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 6 月 5 日

第 4 章　デザイン行列を用いた回帰分析入門

ポアソン分布に従うカウント･データの解析を統計ソフトで行う場合に，2 群比較でも多群比較でも，2 本の回帰直線の場合でも，何らかのデザイン行列を設定し，ポアソン回帰で解析する方法を第 3 章で例示してきた．そこで，Excel の行列関数の使い方の入門を兼ね，デザイン行列を活用した回帰分析に必要な行列計算の方法を伝統的なシグマを用いた計算方法と対比して示す．さらに，デザイン行列を用いた回帰分析から導出されるパラメータの共分散行列を活用した回帰直線の 95%信頼区間，個別データの 95%信頼区間の導出方法を示す．逆推定値に対する 95%信頼区間の推定についても伝統的な偏差平方和による計算方法に代え，パラメータの共分散行列を活用したデルタ法による近似 95%信頼区間の導出法，正確な 95%信頼区間および個別データに対応するの正確な 95%信頼区間の導出法について示す．

第 4 章　目 次

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

4. デザイン行列を用いた回帰分析入門

ポアソン分布に従うカウント・データの解析を統計ソフトで行う場合に，2 群比較でも多群比較でも，2 本の回帰直線の場合でも，何らかのデザイン行列を設定し，ポアソン回帰で解析する方法を第 3 章で例示してきた．そこで，Excel の行列関数の使い方の入門を兼ね，デザイン行列を活用した回帰分析に必要な行列計算の方法を伝統的なシグマを用いた計算方法と対比して示す．さらに，デザイン行列を用いた回帰分析から導出されるパラメータの共分散行列を活用した回帰直線の 95%信頼区間，個別データの 95%信頼区間の導出方法を示す．逆推定値に対する 95%信頼区間の推定についても伝統的な偏差平方和による計算方法に代え，パラメータの共分散行列を活用したデルタ法による近似 95%信頼区間の導出法，正確な 95%信頼区間および個別データに対応するの正確な 95%信頼区間の導出法について示す．

4.1. Excel によるデザイン行列を用いた行列計算

多くの統計解析の教科書で，説明変数を X，反応変数を Y とする回帰分析が取り上げられている．そのほとんどが，シグマを用いた偏差平方和をベースにした計算法で説明されている．そして，推定された回帰直線の 95%信頼区間は，これこれの式で与えられるとの記述に遭遇する．ポアソン回帰の場合は，どうなるのだろうか．どちらの場合でも共通な方法は，あるのだろうか．幸い，デザイン行列をベースにした計算方法は，各種の回帰直線の 95%信頼区間の計算方法の基本は同じであることは，これまでも例示してきた．

ドレーパ・スミス著，中村訳（1968）「応用回帰分析」の第 1 章には，シグマを使った回帰分析，第 2 章には，同じデータについてデザイン行列を用いた計算法が丁寧に例示され，重回帰分析へ橋渡しがなされている．さらに，Draper and Smith（1998），Applied Regrettion Anarysiis 3rd ed. も参考にし，シグマを使った計算との関係を示すことにより，シグマを用いた計算になれ親しんだ人達にデザイン行列を用いた解析の利便性について示す．用いるデータは，第 1.4 節のポアソン回帰の導入で用いた人工データとし，Excel の行列関数による計算法を主体にする．

シグマが出てきただけで読み飛ばしたくなるような人たちの気持ちは，その式を見ても実際のデータで自ら計算を行う意欲がわかないからだと推測する．ましてや，行列の計算式が

出てきたときに，実際のデータで計算することができないものに対し，興味を持つことができないと思われる．そこで，統計ソフトの使い方，結果の見方さえ習得すれば十分だと思われている人達を念頭にして，Excel の基本的な計算機能を用い，行列計算による回帰分析を丁寧に解説する．

デザイン行列を用いた回帰式の表記

反応変数 Y_i の列ベクトルを $\boldsymbol{Y}$ とし，説明変数のデザイン行列を $\boldsymbol{X}$，推定したいパラメータをベクトル $\boldsymbol{\beta}$，誤差ベクトルを $\boldsymbol{\varepsilon}$ とする．データは，表 1.6 の人工データである［ドブソン（2008)］．まず，Excel シートのある場所に，これと同じものを再現してもらいたい．ここに示したのは，Excel シート上の該当部分をコピーして，ワード上で「図（拡張メタファイル）」によりペーストしたものである．

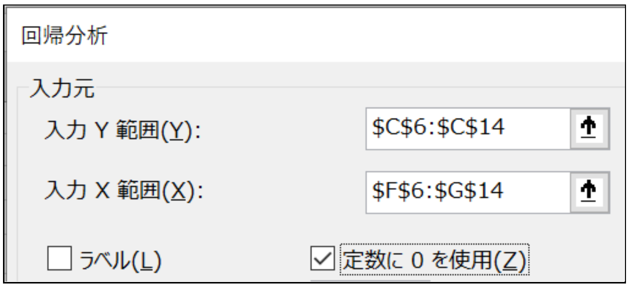
				$\boldsymbol{X}_0$	$\boldsymbol{X}_1$				
$\boldsymbol{Y}$ =	2		$\boldsymbol{X}$ =	1	-1	$\boldsymbol{\beta}$ =	β_0	$\boldsymbol{\varepsilon}$=	ε_1
	3			1	-1		β_1		ε_2
	6			1	0				ε_3
	7			1	0				ε_4
	8			1	0				ε_5
	9			1	0				ε_6
	10			1	1				ε_7
	12			1	1				ε_8
	15			1	1				ε_9
	9×1			9×2			2×1		9×1

フォントのサイズは 10 ポイント，半角英数字のフォントは Times New Roman とする．ギリシャ文字 β は「ベータ」と入力すると変換され，フォントを Times New Roman に変更してイタリックにすると「β」となる．「いぷしろん」で「ε」を生成し，「ε1」の「1」を「セル書式の設定」で「下付き」にする．このように統計で使うギリシャ文字を Excel で扱えるようになることから初めてもらいたい．「デザイン行列 $\boldsymbol{X}$」の「$\boldsymbol{X}$」は，「X」を太文字（B）とし，さらにイタリック（ *I* ）としたものである．

デザイン行列 $\boldsymbol{X}$ の一般的な表記は，角括弧 [・・・] あるいは括弧（・・・）で挟むのであるが，Excel シート上では表記しづらいので，「太い外枠 $\boxed{・・・}$」で矩形データ全体を囲む表記を用いる．デザイン行列 $\boldsymbol{X}$ は，表 1.6 の列方向の人工のデータを表 1.8 の行方向の形式で示したもので，大きさが 9 行×2 列で，9 行×1 の列ベクトル $\boldsymbol{X}_0$ と $\boldsymbol{X}_1$ を並べたものである．ベクトル $\boldsymbol{X}_0$ の値を全て 1 としているのは，回帰直線の Y 切片を推定するためである．列ベクトル $\boldsymbol{\beta}$ は，推定したい切片「β_0」と傾き「β_1」である．

多くの回帰分析の統計ソフトでは，モデル式で，説明変数 $\boldsymbol{X}_1$ のみを与え，切片を得るための変数 $\boldsymbol{X}_0$ の設定を必要としない．原点を通る回帰直線を求めたい場合には，切片を含まない 9 行×1 列の列ベクトル $\boldsymbol{X}_1$ のみのデザイン行列であり，統計ソフトでは，「切片を含めない」などのオプションを設定する．Excel のアドインで提供されている「データ分析ツール」の一つである「回帰分析」では，次に示すように，

Excel
➔ データ分析
➔ 回帰分析



「定数に 0 を使用」オプションを選択する．デザイン行列は，何を推定したいかにより，変幻自在であることは，第 3 章で詳細に示した．通常の回帰分析では，「定数に 0 を使用」オプションを選択せず，列ベクトル $\boldsymbol{X}_1$ の範囲のみを設定する．そして，列ベクトル $\boldsymbol{Y}$ の範囲を設定して，回帰分析を行なう．このように，Excel でも他の統計ソフトでも回帰分析では，切片を含めないことが暗黙の前提である．そのため，原点を通る回帰直線のあてはめのために「切片を含めない」とのオプションが必要となる．

列ベクトル $\boldsymbol{Y}$，デザイン行列 $\boldsymbol{X}$，列ベクトル $\boldsymbol{\beta}$，列ベクトル $\boldsymbol{\varepsilon}$ を用いた式

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{4.1}$$

は，何を意味するのだろうか．行列 $\boldsymbol{X}$ と列ベクトル $\boldsymbol{\beta}$ の積 $\boldsymbol{X\beta}$ は，デザイン行列 $\boldsymbol{X}$ のある行と，列ベクトル $\boldsymbol{\beta}$ のセル同士を順番に掛けて足した「積和」の計算を意味する．これをデザイン行列 $\boldsymbol{X}$ の 1 行目から 9 行目まで繰り返してして，大きさが 9×1 の新たな列ベクトルが生成される．右辺の $\boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ は，大きさが 9×1 の列ベクトル同士の足し算で，同じ行同士の足し算となり，大きさが 9×1 の列ベクトルとなる．

	$\boldsymbol{X\beta}$		$\boldsymbol{\varepsilon}$		$\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$	
$\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$ =	$1\beta_0-1\beta_1$	+	ε_1	=	$1\beta_0-1\beta_1$	$+\varepsilon_1$
	$1\beta_0-1\beta_1$		ε_2		$1\beta_0-1\beta_1$	$+\varepsilon_2$
	$1\beta_0+0\beta_1$		ε_3		$1\beta_0+0\beta_1$	$+\varepsilon_3$
	$1\beta_0+0\beta_1$		ε_4		$1\beta_0+0\beta_1$	$+\varepsilon_4$
	$1\beta_0+0\beta_1$		ε_5		$1\beta_0+0\beta_1$	$+\varepsilon_5$
	$1\beta_0+0\beta_1$		ε_6		$1\beta_0+0\beta_1$	$+\varepsilon_6$
	$1\beta_0+1\beta_1$		ε_7		$1\beta_0+1\beta_1$	$+\varepsilon_7$
	$1\beta_0+1\beta_1$		ε_8		$1\beta_0+1\beta_1$	$+\varepsilon_8$
	$1\beta_0+1\beta_1$		ε_9		$1\beta_0+1\beta_1$	$+\varepsilon_9$

列ベクトル $\boldsymbol{Y}$ と $\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$ を等号で結んだ式

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$$

は，$\boldsymbol{Y}$ の行方向に展開して

$$\begin{aligned} Y_1 &= 1\beta_0 - 1\beta_1 + \varepsilon_1 \\ Y_2 &= 1\beta_0 - 1\beta_1 + \varepsilon_2 \\ &\ \vdots \\ Y_9 &= 1\beta_0 + 1\beta_1 + \varepsilon_9 \end{aligned} \tag{4.2}$$

となる．また，添え字を使った式

$$Y_i=\beta_0 X_{0,i}+\beta_1 X_{1,i}+\varepsilon_i \qquad (i=1,\ 2,\dots,\ 9) \tag{4.3}$$

としても，同じ回帰モデルである．デザイン行列を使った表記は，最も簡潔な表記となっている．その意味することを Excel で表記した矩形データと対比し，連想することにより行列計算が身近なものとなることを期待する．

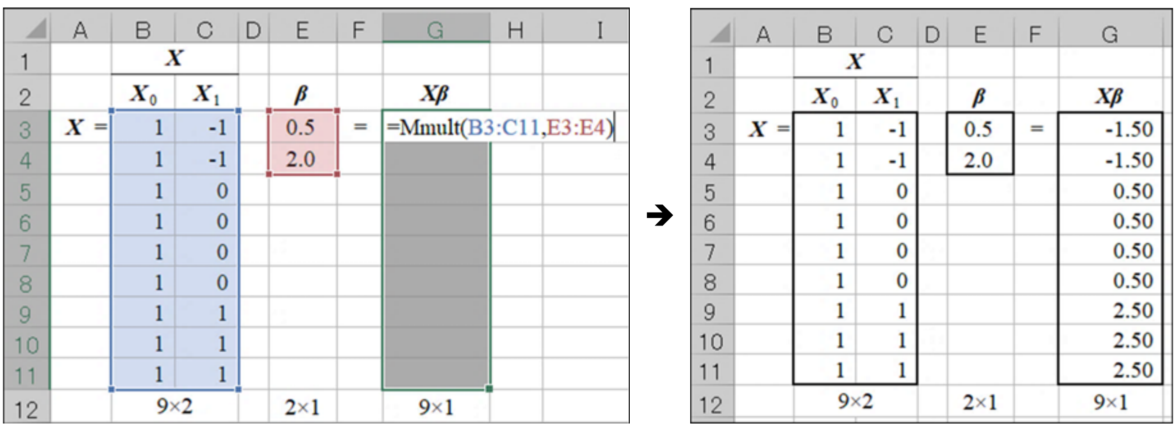
行列計算の実際

Excel シート上の列ベクトル $\boldsymbol{\beta}$ に適当な数値を入れて，実際に $\boldsymbol{X\beta}$ の計算にチャレンジしてみよう．手順は，以下に示すように，

1） $\boldsymbol{X\beta}$ の計算結果となる 9×1 の矩形を枠線で囲む．

2）9×1 の矩形を選択し，最初の行に行列の積の関数で「=Mmult($\boldsymbol{X}$ の範囲を選択，$\boldsymbol{\beta}$ の範囲の選択)」のように入力する．

3)「コントロールキー」と「シフトキー」同時に押しながら「エンター」すると行列の積の計算が行なわれる．



	A	B	C	D	E	F	G	H	I
1		$\boldsymbol{X}$							
2		$\boldsymbol{X}_0$	$\boldsymbol{X}_1$		$\boldsymbol{\beta}$		$\boldsymbol{X\beta}$		
3	$\boldsymbol{X}$ =	1	-1		0.5	=	=Mmult(B3:C11,E3:E4)		
4		1	-1		2.0				
5		1	0						
6		1	0						
7		1	0						
8		1	0						
9		1	1						
10		1	1						
11		1	1						
12		9×2			2×1		9×1		

➔

	A	B	C	D	E	F	G
1		$\boldsymbol{X}$					
2		$\boldsymbol{X}_0$	$\boldsymbol{X}_1$		$\boldsymbol{\beta}$		$\boldsymbol{X\beta}$
3	$\boldsymbol{X}$ =	1	-1		0.5	=	-1.50
4		1	-1		2.0		-1.50
5		1	0				0.50
6		1	0				0.50
7		1	0				0.50
8		1	0				0.50
9		1	1				2.50
10		1	1				2.50
11		1	1				2.50
12		9×2			2×1		9×1

