相関行列を求める課題もあり，Excel による行列計算の入門に含めることもできるが，割愛する（添付の Excel シートには含まれている）．

分析ツールを使う場合

Excel の分析ツールで，データの共分散行列および相関行列を計算することができる．ただし，表 12.3 に示すようにデータの共分散行列は，母集団を仮定した場合であり，標本を仮定した場合ではないので，表 12.2 で示した結果とは，微妙に異なる．なお，データ相関行列は，どちらを仮定した場合でも一致する．

表 12.3　Excel の分析ツールによる共分散行列および相関行列

Excel データの共分散行列(母集団)				
	列 1	列 2	列 3	列 4
列 1	**0.2611**			
列 2	0.0835	**0.0965**		
列 3	0.1792	0.0810	**0.2164**	
列 4	0.0547	0.0404	0.0716	**0.0383**

Excel データの相関行列				
	列 1	列 2	列 3	列 4
列 1	**1**			
列 2	0.5259	**1**		
列 3	0.7540	0.5605	**1**	
列 4	0.5465	0.6640	0.7867	**1**

手順 5）によりデータの相関行列から共分散行列を作成することを示したが，そのためには，上三角行列を代入文で埋める必要があるが，手作業的な操作となるので省略する．

共分散関数を使う場合

変数間の共分散については，Excel の Covariance.S()関数を使うと標本に対する共分散が得られる．ただし，2 変数間なので，共分散行列にするためには一工夫する必要がある．

表 12.4　Excel の Covariance.S() 関数による共分散行列

Covariance.S() 関数				
	列 1	列 2	列 3	列 4
列 1	**0.2664**	0.0852	0.1829	0.0558
列 2	0.0852	**0.0985**	0.0827	0.0412
列 3	0.1829	0.0827	**0.2208**	0.0731
列 4	0.0558	0.0412	0.0731	**0.0391**

列 1 と列 1 に

=Covariance.S(C5 : C54,　C$5 : C$54)

（$\boldsymbol{X}_1$ に固定，$\boldsymbol{X}_1$ の行方向を固定）

のように Covariance.S() 関数をセットし，フィルハンドルで列方向に計算式をコピーする．次に列 2 と列 2 に

=Covariance.S(D5 : D54,　D$5 : D$54)

（$\boldsymbol{X}_2$ に固定，$\boldsymbol{X}_2$ の行方向を固定）

のように関数をセットし，フィルハンドルで列の左右に計算式をコピーする．これを繰り返す．このように，いくつかあるの計算手段を知ったうえで，目的に応じて簡便で汎用性の高い方法を選択することを勧める．

Correl() 関数を使い相関係数行列を直接作成することもできるが，Covariance.S() 関数と同様なので割愛する．

12.3. 偏差平方和ベースの重回帰分析

先人たちによって，計算手段が限られていた時代に様々な工夫により，偏差平方和を主体にした単回帰分析を拡張し重回帰分析の定式化がなされてきた．多くの多変量解析および重回帰分析の書物が出版されているなかで，版を重ね読み継がれてきたのは，奥野・久米・芳賀・吉沢 著（1981)，「多変量解析法 改訂版」である．

奥野ら(1981）は，重回帰分析について行列計算を用いずに，シグマを用いた偏差平方和をベースにした方法で多様な事例を通して丁寧に示している．しかし，それらの式の意味を理解し習得するために，Excel でシグマを用いた計算を行うことは絶望的ですらある．そこで，シグマで示されている計算式を Exce の行列関数を用いて行うことにし，デザイン行列を活用した計算方法と対比する．

奥野ら(1981）の「第 4 章 偏回帰係数の解釈」の「表 4.1 材料，工数と生産量の関係」を用いて，行列計算により偏差平方和ベースの重回帰分析の手順を示す．あるガラス加工工程で，投入材料 x_1，使用工数 x_2 と生産量 y の関係を調べたところ，表 12.5 に示すデータが得られた．

表 12.5　材料，工数と生産量の関係［奥野ら(1981)，表 4.1］

	材料	工数	生産量			材料	工数	生産量
No.	x_1 (m^2)	x_2 (hr)	y (m^2)		No.	x_1 (m^2)	x_2 (hr)	y (m^2)
1	54	29	50		12	82	50	73
2	61	39	51		13	75	39	74
3	52	26	52		14	92	60	78
4	70	48	54		15	96	62	82
5	63	42	53		16	92	61	80
6	79	62	60		17	91	50	87
7	68	45	59		18	85	43	84
8	65	30	65		19	106	72	88
9	79	51	67		20	96	52	92
10	76	44	70		計	1,553	941	1,389
11	71	36	70		平均	77.65	47.05	69.45

手順 1）材料 x_1，工数 x_2 の 20 行分のデータを行列 $\boldsymbol{X}$（20×2）とし，平均値をベクトル $\bar{\boldsymbol{x}}$（2×1）とし，次式で偏差平方和行列 $\boldsymbol{S}_{xx}$ を計算する．

$$\boldsymbol{S}_{xx} = (\boldsymbol{X} - \bar{\boldsymbol{x}})^T (\boldsymbol{X} - \bar{\boldsymbol{x}}) = \begin{bmatrix} \mathbf{4218.55} & 3009.35 \\ 3009.35 & \mathbf{2856.95} \end{bmatrix}$$

手順 2）材料 x_1，工数 x_2 と生産量 y との偏差平方和を計算する．20 行分の生産量 y をベクトル $\boldsymbol{y}$，その平均を $\bar{y}$ とし，次式で偏差平方和ベクトル $\boldsymbol{S}_{xy}$ を計算する．

$$\boldsymbol{S}_{xy} = (\boldsymbol{X} - \bar{\boldsymbol{x}})^T (\boldsymbol{y} - \bar{y}) = \begin{array}{|c|}\hline 3499.15 \\ 1860.55 \\ \hline \end{array}$$

手順 3） $\boldsymbol{S}_{xx}$ の逆行列 $\boldsymbol{S}^{xx}$ を計算する．

$$\boldsymbol{S}^{xx} = (\boldsymbol{S}_{xx})^{-1} = \begin{array}{|cc|}\hline \mathbf{0.000954} & -0.001004 \\ -0.001004 & \mathbf{0.001408} \\ \hline \end{array}$$

手順 4）推定値 $\hat{\beta}_1$， $\hat{\beta}_2$ のベクトル $\hat{\boldsymbol{\beta}}$ （2×1）を次式で計算する．

$$\hat{\boldsymbol{\beta}} = \boldsymbol{S}^{xx}\boldsymbol{S}_{xy} = \begin{array}{|c|}\hline 1.4679 \\ -0.8950 \\ \hline \end{array}$$

手順 5）切片 $\hat{\beta}_0$ を次式で計算する．

$$\hat{\beta}_0 = \bar{y} - \bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}} = \boxed{-2.4245}$$

手順 6）以上の計算から回帰式が得られる．

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = -2.4245 + 1.4679 x_1 - 0.8950 x_2$$

奥野ら（1981）に「この式で奇妙なことは，x_2 の係数がマイナスになっていることである．これをそのまま解釈すれば，工数を減らせば生産量（絶対量）が増加するということだから，こんなうまい話はない．はたしてそうであろうか？」このような疑問が示されている．

手順 7）重相関係数 R^2 を次式で計算する．．

$$R^2 = (\hat{\boldsymbol{\beta}}^T \boldsymbol{S}_{xy}) / S_{yy} = \boxed{0.9904}$$

ただし， $S_{yy} = (\boldsymbol{y} - \bar{y})^T (\boldsymbol{y} - \bar{y}) = \boxed{3504.95}$

手順 8）誤差分散 $\hat{\sigma}^2$ を次式で計算する．．

$$\hat{\sigma}^2 = (\boldsymbol{y} - \hat{\boldsymbol{y}})^T (\boldsymbol{y} - \hat{\boldsymbol{y}}) / (n-3) = \boxed{1.9803}$$

ただし， $\hat{\boldsymbol{y}} = \hat{\beta}_0 + \boldsymbol{X}\hat{\boldsymbol{\beta}}$ ， 自由度： $(n-3) = 20 - 3 = 17$

手順 9）推定値 $\hat{\beta}_1$， $\hat{\beta}_2$ の分散 $Var(\hat{\beta}_1)$， $Var(\hat{\beta}_2)$ を次式で計算する．

$$Var(\hat{\beta}_1) = \boldsymbol{S}^{11}\hat{\sigma}^2 = \boxed{0.0019}$$

$$Var(\hat{\beta}_2) = \boldsymbol{S}^{22}\hat{\sigma}^2 = \boxed{0.0028}$$

ただし， $\boldsymbol{S}^{ii}$ は， $\boldsymbol{S}^{xx}$ の対角要素

手順 10）推定値 $\hat{\beta}_0$ の分散 $Var(\hat{\beta}_0)$ を次式で計算する．

$$Var(\hat{\beta}_0) = \left(\frac{1}{n} + \bar{\boldsymbol{x}}\boldsymbol{S}^{xx}\bar{\boldsymbol{x}}^T \right) \hat{\sigma}^2 = \boxed{3.1236}$$

手順 11）分散分析表のための平方和 S_T， S_R， S_e を次式で計算する．

$$S_T = S_{yy} = \boxed{3504.95}$$

$$S_R = (\boldsymbol{S}_{xy})^T \hat{\boldsymbol{\beta}} = \boxed{3504.95}$$

$$S_e = S_{yy} - S_R = \boxed{33.6658}$$

手順 12）以上の Excel での計算シートを表 12.6 示す．その結果は，表 12.7 に示す Excel の回帰分析による分散分析表およびパラメータの推定値と一致することが確認される．

表 12.6　偏差平方和ベースの重回帰の Excel の計算シート

	$(\boldsymbol{X}\text{-}(\boldsymbol{x}^-))^T(\boldsymbol{X}\text{-}(\boldsymbol{x}^-))$			$(\boldsymbol{X}\text{-}(\boldsymbol{x}^-))^T(\boldsymbol{y}\text{-}(y^-))$			$(\boldsymbol{y}\text{-}(y^-))^T(\boldsymbol{y}\text{-}(y^-))$	
$\boldsymbol{S}_{xx}=$	**4218.55**	3009.35	$\boldsymbol{S}_{xy}=$	3499.15		$S_{yy}=$	3504.95	
	3009.35	**2856.95**		1860.55				
	$(\boldsymbol{S}_{xx})^{-1}$			$\boldsymbol{S}^{xx}\boldsymbol{S}_{xx}$			$\boldsymbol{S}^{11}\boldsymbol{\sigma}\hat{}^2$	SE
$\boldsymbol{S}^{xx}=$	**0.000954**	-0.001004	$\boldsymbol{\beta}\hat{}=$	**1.4679**		$Var(\beta_1\hat{})=$	0.0019	**0.0435**
	-0.001004	**0.001408**		**-0.8950**			$\boldsymbol{S}^{22}\boldsymbol{\sigma}\hat{}^2$	
						$Var(\beta_2\hat{})=$	0.0028	**0.0528**
	$(\boldsymbol{\beta}\hat{}^T\boldsymbol{S}_{xx})/\boldsymbol{S}_{xy}$			$(y^-)\text{-}(\boldsymbol{x}^-)\boldsymbol{\beta}\hat{}$			$(1/n+(\boldsymbol{x}^-)\boldsymbol{S}_{xx}(\boldsymbol{x}^-)^T)\sigma\hat{}^2$	
$R^2=$	0.9904		$\beta_0\hat{}=$	**-2.4245**		$Var(\beta_0\hat{})=$	3.1236	**1.7674**
	$(\boldsymbol{y}\text{-}\boldsymbol{y}\hat{})^T(\boldsymbol{y}\text{-}\boldsymbol{y}\hat{})/(n\text{-}3)$							
$\boldsymbol{\sigma}\hat{}^2=$	1.9803		$S_T=$	**3504.95**	S_{yy}			
	$\boldsymbol{y}\hat{}=\beta_0\hat{}+\boldsymbol{X}\boldsymbol{\beta}\hat{}$		$S_R=$	**3471.28**	$(S_{xy})^T\boldsymbol{\beta}\hat{}$			
			$S_e=$	**33.67**	$S_{yy}\text{-}S_R$			

表 12.7　Excel の回帰分析による分散分析表およびパラメータの推定値

分散分析表					
	自由度	変動	分散	分散比	有意 F
回帰	2	3471.2842	1735.6421	876.4369	0.0000
残差	17	33.6658	1.9803		
合計	19	3504.9500			
	係数	標準誤差	t	P-値	
切片	-2.4245	1.7674	-1.3718	0.1880	
X 値 1	1.4679	0.0435	33.7791	0.0000	
X 値 2	-0.8950	0.0528	-16.9485	0.0000	

注）Excel の回帰分析で簡単にできることを，ごちゃごちゃと計算するのは，私にとってもストレスであるが，基本を身に付けなければ，第 12.4 節で示す 2 次式の回帰曲線の 95%信頼区間を描くこととすらできない．先人たちの様々な工夫から学ぶことも，新たな応用力を付けるために必要である．

手順 13）回帰の推定値 $\hat{y}_i$ および分散 $Var(\hat{y}_i)$ を次式で求める．

$\boldsymbol{x}_1$=	54	29

$$\hat{y}_1 = \hat{\beta}_0 + \boldsymbol{x}_1\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,1} + \hat{\beta}_2 x_{2,1} = \boxed{50.8883}$$

$$Var(\hat{y}_1) = \left[\frac{1}{n} + (\boldsymbol{x}_1 - \bar{\boldsymbol{x}})\boldsymbol{S}^{xx}(\boldsymbol{x}_1 - \bar{\boldsymbol{x}})^T\right]\hat{\sigma}^2 = \boxed{0.3655}$$

