第 3 回 続高橋セミナー

応用回帰分析 1

ー 各種の重み付き回帰における逆推定 ー

2013 年 2 月 14 日

高橋 行雄
BioStat 研究所(株)

要約： 広い範囲の x について検量線を引いたときに，x が大きくなるにつれて，y の変動が大きくなるので，x の逆数を重みにした回帰分析を適用したい．その場合に，回帰直線の 95%信頼区間はどのように求めたらよいのだろうか．多くの検査機器には，様々な重みを選択できるような回帰分析のプログラムが組み込まれているが，未知検体の逆推定の 95%信頼区間までは計算してくれない．計算式を知りたいと思い文献を調べたが見つからない．そこで，応用回帰分析 1 として，各種の重み付き回帰における逆推定の問題に対するさまざまな課題について丁寧な解説書を作成することにした．読者が追試できるように，計算式の誘導を丁寧に示し，多くの数値例を掲載した．

目 次

1. はじめに ---------- 1
2. Excel の行列関数による回帰分析の基礎と応用 ---------- 4
3. 原点を通る回帰式 ---------- 25
4. べき乗を含む回帰分析 ---------- 33
5. 分散が均一とはみなせない場合 ---------- 38
6. 分散が x の関数で現される場合 ---------- 49
7. 分散が x の関数で原点を通る回帰直線---------- 71
8. 曲線となる検量線 ---------- 78
9. 文献に対する批判的吟味 ---------- 92
索引---------- 135
Excel, JMP ファイル 一覧 ---------- 137

非売品，無断複製を禁ずる

連絡先:高橋 行雄，〒105-0014 東京都 港区 芝 1-12－3
takahashi.stat@nifty.com ， FAX : 03-342-8035

目　次

1.　はじめに ---------- 1
2.　Excel の行列関数による回帰分析の基礎と応用 ---------- 4
2.1.　回帰分析の行列による定式化 ---------- 4
2.2.　回帰パラメータの推定 ---------- 5
2.3.　各種の分散の推定 ---------- 7
誤差分散の推定 ---------- 7
回帰パラメータの分散の推定 ---------- 7
2.4.　各種の 95%信頼区間 ---------- 8
回帰直線の 95%信頼区間 ---------- 8
個別データの 95%信頼区間 ---------- 10
複数データの平均値に対する 95%信頼区間 ---------- 11
2.5.　逆推定の定式化 ---------- 12
回帰直線の逆推定 ---------- 12
個別データの逆推定 ---------- 15
繰返し m 回の平均値 95%信頼区間 ---------- 16
2.6.　Excel による計算方法および作図方法 ---------- 16
2.7.　統計ソフト JMP による検証 ---------- 20
2.8.　補足 ---------- 22
正規方程式 ---------- 22
回帰パラメータの分散 ---------- 23
回帰直線の 95%信頼区間の幅が最小になる点 ---------- 24
3.　原点を通る回帰式 ---------- 25
3.1.　回帰分析の行列による表記 ---------- 25
各種の 95%信頼区間の推定 ---------- 26
3.2.　各種の逆推定 ---------- 27
回帰直線 ---------- 27
個別データ ---------- 28
複数データの平均値 ---------- 28
3.3.　統計ソフト JMP による検証 ---------- 29
3.4.　Excel による逆推定の 95%信頼幅の作図法 ---------- 31
3.5.　補足 ---------- 32
正規方程式 ---------- 32
4.　べき乗を含む回帰分析 ---------- 33
4.1.　2 次式のあてはめ ---------- 33
回帰式の 95%信頼区間 ---------- 33
逆推定 ---------- 34

4.2. 2 次式から 1 次の項を除いたモデル 35
各種の 95%信頼区間 35
5. 分散が均一とはみなせない場合 38
5.1. 回帰分析の前提条件 38
5.2. 分散が x^2 または x に比例する場合 39
5.3. 分散が x^2 に比例するか x に比例するかの判断 42
5.4. 直線をあてはめることが困難な場合 43
5.5. 各種の誤差構造を持つデータの生成 45
6. 分散が x の関数で現される場合 49
6.1. 分散が x^2 に比例する回帰直線 49
回帰分析の行列による表記 49
重み付き回帰分析 49
各種の 95%信頼区間 51
各種の逆推定の 95%信頼区間 54
統計ソフトによる検証 55
正規方程式の誘導 57
誤差分散が最小となる x_0 58
6.2. 分散が x に比例する場合 60
各種の逆推定 62
6.3. 分散がおおよそ x^2 に比例 63
Excel による行列計算 65
6.4. 分散がおおよそ x に比例する場合 67
Excel による行列計算 68
6.5. 各種の重みに対する逆推定の 95%信頼区間の計算式 70
7. 分散が x の関数で原点を通る回帰直線 71
7.1. 分散が x^2 に比例し原点を通る場合 71
回帰分析の行列による表記 71
各種の 95%信頼区間 72
各種の逆推定の 95%信頼区間 73
7.2. 分散が x に比例し原点を通る直線 74
各種の逆推定の 95%信頼区間 76
8. 曲線となる検量線 78
8.1. データの例示 78
Excel による重みつき回帰の実用的な解析手順 79
用いる回帰式 80
各種の逆推定 82
2 次式の解を求めるためのパラメータの変更方法 83
8.2. 回帰式のべき λ の最尤法による推定 84
8.3. 重みに用いる x のべき乗の最尤法による推定 89

9. 文献に対する批判的吟味 92
9.1. ［付録 1］ 日本規格協会（2001） 94
序文（原著の引用） 94
比例する残差標準偏差を仮定した校正関数（原著の引用） 94
基本的方法（原著の引用） 95
（高橋の追試） 96
未知検体の 95%信頼区間の例示 98
9.2. ［付録 2］ Draper ら（1998） 102
INVERSE REGRESSION（原著の引用） 102
（高橋のコメント） 103
WEIGHTED LEAST SQUARES（原著の引用） 103
（高橋の追加） 105
9.3. ［付録 3］ Snedecor ら（1967），畑村ら共訳（1972） 109
Xが 標本ごとに変わる場合（原著の引用） 109
（高橋追加） 111
9.4. ［付録 4］ 芳賀ら（1996） 113
前書き（原著の引用） 113
いろいろの平均（原著の引用） 113
Proc reg による重みつき回帰（原著の要約） 114
（高橋追加） 114
誤差分散がηの関数である場合（原著の引用） 114
（高橋追加） 115
ロジスティック回帰における分散の推定の考え方 117
通常の回帰分析における分散の推定 117
9.5. ［付録 5］ 林ら（2000） 119
異なる重み（原著の要約） 119
9.6. ［付録 6］ 竹内啓（1963） 120
回帰直線自体についての推論（原著の引用） 120
（高橋のコメント） 121
9.7. ［付録 7］ Miller ら（2000），宗森ら訳（2004） 123
濃度の計算とその偶然誤差（原著の引用） 123
（高橋の追加 1） 124
重み付き回帰線（原著の引用） 126
（高橋の追加 2） 128
Excel による解析 130
曲線のあてはめ（原著の引用） 131
（高橋の追加 3） 132

1. はじめに

さまざまな身近な課題に対して回帰分析を応用する場合に，その方法が示された文献を探しあてたとしても計算式は示されているが，具体的な数値例がない．実務に使いたいと考えたとしても，計算結果が正しいとの確認ができず，使うことをためらわれる．結果が正しいことを検証レポートにまとめることもできず，使うことを断念することになってしまう．また，身近にある統計ソフト，あるいは，Excel の分析ツールなどで計算しようと思っても，抽象的な記載で，実際の計算方法がよくわからない場合も起きる．また，実際に計算できるのか，できないかも，判断が付かないこともしばしば起きるであろう．

単回帰分析を行なって回帰パラメータ（切片，傾き）求め，未知検体について m 個の測定値の平均 $\bar{y}_0$ を得たときに，回帰直線上で $\bar{y}_0$ となる x_0 の推定値，その 95%信頼区間を求めたいとしよう．逆推定値 $\hat{x}_0$ を求めることは簡単な数値計算で求められるが，その 95%信頼区間 （$\hat{x}_{L95m}$，$\hat{x}_{U95m}$）はどのように計算したらよいのだろうか．回帰式の 95%信頼区間なのか，個別データの 95%信頼区間のどちらでもないと思われるが，数値例が載っている文献を探してもても見つからない．

