

第 12 回 続高橋セミナー
層別因子を含む探索的な回帰分析入門
2024 年 2 月 5 日

第 5 章 デザイン行列を用いた回帰分析の基礎

シグマが出てきただけで読み飛ばしたくなる人たちの気持ちは、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らないからだと思われる。ましてや、行列の計算式が出てきたときに、実際のデータで自ら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だけでなく嫌悪感さえ抱くと推測する。そのために、多くの人達が、統計ソフトの使い方、結果の見方さえ習得すれば、十分だと思ってしまいがちになるのではないだろうか。このような人達を念頭にし、シグマによって記述されている回帰分析を題材にし、Excel の基本的な計算機能を用いることにより、何の苦痛もなくサクサクとシグマの計算ができる事を示す。さらに、Excel シート上に展開された矩形データが行列の基本であることを導入し、Excel の行列関数により、複雑なシグマで表記された計算が、瞬時にできることを示す。本章では、これらの計算手段を用いることにより、伝統的なシグマ表記の回帰分析に代わる、新しい解析手順を導入する。

第 5 章 目 次

5.	デザイン行列を用いた回帰分析入門 -----	155
5.1.	回帰分析における各種の推定 -----	155
5.2.	デザイン行列を用いた行列計算の基礎 -----	157
	Excel シート上にデザイン行列 X を設定、 行列計算の実際、 デザイン行列の 転置と積和、 シグマを用いた積和の計算、 デザイン行列 X の積和、 デザイン 行列 X と反応ベクトル Y との積和、 英語版 Wikipedia に見るデザイン行列の定義	
5.3.	偏差平方和による回帰パラメータの推定 -----	164
	回帰パラメータの推定、 正規方程式、 正規方程式の解による パラメータの推定、 偏差平方を用いたパラメータの推定の実際	

次ページ続く

5.4. デザイン行列 vs. 偏差平方和を用いた回帰分析 -----	169
<p>行列計算によるパラメータ推定, デザイン行列 X と偏差平方和での 推定式の相違, 偏差平方和ベースのパラメータの分散の推定, シグマを用いた回帰分析の功罪</p>	
5.5. Excel の行列計算による回帰分析の実際 -----	174
<p>Excel の行列計算による回帰パラメータの推定, 回帰パラメータの 分散および共分散の推定, Excel の関数を用いた分散分析表の作成, パラメータの共分散行列の活用, 回帰直線の 95%信頼区間, 伝統的な方法, 現実的な対応, 平方和の分解に対する補足</p>	
5.6. 逆推定値に対する各種の 95%信頼区間の推定 -----	183
<p>逆推定とは何か, デルタ法による逆推定値に対する近似 95%信頼区間, 逆推定値に対する正確な 95%信頼区間, 個別データに対する逆推定値の 正確な 95%信頼区間, Excel ソルバーを用いた逆推定の正確な 95%信頼区間, 面積の標準偏差を推定したい</p>	
5.7. JMP による回帰分析と逆推定 -----	191
<p>JMP の「二変量の関係」による回帰分析, 回帰直線の 95%信頼区間の 計算式, 「モデルのあてはめ」による逆推定値の 95%信頼区間, 逆推定値の 95%信頼区間の直接推定</p>	
<p>文献索引, 索引, 解析ファイル一覧 (195)</p>	

第 12 回 続・高橋セミナー 層別因子を含む探索的な回帰分析入門

目次 (全章)

はじめに -----	1
1. 層別因子を含む各種の回帰分析の実例 -----	7
2. デザイン行列を活用した 1 因子実験データの解析 -----	55
3. 繰り返しが不揃いの 2 因子実験データの解析 -----	89
4. 欠測値がある直交表の線形モデルによる解析 -----	121
5. デザイン行列を用いた回帰分析の基礎 -----	155
6. 伝統的な共分散分析からの脱却 -----	195
7. 共変量を含む 3 因子実験データの探索的解析 -----	219
8. 交絡変数と共変量を含む 2 群比較 -----	243
9. 前後差の前値は常に共変量 -----	281
10. 層別因子を含むロジスティック曲線のあてはめ -----	327
11. 各種のシグモイド曲線を用いた逆推定 -----	367
12. ミカエリス・メンテン式をめぐる新たな統計解析 -----	405
13. ロジスティック曲線のさらなる活用 -----	455
文献, 文献索引, 索引, 解析用ファイル一覧 -----	489

5. デザイン行列を用いた回帰分析の基礎

シグマが出てきただけで読み飛ばしたくなる人たちの気持ちは、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らないからだと思われる。ましてや、行列の計算式が出てきたときに、実際のデータで自ら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だけでなく嫌悪感さえ抱くと推測する。そのために、多くの人達が、統計ソフトの使い方、結果の見方さえ習得すれば、十分だと思ってしまいがちになるのではないだろうか。このような人達を念頭にし、シグマによって記述されている回帰分析を題材にし、Excel の基本的な計算機能を用いることにより、何の苦痛もなくサクサクとシグマの計算ができる事を示す。さらに、Excel シート上に展開された矩形データが行列の基本であることを導入し、Excel の行列関数により、複雑なシグマで表記された計算が、瞬時にできることを示す。本章では、これらの計算手段を用いることにより、伝統的なシグマ表記の回帰分析に代わる、新しい解析手順を導入する。

5.1. 回帰分析における各種の推定

第1章では、各種の質的変数を含む回帰分析の事例について、Excel の分析ツールの一つである「回帰分析」を用いた解析方法を示し、統計ソフトによる検証も合わせて示してきた。Excel の統計解析の機能は、限定的であり正確性について MaCullough ら (2008), 「On the accuracy of statistical procedures in Microsoft Excel 2007」での Excel 2007 に対する批判があることは十分承知している。幸いなことに、Excel の回帰分析に限ってみれば、各種の推定結果のグラフ表示への連携などにより、総合的には統計ソフトに勝るとも劣らないことを示してきた。

質的変数を含む場合に、いわゆるダミー変数を使うことにより、統計ソフトの回帰分析による解析ができることは、良く知られている。これまでの章では、質的変数に関するダミー変数と共変量の積で生成された交互作用の変数を含むデザイン行列 \mathbf{X} を用い、Excel の回帰分析を適用し、さらに、行列関数による回帰パラメータの共分散行列 $\Sigma(\hat{\beta})$ あるいは $\Sigma(\hat{\theta})$ の推定方法を示してきた。回帰分析に派生する各種の推定値の分散を求めるために、このパラメータの共分散行列 $\Sigma(\hat{\beta})$ が中心的な役割を演じていることを強調してきた。

各種の推定値に対して 95%信頼区間を計算し、その結果のグラフ表示に散布図および折れ線グラフを活用し、統計ソフトでは対応できない複合的なグラフ表示が、同じ計算シート内で容易にできることも示してきた。特に交互作用がある場合のこれらのグラフ表示は、結果の統計的な解釈をする際に助けになる。

質的変数が含まれる場合に Excel の「回帰分析」を用いる場合は、自らダミー変数を生成する必要がある煩わしい。多くの統計ソフトでは、内部でダミー変数に展開し、交互作用を含めた場合でも内部でしかるべき変数に展開するので、大変便利である。ただし、第 1.4 節の SAS による結果の検証では、各種の推定をする際には、内部で生成されたデザイン行列を念頭にし、推定のため `estimate` ステートメントによる細かな設定が必要で、容易なことではないことも示してきた。

これらの状況を鑑みるにつれ、回帰分析から派生する各種の推定の課題に対処するためには、Excel によるデザイン行列 X の生成、回帰分析の実施、各種の推定、推定値の 95%信頼区間のグラフ化といった一連の解析マナーの習得が必要不可欠と思われる。その結果とし、統計ソフトを自在に使いこなし、活用するために必須の知識と経験が得られるのではないだろうか。

第 1 章では、全く触れなかったが、ほとんど全ての統計に関する成書での回帰分析（重回帰分析を含む）の計算手順は、手計算時代の簡便な画一的な計算法が踏襲されている。そのために、第 1.2 節に示したように共分散分析の解析法もその影響を受け、第 6 章で詳しく述べるが、分散分析と回帰分析の複合的な解析法が示されており、理解し応用することが困難であると認識される。そこで、第 1 章で導入した事例を用いて、デザイン行列 X を用いた回帰分析についての基礎を最初に示し、合わせて伝統的な手計算による解析法を対比し、その問題点を浮き彫りにする。

ドレーパ・スミス著、中村慶一訳（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、同じデータを用いたデザイン行列 X 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 N.R. and Smith H. (1998), *Applied Regression Analysis* 3rd ed. も参考にし、シグマを使った計算に対比したデザイン行列 X を用いた計算法の関係を並列的に示すことにより、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 X を用いた解析の利便性を強調したい。用いるデータは、第 1.3 節の抗うつ薬の臨床試験のデータである。

5.2. デザイン行列 X を用いた行列計算の基礎

Excel シート上にデザイン行列 X を設定


反応変数 Y_i の列ベクトルを \mathbf{Y} ，説明変数のデザイン行列を \mathbf{X} ，推定したいパラメータのベクトルを $\boldsymbol{\beta}$ ，誤差の列ベクトルを $\boldsymbol{\varepsilon}$ とする．データは，表 1.20 の抗うつ薬のプラセボ群のデータである．まず，Excel シートの好きな場所に，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 シート上の該当セルをコピーして，ワード上で「図（拡張メタファイル）」によりペーストした結果が示されている．

[illegible]

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全角の「 β 」となる．半角の「 β 」にするためにフォントを Times New Roman に変更してイタリック **I** を選択すると「 β 」となる．さらに太字にするためにボールド **B** を選択とするとベクトル「 $\boldsymbol{\beta}$ 」となる．

同様に「いぷしろん」で全角の「 ε 」に変換し、フォントを Times New Roman に変更してイタリックにすると「 ε 」となる. 添え字の「1」を加えてイタリックを外して「 ε_1 」とし、「1」を「セル書式の設定」で「下付き」にすると「 ε_1 」となる. 「 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{10}$ 」は、フィルハンドルで「 ε_1 」を引き延ばすと自動的に得られる. パラメータの β_0 と β_1 も同様の手順で作成する.

このように統計で使うギリシャ文字を Excel で自由に扱えるようになることから初めてもらいたい。デザイン行列の「**X**」は、半角の大文字「X」をボード「**X**」とし、さらにイタリックにすると「**X**」となる。

デザイン行列 **X** の一般的な表記は、角括弧 $[\dots]$ あるいは括弧 (\dots) 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(I)」を用いて矩形データ全体を

囲む表記 $\boxed{\dots}$ を用いる．デザイン行列 \mathbf{X} は，大きさが 10 行×2 列で，10 行×1 列のベクトル \mathbf{X}_0 と \mathbf{X}_1 を並べたものである．列ベクトル \mathbf{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 」を含んでいる．

多くの統計ソフトの回帰分析では，モデル式の設定で，説明変数 \mathbf{X}_1 のみを与え，切片を推定するための変数 \mathbf{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 10 行×1 列の列ベクトル \mathbf{X}_1 のみのデザイン行列とする必要がある．そのために，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ことにより原点を通る回帰直線を推定することができる．Excel の「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，



「定数に 0 を使用」オプションをオンと設定する．通常の回帰分析では，「定数に 0 を使用」オプションを選択（オフのまま）せず，列ベクトル \mathbf{X}_1 の範囲のみを設定する．そして，列ベクトル \mathbf{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の回帰分析では，切片を含めない設定を暗黙の前提としている．そのため，原点を通る回帰直線のあてはめるために「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 \mathbf{Y} ，デザイン行列 \mathbf{X} ，列ベクトル $\boldsymbol{\beta}$ ，列ベクトル $\boldsymbol{\varepsilon}$ を用いた回帰モデル式

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5.1)$$

は，何を意味するのだろうか．デザイン行列 \mathbf{X} と列ベクトル $\boldsymbol{\beta}$ の積 $\mathbf{X}\boldsymbol{\beta}$ は，デザイン行列 \mathbf{X} のある行のセルと，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 \mathbf{X} の 1 行目から 10 行目まで繰り返し，大きさが 10×1 の新たな列ベクトルが生成される．右辺の $(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\varepsilon}$ は，大きさが 10×1 の列ベクトル同士の足し算で，同じ行のセル同士の足し算となり，大きさが 10×1 の列ベクトルとなる．

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

$$\mathbf{Y} = (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

は， \mathbf{Y} の行方向に展開して

		$(X\beta)$		ε		$(X\beta + \varepsilon)$
$Y = X\beta + \varepsilon =$		$1\beta_0 + 18\beta_1$	+	ε_1	=	$1\beta_0 + 18\beta_1 + \varepsilon_1$
		$1\beta_0 + 21\beta_1$		ε_2		$1\beta_0 + 21\beta_1 + \varepsilon_2$
		$1\beta_0 + 24\beta_1$		ε_3		$1\beta_0 + 24\beta_1 + \varepsilon_3$
		$1\beta_0 + 24\beta_1$		ε_4		$1\beta_0 + 24\beta_1 + \varepsilon_4$
		$1\beta_0 + 18\beta_1$		ε_5		$1\beta_0 + 18\beta_1 + \varepsilon_5$
		$1\beta_0 + 27\beta_1$		ε_6		$1\beta_0 + 27\beta_1 + \varepsilon_6$
		$1\beta_0 + 21\beta_1$		ε_7		$1\beta_0 + 21\beta_1 + \varepsilon_7$
		$1\beta_0 + 21\beta_1$		ε_8		$1\beta_0 + 21\beta_1 + \varepsilon_8$
		$1\beta_0 + 26\beta_1$		ε_9		$1\beta_0 + 26\beta_1 + \varepsilon_9$
		$1\beta_0 + 20\beta_1$		ε_{10}		$1\beta_0 + 20\beta_1 + \varepsilon_{10}$

$$\left. \begin{aligned} Y_1 &= 1\beta_0 + 18\beta_1 + \varepsilon_1 \\ Y_2 &= 1\beta_0 + 21\beta_1 + \varepsilon_2 \\ &\vdots \\ Y_{10} &= 1\beta_0 + 20\beta_1 + \varepsilon_{10} \end{aligned} \right\} \quad (5.2)$$