全ての $\hat{y}_i$ と $Var(\hat{y}_i)$ を求め，結果を表 12.8 に示す．

表 12.8　偏差平方和ベースの重回帰の推定値と分散

	材料	工数	生産量	推定値	分散
No.	x_1	x_2	y	y^	$Var(y$^$)$
1	**54**	**29**	50	**50.89**	**0.3655**
2	61	39	51	52.21	0.2700
3	52	26	52	50.64	0.4290
4	70	48	54	57.37	0.2410
5	63	42	53	52.46	0.2811
:					
18	85	43	84	83.86	0.3652
19	106	72	88	88.74	0.5386
20	96	52	92	91.96	0.4419
計	1,553	941	1,389		
平均	77.65	47.05	69.45		

奥野ら（1981）の重回帰分析の計算手順は，単回帰分析で一般的となっている偏差平方和に基づく計算手順を重回帰分析に拡張したものである．計算手段が乏しかった時代の標準的の方法であり，生物統計の名著であるスネデガー・コクラン（1972）でも，医学統計の名著であるアーミテージら（2001）でも説明変数の偏差平方和に基づく重回帰分析の手順として示されていることから，世界的に標準的な方法として普及してきたと理解される．

このような偏差平方和ベースの重回帰分析の解析手順は，計算手段が乏しく有効数字の桁数が 7 程度であった単精度実数の時代の標準的な手順として理解できる．それぞれの平均を差し引いた後に，標準偏差を計算し基準化して計算精度を確保することは，数値計算の常識でもあった．現在の Excel は倍精度実数での計算が標準であり，平均値を引かなければ計算精度が保てないことはなくなった．ただし，多項式回帰を行う際に，べき乗の項については，桁数のインフレーションを防ぐ何らかの手立てをすることは必要である．

12.4.　デザイン行列ベースの重回帰分析

Excel によるデザイン行列ベースの重回帰分析

「第 4.5 節 デザイン行列を用いた回帰分析」,「第 4.6 節 分散分析表」では，単回帰分析に対しても偏差平方和に基づく方法ではなく，切片を含むデザイン行列を用いた行列計算による回帰分析の解析法を示した．この方法は，切片の推定値を別途計算するのではなく，説明変数のパラメータ推定も併せて行うので，手順としてはスマートであり，ポアソン回帰などへの拡張が容易である．表 12.5 に示した奥野ら（1981）のデータを表 12.9 に示すように切片 x_0 を加えたデザイン行列 $\boldsymbol{X}$ に対し，Excel シート上で重回帰を行い，偏差平方和ベースの重回帰分析の方法と比較してもらいたい．

表 12.9　Excel によるデザイン行列ベースの重回帰分析

	デザイン行列 $\boldsymbol{X}$				推定値	分散					
No.	x_0	x_1	x_2	y	y^	$Var(y$^$)$					
1	1	54	29	50	**50.89**	**0.3655**	$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$=	**1.5773**	-0.0268	0.0117	
2	1	61	39	51	52.21	0.2700		-0.0268	**0.0010**	-0.0010	
3	1	52	26	52	50.64	0.4290		0.0117	-0.0010	**0.0014**	
4	1	70	48	54	57.37	0.2410					
5	1	63	42	53	52.46	0.2811	$\boldsymbol{\beta}$^=	**-2.4245**	: $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$		
6	1	79	62	60	58.05	0.6454		**1.4679**			
7	1	68	45	59	57.12	0.2079		**-0.8950**			
8	1	65	30	65	66.14	0.3538					
9	1	79	51	67	67.90	0.1248	$\boldsymbol{y}$^=	$\boldsymbol{X\beta}$^			
10	1	76	44	70	69.76	0.1101					
11	1	71	36	70	69.58	0.2307	σ^{2}^=	1.9803	: $(\boldsymbol{y}-\boldsymbol{y}$^$)^T(\boldsymbol{y}-\boldsymbol{y}$^$)/17$		
12	1	82	50	73	73.20	0.1080					
13	1	75	39	74	72.76	0.2081	$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)$=	**3.1236**	-0.0530	0.0233	: $Var(\beta_0$^$)$
14	1	92	60	78	78.92	0.2162		-0.0530	**0.0019**	-0.0020	: $Var(\beta_1$^$)$
15	1	96	62	82	83.01	0.2667		0.0233	-0.0020	**0.0028**	: $Var(\beta_2$^$)$
16	1	92	61	80	78.03	0.2341					
17	1	91	50	87	86.41	0.3032	$Var(y_i$^$)$=	$\boldsymbol{x}_i\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)\boldsymbol{x}_i^T$			
18	1	85	43	84	83.86	0.3652					重相関
19	1	106	72	88	88.74	0.5386	S_T=	**3504.95**	: SumSq($\boldsymbol{y}$-$\bar{y}$)		$\boldsymbol{R}^2$=S_R/S_T
20	1	96	52	92	91.96	0.4419	S_R=	**3471.28**	: SumSq($\boldsymbol{y}$^-$\bar{y}$)		0.9904
平均	1.00	77.65	47.05	69.45			S_e=	**33.67**	: SumSq($\boldsymbol{y}$-$\boldsymbol{y}$^)		

手順 a）切片 x_0，材料 x_1，工数 x_2 のデザイン行列 $\boldsymbol{X}$ （20×3）に対し積和行列を求め，その逆行列を計算する．

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$=	**1.5773**	-0.0268	0.0117
	-0.0268	**0.0010**	-0.0010
	0.0117	-0.0010	**0.0014**

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

以下，Excel での計算式は省略する．

手順 b）推定値 $[\hat{\beta}_0 \quad \hat{\beta}_1 \quad \hat{\beta}_3]^T$ のベクトル $\hat{\boldsymbol{\beta}}$ （3×1）を次式で計算する．

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \begin{array}{|r|}\hline -2.4245 \\ 1.4679 \\ -0.8950 \\ \hline \end{array}$$

手順 c）推定された $\hat{\boldsymbol{\beta}}$ を用いて推定値 $\hat{\boldsymbol{y}}$ を計算する．

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$$

$$\begin{aligned}\hat{y}_1 &= \hat{\beta}_0 x_{0,1} + \hat{\beta}_1 x_{1,1} + \hat{\beta}_2 x_{2,1} \\ &= -2.4245\times 1 + 1.4679\times 54 - 0.8950\times 29 \\ &= 50.89\end{aligned}$$

手順 d）誤差分散 $\hat{\sigma}^2$ を次式で計算する．.

$$\begin{aligned}\hat{\sigma}^2 &= (\boldsymbol{y}-\hat{\boldsymbol{y}})^T(\boldsymbol{y}-\hat{\boldsymbol{y}})/(n-3) \\ &= \mathrm{SumSq}(\boldsymbol{y}\text{の範囲} - \hat{\boldsymbol{y}}\text{の範囲})/17 \\ &= 1.9803\end{aligned}$$

手順 e）パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を計算する．対角要素がパラメータの分散になる．

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2 = \begin{array}{|r|r|r|l|}\hline \mathbf{3.1236} & -0.0530 & 0.0233 & : Var(\hat{\beta_0}) \\ -0.0530 & \mathbf{0.0019} & -0.0020 & : Var(\hat{\beta_1}) \\ 0.0233 & -0.0020 & \mathbf{0.0028} & : Var(\hat{\beta_2}) \\ \hline \end{array}$$

手順 f）分散分析表のための平方和 S_T， S_R， S_e を次式で計算する．

$$S_T = \mathrm{SumSq}(\boldsymbol{y}\text{の範囲} - \bar{y}) = 3504.95$$

$$S_R = \mathrm{SumSq}(\hat{\boldsymbol{y}}\text{の範囲} - \bar{y}) = 3504.95$$

$$S_e = \mathrm{SumSq}(\boldsymbol{y}\text{の範囲} - \hat{\boldsymbol{y}}\text{の範囲}) = 33.67$$

手順 g）重相関係数 R^2 を次式で計算する．.

$$R^2 = S_R / S_T = 0.9904$$

手順 h）回帰の推定値 $\hat{y}_1$ の分散 $Var(\hat{y}_1)$ を次式で求め，

$\boldsymbol{x}_1$=	1	54	29

$$Var(\hat{y}_1) = \boldsymbol{x}_1\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\boldsymbol{x}_1^T = 0.3655$$

$\boldsymbol{x}_1$			$\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{})$			$\boldsymbol{x}_i^T$		
1	54	29	**3.1236**	-0.0530	0.0233	1	=	0.3655
			-0.0530	**0.0019**	-0.0020	54		
			0.0233	-0.0020	**0.0028**	29		

フィルハンドルを用いて計算式をコピーして全ての $Var(\hat{y}_i)$ を求める．

2 変量の重回帰分析について，奥野ら（1981）に忠実に偏差平方和をベースにした解析を Excel の行列関数を用いて行い，それと対比する形でデザイン行列をベースにした解析法を対

比した．手順数は 13 から 8 に減少し，数式も簡素化された．これは，偏差平方和をベースにした場合に，常に平均値のベクトルを考慮した式となり，さらに，切片の推定値 $\hat{\beta}_0$ を別計算で求める煩雑さが付きまとっている．慣れ親しんだ 1 変量の回帰分析の手順を多変量に拡張することは可能であることを示した．ただし，ポアソン回帰を含んだ一般化線形モデルへの拡張を視野にした場合には，切片を含めたデザイン行列をベースにした回帰分析の方法が理解の助けになる．

等高線図

手順 6）および手順 c）で推定された回帰式，

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 = -2.4245 + 1.4679 x_1 - 0.8950 x_2$$

に対し，x_2 のパラメータがマイナスとなっており，そのまま解釈すれば，工数を減らせば生産量（絶対量）が増加するということだから，こんなうまい話はない．「はたしてそうであろうか」との疑問が起きる．

この疑問に答えるために，2 変量の等高線図を用いる必要がある．図 12.1 に X 軸に材料 x_1 を，Y 軸に工数 x_2 とし，生産量 y の値を散布図として示す．工数 x_2 を 50 hr に固定し，材料 x_1 の 70 m^2～90 m^2 の等高線を読むと生産量 y は，60，70，80 m^2 と増加している．次に，材料 x_1

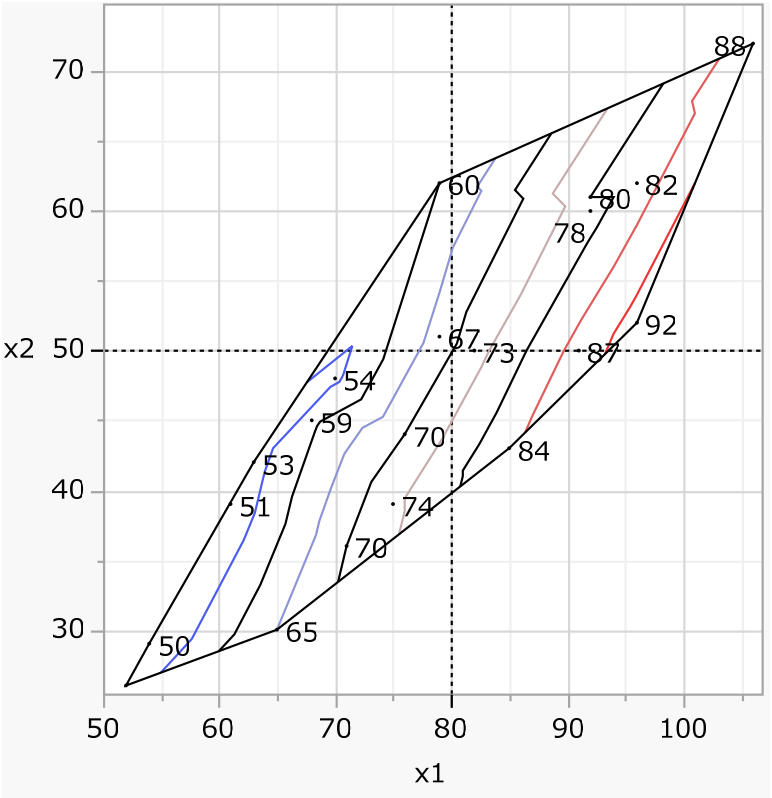


図 12.1 材料 x_1 と工数 x_2 に対する生産量 y の JMP による等高線図

を 80 m^2 に固定し，工数 x_2 の 40 hr～60 hr の等高線を読むと生産量 y は，おおよそ 80，70，60 m^2 と減少している．

等高線から，工数 x_2 を増やすことなく材料 x_1 を増やせば，生産量 y を増やすこと読み取れる．材料 x_1 が同じでも工数 x_2 を減らすことにより生産量をわずかであるが増やすことができる．材料 x_1 が同じでも工数 x_2 を増加させると生産量が落ちることは，何か別の測定していない変数の影響が潜んでいることが伺われる．