標準的な回帰分析の教科書から，x_0 に対する回帰直線の Y 軸方向の 95%信頼区間 （$\hat{y}_{L95R}$，$\hat{y}_{U95R}$）を求める式，個別データの 95%信頼区間 （$\hat{y}_{L95I}$，$\hat{y}_{U95I}$）を求める式は，多くの文献にあり，多くの統計ソフトで計算可能のである．しかしながら，逆推定値 $\hat{x}_0$ の場合，必要なのは X 軸方向の個別データの 95%信頼区間であり，Y 軸方向の 95%信頼区間ではないので，使うことができない．

広い範囲の x について検量線を引いたときに，x が大きくなるにつれて，y の変動が大きくなるので，分散の逆数を重みにした回帰分析が必要だといわれている．しかしながら，多くの分析機器に搭載されている計算ソフトには，重みに選択できるのは，$1/x$，$1/x^2$，$1/y$，$1/y^2$，$1/\hat{y}$ などで，「分散の逆数」は見当たらない．疑問は，さらにふくらみ，どれを選んだら良いのか，身近な教科書を調べても何も記載がない．その場合に，回帰直線あるいは個別データの 95%信頼区間，は，どのように求めたらよいのだろうか．文献を調べても見つからない，さらに，未知検体の 95%信頼区間は，どのように計算したらよいのか文献を調べたが全く見つからない．

検量線に関連する諸問題は，統計解析の基本である回帰分析に関する応用問題である．それにもかかかわらず，参考となる教科書や論文が極めて少ないことを実感した．最も基本の単回帰分析の場合にも応用問題に直面した場合に，それを自ら解決するためには，一般的な教科書で示されている Σ(シグマ)を使った計算式は，手計算に対応した式となっているために，各種の応用問題に対する見通しを付け難い．

回帰分析を応用するためには，Excel の行列関数を用いた定式化が見通しのよい学習の助けに

なる. 簡単な問題に対しては, **Σ** 記号を用いた定式化での学習もよいが, さまざまな応用をするためには行列を用いた計算が必須となる. 自ら応用問題を目指すならば, 最初から行列を用いた回帰分析を学習することを薦める. また, 計算結果が正しいかの自己検証には, Excel のグラフで計算結果を図示し, それらの結果の相互関係に矛盾がないことを示すとよい.

簡単な回帰分析であれば Excel の分析ツール, あるいは, 回帰分析用の Linest 関数で解決できるが,「逆推定」,「重み付き回帰分析」などには対応できない. これらの応用問題に対し, 一般的に使用されている統計ソフトの活用も考慮すべきである. そのためには, 無料の R 言語を用いることも考えられるが, GUI に優れた統計ソフト JMP の使用を薦める.

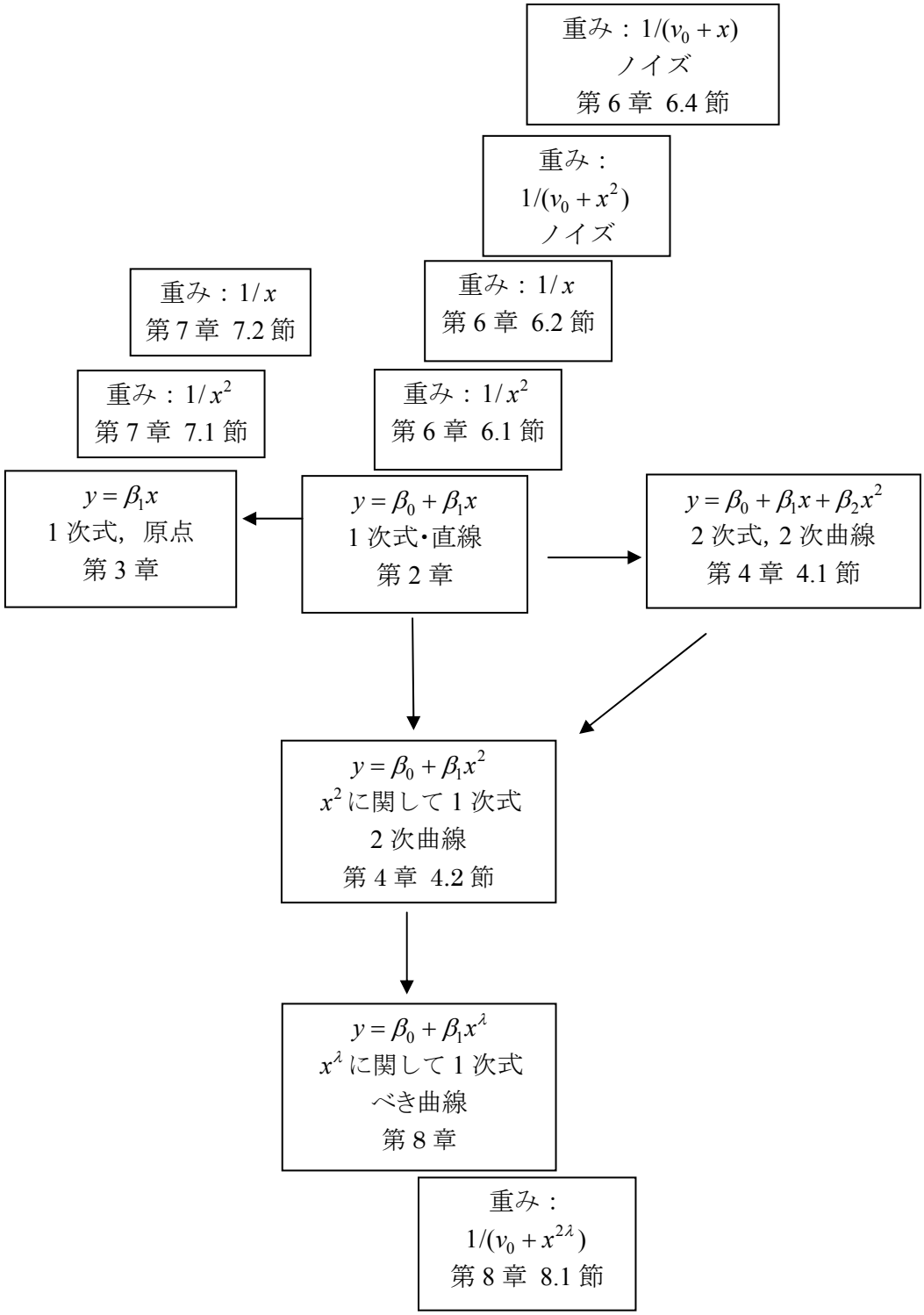
さらに, 測定機器にこれらの応用問題を組み込む際には, C 言語などで行列関数をサブプログラム化して, 多くの一般ユーザが日常的に活用できるようになってもらいたいと願っている.

さまざまな検体に含まれる化学物質の濃度の測定を請負って, 顧客に測定結果を報告することを業務にしている「分析センター」は, その測定結果の点推定値のみならず, その 95%信頼区間を日常的に表示するようになることを願っている.

一般ユーザにとって多くの選択肢があり, 自己責任でどれかを選択できるようにしておくことは, 何ら問題はないのであるが, その選択に際して適切な判断基準, さらには, 適宜ユーザ教育なども恒常的に開催すことも不可欠と考えている.

数少ない文献は参考になる面も多くあったが, 改めて精査すると不具合が散見しており, 単に引用するだけではなく, 批判的な吟味の必要性を痛感した. そのために, 原著の該当部分を数ページにわたり原文をそのまま引用した. 原著の図ではなく, すべて新たに作成し直すとともに, 個別データの 95%信頼区間など原書にない結果を含め, 何が不具合なのかが判断しやすいように試みた.