デザイン行列の転置

デザイン行列 $\boldsymbol{X}$ 同志の掛け算をしたい．掛け算は，左側の行列 $\boldsymbol{A}$ の「行」と右側の行列 $\boldsymbol{B}$ の「列」を順番に掛けて加えた和(積和)の行列として定義されている．行列の積和の計算は，“行”方向と“列”方向であって，“列”方向と“行”の“列行”ではなく，あくまで“行・列”の順番である．

$\boldsymbol{A}$				$\boldsymbol{B}$				$\boldsymbol{AB}$	
→				↓			=	$A_{11}B_{11}+A_{12}B_{21}+A_{13}B_{31}$	
→								$A_{21}B_{11}+A_{22}B_{21}+A_{23}B_{31}$	
→								$A_{31}B_{11}+A_{32}B_{21}+A_{33}B_{31}$	

(9 行 2 列)の $\boldsymbol{X}$ と（9 行 2 列)の $\boldsymbol{X}$ のままだと積 $\boldsymbol{XX}$ は，（2 列 vs. 9 行）と列と行の数が異なり積和の計算ができない．そこで，最初の $\boldsymbol{X}$ について行と列を入れ替え，（2 行 9 列)の $\boldsymbol{X}^T$ (転置行列)とする．すると，積 $\boldsymbol{X}^T\boldsymbol{X}$ は，（9 列 vs. 9 行）と内側の数が一致し積和の計算ができる．このように，行列の積が成り立つのは， $\boldsymbol{X}^T$（2 行 9 列）と $\boldsymbol{X}$（9 行 2 列）のように隣り合う行列の内側が，9 列と 9 行のように大きさが完全に一致する必要がある．行列の積の結果は，外側の（2 行 2 列）の大きさとなる．

デザイン行列の積和

行と列の入れ替えは，転置（Transpose）といい， $\boldsymbol{X}^T$ のように表記する．なお， $\boldsymbol{X}'$ あるいは ${}^t\boldsymbol{X}$ と表記する場合もあり，様々である．行列の積は，Excel の Mmult()関数を使い

$\boldsymbol{X}_0{}^T$ 行と $\boldsymbol{X}_0$ 列の 積和＝9 を $(\boldsymbol{X}^T\boldsymbol{X})$ の 1 行 1 列目へ
$\boldsymbol{X}_1{}^T$ 行と $\boldsymbol{X}_0$ 列の 積和＝1 を $(\boldsymbol{X}^T\boldsymbol{X})$ の 2 行 1 列目へ
$\boldsymbol{X}_0{}^T$ 行と $\boldsymbol{X}_1$ 列の 積和＝1 を $(\boldsymbol{X}^T\boldsymbol{X})$ の 1 行 2 列目へ
$\boldsymbol{X}_1{}^T$ 行と $\boldsymbol{X}_1$ 列の 積和＝5 を $(\boldsymbol{X}^T\boldsymbol{X})$ の 2 行 2 列目へ

										$\boldsymbol{X}$				
					$\boldsymbol{X}^T$					$\boldsymbol{X}_0$	$\boldsymbol{X}_1$		$\boldsymbol{X}^T\boldsymbol{X}$	
$\boldsymbol{X}_0{}^T$	1	1	1	1	1	1	1	1	1	1	-1	＝	9	1
$\boldsymbol{X}_1{}^T$	-1	-1	0	0	0	0	1	1	1	1	-1		1	5
										1	0			
		$\boldsymbol{X}^T$	:	=Transpose($\boldsymbol{X}$の範囲)						1	0		=Mmult($\boldsymbol{X}^T$の範囲, $\boldsymbol{X}$の範囲)	
				Excelの関数						1	0		Excelの関数	
										1	0			
										1	1			
										1	1			
										1	1			
					2×9					9×2			2×2	

として，（2×2）の $\boldsymbol{X}^T\boldsymbol{X}$ 行列が生成される．なお， $\boldsymbol{X}^T\boldsymbol{X}$ ではなく $\boldsymbol{X}\boldsymbol{X}^T$ とすると，（9×2）と（2×9）となり，（9×9）の行列となる．このように行列の計算に際しては，行列のサイズを欄外に示しておくと計算ミスが少なくなる．

シグマ流の積和の計算

行列を用いた表記は，慣れれば簡潔で見通しがいいのだが，行列計算の方法を連想しやすくなるように，シグマを用いた計算方法を示す．Excel の SumProduct()関数は，複数の配列の要素ごとの積和の計算をする便利な関数であり，次に示すように行列 $\boldsymbol{X}$ の 1 列目を $\boldsymbol{X}_0$ とし，2 列目を $\boldsymbol{X}_1$ としたときに，2×2 の積和行列の 1 列 2 行目は，

SumProduct($\boldsymbol{X}_0$ の範囲，$\boldsymbol{X}_1$ の範囲)

として計算できる．同じ列同士の場合は，SumSq($\boldsymbol{X}_0$ の範囲) を使うこともできる．実際にデザイン行列 $\boldsymbol{X}$ の積和行列は，次に示すように求めることができるが，煩雑であり利便性に欠ける．

$\boldsymbol{X}_0$	$\boldsymbol{X}_1$		積和行列		
1	-1	＝	9	1	
1	-1		1	5	
1	0				
1	0		=SumSq($\boldsymbol{X}_0$の範囲) = 9		
1	0			=SumProduct($\boldsymbol{X}_0$の範囲, $\boldsymbol{X}_1$の範囲) = 1	
1	0		=SumProduct($\boldsymbol{X}_1$の範囲, $\boldsymbol{X}_0$の範囲) = 1		
1	1			=SumSq($\boldsymbol{X}_1$の範囲) = 5	
1	1				
1	1				
9×2					

行列の積

デザイン行列の転置は，Transpose()関数を使い，デザイン行列の掛け算は，Mmult()関数を使い $\boldsymbol{X}^T\boldsymbol{X}$ は，

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

のように Transpose()関数を Mmult()関数の入れ子にして計算することができる．

転置したデザイン行列 $\boldsymbol{X}^T$ と $\boldsymbol{X}$ の積をシグマ記号 Σ で表せば，行ベクトル $\boldsymbol{X}_0^T$ と列ベクトル $\boldsymbol{X}_0$ の積和は $\boldsymbol{n}$，行ベクトル $\boldsymbol{X}_0^T$ と列ベクトル $\boldsymbol{X}_1$ の積和は $\Sigma_i X_i$，行ベクトル $\boldsymbol{X}_1^T$ と列ベクトル $\boldsymbol{X}_1^T$ の積和は $\Sigma_i X_i^2$ となり，(2 行 2 列) の行列としてまとめて現したことになる．ここでは，$X_{1,i}$ とすべきところを略して X_i としている． $\boldsymbol{X}_0^T$ と $\boldsymbol{X}_0$ の積和は，行列 $\boldsymbol{X}$ の行の数 $\boldsymbol{n}$ となり，$(\boldsymbol{X}^T\boldsymbol{X})$ の 1 行 1 列目となっている．

											$\boldsymbol{X}$				
					$\boldsymbol{X}^T$						$\boldsymbol{X}_0$	$\boldsymbol{X}_1$		$\boldsymbol{X}^T\boldsymbol{X}$	
$\boldsymbol{X}_0^T$	1	1	1	1	1	1	1	1	1		1	X_1	=	n	$\Sigma_i X_i$
$\boldsymbol{X}_1^T$	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9		1	X_2		$\Sigma_i X_i$	$\Sigma_i X_i^2$
											1	X_3			
											1	X_4			
											1	X_5			
											1	X_6			
											1	X_7			
											1	X_8			
											1	X_9			
					2×9						9×2			2×2	

デザイン行列 *X* と反応 *Y* との積

次に，反応変数の列ベクトル$\boldsymbol{Y}$とデザイン行列$\boldsymbol{X}$の積和を示す．$\boldsymbol{X}^T\boldsymbol{Y}$は，転置されたデザイン行列$\boldsymbol{X}$と列ベクトル$\boldsymbol{Y}$の積で，$\boldsymbol{X}_0^T$と$\boldsymbol{Y}$についての積和は

$$1\times 2+1\times 3+\cdots+1\times 15=72$$

となり，$\boldsymbol{X}_1^T$と$\boldsymbol{Y}$についての積和は

$$(-1)\times 2+(-1)\times 3+\cdots+1\times 15=32$$

となる．しつこいことは承知しているが，実際に手を動かして行列計算を体験することが，最初の一歩である．

					$\boldsymbol{X}^T$						$\boldsymbol{Y}$		$\boldsymbol{X}^T\boldsymbol{Y}$	
$\boldsymbol{X}_0^T$	1	1	1	1	1	1	1	1	1		2	=	72	$\Sigma_i Y_i$
$\boldsymbol{X}_1^T$	-1	-1	0	0	0	0	1	1	1		3		32	$\Sigma_i X_i Y_i$
											6			
		$\boldsymbol{X}^T$	:	=Transpose($\boldsymbol{X}$の範囲)							7		=Mmult($\boldsymbol{X}^T$の範囲, $\boldsymbol{Y}$の範囲)	
				Excelの関数							8		Excelの関数	
											9			
											10			
											12			
											15			
					2×9						9×1		2×1	

シグマ記号で示せば，転置したデザイン行列$\boldsymbol{X}^T$と$\boldsymbol{Y}$の積は，$\boldsymbol{X}_0^T$と$\boldsymbol{Y}$の積和で$\Sigma_i Y_i$，$\boldsymbol{X}_1^T$ $\boldsymbol{Y}$の積和で$\Sigma_i(X_iY_i)$となる．

4.2. 偏差平方和ベースの回帰パラメータ推定

回帰式のパラメータ推定

求めたい回帰式を，

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i\ ,\quad i = 1, 2, ..., n \tag{4.4}$$

としたときに，真の直線からの偏差 ε_i の平方和は，

$$S_e = \sum_{i=1}^{n} \varepsilon_i^{\ 2} = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2 \tag{4.5}$$

である．通常の回帰分析において ε_i は，平均が 0，分散が σ^2 の正規分布に従うと仮定するのだが，実際にはどんな分布であっても最小 2 乗法での計算は可能である．このことが，最小 2 乗法による回帰分析が，ゆうずうむげに無批判的に重用されている最大の理由である．しかし，基本中の基本であるので，丁寧に説明する．偏差平方和 S_e をパラメータ β_0 と β_1 で偏微分すると，式（4.6）が得られる．

$$\begin{aligned} \frac{\partial S_e}{\partial \beta_0} &= -2\sum_{i=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_i\right) \\ \frac{\partial S_e}{\partial \beta_1} &= -2\sum_{i=1}^{n}\left[\left(Y_i - \beta_0 - \beta_1 X_i\right) X_i\right] \end{aligned} \tag{4.6}$$

偏微分に不慣れな場合には，段階的な学習が必要である．それぞれの ε_i^2 は，

$$\begin{aligned} \varepsilon_i^{\ 2} &= (Y_i - \beta_0 - \beta_1 X_i)^2 \\ &= Y_i^2 + \beta_0^{\ 2} + \beta_1^{\ 2} X_i^{\ 2} - 2Y_i\beta_0 - 2\beta_1 X_i Y_i + 2\beta_0\beta_1 X_i \end{aligned}$$

と式を展開できるので， β_0 での偏微分は，他の変数を定数と見なした微分なので，

$$\begin{aligned} \frac{\partial \varepsilon_i^{\ 2}}{\partial \beta_0} &= \frac{\partial (Y_i^2 + \beta_0^{\ 2} + \beta_1^{\ 2} X_i^{\ 2} - 2Y_i\beta_0 - 2\beta_1 X_i Y_i + 2\beta_0\beta_1 X_i)}{\partial \beta_0} \\ &= \quad 0 \ + 2\beta_0 \ + \quad 0 \quad - 2Y_i \quad - \quad 0 \quad + \ 2\beta_1 X_i \\ &= 2(\beta_0 - Y_i + \beta_1 X_i\ ) \\ &= -2(Y_i - \beta_0 - \beta_1 X_i) \end{aligned}$$

のように要素に分解し，シグマで再統合すれば式（4.6）の 1 行目となる． β_1 での偏微分も同様である．第 2.3 節でも示したが，偏微分を含む数学の学習をサポートするソフトを活用しつつ，合成関数の偏微分，さらに，一般化線形モデルでの対数尤度の偏微分に慣れ親しんでもらいたい．

式（4.6）を0と置くと，β_0とβ_1の推定値としての$\hat{\beta}_0$と$\hat{\beta}_1$を求めることができる．

$$
\begin{aligned}
&-2\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right)=0\\
&-2\sum_{i=1}^{n}\left[\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right)X_i\right]=0
\end{aligned}
\tag{4.7}
$$

正規方程式

式（4.7）を$\hat{\beta}_0$と$\hat{\beta}_1$について解くために，両辺を-2で割り式を整理すると，

$$
\begin{aligned}
&\sum_{i=1}^{n}Y_i-n\hat{\beta}_0-\hat{\beta}_1\sum_{i=1}^{n}X_i\ =0\\
&\sum_{i=1}^{n}X_iY_i-\hat{\beta}_0\sum_{i=1}^{n}X_i-\hat{\beta}_1\sum_{i=1}^{n}X_i^2=0
\end{aligned}
\tag{4.8}
$$

を得る．推定値$\hat{\beta}_0$と$\hat{\beta}_1$が，含まれない項を右辺に移して整理すると，

$$
\begin{aligned}
&(1)\quad n\hat{\beta}_0+\hat{\beta}_1\sum_{i=1}^{n}X_i\ =\sum_{i=1}^{n}Y_i\\
&(2)\quad \hat{\beta}_0\sum_{i=1}^{n}X_i+\hat{\beta}_1\sum_{i=1}^{n}X_i^2=\sum_{i=1}^{n}X_iY_i
\end{aligned}
\tag{4.9}
$$

が得られる．これらの方程式は，正規方程式と呼ばれている．式（4.9）の左辺は，$\boldsymbol{X}^T\boldsymbol{X}$と$\hat{\boldsymbol{\beta}}$の積に等しく，右辺は$\boldsymbol{X}^T\boldsymbol{Y}$と等しいので，

$\boldsymbol{X}^T\boldsymbol{X}$			$\boldsymbol{\beta}$^		$\boldsymbol{X}^T\boldsymbol{Y}$
n	$\Sigma_i X_i$		β_0^	=	$\Sigma_i Y_i$
$\Sigma_i X_i$	$\Sigma_i X_i{}^2$		β_1^		$\Sigma_i X_i Y_i$
2×2			2×1		2×1

が成り立ち，

$$
(\boldsymbol{X}^T\boldsymbol{X})\hat{\boldsymbol{\beta}}=\boldsymbol{X}^T\boldsymbol{Y}
\tag{4.10}
$$