予測プロファイル

等高線図を用いた検討方法をさらに細かく「予測プロファイル」を作成し検討する．表 12.10 に示すように，等高線図を参考にして材料 x_1 を 80 m^2 に固定し，工数 x_2 を（30，40, . . . , 70）と変化させて，生産量 y を推定する．材料 x_1 を 80 m^2，工数 x_2 を 30 hr とした場合に，

$$\begin{aligned}\hat{y}_1 &= \hat{\beta}_0 + \hat{\beta}_1 x_{1,1} + \hat{\beta}_2 x_{2,1} \\ &= -2.4245 + 1.4679 \times 80 - 0.8950 \times 30 \\ &= 88.16\end{aligned}$$

が推定される．分散は，手順 h）で示した方法で，

$$Var(\hat{y}_1) = \boldsymbol{x}_1 \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \boldsymbol{x}_1^T =$$

1	80	30	**3.1236**	-0.0530	0.0233	1	=	1.0795
			-0.0530	**0.0019**	-0.0020	80		
			0.0233	-0.0020	**0.0028**	30		

にて計算する．95%信頼区間は，

$$\begin{aligned}(L95\%, U95\%) &= \hat{y}_1 \pm t(0.05,17)\sqrt{Vay(\hat{y}_1)} \\ &= 80.82 \pm 2.1098\sqrt{1.0795} \\ &= (85.97, 90.35)\end{aligned}$$

表 12.10　予測プロファイルのための予測値と 95%信頼区間の計算

	切片	材料	工数	推定値	分散	95%信頼区間					
i	x_0	x_1	x_2	y^	$Var(y$^$)$	$L95\%$	$U95\%$				
1	1	**80**	30	88.16	1.0795	85.97	90.35	$\boldsymbol{\beta}$^=	-2.4245		
2	1	**80**	40	79.21	0.3139	78.03	80.39		1.4679		
3	1	**80**	50	70.26	0.1061	69.57	70.95		-0.8950		
4	1	**80**	60	61.31	0.4560	59.88	62.73				
5	1	**80**	70	52.36	1.3636	49.90	54.82	$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)$=	**3.1236**	-0.0530	0.0233
6	1	60	**50**	40.90	0.9187	38.88	42.92		-0.0530	**0.0019**	-0.0020
7	1	70	**50**	55.58	0.3236	54.38	56.78		0.0233	-0.0020	**0.0028**
8	1	80	**50**	70.26	0.1061	69.57	70.95				
9	1	90	**50**	84.94	0.2664	83.85	86.03	$Var(y_i$^$)$= $\boldsymbol{x}_i \boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)\boldsymbol{x}_i^T$			
10	1	100	**50**	99.62	0.8043	97.73	101.51				

となる．さらに，工数 x_2 を 50 hr に固定し，材料 x_1 を（60，70, . . . 100）と変化させて，生産量 $\hat{y}$ と 95%信頼区間を計算する．

表 12.10 で計算した推定値と 95%信頼区間を，Excel の散布図を用いて図 12.2 に「予測プロファイル」を作成する．表 12.10 の x_1 および x_2 を変化させると，図 12.2 の予測プロファイル」も連動して変化する．

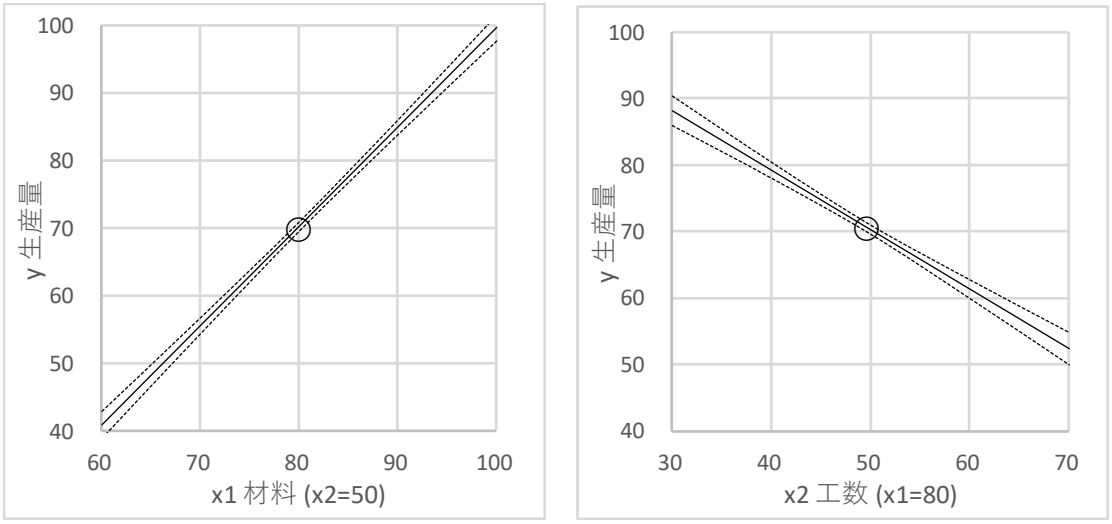


図 12.2　予測プロファイル
図左：工数 x_2 を 50 hr に固定し材料 x_1 を変化せた場合の生産量の変化
図右：材料 x_1 を 80 m^2 に固定し工数 x_2 を変化せた場合の生産量の変化

偏差平方和ベース vs デザイン行列ベース

ドレーパー・スミス（1968）では，単回帰分析について，偏差平方和をベースにした解析を示した後に，重回帰分析への導入を意図し，同じデータを用いてデザイン行列ベースの単回帰分析を示している．Excel の分析ツールの「回帰分析」により，誰にでも手軽に重回帰分析が行えるようになったのは，素晴らししいことと思う．しかし，Excel の「回帰分析」に欠けているのは，パラメータの共分散行列の出力がないことである．単回帰分析の 95%信頼区間をグラフに示したいと思った場合には，説明変数 x について偏差平方和 S_{xx} を別途計算し，分散の公式を使って，何とかできる範囲ではある．さて，説明変数が複数ある場合には，どうしたらよいのだろうかと，途方に暮れることになる．

偏差平方和をベースした場合には，計算手順の中で，偏差平方和の逆行列を

手順 3） $\boldsymbol{S}_{xx}$ の逆行列 $\boldsymbol{S}^{xx}$ を計算する．

$$\boldsymbol{S}^{xx}=(\boldsymbol{S}_{xx})^{-1}=\begin{array}{|c|c|}\hline \mathbf{0.000954} & -0.001004 \\ \hline -0.001004 & \mathbf{0.001408} \\ \hline \end{array}$$

とあり，これに $\hat{\sigma}^2=1.9803$ を掛ければ，

	$\hat{\beta_0}$	$\hat{\beta_1}$	$\hat{\beta_2}$
$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=$			
		0.0019	-0.0020
		-0.0020	**0.0028**

切片を含まないパラメータについての共分散行列となり，さらに，切片の分散

$$Var(\hat{\beta}_0)=\left(\frac{1}{n}+\bar{\boldsymbol{x}}\boldsymbol{S}^{xx}\bar{\boldsymbol{x}}^T\right)\hat{\sigma}^2=3.1236$$

の結果を加えても，

	$\hat{\beta_0}$	$\hat{\beta_1}$	$\hat{\beta_2}$
$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=$	**3.1236**	未	未
	未	**0.0019**	-0.0020
	未	-0.0020	**0.0028**

のように，切片と他のパラメータの共分散が欠けた共分散行列しかできない．

現実的な対応として，Excel の「回帰分析」を使い，パラメータの推定値と誤差分散 $\hat{\sigma}^2$ 求める．パラメータに関する共分散行列は，説明変数に切片を加えたデザイン行列 $\boldsymbol{X}$ を設定し，これまでも示してきた計算式

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$$
$$=\text{Minverse(}\ \text{Mmult(Transpose(}\boldsymbol{X}\text{の範囲),}\boldsymbol{X}\text{の範囲))*}\hat{\sigma}^2$$

$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=$	**3.1236**	-0.0530	0.0233
	-0.0530	**0.0019**	-0.0020
	0.0233	-0.0020	**0.0028**

によって，簡単に求めることができる．いずれにしても，切片を含むデザイン行列 $\boldsymbol{X}$ を使うことが，各種の推定値に 95%信頼区間を計算する際に利便性の向上になる．

統計教育の現場での葛藤

統計教育に携わってきた人達から，「シグマを使うと嫌われ，ましてや行列を出すとそっぽを向かれる」を散々聞かされてきたが，Excel で行列を扱い始めると，シグマを使った計算式は，冗長で見たくも計算したくもないと思うようになるのではないだろうか．私も久々に奥野ら(1981）のシグマを用いた計算式に遭遇してめまいを感じた．救いは，サマリーとして行列表記が適宜挿入されていて，これにより偏差平方和ベースの計算を Excel の行列計算で容

易に行うことができた．この経験により，切片を含むデザイン行列ベースの回帰分析の簡潔な計算方法の優位性を認識しつつ，先人たちの苦悩を再認識した．

私は，幸いなことに切片を含むデザイン行列を用いた方法に慣れ親しんできたので，奥野ら(1981）の偏差平方和ベースによる重回帰分析の解析法には，ほとんど関心がなかった．あらためて Eecel での計算をし，計算手段が乏しい時代の先人たちの苦労を垣間見たのである．他方，Excle の分析ツールの回帰分析（重回帰分析）のみならず，全ての統計ソフトでの重回帰分析は，切片を含まない変数の指定を前提にしている．これは，ごく自然のことと思うのであるが，解析結果に含まれる切片について認識不足にもなる原因である．

デザイン行列ベースの重回帰

偏差平方和ベースおよびザイン行列ベースの重回帰について Excel を用いて例示してきたのであるが，詳しくは，新村(1983a,b）の「行列計算による重回帰分析（1）および（2)」を参照されたい．この論文は，応答変数 y と説明変数 x_1 から x_4 までの 4 個の説明変数からなる 7 個の観測データを用いて行列表現による重回帰分析について詳しく解説している．

また，偏差平方和ベースの重回帰分析について,「規準化データによる重回帰」として例示と文献の引用もあり，時代的な背景の理解に役に立つ．データの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{x})$，推定値 $\hat{\boldsymbol{\beta}}$ の分散行列（パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ ）の明示的な使い分けなどについても示唆された．さらに，テコ比，スチューデント化残差の具体的な計算事例についても，Excel での計算を行う際に参考にした．

この時代の重回帰分析は，逆行列を求めるために掃き出し計算が用いられており，奥野ら（1981）にも丁寧な解説がある．本書では, Excel の Minverse() 関数を使うことを前提にし,，逆行列の計算はブラック・ボックスのままにしてきた．詳しくは，新村(1983c）,「重回帰分析における掃き出し演算子」を参照のこと．なお，私にとっても掃き出し計算による逆行列の計算は, Fortran を使っていたころ慣れ親しんできたので, なつかしく思うのであるが, Excel のソルバーを含む基本の計算機能だけでは逆行列の計算は実現できなかったので，深入りはしない．

12.5.　2 次曲線の 95%信頼区間

通常の回帰分析において，回帰直線の 95%信頼区間の計算式，95%予測区間（個別データの 95%信頼区間）の計算式は，ほとんどの統計の教科書で示されていて，統計ソフトでもこれらの信頼区間のグラフ表示も標準的にサポートされている．反応が直線でなく曲線となるような場合に，2 次式あるいは 3 次式のあてはめを検討することも一般的に薦められている．

ところで，2 次式の 95%信頼区間の計算はどのようにしたらよいのだろうか．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を使って，デザイン行列 $\boldsymbol{X}$ の行ベクトル $\boldsymbol{x}_i$ から，分散 $Var(\hat{y}_i)$ を次式

$$Var(\hat{y}_i) = \boldsymbol{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \boldsymbol{x}_i^T$$

で計算し

$$95\%\mathrm{CL} = \hat{y}_i \pm t(0.05, df)\sqrt{Var(\hat{y}_i)}$$

によって 95%信頼区間を計算すればよい．本書でも，各種の事例で推定値の 95%信頼区間を計算しグラフ表示をしてきたのであるが，初心に戻り 2 次式の 95%信頼区間を実際に計算し，Excel で散布図上に描いてみる．

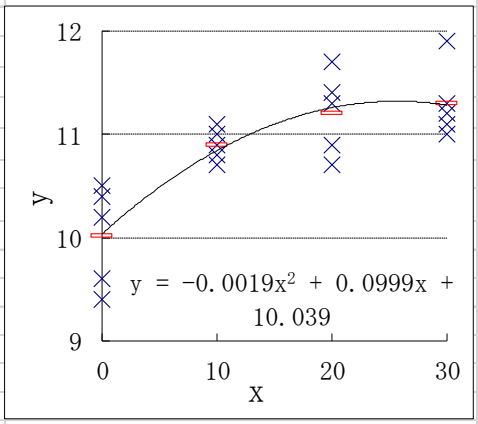
芳賀の事例

事例としては，表 12.11 に芳賀(2009)，「医薬品開発のための統計解析　第 2 部　実験計画法」の第 2.2 節「非線形の関係」で使われているデータを示す．なお，芳賀(2009）に掲載されている Excel の図表類は，https://scientist-press.com/download/（2020 年 5 月 4 日アクセス）から得られる．

表 12.11　2 次曲線となるデータ［芳賀(2009)，表示 2.2.1 再掲］

			データ				
水準	n	平均	1	2	3	4	5
0	5	10.02	10.5	9.6	10.4	10.2	9.4
10	5	10.90	10.8	10.7	11.1	10.9	11.0
20	5	11.20	11.4	10.7	10.9	11.3	11.7
30	5	11.30	11.9	11.2	11.0	11.1	11.3
全体	20	10.855					
水準数	4						