本書で取り上げた各種の重み付き回帰



2. Excelの行列関数による回帰分析の基礎と応用

2.1. 回帰分析の行列による定式化

5 個のデータ $(x, y) = (1, 0), (1, 2), (2, 3), (3, 4), (3, 6)$ について,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma_e^2),\quad i = 1, 2, ..., 5 \tag{2.1}$$

誤差 ε_i が互いに独立に平均, 分散が σ_e^2 の正規分布に従うと仮定した最小 2 乗法によって回帰分析を行ない, 回帰パラメータ(β_0, β_1)を推定したい. Excel の行列関数を用いて計算する場合には, 式(2.1)を用いるのではなく, β_0 に x_{0i} なる変数を明示した

$$y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma_e^2),\ i = 1, 2, ..., 5 \tag{2.2}$$

を用いる. ただし, 全ての i について $x_{0i} = 1$ とする. 一般的には, x_{0i} はすべて 1 なので, これを省略して $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, あるいは, $y = a + bx$, $y = ax + b$ と記述されることもある. しかしながら, さまざまな回帰分析の応用を考えるためには, 線形回帰モデルの基本である式(2.2)を用いることが見通しのよい計算手順を示すことができる. .

式(2.2)は,

$$y_i = \sum_{k=0}^{1} \beta_k x_{ki} + \varepsilon_i,\quad i = 1, 2, ..., 5 \tag{2.3}$$

と Σ 記号を用いて表すことができる. さらに, 行列を用いて現すと, データ数 $n = 5$, 変数の数, ここでは, x_{0i} も含めて変数の数 $p = 2$ の場合も明示することなく, 一般式

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.4}$$

で表すことができる. Excel シート上で, 式(2.4)に数値を入れて示そう.

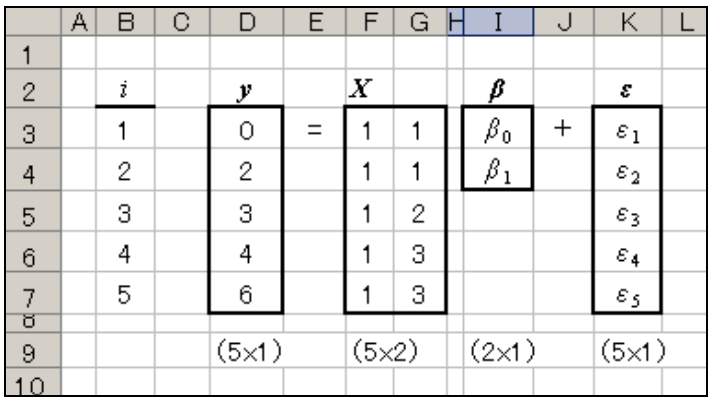


図 2.1 Exxel シート上での行列表現

行列 $\boldsymbol{y}$ は, 大きさ(5×1), $\boldsymbol{X}$ は(5×2), $\boldsymbol{\beta}$ は(2×1), $\boldsymbol{\varepsilon}$ は(5×1)の長方形(矩形)で表現されている. $\boldsymbol{X\beta}$ は, 行列の積で, $\boldsymbol{X}$ の行方向の i に対して, $\boldsymbol{\beta}$ の列方向の積和を, 次に示すように計算する. 行列の積(5×2)(2×1)は, それぞれの内側の大きさが等しくなければならない. 行列の積 $\boldsymbol{X\beta}$ の結果は, それぞれの内側の大きさ 2 が消えて, (5×1)と行列となる.

$$\boldsymbol{X\beta} \qquad\qquad \boldsymbol{X\beta}+\boldsymbol{\varepsilon}$$

$$\begin{bmatrix} 1\times\beta_0 + 1\times\beta_1 \\ 1\times\beta_0 + 1\times\beta_1 \\ 1\times\beta_0 + 2\times\beta_1 \\ 1\times\beta_0 + 3\times\beta_1 \\ 1\times\beta_0 + 3\times\beta_1 \end{bmatrix} \qquad \begin{bmatrix} 1\times\beta_0 + 1\times\beta_1 + \varepsilon_1 \\ 1\times\beta_0 + 1\times\beta_1 + \varepsilon_2 \\ 1\times\beta_0 + 2\times\beta_1 + \varepsilon_3 \\ 1\times\beta_0 + 3\times\beta_1 + \varepsilon_4 \\ 1\times\beta_0 + 3\times\beta_1 + \varepsilon_5 \end{bmatrix}$$

（5x1）　　（5x1）

これらの行列を，

$\boldsymbol{y}$:　測定値（反応・応答）行列（ベクトル）$\boldsymbol{y}$,

$\boldsymbol{X}$:　計画行列，デザイン行列 $\boldsymbol{X}$,

$\boldsymbol{\varepsilon}$:　誤差行列（ベクトル）$\boldsymbol{\varepsilon}$,

と言うことにする．大きさ（1×1）の場合はスカラーと区別することもあるが，全て行列として扱うことにする．

行列の和は，行列の対応するセルごとの和となる．したがって，行列の大きさは等しくなければならない．大きさ（5×1）の誤差 $\boldsymbol{\varepsilon}$ を $\boldsymbol{X\beta}$ に加えても行列の大きさは変わらずに（5×1）になる．

2.2. 回帰パラメータの推定

回帰パラメータ $\boldsymbol{\beta}$ を最小 2 乗法で推定したい．最小 2 乗法とは，誤差 $\boldsymbol{\varepsilon}$ のそれぞれの要素 i ごとに平方し，すべて加えた偏差平方和 S_e が最小になるように，回帰パラメータ $\boldsymbol{\beta}$ の推定値 $\hat{\boldsymbol{\beta}}$（ベータ・ハット）を求めるための解析法である．$\hat{\boldsymbol{\beta}}$ を求める方法については，2.8 節の補則で詳しく示すが，次式で与えられる．

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2.5}$$

ここで，$\boldsymbol{X}^T$ はもとの $\boldsymbol{X}$ の行と列を入れ替えた転置行列（Excel：Transpose 関数，トランスポーズ関数）という．転置は，（5×2）の長方形を裏返しつつ右上方に 90 度に回転して貼り付けるイメージである．$\boldsymbol{X}$ は（5×2）の大きさなので，$\boldsymbol{X}^T$ の大きさは（2×5）となる．行列の積 $\boldsymbol{X}^T\boldsymbol{X}$ は大きさが（2×5）と（5×2）なので，内側の 5 が消えて（2×2）の正方行列となる．

行列の積の計算（Excel：Mmult 関数，マトリックス・マルチプル関数）は，前項の i 列目に，次項の i 行目のセル同士の積を加えたものである．$\boldsymbol{X}^T\boldsymbol{X}$ の 2 行 2 列目の 24 は，$\boldsymbol{X}$ の 2 列目と $\boldsymbol{X}^T$ の 2 行目のセルの積和として計算されたもので，

$$(1\times1)+(1\times1)+(2\times2)+(3\times3)+(3\times3)=24$$

の計算結果である．Excel で実際に行列計算を行なう手順については，2.6 節を参照のこと．

$$\boldsymbol{X}^T \quad \boldsymbol{X} \quad = \quad \boldsymbol{X}^T\boldsymbol{X}$$

$$\begin{array}{c} \boldsymbol{X}^T \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 3 \end{bmatrix} \\ (2\text{x}5) \end{array} \begin{array}{c} \boldsymbol{X} \\ \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \\ (5\text{x}2) \end{array} = \begin{array}{c} \boldsymbol{X}^T\boldsymbol{X} \\ \begin{bmatrix} 5 & 10 \\ 10 & 24 \end{bmatrix} \\ (2\text{x}2) \end{array}$$