が得られる．

正規方程式の解

正規方程式（4.9）を$\hat{\beta}_0$と$\hat{\beta}_1$について解くために，(1) の両辺に$1/n$を掛け，さらにΣX_iを掛けると次式を得る．(2) はΣの範囲の表示を外してある．以後，Σ記号の添え字iは省略する．

$$
\begin{aligned}
&(1)\quad \hat{\beta}_0\Sigma_iX_i+\hat{\beta}_1\frac{(\Sigma_iX_i)^2}{n}=\frac{(\Sigma_iX_i)(\Sigma_iY_i)}{n}\\
&(2)\quad \hat{\beta}_0\Sigma X_i+\hat{\beta}_1\Sigma X_i^2=\Sigma X_iY_i
\end{aligned}
\tag{4.11}
$$

式（4.11）の（2）式からの（1）式を引いて，$\hat{\beta}_1$について解くと

$$\begin{aligned}\hat{\beta}_1 &= \frac{\Sigma X_i Y_i - \dfrac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\Sigma X_i^2 - \dfrac{(\Sigma X_i)^2}{n}} \\ &= \frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2} \\ &= \frac{S_{XY}}{S_{XX}}\end{aligned} \tag{4.12}$$

のように，多くのテキストで示されている結果が得られる．ここで，S_{XX}は，$\boldsymbol{X}$の平均$\bar{X}$からの偏差の平方和の略語であり，S_{XY}は，$\boldsymbol{X}$の平均$\bar{X}$からの偏差と$\boldsymbol{Y}$の平均$\bar{Y}$からの偏差の積和の略号である．この式の変形は，$\bar{X}=(\Sigma X_i)/n$，$\bar{Y}=(\Sigma Y_i)/n$などの関係を用い，式（4.12）の 2 行目の分子 $\Sigma(X_i-\bar{X})(Y_i-\bar{Y})$ について

$$\begin{aligned}\Sigma(X_i-\bar{X})(Y_i-\bar{Y}) &= \Sigma X_i Y_i - \bar{X}\Sigma Y_i - \bar{Y}\Sigma X_i + n\bar{X}\bar{Y} \\ &= \Sigma X_i Y - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} \\ &= \Sigma X_i Y_i - n\bar{X}\bar{Y} \\ &= \Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}\end{aligned} \tag{4.13}$$

と，展開し，整理すると式（4.12）1 行目の分子となることを利用している．式（4.12）の 2 行目の分母$\Sigma(X_i-\bar{X})^2$については，式（4.13）のYをXに置き換えることにより式（4.12）1 行目の分母となる．

$$\begin{aligned}\Sigma(X_i-\bar{X})^2 &= \Sigma X_i^2 - 2\bar{X}\Sigma X_i + n\bar{X}^2 \\ &= \Sigma X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \Sigma X_i^2 - n\left(\frac{\Sigma X_i}{n}\right)^2 \\ &= \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\end{aligned} \tag{4.14}$$

このように，平均値を差し引いた平方和の計算は，手計算あるいは電卓の時代では，端数が出るので計算がめんどうであり，もとの数字X_iの 2 乗和の計算で済ませられるとの理由で，標準的な計算法として普及し，多くの統計の統計の教科書に引き継がれている．Excel を用いた場合には，平均値を差し引いた平方和の計算の方が簡潔で扱いやすい．

正規方程式（4.9）の 1 行目は，

$$n\hat{\beta}_0 + \hat{\beta}_1\Sigma X_i = \Sigma Y_i \tag{4.15}$$

なので，$\hat{\beta}_0$ について解くと，$\hat{\beta}_1$ を含む次式が得られる．

$$\begin{aligned}\hat{\beta}_0 &= \frac{\Sigma Y_i - \hat{\beta}_1 \Sigma X_i}{n} \\ &= \bar{Y} - \hat{\beta}_1 \bar{X}\end{aligned} \tag{4.16}$$

まとめると，回帰式の切片 $\hat{\beta}_0$ と傾き $\hat{\beta}_1$ の推定値は，

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{4.17}$$

と簡潔な式となる．回帰パラメータの導出だけであれば，この簡便な式によってパラメータ $\hat{\beta}_1$ および $\hat{\beta}_0$ を計算することができ，表 4.1 に示すように Excel を用いて実に簡便にパラメータ $\hat{\beta}_1$ および $\hat{\beta}_0$ を推定することができる．ただし，説明変数が増えて 2 変数以上となった場合にも拡張は可能であるが，第 12.3 節に示すように煩雑な式の展開が求められ，3 変数以上に対しては，さらに煩雑になり，デザイン行列を用いた定式化の方が簡潔である．

偏差平方を用いたパラメータの推定の実際

表 4.1 に示す Excel による計算シートの結果を用いて，回帰パラメータは，

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{24.0000}{4.8889} = 4.9091$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 8.0000 - 4.9091 \times 0.1111 = 7.4545$$

として求められる．

表 4.1　回帰パラメータの Excel シート上での推定

i	Y	X	Y偏差	X偏差	X偏差^2	XY偏差			
1	2	-1	-6.0000	-1.1111	1.2346	6.6667		β_1^=	4.9091
2	3	-1	-5.0000	-1.1111	1.2346	5.5556		β_0^=	7.4545
3	6	0	-2.0000	-0.1111	0.0123	0.2222			
4	7	0	-1.0000	-0.1111	0.0123	0.1111			
5	8	0	0.0000	-0.1111	0.0123	0.0000			
6	9	0	1.0000	-0.1111	0.0123	-0.1111			
7	10	1	2.0000	0.8889	0.7901	1.7778			
8	12	1	4.0000	0.8889	0.7901	3.5556			
9	15	1	7.0000	0.8889	0.7901	6.2222			
	8.0000	0.1111	0.0000	0.0000	4.8889	24.0000			
	平均	平均	合計	合計	平方和	平方和			
	Y^{-}	X^{-}	Y_i - Y^{-}	X_i - X^{-}	S_{XX}	S_{XY}			

回帰分析は，この様に定式化され，ほとんどの統計の教科書で取り上げられているが，いわゆる有意差検定よりもかなり高級であり，きちっと理解し，さらなる応用のために学習し

たいと思っても難解であり，教科書に示されている計算公式の範囲内に多くの人達が留まざるを得なくなってしまう．伝統的な回帰分析の解法は，更に回帰分析を活用したいと思う人達の学習意欲をへし折るような，まさにガラスの天井のごとくである．

ガラスの天井を超えるためには，回帰分析を Excel の行列計算で実行できるようになることが最初の一歩である．とは言え，いきなりパラメータが 3 以上の場合に取り組むと，敷居が高すぎて挫折しかねない．段階的な学習としては，式（4.10）で示した正規方程式

$$(\boldsymbol{X}^T\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$$

$\boldsymbol{X}^T\boldsymbol{X}$		$\boldsymbol{\beta}$^		$\boldsymbol{X}^T\boldsymbol{Y}$
n	ΣX_i	β_0^	＝	ΣY_i
ΣX_i	$\Sigma X_i^{\ 2}$	β_1^		$\Sigma X_i Y_i$
2×2		2×1		2×1

に立ち戻り，学習することが望ましいのであるが，推奨できる日本語の成書は，残念ながら絶版となっているドレーパ・スミス（1986）しか見当たらない．なお，Net 書店では中古本が手に入る場合もあるが，第 3 版の原著も，Net 書店で手軽に入手できる．

そこで，「ポアソン回帰」の基礎である通常の「回帰分析」について，従来の偏差平方和をベースした解析とデザイン行列をベースにした行列計算について，相互の関連を丁寧に示すことにした．そして，読者が偏差平方和をベースの回帰分析から，デザイン行列をベースにした回帰分析を主体にした解析法に親しみを感じてもらいたいと願っている．

4.3. デザイン行列を用いた回帰パラメータの推定

行列計算による回帰パラメータの推定

前節では，偏差平方和に基づく回帰分析のパラメータを推定する方法を示し，式（4.9）で正規方程式を行列を用いて，

$$\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$$

となることを式（4.10）で示した．

推定値 $\hat{\boldsymbol{\beta}}$ を得るために，行列計算では両辺を $(\boldsymbol{X}^T\boldsymbol{X})$ で割ることができない．逆行列の定義により $(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{X}) = \boldsymbol{I}$ が単位行列となるので，逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ を式（4.10）両辺に掛けて，

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

$$\boldsymbol{I}\,\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad (4.18)$$

を得る．2×2 の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ ならば，$(\boldsymbol{X}^T\boldsymbol{X})$ の行列式 D を計算し，次のようにして計算することができる．実際に手軽に手計算できるのは，2×2 の場合までで，3×3 以上の場合には，行列式の Mdetarm()関数を用いても煩雑であり勧められない．

X^TX				$(X^TX)^{-1}$			$(X^TX)^{-1}$	
n	ΣX_i	-1	=	$\Sigma X_i^2/D$	$-\Sigma X_i/D$			
ΣX_i	ΣX_i^2			$-\Sigma X_i/D$	n/D			
				$D=(n\Sigma X_i^2)-(\Sigma X_i)^2$				
9	1	-1	=	$5/D$	$-1/D$	=	0.1136	-0.0227
1	5			$-1/D$	$9/D$		-0.0227	0.2045
				D=45-1=44				
				=Mdeterm(X^TXの範囲)				

実際に逆行列を前掛けすると

$(X^TX)^{-1}$		X^TX			$(X^TX)^{-1}(X^TX)$	
0.1136	-0.0227	9	1	=	1.0000	0.0000
-0.0227	0.2045	1	5		0.0000	1.0000

のように単位行列になることが確認され，単位行列と $\hat{\boldsymbol{\beta}}$ の積は，

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\boldsymbol{X})$			β^		β^
1.0000	0.0000		β^$_0$	=	β^$_0$
0.0000	1.0000		β^$_1$		β^$_1$

と元の $\hat{\boldsymbol{\beta}}$ となる．

逆行列は，Excel の Minverse()関数によって簡単に求めることができる．推定値 $\hat{\boldsymbol{\beta}}$ を求めるために必要な， $\boldsymbol{X}^T\boldsymbol{Y}$ は，第 4.1 節の「デザイン行列 $\boldsymbol{X}$ と反応 $\boldsymbol{Y}$ との積」の項の結果を用い，これらの計算結果を組み合わせて，次のような手順で推定値 $\hat{\boldsymbol{\beta}}$ が求めることができる．

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

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$			$\boldsymbol{X}^T\boldsymbol{Y}$		β^
0.1136	-0.0227		72	=	7.4545
-0.0227	0.2045		32		4.9091
=Minverse($\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$の範囲)					= Mmult (($\boldsymbol{X}^T\boldsymbol{X}$)$^{-1}$の範囲, $\boldsymbol{X}^T\boldsymbol{Y}$の範囲)
			= Mmult (Transpose($\boldsymbol{X}$の範囲), $\boldsymbol{Y}$の範囲)		

デザイン行列と偏差平方和での推定式の相違

行列計算によって推定値 $\hat{\boldsymbol{\beta}}$ が得られたのであるが，正規方程式（4.17）から，導出された $\hat{\beta}_0$ と $\hat{\beta}_1$ とは，異なる式となっているので，$(\Sigma X_i) = n\bar{X}$ ，$(\Sigma Y_i) = n\bar{Y}$ などの関係を用いて行列計算の式を整理すると，一致することが確認される．

$$\begin{aligned}
\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\Sigma X_i^2}{D} & \dfrac{-\Sigma X_i}{D} \\ \dfrac{-\Sigma X_i}{D} & \dfrac{n}{D} \end{bmatrix} \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{(\Sigma X_i^2)(\Sigma Y_i) - (\Sigma X_i)(\Sigma X_i Y_i)}{n\Sigma X_i^2 - (\Sigma X_i)^2} \\ \dfrac{-\Sigma X_i \Sigma Y_i + n\Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\bar{Y}(\Sigma X_i^2) - \bar{X}(\Sigma X_i Y_i)}{\Sigma X_i^2 - n\bar{X}^2} \\ \dfrac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2} \end{bmatrix}
\end{aligned} \tag{4.19}$$

ここで，式（4.19）の最後の行列の 2 行目は，式（4.17）で導出された $\hat{\beta}_1$ の推定値に一致する．

$$\hat{\beta}_1 = \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2} = \frac{S_{XY}}{S_{XX}} \tag{4.20}$$

さて，式（4.17）で導出された $\hat{\beta}_0$ の推定式とは，式（4.19）の最後の行列の 1 行目は明らかに異なる．そこで，式（4.17）からスタートして，

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1\bar{X} \\
&= \frac{\bar{Y}S_{XX} - \bar{X}S_{XY}}{S_{XX}} \\
&= \frac{\left(\bar{Y}(\Sigma X_i^2) - \bar{Y}\dfrac{(\Sigma X_i)^2}{n}\right) - \left(\bar{X}(\Sigma X_i Y_i) - \bar{X}\dfrac{(\Sigma X_i)(\Sigma Y_i)}{n}\right)}{S_{XX}} \\
&= \frac{\left(\bar{Y}(\Sigma X_i^2) - n\bar{Y}\bar{X}^2\right) - \left(\bar{X}(\Sigma X_i Y_i) - n\bar{Y}\bar{X}^2\right)}{S_{XX}} \\
&= \frac{\bar{Y}(\Sigma X_i^2) - \bar{X}(\Sigma X_i Y_i)}{\Sigma X_i^2 - n\bar{X}^2}
\end{aligned} \tag{4.21}$$

となり，式（4.19）の最後の行列の 1 行目が導出される．私にとっても見るのも嫌になる数式の変形であり，シグマによる計算と行列による回帰係数の計算結果が一致することを数式で示すことは，難儀である．実用上は，事例により数値計算の結果が一致するは，これまでの結果で明らかである．

したがって，偏差平方和ベースの回帰分析およびデザイン行列ベースの計算方法を両建てで説明することは冗長であり．気持ちとしては避けたかったのであるが，偏差平方和ベースの回帰分析を「ガラスの天井」と言い切るため，あえて両者の関係について丁寧に示した．第 12.3 節の「偏差平方和ベースの重回帰分析」と第 12.4 節の「デザイン行列ベースの重回帰分析」で，切片と他の回帰パラメータとの間の共分散が含まれていないことの功罪についてさらに言及している．

4.4. 偏差平方和ベースの回帰パラメータの分散の推定

回帰パラメータ $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$ は，式（4.12）の正規方程式の解を用いて，

$$
\begin{aligned}
Var(\beta_1) &= Var\left[\frac{\Sigma(X_i-\bar{X})(Y_i-\bar{Y})}{\Sigma(X_i-\bar{X})^2}\right] \\
&= Var\left[\frac{\Sigma(X_i-\bar{X})Y_i}{\Sigma(X_i-\bar{X})^2}\right] \\
&= \frac{Var(Y_i)}{\Sigma(X_i-\bar{X})^2} \\
&= \frac{\sigma^2}{\Sigma(X_i-\bar{X})^2}
\end{aligned} \tag{4.22}
$$