として展開できる．また，添え字を使った式

$$Y_i = \beta_0 X_{0,i} + \beta_1 X_{1,i} + \varepsilon_i \quad (i=1, 2, \dots, 10) \quad (5.3)$$

としても，同じ回帰モデルである．デザイン行列 X を使った式 (5.1) 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 β に適当な数値 ($-11.0, 0.8$) を入れて，実際に $X\beta$ 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，

	A	B	C	D	E	F	G	H	I
1		X							
2		X₀	X₁		β		Xβ		
3	X =	1	18		-11.0	=	=Mmult(B3:C12,E3:E4)		
4		1	21		0.8				
5		1	24						
6		1	24						
7		1	18						
8		1	27						
9		1	21						
10		1	21						
11		1	26						
12		1	20						
13		10×2			2×1		10×1		



	A	B	C	D	E	F	G
1		X					
2		X₀	X₁		β		Xβ
3	X =	1	18		-11.0	=	3.40
4		1	21		0.8		5.80
5		1	24				8.20
6		1	24				8.20
7		1	18				3.40
8		1	27				10.60
9		1	21				5.80
10		1	21				5.80
11		1	26				9.80
12		1	20				5.00
13		10×2			2×1		10×1

=Mmult (B3:C12, E3:E4)

=Mmult (X の範囲, β の範囲)

- 1) $X\beta$ の計算結果となる 10×1 の矩形を選択し太い外枠で囲み、上部に「 $X\beta$ 」を入力する。
- 2) 10×1 の矩形を選択し、数式バーを選択し、 $X\beta$ の最初の行に、行列の積の関数「=Mmult(X の範囲を選択, β の範囲を選択)」のように入力する。なお、「 X の範囲を選択」は、マウスで 10×2 の行列 X の範囲の選択を意味する。
- 3) $X\beta$ の最初の行で計算式の入力が終わったら「コントロールキー」と「シフトキー」を同時に押しながら「エンター」すると行列の積和の計算が行なわれる。

デザイン行列 X の転置と積和

行列 A と行列 B の掛け算は、左側の行列 A の「行」のセルと右側の行列 B の「列」のセル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”方向の“列・行”ではなく、あくまで“行・列”の順番である。

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

(10×2) のデザイン行列 X と (10×2) X の積 XX の内側のサイズは、(2 列 vs. 10 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 X について行と列を入れ替え、(2×10) の X^T (転置行列) とする。行列の転置は、 X の範囲をコピーし、転置したい先頭位置に対し、「形式を選択し貼り付け」、「行/列の入れ替え」で作成する。

積 $X^T X$ は、(2×10 vs. 10×2) 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、 X^T (2×10) と X (10×2) のように隣り合う行列の内側が、10 列と 10 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 $X^T X$ の結果は、外側の (2×2) の大きさとなる。

行と列の入れ替えは、転置 (Transpose) といい、 X^T のように表記する。なお、 X' あるいは 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

X_0^T 行と X_0 列の 積和=10 を ($X^T X$) の 1 行 1 列目へ
 X_1^T 行と X_0 列の 積和=220 を ($X^T X$) の 2 行 1 列目へ
 X_0^T 行と X_1 列の 積和=220 を ($X^T X$) の 1 行 2 列目へ
 X_1^T 行と X_1 列の 積和=4928 を ($X^T X$) の 2 行 2 列目へ

(2×10) の \mathbf{X}^T と (10×2) の \mathbf{X} の行列の積の計算結果として、(2×2) の $(\mathbf{X}^T \mathbf{X})$ 行列が生成される。なお、 $\mathbf{X}^T \mathbf{X}$ ではなく $\mathbf{X} \mathbf{X}^T$ とすると、(10×2) と (2×10) の積となり、(10×10) 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う理解が深まり計算ミスも少なくなる。

[illegible]

シグマを用いた積和の計算

行列関数を用いた計算方法に対し、シグマを用いた積和の計算方法を対比する。Excel の SumProduct() 関数は、複数の配列のセルごとの積和の計算をする便利な関数である。次に示すように行列 \mathbf{X} の 1 列目を \mathbf{X}_0 とし、2 列目を \mathbf{X}_1 とした場合の積和 220 を 2×2 の積和行列の 1 行 2 列に代入する。行列 \mathbf{X} のセルごとの変数を $x_{i,j}$, $i=1,2,\dots,10$, $j=0,1$ としたときに

$$\sum_{i=1}^{10} x_{i,0} x_{i,1} = \text{SumProduct}(\mathbf{X}_0 \text{の範囲}, \mathbf{X}_1 \text{の範囲}) = 220 \quad (5.4)$$

として計算されている．同じ列同士の場合は， $\text{SumSq}(\mathbf{X}_0)$ の範囲) を使うこともできる．実際にデザイン行列 \mathbf{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ので，まったく薦められない．

X_0	X_1		積和行列			
1	18	=	10	220		
1	21		220	4928		
1	24					
1	24		=SumSq (X_0 の範囲) = 10			
1	18		=SumProduct (X_0 の範囲, X_1 の範囲) = 220			
1	27		=SumProduct (X_1 の範囲, X_0 の範囲) = 220			
1	21		=SumSq (X_1 の範囲) = 4928			
1	21					
1	26					
1	20					
10×2						

デザイン行列 X の積和

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 $X^T X$ は、

$$X^T X = \text{Mmult}(\text{Transpose}(X \text{ の範囲}), X \text{ の範囲}) \quad (5.5)$$

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

転置したデザイン行列 X^T と X の積をシグマ記号 Σ で表せば、

行ベクトル X_0^T と列ベクトル X_0 の積和は n ,

行ベクトル X_0^T と列ベクトル X_1 の積和は $\Sigma_i X_i$

行ベクトル X_1^T と列ベクトル X_0 の積和は $\Sigma_i X_i$

行ベクトル X_1^T と列ベクトル X_1 の積和は $\Sigma_i X_i^2$

となり、 2×2 の行列としてまとめて表すことができる。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。 X_0^T と X_0 の積和は、行列 X の行の数 n となり、 $(X^T X)$ の 1 行 1 列目となっている。

X^T										X		$X^T X$	
X_0^T	1	1	1	1	1	1	1	1	1	X_0	X_1	n	ΣX_i
X_1^T	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_1	ΣX_i	ΣX_i^2
											1	X_2	
											1	X_3	
											1	X_4	
											1	X_5	
											1	X_6	
											1	X_7	
											1	X_8	
											1	X_9	
											1	X_{10}	
2×10										10×2		2×2	

デザイン行列 X と反応ベクトル Y との積和

次に、反応変数の列ベクトル Y とデザイン行列 X の積和を示す。 $X^T Y$ は、転置されたデザイン行列 X と列ベクトル Y の積で、 X_0^T と Y についての積和は

$$1 \times 4 + 1 \times 7 + \cdots + 1 \times 8 = 77$$

となり、 X_1^T と Y についての積和は

$$18 \times 4 + 21 \times 7 + \cdots + 20 \times 8 = 1772$$

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 Excel の行列計算を体験することが、スキルアップの最初の一步である。

5.3. 偏差平方和による回帰パラメータの推定

回帰パラメータの推定

求めたい回帰式を,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i=1,2,\dots,n \quad (5.6)$$

としたときに, 誤差 ε_i の平方和 Q は,

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2 \quad (5.7)$$

となる. 通常回帰分析において誤差 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 乗法での計算が可能である. そのために, 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使われ続け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. 誤差 ε_i の平方和 Q をパラメータ β_0 および β_1 で偏微分すると, 式 (5.8) が得られる.

$$\left. \begin{aligned} \frac{\partial Q}{\partial \beta_0} &= -2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{\partial Q}{\partial \beta_1} &= -2 \sum_{i=1}^n [(Y_i - \beta_0 - \beta_1 X_i) X_i] \end{aligned} \right\} \quad (5.8)$$

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

$$\left. \begin{aligned} \varepsilon_i^2 &= (Y_i - \beta_0 - \beta_1 X_i)^2 \\ &= Y_i^2 + \beta_0^2 + \beta_1^2 X_i^2 - 2\beta_0 Y_i - 2\beta_1 X_i Y_i + 2\beta_0 \beta_1 X_i \end{aligned} \right\} \quad (5.9)$$

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,

$$\left. \begin{aligned} \frac{\partial \varepsilon_i^2}{\partial \beta_0} &= \frac{\partial (Y_i^2 + \beta_0^2 + \beta_1^2 X_i^2 - 2\beta_0 Y_i - 2\beta_1 X_i Y_i + 2\beta_0 \beta_1 X_i)}{\partial \beta_0} \\ &= \begin{matrix} 0 & +2\beta_0 & + & 0 & -2Y_i & - & 0 & + & 2\beta_1 X_i \end{matrix} \\ &= 2(\beta_0 - Y_i + \beta_1 X_i) \\ &= -2(Y_i - \beta_0 - \beta_1 X_i) \end{aligned} \right\} \quad (5.10)$$

のように要素に分解し, シグマで再統合すれば式 (5.8) の 1 行目となる. β_1 での偏微分も同様である.

式 (5.8) を 0 と置き, β_0 と β_1 の推定値としての $\hat{\beta}_0$ と $\hat{\beta}_1$ を求める.

$$\left. \begin{aligned} -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) &= 0 \\ -2 \sum_{i=1}^n [(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i] &= 0 \end{aligned} \right\} \quad (5.11)$$

正規方程式

式 (5.11) 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,

$$\left. \begin{aligned} \sum_{i=1}^n Y_i - n \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^n X_i &= 0 \\ \sum_{i=1}^n X_i Y_i - \hat{\beta}_0 \sum_{i=1}^n X_i - \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= 0 \end{aligned} \right\} \quad (5.12)$$

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

$$\left. \begin{aligned} (1) \quad n \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i &= \sum_{i=1}^n Y_i \\ (2) \quad \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i Y_i \end{aligned} \right\} \quad (5.13)$$

が得られる. この方程式は, 正規方程式と呼ばれている. 式 (5.13) の左辺は, $\mathbf{X}^T \mathbf{X}$ と $\hat{\boldsymbol{\beta}}$ の積に等しく, 右辺は $\mathbf{X}^T \mathbf{Y}$ と等しいので, 行列で表わすと

$\mathbf{X}^T \mathbf{X}$			$\hat{\boldsymbol{\beta}}$		$\mathbf{X}^T \mathbf{Y}$
n	ΣX_i		$\hat{\beta}_0$	=	ΣY_i
ΣX_i	ΣX_i^2		$\hat{\beta}_1$		$\Sigma (X_i Y_i)$
2×2			2×1		2×1

となり, 行列表記での正規方程式

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y} \quad (5.14)$$

が得られる.

正規方程式の解によるパラメータの推定

正規方程式 (5.13) 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.

$$\left. \begin{aligned} (1) \quad \hat{\beta}_0 \Sigma_i X_i + \hat{\beta}_1 \frac{(\Sigma_i X_i)^2}{n} &= \frac{(\Sigma_i X_i)(\Sigma_i Y_i)}{n} \\ (2) \quad \hat{\beta}_0 \Sigma X_i + \hat{\beta}_1 \Sigma X_i^2 &= \Sigma (X_i Y_i) \end{aligned} \right\} \quad (5.15)$$

式 (5.15) の (2) 式から (1) 式を引いて, $\hat{\beta}_1$ について解くと

$$\left. \begin{aligned} \hat{\beta}_1 &= \frac{\Sigma(X_i Y_i) - \frac{(\Sigma X_i)(\Sigma Y_i)}{n}}{\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}} \\ &= \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \\ &= \frac{S_{XY}}{S_{XX}} \end{aligned} \right\} \quad (5.16)$$

のように, 多くの統計の教科書で示されている結果が得られる. ここで, S_{XX} は, \mathbf{X} の平均 \bar{X} からの偏差の平方和であり, S_{XY} は, \mathbf{X} の平均 \bar{X} からの偏差と \mathbf{Y} の平均 \bar{Y} からの偏差の積和である. この式をさらに, $\bar{X} = (\Sigma X_i) / n$, $\bar{Y} = (\Sigma Y_i) / n$ などの関係を用い, 式 (5.16) の 2 行目の分子 $\Sigma(X_i - \bar{X})(Y_i - \bar{Y})$ について

$$\left. \begin{aligned} \Sigma(X_i - \bar{X})(Y_i - \bar{Y}) &= \Sigma(X_i Y_i) - \bar{X} \Sigma Y_i - \bar{Y} \Sigma X_i + n \bar{X} \bar{Y} \\ &= \Sigma(X_i Y_i) - n \bar{X} \bar{Y} - n \bar{X} \bar{Y} + n \bar{X} \bar{Y} \\ &= \Sigma(X_i Y_i) - n \bar{X} \bar{Y} \\ &= \Sigma(X_i Y_i) - \frac{(\Sigma X_i)(\Sigma Y_i)}{n} \end{aligned} \right\} \quad (5.17)$$

と, 展開し, 整理すると式 (5.16) の 1 行目の分子となることを利用している. 式 (5.16) の 2 行目の分母 $\Sigma(X_i - \bar{X})^2$ については, 式 (5.17) の Y を X に置き換えることにより式 (5.16) の 1 行目の分母となる.

$$\left. \begin{aligned} \Sigma(X_i - \bar{X})^2 &= \Sigma X_i^2 - 2 \bar{X} \Sigma X_i + n \bar{X}^2 \\ &= \Sigma X_i^2 - 2n \bar{X}^2 + n \bar{X}^2 \\ &= \Sigma X_i^2 - n \left(\frac{\Sigma X_i}{n} \right)^2 \\ &= \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n} \end{aligned} \right\} \quad (5.18)$$

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ろであり, 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, 多くの統計の教科書に引き継がれている. ただし, 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 (5.13) の 1 行目は,

$$n \hat{\beta}_0 + \hat{\beta}_1 \Sigma X_i = \Sigma Y_i \quad (5.19)$$

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.

$$\left. \begin{aligned} \hat{\beta}_0 &= \frac{\sum Y_i - \hat{\beta}_1 \sum X_i}{n} \\ &= \bar{Y} - \hat{\beta}_1 \bar{X} \end{aligned} \right\} \quad (5.20)$$

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、

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \quad (5.21)$$

と簡潔な式となる. パラメータの導出だけであれば、この簡便な式 (5.21) 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 5.1 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.