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ は逆行列（Excel : Minverse 関数，マトリックス・インバース関数）といわれるもので，$(\boldsymbol{X}^T\boldsymbol{X})$ と $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ の積を計算したときに，その行列の対角要素が全て 1 となる単位行列 $\boldsymbol{I}$ となる．逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ は，Minverse 関数で計算した結果である．$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ の 2 行目と $\boldsymbol{X}^T\boldsymbol{X}$ の 2 列目の積和は，$(10\times(-0.5))+(24\times 0.25)=1$ となることが確認できる．なお，大きさが（1×1）の場合には通常の逆数と同じである．逆行列の計算方法については，2.8 節の補則を参照のこと．

$$\begin{array}{c} \boldsymbol{X}^T\boldsymbol{X} \\ \begin{bmatrix} 5 & 10 \\ 10 & 24 \end{bmatrix} \\ (2\text{x}2) \end{array} \begin{array}{c} (\boldsymbol{X}^T\boldsymbol{X})^{-1} \\ \begin{bmatrix} 1.20 & -0.50 \\ -0.50 & 0.25 \end{bmatrix} \\ (2\text{x}2) \end{array} = \begin{array}{c} \boldsymbol{I} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ (2\text{x}2) \end{array}$$

逆行列の結果を用いて行列の積を Mmult 関数で推定値 $\hat{\boldsymbol{\beta}}$ を求めてみよう．逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ の大きさは（2×2），$\boldsymbol{X}^T$ は（2×5）なので $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\ \boldsymbol{X}^T$ の大きさは（2×5）となる．実際に手を動かして計算して行列の積の計算に慣れてもらいたい．さらに $\boldsymbol{y}$ は（5×1）なので，推定値 $\hat{\boldsymbol{\beta}}$ の大きさは（2×1）となる．

$$\begin{array}{c} (\boldsymbol{X}^T\boldsymbol{X})^{-1} \\ \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix} \\ (2\text{x}2) \end{array} \begin{array}{c} \boldsymbol{X}^T \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 3 \end{bmatrix} \\ (2\text{x}5) \end{array} \begin{array}{c} \boldsymbol{y} \\ \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \\ 6 \end{bmatrix} \\ (5\text{x}1) \end{array} = \begin{array}{c} \boldsymbol{\beta}\wedge \\ \begin{bmatrix} -1 \\ 2 \end{bmatrix} \\ (2\text{x}1) \end{array}$$

回帰分析のパラメータ，切片 $\hat{\beta}_0=-1$，傾き $\hat{\beta}_1=2$ が得られた．これを用いて，x_{1i} に対する回帰直線上の推定値 $\hat{y}_i$ は

$$\hat{y}_i=\hat{\beta}_0 x_{0i}+\hat{\beta}_1 x_{1i}\ ,\ \ i=1,2,\ldots,5 \tag{2.6}$$

となる．行列で現すと

$$\hat{\boldsymbol{y}}=\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{2.7}$$

となり，Excel の行列計算では，Y 軸の切片が -1，傾きが 2 の回帰直線上の推定値 $\hat{\boldsymbol{y}}$（y ハット）が得られる．

$$\underset{(5\times2)}{\boldsymbol{X}}\ \underset{(2\times1)}{\boldsymbol{\beta}\hat{}} = \underset{(5\times1)}{\boldsymbol{y}\hat{}}\ ,\qquad \underset{(5\times1)}{\boldsymbol{y}} - \underset{(5\times1)}{\boldsymbol{y}\hat{}} = \underset{(5\times1)}{\boldsymbol{e}}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \\ 5 \\ 5 \end{bmatrix},\qquad \begin{bmatrix} 0 \\ 2 \\ 3 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 3 \\ 5 \\ 5 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

測定された y_i と推定値 $\hat{y}_i$ の差は，誤差 e_i と現す．推定されたものなので誤差 $\hat{\varepsilon}_i$ と書くべきではあるが，推定値にギリシャ文字にハットを付ける代わりにアルファベットを使うことも一般的であり，誤差については $\boldsymbol{e}$ を用いることにする．回帰係数 β_0 の推定値を $\hat{\beta}_0$ ではなく b_0 とする教科書もある．

2.3. 各種の分散の推定

誤差分散の推定

回帰直線からの測定値 y_i と回帰直線上の推定値 $\hat{y}_i$ との誤差を

$$e_i = y_i - \hat{y}_i\ ,\quad i = 1,2,...,5 \tag{2.8}$$

としたときに，それらの誤差平方和 S_e は，次のようになる．

$$S_e = \sum_{i=1}^{5}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{5} e_i^2\ ,\ i = 1,2,...,5 \tag{2.9}$$

行列で現すと，$S_e = \boldsymbol{e}^T\boldsymbol{e}$ となる．

$$\underset{(1\times5)}{\boldsymbol{e}^T}\ \underset{(5\times1)}{\boldsymbol{e}} = \underset{(1\times1)}{S_e}$$

$$\begin{bmatrix} -1 & 1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \end{bmatrix}$$

誤差分散 $\hat{\sigma}_e^2$ は，データ数 $n = 5$ から推定に用いたパラメータの数 $p = 2$ を差し引いた自由度 $f_e = n - p = 3$ で，誤差平方和 S_e を割って得られる．

$$\hat{\sigma}_e^2 = \frac{S_e}{n-p} = \frac{4}{5-2} = 1.3333 \tag{2.10}$$

回帰パラメータの分散の推定

切片 $\hat{\beta}_0$ の分散 $Var(\hat{\beta}_0)$，傾き $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$，$\hat{\beta}_0$ と $\hat{\beta}_1$ の共分散 $Cov(\hat{\beta}_0, \hat{\beta}_1)$，これらを分散共分散行列 $\boldsymbol{\Sigma}$ として表す．なお，x_i を用いた式は，煩雑で本質的な内容の理解を妨げるとの理由から本文中でも，実際の計算でも使っていないので，一括して 2.8 節の補足の回帰パラメータで示した．

$$\boldsymbol{\Sigma} = \begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) \\ Cov(\hat{\beta}_0, \hat{\beta}_1) & Var(\hat{\beta}_1) \end{bmatrix} \tag{2.11}$$

分散共分散行列 $\boldsymbol{\Sigma}$ は，デザイン行列 $\boldsymbol{X}$ に誤差分散 $\hat{\sigma}_e^2$ を掛けて求められる．

$$\boldsymbol{\Sigma} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\sigma}_e^2 \tag{2.12}$$

行列とスカラーの積の計算は，行列の全てのセルに対してスカラーを掛け合わせる．

$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		$\sigma_e{}^{2\wedge}$	=	$\boldsymbol{\Sigma}$	
1.20	−0.50	1.333		1.600	−0.667
−0.50	0.25			−0.667	0.333
(2x2)		(1x1)		(2x2)	

$\boldsymbol{\Sigma}$ の対角要素が，$Var(\hat{\beta}_0) = 1.600$ と $Var(\hat{\beta}_1) = 0.333$ となり，2 列目 1 行目が共分散 $Cov(\hat{\beta}_0, \hat{\beta}_1) = -0.667$ となる．共分散行列 $\boldsymbol{\Sigma}$ を用いることにより，回帰直線の 95%信頼区間，個別データの 95%信頼区間，3 個のデータの平均値に対する 95%信頼区間が容易に計算できる．

分散共分散行列 $\boldsymbol{\Sigma}$ の対角要素の平方根が，回帰パラメータの標準誤差 SE となる．標準的な回帰分析で出力される回帰パラメータに対する t 統計量は，$t_k = \hat{\beta}_k / \sqrt{Var(\hat{\beta}_k)}$ で計算される．さらに，自由度 3 の t 統計量の両側確率 p は Excel の Tdist 関数（t□distribution，分布関数）で計算する．

表 2.1　回帰パラメータの推定

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k$^	$Var(\beta_k\hat{})$	SE	t	p
0	x_0:切片	β_0	−1.0	1.6000	1.2649	−0.7906	0.4869
1	x_1:傾き	β_1	2.0	0.3333	0.5774	3.4641	0.0405

2.4. 各種の95%信頼区間

回帰直線の 95%信頼区間

回帰直線の 95%信頼区間は，回帰直線上の推定値 $\hat{y}_i$ の分散 $Var(\hat{y}_i)$ の平方に，自由度が $f_e = n - p = 3$ の t 分布の両側 5%点を掛け，$\hat{y}_i$ にプラス・マイナスした，

$$\hat{y}_i \text{ の } 95\%\text{CL} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)} \tag{2.13}$$