となる．分子の変形は，

$$
\begin{aligned}
\Sigma(X_i-\bar{X})(Y_i-\bar{Y}) &= \Sigma(X_iY_i - X_i\bar{Y} - \bar{X}Y_i + \bar{X}\bar{Y}) \\
&= \Sigma(X_iY_i - \bar{X}Y_i) - \Sigma X_i\bar{Y} + \Sigma\bar{X}\bar{Y} \\
&= \Sigma(X_iY_i - \bar{X}Y_i) - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} \\
&= \Sigma(X_i-\bar{X})Y_i
\end{aligned}
$$

を用いて，$\Sigma(X_i-\bar{X})$ を Y_i に関してコンスタント化するためである．$\hat{\beta}_0$ の分散 $Var(\hat{\beta}_0)$ も，正規方程式の解を用いて，$\bar{Y}$ と $\hat{\beta}_1$ が無相関であり，$\bar{X}$ はコンスタントなので，

$$
\begin{aligned}
Var(\hat{\beta}_0) &= Var(\bar{Y} - \hat{\beta}_1\bar{X}) \\
&= Var(\bar{Y}) + \bar{X}^2 Var(\hat{\beta}_1) \\
&= \frac{\sigma^2}{n} + \frac{\sigma^2\bar{X}^2}{\Sigma(X_i-\bar{X})^2} \\
&= \frac{\sigma^2\Sigma X_i^{\ 2}}{n\Sigma(X_i-\bar{X})^2}
\end{aligned} \tag{4.23}
$$

となる．共分散 $Cov(\hat{\beta}_0, \hat{\beta}_1)$ は，

$$
\begin{aligned}
Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\bar{Y} - \hat{\beta}_1\bar{X}, \hat{\beta}_1) \\
&= -\bar{X}\, Var(\hat{\beta}_1) \\
&= \frac{-\bar{X}\sigma^2}{\Sigma(X_i-\bar{X})^2}
\end{aligned} \tag{4.24}
$$

となる．

回帰パラメータ $\hat{\boldsymbol{\beta}}$ の共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ は，これらの計算式から

$$
\begin{aligned}
\boldsymbol{\Sigma}(\boldsymbol{\beta}) &= \begin{bmatrix} Var(\beta_0) & Cov(\beta_0, \beta_1) \\ Cov(\beta_0, \beta_1) & Var(\beta_1) \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{\Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2} & \dfrac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} \\ \dfrac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} & \dfrac{1}{\Sigma(X_i - \bar{X})^2} \end{bmatrix} \sigma^2 \\
&= (\boldsymbol{X}^T \boldsymbol{X})^{-1} \sigma^2
\end{aligned}
\qquad (4.25)
$$

となる．分散共分散 $Var(\boldsymbol{\beta})$ は，デザイン行列 $\boldsymbol{X}^T$ と $\boldsymbol{X}$ の積の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ に誤差分散 σ^2 を掛けた結果に一致する．ただし，誤差分散 σ^2 は，未知なので，推定誤差 ε_i の平方和を自由度 $(n-2)$ で割った $\hat{\sigma}^2$ を用いて計算する．

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2} \\
&= \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{n-2}
\end{aligned}
\qquad (4.26)
$$

これまで，$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ については，次の式を示してきたのであるが，式（4.25）と異なるので，式の変形を行う．$(\boldsymbol{X}^T\boldsymbol{X})$ の行列式は，

$\boldsymbol{X}^T\boldsymbol{X}$				$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$	
n	ΣX_i	-1	=	$\Sigma X_i^2/D$	$-\Sigma X_i/D$
ΣX_i	ΣX_i^2			$-\Sigma X_i/D$	n/D
				$D=(n\Sigma X_i^2)-(\Sigma X_i)^2$	

としてきた．ただし，D は，

$$
\begin{aligned}
D &= n\Sigma X_i^2 - (\Sigma X_i)^2 \\
&= n(\Sigma X_i^2 - n\bar{X}^2) \\
&= n\Sigma(X_i - \bar{X})^2
\end{aligned}
\qquad (4.27)
$$

と変形できるので，

$$
(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \begin{bmatrix} \dfrac{\Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2} & \dfrac{-\Sigma X_i}{n\Sigma(X_i - \bar{X})^2} \\ \dfrac{-\Sigma X_i}{n\Sigma(X_i - \bar{X})^2} & \dfrac{n}{n\Sigma(X_i - \bar{X})^2} \end{bmatrix} = \begin{bmatrix} \dfrac{\Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2} & \dfrac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} \\ \dfrac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} & \dfrac{1}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}
\qquad (4.28)
$$

と一致することが確かめられる．このように，従来のシグマを用いたパラメータの分散の計算式とデザイン行列を用いた計算結果が同じであることが示された．したがって，従来の計算式ではなく，デザイン行列ベースの計算手順で簡単に得れるパラメータの共分散行列を基本とすることは，ガラスの天井を超えるための必須の知識である．

4.5. デザイン行列を用いた回帰分析の実際

パラメータの推定

表 4.2 にこれまで示してきた計算方法を総合し, デザイン行列を用いた回帰分析を例示す. 全ての計算過程を 1 枚の Excel シートで示そうとしたために, やや見づらくなっているので, 丁寧な解説を付け加える．なお，Excel の分析ツールの回帰分析を活用した後述する表 4.6 と比較してもらいたい.

表 4.2　デザイン行列を用いた回帰分析

i	$\boldsymbol{X}$		$\boldsymbol{Y}$	$\boldsymbol{Y}$^	$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{Y}$^							
1	1	-1	2	2.5455	-0.5455							
2	1	-1	3	2.5455	0.4545							
3	1	0	6	7.4545	-1.4545							
4	1	0	7	7.4545	-0.4545							
5	1	0	8	7.4545	0.5455							
6	1	0	9	7.4545	1.5455							
7	1	1	10	12.3636	-2.3636		項	推定値	分散	SE	t 値	p 値
8	1	1	12	12.3636	-0.3636		β_0^=	7.4545	0.2952	0.5433	13.72	0.0000
9	1	1	15	12.3636	2.6364		β_1^=	4.9091	0.5313	0.7289	6.73	0.0003
								$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$		sqrt(分散)		T.Dist.2T(t, 7)
	9.00	1.00	0.1136	-0.0227	72.0000		$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$=	18.1818	共分散	0.2952	-0.0590	
	1.00	5.00	-0.0227	0.2045	32.0000		σ^2=	2.5974	行列	-0.0590	0.5313	
	$\boldsymbol{X}^T\boldsymbol{X}$		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		$\boldsymbol{X}^T\boldsymbol{Y}$					$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^2		

デザイン行列 $\boldsymbol{X}$ に対し，$\boldsymbol{X}^T\boldsymbol{X}$ の結果が表の下段に 2×2 の矩形内に Excel の Mmult()関数および Transpose()関数を用いて

= Mmult(Transpose($\boldsymbol{X}$の範囲), $\boldsymbol{X}$の範囲)) =

9.00	1.00
1.00	5.00

と計算されている．その横に $\boldsymbol{X}^T\boldsymbol{X}$ の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ が，Minverse()関数を用いて

= Minverse($\boldsymbol{X}^T\boldsymbol{X}$の範囲) =

0.1136	-0.0227
-0.0227	0.2045

となり，デザイン行列 $\boldsymbol{X}$ の転置行列列 $\boldsymbol{X}^T$ と列ベクトル $\boldsymbol{Y}$ との積 $\boldsymbol{X}^T\boldsymbol{Y}$ が

= Mmult(Transpose($\boldsymbol{X}$の範囲),$\boldsymbol{Y}$の範囲) =

72.0000
32.0000

として計算されている．回帰パラメータの推定値 $\hat{\boldsymbol{\beta}}$ は，表の中段の「推定値」の欄に

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

7.4545
4.9091

と $\hat{\beta}_0 = 7.4545$，$\hat{\beta}_1 = 4.9091$ として計算されている．誤差分散は，回帰の推定値 $\hat{\boldsymbol{Y}}$ を

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

で求め，誤差ベクトル $\boldsymbol{\varepsilon}$ が列ベクトル $\boldsymbol{Y}$ と推定ベクトル $\hat{\boldsymbol{Y}}$ の差

$$\hat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (\boldsymbol{Y}\text{の範囲} - \hat{\boldsymbol{Y}}\text{の範囲})$$

として求められている．誤差平方和

$$S_e = \hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} = \text{Mmult}(\text{Transpose}(\hat{\boldsymbol{\varepsilon}}\text{の範囲}), \hat{\boldsymbol{\varepsilon}}\text{の範囲})$$

を計算し，データ数 n からパラメータの数 2 を引いた自由度で割った平均平方が誤差分散 $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}}{n-2} = \frac{18.1818}{9-2} = 2.5974$$

として計算されている．

デザイン行列を用いた回帰分析の最大の利点は，回帰パラメータ $\hat{\beta}_0$ と $\hat{\beta}_1$ の分散および共分散が 2×2 の行列として簡単に得られることである．式（4.25）からパラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ は，

$$\begin{aligned}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) &= \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2 \end{aligned} \tag{4.29}$$

であることを示した．デザイン行列の積和の逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ および誤差分散 $\hat{\sigma}^2$ は，

0.1136	-0.0227
-0.0227	0.2045
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$	

$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$=	18.1818
σ^2=	2.5974

として計算されているので，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ は，

$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ =

0.1136	-0.0227	2.5974	=	0.2952	-0.0590
-0.0227	0.2045			-0.0590	0.5313
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		σ^2		$\boldsymbol{\Sigma}$(β^)=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$σ^2	

と計算される．分散 $Var(\hat{\beta}_0) = 0.2952$，$Var(\hat{\beta}_1) = 0.5313$ は，$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$ の対角要素となっているので，次のように推定値の分散が得られる．

項	推定値	分散
β0^=	7.4545	0.2952
β1^=	4.9091	0.5313

更に標準誤差 SE を分散の平方根 sqrt()関数で求め，推定値/SE で t 値を計算し，t 分布の両側確率の p 値を T.dist.2T()関数

$$p\text{ 値} = \text{T.dist.2T}(t\text{値}, (9-2))$$

で計算し，回帰パラメータについての推定およびt検定が次のように行える．

項	推定値	分散	SE	t 値	p 値
β_0^=	7.4545	0.2952	0.5433	13.72	0.0000
β_1^=	4.9091	0.5313	0.7289	6.73	0.0003
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$		sqrt(分散)		T.Dist.2T(t, 7)	

表 4.1 に示した単に回帰パラメータを計算した Excel シートより計算は複雑ではあるが，同程度の Excel シート 1 枚の中に，基本的な関数のみで，回帰分析の基本的な結果が網羅されている．なお，回帰分析に付きものの，分散分析表は，行列計算で求めるよりも推定値ベースで求めることが簡潔であり，本質的でもある．

分散分析表

一般的に最小 2 乗法による回帰分析には，分散分析表が付きものである．分散分析表は，各種の偏差平方和をベースに構成されている．表 4.2 には，反応Y_iに対して，推定された回帰パラメータ$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$との偏差$\hat{\varepsilon}_i = Y_i - \hat{Y}_i$を計算し，その平方和を求めた．この偏差平方和を，誤差平方和S_eといい，推定された回帰直線からのズレ（誤差平方和）の大きさを表している．

$$S_e = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (4.30)$$

反応Y_i自体についての統計量として，平均と分散が代表的であるが，平均$\overline{Y}$からの偏差を，$d_i = Y_i - \overline{Y}$　としたときの偏差平方和を 全体の平方和の意味でS_Tとする．

$$S_T = \sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2 \quad (4.31)$$

表 4.3　各種の平方和の計算

i	$\boldsymbol{X}$		Y_i	$\overline{Y}$	$Y_i - \overline{Y}$	$\hat{Y}_i$	$\hat{Y}_i - \overline{Y}$	$Y_i - \hat{Y}_i$			
1	1	-1	2	8.00	-6.00	2.55	-5.45	-0.55		β_0^=	7.4545
2	1	-1	3	8.00	-5.00	2.55	-5.45	0.45		β_1^=	4.9091
3	1	0	6	8.00	-2.00	7.45	-0.55	-1.45			
4	1	0	7	8.00	-1.00	7.45	-0.55	-0.45			
5	1	0	8	8.00	0.00	7.45	-0.55	0.55			
6	1	0	9	8.00	1.00	7.45	-0.55	1.55			
7	1	1	10	8.00	2.00	12.36	4.36	-2.36			
8	1	1	12	8.00	4.00	12.36	4.36	-0.36			
9	1	1	15	8.00	7.00	12.36	4.36	2.64			
				8.00	136.00		117.82	18.18			
				$\overline{Y}$	S_T		S_R	S_e			
自由度			9	1	9−1=8	2	2−1=1	9-2=7		2	

回帰直線の推定値は，$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$であるが，平均$\overline{Y}$からの偏差$R_i = \hat{Y}_i - \overline{Y}$の平方和が回帰の平方和$S_R$となる．

$$S_R = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y}_i)^2 \quad (4.32)$$

この様に，元の反応 Y_i についての平方和には，

$$\begin{aligned} S_T &= \sum\nolimits_{i=1}^{n}(Y_i - \bar{Y})^2 \\ &= \sum\nolimits_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum\nolimits_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \\ &= S_e + S_R \end{aligned} \tag{4.33}$$

との関係があり，平方和の分解とも言われている（本節の末尾に式の証明を示す）．実際に，これらの平方和を計算した結果を表 4.4 に示す．

回帰パラメータの推定には, デザイン行列を使うことを勧めるが, 分散分析表の作成には, 各種の平方和の計算に基づいた方法が，わかりやすい．平方和の計算は，SumSq()関数の使用が効率的である．

$$S_T = \sum\nolimits_{i=1}^{9}(Y_i - \bar{Y})^2 = \text{SumSq}(\boldsymbol{Y}\text{の範囲} - \bar{Y}) = 136.00\,, \qquad df = 9-1=8$$

$$\hat{\boldsymbol{Y}} = \text{Mmult}(\boldsymbol{X}\text{の範囲}, \boldsymbol{\beta}\text{の範囲})\,, \qquad df = 2$$

$$S_R = \sum\nolimits_{i=1}^{9}(\hat{Y}_i - \bar{Y})^2 = \text{SumSq}(\hat{\boldsymbol{Y}}\text{の範囲} - \bar{Y}) = 117.82\,, \qquad df = 2-1=1$$

$$S_e = \sum\nolimits_{i=1}^{9}(\hat{Y}_i - Y_i)^2 = \text{SumSq}(\hat{\boldsymbol{Y}}\text{の範囲} - \boldsymbol{Y}\text{の範囲}) = 18.18\,, \qquad df = 9-2=7$$