偏差平方和を用いたパラメータの推定の実際

表 5.1 に示す Excel による計算シートの結果を用いて、回帰直線のパラメータは、

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{78.00}{88.00} = 0.8864$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 7.70 - 0.8864 \times 22.00 = -11.8000$$

として求められる.

表 5.1 偏差平方和を用いた回帰パラメータの 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	4	18	-3.70	-4.00	16.00	14.80	$\hat{\beta}_1 =$	0.8864
2	7	21	-0.70	-1.00	1.00	0.70	$\hat{\beta}_0 =$	-11.8000
3	10	24	2.30	2.00	4.00	4.60		
4	10	24	2.30	2.00	4.00	4.60		
5	2	18	-5.70	-4.00	16.00	22.80		
6	11	27	3.30	5.00	25.00	16.50		
7	9	21	1.30	-1.00	1.00	-1.30		
8	5	21	-2.70	-1.00	1.00	2.70		
9	11	26	3.30	4.00	16.00	13.20		
10	8	20	0.30	-2.00	4.00	-0.60		
	7.70	22.00	0.00	0.00	88.00	78.00		
	平均	平均	合計	合計	平方和	平方和		
	\bar{Y}	\bar{X}	$Y_i - \bar{Y}$	$X_i - \bar{X}$	S_{XX}	S_{XY}		

Excel シート上で数値を入力すると表示形式が「標準」となり、右側の罫線に貼り付いてしまい、見栄えが悪い。「セル書式設定」で表示形式を「数値」に変更し、小数点以下の桁数を調整する見栄えが向上する.

計算シートを用いなくとも、Excel の関数を用いて平方和の計算を

$$\bar{X} = \text{Average}(X\text{の範囲}) = 7.70$$

$$\bar{Y} = \text{Average}(Y\text{の範囲}) = 22.00$$

$$S_{XX} = \text{DevSq}(X\text{の範囲}) = 88.00$$

$$S_{XY} = \text{SumProduct}(X\text{の範囲}-\bar{X}, Y\text{の範囲}-\bar{Y}) = 78.00$$

のように直接推定することもできる。さらに、Excel には、 $\hat{\beta}_1$ および $\hat{\beta}_0$ を推定する便利な関数

$$\hat{\beta}_1 = \text{Slope}(Y\text{の範囲}, X\text{の範囲}) = 0.8864$$

$$\hat{\beta}_0 = \text{Intercept}(Y\text{の範囲}, X\text{の範囲}) = -11.8000$$

もある。

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えなくなってしまう。単回帰分析に引き続き、2 次式あるいは 3 次式をあてはめたいと思っても、パラメータを推定する式を見つけることができるのだろうか。見つけることができなければ、自ら導出することができるのだろうか。伝統的な回帰分析の解法は、更に回帰分析を拡張し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。第 2.4 節で示したように、デザイン行列 X を用いた回帰分析を適用すれば、同一手順ですべて解決することができる。

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步である。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 (5.15) で示した正規方程式を行列表記に対応付け

$$(X^T X) \hat{\beta} = X^T Y \quad \longleftrightarrow \quad \begin{array}{|c|c|} \hline X^T X & \hat{\beta} \\ \hline \hline n & \Sigma X_i \\ \hline \Sigma X_i & \Sigma X_i^2 \\ \hline 2 \times 2 & 2 \times 1 \\ \hline \end{array} = \begin{array}{|c|c|} \hline X^T Y & \\ \hline \hline \Sigma Y_i & \\ \hline \Sigma X_i Y_i & \\ \hline 2 \times 1 & \\ \hline \end{array}$$

学習することが望ましい。ただし、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・スミス著、中村慶一訳 (1968)、「応用回帰分析」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、Draper, N.R., Smith, H. (1989), Applied Regression Analysis 3rd.ed. もネット書店で手軽に入手できる。

そこで、「偏差平方和を用いた回帰分析」と「デザイン行列を用いた回帰分析」について、相互の関連を丁寧に示すことにする。そして、読者が「偏差平方和による回帰分析」から、「デザイン行列 X を用いた回帰分析」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。

5.4. デザイン行列 vs. 偏差平方和を用いた回帰分析

行列計算によるパラメータ推定

前節では、偏差平方和を用いた回帰分析のパラメータの推定方法を示し、式 (5.13) のシグマ表記の正規方程式は、式 (5.14) に示した様に行列を用いて表すと、

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

となる。推定値 $\hat{\beta}$ を得るために、行列計算では両辺を $(X^T X)$ で割ることができない。そこで、逆行列の定義により $(X^T X)^{-1}(X^T X) = I$ が単位行列となることを活用し、逆行列 $(X^T X)^{-1}$ を式 (5.14) 両辺に掛けると

$$\left. \begin{aligned} (X^T X)^{-1}(X^T X) \hat{\beta} &= (X^T X)^{-1} X^T Y \\ I \hat{\beta} &= (X^T X)^{-1} X^T Y \\ \hat{\beta} &= (X^T X)^{-1} X^T Y \end{aligned} \right\} \quad (5.22)$$

のように $\hat{\beta}$ を推定することができる。2×2 の逆行列 $(X^T X)^{-1}$ ならば、 $(X^T X)$ 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。

$X^T X$			$(X^T X)^{-1}$		
n	ΣX_i	$^{-1} =$	$\Sigma X_i^2 / D$	$-\Sigma X_i / D$	
ΣX_i	ΣX_i^2		$-\Sigma X_i / D$	n / D	
			$D = (n \Sigma X_i^2) - (\Sigma X_i)^2$		
$X^T X$			$(X^T X)^{-1}$		$(X^T X)^{-1}$
10	220	$^{-1} =$	$4928 / D$	$-220 / D$	5.6000
220	4928		$-220 / D$	$10 / D$	-0.2500
			$D = 49280 - 48400 = 880$		0.0114
			$= \text{Mdetarm}(X^T X \text{ の範囲})$		

実際に逆行列を前掛けすると

$(X^T X)^{-1}$			$X^T X$		$(X^T X)^{-1}(X^T X)$
5.6000	-0.2500	$=$	10	220	1.0000
-0.2500	0.0114		220	4928	0.0000
					0.0000
					1.0000

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

$(X^T X)^{-1}(X^T X)$			$\hat{\beta}$		$\hat{\beta}$
1.0000	0.0000	$=$	β_0^{\wedge}		β_0^{\wedge}
0.0000	1.0000		β_1^{\wedge}		β_1^{\wedge}

と元の $\hat{\beta}$ と同じになる。

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 $\hat{\beta}$ を求めるために必要な $X^T Y$ は、第 5.2 節の「デザイン行列 X と反応ベクトル 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、次のような手順で推定値 $\hat{\beta}$ を簡単に求めることができる。

$\hat{\beta} = (X^T X)^{-1} X^T Y =$	$(X^T X)^{-1}$	$X^T Y$	$\hat{\beta}$
	5.6000 -0.2500	77 1772	-11.8000 0.8864
	=Minverse ($X^T X$ の範囲)		
		=Mmult (Transpose (X の範囲), Y の範囲)	
			=Mmult (($X^T X$) ⁻¹ の範囲, $X^T Y$ の範囲)

デザイン行列 X での推定式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 $\hat{\beta}$ が得られるのであるが、正規方程式 (5.22) から、導出された $\hat{\beta}_0$ と $\hat{\beta}_1$ の計算式は、式 (5.16) から式 (5.21) に示した式とは異なるので、 $(\Sigma X_i) = n\bar{X}$, $(\Sigma Y_i) = n\bar{Y}$ 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。

$$\begin{aligned}
 \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\Sigma X_i^2}{D} & \frac{-\Sigma X_i}{D} \\ \frac{-\Sigma X_i}{D} & \frac{n}{D} \end{bmatrix} \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix} \\
 &= \begin{bmatrix} \frac{(\Sigma X_i^2)(\Sigma Y_i) - (\Sigma X_i)(\Sigma X_i Y_i)}{n\Sigma X_i^2 - (\Sigma X_i)^2} \\ \frac{-\Sigma X_i \Sigma Y_i + n\Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} \end{bmatrix} \\
 &= \begin{bmatrix} \frac{\bar{Y}(\Sigma X_i^2) - \bar{X}(\Sigma X_i Y_i)}{\Sigma X_i^2 - n\bar{X}^2} \\ \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2} \end{bmatrix}
 \end{aligned} \tag{5.23}$$

ここで、式 (5.23) の最後の行列の 2 行目は、式 (5.21) で導出された $\hat{\beta}_1$ の推定値に一致する。

$$\hat{\beta}_1 = \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2} = \frac{S_{XY}}{S_{XX}} \tag{5.24}$$

さて、式 (5.21) で導出された $\hat{\beta}_0$ の推定式とは、式 (5.23) の最後の行列の 1 行目は明らかに異なる。そこで、式 (5.21) からスタートし、式の変形を行なうと

$$\begin{aligned}
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X} \\
&= \frac{\bar{Y}S_{XX} - \bar{X}S_{XY}}{S_{XX}} \\
&= \frac{\left(\bar{Y}(\Sigma X_i^2) - \bar{Y} \frac{(\Sigma X_i)^2}{n} \right) - \left(\bar{X}(\Sigma X_i Y_i) - \bar{X} \frac{(\Sigma X_i)(\Sigma Y_i)}{n} \right)}{S_{XX}} \\
&= \frac{(\bar{Y}(\Sigma X_i^2) - n\bar{Y}\bar{X}^2) - (\bar{X}(\Sigma X_i Y_i) - n\bar{Y}\bar{X}^2)}{S_{XX}} \\
&= \frac{\bar{Y}(\Sigma X_i^2) - \bar{X}(\Sigma X_i Y_i)}{\Sigma X_i^2 - n\bar{X}^2}
\end{aligned} \quad (5.25)$$

となり、式 (5.23) 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ことは、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、偏差平方和を用いた回帰分析を「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。

偏差平方和を用いたパラメータの分散の推定

パラメータ $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$ は、式 (5.16) の正規方程式の解を用いて、

$$\begin{aligned}
Var(\beta_1) &= Var \left[\frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \right] \\
&= Var \left[\frac{\Sigma(X_i - \bar{X})Y_i}{\Sigma(X_i - \bar{X})^2} \right] \\
&= \frac{Var(Y_i)}{\Sigma(X_i - \bar{X})^2} \\
&= \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2}
\end{aligned} \quad (5.26)$$

となる。分子の変形は、

$$\begin{aligned}
\Sigma(X_i - \bar{X})(Y_i - \bar{Y}) &= \Sigma(X_i Y_i - X_i \bar{Y} - \bar{X} Y_i + \bar{X} \bar{Y}) \\
&= \Sigma(X_i Y_i - \bar{X} Y_i) - \Sigma X_i \bar{Y} + \Sigma \bar{X} \bar{Y} \\
&= \Sigma(X_i Y_i - \bar{X} Y_i) - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} \\
&= \Sigma(X_i - \bar{X})Y_i
\end{aligned} \quad (5.27)$$

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,

$$\left. \begin{aligned} Var(\hat{\beta}_0) &= Var(\bar{Y} - \hat{\beta}_1 \bar{X}) \\ &= Var(\bar{Y}) + \bar{X}^2 Var(\hat{\beta}_1) \\ &= \frac{\sigma^2}{n} + \frac{\bar{X}^2 \sigma^2}{\Sigma(X_i - \bar{X})^2} \\ &= \frac{\sigma^2 \Sigma X_i^2}{n \Sigma(X_i - \bar{X})^2} \end{aligned} \right\} \quad (5.28)$$

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

$$\left. \begin{aligned} Cov(\hat{\beta}_0, \hat{\beta}_1) &= Cov(\bar{Y} - \hat{\beta}_1 \bar{X}, \hat{\beta}_1) \\ &= -\bar{X} Var(\hat{\beta}_1) \\ &= \frac{-\bar{X} \sigma^2}{\Sigma(X_i - \bar{X})^2} \end{aligned} \right\} \quad (5.29)$$

となる.

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

$$\left. \begin{aligned} \Sigma(\hat{\beta}) &= \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\Sigma X_i^2}{n \Sigma(X_i - \bar{X})^2} & \frac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\Sigma(X_i - \bar{X})^2} & \frac{1}{\Sigma(X_i - \bar{X})^2} \end{bmatrix} \sigma^2 \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \sigma^2 \end{aligned} \right\} \quad (5.30)$$

となる. したがって, パラメータの分散共分散行列は, デザイン行列 \mathbf{X}^T と \mathbf{X} の積の逆行列 $(\mathbf{X}^T \mathbf{X})^{-1}$ 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, 残差の推定値 $\hat{\varepsilon}_i$ の平方和を自由度 $(n-2)$ で割った誤差分散の推定値 $\hat{\sigma}^2$ を用いて計算する.

$$\left. \begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2} \\ &= \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{n-2} \end{aligned} \right\} \quad (5.31)$$

これまで, $(\mathbf{X}^T \mathbf{X})^{-1}$ については, 次の式を示してきたのであるが, 式 (5.30) と異なるので, 式の変形を行う.

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

$(\mathbf{X}^T \mathbf{X})$ の行列式 D は,

$$\left. \begin{aligned} D &= n \Sigma X_i^2 - (\Sigma X_i)^2 \\ &= n(\Sigma X_i^2 - n \bar{X}^2) \\ &= n \Sigma (X_i - \bar{X})^2 \end{aligned} \right\} \quad (5.32)$$

と変形できるので,

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{bmatrix} \frac{\Sigma X_i^2}{n \Sigma (X_i - \bar{X})^2} & \frac{-\Sigma X_i}{n \Sigma (X_i - \bar{X})^2} \\ \frac{-\Sigma X_i}{n \Sigma (X_i - \bar{X})^2} & \frac{n}{n \Sigma (X_i - \bar{X})^2} \end{bmatrix} = \begin{bmatrix} \frac{\Sigma X_i^2}{n \Sigma (X_i - \bar{X})^2} & \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} \\ \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} & \frac{1}{\Sigma (X_i - \bar{X})^2} \end{bmatrix} \quad (5.33)$$

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 \mathbf{X} 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 \mathbf{X} を用いた計算手順で簡単に得れるパラメータの共分散行列 $\Sigma(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$ を基本とすることは, 1 変数の回帰分析にしか適用できない偏差平方和を用いた計算方法によって作り出されているガラスの天井を超えるための必須の知識である.