として求められる．$Var(\hat{y}_i)$ は，$\hat{y}_i = \hat{\beta}_0 x_{0i} + \hat{\beta}_0 x_{1i}$ として計算されているので，分散に関する基本演算から，次のように，

$$\begin{aligned} Var(\hat{y}_i) &= Var(\hat{\beta}_0 x_{0i} + \hat{\beta}_1 x_{1i}) \\ &= Var(\hat{\beta}_0 x_{0i}) + 2Cov(\hat{\beta}_0 x_{0i}, \hat{\beta}_1 x_{1i}) + Var(\hat{\beta}_1 x_{1i}) \\ &= x_{0i}^2 Var(\hat{\beta}_0) + 2x_{0i}x_{1i}Cov(\hat{\beta}_0, \hat{\beta}_1) + x_{1i}^2 Var(\hat{\beta}_1) \end{aligned} \tag{2.14}$$

求められる．一般式で書けば，

$$Var(\hat{y}_i) = \sum_{k=0}^{p-1}\sum_{k'=0}^{p-1} x_{ki}x_{k'i}Cov(\hat{\beta}_k, \hat{\beta}_{k'}) \tag{2.15}$$

となる．ただし，$Cov(\hat{\beta}_0, \hat{\beta}_0)$ のように回帰パラメータの添字が同じ場合は，

$$Cov(\hat{\beta}_0, \hat{\beta}_0) = Var(\hat{\beta}_0) \tag{2.16}$$

とする．多くのでは，$Var(\hat{y}_i)$ を x_i を用いて，

$$Var(\hat{\beta}_0 + \hat{\beta}_1 x_i) = \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right) \hat{\sigma}_e^2 \tag{2.17}$$

などのような式が示されているが，本書では，$Var(\hat{y}_i)$ のままとし，実際に計算する場合には，次に示す行列計算式を用いている．

第 i 番目デザイン行列（x_{0i}，x_{1i}）を行列（ベクトル）$\boldsymbol{x}_i$ で現したときに，$\hat{y}_i = \hat{\beta}_0 x_{0i} + \hat{\beta}_0 x_{1i}$ の分散は，

$$Var(\hat{y}_i) = \boldsymbol{x}_i \boldsymbol{\Sigma} \boldsymbol{x}_i^T \tag{2.18}$$

のように，共分散行列 $\boldsymbol{\Sigma}$ に $\boldsymbol{x}_i$ を両側から挟みこむ形式（2 次形式）で計算することができる．第 5 番目の $\boldsymbol{x}_5 = (x_{0,5}, x_{1,5}) = (1,3)$ について $\hat{y}_5 = \hat{\beta}_0 x_{0,5} + \hat{\beta}_0 x_{1,5}$ の分散 $Var(\hat{y}_5)$ の行列計算の結果を次に示す．

$$\underset{(1\times2)}{\boxed{\begin{matrix} 1 & 3 \end{matrix}}} \quad \underset{(2\times2)}{\boxed{\begin{matrix} 1.600 & -0.667 \\ -0.667 & 0.333 \end{matrix}}} \quad \underset{(2\times1)}{\boxed{\begin{matrix} 1 \\ 3 \end{matrix}}} \quad = \quad \underset{(1\times1)}{\boxed{0.600}}$$

$$\boldsymbol{x}_5 \qquad \boldsymbol{\Sigma} \qquad \boldsymbol{x}_5^T \qquad = \quad Var(y_5\hat{\ })$$

式（2.14）でも行列計算と同じ結果が求められる．

$$\begin{aligned} Var(\hat{y}_5) &= Var(\hat{\beta}_0 x_{0,5} + \hat{\beta}_1 x_{1,5}) \\ &= Var(\hat{\beta}_0 + 3\hat{\beta}_1) \\ &= Var(\hat{\beta}_0) + 2 \times 3Cov(\hat{\beta}_0, \hat{\beta}_1) + 3^2 Var(\hat{\beta}_1) \\ &= 1.600 - 6 \times 0.667 + 9 \times 0.333 = 0.600 \end{aligned}$$

自由度 3 の t 分布の両側 5%点を Excel の Tinv 関数（t inverse,インバース関数）で計算すると $t_{0.05} = 3.182$ となるので，回帰直線の 95%信頼区間は，

$$\begin{aligned} \hat{y}_5 \text{の} 95\% CL &= \hat{y}_5 \pm t(3, 0.05)\sqrt{Var(\hat{y}_5)} \\ &= 3 \pm 3.182\sqrt{0.600} \\ &= (2.535,\ 7.465) \end{aligned}$$

が得られる．図 2.2 には，切片が 1，傾きが 2 の回帰直線，回帰直線の 95%信頼曲線が示され，$x = 3.5$ の位置に 95%信頼区間が引かれている．なお，Excel での作図については，2.6 節で詳しく説明する．

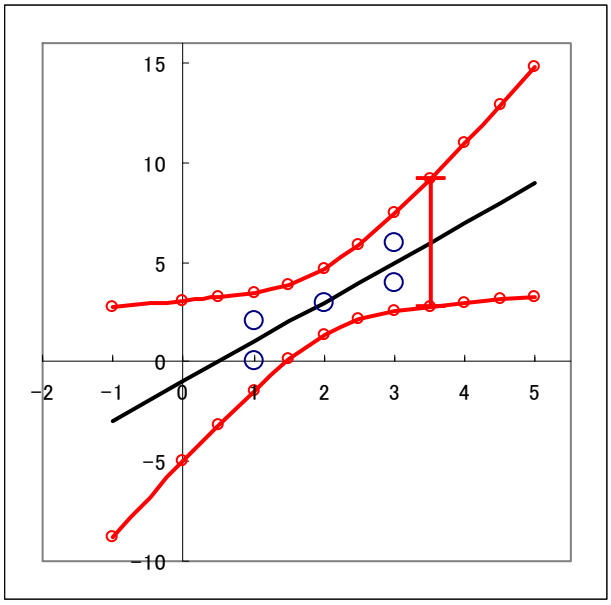


図 2.2　回帰直線の 95%信頼区間

個別データの 95%信頼区間

個別データについての 95%信頼区間は，回帰直線上の $\hat{y}_i$ の分散に，誤差 $y_i - \hat{y}_i = e_i$ の分散を加えればよい．推定値 $\hat{y}_i$ と誤差 e_i の相関は 0 となるので共分散も 0 となり，全体の分散は，それぞれの分散を加える（分散の加法性）．誤差 e_i の分散は $\hat{\sigma}_e^2$ なので，

$$\begin{aligned} Var(\hat{y}_i + e_i) &= Var(\hat{y}_i) + Var(e_i) \\ &= Var(\hat{y}_i) + \hat{\sigma}_e^2 \end{aligned} \tag{2.19}$$

となり，個別データの 95%信頼区間は，回帰直線の分散 $Var(\hat{y}_i)$ に，個別データの分散 $\hat{\sigma}_e^2$ を加えた分散を用いればよい．

$$\text{個別データの}95\%CL = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + \hat{\sigma}_e^2} \tag{2.20}$$

実際 $\boldsymbol{x}_5 = (x_{0,5}, x_{1,5}) = (1, 3)$ の場合に回帰直線の分散は $Var(\hat{y}_5) = 0.60$，個別データの分散は $\hat{\sigma}_e^2 = 1.333$ なので，

$$\begin{aligned} \text{個別のデータの}95\%CL &= 3.0 \pm 3.182\sqrt{0.60 + 1.333} \\ &= (0.575,\ 9.425) \end{aligned} \tag{2.21}$$

となる．図 2.3 には，回帰直線，回帰直線の 95%信頼曲線（内側），個別データの 95%信頼曲線（外側）に示され，$x = 3.5$ の位置に 95%信頼区間が引かれている．

データの存在範囲を超えると個別データの 95%信頼曲線は，回帰直線の 95%信頼曲線に近づく．これは，個別データの分散は，回帰直線の分散 $Var(\hat{y}_i)$ に一定量の誤差分散 $\hat{\sigma}_e^2$ を加えており，$Var(\hat{y}_i)$ が x に関連して大きくなった場合に，$\hat{\sigma}_e^2$ を加えて平方根をとった場合に，$\hat{\sigma}_e^2$ の影響がが相対的に小さくなるためである．