			残差				
水準	標準偏差	効果	1	2	3	4	5
0	0.49	−0.84	0.48	−0.42	0.38	0.18	−0.62
10	0.16	0.04	−0.10	−0.20	0.20	0.00	0.10
20	0.40	0.34	0.20	−0.50	−0.30	0.10	0.50
30	0.35	0.45	0.60	−0.10	−0.30	−0.20	0.00



分散分析表

要因	平方和	自由度	平均平方	F 比	p値
水準間	5.082	3	1.694	12.27	0.0002
残差	2.208	16	0.138	1.00	
全体	7.290	19			
(検算)	7.290	19			

高橋の注）
この分散分析表は，
4水準の一元配置の分散分析表であり，
2次式のあてはめた場合の分散分析表ではない．

芳賀は，2 次式の回帰パラメータを求めるために Excel の LinEst() 関数を用いており，詳しい解説がなされている．2 次式の回帰曲線に対する 95%信頼区間は，JMP の「二変量の関係」によるグラフで示されているが，その計算方法については示されていない．そこで，Excel の「分析ツール：回帰分析」を用いて回帰パラメータおよび誤差分散を計算し，パラメータの共分散行列を Excel の行列関数を用いて，2 次式の回帰の 95%信頼区間をグラフ化する手順を示す．

Excel による 2 次式のあてはめ

表 12.11 のデータを表 12.12 に示すようにデザイン行列の形に整え，Excel の分析ツールの回帰分析で得られた分散分析表と回帰パラメータの推定値を示す．分散分析表の「残差」の行の分散の列が誤差分散 $\hat{\sigma}^2 = 0.1320$ となる．得られた列ベクトルの回帰パラメータの推定値を

$$\hat{\boldsymbol{\beta}} = [\,10.0390 \quad 0.0999 \quad -0.0020\,]^T$$

のように行ベクトルに転置記号「T」を付けて示す．デザイン行列 $\boldsymbol{X}$ の転置行列 $\boldsymbol{X}^T$ を Transpose() 関数で求め，$\boldsymbol{X}^T$ と $\boldsymbol{X}$ の積を Mmult()関数で計算する．

表 12.12　2 次式のあてはめ

	デザイン行列 $\boldsymbol{X}$					Excel 分析ツール 回帰分析			
i	切片	x	x^2	y		分散分析表			
1	1	0	0	10.5			自由度	変動	分散
2	1	0	0	9.6		回帰	2	5.0454	2.5227
3	1	0	0	10.4		残差	17	2.2441	**0.1320**
4	1	0	0	10.2		合計	19	7.2895	
5	1	0	0	9.4			係数	標準誤差	
6	1	10	100	10.8		切片	**10.0390**	0.1584	
7	1	10	100	10.7		X 値 1	**0.0999**	0.0254	
8	1	10	100	11.1		X 値 2	**-0.0020**	0.0008	
9	1	10	100	10.9					
10	1	10	100	11.0		行列計算			
11	1	20	400	11.4		$\boldsymbol{X}^T\boldsymbol{X} =$	**20**	300	7000
12	1	20	400	10.7			300	**7000**	180000
13	1	20	400	10.9			7000	180000	**4900000**
14	1	20	400	11.3					
15	1	20	400	11.7		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}=$	**0.1900**	-0.0210	0.0005
16	1	30	900	11.9			-0.0210	**0.0049**	-0.0002
17	1	30	900	11.2			0.0005	-0.0002	**5.00E-06**
18	1	30	900	11.0					
19	1	30	900	11.1		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma\text{^}2=$	**2.508E-02**	-2.772E-03	6.600E-05
20	1	30	900	11.3		$\boldsymbol{\Sigma}(\boldsymbol{\beta}\text{^})$	-2.772E-03	**6.468E-04**	-1.980E-05
							6.600E-05	-1.980E-05	**6.600E-07**

							X						
	X^T						切片	x	x^2		X^TX		
$X^TX=$	1	1	1	・・・	1		1	0	0	=	**20**	300	7000
	0	0	0		30		1	0	0		300	**7000**	180000
	0	0	0		900		1	0	0		7000	180000	**4900000**
							:						
							1	30	900				
					=Mmult(Transpose(Xの範囲), Xの範囲)								

X^TX の逆行列を Minverse() 関数で求め，

		$(X^TX)^{-1}$		
$(X^TX)^{-1}$ =		**0.1900**	-0.0210	0.0005
		-0.0210	**0.0049**	-0.0002
		0.0005	-0.0002	**5.00E-06**
		= Minverse(X^TXの範囲)		

Excel の分散分析表の「分散」列の「残差」の行の誤差分散 $\hat{\sigma}^2 = 0.1320$ を掛けて，パラメータの共分散行列 $\Sigma(\hat{\beta})$

	$(X^TX)^{-1}$			$\sigma^{\wedge 2}$		$\Sigma(\beta^{\wedge})$		
$(X^TX)^{-1}\sigma^{\wedge 2}=$	**0.1900**	-0.0210	0.0005	0.1320	=	**2.508E-02**	-2.772E-03	6.600E-05
	-0.0210	**0.0049**	-0.0002			-2.772E-03	**6.468E-04**	-1.980E-05
	0.0005	-0.0002	**5.00E-06**			6.600E-05	-1.980E-05	**6.600E-07**

を計算する．パラメータの共分散行列 $\Sigma(\hat{\beta})$ の対角要素の平方根が，回帰パラメータの標準誤差となっている．

$Ver(\beta^{\wedge})=$	0.025081118	$SE(\beta^{\wedge})=$	0.1584
	0.000646829		0.0254
	0.000000660		0.0008

段階を追って計算式を示したが，

$$\Sigma(\hat{\beta}) = \text{Minverse}(\ \text{Mmult}(\text{Transpose}(X\text{の範囲}), X\text{の範囲})\ ) * \hat{\sigma}^2$$

のようにまとめて計算することができる．

推定値の 95%信頼区間

Excel で推定された回帰パラメータを用いて推定値 $\hat{y} = X\hat{\beta}$ を計算する．デザイン行列 X の i 行目のベクトル x_i を用いて $\hat{y}_i = x_i\hat{\beta}$ として計算し，フィルハンドルを用いて計算式をコピーする方法でも良い．

		X			$\beta^{\wedge}$		$y^{\wedge}$
$X\beta^{\wedge}=$	1	0	0		10.0390	=	10.04
	1	10	100		0.0999		10.84
	1	20	400		-0.0020		11.26
	1	30	900				11.28
	=Mmult(Xの範囲，$\beta^{\wedge}$の範囲)						

表 12.13 に示すように，推定値 $\hat{y}_i$ の分散 $Var(\hat{y}_i)$ は， $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を挟む $\boldsymbol{x}_i$ の 2 次形式 $\boldsymbol{x}_i\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\boldsymbol{x}_i^T$ によって計算することができる．

$Var(\hat{y}_6) = \boldsymbol{x}_6\boldsymbol{\Sigma}\boldsymbol{x}_6^T =$

$\boldsymbol{x}_6$			$\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{})$			$\boldsymbol{x}_6^T$		$Var(y_6\hat{})$
1	10	100	**2.51E-02**	-2.77E-03	6.60E-05	1	=	0.0145
			-2.77E-03	**6.47E-04**	-1.98E-05	10		
			6.60E-05	-1.98E-05	**6.60E-07**	100		

推定値の $Var(\hat{y}_i)$ 分散の平方根を用いて，95%信頼区間および 95%予測区間を

$$\text{信頼区間} = \hat{y}_i \pm t(0.05,17)\sqrt{Var(\hat{y}_i)} = (10.5888, 11.0972)$$

$$\text{予測区間} \ \hat{y}_i \pm t(0.05,17)\sqrt{Var(\hat{y}_i) + \hat{\sigma}^2} = (10.0354, 11.6506)$$

として計算することができる．回帰式の滑かな線グラフにするために，x の範囲を広げ（–10, –5, . . . , 40）について，推定値 $\hat{y}$，95%信頼区間，95%予測区間を計算する．

表 12.13　散布図に上書きする 2 次曲線と信頼区間の計算シート

			―― $\boldsymbol{X}$ ――					信頼区間		予測区間	
i	x	y	切片	x	x^2	$y\hat{}$	$Var(y\hat{})$	L95%	U95%	L95%	U95%
1	0	10.5	1	-10	100	8.85	0.2046	7.8907	9.7993	7.6209	10.0691
2	0	9.6	1	-5	25	9.49	0.0776	8.9029	10.0786	8.5247	10.4568
3	0	10.4	1	0	0	10.04	0.0251	9.7049	10.3731	9.2028	10.8752
4	0	10.2	1	5	25	10.49	0.0123	10.2558	10.7237	9.6883	11.2912
5	0	9.4	1	10	100	10.84	**0.0145**	**10.5888**	**11.0972**	**10.0354**	**11.6506**
6	10	10.8	1	15	225	11.10	0.0169	10.8244	11.3731	10.2846	11.9129
7	10	10.7	1	20	400	11.26	0.0145	11.0028	11.5112	10.4494	12.0646
8	10	11.1	1	25	625	11.32	0.0123	11.0838	11.5517	10.5163	12.1192
9	10	10.9	1	30	900	11.28	0.0251	10.9469	11.6151	10.4448	12.1172
10	10	11.0	1	35	1225	11.15	0.0776	10.5589	11.7346	10.1807	12.1128
11	20	11.4	1	40	1600	10.92	0.2046	9.9607	11.8693	9.6909	12.1391
12	20	10.7									
13	20	10.9				$\boldsymbol{\beta}\hat{}$		$\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{})=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma^2$			
14	20	11.3			切片	**10.0390**		**2.51E-02**	-2.77E-03	6.60E-05	
15	20	11.7			x	**0.0999**		-2.77E-03	**6.47E-04**	-1.98E-05	
16	30	11.9			x^2	**-0.0020**		6.60E-05	-1.98E-05	**6.60E-07**	
17	30	11.2									
18	30	11.0					$\sigma\hat{}^2$		$t(0.05, 17)$		
19	30	11.1					**0.1320**		**2.1098**		
20	30	11.3									

Excel の「散布図」による 2 次式のグラフを図 12.3 に示す．最初に x と y の散布図を描き，その上に「データの選択」機能を用いて，$\hat{y}$，信頼区間の下限・上限，予測区間下限・上限を上書きし，「データ行列の書式設定」機能を用いて整える．

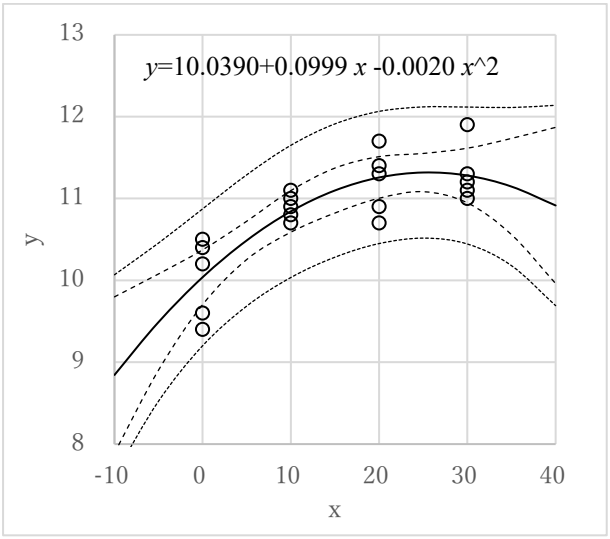


図 12.3　Excel の散布図を用いた 2 次式の 95%信頼区間および 95%予測区間

JMP の「二変量の関係」による 2 次式のあてはめ

JMP の「二変量の関係」には，多項式に対する 95%信頼区間を上書きする機能があるので，結果を示す．芳賀（2009）には，この JMP で作成した図が示されているが，計算方法は示されていない．

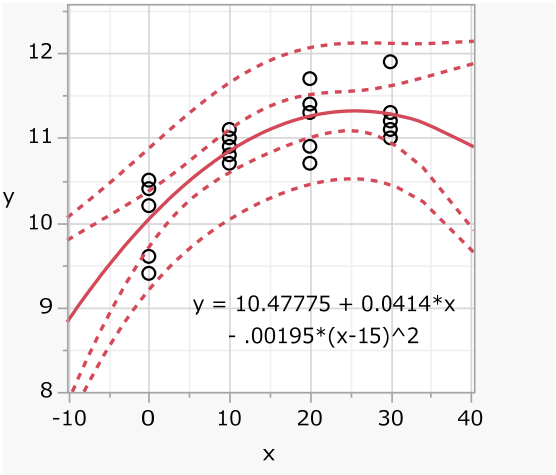
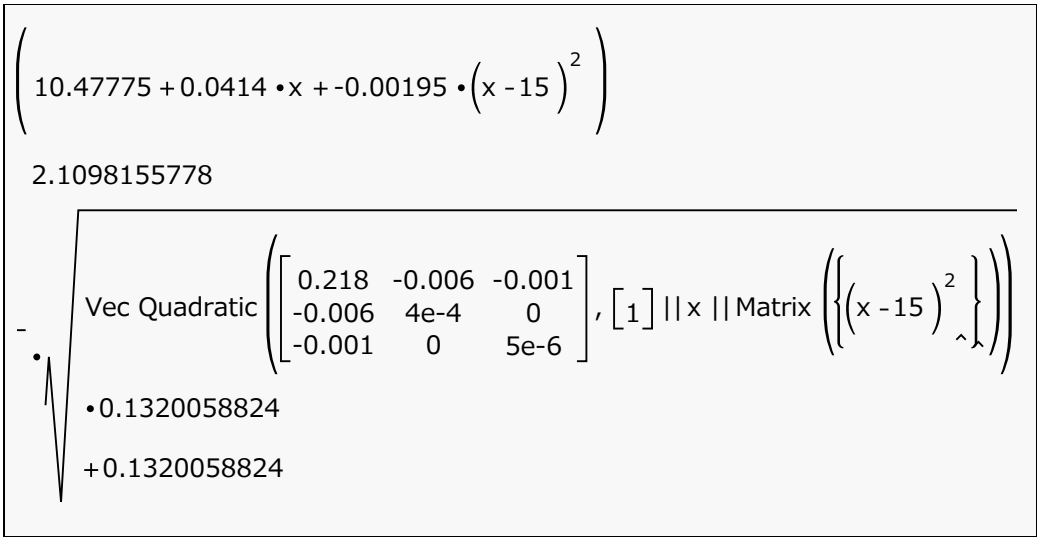