ここに示したパラメータの共分散行列 $\Sigma(\hat{\beta}) = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$ は, 1 変数のデザイン行列 \mathbf{X} のみならず, 2 次式の回帰 $y = \beta_0 + \beta_1 x + \beta_2 x^2$, 2 変数の重回帰 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, ダミー変数 a を用いた共分散分析 $y = \beta_0 + \beta_1 a + \beta_2 x$, 交互作用を含む 2 因子実験など, 一般線形モデルのパラメータの共分散行列においても, デザイン行列 \mathbf{X} のサイズが異なるだけで, 計算手順はすべて同じである.

シグマを用いた回帰分析の功罪

得られたデータに対し 1 次式 (回帰直線) をあてはめてみたが, 直線とはいいがたいので, 2 次式 (回帰曲線) をあてはめたい. パラメータの推定は, 回帰分析の説明変数 x に加えて x^2 を加えることにより得られ, Excel の散布図上に 2 次曲線の推定値を重ね書きすることができた. さらに, 2 次曲線の 95% 信頼区間を重ね書きしたと思い, 計算式がどこにあるか手を尽くして探したが見つからない. 1 次式 (回帰直線) の計算式を参考にし, 拡張しようと思ってもまったく分からない. 統計ソフトの多項式回帰回帰を使えば, 95% 信頼区間を図示できることは分かったが, 計算式が分からない. このように, シグマを用いた回帰分析の方法は, 拡張性が全くないので, デザイン行列 \mathbf{X} を活用した第 1.4 節, 「1 因子実験の量的変数に対する多項式回帰」に詳細が示されている.

5.5. Excel の行列計算による回帰分析の実際

Excel の行列計算による回帰パラメータの推定

表 5.2 に示すのは、これまで示してきた計算方法を総合し、デザイン行列 X を用いた回帰分析の計算事例である。全ての計算過程を 1 枚の Excel シートで示したために、やや見づらくなっているのですが、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」を活用した表 5.6 と比較してもらいたい。

表 5.2 Excel の行列計算を用いた回帰分析

i	X		Y	Y^{\wedge}	$\varepsilon^{\wedge} = Y - Y^{\wedge}$						
1	1	18	4	4.1545	-0.1545						
2	1	21	7	6.8136	0.1864						
3	1	24	10	9.4727	0.5273						
4	1	24	10	9.4727	0.5273						
5	1	18	2	4.1545	-2.1545						
6	1	27	11	12.1318	-1.1318						
7	1	21	9	6.8136	2.1864						
8	1	21	5	6.8136	-1.8136	項	推定値	分散	SE	t 値	p 値
9	1	26	11	11.2455	-0.2455	$\beta_0^{\wedge} =$	-11.8000	13.2745	3.6434	-3.2387	0.0119
10	1	20	8	5.9273	2.0727	$\beta_1^{\wedge} =$	0.8864	0.0269	0.1641	5.4005	0.0006
						$(X^T X)^{-1} X^T Y$		sqrt(分散)	T.dist.2T (t , 10-2)		
	10	220	5.6000	-0.2500	77	$\varepsilon^{\wedge T} \varepsilon^{\wedge} =$	18.9636	共分散	13.2745	-0.5926	
	220	4928	-0.2500	0.0114	1772	$\sigma^{\wedge 2} =$	2.3705	行列	-0.5926	0.0269	
	$(X^T X)$		$(X^T X)^{-1}$		$(X^T Y)$	$\Sigma(\beta^{\wedge}) = (X^T X)^{-1} \sigma^{\wedge 2}$					

デザイン行列 X に対し、 $X^T X$ 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

$$X^T X = \text{Mmult}(\text{Transpose}(X \text{ の範囲}), X \text{ の範囲}) = \begin{bmatrix} 10 & 220 \\ 220 & 4928 \end{bmatrix}$$

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

$$(X^T X)^{-1} = \text{Minverse}(X^T X \text{ の範囲}) = \begin{bmatrix} 5.6000 & -0.2500 \\ -0.2500 & 0.0114 \end{bmatrix}$$

となり、デザイン行列 X の転置行列 X^T と列ベクトル Y との積 $X^T Y$ が

$$X^T Y = \text{Mmult}(\text{Transpose}(X \text{ の範囲}), Y \text{ の範囲}) = \begin{bmatrix} 77 \\ 1772 \end{bmatrix}$$

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

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \text{Mmult}((X^T X)^{-1} \text{ の範囲}, X^T Y \text{ の範囲}) = \begin{bmatrix} -11.8000 \\ 0.8864 \end{bmatrix}$$

$\hat{\beta}_0 = -11.8000$, $\hat{\beta}_1 = 0.8864$ として計算されている. 誤差分散は, 回帰の推定値 \hat{Y} を

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

で求め, 残差ベクトル $\hat{\epsilon}$ が列ベクトル Y と推定ベクトル \hat{Y} の差

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

として求めている. 残差平方和

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

$$\text{または, } S_e = \text{SumSq}(\hat{\epsilon}\text{の範囲})$$

を計算し, データ数 $n = 10$ 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の推定値 $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}^T \hat{\epsilon}}{n-2} = \frac{18.9636}{10-2} = 2.3705$$

として計算されている.

回帰パラメータの分散および共分散の推定

デザイン行列 X 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 (5.30) からパラメータの共分散行列 $\Sigma(\hat{\beta})$ は,

$$\begin{aligned} \Sigma(\hat{\beta}) &= \begin{bmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \end{bmatrix} \\ &= (X^T X)^{-1} \hat{\sigma}^2 \end{aligned} \quad (5.34)$$

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

5.6000	-0.2500	$\hat{\epsilon}^T \hat{\epsilon} =$	18.9636
-0.2500	0.0114	$\hat{\sigma}^2 =$	2.3705
$(X^T X)^{-1}$			

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

$$\Sigma(\hat{\beta}) = \begin{bmatrix} 5.6000 & -0.2500 \\ -0.2500 & 0.0114 \end{bmatrix} \begin{bmatrix} 2.3705 \\ \sigma^2 \end{bmatrix} = \begin{bmatrix} 13.2745 & -0.5926 \\ -0.5926 & 0.0269 \end{bmatrix} = \Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$$

と計算される. 分散 $\text{Var}(\hat{\beta}_0) = 13.2745$, $\text{Var}(\hat{\beta}_1) = 0.0269$ は, $(X^T X)^{-1} \hat{\sigma}^2$ 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
$\hat{\beta}_0 =$	-11.8000	13.2745
$\hat{\beta}_1 =$	0.8864	0.0269

更に標準誤差 SE を分散の平方根 $\text{sqrt}()$ 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 $\text{T.dist.2T}()$ 関数

$$p \text{ 値} = T.\text{dist.}2T(t\text{値}, (10-2))$$

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
$\beta_0^{\wedge} =$	-11.8000	13.2745	3.6434	-3.2387	0.0119
$\beta_1^{\wedge} =$	0.8864	0.0269	0.1641	5.4005	0.0006
$(X^T X)^{-1} X^T Y$		sqrt(分散)		T.dist.2T(t , 10-2)	

表 5.1 で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、実用的でもある。

Excel の関数を用いた分散分析表の作成

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。表 5.2 では、反応 Y_i に対して、推定された回帰パラメータを用いた推定値 $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ との残差 $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ を計算し、その平方和を求めた。この残差平方和 S_e は、推定された回帰直線からのズレの大きさを表している。

$$S_e = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (5.35)$$

表 5.3 に示すように、平均 \bar{Y} からの偏差を $(Y_i - \bar{Y})$ としたときの偏差平方和を「全体の偏差平方和」の意味で S_T とする。

$$S_T = \sum_{i=1}^n (Y_i - \bar{Y})^2 \quad (5.36)$$

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 \bar{Y} からの偏差 $R_i = \hat{Y}_i - \bar{Y}$ の平方和が回帰の平方和 S_R となる。

$$S_R = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \quad (5.37)$$

表 5.3 分散分析表作成のための平方和の計算

i	X		Y_i	\bar{Y}	$Y_i - \bar{Y}$	Y_i^{\wedge}	$Y_i^{\wedge} - \bar{Y}$	$Y_i - Y_i^{\wedge}$		
1	1	18	4	7.70	-3.70	4.1545	-3.5455	-0.1545		
2	1	21	7	7.70	-0.70	6.8136	-0.8864	0.1864	$\beta_0^{\wedge} =$	-11.8000
3	1	24	10	7.70	2.30	9.4727	1.7727	0.5273	$\beta_1^{\wedge} =$	0.8864
4	1	24	10	7.70	2.30	9.4727	1.7727	0.5273		
5	1	18	2	7.70	-5.70	4.1545	-3.5455	-2.1545		
6	1	27	11	7.70	3.30	12.1318	4.4318	-1.1318		
7	1	21	9	7.70	1.30	6.8136	-0.8864	2.1864		
8	1	21	5	7.70	-2.70	6.8136	-0.8864	-1.8136		
9	1	26	11	7.70	3.30	11.2455	3.5455	-0.2455		
10	1	20	8	7.70	0.30	5.9273	-1.7727	2.0727		
				7.70	88.10		69.1364	18.9636		
				\bar{Y}	S_T		S_R	S_e		
自由度			10	1	10-1=9	2	2-1=1	10-2=8	2	

この様に、元の反応 Y_i についての平方和には、

$$\left. \begin{aligned} S_T &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \\ &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \\ &= S_e + S_R \end{aligned} \right\} \quad (5.38)$$

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 5.4 に示す。

回帰パラメータの推定には、デザイン行列 \mathbf{X} 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。

$$S_T = \sum_{i=1}^{10} (Y_i - \bar{Y})^2 = \text{SumSq}(\mathbf{Y} \text{ の範囲} - \bar{Y}) = 88.10, \quad df = 10 - 1 = 9$$

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

$$S_R = \sum_{i=1}^{10} (\hat{Y}_i - \bar{Y})^2 = \text{SumSq}(\hat{\mathbf{Y}} \text{ の範囲} - \bar{Y}) = 69.1364, \quad df = 2 - 1 = 1$$

$$S_e = \sum_{i=1}^{10} (\hat{Y}_i - Y_i)^2 = \text{SumSq}(\hat{\mathbf{Y}} \text{ の範囲} - \mathbf{Y} \text{ の範囲}) = 18.9636, \quad df = 10 - 2 = 8$$

計算結果を、表 5.4 の分散分析表にまとめる。計算原理を習得した後は、表 5.6 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はこだわる必要はない。

表 5.4 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	69.1364	2-1=1	69.1364	29.1659	0.0006
残差	S_e	18.9636	10-2=8	2.3705		
全体	S_T	88.1000	10-1=9			

反応 \mathbf{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 \mathbf{Y} から求めた算術平均 \bar{Y} を用いているので、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

$$\left. \begin{aligned} df_T &= df_R + df_e \\ &= 1 + (n - 2) = n - 1 \end{aligned} \right\} \quad (5.39)$$

が成り立つ。分散分析表における自由度については、各種の便宜的な説明が行なわれているが、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 69.1364 を残差の平均平方 2.3705 で割って 29.1659 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 8 の F 分布の右側 (Right Tail) 確率で、 $F.\text{dist.RT}()$ 関数を使い、 $F.\text{dist.RT}(29.1659, 1, 8)$ により計算されている。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。Excel の 2007 年以前にも多くの分布関数が提供されていたのであるが、分布関数の引数の使い方などが不統一であり、計算ミスを起こしやすく、Excel に対する不信感の原因でもあった。それ以後に改善がはかられ、Excel の分布関数を使うときに「2007 年以前のバージョンと互換性があります。」とのメッセージが出た場合には、使用を避けることが望ましい。

回帰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 5.2 に示したようにパラメータの共分散行列 $\Sigma(\hat{\beta})$ の計算結果が含まれており、この行列 $\Sigma(\hat{\beta})$ 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 $\Sigma(\hat{\beta})$ 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の推定値 $\hat{\sigma}^2$ が得られるので、切片を含めた説明変数をデザイン行列 X とし、行列計算で $(X^T X)^{-1}$ を求め

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

により、パラメータの共分散行列 $\Sigma(\hat{\beta})$ 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。合成分散の一般的式により、

$$\left. \begin{aligned} \text{Var}(\hat{Y}_i) &= \text{Var}(\hat{\beta}_0 + \hat{\beta}_1 X_i) \\ &= \text{Var}(\hat{\beta}_0) + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) X_i + \text{Var}(\hat{\beta}_1) X_i^2 \end{aligned} \right\} \quad (5.40)$$

により、計算する。任意の行ベクトルを

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

とすれば、次の 2 次形式により、

$$\begin{aligned}
Var(\hat{y}) &= [(X^T X)^{-1} \hat{\sigma}^2] \\
&= [1 \ x] \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} \\
&= [1 \ x] \begin{bmatrix} Var(\hat{\beta}_0) + Cov(\hat{\beta}_0, \hat{\beta}_1)x \\ Cov(\hat{\beta}_0, \hat{\beta}_1) + Var(\hat{\beta}_1)x \end{bmatrix} \\
&= Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1)x + Var(\hat{\beta}_1)x^2
\end{aligned} \tag{5.41}$$

合成分散の一般式に一致する。

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 5.5 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 $\hat{y} = \mathbf{x}\hat{\boldsymbol{\beta}}$ 、分散 $Var(\hat{y})$ を

$$Var(\hat{y}) = \mathbf{x}[(X^T X)^{-1} \hat{\sigma}^2] \mathbf{x}^T = \mathbf{x}\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{x}^T \tag{5.42}$$