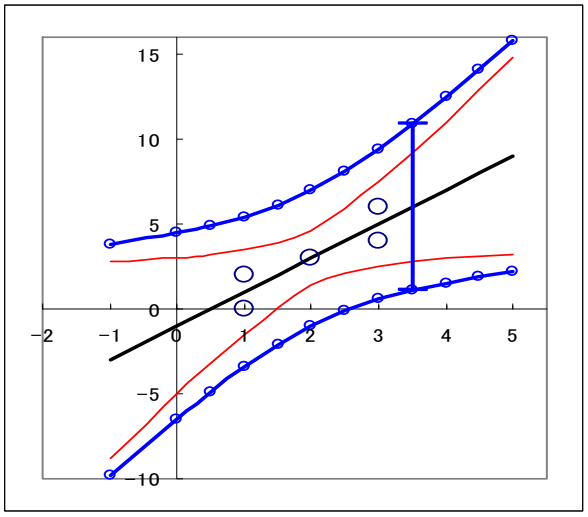


図 2.3　個別データについての 95%信頼区間

なお，個別データの 95%信頼区間の計算公式に，式(2.17)に 1 を加えた

$$Var(\hat{y}_i)+\hat{\sigma}_e^2=\left(1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)\hat{\sigma}_e^2 \tag{2.22}$$

が一般的に使われているが，本書では，式の内容把握がしやすい $Var(\hat{y}_i)+\hat{\sigma}_e^2$ で統一している．

複数データの平均値に対する 95%信頼区間

同じ x_i について複数データの平均値 $\bar{y}_{i\cdot}$ に対する 95%信頼区間は，どうなるのであろうか．このデータの構造は，$(\hat{y}_i+\varepsilon_{ij})$，$j=1,2,\ldots,m$ としたときに，その平均は，

$$\begin{aligned}\bar{y}_{i\cdot}&=\frac{y_{i1}+\cdots+y_{im}}{m}\\&=\frac{\hat{y}_{i1}+e_{i1}}{m}+\cdots+\frac{\hat{y}_{im}+e_{im}}{m}\\&=\hat{y}_i+\frac{e_{i1}}{m}+\cdots+\frac{e_{im}}{m}\end{aligned} \tag{2.23}$$

なので，その分散 $Var\ (\bar{y}_{i\cdot})$ は，

$$\begin{aligned}Var(\bar{y}_{i\cdot})&=Var(\hat{y}_i+\frac{e_{i1}}{m}+\cdots+\frac{e_{im}}{m})\\&=Var(\hat{y}_i)+\frac{1}{m^2}\hat{\sigma}_e^2+\cdots+\frac{1}{m^2}\hat{\sigma}_e^2\\&=Var(\hat{y}_i)+\frac{\hat{\sigma}_e^2}{m}\end{aligned} \tag{2.24}$$

となり，$Var(\hat{y}_i)$ に，誤差分散 σ_e^2/m を加えたものを用いればよい．したがって，その 95%信頼区間は，

$$m\text{個の平均の}95\%CL=\hat{y}_i\pm t(f_e,0.05)\sqrt{Var(\hat{y}_i)+\frac{\hat{\sigma}_e^2}{m}} \tag{2.25}$$

となる．個別データの 95%信頼区間の計算式の誤差分散 $\hat{\sigma}_e^2$ に変えて σ_e^2/m として 95%信頼区間を

計算する．第 5 番目の $\boldsymbol{x}_5 = (1, 3)$ の 95%信頼区間は，

$$m\text{個のデータの}95\%CL = 3.0 \pm 3.182\sqrt{0.60 + \frac{1.333}{3}}$$
$$= (1.748,\ 8.252\ )$$

となる．図 2.4 の 3 本の曲線の中央が m 個のデータの 95%信頼曲線となり，データ数 m が大きくなるにつれて回帰直線の 95%信頼幅の曲線に近づく．

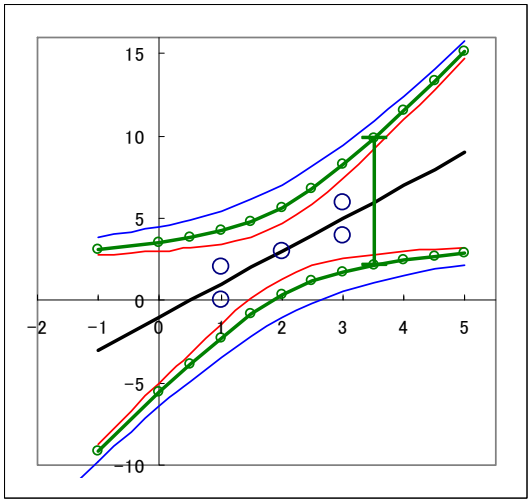


図 2.4 複数データ（m=3）の平均値の 95%信頼区間

2.5. 逆推定の定式化

回帰直線の逆推定

回帰分析における逆推定とは，ある y_0 が与えられたときに，あらかじめ求められた回帰直線上の $\hat{x}_0$ を推定する問題である．ある y_0 は，推定したい $\hat{x}_0$ を使って回帰式 $y_0 = \hat{\beta}_0 + \hat{\beta}_1\hat{x}_0$ で関連付けられる．この式を $\hat{x}_0$ について解けば，

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} \tag{2.26}$$

となるので，$y_0 = 3.5$ の場合であれば，切片 $\hat{\beta}_0 = -1$，傾き $\hat{\beta}_1 = 2$ であれば，

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{3.5 - (-1)}{2} = 2.25$$

となる．推定された $\hat{x}_0$ の 95%信頼区間を求めたい．図 2.2 に上書きした 95%信頼区間のように，x_0 が与えられたとすれば，Y 軸方向の 95%信頼区間について式(2.13)にすでに示した．問題は y_0 が与えられたときに，推定された $\hat{x}_0$ について，X 軸方向についての 95%信頼区間を求めたいのである．図 2.5 に示すように，y_0 を通る水平線を引き，95%信頼区間との交点から，垂直方向に線を引いて X 軸の目盛りを読むことにより，図からおおよその 95%信頼区間を求めることができる．そのような手作業ではなく，計算式を用いて推定したい．

95%信頼区間の下限の推定値を $\hat{x}_{L95}$ としたときに，$\hat{x}_{L95}$ における回帰直線上の推定値は，

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

で与えられる．その推定値 $\hat{y}_{\hat{x}_{L95}}$ の 95%信頼区間は，式(2.13)と同様に，自由度 f_e の t 分布の両側 5%点を t_α としたときに，

$$\hat{x}_0 \text{ の } 95\%\text{信頼区間} = \hat{y}_{\hat{x}_{L95}} \pm t_\alpha \sqrt{Var(\hat{y}_{\hat{x}_{L95}})} \tag{2.27}$$

で与えられる．その上側の 95%信頼曲線が，y_0 を通る X 軸に平行な直線との交点は，

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} = y_0 \tag{2.28}$$

で与えられる．この式を $\hat{x}_{L95}$ について解くことにより，$\hat{x}_0$ の下側の 95%信頼区間の式が求められる．

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

$$Var(\hat{\beta}_0) + 2\hat{x}_{L95} Cov(\hat{\beta}_0, \hat{\beta}_1) + \hat{x}_{L95}^2 Var(\hat{\beta}_1) = \frac{(y_0 - \hat{\beta}_0)^2 - 2(y_0 - \hat{\beta}_0)\hat{\beta}_1 \hat{x}_{L95} + \hat{\beta}_1^2 \hat{x}_{L95}^2}{t_\alpha^2} \tag{2.29}$$