図 12.4　JMP による 2 次式の 95%信頼区間および 95%予測区間

JMP には，95%信頼区間および個別データの 95%信頼区間の計算式を JMP ファイルに出力する機能があり，これを用いて JMP の内部でどのような計算をしているかを可視化することができる．表 12.14 に示すように，推定値 $\hat{y}$ の計算が最初に示されているが，2 次の項は，$(x-15)^2$ となっており，x から平均値 15 を引くことにより，2 乗の桁数が増えないように「多項式の中心化」が行われている．次に 95%信頼区間の計算のために自由度が 17 の両側 5%の t 値が 2.1098 と示され，標準誤差の計算のために分散の平方根が現われ，その中で 2 次形式

表 12.14　JMP で生成された個別データの下側 95%信頼区間雄の計算式



の VecQuadratic()関数の第 1 引数とし積和行列 $(\boldsymbol{X}-15)^T(\boldsymbol{X}-15)$ の逆行列，第 2 引数として [1 x　$(x\text{-}15)^2$]が示されている．これに誤差分散 $\hat{\sigma}^2=0.1320$ を掛け推定値の回帰の分散 $Var(\hat{y})$ を求めている．さらに，誤差分散 $\hat{\sigma}^2=0.1320$ を加え，個別データの分散としている．

$$\sqrt{Var(個別\hat{y})}=\sqrt{\text{VecQuadratic}\{[(\boldsymbol{X}-15)^T(\boldsymbol{X}-15)]^{-1},[1\ \ x\ \ (x-15)^2]\}\hat{\sigma}^2+\hat{\sigma}^2}$$

なお，$(x-15)^2$ と中心化した場合の $[(\boldsymbol{X}-15)^T(\boldsymbol{X}-15)]^{-1}$ を Excel で別途計算すると，

$[(\boldsymbol{X}-15)^T(\boldsymbol{X}-15)]^{-1}=$	**0.218125**	-0.006000
	-0.006000	**0.000400**
	-0.000625	0.000000

が得られ，JMP の計算式に一致することが確認される．このように，統計ソフトの内部では，行列計算が行われている．通常，ユーザが目にすることはできないが，JMP では，表 12.14 にその一部に示すように可視化することができるようになっていることは，統計ソフトを通じての学習効果が期待される．

2 次式の 95%信頼区間の計算式を行列表記でなくシグマで表わすことも，分散共分散の要素を使って表わすこともできるが，冗長であり示すこともためらわれ，さらに，それらの式を使った計算を例示することもためらわれる．パラメータの共分散行列を使った 2 次形式で示したら読者からそっぽを向かれる．このような事情により，2 次式の 95%信頼区間の計算方法が，ブラック・ボックス化している要因となっていると思われる．いずれにしても，容易かつ可視化に優れる Excel の行列計算が，ブレイク・スルーのための立役者となることを期待したい．

「自然科学の統計学」での事例

東京大学教養学部統計学教室編（1992），「基礎統計学 III 自然科学の統計学」の第 2 章の表 2.3 に「2 次多項式のデータ：液体のある成分と曇り点の関係」が示されている．このデータは，ある溶液の成分（I-8）の比率とその溶液の曇り点（透明な溶液が温度の変化によって曇りを生じさせる温度）の関係である．元のデータを x の小さい順に並び替え，切片および 2 乗項を付け加え，Excel の「回帰分析」の結果を加えたものを表 12.15 に示す．さらに，分散分析表の残差の分散から $\hat{\sigma}^2 = 0.155412$ とし，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$ を Excel の行列関数で計算した結果を加えてある．

表 12.15　2 次多項式のデータ：液体のある成分と曇り点の関係

	——$\boldsymbol{X}$——			曇り点
No.	切片	x	x^2	y
1	1	0	0	21.9
2	1	0	0	22.1
3	1	0	0	22.8
4	1	1	1	24.5
5	1	2	4	26.0
6	1	2	4	26.1
7	1	3	9	26.8
8	1	3	9	27.3
9	1	4	16	28.2
10	1	4	16	28.5
11	1	5	25	28.9
12	1	6	36	29.8
13	1	6	36	30.0
14	1	6	36	30.3
15	1	7	49	30.4
16	1	8	64	31.4
17	1	8	64	31.5
18	1	9	81	31.8
19	1	10	100	33.1

分散分析表			
	自由度	変動	分散
回帰	3	15064.41	5021.47
残差	16	2.486599	**0.155412**
合計	19	15066.90	
	係数	標準誤差	t
切片	0	#N/A	#N/A
X 値 1	**22.5612**	0.1984	113.6984
X 値 2	**1.6680**	0.0990	16.8568
X 値 3	**-0.0680**	0.0103	-6.5911

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$

0.253356	-0.097147	0.007903
-0.097147	**0.063004**	-0.006266
0.007903	-0.006266	**0.000684**

$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$

0.039375	-0.015098	0.001228
-0.015098	**0.009792**	-0.000974
0.001228	-0.000974	**0.000106**

「自然科学の統計学」では，デザイン行列 $\boldsymbol{X}$ を用いて $\boldsymbol{X}^T\boldsymbol{X}$，$\boldsymbol{X}^T\boldsymbol{Y}$，$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ の計算過程が示され，2 次式のパラメータの推定値 $\hat{\boldsymbol{\beta}}$ も $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ の計算結果として示されている．残念なのは，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ についての記載がないことである．そのためであろうか，2 次曲線の推定値の分散 $Var(\hat{y})$ についても例示がない．もちろん，一般論としてのパラメータの共分散行列については丁寧な説明があるが，それを用いた数値例は示されていない．

2 次式の推定値に対する分散（4 次式）を実際に求めるためには，行列計算なしには計算事例として示し難いことは確かである．他方，2 変数以上の回帰分析における推定値の分散は，直観的な意味付けができにくいのであるが，2 次式を含めて多項式回帰の推定値の分散を計

算し，その 95%信頼区間を求め，グラフ化することとの意義は明確であり，また入門的でもある．表 12.15 の結果を用いて，表 12.16 に示すように 2 次曲線の 95%信頼区間および個別データの 95%信頼区間の計算を行い，その結果を図 12.5 に示す．

表 12.16　2 次多項式の 95%信頼区間および予測区間

	I-8(%)	曇り点		――X'――					信頼区間		予測区間	
No.	x	y		切片	x'	x'^2	y^	Var(y^)	L95%	U95%	L95%	U95%
1	0	21.9		**1**	**-1**	**1**	**20.83**	**0.0839**	**20.21**	**21.44**	**19.79**	**21.86**
2	0	22.1		1	0	0	22.56	0.0394	22.14	22.98	21.63	23.50
3	0	22.8		1	1	1	24.16	0.0196	23.86	24.46	23.27	25.05
4	1	24.5		1	2	4	25.63	0.0141	25.37	25.88	24.75	26.50
5	2	26.0		1	3	9	26.95	0.0150	26.69	27.21	26.08	27.83
6	2	26.1		1	4	16	28.15	0.0171	27.87	28.42	27.27	29.03
7	3	26.8		1	5	25	29.20	0.0176	28.92	29.48	28.32	30.08
8	3	27.3		1	6	36	30.12	0.0162	29.85	30.39	29.24	31.00
9	4	28.2		1	7	49	30.91	0.0153	30.64	31.17	30.03	31.78
10	4	28.5		1	8	64	31.56	0.0199	31.26	31.85	30.67	32.44
11	5	28.9		1	9	81	32.07	0.0373	31.66	32.48	31.14	33.00
12	6	29.8		1	10	100	32.45	0.0775	31.86	33.04	31.42	33.47
13	6	30.0		1	11	121	32.69	0.1532	31.86	33.52	31.51	33.86
14	6	30.3										
15	7	30.4				$\hat{\beta}$			$\Sigma(\hat{\beta})=(X^TX)^{-1}\sigma^{\wedge 2}$			
16	8	31.4			切片	**22.5612**			**0.0394**	-0.0151	0.0012	
17	8	31.5			x	**1.6680**			-0.0151	**0.0098**	-0.0010	
18	9	31.8			x^2	**-0.0680**			0.0012	-0.0010	**0.0001**	
19	10	33.1		t(0.05,16)=		**2.1199**		$\sigma^{\wedge 2}$=	**0.1554**			

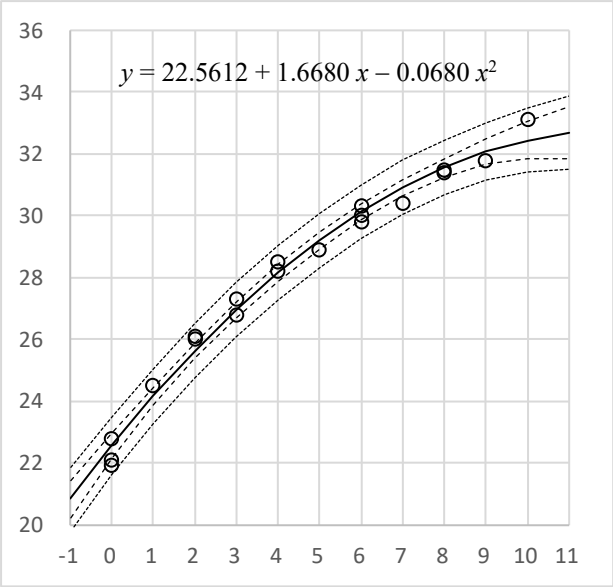


図 12.5　Excel の散布図を用いた 2 次式の推定曲線と 95%信頼区間および予測区間

「自然科学の統計学」の 38 ページに行列計算の結果が示されている．Excel の行列計算の計算式および結果を示すので，実際に手を動かして計算することが，95%信頼区間を求めるための練習となる．

X^TX			β		X^TY
19.0	84.0	550.0	β_0	=	531.4
84.0	**550.0**	4068.0	β_1		2536.1
550.0	4068.0	**32374.0**	β_2		16994.1

X^TX = Mmult(Transpose(Xの範囲), Xの範囲)

X^TY = Mmult(Transpose(Xの範囲), Yの範囲)

$(X^TX)^{-1}$		
0.253356	-0.097147	0.007903
-0.097147	**0.063004**	-0.006266
0.007903	-0.006266	**0.000684**

$(X^TX)^{-1}$ = Minverse(X^TXの範囲)

β^		$(X^TX)^{-1}$			X^TY
22.5612	=	**0.253356**	-0.097147	0.007903	531.4
1.6680		-0.097147	**0.063004**	-0.006266	2536.1
-0.0680		0.007903	-0.006266	**0.000684**	16994.1

$\hat{\beta}$ == Mmult($(X^TX)^{-1}$の範囲, X^TYの範囲)

σ^2
0.155412

$\hat{\sigma}^2$ = SumSq(Yの範囲 − Mmult(Xの範囲, $\hat{\beta}$の範囲)) / (19 − 3)

$\hat{y}$

$\Sigma(\beta$^$)=(X^TX)^{-1}\sigma$^2		
0.039375	-0.015098	0.001228
-0.015098	**0.009792**	-0.000974
0.001228	-0.000974	**0.000106**

$\Sigma(\hat{\beta})$ = ($(X^TX)^{-1}$の範囲) * $\hat{\sigma}^2$

$Var(y$^$)$		切片	x	x^2	$\Sigma(\beta$^$)=(X^TX)^{-1}\sigma$^2			x_1^T
0.0839	=	1	-1	1	**0.039375**	-0.015098	0.001228	1
					-0.015098	**0.009792**	-0.000974	-1
					0.001228	-0.000974	**0.000106**	1

$Var(\hat{y}_1)$ = Mmult(Mmult(x_1の範囲, $\Sigma(\hat{\beta})$の範囲), Transpose(x_1の範囲))

以下 略 （フィルハンドルでコピー）

95%信頼区間 = $\hat{y}_1 \pm t(0.05,16)\sqrt{Var(\hat{y}_1)}$ = **20.21** **21.44**

95%予測区間 = $\hat{y}_1 \pm t(0.05,16)\sqrt{Var(\hat{y}_1)+\hat{\sigma}^2}$ = **19.79** **21.86**