で計算する。

表 5.5 回帰直線の 95%信頼区間

切片	x	\hat{y}	$Var(\hat{y})$	$L_{95\%}$	$U_{95\%}$	個別 $L_{95\%}$	個別 $U_{95\%}$
1	14	0.61	1.96	-2.62	3.84	-4.19	5.41
1	16	2.38	1.21	-0.15	4.92	-1.98	6.74
1	18	4.15	0.67	2.27	6.04	0.13	8.17
1	20	5.93	0.34	4.57	7.28	2.13	9.73
1	22	7.70	0.24	6.58	8.82	3.98	11.42
1	24	9.47	0.34	8.12	10.83	5.67	13.27
1	26	11.25	0.67	9.36	13.13	7.23	15.27
1	28	13.02	1.21	10.48	15.55	8.66	17.38
1	30	14.79	1.96	11.56	18.02	9.99	19.59
1	32	16.56	2.93	12.62	20.51	11.25	21.87
T.inv.2T(0.05, 10-2)=				2.3060			

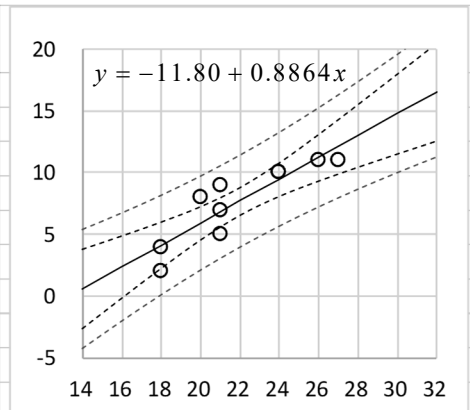


図 5.1 回帰直線の 95%信頼区間

表 5.5 の 1 行目の $\mathbf{x}_1 = [1 \ 14]$ に対する推定値 \hat{y}_1 は、

$$\begin{aligned}
\hat{y}_1 &= \mathbf{x}\hat{\boldsymbol{\beta}} \\
&= [1 \ 14] \begin{bmatrix} -11.8000 \\ 0.8864 \end{bmatrix} = 0.6091
\end{aligned}$$

となり、2 次形式により回帰直線の推定値 \hat{y}_1 の分散は、

$$\begin{aligned}
Var(\hat{y}_1) &= \mathbf{x}[(X^T X)^{-1} \hat{\sigma}^2] \mathbf{x}^T \\
&= [1 \ 14] \begin{bmatrix} 13.2745 & -0.5926 \\ -0.5926 & 0.0269 \end{bmatrix} \begin{bmatrix} 1 \\ 14 \end{bmatrix} = 1.9610
\end{aligned}$$

として求められ、

95%信頼区間は,

$$\begin{aligned} L95\% &= \hat{y}_1 - t_{0.05}(10-2)\sqrt{Var(\hat{y}_1)} \\ &= 0.6091 - 2.3060 \times \sqrt{1.9610} = -2.6201 \\ U95\% &= 0.6091 + 2.3060 \times \sqrt{1.9610} = 3.8383 \end{aligned}$$

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 σ^2 を加えた分散を使う.

$$\begin{aligned} \text{個別}L95\% &= \hat{y}_1 - t_{0.05}(10-2)\sqrt{Var(\hat{y}_1) + \sigma^2} \\ &= 0.6091 - 2.3060 \times \sqrt{1.9610 + 2.3705} = -4.1902 \\ \text{個別}U95\% &= 0.6091 + 2.3060 \times \sqrt{1.9610 + 2.3705} = 5.4084 \end{aligned}$$

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すべて計算される. 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

多くの成書で回帰直線の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 についての回帰式を

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 \quad (5.43)$$

とする. ところで $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ なので, $\hat{\beta}_0$ について解くと, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ となり, 式(5.43)に $\hat{\beta}_0$ を代入して

$$\left. \begin{aligned} \hat{y}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_0 \\ &= \bar{y} + \hat{\beta}_1 (x_0 - \bar{x}) \end{aligned} \right\} \quad (5.44)$$

が得られ, \hat{y}_0 の分散 $Var(\hat{y}_0)$ は, \bar{y} と $\hat{\beta}_1$ の共分散が0なので, 式(5.26)の $Var(\hat{\beta}_1)$ を用いて, 次式で与えられる. $Var(\bar{y})$ は, n 個データの平均なので, $Var(\bar{y}) = \sigma^2 / n$ となり,

$$\left. \begin{aligned} Var(\hat{y}_0) &= Var[\bar{y} + \hat{\beta}_1 (x_0 - \bar{x})] \\ &= Var(\bar{y}) + (x_0 - \bar{x})^2 Var(\hat{\beta}_1) \\ &= \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right] \sigma^2 \end{aligned} \right\} \quad (5.45)$$

分散が求められれば, 標準的な方法で95%信頼区間を求めることができる. また, 個別データに対する分散 $Ver(\hat{y}_{0, \text{個別}})$ は,

$$\left. \begin{aligned} Ver(\hat{y}_{0, \text{個別}}) &= Ver(\hat{y}_0) + \sigma^2 \\ &= \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right] \sigma^2 \end{aligned} \right\} \quad (5.46)$$

で与えられる.

このように、多くの成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、単回帰分析のみに対するものであり、2変数以上の問題に対して応用することができない。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」そのものである。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 $\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$ であり、第2.4節で例示したように、変数の数が増えても95%信頼区間の計算方法は2次形式の適用方法同じである。ただし、行列計算の式を使わない場合には、95%信頼区間の計算式は、4次式になり、計算式を示すこと自体が難儀でもあり、それを理解し応用しようとする読者にとっても優しくない。そのために、式を示さずに図による表示に留められるのが常である。

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行列 $\Sigma(\hat{\beta})$ 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5.6に示すように、Excel の「回帰分析」を使い、分散分析表から残差の分散（誤差分散の推定値） $\hat{\sigma}^2 = 2.3705$ を、係数の表から $\hat{\beta}_0 = -11.8000$ 、 $\hat{\beta}_1 = 0.8864$ を得る。表5.2で示したと同様にパラメータの共分散行列 $\Sigma(\hat{\beta})$ は、デザイン行列 X を用い

表 5.6 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	X		Y	分散分析表（定数に0を使用）off					
1	1	18	4		自由度	変動	分散	分散比	有意 F
2	1	21	7	回帰	1	69.1364	69.1364	29.1659	0.0006
3	1	24	10	残差	8	18.9636	2.3705		
4	1	24	10	合計	9	88.1000			
5	1	18	2						
6	1	27	11		係数	標準誤差	t	P -値	下限 95% 上限 95%
7	1	21	9	切片	-11.8000	3.6434	-3.2387	0.0119	-20.2018 -3.3982
8	1	21	5	x	0.8864	0.1641	5.4005	0.0006	0.5079 1.2648
9	1	26	11						
10	1	20	8						
	10.0	220.0		5.6000	-0.2500		13.2745	-0.5926	$t_{0.05}(8) =$ 2.3060
	220.0	4928.0		-0.2500	0.0114		-0.5926	0.0269	
	$X^T X$			$(X^T X)^{-1}$			$\Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2$		

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

$$= \begin{array}{|c|c|} \hline 10.0 & 220.0 \\ \hline 220.0 & 4928.0 \\ \hline \end{array}$$

$$X^T X$$

$$\begin{aligned}
 \Sigma(\hat{\beta}) &= (X^T X)^{-1} \hat{\sigma}^2 \\
 &= \text{Minverse}(X^T X \text{の範囲}) \times \hat{\sigma}^2 \\
 &= \begin{array}{|c|c|} \hline 5.6000 & -0.2500 \\ \hline -0.2500 & 0.0114 \\ \hline \end{array} \times \begin{array}{|c|} \hline 2.3705 \\ \hline \end{array} \\
 &\quad (X^T X)^{-1} \\
 &= \begin{array}{|c|c|} \hline 13.2595 & -0.5919 \\ \hline -0.5919 & 0.0269 \\ \hline \end{array} \\
 &\quad \Sigma(\hat{\beta}) = (X^T X)^{-1} \hat{\sigma}^2
 \end{aligned}$$

によって計算できる．これらを用いて，表 5.5 で示した回帰直線の 95%信頼区間の計算が可能となる．

平方和の分解に対する補足

式 (5.38) 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 (5.38) 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．

$$\left. \begin{aligned}
 S_T &= \sum_{i=1}^n (Y_i - \bar{Y})^2 \\
 &= \sum_{i=1}^n [(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})]^2 \\
 &= \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 + \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + 2 \sum_{i=1}^n [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})] \\
 &= S_e + S_R + 2 \sum_{i=1}^n [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})] \\
 &= S_e + S_R
 \end{aligned} \right\} \quad (5.47)$$

第3項は，次のように展開して0となる．

$$\left. \begin{aligned}
 \sum_{i=1}^n [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})] &= \sum_{i=1}^n \hat{\varepsilon}_i \hat{Y}_i - \bar{Y} \sum_{i=1}^n \hat{\varepsilon}_i \\
 &= \sum_{i=1}^n \hat{\varepsilon}_i (\hat{\beta}_0 + \hat{\beta}_1 X_i) \\
 &= \hat{\beta}_0 \sum_{i=1}^n \hat{\varepsilon}_i + \hat{\beta}_1 \sum_{i=1}^n \hat{\varepsilon}_i X_i \\
 &= 0
 \end{aligned} \right\} \quad (5.48)$$

これは，式 (5.11) 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．

$$\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_{i=1}^n (Y_i - \hat{Y}_i) = \sum_{i=1}^n \hat{\varepsilon}_i = 0 \quad (5.49)$$

$$\sum_{i=1}^n [(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i] = \sum_{i=1}^n [(Y_i - \hat{Y}_i) X_i] = \sum_{i=1}^n \hat{\varepsilon}_i X_i = 0 \quad (5.50)$$

5.6. 逆推定値に対する各種の 95%信頼区間の推定

逆推定とは何か

前節の図 5.1 を図 5.2 に再掲し，得られた回帰直線に対して反応が $y_0 = 6$ となる x_0 を逆推定するための矢印を重ね書きしている．逆推定値は，回帰直線の式

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

を \hat{x}_0 について解くと

$$\begin{aligned} \hat{x}_0 &= \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \\ &= \frac{6.0 - (-11.800)}{0.8864} = 20.0821 \end{aligned} \quad (5.51)$$

となり，容易に逆推定値 $\hat{x}_0 = 20.0821$ 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一般式が使えないからである．

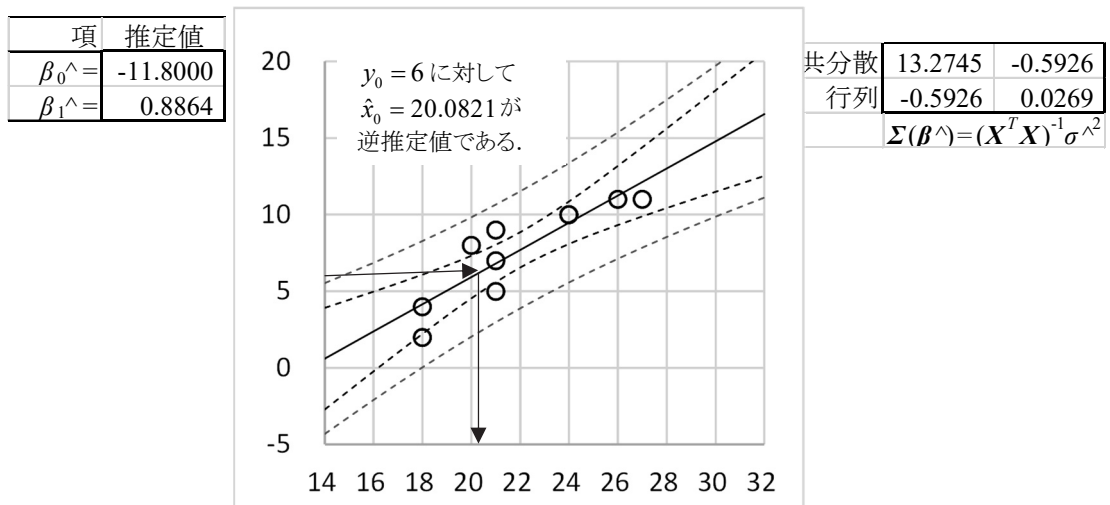


図 5.2 逆推定の例示

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 (2016)，「医薬品開発のための統計学，第 3 部 非線形モデル 改訂版，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―」が詳しい．また，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回目の公表資料で、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。逆推定値の分散の推定について示されていないが、アーミテイジ著、椿・椿訳（2001）、「医学研究のための統計的方法、第3.6節 分散に関するその他の公式」に「比の分散」、「積の分散」、「一般の関数の分散」についても参考になる。

デルタ法による逆推定値に対する近似 95%信頼区間

推定されたパラメータの比についての合成分散は、いわゆるデルタ法によって求めることができる。そのために、求めたい逆推定値 \hat{x}_0 の式 (5.51) に対して、パラメータ $\hat{\beta}_0$ と $\hat{\beta}_1$ で偏微分し、

$$d_0 = \frac{\partial \hat{x}_0}{\partial \hat{\beta}_0} = \frac{\partial (y_0 - \hat{\beta}_0) / \hat{\beta}_1}{\partial \hat{\beta}_0} = -\frac{1}{\hat{\beta}_1} \quad (5.52)$$

$$d_1 = \frac{\partial \hat{x}_0}{\partial \hat{\beta}_1} = \frac{\partial (y_0 - \hat{\beta}_0) / \hat{\beta}_1}{\partial \hat{\beta}_1} = \frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2} \quad (5.53)$$

を得る。これらの偏微分式を行ベクトル \mathbf{d}

$$\mathbf{d} = [d_0 \ d_1] \quad (5.54)$$

としたとき、 \hat{x}_0 の分散 $Var(\hat{x}_0)$ は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する \mathbf{d} の2次形式

$$\begin{aligned} Var(\hat{x}_0) &= \mathbf{d} \Sigma(\hat{\beta}) \mathbf{d}^T \\ &= [d_0 \ d_1] \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \end{bmatrix} \end{aligned} \quad (5.55)$$

で求めることができる。

表 5.7 逆推定の近似 95%信頼区間

y_0	x_0^{\wedge}	d_0	d_1	$Var(x_0^{\wedge})$	L95%	U95%	個別L95	個別U95
2	15.5692	-1.1282	-17.5653	1.7196	12.5453	18.5932	?	?
4	17.8256	-1.1282	-20.1110	0.8992	15.6390	20.0123		
6	20.0821	-1.1282	-22.6567	0.4278	18.5737	21.5904		
8	22.3385	-1.1282	-25.2024	0.3057	21.0636	23.6134		
10	24.5949	-1.1282	-27.7481	0.5326	22.9120	26.2778		
12	26.8513	-1.1282	-30.2938	1.1087	24.4232	29.2793		
T.inv.2T (0.05, 10-2)=					2.3060			