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

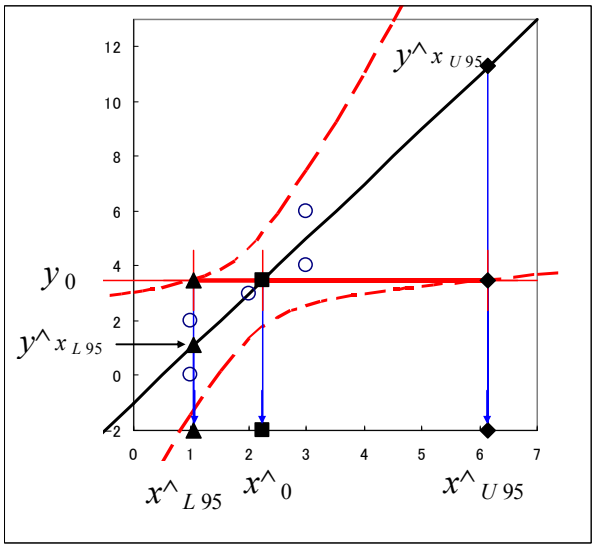


図 2.5 逆推定の模式図

式は複雑であるが，$\hat{x}_{L95}$ に関して 2 次式

$$a + b\hat{x}_{L95} + c\hat{x}_{L95}^2 = 0 \tag{2.30}$$

ただし，

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

$$b = 2Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} \tag{2.31}$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^{\,2}}{t_\alpha^2}$$

となる．従って，2 次式の解の公式

$$\hat{x}_{L95} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \tag{2.32}$$

を用いて x_{L95} を求めればよい．解は 2 あるが，これは，上側の 95%信頼区間を $\hat{x}_{U95}$ としたときに回帰直線の下側の 95%信頼曲線が y_0 となるのは，

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{U95}) - t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{U95})} = y_0 \tag{2.33}$$

で与えられる．この式を $\hat{x}_{U95}$ について解くと式(2.26)の $\hat{x}_{L95}$ を $\hat{x}_{U95}$ に置き換えた式が得られ，次の 2 次式を解くことになる．

$$a + b\,\hat{x}_{U95} + c\,\hat{x}_{U95}^2 = 0 \tag{2.34}$$

式の変形の過程で，±がついた項を 2 乗するために同じ係数を持った 2 次式が得られ，どちらでも同じ 2 つの解となる．したがって，どちらの式を用いても 2 つの解の小さい方が $\hat{x}_{L95}$ となり，大きい方が $\hat{x}_{U95}$ となる．

Excel を用いて計算をしてみよう．

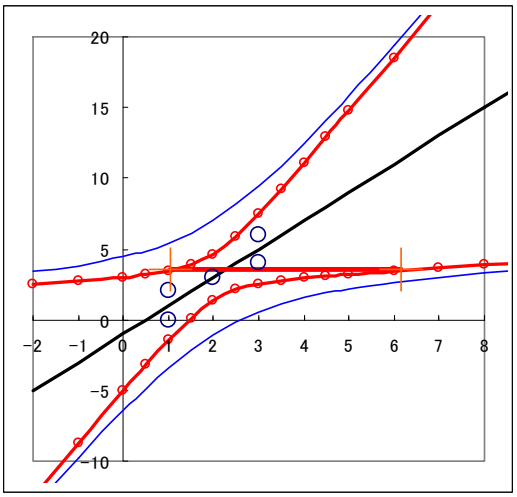
β^	$t_{0.05}$=	3.182	f_e=	3
−1	Σ=	1.600	−0.667	
2		−0.667	0.333	
	$\sigma_e{}^2$ ^=	1.333		

$$a = Var(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} = 1.6 - \frac{(3.5 - (-1))^2}{3.182^2} = -0.399$$

$$b = 2Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2 \times (-0.667) + \frac{2 \times (3.5 - (-1)) \times 2}{3.182^2} = 0.444$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^{\,2}}{t_\alpha^2} = 0.333 - \frac{2^2}{3.182^2} = -0.062$$

$$(\hat{x}_{L95}, \hat{x}_{U95}) = \frac{-0.444 \pm \sqrt{0.444^2 - 4 \times (-0.399) \times (-0.062)}}{2 \times (-0.062)} = (1.054, 6.151)$$



逆推定 y_0 =	3.500	x_0^ =	2.250
回帰直線の95%CL			
a =	−0.399	x_{L95R}^ =	1.054
b =	0.444	x_{U95}^ =	6.151
c =	−0.062		

図 2.6　回帰直線の逆推定の 95%信頼区間

個別データの逆推定

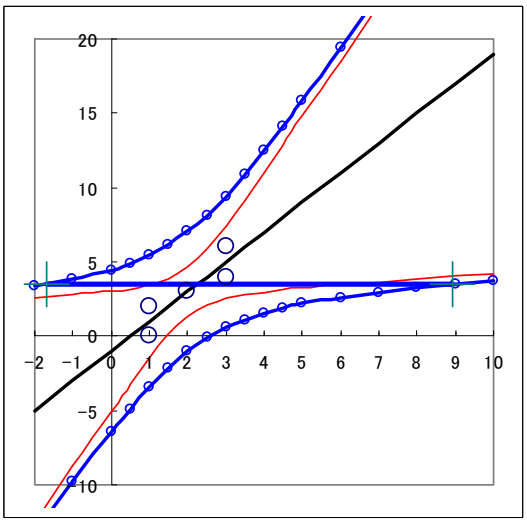
個別データの場合は，回帰直線の 95%信頼区に対して，式@@で示したように，誤差分散 $\hat{\sigma}_e^2$ が加わるので，同様に 95%信頼区間の下限の推定値を $\hat{x}_{L95I}$ としたときに，$\hat{x}_{L95I}$ 点での 95%信頼区間の上限が y_0 となるは，

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95I}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95I}) + \hat{\sigma}_e^2} = y_0 \tag{2.35}$$

である．この式を $\hat{x}_{L95I}$ について解くと 2 次式の定数項 a に $\hat{\sigma}_e^2$ が加わるだけで，1 次の係数 b，および，2 次の係数 c は変化しない．実際に計算すると，

$$a' = a + \hat{\sigma}_e^2 = -0.399 + 1.333 = 0.934$$

が得られる．



個別データの95%CL			
a' =	0.934	x_{L95I}^ =	−1.702
b =	0.444	x_{U95I}^ =	8.907
c =	−0.062		

図 2.7　個別データの逆推定の 95%信頼区間

繰返し m 回の平均値 95%信頼区間

同じ x_i について m 個のデータの平均値 $\bar{y}_{i\bullet}$ の 95%信頼区間を考える．データの構造は，$(\hat{y}_i + \varepsilon_{ij})$，$j = 1, 2, \ldots, m$ であり，その平均は，

繰り返し測定された数回のデータの平均値についての逆推定は，m 個のデータの平均値の 95%信頼区間の式から誘導される．

$$m\text{個の平均の}95\%\text{CL} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + \frac{\hat{\sigma}_e^2}{m}} \tag{2.36}$$

個別データの分散 σ_e^2 をデータ数 m で割った σ_e^2 / m となるので，2 次式の a 項を次のように a'' と置

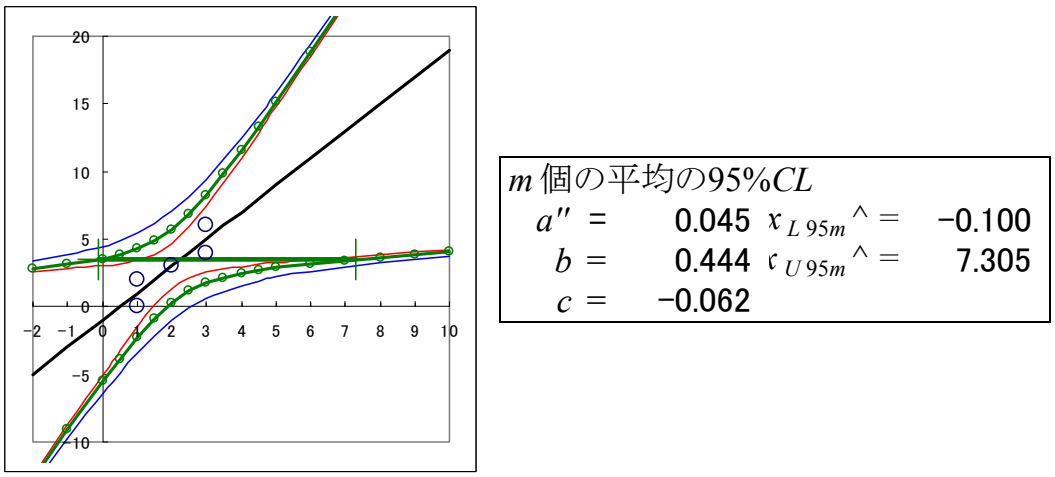


図 2.8　繰返し m 回の平均値の逆推定の 95%信頼区間

き換えればよい．

$$a'' = a + \frac{\hat{\sigma}_e^2}{m} = -0.399 + \frac{1.333}{3} = 0.045$$

2.6. Excel による計算方法および作図方法

Excel での行列計算の基本を，転置行列の作成を用いて説明する．大きさ (5×2) デザイン行列 $\boldsymbol{X}$ を転置して，大きさ (2×5) 転置行列 $\boldsymbol{X}^T$ を作成したい．