計算結果を，表 4.4 の分散分析表にまとめる．計算原理を習得したの後は，表 4.6 に示すように Excel の分析ツールの回帰分析によっても同じ結果が得られるので，自ら計算することにこだわる理由はない．

表 4.4　回帰に対する分散分析表

要因		平方和	自由度	平均平方	F 値	p 値
回帰	S_R	117.82	2-1=1	117.82	45.36	0.0003
誤差	S_e	18.18	9-2=7	2.60		
全体	S_T	136.00	9-1=8			

反応 $\boldsymbol{Y}$ の自由度は，データ数 n であるが，全体の平方和 S_T の自由度は，計算のために反応 $\boldsymbol{Y}$ から求めた $\bar{Y}$ を使っているので，自由度が 1 つ分減少して $df_T = n-1$ となる．誤差平方和 S_e は，回帰の推定値 $\hat{Y}_i$ の計算のために $\hat{\beta}_0$ と $\hat{\beta}_1$ を用いているので自由度が 2 つ減り $df_e = n-2$ となる．回帰の平方和 S_R は，自由度が 2 の $\hat{Y}_i$ に対し，自由度 1 の $\bar{Y}$ の差の平方和なので 1 つ減り $df_R = 1$ となる．平方和と同様に自由度の推定にも

$$\begin{aligned} df_T &= df_R + df_e \\ &= 1 + (n-2) = n-1 \end{aligned} \tag{4.34}$$

が成り立つ．分散分析における自由度については，各種の便宜的な説明が行われているが，偏差平方和の定義式に立ち返ることにより，自由度の本質的な理解となる．

F 値は，回帰の平均平方 117.82 を誤差の平均平方 2.60 で割って 45.36 が計算されている．この F 値は，分子の自由度 1，分母の自由度 7 の F 分布の上側確率で，F.dist.RT()関数を使い，=F.dist.RT(45.36, 1, 7)で計算されている．これらの Excel の確率分布の計算になれることも大切である．

分散分析表は，最小 2 乗法による分析に際しては，結果を概観するために有益である．しかし，最尤法によるポアソン回帰では，対数尤度を用いた要約となり，偏差平方和をベースにした分散分析表の適用ができないが，基本的な考え方は同様であり，第 11.3 節の表 11.7 に「通常の回帰とポアソン回帰の対比」として対応関係を示してあるので，参照されたい．

パラメータの共分散行列の活用

Excel のデータ分析ツールの回帰分析により，パラメータの推定などが手軽にできることから Excel の行列関数を使った回帰分析をする意義はないと思われるかもしれない．多くの Excel ユーザの悩みは，回帰直線の 95%信頼区間および 95%予測区間（個別データの 95%信頼区間）を散布図上に描きたいと思っても手軽に解決する機能が見いだせないことにある．

デザイン行列を用いた行列計算による回帰分析の計算過程では，表 4.2 に示したようにパラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ の計算が含まれており，この行列を使って，パラメータの標準誤差 SE を計算した．パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ があれば，95%信頼区間を散布図上に描くのは容易である．Excel のデータ分析ツールの回帰分析で出力される分散分析表の誤差分散 $\hat{\sigma}^2$ が得られるので，切片を含めた説明変数をデザイン行列 $\boldsymbol{X}$ とし，行列計算で $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ を求め

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

により，パラメータの共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ が容易に得られる．

回帰直線の 95%信頼区間

回帰直線の 95%信頼区間を求めるためには，回帰の推定値 $\hat{Y}_i$ の分散が必要となる．合成分散の一般的な公式により，

$$\begin{aligned} Var(\hat{Y}_i) &= Var(\hat{\beta}_0 + \hat{\beta}_1 X_i) \\ &= Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1)X_i + Var(\hat{\beta}_1)X_i^2 \end{aligned} \tag{4.35}$$

により，計算する．任意の行ベクトルを

$$\boldsymbol{x} = [1 \quad x]$$

とすれば，次の 2 次形式による計算方法で，

$$
\begin{aligned}
Var(\hat{y}) &= \boldsymbol{x}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2]\boldsymbol{x}^T \\
&= \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} \\
&= \begin{bmatrix} 1 & x \end{bmatrix} \begin{bmatrix} Var(\hat{\beta}_0) + Cov(\hat{\beta}_0, \hat{\beta}_1)x \\ Cov(\hat{\beta}_0, \hat{\beta}_1) + Var(\hat{\beta}_1)x \end{bmatrix} \\
&= Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1)x + Var(\hat{\beta}_1)x^2
\end{aligned}
\tag{4.36}
$$

合成分散の一般的な公式に一致する．

Excel できれいな 95%信頼区間の滑かな曲線を描くためには，表 4.5 に示すように，描きたい X 軸の範囲内で，適当な間隔の x を設定し，推定値を $\hat{y} = \boldsymbol{x}\hat{\boldsymbol{\beta}}$，分散 $Var(\hat{y})$ を

$$
Var(\hat{y}) = \boldsymbol{x}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2]\boldsymbol{x}^T \tag{4.37}
$$

で計算する．表 1.10 のポアソン回帰での 95%信頼区間と比べると，X の小さい方の幅が広がり，大きい方の幅が狭くなっている．これは，回帰直線からの残差の系統的な広がりを全く考慮せずに共通の分散としたためである．

表 4.5　回帰直線の 95%信頼区間

切片	x	y^	$Var(y$^$)$	L95%	U95%	個別 L95%	個別 U95%
1	-2	-2.36	2.66	-6.22	1.49	-7.78	3.06
1	-1.8	-1.38	2.23	-4.91	2.15	-6.58	3.81
1	-1.6	-0.40	1.84	-3.61	2.81	-5.38	4.58
1	-1.4	0.58	1.50	-2.32	3.48	-4.21	5.37
1	-1.2	1.56	1.20	-1.03	4.16	-3.05	6.17
1	-1	2.55	0.94	0.25	4.84	-1.90	7.00
1	-0.8	3.53	0.73	1.51	5.55	-0.79	7.84
1	-0.6	4.51	0.56	2.74	6.27	0.31	8.71
1	-0.4	5.49	0.43	3.95	7.04	1.38	9.60
1	-0.2	6.47	0.34	5.09	7.85	2.42	10.53
1	0	7.45	0.30	6.17	8.74	3.43	11.48
:							
1	1.8	16.29	1.80	13.11	19.47	11.33	21.25
1	2	17.27	2.18	13.78	20.77	12.10	22.44

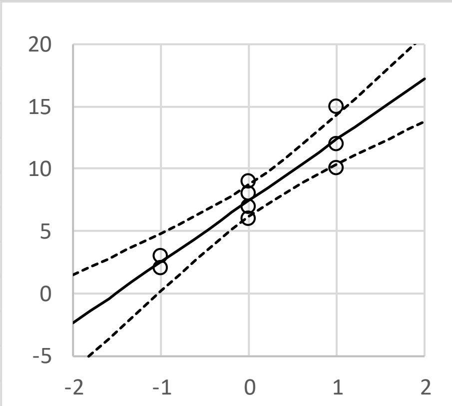


表 4.5 の 1 行目の $\boldsymbol{x}_1 = \begin{bmatrix} 1 & -2 \end{bmatrix}$ に対する推定値 $\hat{y}$ は，

$$
\begin{aligned}
\hat{y} &= \boldsymbol{x}\hat{\boldsymbol{\beta}} \\
&= \begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 7.4545 \\ 4.9091 \end{bmatrix} \\
&= -2.3636
\end{aligned}
$$

となり，分散は，

$$
\begin{aligned}
Var(\hat{y}) &= \boldsymbol{x}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2]\boldsymbol{x}^T \\
&= \begin{bmatrix} 1 & -2 \end{bmatrix}\begin{bmatrix} 0.2952 & -0.0590 \\ -0.0590 & 0.5313 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} \\
&= 2.6564
\end{aligned}
$$

95%信頼区間は,

$$
\begin{aligned}
L95\% &= \hat{y} - t_{0.05}(9-2)\sqrt{Var(\hat{y})} \\
&= -2.3636 - 2.3646 \times \sqrt{2.6564} \\
&= -6.2176
\end{aligned}
$$

$$
\begin{aligned}
U95\% &= -2.3636 + 2.3646 \times \sqrt{2.6564} \\
&= 1.4904
\end{aligned}
$$

として計算されている.

個別データの 95%信頼区間は，回帰直線の分散 $Var(\hat{Y}_i)$ に 1 個分のデータの分散 $\hat{\sigma}^2$ を加えた分散を使う.

$$
\begin{aligned}
\text{個別}L95\% &= \hat{y} - t_{0.05}(9-2)\sqrt{Var(\hat{y}) + \hat{\sigma}^2} \\
&= -2.3636 - 2.3646 \times \sqrt{2.6564 + 2.5974} \\
&= -7.7837
\end{aligned}
$$

$$
\begin{aligned}
\text{個別}L95\% &= -2.3636 + 2.3646 \times \sqrt{2.6564 + 2.5974} \\
&= 3.0564
\end{aligned}
$$

第 1 行目で作成した計算式は，行方向にフィルハンドルでコピーすることにより計算される．この結果を Excel 散布図にまとめた結果が示されている．Excel 散布図は，きめ細かな設定ができる優れものである.

伝統的な方法

多くの教科書で回帰直線の 95%信頼区間についての記述は，ほとんどが，以下の形式で示されている（過度な標準化となっている）．ある x_0 について

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_0 \tag{4.38}$$

から， $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1\overline{x}$ なので， $\hat{\beta}_0$ について解くと， $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$ となり，これを $\hat{\beta}_0$ を代入して

$$
\begin{aligned}
\hat{Y} &= \overline{y} - \hat{\beta}_1\overline{x} + \hat{\beta}_1 x_0 \\
&= \overline{y} + \hat{\beta}_1(x_0 - \overline{x})
\end{aligned}
\tag{4.39}
$$

が得られる．$\hat{Y}$ の分散 $Var(\hat{Y})$ は，$\overline{y}$ と $\hat{\beta}_1$ の共分散が 0 なので，式（4.22）の $Var(\hat{\beta}_1)$ を用いて，次式で与えられる.

$$
\begin{aligned}
Var(\hat{Y}) &= Var[\bar{y} + \hat{\beta}_1(x_0 - \bar{x})] \\
&= Var(\bar{y}) + (x_0 - \bar{x})^2 Var(\hat{\beta}_1) \\
&= \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}\right]\hat{\sigma}^2
\end{aligned} \qquad (4.40)
$$

また，個別データに対する分散 $Ver(\hat{Y}_{個別})$ は，

$$
\begin{aligned}
Ver(\hat{Y}_{個別}) &= Ver(\hat{Y}) + \hat{\sigma}^2 \\
&= \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}\right]\hat{\sigma}^2
\end{aligned} \qquad (4.41)
$$

で与えられる．

単回帰分析の場合には，これらのシグマでの式を用いればよいのであるが，最尤法によるポアソン回帰の場合には，そもそも偏差平方和の計算をしないので，これらの式は使えない．回帰分析でも，変数が増えた場合には，これらの式は使えない．

このように，多くの教科書で定式化されている単回帰についてシグマを用いた計算公式は，単回帰分析でのみに対するものであり，他の問題に対して応用することが出来ない．このような状況は，再度繰り返すが，多くの読者に対して応用力を封じ込めるような「ガラスの天井」のごとくである．Excel の「データ分析」の「回帰分析」，回帰分析のための LinEst()関数は，重回帰分析もサポートする優れものではあるが，信頼区間の計算についてはサポートされていないので，他の問題への応用力をさらに阻害してしまう．

例えば 2 次式 $y = \beta_0 + \beta_1 x + \beta_2 x^2$ をあてはめ，その 95%信頼区間を描きたいとしても，従来の計算方法では，解決の糸口はつかめない．行列計算の場合ならば，変数の数が増えてもパラメータの共分散行列は，常に $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}^2$ であり，変数の数によらず 95%信頼区間の計算方法は同じである．

第 2 章でも示したように，最尤法によるポアソン回帰の場合であっても，パラメータの分散共分散行列（共分散行列，と分散を省略する場合もある）を用いた信頼区間の計算の考え方は，通常の回帰分析の場合と全く同じで，汎用的である．

また，第 12.5 節では，デザイン行列をベースにした行列計算により，2 次曲線の 95%信頼区間を Excel の散布図で作成する手順を詳細に示している．単に 2 次式の回帰パラメータを出すだけならば，Excel の「回帰分析」を使えばいいのだが，2 次曲線の 95%信頼区間の計算はどのようにしたらよいのだろうか．

東京大学教養学部統計学教室編（1992），「自然科学の統計学」の「第 2 章 線形モデルと最小二乗法」には，2 次式についての正規方程式が展開され，パラメータの推定値が求められ，2 次曲線が例示されている．しかし，95%信頼区間の計算式の例示は見いだせない．もちろん，最小二乗推定量の分散の一般式としての記述はあるものの，残念ながら事例としては，単回帰分析の伝統的な計算方法が示され，95%信頼区間の計算式も偏差平方和を用いた伝統的な記述となっている．そこで，第 12.5 節でパラメータの共分散行列の活用事例として 2 次式の95%信頼区間および予測区間を推定しグラフ化した結果を示す．

現実的な対応

Excel の「回帰分析」は手軽に使える優れものであるので，パラメータの共分散分行列の計算を付け加えることにより，95%信頼区間の計算を自在にできるようになる．表 4.6 に示すように，Excel の「回帰分析」を使い，分散分析表から残差分散 $\hat{\sigma}^2 = 2.5974$ を，係数の表から

表 4.6　分析ツールの回帰分析およびパラメータの共分散行列の計算

i	$\boldsymbol{X}$		$\boldsymbol{Y}$		分散分析表(分析ツールの回帰分析)						
1	1	-1	2			自由度	変動	分散	分散比	有意 F	
2	1	-1	3		回帰	1	117.8182	117.8182	45.3600	0.0003	
3	1	0	6		残差	7	18.1818	2.5974			
4	1	0	7		合計	8	136.0000				
5	1	0	8								
6	1	0	9			係数	標準誤差	t	P-値	下限 95%	上限 95%
7	1	1	10		切片	7.4545	0.5433	13.7212	0.0000	6.1699	8.7392
8	1	1	12		X 値 1	4.9091	0.7289	6.7350	0.0003	3.1855	6.6327
9	1	1	15								
	9.00	1.00			0.1136	-0.0227		0.2952	-0.0590	$t_{0.05}(7)$=	2.3646
	1.00	5.00			-0.0227	0.2045		-0.0590	0.5313		
	$\boldsymbol{X}^T\boldsymbol{X}$				$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$			$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^)=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^2			