表 5.7 の1行目の $y_0 = 2$ に対する逆推定値 \hat{x}_0 は、表 5.6 または図 5.2 に示されている $\hat{\beta}_0$ と $\hat{\beta}_1$ の推定値を用いて

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{2 - (-11.80)}{0.8864} = 15.5692$$

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

$$d_0 = -\frac{1}{\hat{\beta}_1} = -\frac{1}{0.8864} = -1.1282$$

$$d_1 = \frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2} = \frac{-(2 - (-11.80))}{0.8864^2} = -17.5653$$

$$\mathbf{d} = [-1.1282 \quad -17.5653]$$

なので、分散 $Var(\hat{x}_0)$ は、

$$Var(\hat{x}_0) = \mathbf{d} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}) \mathbf{d}^T$$

$$= [-1.1282 \quad -17.5653] \begin{bmatrix} 13.2745 & -0.5926 \\ -0.5926 & 0.0269 \end{bmatrix} \begin{bmatrix} -1.1282 \\ -17.5653 \end{bmatrix}$$

$$= 1.7196$$

と推定されている。下側 95%点と上側 95%点は、

$$L95\% = \hat{x}_0 - t_{0.05}(10-2)\sqrt{Var(\hat{x}_0)}$$

$$= 15.5692 - 2.3060 \times \sqrt{2.3705} = 12.5453$$

$$U95\% = \hat{x}_0 + t_{0.05}(10-2)\sqrt{Var(\hat{x}_0)}$$

$$= 15.5692 + 2.3060 \times \sqrt{2.3705} = 18.5932$$

となる。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 5.2 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、

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{6.0 - (-11.800)}{0.8864} = 20.0821$$

と推定される。図 5.3 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、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。

交点の正確な 95%信頼区間の下限の推定値 \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

$$\hat{y}_{L95} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} \quad (5.56)$$

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。

$$y_0 = \hat{y}_{L95} + t_{0.05} \sqrt{Var(\hat{y}_{L95})} \quad (5.57)$$

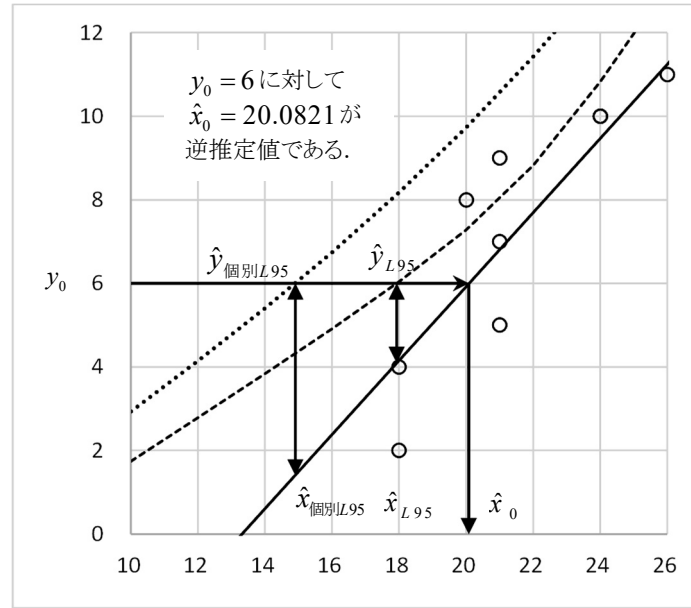


図 5.3 逆推定の 95%信頼区間の算出の例示

推定したいのは、正確な 95%信頼区間の下限の推定値 \hat{x}_{L95} なので、 \hat{y}_{L95} を $\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}$

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} + t_{0.05} \sqrt{\text{Var}(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} \quad (5.58)$$

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．

$$\left(\frac{y_0 - \hat{\beta}_0 - \hat{\beta}_1 \hat{x}_{L95}}{t_{0.05}} \right)^2 = \text{Var}(\hat{\beta}_0) + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \hat{x}_{L95} + \text{Var}(\hat{\beta}_1) \hat{x}_{L95}^2 \quad (5.59)$$

左辺を右辺に移して \hat{x}_{L95} について整理すると、

$$\underbrace{\left[\text{Var}(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_{0.05}^2} \right]}_a + \underbrace{\left[2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_{0.05}^2} \right]}_b \hat{x}_{L95} + \underbrace{\left[\text{Var}(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_{0.05}^2} \right]}_c \hat{x}_{L95}^2 = 0 \quad (5.60)$$

が得られる．この複雑な式は、幸い正確な 95%信頼区間の下限の推定値 \hat{x}_{L95} に関する 2 次式

$$a + b\hat{x}_{L95} + c\hat{x}_{L95}^2 = 0 \quad (5.61)$$

になるので、2 次式の解の公式

$$\hat{x}_{L95} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \quad (5.62)$$

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

個別データに対する逆推定値の正確な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、個別データの下側95%点の推定値 $\hat{x}_{\text{個別}L95}$ に対する回帰の推定値

$$\hat{y}_{\text{個別}L95} = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{個別}L95} \quad (5.63)$$

である。個別のデータの上側95%点 $\hat{y}_{\text{個別}L95}$ は、 $y_0 = 6$ に等しいので、次の等式が成り立つ。

$$y_0 = \hat{y}_{L95} + t_{0.05} \sqrt{\text{Var}(\hat{y}_{\text{個別}L95}) + \hat{\sigma}^2} \quad (5.64)$$

推定したいのは、 $\hat{x}_{\text{個別}L95}$ なので、

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{個別}L95} + t_{0.05} \sqrt{\text{Var}(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{\text{個別}L95}) + \hat{\sigma}^2} \quad (5.65)$$

と置き換える。式(5.64)を $\hat{x}_{\text{個別}L95}$ について解くと式(5.60)と同様な結果となるが、2次式のパラメータ a について

$$\text{個別}a' = a + \hat{\sigma}^2$$

となるだけで、他は同様である。

これらから、表5.8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、個別データの逆推定値の95%信頼区間を推定することができる。

表5.8 逆推定の正確な2種類の95%信頼区間の推定値

y_0	2次式のパラメータ			逆推定	95%信頼区間		個別	個別95% CL	
	a	b	c	\hat{x}_0	\hat{x}_{L95}	\hat{x}_{U95}	a'	個別L95	個別U95
2	-22.5382	3.4152	-0.1208	15.5692	10.4967	17.7739	-20.1678	8.4028	19.8678
4	-33.6710	4.0820	-0.1208	17.8256	14.3037	19.4860	-31.3005	11.7630	22.0267
6	-46.3081	4.7487	-0.1208	20.0821	17.9324	21.3764	-43.9377	14.9018	24.4070
8	-60.4497	5.4154	-0.1208	22.3385	21.0020	23.8258	-58.0792	17.7647	27.0632
10	-76.0957	6.0822	-0.1208	24.5949	23.2245	27.1224	-73.7252	20.3340	30.0130
12	-93.2461	6.7489	-0.1208	26.8513	25.0382	30.8279	-90.8756	22.6413	33.2247

表5.7の1行目の $y_0 = 2$ の逆推定値 \hat{x}_0 は、表5.2の $\hat{\beta}_0$ と $\hat{\beta}_1$ の推定値を用いて

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{2 - (-11.80)}{0.8864} = 15.5692$$

である。2次式のパラメータ a 、 b 、 c は、

$$a = \text{Var}(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} = 13.2745 - \frac{(2 - (-11.80))^2}{2.3060^2} = -22.5382$$

$$b = 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2 \times (-0.5926) + \frac{2 \times (2 - (-11.80)) \times 0.886}{2.3060^2} = 3.4152$$

$$c = \text{Var}(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_\alpha^2} = 0.0269 - \frac{0.8864^2}{2.3060^2} = -0.1208$$

表 5.2 :

項	推定値	$\varepsilon^T \varepsilon =$	18.9636	共分散	13.2745	-0.5926
$\hat{\beta}_0 =$	-11.8000	$\hat{\sigma}^2 =$	2.3705	行列	-0.5926	0.0269
$\hat{\beta}_1 =$	0.8864	$t_{0.05} =$	2.3060		$\hat{\Sigma} = (\mathbf{X}^T \mathbf{X})^{-1} \hat{\sigma}^2$	
$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$						

となる。2 次式の解は、2 つあり下側 95%点 \hat{x}_{L95} と上側 95%点 \hat{x}_{U95} が同時に求められ、

$$\begin{aligned} \hat{x}_{L95} &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \\ &= \frac{-3.4152 \pm \sqrt{3.4152^2 - 4 \times (-22.5382) \times (-0.1208)}}{2 \times (-0.1208)} = (10.4967, 17.7739) \end{aligned}$$

となる。個別の下側 95%点は、

$$\text{個別}a' = a + \hat{\sigma}^2 = -22.5382 + 2.3705 = -20.1678$$

なので、

$$\begin{aligned} \hat{x}_{\text{個別}L95} &= \frac{-b \pm \sqrt{b^2 - 4a'c}}{2c} \\ &= \frac{-3.4152 \pm \sqrt{3.4152^2 - 4 \times (-20.1678) \times (-0.1208)}}{2 \times (-0.1208)} = (8.4028, 19.8678) \end{aligned}$$

と計算される。図 5.3 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 5.8 から該当部分を抜粋した結果を表 5.9 に示すように、 $\hat{x}_0 = -0.296$ 、 $\hat{x}_{L95} = -0.675$ 、 $\hat{x}_{\text{個別}L95} = -1.243$ となる。

表 5.9 近似および正確 95%信頼区間の逆推定値の比較

		逆推定	95%信頼区間		個別95% CL	
	y_0	\hat{x}_0	\hat{x}_{L95}	\hat{x}_{U95}	個別L95	個別U95
正確	6	20.0821	17.9324	21.3764	14.9018	24.4070
近似	"	"	18.5737	21.5904	?	?

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パラメータの推定だけでなく、任意に設定した推定値となるようなパラメータを推定することができる。この方法を使うことにより、表 5.5 に示した回帰直線の 95%信頼区間の計算シートを用い、95%信頼区間の曲線が y_0 となるような逆推定値 \hat{x}_0 を Excel ソルバーで探索する。

表 5.10 に示すのは、表 5.5 をコピーした表に対し、1 行目の「U95%」が「6」となるように 1 行目の「x」をソルバーで探索した結果であり、表頭の「x」を「逆推定値 \hat{x}_0 」に代えた結

果である．さらに2行目の「L95%」が「6」となるように2行目の「 x 」をソルバーで探索した結果である．さらに，3行目の「 \hat{y} 」が「6」となるように3行目の「 x 」をソルバーで探索した結果は， $y_0 = 6$ となる逆推定値 $\hat{x}_{(y_0=6)} = 20.0821$ が推定されている．この逆推定値 $\hat{x}_{(y_0=6)}$ に対する95%信頼区間は，1行目目と2行目が95%信頼区間 (17.9324, 21.3764) となっている．

表 5.10 Excel ソルバーによる逆推定の実際

逆推定値	切片	逆推定値 x_0^{\wedge}	y^{\wedge}	$Var(y^{\wedge})$	L95%	U95%	個別 L95%	個別 U95%
L95%	1	17.9324	4.0946	0.6827	2.1892	6.0000	: 目標値(6)にセット	
U95%	1	21.3764	7.1473	0.2475	6.0000	8.2945		
$x^{\wedge}_{(y_0=6)}$	1	20.0821	6.0000	0.3361	4.6630	7.3370		
個別L95%	1	14.9017	1.4084	1.5943			-3.1833	6.0000
個別U95%	1	24.4070	9.8335	0.3931			6.0000	13.6670
$x^{\wedge}_{(y_0=6)}$	1	20.0821	6.0000	0.3361			2.2062	9.7938
T.inv.2T(0.05, 10-2)=					2.3060			

Excel のソルバーで「目標セルの設定」で当該のセルを選択し，「指定値」が「6」となるように元の表の「 x 」を変化させた結果で，表頭の「 x 」は，「逆推定値 \hat{x}_0 」に変更してある．

個別のU95%およびL95%についても同様な手順により，逆推定値 $\hat{x}_{(y_0=6)}$ に対する95%信頼区間は，4行目目と5行目が個別の95%信頼区間 (14.9017, 24.4070) となっている．これらは，表 5.9 に示した推定結果に一致する．

面積の標準偏差を推定したい (<https://mathwords.net/gosadenpa>)

長方形の横の長さを測定し $x=10\text{ cm}$ を得た．標準偏差は 0.3 cm だとする．同様に，縦の長さは $y=5\text{ cm}$ で少しでも正確に測定できて標準偏差は 0.2 cm だとする．面積は， $z=xy=10 \times 5 = 50\text{ cm}^2$ である．面積 z の標準偏差を誤差伝播の公式を用いて推定したい．

横の長さ x の分散は， $Var(x)=0.3^2$ ，縦の長さの分散は， $Var(y)=0.2^2$ であり， x と y の共分散は，互いに独立して計測しているので $Cov(x,y)=0$ である．面積 $z=xy$ の分散を推定したい．面積 z を x と y で偏微分すると $d_1=\partial z/\partial x=y=5$ ， $d_2=\partial z/\partial y=x=10$ となるので，分散共分散行列を Σ ，偏微分ベクトルを $\mathbf{d}=[d_1\ d_2]$ としたときに，面積 z の分散は， Σ に関する \mathbf{d} の2次形式 $Var(z) = \mathbf{d}^T \Sigma \mathbf{d}$ によって推定できる（誤差伝播の公式の一般化）．面積 z の標準偏差の推定値は，分散の平方根なので， $SD(z) = \sqrt{Var(z)} = \sqrt{6.25} = 2.5\text{ cm}^2$ となる．

表 5.11 面積の標準偏差の推定（誤差伝播の公式での計算に一致）

		分散共分散行列 Σ		\mathbf{d}		Σ		\mathbf{d}^T		$Var(z)$	$SD(z)$
$x=$	10	0.3 ²	0	5	10	0.09	0	5	=	6.25	2.5
$y=$	5	0	0.2 ²	d_1	d_2	0	0.04	10			
$z=xy=$	50	x	y								