以下 略 （フィルハンドルでコピー）

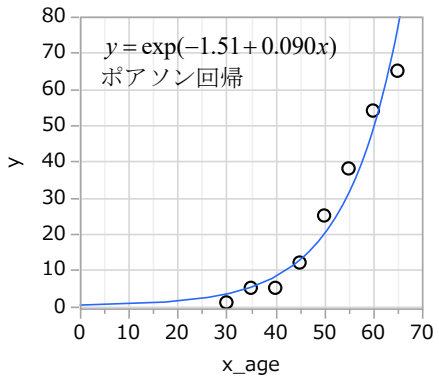
12.6.　対数リンクでのポアソン回帰の 95%信頼区間

第 1.5 節でオフセットを考慮した「冠動脈心疾患の死亡者数」の事例に対し，第 2.6 節では，オフセットを考慮した対数リンクのポアソン回帰について Excel によるニュートン・ラフソン法による最尤法を例示した．第 5.2 節では，同じ事例について反復重み付き回帰によるオフセット無しで対数リンクによるポアソン回帰を適用し，95%信頼区間を含むグラフを示し，さらに 2 次式によるポアソン回帰も例示した．どちらも Excel による例示となっている．

ここでは，オフセット無しの対数リンクに対するポアソン回帰として取り上げる．回帰パラメータは，統計ソフトでの計算結果を用いることを想定し，Excel でパラメータの共分散行列を計算し，95%信頼区間および 95%予測区間のグラフ表示の方法について示す．表 12.17 に示すように第 5.3 節で取り上げた冠動脈心疾患の死亡者数［ドブソン(2008)］を，オフセットなしの対数リンクの事例として取り上げる．

表 12.17　オーストラリアのある地方の冠動脈心疾患の死亡者数（表 5.8 再掲）

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



第 5.2 節では，対数リンクに対する反復重み付き回帰の入門とし，ポアソン回帰曲線の 95%信頼区間の求め方についても示したのであるが，ここでは，ポアソン回帰のパラメータがすでに得られていることを前提にし，事後的に 95%信頼区間を求める方法に焦点をあてる．

表 12.18 に示すように，ポアソン回帰係数 $\hat{\beta}_0 = -1.5078$，$\hat{\beta}_1 = 0.0899$ がパラメータの推定値として得られているとする．回帰曲線の推定値は，

$$\begin{aligned}\hat{y}_i &= \exp(\hat{\beta}_0 + \hat{\beta}_1 x_i) \\ &= \exp(-1.5078 + 0.0899 x_i) \\ &= 3.2831\end{aligned}$$

として，計算されている．

表 12.18　冠動脈心疾患の死亡者数（対数リンク）

	デザイン行列		死亡	回帰	対数		対数	対数		
	$\boldsymbol{X}$		者数	推定値	尤度	重み	推定値	分散	95%信頼区間	
i	$\boldsymbol{X}_0$	$\boldsymbol{X}_1$	$\boldsymbol{y}$	$\boldsymbol{y}$^	$\ln L_i$	$\boldsymbol{w}$^=$\boldsymbol{y}$^	$\ln \boldsymbol{y}$^	$Var(\ln \boldsymbol{y}$^)	L95%	U95%
1	1	30	1	3.2830	-2.0943	3.2830	1.1888	**0.0532**	2.0887	5.1603
2	1	35	5	5.1458	-1.7424	5.1458	1.6382	0.0372	3.5268	7.5081
3	1	40	5	8.0656	-2.4151	8.0656	2.0876	0.0243	5.9418	10.9486
4	1	45	12	12.6422	-2.1849	12.6422	2.5370	0.0147	9.9675	16.0346
5	1	50	25	19.8154	-3.1575	19.8154	2.9865	0.0083	16.5665	23.7014
6	1	55	38	31.0589	-3.4635	31.0589	3.4359	0.0052	26.9557	35.7866
7	1	60	54	48.6819	-3.1954	48.6819	3.8853	0.0053	42.1830	56.1820
8	1	65	65	76.3044	-3.8895	76.3044	4.3347	0.0087	63.5534	91.6138
	1	70		119.6002			4.7842	0.0153	93.8501	152.415
			β_0^=	**-1.5078**	**-22.1425**	$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^)=	**0.2178**	-0.0037		
			β_1^=	**0.0899**	$\ln L$		-0.0037	**0.0001**		
							$[(\boldsymbol{X}*\boldsymbol{w}$^$)^T \boldsymbol{X}]^{-1}$			

このパラメータの推定値は，Excel の関数 Poisson.dist() 関数で対数尤度 $\ln L$ を Excel のソルバーで最大化して求めたものであるが，後の手順では，パラメータの推定値のみを使う．

$$\ln L = \sum_i \ln(\text{Poisson.dist}(y_i\ \hat{y}_i, false)) = -22.1425$$

第 5 章の反復重み付き回帰で示したように，対数リンクの場合にパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を求めるためには，重み $\hat{w}_i = \hat{y}_i$ を対角要素とした行列 $\boldsymbol{W}$ としたデザイン行列の積和行列の逆数を計算して得られる．

1	1	1	1	1	1	1	1
30	35	40	45	50	55	60	65
			$\boldsymbol{X}^T$				

3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	**5.1**	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	**8.1**	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	**12.6**	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	**19.8**	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	**31.1**	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	**48.7**	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	**76.3**
			$\boldsymbol{W}$				

1	30
1	35
1	40
1	45
1	50
1	55
1	60
1	65
	$\boldsymbol{X}$

=

205.0	11750.1
11750.1	**688912.4**
$\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$	

0.2178	-0.0037
-0.0037	**0.0001**
$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)=(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$	

簡便的には，表 12.18 に示したように $[(\boldsymbol{X}*\hat{\boldsymbol{w}})^T\boldsymbol{X}]^{-1}$ として計算することができる．実際に確認して見ると，次に示すように同じ結果が得られる．このような技巧的な計算は Excel の行列関数にベクトルを対角化する関数，逆に行列の対角要素をベクトル化する関数がないためである．

1	30	*	3.2831	=	3.28	98.49
1	35		5.1460		5.15	180.11
1	40		8.0658		8.07	322.63
1	45		12.6425		12.64	568.91
1	50		19.8159		19.82	990.79
1	55		31.0596		31.06	1708.28
1	60		48.6830		48.68	2920.98
1	65		76.3062		76.31	4959.90
X			***w***		***X*w***	

3.28	5.15	8.07	12.64	19.82	31.06	48.68	76.31
98.49	180.11	322.63	568.91	990.79	1708.28	2920.98	4959.90
			$(\boldsymbol{X}*\boldsymbol{w})^T$				

1	30	=	**205.0** 11750.1
1	35		11750.1 **688912.4**
1	40		$(\boldsymbol{X}*\boldsymbol{w})^T\boldsymbol{X}$
1	45		
1	50		**0.2178** -0.0037
1	55		-0.0037 **0.0001**
1	60		$[(\boldsymbol{X}*\boldsymbol{w})^T\boldsymbol{X}]^{-1}$
1	65		$\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{})$
X			

パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を用いて，回帰直線の 95%信頼区間を求める．デザイン行列 $\boldsymbol{X}$ それぞれの i 行目ごとに， $\boldsymbol{x}_i = [x_{0,i} \quad x_{1,i}]$ としたとき $\hat{y}_i$ の分散は，次の 2 次形式で求められるので，

$$Var(\ln \hat{y}_i) = \boldsymbol{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \boldsymbol{x}_i^T$$

$\hat{y}_i$ の 95%信頼区間は，

$$(L95\%, U95\%) = \exp\left[\ln \hat{y}_i \pm 1.96\sqrt{Var(\ln \hat{y}_i)} \right]$$

で求められる．元のスケールは，対数の 95%信頼区間について指数を取って計算したものである．

共分散行列を用いた計算の実例を，$i = 1$ の場合について示す．まず，対数についての $Var(\hat{y}_i)$ は，2 次形式の計算法で，0.0532 が得られる．

$Var(y_1\hat{})$=	1	30	**0.2178**	-0.0037	1	=	**0.0532**
	$\boldsymbol{x}_{1,i}$		-0.0037	**0.0001**	30		
			$\boldsymbol{\Sigma}(\boldsymbol{\beta}\hat{}) = [(\boldsymbol{X}*\boldsymbol{w}\hat{})^T\boldsymbol{X}]^{-1}$		$\boldsymbol{x}_{1,i}^T$		

信頼区間の計算は，

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

で得られる．指数を取って元のスケールでは，

$$\hat{y}_1 = \exp(\ln \hat{y}_1) = \exp(1.1888) = 3.283$$

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

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

となる．これらの計算結果より，図 12.6 に 95%信頼区間を示す．

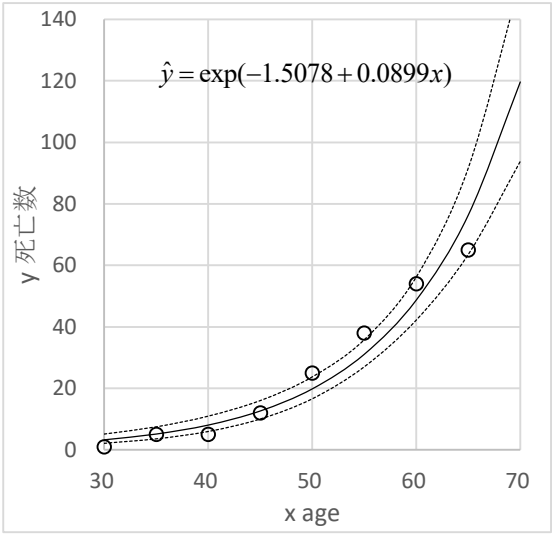


図 12.6　死亡数の 95%信頼区間

大変な計算と思われるかもしれなので，段階的な計算手順を示す．

手順 1）表 12.17 に示す冠動脈疾患の死亡者データに対し，使い慣れた統計ソフトで対数リンクのポアソン回帰を行い，回帰パラメータ $\hat{\beta}_0 = -1.5078$，$\hat{\beta}_1 = 0.0899$ を得る．

手順 2）推定値 $\ln \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_0 x_i$ を計算し，その指数を計算し，表 12.17 右に示す散布図に指数の線グラフを上書きする．Excel でも統計ソフトでも，このグラフを書けることが必須である．

手順 3）統計ソフトでパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ が得られない場合は，表 12.19 に示すように Excel で計算する．

共分散行列	
0.2178	-0.0037
-0.0037	**0.0001**

手順 4）推定値 $\ln \hat{y}_1$ の分散 $Var(\ln \hat{y}_1)$ を計算し，計算式をコピーする．

手順 5）95%信頼区間を計算する．

手順 6）95%信頼区間を散布図に上書きし，形式を整える．

表 12.19　対数リンクでの 95%信頼区間の計算

デザイン		回帰		推定値	重み=推定値			分散	95%信頼区間	
行列		パラメータ		$\ln \boldsymbol{y}$^	$\boldsymbol{y}$^=$\boldsymbol{w}$^	共分散行列		$Var(\ln \boldsymbol{y}$^)	L95%	U95%
1	30	**-1.5078**	=	1.1888	3.2831	**0.2178**	-0.0037	0.0532	2.0888	5.1604
1	35	**0.0899**		1.6382	5.1460	-0.0037	**0.0001**	0.0372	3.5269	7.5082
1	40			2.0876	8.0658			0.0243	5.9420	10.9489
1	45			2.5371	12.6425			0.0147	9.9677	16.0349
1	50			2.9865	19.8159			0.0083	16.5670	23.7019
1	55			3.4359	31.0596			0.0052	26.9563	35.7874
1	60			3.8853	48.6830			0.0053	42.1841	56.1832
1	65			4.3348	76.3062			0.0087	63.5550	91.6157
$\boldsymbol{X}$		$\boldsymbol{\beta}$		=$\boldsymbol{X\beta}$	=exp($\ln \boldsymbol{y}$^)	$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^)=		=$\boldsymbol{x}_i\boldsymbol{\Sigma}\boldsymbol{x}_i^T$	=exp($\ln \boldsymbol{y}$^±1.96*	
						$[(\boldsymbol{X}*\boldsymbol{w}$^$)^T\boldsymbol{X}]^{-1}$			sqrt($Var(\ln \boldsymbol{y}$^))	

Excel での計算は以下の通り.

$$\ln(\hat{\boldsymbol{y}}) = \text{Mmult}(\boldsymbol{X}\text{の範囲}, \hat{\boldsymbol{\beta}}\text{の範囲})$$

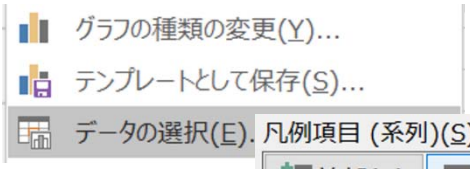
$$\hat{\boldsymbol{y}} = \exp(\ln(\hat{\boldsymbol{y}})\text{の範囲})$$

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = \text{Minverse}(\text{Mmult}(\text{Transpose}(\boldsymbol{X}\text{の範囲}*\hat{\boldsymbol{w}}\text{の範囲}), \boldsymbol{X}\text{の範囲}))$$

$$Var(\ln \hat{y}_i) = \text{Mmult}(\text{Mmult}(\boldsymbol{x}_i\text{の範囲}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\text{の範囲}), \text{Transpose}(\boldsymbol{x}_i))$$