$\hat{\beta}_0 = 7.4545$， $\hat{\beta}_1 = 4.9091$ を得る．パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ は，デザイン行列 $\boldsymbol{X}$ を用い

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

＝

9.00	1.00
1.00	5.00
$\boldsymbol{X}^T\boldsymbol{X}$	

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

＝

0.1136	-0.0227	×	2.5974
-0.0227	0.2045		
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$			

$=$

0.2952	-0.0590
-0.0590	0.5313

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

によって計算できる．これらを用いて，表 4.5 で示した回帰直線の 95%信頼区間の計算が可能となる．

表 4.5 に示した Excel によるグラフ作成の手順は，すでに示したが，ここでは，できる限り表 4.7 に示すように行列計算を使った計算方法で示す．もちろん，個別データに対しての計算式をフィルハンドルでコピーしても同じ結果が得られるのであるが，データ数が変化したときなど，行列計算による計算式の方が，変更に際して見通しがよい．

手順 1）表 4.6 の（$x_{1,i}$，y_i）により基本の散布図を作成する．X 軸の範囲を（–2～+2)，Y 軸の範囲を（–5～+20）とセットする．

手順 2）X 軸の範囲を（–2～+2）とし，適当な増分の x' で作表する．切片 1 も含めておく．

手順 3）推定値 $\hat{\boldsymbol{y}}'$ を計算する．

$$\hat{\boldsymbol{y}}' = \boldsymbol{X}'\hat{\boldsymbol{\beta}}$$
$$= \mathrm{Mmultt}(\boldsymbol{X}'\text{の範囲}, \hat{\boldsymbol{\beta}}\text{の範囲})$$

手順 4）推定値 $\hat{y}_i'$ の分散を計算し，フィルハンドルで計算式をコピーする．

$$Var(\hat{y}_i') = \boldsymbol{x}_i'\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\boldsymbol{x}_i'^T$$
$$= \mathrm{Mmult}(\mathrm{Mmult}(\boldsymbol{x}_i'\text{の範囲}, \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\text{の範囲}), \mathrm{Transpose}(\boldsymbol{x}_i'\text{の範囲}))$$

手順 5）回帰の推定値の 95%信頼区間を一括で計算する

$$\boldsymbol{L}95\% = \hat{\boldsymbol{y}}' - t_{0.05}(9-2)\sqrt{Var(\hat{\boldsymbol{y}}')}$$
$$= \hat{\boldsymbol{y}}'\text{の範囲} - t_{0.05}(9-2)\times \mathrm{sqrt}(Var(\hat{\boldsymbol{y}}')\text{の範囲})$$
$$\boldsymbol{U}95\% = \hat{\boldsymbol{y}}' + t_{0.05}(9-2)\sqrt{Var(\hat{\boldsymbol{y}}')}$$
$$= \hat{\boldsymbol{y}}'\text{の範囲} + t_{0.05}(9-2)\times \mathrm{sqrt}(Var(\hat{\boldsymbol{y}}')\text{の範囲})$$

手順 6）個別の回帰の推定値の 95%信頼区間を一括で計算する

$$\text{個別}\boldsymbol{L}95\% = \hat{\boldsymbol{y}}' - t_{0.05}(9-2)\sqrt{Var(\hat{\boldsymbol{y}}') + \hat{\sigma}^2}$$
$$= \hat{\boldsymbol{y}}'\text{の範囲} - t_{0.05}(9-2)\times \mathrm{sqrt}(Var(\hat{\boldsymbol{y}}')\text{の範囲}+\hat{\sigma}^2)$$
$$\text{個別}\boldsymbol{U}95\% = \hat{\boldsymbol{y}}' + t_{0.05}(9-2)\sqrt{Var(\hat{\boldsymbol{y}}') + \hat{\sigma}^2}$$
$$= \hat{\boldsymbol{y}}'\text{の範囲} + t_{0.05}(9-2)\times \mathrm{sqrt}(Var(\hat{\boldsymbol{y}}')\text{の範囲}+\hat{\sigma}^2)$$

手順 7）Excel グラフの「データの選択」機能を使い，推定値 $\hat{\boldsymbol{y}}'$，95%信頼区間 $L95\%$，$U95\%$，個別$L95\%$，個別$U95\%$ を追加し，「データ系列の書式設定」で適当な線グラフとする．

表 4.7　Excel の行列計算機能を用いた上側 95%信頼区間の一括計算

TEXT　　=Q5:Q13+T.INV.2T(0.05,7)*SQRT(R5:R13)

T.INV.2T(確率, 自由度)

	O	P	Q	R	S	T	U	V
3								
4	切片	*x*	*y*^	*Var* (*y*^)	*L* 95%	*U*95%	個別*L* 95%	個別*U*95%
5	1	-2.0	-2.3636	2.6564	-6.2176	T.INV.2T(0.(	-7.7837	3.0564
6	1	-1.5	0.0909	1.6677	-2.9627	3.1445	-4.7925	4.9743
7	1	-1.0	2.5455	0.9445	0.2474	4.8435	-1.9048	6.9957
8	1	-0.5	5.0000	0.4870	3.3498	6.6502	0.8471	9.1529
9	1	0.0	7.4545	0.2952	6.1699	8.7392	3.4329	11.4762
10	1	0.5	9.9091	0.3689	8.4728	11.3454	5.8365	13.9817
11	1	1.0	12.3636	0.7084	10.3734	14.3538	8.0643	16.6630
12	1	1.5	14.8182	1.3135	12.1082	17.5282	10.1419	19.4944
13	1	2.0	17.2727	2.1842	13.7781	20.7674	12.1020	22.4434

平方和の分解に対する補足

式（4.33）で，全体の平方和 S_T が，回帰の平方和 S_R と誤差平方和 S_e の和に分解できるとしたが，式の展開を省略して結論だけを示したので補足をする．式（4.33）では，第 3 項があり，これがゼロとなることを示さなかった．

$$\begin{aligned} S_T &= \sum_{i=1}^{n}(Y_i-\bar{Y})^2 \\ &= \sum_{i=1}^{n}[(Y_i-\hat{Y}_i)+(\hat{Y}-\bar{Y})]^2 \\ &= \sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2+\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2+2\sum_{i=1}^{n}[(Y_i-\hat{Y}_i)(\hat{Y}-\bar{Y})] \\ &= S_e+S_R+2\sum_{i=1}^{n}[(Y_i-\hat{Y}_i)(\hat{Y}-\bar{Y})] \\ &= S_e+S_R \end{aligned} \tag{4.42}$$

第 3 項は，次のように展開して 0 となる．

$$\begin{aligned} \sum_{i=1}^{n}[(Y_i-\hat{Y}_i)(\hat{Y}_i-\bar{Y})] &= \sum_{i=1}^{n}\hat{\varepsilon}_i\hat{Y}_i-\bar{Y}\sum_{i=1}^{n}\hat{\varepsilon}_i \\ &= \sum_{i=1}^{n}\hat{\varepsilon}_i(\hat{\beta}_0+\hat{\beta}_1X_i) \\ &= \hat{\beta}_0\sum_{i=1}^{n}\hat{\varepsilon}_i+\hat{\beta}_1\sum_{i=1}^{n}\hat{\varepsilon}_iX_i \\ &= 0 \end{aligned} \tag{4.43}$$

これは，式（4.7）の正規方程式を ε_i で置き換えた次の式が，0 となることを用いている．

$$\sum_{i=1}^{n}\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right)=\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)=\sum_{i=1}^{n}\hat{\varepsilon}_i=0 \tag{4.44}$$

$$\sum_{i=1}^{n}\left[\left(Y_i-\hat{\beta}_0-\hat{\beta}_1X_i\right)X_i\right]=\sum_{i=1}^{n}\left[\left(Y_i-\hat{Y}_i\right)X_i\right]=\sum_{i=1}^{n}\hat{\varepsilon}_iX_i=0 \tag{4.45}$$

4.6. 逆推定値に対する各種の 95%信頼区間の推定

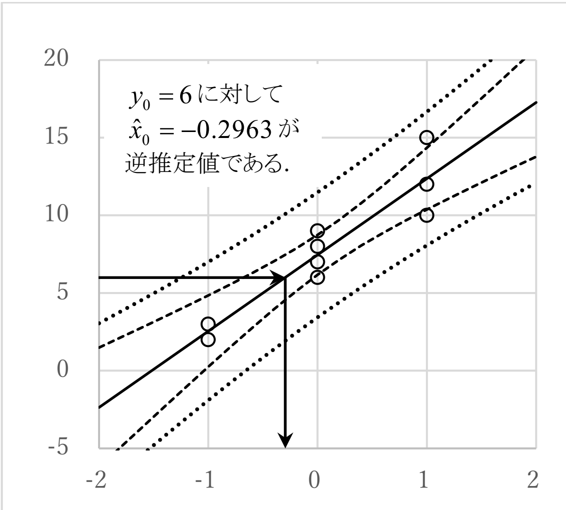
図 4.1 に示すように逆推定は，検量線に対して未知検体の濃度を知りたい場合に，未知検体から反応 y_0 が得られた場合の濃度 $\hat{x}_0$ を推定する問題である．検量線が直線の場合に

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_0$$

の関係から，

$$\begin{aligned}\hat{x}_0 &= \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \\ &= \frac{6.0 - 7.4545}{7.9091} = -0.2963\end{aligned} \tag{4.46}$$

項	推定値
β_0^=	7.4545
β_1^=	4.9091



共分散	0.2952	-0.0590
行列	-0.0590	0.5313
	$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^2	

図 4.1　逆推定の例示

と，容易に濃度 $\hat{x}_0$ を得ることができる．難しいのは，$\hat{x}_0$ の 95%信頼区間の算出である．これは，推定された逆推定値 $\hat{x}_0$ が回帰パラメータの比で表されており，一般的な線形式に対する合成分散の計算公式が使えないからである．

逆推定値の 95%信頼区間の求め方について，参考になるのは，竹内（1979），「数理統計学，第 29 章 IV 回帰直線自体についての推論」である．JMP を用いた逆推定については，芳賀（2010），「医薬品開発のための統計学，第 3 部 非線形モデル，第 1 章 (4) JMP による逆推定の解析」が詳しい．パラメータの共分散分散行列を活用については，Collett (2003)，「Modeling Binary Data 2nd. ed., 4.2.1 Approximate standard error of an estimated effective dose」が参考になる．逆推定の 95%信頼区間に関して，高橋（2013a），「応用回帰分析 I －各種の重み付き回帰における逆推定－」も，第 9 章で文献に対する「文献に対する批判的吟味」を行っている．高橋（2013b），「回帰分析・再入門 －統計ソフトが対応していない生物統計の各種の課題を Excel

でサクサク解こう」は，「基礎セミナー じっくり勉強すれば身につく統計入門」シリーズの第 7 回目の資料集で，本章でのデザイン行列を用いた回帰分析についてスライドを用いて説明をしている．

デルタ法による近似 95%信頼区間

任意の合成分散式に対しては，デルタ法によって合成分散を求めることができる．そのために，求めたい $\hat{x}_0$ の式（4.46）に対して，$\hat{\beta}_0$ と $\hat{\beta}_1$ で偏微分し，

$$d_0 = \frac{\partial \hat{x}_0}{\partial \hat{\beta}_0} = \frac{1}{\hat{\beta}_1} \tag{4.47}$$

$$d_1 = \frac{\partial \hat{x}_0}{\partial \hat{\beta}_1} = \frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^{\,2}} \tag{4.48}$$

これらの偏微分式を行ベクトル

$$\boldsymbol{d} = [d_0 \ d_1] \tag{4.49}$$

としたとき，$\hat{x}_0$ の分散 $Var(\hat{x}_0)$ は，

$$\begin{aligned} Var(\hat{x}_0) &= \boldsymbol{d}\,\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\,\boldsymbol{d}^T \\ &= [d_0 \ d_1]\begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_0) \end{bmatrix}\begin{bmatrix} d_0 \\ d_1 \end{bmatrix} \end{aligned} \tag{4.50}$$

の 2 次形式の計算で求めることができる．

表 4.8　逆推定の近似 95%信頼区間

y_0	x_0^	d_0	d_1	$Var(X_0$^$)$	$L95\%$	$U95\%$	個別$L95$	個別$U95$
2	-1.1111	0.2037	0.2263	0.0340	-1.5473	-0.6750	?	?
3	-0.9074	0.2037	0.1848	0.0260	-1.2884	-0.5265		
6	-0.2963	0.2037	0.0604	0.0127	-0.5631	-0.0295		
7	-0.0926	0.2037	0.0189	0.0120	-0.3514	0.1663		
8	0.1111	0.2037	-0.0226	0.0131	-0.1592	0.3814		
9	0.3148	0.2037	-0.0641	0.0160	0.0159	0.6137		
10	0.5185	0.2037	-0.1056	0.0207	0.1782	0.8589		
12	0.9259	0.2037	-0.1886	0.0357	0.4792	1.3726		
15	1.5370	0.2037	-0.3131	0.0719	0.9032	2.1709		

表 4.8 の 1 行目の $y_0 = 2$ に対する逆推定値 $\hat{x}_0$ は，表 4.2 の $\hat{\beta}_0$ と $\hat{\beta}_1$ の推定値を用いて

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{2 - 7.4545}{4.9091} = -1.1111$$

であり，$\hat{x}_0$ に対する $\hat{\beta}_0$ と $\hat{\beta}_1$ で偏微分式は，

$$d_0 = \frac{1}{\hat{\beta}_1} = \frac{1}{4.9091} = 0.2037$$

$$d_1 = \frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2} = \frac{-(2-7.4545)}{4.9091^2} = 0.2263$$

$$\boldsymbol{d} = [0.2037 \;\; 0.2263]$$

なので，分散$Var(\hat{x}_0)$は，

$$\begin{aligned} Var(\hat{x}_0) &= \boldsymbol{d}\,\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})\boldsymbol{d}^T \\ &= \begin{bmatrix} 0.2037 & 0.2263 \end{bmatrix} \begin{bmatrix} 0.2952 & -0.0590 \\ -0.0590 & 0.5313 \end{bmatrix} \begin{bmatrix} 0.2037 \\ 0.2263 \end{bmatrix} \\ &= 0.0340 \end{aligned}$$

と計算されている．下側 95%点と上側 95%点は，

$$\begin{aligned} L95\% &= \hat{x}_0 - t_{0.05}(9-2)\sqrt{Var(\hat{x}_0)} \\ &= -1.1111 - 2.3646 \times \sqrt{0.0340} = -1.5473 \\ U95\% &= \hat{x}_0 + t_{0.05}(9-2)\sqrt{Var(\hat{x}_0)} \\ &= -1.1111 + 2.3646 \times \sqrt{0.0340} = -0.6750 \end{aligned}$$