5.7. JMP の回帰分析による逆推定

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グ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。

JMP の「二変量の関係」による回帰分析

表 5.12 に示すのは、JMP の「二変量の関係」による回帰分析の結果である。「分散分析」の結果は、表 5.4 に相当し、「パラメータの推定値」は、表 5.2 に相当する。図 5.4 に示すのは、デフォルトで出力される回帰直線の散布図に 95%信頼区間、および、個別データの 95%信頼区間を重ね書きした結果であり、図 5.1 に相当する。

表 5.12 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	69.1364	69.1364	29.1659
誤差	8	18.9636	2.3705	p値(Prob>F)
全体(修正済み)	9	88.1000		0.0006*
パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob> t)
切片	-11.8000	3.6434	-3.2387	0.0119*
x	0.8864	0.1641	5.4005	0.0006*

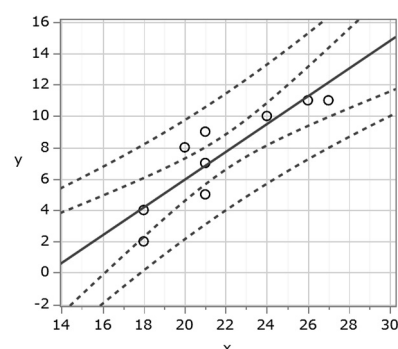


図 5.4 95%信頼区間の出力

回帰直線の推定値の 95%信頼区間の計算式

図 5.4 に示した 2 種類の 95%信頼区間は、どのような計算式が使われているのでしょうか。詳細は省くが、「平均の信頼限界の計算式」を選択すると表 5.13 に示すように、「平均 y の下側 95%」および「平均 y の上側 95%」が、元の JMP ファイルに新たな変数として追加される。さらに、「個別の信頼限界の計算式」を選択すると「個別 y の下側 95%」および「個別 y の上側 95%」が新たな変数として追加される。これらの追加された変数には、計算式が含まれており、内部での計算式を確認できるようになっている。これらの JMP 内部での計算式は、過度に標準化されている平方和を用いた式 (5.45) および 式 (5.46) と異なることが明らかである。これらの式は、 x_i に関する偏差平方和 $S_x = \sum_i (x_i - \bar{x})^2$ を用いた極めてシンプルな式で、魅力的ではあるものの、2 変数以上に拡張できない脆弱性を合わせ持っている。

表 5.13 JMP の「二変量の関係」による 95%信頼区間の出力

		x	y	予測値 y	平均 yの 下側95%	平均 yの 上側95%	個別 yの 下側95%	個別 yの 上側95%
○	1	18	4	4.15	2.27	6.04	0.13	8.17
○	2	21	7	6.81	5.63	8.00	3.07	10.56
○	3	24	10	9.47	8.12	10.83	5.67	13.27
○	4	24	10	9.47	8.12	10.83	5.67	13.27
○	5	18	2	4.15	2.27	6.04	0.13	8.17
○	6	27	11	12.13	9.93	14.33	7.95	16.31
○	7	21	9	6.81	5.63	8.00	3.07	10.56
○	8	21	5	6.81	5.63	8.00	3.07	10.56
○	9	26	11	11.25	9.36	13.13	7.23	15.27
○	10	20	8	5.93	4.57	7.28	2.13	9.73

表 5.14 に示すのは、JMP のデータファイル上で「平均 y の上側 95%」の計算式を表示させた結果である。計算式の `Vec Quadratic ()` 関数は、デザイン行列 X の積和に対する逆行列 $(X^T X)^{-1}$ に対し、デザイン行列 X の行ベクトルとの式 (5.41) に示した 2 次形式の計算を行う関数であり、誤差分散 $\sigma^2 = 2.3060$ 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。 $(X^T X)^{-1}$ に σ^2 を掛けた結果は、パラメータの共分散行列 $\Sigma(\hat{\beta})$ であり、推定値 \hat{y}_i の分散の計算にも $\Sigma(\hat{\beta})$ 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。

表 5.14 JMP の「二変量の関係」による上側 95%の計算式

$$-11.8 + 0.8863636364 \cdot x + 2.3060041352 \cdot \sqrt{\text{Vec Quadratic} \left(\begin{bmatrix} 5.6 & -0.25 \\ -0.25 & 0.011 \end{bmatrix}, \begin{bmatrix} 1 \end{bmatrix} || x \right)} \cdot 2.3704545455$$

表 5.6 で示されている Excel による回帰分析と対比すると VecQuadraticc () 関数は,

$$Var(\hat{y}_i) = \begin{bmatrix} 1 & x \\ 5.6000 & -0.2500 \\ -0.2500 & 0.0114 \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} * \begin{bmatrix} 2.3705 \\ \hat{\sigma}^2 \end{bmatrix}$$

と解釈される。回帰の 95%信頼区間の上側の計算式は、

$$\begin{aligned} L95\% &= \hat{\beta}_0 + \hat{\beta}_1 x_i + t_{0.05}(8) \sqrt{Var(\hat{y}_i)} \\ &= -11.80 + 0.8864x + 2.3060 \sqrt{Var(\hat{y}_i)} \end{aligned}$$

であることが、読み取れる.

表 5.15 に示すのは、「個別 y の上側 95%」の計算式である。この式は、

$$\text{個別}U95\% = \hat{\beta}_0 + \hat{\beta}_1 x_i + t_{0.05}(8) \sqrt{\text{Var}(\hat{y}_i) + \hat{\sigma}^2}$$

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

のように 2 次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 X を用いた計算となっている。

表 5.15 JMP の「二変量の関係」による個別データの上側 95%の計算式

$$\begin{aligned} & -11.8 \\ & + 0.8863636364 \cdot x \\ & 2.3060041352 \\ & + \sqrt{\text{Vec Quadratic} \left(\begin{bmatrix} 5.6 & -0.25 \\ -0.25 & 0.011 \end{bmatrix}, [1 \ || \ x] \right) \cdot 2.3704545455 + 2.3704545455} \end{aligned}$$

JMP の「モデルのあてはめ」による逆推定値の 95%信頼区間

逆推定を行い、Excel での計算結果との相互検証を行う。JMP の「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 5.17 に結果を示す。

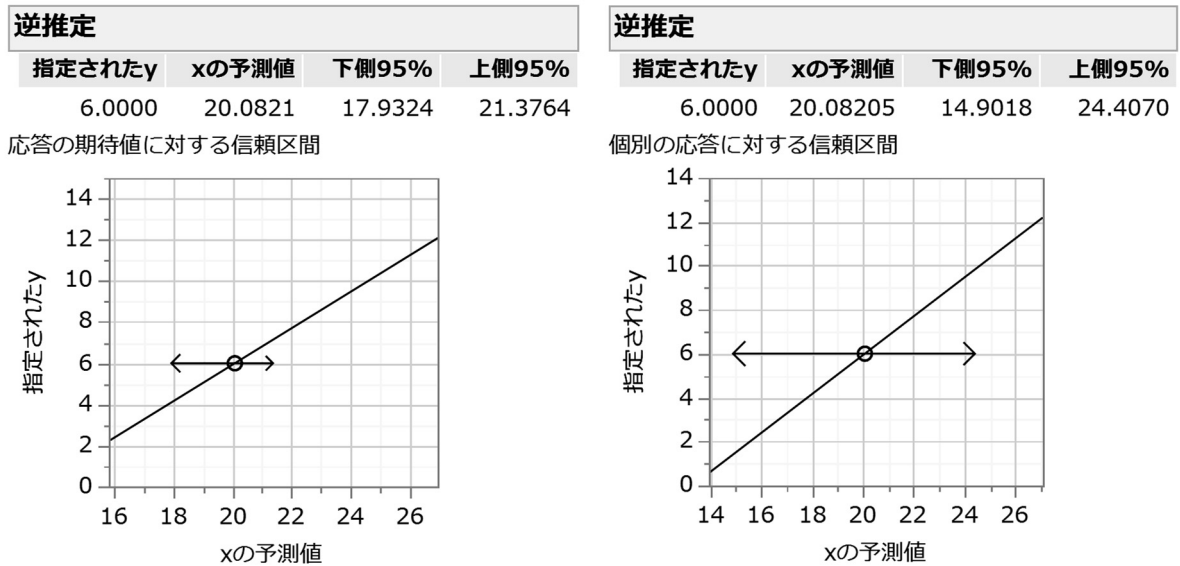
表 5.16 JMP の「逆推定」の設定

The image shows the JMP software interface. On the left, the 'Model Fit' menu is open, showing options like 'Regression Report', 'Predicted Values', 'Factor Screening', etc. The 'Predicted Values' option is selected. On the right, the 'Inverse Estimation' dialog box is open. It contains the following information:

- Header: 逆推定 (Inverse Estimation)
- Instruction: 逆推定したいY値を1つ以上、指定してください。 (Specify one or more Y values you want to inverse estimate.)
- Fields:
 - x (予測対象) (Prediction Target):
 - 信頼水準 (Confidence Level): 0.95
 - 両側 (Two-sided) dropdown menu.
 - y (Response): 6
- Checkbox: ☒ 応答変数の期待値ではなく、個々の値に対する信頼区間 (Not the expected value of the response variable, but the confidence interval for individual values).
- Buttons: OK, キャンセル (Cancel), ヘルプ (Help).

Below the dialog box, there is a text box explaining the process: Yの値(およびその他の説明変数の値)から、Xの値を予測したい場合。信頼区間も求められる。 (When you want to predict the value of X from the value of Y (and other explanatory variables), the confidence interval can also be obtained.)

表 5.17 JMP の「モデルのあてはめ」による逆推定の結果とグラフ表示



なお, JMP の逆推定による結果は, 表 5.9 の正確な 95%信頼区間の結果と完全に一致する.

表 5.9 抜粋

	逆推定	95%信頼区間		個別95% CL	
y_0	\hat{x}_0	\hat{x}_{L95}	\hat{x}_{U95}	個別L95	個別U95
正確 6	20.0821	17.9324	21.3764	14.9018	24.4070

逆推定値の 95%信頼区間の直接推定

推定された回帰直線のパラメータを用い, ある y_0 に対する逆推定値を $\hat{x}_0 = (y_0 - \hat{\beta}_0) / \hat{\beta}_1$ によって求め, その分散の推定のために \hat{x}_0 をパラメータで偏微分し, パラメータの共分散行列 $\Sigma(\hat{\beta})$ に関する 2 次形式の計算を必要とし, 煩雑な手順を示した. そこで, 逆推定値 \hat{x}_0 およびその分散を直接推定する方法を示す.

一般的に回帰直線は, 切片 β_0 と傾き β_1 をパラメータとして推定しているのだが, ある y_0 とその逆推定値 \hat{x}_0 が原点となるようなオフセット式を想定しよう. Y 軸となる反応 y に対し $y' = y - y_0$, X 軸の変数 x に対し $x' = x - \hat{x}_0$ とすると, 原点を通る回帰直線が

$$\left. \begin{aligned} y_i - y_0 &= \beta_1(x_i - \hat{x}_0) + \varepsilon_i \\ y_i &= y_0 + \beta_1(x_i - \hat{x}_0) + \varepsilon_i \end{aligned} \right\} \quad (5.66)$$

のように想定される. ただし, y_0 は, 与えられた定数であるが, 逆推定値 \hat{x}_0 は, データから推定される推定値であるので, $\beta_2 = \hat{x}_0$ と置き換え, 傾き β_1 と逆推定値 β_2 を推定し, それらの分散から, 逆推定値 \hat{x}_0 の 95%信頼区間を求めることができる.

式 (5.66) は、線形モデル式ではないので、残差 $\varepsilon_i = y_i - \hat{y}_i$ に対し線形モデルを適用する方法でパラメータの推定ができる。この方法は、第10章から第13章で活躍するのであるが、手短かに計算方法を示す。回帰式 (5.66) をパラメータ β_1 および β_2 で偏微分

$$z_{1,i} = \frac{\partial y_i}{\partial \beta_1} = x_i - \beta_2, \quad z_{2,i} = \frac{\partial y_i}{\partial \beta_2} = -\beta_1 \quad (5.67)$$

し、10行×2列の偏微分行列 \mathbf{Z} とする。表 5.18 に示すように、 $y_0 = 6.0$ を設定し、初期値 $\hat{\beta}_1 = 1$ 、 $\hat{\beta}_2 = 20$ に対し、式 (5.66) により推定値 \hat{y}_i を計算する。残差 $\varepsilon_i = y_i - \hat{y}_i$ について平方和 S_e を計算する。Excel のソルバーで、 S_e を最小にするように $(\hat{\beta}_1, \hat{\beta}_2)$ を変化させた結果が示され、逆推定値 $\hat{\beta}_2 = 20.0820$ が推定されている。偏微分行列 \mathbf{Z} を“デザイン行列 \mathbf{X} ”と見なすことにより、 $(\hat{\beta}_1, \hat{\beta}_2)$ に対する共分散行列 $\Sigma(\hat{\beta})$ が得られるので、その対角要素から $Var(\hat{\beta}_2) = 0.4278$ が得られ、一般的な方法により近似 95%信頼区間が (18.5737, 21.5904) が推定されている。

表 5.18 逆推定値の 95%信頼区間の直接推定

$y_0=$		6.0				共分散行列			近似95%信頼区間	
		初期値		推定値		$\Sigma(\beta^{\wedge})=(Z^T Z)^{-1}\sigma^{\wedge^2}$		$SE(\beta_i)$	$L95\%$	$U95\%$
		$\beta_1^{\wedge}=$	1.0	$\beta_1^{\wedge}=$	0.8864	0.0269	-0.0583	0.1641	0.5079	1.2648
		$\beta_2^{\wedge}=$	20.0	$\beta_2^{\wedge}=$	20.0820	-0.0583	0.4278	0.6541	18.5737	21.5904
				$S_e=$	18.9636	偏微分行列 Z		$t_{0.05}(10-2)=$	2.3060	
i	x	y	y^{\wedge}	$\varepsilon=y-y^{\wedge}$	$z_1=\partial_{\beta_1}\hat{y}$	$z_2=\partial_{\beta_2}\hat{y}$				
1	18	4	4.1546	-0.1546	-2.0820	0.8864				
2	21	7	6.8137	0.1863	0.9180	0.8864				
3	24	10	9.4727	0.5273	3.9180	0.8864				
4	24	10	9.4727	0.5273	3.9180	0.8864				
5	18	2	4.1546	-2.1546	-2.0820	0.8864				
6	27	11	12.1318	-1.1318	6.9180	0.8864				
7	21	9	6.8137	2.1863	0.9180	0.8864				
8	21	5	6.8137	-1.8137	0.9180	0.8864				
9	26	11	11.2455	-0.2455	5.9180	0.8864				
10	20	8	5.9273	2.0727	-0.0820	0.8864				