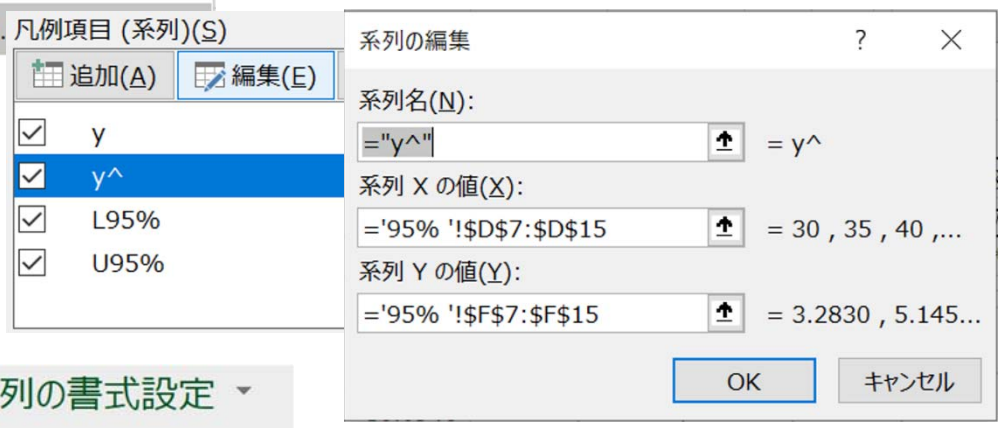
$$(L95\%, U95\%) = \exp(\ln \hat{y}_i \pm 1.96 * \text{sqrt}(Var(\ln \hat{y}_i))$$

Excel の散布図の活用のヒント

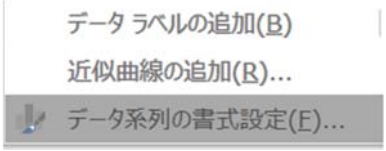


データの選択

追加➔編集➔X の選択，Y の選択



データ系列の書式設定



12.7.　オフセットを含むポアソン回帰の各種の 95%信頼区間

第 3.6 節では，喫煙習慣による年齢階層ごとの冠動脈疾患による死亡データについて，各種の統計モデルをあてはめ，尤度比検定を用いて探索的な解析の手順を示した．そのために，パラメータの共分散行列を用いたが，95%信頼区間については示さなかった．データを表 12.20 に再掲する［ドブソン(2008)］.

表 12.20　年齢階層毎の喫煙習慣による冠動脈心疾患による死亡数（表 3.32 再掲）

年齢		非喫煙者（$x_{smoke}=0$）			喫煙者（$x_{smoke}=1$）		
範囲	歳	死亡	人年	10万人比	死亡	人年	10万人比
35-44	40	2	18,790	10.6	32	52,407	61.1
45-54	50	12	10,673	112.4	104	43,248	240.5
55-64	60	28	5,710	490.4	206	28,612	720.0
65-74	70	28	2,585	1083.2	186	12,663	1468.8
75-84	80	31	1,462	2120.4	102	5,317	1918.4

第 3.6 節では，年齢階層を無視した（非喫煙・喫煙）のポアソン回帰による 2 群間比較に引き続き，年齢をモデルに加えたモデル，年齢の 2 乗の項を加えたモデル，さらに，年齢と喫煙習慣の交互作用，年齢の 2 乗と喫煙習慣の交互作用をモデルに加え，尤度比検定によるモデル選択を行った.

2 次式のあてはめ

基本の主効果モデル

$$y_i = n_i \exp(\beta_0 + \beta_1 x_{smoke} + \beta_2 x_{age}) + \varepsilon_i$$

に対し，年齢の 2 乗と年齢と喫煙習慣の交互作用を加えたモデル

$$y_i = n_i \exp(\beta_0 + \beta_1 x_{smoke} + \beta_2 x_{age} + \beta_3 x^2_{(age/10)} + \beta_4 x_{smoke\times age}) + \varepsilon_i$$

が尤度比検定によって選択された.

図 12.7 左の基本モデルのあてはめは，年齢が高くなるにつれて，死亡者数が指数関数的に増大し続けることになり，モデルとしては不適切である．さらに，図 12.7 右の対数目盛での直線に対し，プロットされた点は，上に凸となっており，あてはめは支持されない.

直線のあてはめが支持されない場合は，便宜的な方法ではあるが，年齢について 2 乗の項を加えて 2 次式をあてはめて様子をみることになる．図 12.8 に示すように，×印の喫煙者へのあてはまりは良くなったが，〇印の非喫煙者の 80 歳代に対しては，過小評価となっていてモデルとしては不十分である.

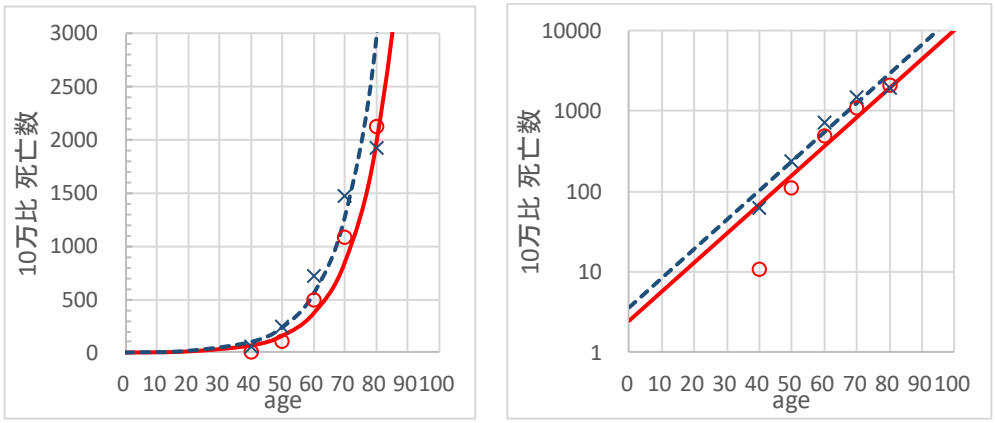


図 12.7　2 本のポアソン回帰直線（図 3.4 再掲）

○非喫煙：$\hat{y}_i = 100{,}000 \times \exp(-10.6260 + 0.4064 \times 0 + 0.0836 \times x_{age})$

×喫煙：　$\hat{y}_i = 100{,}000 \times \exp(-10.6260 + 0.4064 \times 1 + 0.0836 \times x_{age})$

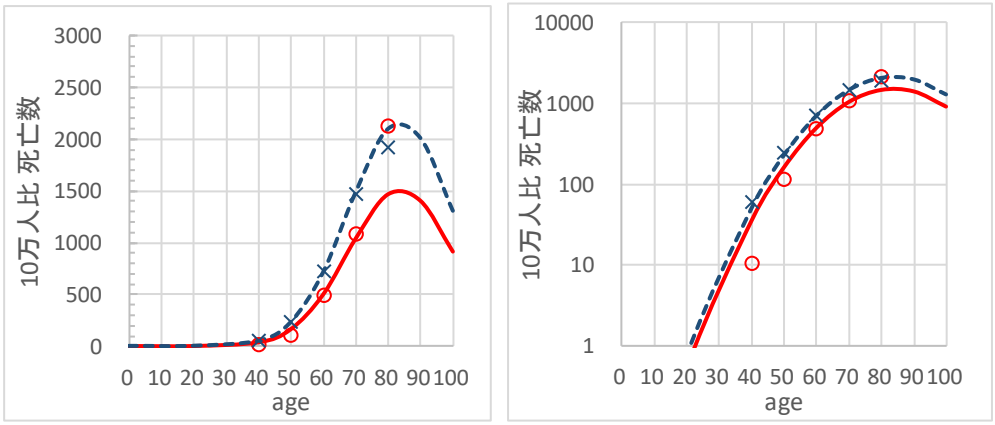


図 12.8　2 本の 2 次曲線（図 3.5 再掲）

○非喫煙：$\hat{y}_i = 100{,}000 \times \exp(-17.8583 + 0.3548 \times 0 + 0.3285 x_{age} - 0.1942 x^2_{(age/10)})$

×喫煙：　$\hat{y}_i = 100{,}000 \times \exp(-17.8583 + 0.3548 \times 1 + 0.3285 x_{age} - 0.1942 x^2_{(age/10)})$

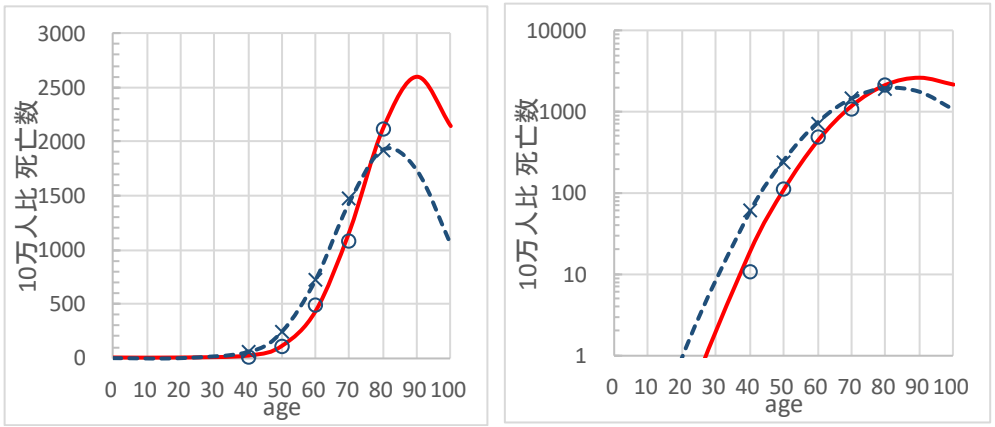


図 12.9　交互作用を含む 2 本の 2 次曲線（図 3.6 再掲）

○非喫煙：$\hat{y}_i = 100{,}000 \times \exp(-19.7003 + 2.3636 \times 0 + 0.3563 x_{age} - 0.1977 x^2_{(age/10)} - 0.0308 \times 0)$

×喫煙：　$\hat{y}_i = 100{,}000 \times \exp(-19.7003 + 2.3636 \times 1 + 0.3563 x_{age} - 0.1977 x^2_{(age/10)} - 0.0308 \times 1 \times x_{age})$

図 12.9 に示すように，喫煙習慣と年齢の交互作用を加えたモデルは，良くあてはまっていると判断される．喫煙習慣と年齢の交互作用を含むオフセット付き対数リンクのポアソン回帰の結果を表 12.21 に示す．推定値 $\hat{y}_i$ は，

$$\hat{y}_i = n_i \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{smoke} + \hat{\beta}_2 x_{age} + \hat{\beta}_3 x^2_{(age/10)} + \hat{\beta}_4 x_{smoke\times age})$$

として求められている．対数尤度 $\ln L_i$ は，

$$\ln L_i = \ln\left[\text{Poisson.dist}(y_i, \hat{y}_i, false)\right]$$

Excel のポアソン関数で計算し，対数尤度 $\ln L = \Sigma_i \ln L_i$ を最大化するようにパラメータ $\hat{\beta}_0$，$\hat{\beta}_1$，$\hat{\beta}_2$，$\hat{\beta}_3$，$\hat{\beta}_4$ に適当な初期値を入れて，Excel のソルバーで変化させた結果を示す．JMP での結果と微妙に異なるが，$\ln L = -28.3517$ と小数点以下 4 桁までで同じ結果となっている．JMP の方が小数点以下 5 桁目で大きくなっているので，JMP で推定したパラメータを以後使うことにする．

表 12.21　交互作用モデル（推定値は JMP による）

	切片	喫煙	年齢	年齢^2	交互.	死亡	人年		推定値	対数尤度		
i	x_0	x_1	x_2	x_3	$x_1 \times x_2$	y	n	10万y	y^=w^	$\ln Li$		JMP
1	1	0	40	16	0	2	18,790	10.6	3.4	-1.65171	$\hat{\beta}_0$=	**-19.7003**
2	1	0	50	25	0	12	10,673	112.4	11.5	-2.17732	$\hat{\beta}_1$=	**2.3636**
3	1	0	60	36	0	28	5,710	490.4	24.7	-2.79350	$\hat{\beta}_2$=	**0.3563**
4	1	0	70	49	0	28	2,585	1,083.2	30.2	-2.67231	$\hat{\beta}_3$=	**-0.1977**
5	1	0	80	64	0	31	1,462	2,120.4	31.1	-2.63870	$\hat{\beta}_4$=	**-0.0308**
6	1	1	40	16	40	32	52,407	61.1	29.6	-2.75042		Excel
7	1	1	50	25	50	104	43,248	240.5	106.8	-3.27928		**-19.6978**
8	1	1	60	36	60	206	28,612	720.0	208.2	-3.59493		**2.3677**
9	1	1	70	49	70	186	12,663	1,468.8	182.8	-3.55962		**0.3561**
10	1	1	80	64	80	102	5,317	1,918.4	102.6	-3.23387		**-0.1975**
			$x_3=(x_2/10)^2$						$\ln L$ =	**-28.35166**		**-0.0308**

2 次式の 95%信頼区間

図 12.9 に示した喫煙習慣と年齢の交互作用を含む 10 万人比での推定死亡曲線に，95%信頼区間を重ね書きするために，パラメータの共分散行列を計算する．対数リンクの場合の重み $\hat{w}_i$ は，推定値 $\hat{y}_i$ に等しいことは，前節で示した．求めるパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ は，

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = [(\boldsymbol{X} * \hat{\boldsymbol{w}})^T \boldsymbol{X}]^{-1}$$