で計算されている．第 1 行目で作成した計算式は，行方向にフィルハンドルでコピーすることにより計算される．なお，この方法では，個別データの 95%信頼区間の計算ができない．

逆推定値に対する正確な 95%信頼区間

逆推定値に対する正確な 95%信頼区間の算出方法には，いくつかの方法があるので，図 4.1 で示した回帰直線の推定値 $\hat{y}$ の 95%信頼区間を活用する方法を示す．ある $y_0 = 6$ に対する逆推定値は，

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{6 - 7.4545}{4.9091} = 0.2693$$

と推定される．図 4.2 に示すように，ある $y_0 = 6$ の水平線 2 つの 95%信頼区間の交点に着目する．水平線と 95%信頼上側曲線の交点の X 軸 $\hat{x}_{L95}$ が，$\hat{x}_0 = -0.269$ の X 軸方向の下側 95%点となるが，このままでは推定できない．何らかの探索的な方法が必要となる．

交点の推定値 $\hat{x}_{L95}$ に対する回帰の推定値 $\hat{y}_{L95}$ は，未知の $\hat{x}_{L95}$ を用いて

$$\hat{y}_{L95} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} \tag{4.51}$$

である．この $\hat{y}_{L95}$ に対する Y 軸方向の上側 95%点は，$y_0 = 6$ と等しいので，次の等式が成り立つ．

$$y_0 = \hat{y}_{L95} + t_{0.05}\sqrt{Var(\hat{y}_{L95})} \tag{4.52}$$

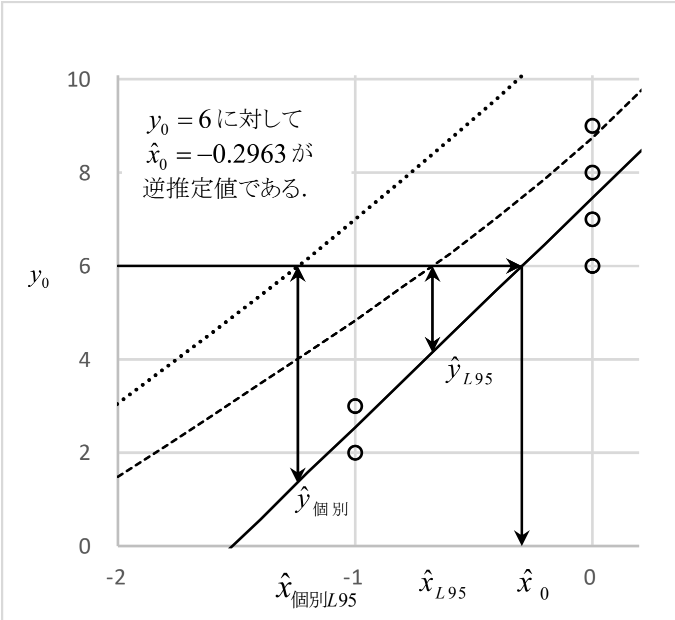


図 4.2　逆推定の 95%信頼区間の算出の例示

推定したいのは，$\hat{x}_{L95}$ なので，$\hat{y}_{L95}$ を $\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95}$

$$y_0=\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95}+t_{0.05}\sqrt{Var(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95})} \tag{4.53}$$

に置き換える．この式を $\hat{x}_{L95}$ について解くことにより，逆推定値 $\hat{x}_{L95}$ の下側 95%点が推定できる．上の式を右辺の $\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95}$ を左辺に移し，右辺の $t_{0.05}$ を左辺の分母とし，両辺を平方すると，次式が得られる．

$$\left(\frac{y_0-\hat{\beta}_0-\hat{\beta}_1\hat{x}_{L95}}{t_{0.05}}\right)^2=Var(\hat{\beta}_0)+2Cov(\hat{\beta}_0,\hat{\beta}_1)\hat{x}_{L95}+Var(\hat{\beta}_1)\hat{x}_{L95}^2 \tag{4.54}$$

左辺を右辺に移して $\hat{x}_{L95}$ について整理すると，

$$\left[Var(\hat{\beta}_0)-\frac{(y_0-\hat{\beta}_0)^2}{t_\alpha^2}\right]+\left[2Cov(\hat{\beta}_0,\hat{\beta}_1)+\frac{2(y_0-\hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2}\right]\hat{x}_{L95}+\left[Var(\hat{\beta}_1)-\frac{\hat{\beta}_1^2}{t_\alpha^2}\right]\hat{x}_{L95}^2=0 \tag{4.55}$$

が得られる．この複雑な式は，幸い $\hat{x}_{L95}$ に関する 2 次式に

$$a+b\hat{x}_{L95}+c\hat{x}_{L95}^2=0 \tag{4.56}$$

なるので，2 次式の解の公式

$$\hat{x}_{L95}=\frac{-b\pm\sqrt{b^2-4ac}}{2c} \tag{4.57}$$

により，$\hat{x}_{L95}$ を求めることができる．解は 2 つあるが，小さい方が $\hat{x}_{L95}$ となり，大きい方が上側 95%点 $\hat{x}_{U95}$ となる．

個別データの正確な 95%信頼区間

個別の上側 95%曲線と $y_0 = 6$ を通る水平線との交点は，$\hat{x}_{個別L95}$ に対する回帰の推定値

$$\hat{y}_{個別L95} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{個別L95} \tag{4.58}$$

の上側 95%点でもある．個別の上側 95%点は，y_0 と等しいので，次の等式が成り立つ．

$$y_0 = \hat{y}_{L95} + t_{0.05}\sqrt{Var(\hat{y}_{個別L95}) + \hat{\sigma}^2} \tag{4.59}$$

推定したいのは，$\hat{x}_{個別L95}$ なので，

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{個別L95} + t_{0.05}\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{個別L95}) + \hat{\sigma}^2} \tag{4.60}$$

と置き換える．式（4.59）を $\hat{x}_{個別L95}$ について解くと式（4.55）と同様な結果となるが，2 次式のパラメータ a について

$$個別 a' = a + \hat{\sigma}^2$$

となるのみで，他は同様である．

これらから，表 4.9 に示すように，与えられた y_0 に対する逆推定値の 95%信頼区間，および，個別データの 95%信頼区間を推定することができる．

表 4.9　逆推定の正確な 2 種類の 95%信頼区間

	2次式のパラメータ			逆推定	95%信頼区間		個別	個別95% CL	
y_0	a	b	c	$x^\wedge_0$	$x^\wedge_{L95}$	$x^\wedge_{U95}$	a'	個別$L95$	個別$U95$
2	-5.0258	-9.6959	-3.7787	-1.1111	-1.8450	-0.7209	-2.4284	-2.2846	-0.2813
3	-3.2536	-7.9399	-3.7787	-0.9074	-1.5433	-0.5579	-0.6562	-2.0150	-0.0862
6	-0.0832	-2.6721	-3.7787	-0.2963	-0.6745	-0.0327	2.5142	-1.2426	0.5354
7	0.2582	-0.9162	-3.7787	-0.0926	-0.4094	0.1669	2.8556	-0.9990	0.7565
8	0.2419	0.8397	-3.7787	0.1111	-0.1652	0.3875	2.8394	-0.7628	0.9850
9	-0.1320	2.5956	-3.7787	0.3148	0.0553	0.6316	2.4654	-0.5343	1.2212
10	-0.8636	4.3516	-3.7787	0.5185	0.2549	0.8967	1.7338	-0.3132	1.4648
12	-3.4000	7.8634	-3.7787	0.9259	0.6129	1.4681	-0.8026	0.1076	1.9733
15	-9.8872	13.1312	-3.7787	1.5370	1.1031	2.3719	-7.2898	0.6936	2.7815

表 4.8 の 1 行目の $y_0 = 2$ の逆推定値 $\hat{x}_0$ は，表 4.2 の $\hat{\beta}_0$ と $\hat{\beta}_1$ の推定値を用いて

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{2 - 7.4545}{4.9091} = -1.1111$$

である．2 次式のパラメータ a，b，c は，

$$a = Var(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} = 0.2952 - \frac{(2 - 7.4545)^2}{2.3646^2} = -5.0258$$

$$b = 2Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2 \times (-0.0590) + \frac{2 \times (2 - 7.4545) \times 4.9091}{2.3646^2} = -9.6959$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_\alpha^2} = 0.5313 - \frac{7.4545^2}{2.3646^2} = -3.7778$$

表 4.2：

β_0^=	7.4545	$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$=	18.1818	共分散	0.2952	-0.0590
β_1^=	4.9091	σ^2=	2.5974	行列	-0.0590	0.5313
$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$		$t_{0.05}$=	2.3646		$\boldsymbol{\Sigma}(\boldsymbol{\beta}$^$)=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma$^2	

となる．下側 95%点 $\hat{x}_{L95}$ は，

$$\begin{aligned}\hat{x}_{L95} &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \\ &= \frac{-(-9.6959) \pm \sqrt{(-9.6959)^2 - 4 \times (-5.0258) \times (-3.7787)}}{2 \times (-3.7778)} = (-1.8450, -0.7209)\end{aligned}$$

となる．個別の下側 95%点は，

$$個別 a' = a + \hat{\sigma}^2 = -5.0258 + 2.5974 = -2.4284$$

なので，

$$\begin{aligned}\hat{x}_{個別L95} &= \frac{-b \pm \sqrt{b^2 - 4a'c}}{2c} \\ &= \frac{-(-9.6958) \pm \sqrt{(-9.6959)^2 - 4 \times (-2.4284) \times (-3.7787)}}{2 \times (-3.7787)} = (-2.2846, -0.2813)\end{aligned}$$

と計算される．図 4.2 の逆推定の 95%信頼区間の算出の例示で用いた $\hat{x}_{L95}$ に関する逆推定値は，表 4.9 から該当部分を抜粋した結果を表 4.10 に示すように，$\hat{x}_0 = -0.296$，$\hat{x}_{L95} = -0.675$，$\hat{x}_{個別L95} = -1.243$ である．

表 4.10　近似および正確な逆推定値の比較

		逆推定	95%信頼区間		個別95% CL	
	y_0	x^$_0$	x^$_{L95}$	x^$_{U95}$	個別L95	個別U95
正確	6	-0.2963	-0.6745	-0.0327	-1.2426	0.5354
近似	"	"	-0.5631	-0.0295	?	?

Excel ソルバーを用いた逆推定の正確な 95%信頼区間

Excel ソルバーには，目標値を最大化または最小化する以外に，設定した推定値となるように変数セルを変更する機能がある．この機能を使うことにより，逆推定値 y_0 の水平線と 95%信頼区間の曲線が交わる点を目標値に設定する．Excel のソルバーで $\hat{x}_{L95}$ または $\hat{x}_{U95}$ を変化させて

$$\hat{\beta}_0 + \hat{\beta}_1\hat{x}_{L95},\quad \hat{\beta}_0 + \hat{\beta}_1\hat{x}_{U95} \;\rightarrow\; y_0$$

y_0 になるような推定値 $\hat{x}_{L95}$ または $\hat{x}_{U95}$ を得る．表 4.2 に示した回帰分析の結果から，

		β_0^=	7.4545	共分散	0.2952	-0.0590	σ^2=	2.5974	
		β_1^=	4.9091	行列	-0.0590	0.5313	$t_{0.05}$=	2.3646	

を用いて，$x=0$ に対して

$$\hat{y}=7.4545+4.9091\times 0=7.4545$$

$$Var(\hat{y})=\begin{bmatrix}1 & 0\end{bmatrix}\begin{bmatrix}0.2952 & -0.0590\\ -0.0590 & 0.5313\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}=0.2952$$

$$U95\%=7.4545+2.3646\times\sqrt{0.2952}=8.7392$$

として計算式をセットし，ソルバーで目標値の探索で用いる．

表 4.11　Excel ソルバーによる逆推定の実際

		β_0^=	7.4545	共分散	0.2952	-0.0590	σ^2=	2.5974	
		β_1^=	4.9091	行列	-0.0590	0.5313	$t_{0.05}$=	2.3646	
		X_0	X_1	Y^	$Var(Y$^$)$	$L95\%$	$U95\%$	個別 $L95\%$	個別 $U95\%$
推定値		1	**0**	7.4545	0.2952	6.1699	**8.7392**	3.4329	11.4762
回帰下限		1	**-0.6745**	4.1433	0.6165		**6.0000**	：目標値(6 にセット)	
回帰上限		1	**-0.0327**	7.2943	0.2996	**6.0000**			
個別下限		1	**-1.2426**	1.3545	1.2622				**6.0000**
個別上限		1	**0.5354**	10.0831	0.3843			**6.0000**	

この $x=0$ に対する推定値全体を下段にコピー&ペーストし，$U95\%=\boxed{8.7392}$ のセルの位置を Excel ソルバーの目標値としてセットし，さらに「指定値=6」とセットする．「変数セルの変更」に $x=\boxed{0}$ の位置としてセットし実行すると，回帰の逆推定値の 95%信頼区間の下限が–0.6745 との結果を得る．

推定値	1	**0**	7.4545	0.2952	6.1699	**8.7392**
回帰下限	1	**-0.6745**	4.1433	0.6165		**6.0000**

次いで，$L95\%$ について同様の手順を繰り返すことにより，$y_0=6$ に対する 95%信頼区間として（–0.6745，–0.0327）が推定される．更に個別の $U95\%$ と $L95\%$ については，（–1.2426，–0.5345）と推定される．これは，表 4.10 に示した指定結果に一致する．

4.7. JMP による回帰分析と逆推定

これまで，Excel を用いて回帰分析のパラメータの共分散行列を活用し，各種の 95%信頼区間について示してきた．統計ソフト JMP の「二変量のあてはめ」によって回帰直線の 95%信頼区間のきれいなフラフを手軽に作成できるのであるが，内部でどのような計算式が使われているかを出力させることが可能であり，大変興味深い．また，JMP の「モデルのあてはめ」を使うことにより，逆推定も手軽に行うことができる．

「二変量の関係」による回帰分析

表 4.12 に JMP の「二変量の関係」による回帰分析の結果を示す．「分散分析」の結果は，表 4.4 で示した結果に相当し，「パラメータの推定値」は，表 4.2 に相当する．さらに，回帰直線の 95%信頼区間，および，個別データの 95%信頼区間を計算式付きで JMP ファイルへ出力することもできる．