式 (5.66) は、一つの変数に対し、一つのパラメータの積の和で定義される線形モデルではないことから、非線形最小2乗法によりパラメータの推定を行なう必要がある。式 (5.66) に対して JMP の非線形回帰を使えば、表 5.17 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5章 文献 索引

大和田 (2010) - 線形モデルと非線形モデルの基本的な考え方ー逆推定の解析, 標準誤差と信頼限界ー	194
Collett (2003) - Modeling Binary Data 2nd. ed.	183
高橋 (2013a) - 応用回帰分析Iー各種の重み付き回帰における逆推定ー	183
高橋 (2013b) - 回帰分析・再入門ー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	183
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	183
アーミテージら著, 椿・椿訳 (2001) - 医学研究のための統計的方法, 第3.6節 分散に関するその他の公式	184
Draper N.R. and Smith H. (1998) - Applied Regression Analysis 3rd ed.	156
ドレーパ・スミス著, 中村慶一訳 (1968) - 応用回帰分析	156
中西 (2016) - 非線形回帰を用いた逆推定の基礎	194
芳賀 (2010) - 医薬品開発のための統計学, 第3部 非線形モデル 改訂版	183
MaCullough ら (2008) - On the accuracy of statistical procedures in Microsoft Excel 2007	155

第5章 索引

あ アーミテージら (2001) - 積の分散	184	あ 「X」 - 「 X 」をボールド&イタリック	157
- 比の分散	184	F.dist.RT () 関数 - F分布の上側確率	178
Average () 関数 - Excel	167	F分布の上側確率 - F.dist.RT () 関数	178
Arithmetic mean - design matrix	163	Minverse () 関数 - Excel	170
「いぶしろん」 - 「ε」に変換	157	Mdetarm () 関数 - 行列式	169
Intecept () 関数 - Excel	168	大和田 (2010) - 逆推定の解析	194
Wikipedia - 計画行列	163	offset from reference group - One-way ANOVA	163
- design matrix	163	オフセット式 - 逆推定値	193
内側の数的一致 - 積和の計算	160	か 回帰の平方和 - S_R	176
Excel - Average () 関数	167	回帰パラメータ - Excel の行列計算	174
- Intecept () 関数	168	- t 検定	176
- Minverse () 関数	170	- 分散および共分散の推定	175
- 解析マナーの習得	156	回帰パラメータの比 - 逆推定値	183
- 行列関数	155	回帰式のパラメータ推定 - 偏差平方和	164
- SumSq () 関数	161	回帰直線 - 95%信頼区間	178
- SumProduct () 関数	161, 168	- 95%信頼区間の計算式	190
- Slope () 関数	168	回帰直線の95%信頼区間 - 伝統的な方法	180
- Mdetarm () 関数	169	回帰分析 - 各種の推定	155
- DevSq () 関数	168	- 計算公式の範囲内	168
- Transpose () 関数	161	- 最小2乗法	164
- 分散分析表の作成	176	- データ分析ツール	158
- データ分析ツール	158	回帰分析の実際 - Excel の行列計算	174
- Mmult () 関数	161	解析マナーの習得 - Excel	156
Excel 2007 に対する批判 - MaCullough ら (2008)	155	各種の推定 - 回帰分析	155
Excel 2007 以前 - 計算ミス	178	ガラスの天井 - 計算公式の範囲内	168
- 互換性・メッセージに注意	178	- 2次式の95%信頼区間	181
Excel シート - 行列計算の実際 $X\beta$	159	逆行列 - 両辺に掛ける	169
- デザイン行列 X を入力	157	逆行列を前掛 - 単位行列	169
Excel ソルバー - 逆推定の実際	189	逆推定 - グラフ表示	193
- 正確な95%信頼区間	188	逆推定の解析 - 大和田 (2010)	194
Excel の回帰分析 - 現実的な対応	181	逆推定の基礎 - 中西 (2016)	194
Excel の行列計算 - 回帰パラメータ	174	逆推定の実際 - Excel ソルバー	189
- 回帰分析の実際	174	逆推定値 - オフセット式	193
S_R - 回帰の平方和	176	- 回帰パラメータの比	183
S_e - 残差平方和	176	- 95%信頼区間	183
S_e と S_R に分解 - 全体の平方和 S_T	182	- 95%信頼区間の直接推定	193
S_T - 全体の偏差平方和	176	- 個別データ	187
$S_T = S_e + S_R$ - 平方和の分解	177	- 正確な95%信頼区間	185, 187

か	逆推定値の95%信頼区間 - 参考	183	さ	算術平均 - デザイン行列	163
	- JMP の回帰分析	190		$\Sigma(\beta^{\wedge})$ - パラメータの共分散行列	178
	- モデルのあてはめ	192		シグマ - 読み飛ばしたくなる	155
	逆推定値の式 - パラメータで偏微分	184		シグマを用いた - 積和の計算	161
	逆推定値の比較 - 95%信頼区間	187		JMP - Vec Quadratic () 関数	191
	95%信頼区間 - 回帰直線	178		JMP の「二変量の関係」-95%信頼区間の出力	191
	- 逆推定値	183		JMP の回帰分析 - 逆推定値の95%信頼区間	190
	- 逆推定値の比較	187		- 個別の信頼限界の計算式	190
	- 個別データ	180		- 二変量のあてはめ	190
	95%信頼区間の計算式 - 回帰直線	190		- 平均の信頼限界の計算式	190
	95%信頼区間の出力 -JMP の「二変量の関係」	191		自由度 - 分散分析表	177
	95%信頼区間の直接推定 - 逆推定値	193		推定値 y^{\wedge} の分散の分散 - 2次形式	179
	共分散行列 - パラメータ	172		Slope () 関数 - Excel	168
	“行・列”の順番 - “列・行”ではなく	160		正確な95%信頼区間 - Excel ソルバー	188
	行ベクトル d - 偏微分式	184		- 逆推定値	185, 187
	行列の基本 - 矩形データ	155		- 2次式の解の公式	186
	行列の計算 - 嫌悪感	155		正規方程式 - 行列表記	165
	行列の表記 - 太い外枠	157		- 行列表記に対応付け	168
	行列関数 - Excel	155		正規方程式の解 - パラメータの推定	165
	行列計算 - パラメータ推定	169		積 $X\beta$ - 「積和」の計算	158
	行列計算の基礎 - デザイン行列 X	157		積の分散 - アーミテージら(2001)	184
	行列計算の実際 $X\beta$ - Excel シート	159		「積和」の計算 - 積 $X\beta$	158
	行列式 - Mdetarm () 関数	169		積和 - デザイン行列 X	162
	行列表記 - 正規方程式	165		積和 XY - デザイン行列 X と Y	162
	- Draper ら(1989)	168		積和の計算 - シグマを用いた	161
	- ドレーパら(1968)	168		- 内側の数が一致	160
	行列表記に対応付け - 正規方程式	168		切片の扱い - 定数に 0 を使用	158
	「 ϵ 」に変換 - 「いぶしろん」	157		cell means model - One-way ANOVA	163
	ギリシャ文字 β - 「ベータ」と入力	157		全体の平方和 S_T - S_e と S_R に分解	182
	近似95%信頼区間 - デルタ法	184		全体の偏差平方和 - S_T	176
	矩形データ - 行列の基本	155	た	Times New Roman - 半角英数字のフォント	157
	グラフ表示 - 逆推定	193		他の変数を定数 - 偏微分 β_0	164
	計画行列 - Wikipedia	163		単位行列 - 逆行列を前掛	169
	- デザイン行列	163		定義 - デザイン行列	163
	計算ミス - Excel 2007 以前	178		t 検定 - 回帰パラメータ	176
	計算公式の範囲内 - 回帰分析	168		定数に 0 を使用 - 切片の扱い	158
	- ガラスの天井	168		Mdetarm () 関数 - Excel	169
	嫌悪感 - 行列の計算	155		T.dist.2T () - t 分布の両側確率	175
	現実的な対応 - Excel の回帰分析	181		t 分布の両側確率 - T.dist.2T ()	175
	抗うつ薬 - プラセボ群	157		design matrix - Arithmetic mean	163
	合成分散 - パラメータの比	184		- Wikipedia	163
	合成分散の一般的式 - 2次形式	178		デザイン行列 - 計画行列	163
	互換性・メッセージに注意 - Excel 2007 以前	178		- 算術平均	163
	誤差伝播の公式 - 面積の標準偏差	189		- 定義	163
	誤差伝播の公式の一般化 - 分散共分散行列	189		デザイン行列 X - 行列計算の基礎	157
	誤差の平方和 Q - 偏微分	164		- 積和	162
	誤差分散 - 残差平方和	172		- 転置と積和	160
	個別データ - 逆推定値	187		- パラメータの共分散行列	173
	- 95%信頼区間	180		- 偏差平方和	170
	個別の信頼限界の計算式 -JMPの回帰分析	190		- 偏微分行列 Z	194
さ	最小2乗法 - 回帰分析	164		デザイン行列 X と Y - 積和 XY	162
	SumSq () 関数 - Excel	161		デザイン行列 X を入力 - Excelシート	157
	SumProduct () 関数 - Excel	161, 168		DevSq () 関数 - Excel	168
	参考 - 逆推定値の95%信頼区間	183		デルタ法 - 近似95%信頼区間	184
	残差平方和 - S_e	176		転置と積和 - デザイン行列 X	160
	- 誤差分散	172			

た	伝統的な方法 - 回帰直線の95%信頼区間	180	分散共分散行列 - 誤差伝播の公式の一般化	189
	統計ソフトの使い方 - 見方さえ習得	155	- 2次形式	189
	Transpose () 関数 - Excel	161	は 分散分析表 - 自由度	177
	Draper ら (1989) - 行列表記	168	分散分析表の作成 - Excel	176
	ドレーパら (1968) - 行列表記	168	データ分析ツール - Excel	158
な	中西 (2016) - 逆推定の基礎	194	- 回帰分析	158
	2次形式 - 合成分散の一般的式	178	平均の信頼限界の計算式 - JMPの回帰分析	190
	- 推定値 y^{\wedge} の分散の分散	179	平方和の分解 - $S_T = S_e + S_R$	177
	- 2次式の95%信頼区間	181	- 補足	182
	- 分散共分散行列	189	「ベータ」と入力 - ギリシャ文字 β	157
	- Vec Quadratic () 関数	191	Vec Quadratic () 関数 - JMP	191
	2次式の95%信頼区間 - ガラスの天井	181	- 2次形式	191
	- 2次形式	181	偏差平方和 - パラメータの推定の実際	167
	2次式の解の公式 - 正確な95%信頼区間	186	- 回帰式のパラメータ推定	164
	二変量のあてはめ - JMP の回帰分析	190	- デザイン行列 X	170
は	芳賀 (2010) - 非線形	194	- パラメータの分散の推定	171
	パラメータ - 共分散行列	172	偏微分 - 誤差の平方和 Q	164
	パラメータで偏微分 - 逆推定値の式	184	偏微分 β_0 - 他の変数を定数	164
	パラメータの共分散行列 - $\Sigma(\beta^{\wedge})$	178	偏微分行列 Z - デザイン行列 X	194
	- デザイン行列 X	173	偏微分式 - 行ベクトル d	184
	パラメータ推定 - 行列計算	169	「X」をボード&イタリック - 「 X 」	157
	- 正規方程式の解	165	補足 - 平方和の分解	182
	パラメータの推定の実際 - 偏差平方和	167	ま MaCullough ら (2008) - Excel 2007 に対する批判	155
	パラメータの比 - 合成分散	184	Mmult () 関数 - Excel	161
	パラメータの分散の推定 - 偏差平方和	171	見方さえ習得 - 統計ソフトの使い方	155
	半角英数字のフォント - Times New Roman	157	面積 $z = x y$ - 分散	189
	非線形 - 芳賀 (2010)	194	面積の標準偏差 - 誤差伝播の公式	189
	比の分散 - アーミティージら (2001)	184	モデルのあてはめ - 逆推定値の95%信頼区間	192
	太い外枠 - 行列の表記	157	や 読み飛ばしたくなる - シグマ	155
	プラセボ群 - 抗うつ薬	157	ら 両辺に掛ける - 逆行列	169
	分散 - 面積 $z = x y$	189	“列・行”ではなく - “行・列”の順番	160
	分散および共分散の推定 - 回帰パラメータ	175	わ One-way ANOVA - offset from reference group	163
			One-way ANOVA - cell means model	163

第5章 解析用ファイル一覧

サイズ	名前 ^		種類
 29 KB	第05章_02_行列計算_入門		Microsoft Excel ワークシート
 41 KB	第05章_03_回帰_偏差平方和		Microsoft Excel ワークシート
 21 KB	第05章_04_回帰_入門逆行列		Microsoft Excel ワークシート
 57 KB	第05章_05_回帰_デザイン行列		Microsoft Excel ワークシート
 87 KB	第05章_06_回帰_逆推定		Microsoft Excel ワークシート
 20 KB	第05章_07_回帰_JMP		Microsoft Excel ワークシート
 11 KB	第05章_07_回帰_逆推定		JMP Data Table

空白ページ

非売品，無断複製を禁ずる

第 12 回 続高橋セミナー
層別因子を含む探索的な回帰分析入門

第 5 章 デザイン行列を用いた回帰分析の基礎

BioStat 研究所(株)
〒105-0014 東京都 港区 芝 1-12-3 の1005
2024 年 2 月 5 日 高橋 行雄
takahashi.stat@nifty.com , FAX : 03-342-8035