として得られ，表 12.22 に結果を示す．

得られたパラメータの推定値 $\hat{\boldsymbol{\beta}}$ を用いて，表 12.23 に示すように 0 歳から 100 歳までの 1 人あたりの対数死亡者の推定値 $\ln \hat{y}_i$ を

$$\ln \hat{y}_i = \boldsymbol{x}_i \hat{\boldsymbol{\beta}}$$

表 12.22　パラメータの共分散行列

		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})=$	$\hat{\beta}_0$	**1.5712**	-0.4352	-0.0439	0.0300	0.0064
	$\hat{\beta}_1$	-0.4352	**0.4306**	0.0074	-0.0016	-0.0063
	$\hat{\beta}_2$	-0.0439	0.0074	**0.0013**	-0.0010	-0.0001
	$\hat{\beta}_3$	0.0300	-0.0016	-0.0010	**0.0007**	0.0000
	$\hat{\beta}_4$	0.0064	-0.0063	-0.0001	0.0000	**0.0001**

=Minverse(Mmult(Transpose($\boldsymbol{X}$の範囲*$\boldsymbol{w}$^の範囲), $\boldsymbol{X}$の範囲))

で求める．その分散 $Var(\ln \hat{y}_i)$ は，

$$Var(\ln \hat{y}_i) = \boldsymbol{x}_i \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \boldsymbol{x}_i^T$$

である．95%信頼区間は，

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

で求められる．これらを 10 万人比にするために，

$$\hat{y}_i = 100{,}000 \exp(\ln \hat{y}_i)$$

$$L95\% = 100{,}000 \exp(\ln L95\%)$$

$$U95\% = 100{,}000 \exp(\ln U95\%)$$

で換算する．

表 12.23　10 万人比における 95%信頼区間

切片	喫煙	年齢	年齢^2	交互	——— 1人当たりの推定値 対数———				———10万人比———		
x_0	x_1	x_2	x_3	$x_1 \times x_2$	$\ln y$^	$Var(\ln y$^$)$	$L95\%$	$U95\%$	y^	$L95\%$	$U95\%$
1	0	0	0	0	-19.7003	1.5712	-22.1571	-17.2435	0.0	0.0	0.0
1	0	20	4	0	-13.3659	0.4393	-14.6650	-12.0668	0.2	0.0	0.6
1	0	40	16	0	-8.6130	0.0851	-9.1847	-8.0412	18.2	10.3	32.2
1	0	50	25	0	-6.8295	0.0342	-7.1920	-6.4670	108.1	75.3	155.4
1	0	60	36	0	-5.4414	0.0152	-5.6828	-5.2001	433.3	340.4	551.6
1	0	70	49	0	-4.4487	0.0112	-4.6564	-4.2409	1169.4	950.0	1439.4
1	0	80	64	0	-3.8513	0.0237	-4.1527	-3.5499	2125.2	1572.2	2872.9
1	0	90	81	0	-3.6492	0.0716	-4.1738	-3.1247	2601.1	1539.4	4395.0
1	0	100	100	0	-3.8426	0.1923	-4.7021	-2.9830	2143.9	907.6	5064.2
1	1	0	0	0	-17.3367	1.1314	-19.4215	-15.2518	0.0	0.0	0.0
1	1	20	4	20	-11.6174	0.2408	-12.5791	-10.6557	0.9	0.3	2.4
1	1	40	16	40	-7.4795	0.0204	-7.7592	-7.1999	56.5	42.7	74.7
1	1	50	25	50	-6.0036	0.0042	-6.1302	-5.8771	247.0	217.6	280.3
1	1	60	36	60	-4.9231	0.0027	-5.0249	-4.8213	727.7	657.2	805.6
1	1	70	49	70	-4.2379	0.0024	-4.3348	-4.1410	1443.8	1310.4	1590.8
1	1	80	64	80	-3.9481	0.0079	-4.1221	-3.7740	1929.2	1621.0	2296.0
1	1	90	81	90	-4.0536	0.0415	-4.4527	-3.6544	1736.0	1164.7	2587.6
1	1	100	100	100	-4.5544	0.1436	-5.2972	-3.8117	1052.0	500.6	2211.2

表 12.21 で求めた 10 万人比 y_i，表 12.23 で求めた 10 万人比の $\hat{y}_i$，$L95\%$，$U95\%$を非喫煙者に対しては図 12.10 に，喫煙者に対しては図 12.11 に示す．

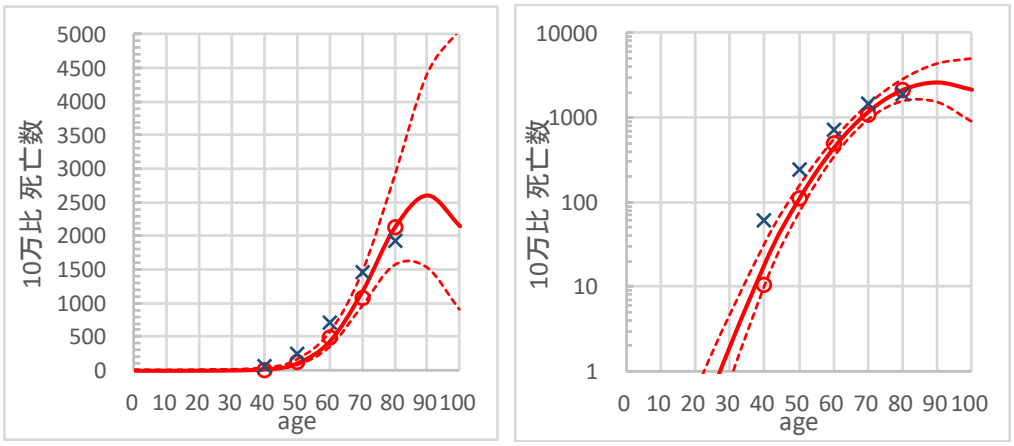


図 12.10　非喫煙者に対する 10 万人比での 95%信頼区間

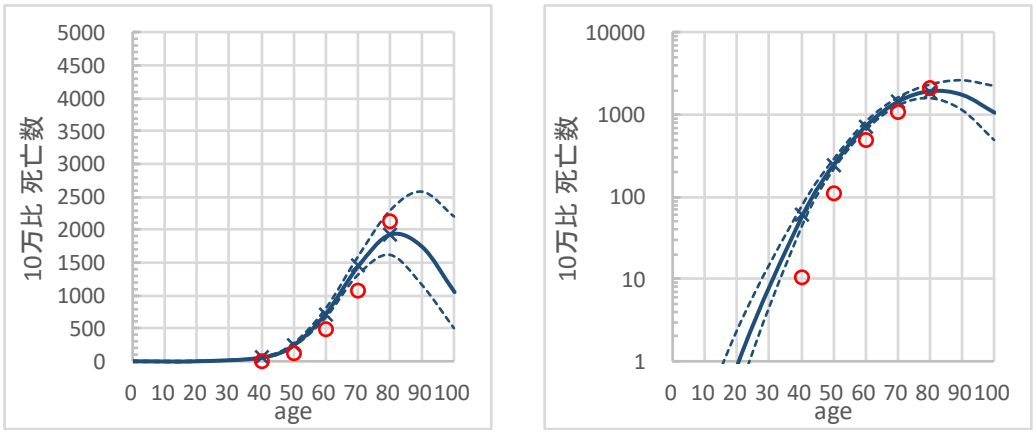


図 12.11　喫煙者に対する 10 万人比での 95%信頼区間

回帰分析の推定値および 95%信頼区間について，あえて外挿を行い対数リンクでの 2 次曲線のあてはめの形状の認識を行った．2 次曲線のあてはめは，対数目盛上で行っているので，指数を取って元に戻した結果も示した．対数目盛上で 80 歳以上の外挿の落ち込み具合は，少ないように見えても，元の目盛り上では，落ち込みが激しい．逆に，40 歳以下の場合は，指数を取ることにより，きれいに 0 人に収束していることが読み取れる．

上限がある場合のシグモイド曲線のあてはめ

2 次曲線のあてはめは，1 次直線のあてはめが適しているかの検討のためにであって，統計モデルとしては，便宜的な方法である．第 2.6 節では，死亡率について上限を新たな変数としたロジスティック回帰でシグモイド曲線をあてはめる方法を示した．この事例であれば，喫煙者と非喫煙者に共通の死亡率を新たな変数とし，形状が同じで，位置が異なるシグモイド

曲線をあてはめ，喫煙者の位置パラメータと非喫煙者の位置パラメータの違いを検討するのが本質的な解析方法である．

表 12.24　死亡率に上限を持つ 2 本のロジスティック回帰の同時あてはめ

					$\hat{\beta_0}$=	**-10.2556**		
					$\hat{\beta_1}$=	**0.7578**		
					$\hat{\beta_2}$=	**0.1480**		対数尤度
					Umax=	**0.0212**	ln*L* =	**-29.7838**
切片	喫煙	年齢	死亡	人年	死亡率	推定値	二項分布	対数尤度
x_0	x_1	x_2	y	n	p	π^	P	lnL_i
1	0	0				0.0000		
1	0	20				0.0000		
1	0	40	2	18,790	0.0001	0.0003	0.0770	-2.5640
1	0	50	12	10,673	0.0011	0.0012	0.1140	-2.1715
1	0	60	28	5,710	0.0049	0.0043	0.0585	-2.8389
1	0	70	28	2,585	0.0108	0.0112	0.0747	-2.5948
1	0	80	31	1,462	0.0212	0.0176	0.0432	-3.1428
1	0	90				0.0203		
1	0	100				0.0210		
1	1	0				0.0000		
1	1	20				0.0000		
1	1	40	32	52,407	0.0006	0.0006	0.0667	-2.7072
1	1	50	104	43,248	0.0024	0.0023	0.0364	-3.3122
1	1	60	206	28,612	0.0072	0.0074	0.0252	-3.6792
1	1	70	186	12,663	0.0147	0.0149	0.0288	-3.5465
1	1	80	102	5,317	0.0192	0.0194	0.0397	-3.2267
1	1	90				0.0208		
1	1	100				0.0211		

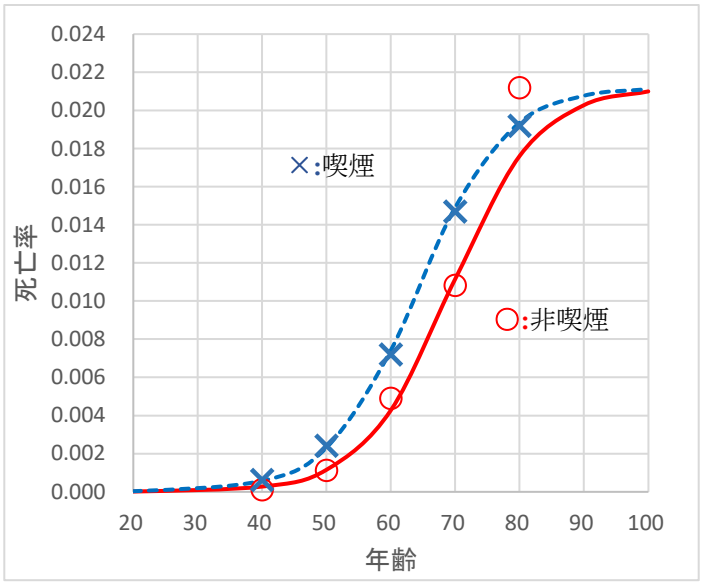


図 12.12　喫煙者と非喫煙者に対する上限があるシグモイド曲線の同時あてはめ

第 12 章　文 献 索　引

奥野・久米・芳賀・吉沢 著(1981) - 多変量解析法 改訂版 390
新村(1983a) - 行列計算による重回帰分析(1) 400
新村(1983b) - 行列計算による重回帰分析(2) 400
新村(1983c) - 重回帰分析における掃き出し演算子 400
東京大学教養学部統計学教室編(1992) - 自然科学の統計学 407
ドブソン著, 田中・森川・山中・富田 訳(2008) - 一般化線形モデル入門, 原著 第2版 410, 415
芳賀(2009) - 医薬品開発のための統計解析 第2部 実験計画法 401

第 12 章　索　引

第 12 章　解析用ファイル一覧

16 KB	第12章02_iris	JMP Data Table
39 KB	第12章02_iris_相関行列	Microsoft Excel ワークシート
22 KB	第12章03_ガラス工程_偏差平方和ベース	Microsoft Excel ワークシート
5 KB	第12章04_ガラス工程	JMP Data Table
32 KB	第12章04_ガラス工程_デザイン行列ベース	Microsoft Excel ワークシート
44 KB	第12章05_芳賀_2次回帰	Microsoft Excel ワークシート
5 KB	第12章05_芳賀_2次式	JMP Data Table
173 KB	第12章05_芳賀2次回帰−DE改2- 1 因子(量)	Microsoft Excel マクロ有効ワークシート
114 KB	第12章05_2次回帰_自然科学の統計学	Microsoft Excel ワークシート
3 KB	第12章06_冠動脈心疾患	JMP Data Table
75 KB	第12章06_冠動脈心疾患	Microsoft Excel ワークシート
4 KB	第12章07_タバコと冠動脈心疾患	JMP Data Table
76 KB	第12章07_タバコと冠動脈心疾患	Microsoft Excel ワークシート

第 9 回 続高橋セミナー

最尤法によるポアソン回帰分析入門＜12 章＞

第 12 章　パラメータの共分散行列の活用

BioStat 研究所（株）

〒105-0014 東京都 港区 芝 1-12－3 の 1005

2020 年 7 月 13 日　　高橋 行雄