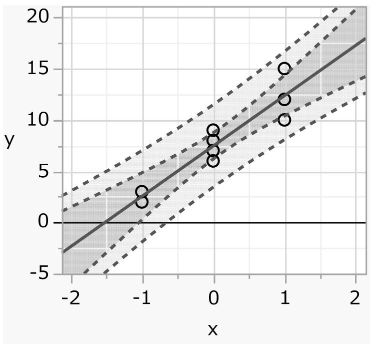
表 4.12　JMP の「二変量の関係」による回帰分析

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	117.8182	117.8182	45.3600
誤差	7	18.1818	2.5974	**p値(Prob>F)**
全体(修正済み)	8	136.0000		0.0003*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	7.4545	0.5433	13.72	<.0001*
x	4.9091	0.7289	6.73	0.0003*



回帰直線の 95%信頼区間の計算式

表 4.14 の「下側 95%」および「上側 95%」の欄は，計算式を含めた出力になっており，内部での計算式を確認できるようになっている．回帰の 95%信頼区間の下限については，表 4.13 に示すような計算式を表示させることができる．

表 4.13　JMP の「二変量の関係」による上側 95%の計算式

$$\left(7.4545454545 + 4.9090909091 \cdot x\right) - 2.3646242516 \cdot \sqrt{\text{Vec Quadratic}\left(\begin{bmatrix}0.114 & -0.023\\ -0.023 & 0.205\end{bmatrix}, [1]\,||\,x\right) \cdot 2.5974025974}$$

表 4.14　JMP の「二変量の関係」による信頼区間の出力

	x	y	予測式 y	下側95%	上側95%	L95% 個別	U95% 個別
1	-1	2	2.52	0.25	4.84	-1.90	7.00
2	-1	3	2.52	0.25	4.84	-1.90	7.00
3	0	6	7.45	6.17	8.74	3.43	11.48
4	0	7	7.45	6.17	8.74	3.43	11.48
5	0	8	7.45	6.17	8.74	3.43	11.48
6	0	9	7.45	6.17	8.74	3.43	11.48
7	1	10	12.39	10.37	14.35	8.06	16.66
8	1	12	12.39	10.37	14.35	8.06	16.66
9	1	15	12.39	10.37	14.35	8.06	16.66
10	-2	•	-2.42	-6.22	1.49	-7.78	3.06
11	-1.5	•	0.05	-2.96	3.14	-4.79	4.97
12	-0.5	•	4.98	3.35	6.65	0.85	9.15
13	0.5	•	9.92	8.47	11.35	5.84	13.98
14	1.5	•	14.85	12.11	17.53	10.14	19.49
15	2	•	17.32	13.78	20.77	12.10	22.44

注）10 行目～15 行目は，グラフ表示の際のため，x をあたえて予測値などの計算をさせている．

計算式内の VecQuadraticc()関数は，デザイン行列の計算 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ 結果に $\boldsymbol{X}$ の行ベクトルとの 2 次形式の計算を行う関数であり，誤差分散 $\hat{\sigma}^2=2.5974$ を掛け，$\hat{Y}_i$ の分散 $Var(\hat{Y}_i)$ の平方根が計算されている．$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ に $\hat{\sigma}^2$ を掛けた結果は，パラメータの共分散行列 $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ となっており，また，推定値 $\hat{Y}_i$ の分散の計算にも $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}})$ を挟んだ 2 次形式になっていて，伝統的な偏差平方和を用いた計算ではないことが確認できる．

1	x		0.1136	-0.0227	1	*	2.5974
			-0.0227	0.2045	x		
			$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$				σ^2

回帰の 95%信頼区間の下限の計算式は,

$$L95\%=\hat{Y}_i-t_{0.05}(7)\sqrt{Var(\hat{Y}_i)}$$

であることが，読み取れる．

個別データの 95%信頼区間は,「$L95\%$ 個別」の欄の計算式から，表 4.15 に示すように 2 次

$$L95\%個別=\hat{Y}_i-t_{0.05}(7)\sqrt{Var(\hat{Y}_i)+\hat{\sigma}^2}$$

$$\sqrt{Var(\hat{Y}_i)+\hat{\sigma}^2}=\sqrt{\text{Vec Quadratic}((\boldsymbol{X}^T\boldsymbol{X})^{-1},[1\ \ X])\hat{\sigma}^2+\hat{\sigma}^2}$$

形式の関数が使われていることがわかる．これらのことから，JMP での回帰分析は，デザイン行列ベースの計算となっていることを垣間見ることができる．

表 4.15　JMP の「二変量の関係」による個別データの下側 95%の計算式

$$\left(7.4545454545 + 4.9090909091 \cdot x\right) - 2.3646242516 \cdot \sqrt{\mathrm{Vec\ Quadratic}\left(\begin{bmatrix}0.114 & -0.023\\ -0.023 & 0.205\end{bmatrix}, \begin{bmatrix}1\end{bmatrix} || x\right) \cdot 2.5974025974 + 2.5974025974}$$

「モデルのあてはめ」による逆推定

逆推定を行い，Excle での計算結果との相互検証を行う.「二変量の関係」での回帰分析は，信頼区間の標示などで豊富な機能があるが，逆推定には対応していない.「モデルのあてはめ」による回帰分析を用いて逆推定を行ない，表 4.17 に結果を示す．

表 4.16　JMP の「逆推定」の設定

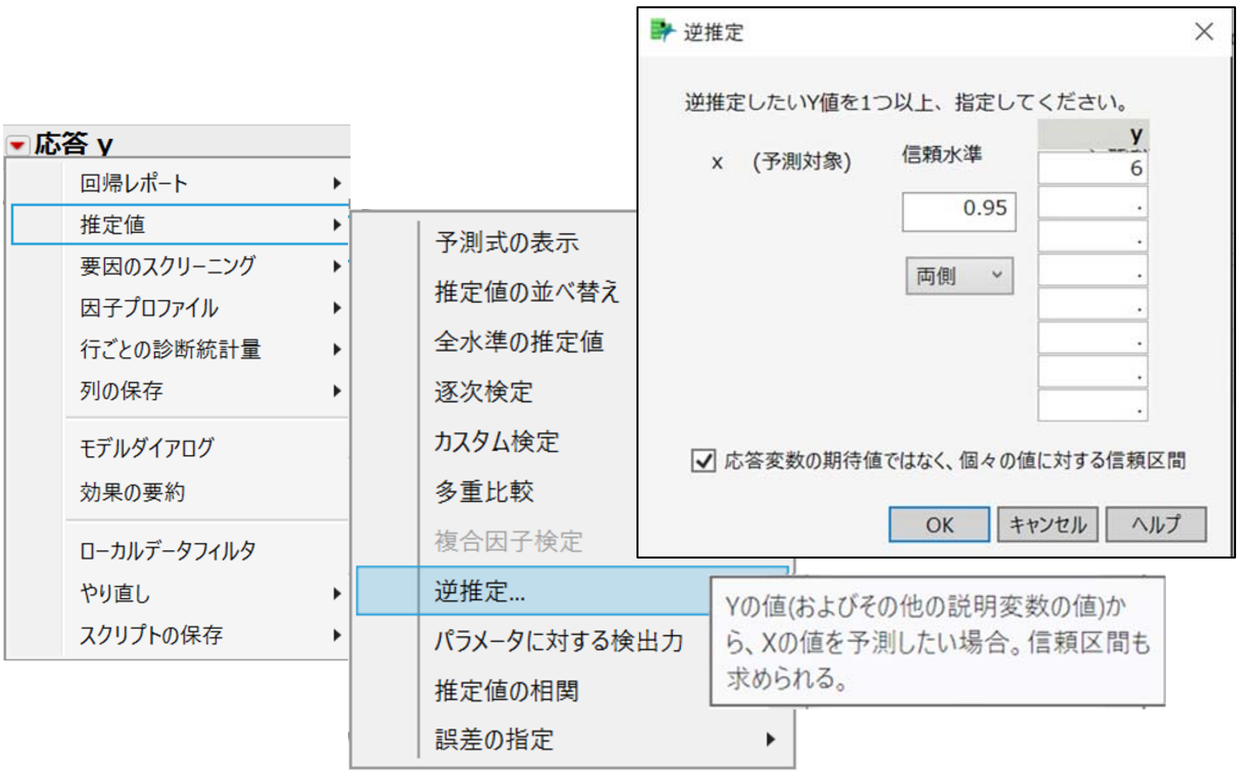


表 4.17　JMP の「モデルのあてはめ」による逆推定

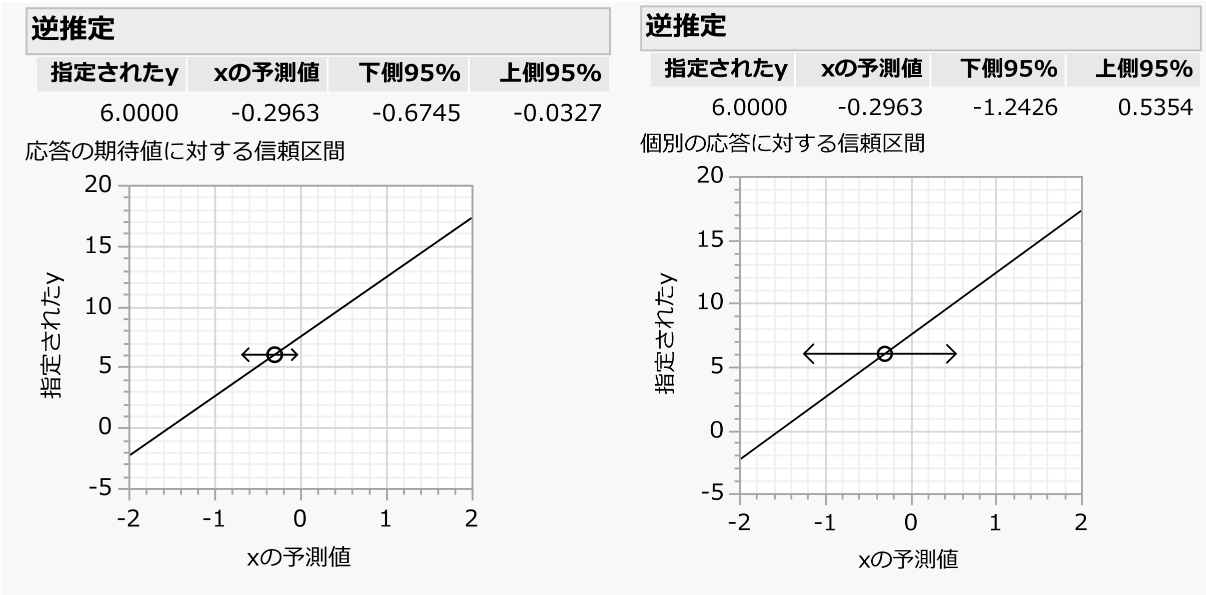


表 4.12 の「分散分析」および「パラメータの推定値」と全く同じ結果が得られることを確認し，$y_0 = 6$ についての逆推定を行う．なお，この結果は，表 4.10 と完全に一致する．

表 4.10 抜粋

	逆推定	95%信頼区間		個別95% CL	
y_0	$x\hat{}_0$	$x\hat{}_{L95}$	$x\hat{}_{U95}$	個別$L95$	個別$U95$
6	-0.2963	-0.6745	-0.0327	-1.2426	0.5354

通常の回帰分析でのパラメータの共分散行列を用いた逆推定の計算方法は，ポアソン回帰，ワイブル回帰のみならず，ロジスティック回帰による 10%あるいは 50%有効量の推定なども同様に適用できる汎用的な方法である．

非線形回帰を用いた逆推定値の 95%信頼区間の直接推定

回帰直線を推定し，得られたパラメータの推定値と共分散行列を用いて，ある y_0 に対する逆推定値 $\hat{x}_0$ を推定する方法を示してきた．通常の回帰分析は，次式

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (4.61)$$

で与えられる．ある y_0 に対する逆推定値 $\hat{x}_0$ を通る回帰式は，

$$Y_i - y_0 = \beta_1 (X_i - \hat{x}_0) + \varepsilon_i \qquad (4.62)$$

で与えられる．

ある y_0 を観測値 Y_i の平均値とし，$\hat{x}_0$ を X_i の平均とすると式（4.62）は，重心を通る回帰式となり，パラメータは傾き β_1 のみとなる．切片 β_0 は，

$$\beta_0 = \overline{Y} + \beta_1 \overline{X} \tag{4.63}$$

で計算できる．ただし，ある y_0 に対する逆推定値 $\hat{x}_0$ は，回帰パラメータが推定されないと求めることができない．式（4.62）を

$$Y_i = y_0 + \beta_1 (X_i - \hat{x}_0) + \varepsilon_i \tag{4.64}$$

のように，ある y_0 を右辺に移項した式とする．この式を用いて，β_1 と $\hat{x}_0$ を同時に推定し，それらの分散も推定したい．

式（4.64）パラメータに関して偏微分すると，

$$\frac{\partial Y_i}{\partial \beta_1} = (X_i - \hat{x}_0)\,, \qquad \frac{\partial Y_i}{\partial \hat{x}_0} = -\beta_1 \tag{4.65}$$

となり，互いに他のパラメータが残り，線形式ではなくなり通常の回帰分析で解くことができない．そのために，偏微分式を新たな変数とした反復回帰が必要となる［ドレーパ・スミス（1968），第 10 章 非線形推定序説］.

式（4.64）に対して JMP の非線形回帰を使えば，表 4.17 右に示した回帰の逆推定 $\hat{x}_0$ の正確な 95%信頼区間を直接推定することができる．詳しくは，芳賀（2010）の非線形についての解説資料，大和田（2010），線形モデルと非線形モデルの基本的な考え方－逆推定の解析，標準誤差と信頼限界－（https://biostat.jp/archive_teireikai_2_download.php?id=19），中西（2016），じっくり勉強すれば身につく統計入門 第 12 回－非線形回帰を用いた逆推定の基礎－（https://scientist-press.com/tokei-nyumon/）を参照のこと．

第 4 章 文 献 索　引

第 4 章 索　引

第 4 章　解析用ファイル一覧

22 KB	第4章01_回帰_入門	Microsoft Excel ワークシート
14 KB	第4章02_回帰_正規方程式	Microsoft Excel ワークシート
13 KB	第4章03_回帰_逆行列	Microsoft Excel ワークシート
41 KB	第4章05_回帰_デザイン行列	Microsoft Excel ワークシート
63 KB	第4章06_回帰_逆推定	Microsoft Excel ワークシート
12 KB	第4章07_回帰_JMP	Microsoft Excel ワークシート
6 KB	第4章07_回帰_逆推定	JMP Data Table

非売品，無断複製を禁ずる

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 4 章＞＞

第 4 章　デザイン行列を用いた回帰分析入門

BioStat 研究所（株）

〒105-0014 東京都 港区 芝 1-12－3 の 1005

2020 年 6 月 5 日　　高橋 行雄