手順 1）E3 セルで左クリックしたまま，I4 セルまで移動し，E3:I4 の範囲を選択する．

手順 2）そのままにして，関数 ＝transpose(を入力し，転置したい行列 $\boldsymbol{X}$ の範囲 B3:C7 を選択する．

手順 3）関数の最後 ）を入力する．

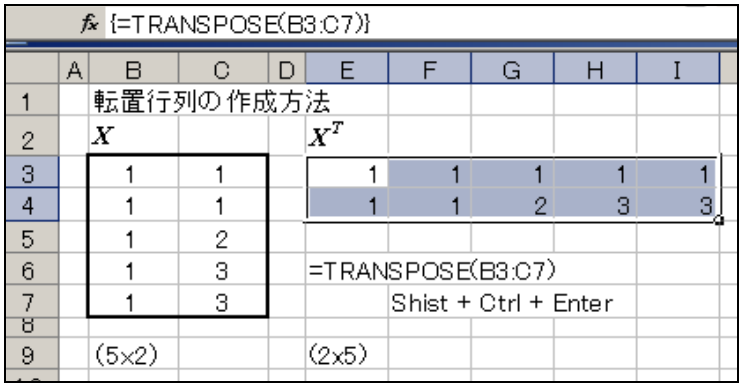


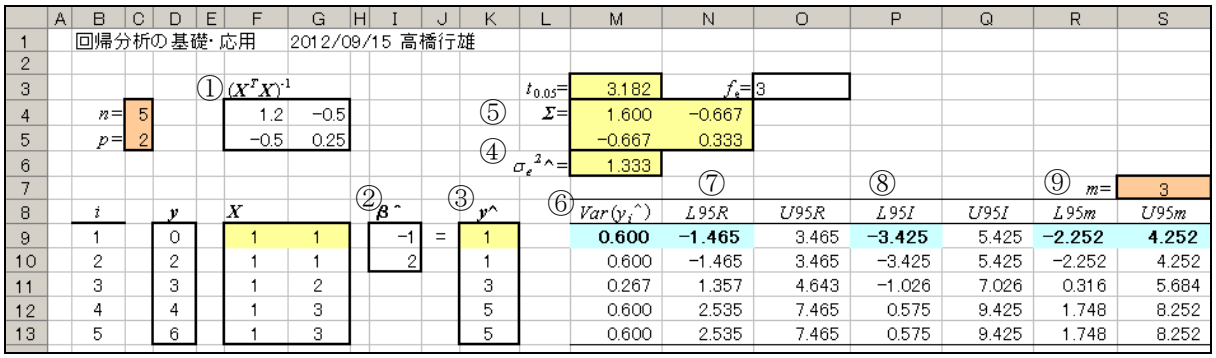
図 2.9　Excel の Transpose 関数の使い方

手順 4）E3:I4 の範囲が選択されていることを確認し，Sift キー と Ctrl キーを同時に押しながら，Enter キーを押す．

手順 5）計算式全体が {**計算式**}：{=TRANSPOSE(B3:C7)} と {} で挟まれていることを確認する．

回帰パラメータ $\hat{\boldsymbol{\beta}}$ の計算は，まず① $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ を計算しておき，次に② $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ を計算する．回帰パラメータ $\hat{\boldsymbol{\beta}}$ が求まれば，$\boldsymbol{y}$ の推定値を③ $\hat{\boldsymbol{y}} = \boldsymbol{X}^T\boldsymbol{\beta}$ として計算できる．誤差分散 $\hat{\sigma}_e^2$ は，誤差ベクトル $\boldsymbol{\sigma} = (\boldsymbol{y} - \hat{\boldsymbol{y}})$ を計算せず④ $\hat{\sigma}_e^2 = (\boldsymbol{y} - \hat{\boldsymbol{y}})^T(\boldsymbol{y} - \hat{\boldsymbol{y}})$ で直接求めている．分散共分散行列は，⑤ $\boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e^2$ で計算している．

与えられたそれぞれの $\boldsymbol{x}_i$ から推定された回直線上の $\hat{y}_i = \boldsymbol{x}_i\hat{\boldsymbol{\beta}}$ の分散 $Var(\hat{y}_i)$ は，⑥ $Var(\hat{y}_i) = \boldsymbol{x}_i^T\boldsymbol{\Sigma}\boldsymbol{x}_i$，回帰直線の 95%信頼区間は，⑦ $\hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)}$，個別データの 95%信頼区間は，⑧ $\hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + \sigma_e^2}$ で，m 個の平均の 95%信頼区間は，⑨ $\hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + \sigma_e^2/m}$ で計算している．



$Var(y_i\hat{})$	=MMULT(MMULT($F8:$G8,$M$3:$N$4),TRANSPOSE($F8:$G8))
$L95R$=	=$K8-$M$2*SQRT($M8)
$L95I$=	=$K8-$M$2*SQRT($M8+$M$5)
$L95\%m$	=$K8-$M$2*SQRT($M8+$M$5/$S$6)
$U95\%m$	=$K8+$M$2*SQRT($M8+$M$5/$S$6)

図 2.10　各種 95%信頼区間の計算シート

これまで，回帰直線，各種の 95%信頼区間について Excel の散布図の機能を用いて示してきた．これらの作図のもとになる Excel のシート上の計算プロセスを示し，それらを図示する方法について基本的な操作方法の解説をする．

Excel で各種の曲線をきれいに書くためには，曲線の始点から終点まで，適用な間隔の座標点が必要である．そのために，x について –1から 5 まで 0.5 刻みに別途データを用意し，$\hat{y}$ の計算，$Var(\hat{y})$，各種 95%の信頼区間の計算シートを用意しておく．

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
8		i		y		X			β ^		y^		$Var(y_i$ ^$)$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
9		1		0		1	1		−1	=	1		**0.600**	**−1.465**	3.465	**−3.425**	5.425	**−2.252**	**4.252**
10		2		2		1	1		2		1		0.600	−1.465	3.465	−3.425	5.425	−2.252	4.252
11		3		3		1	2				3		0.267	1.357	4.643	−1.026	7.026	0.316	5.684
12		4		4		1	3				5		0.600	2.535	7.465	0.575	9.425	1.748	8.252
13		5		6		1	3				5		0.600	2.535	7.465	0.575	9.425	1.748	8.252
14																			
15		図作成のための				1	−2.0				−5.0		5.600	−12.531	2.531	−13.380	3.380	−12.824	2.824
16		計算シート				1	−1.0				−3.0		3.267	−8.752	2.752	−9.826	3.826	−9.131	3.131
17						1	0.0				−1.0		1.600	−5.026	3.026	−6.451	4.451	−5.550	3.550
18						1	0.5				0.0		1.017	−3.209	3.209	−4.879	4.879	−3.847	3.847
19						1	1.0				1.0		0.600	−1.465	3.465	−3.425	5.425	−2.252	4.252
20						1	1.5				2.0		0.350	0.117	3.883	−2.129	6.129	−0.837	4.837
21						1	2.0				3.0		0.267	1.357	4.643	−1.026	7.026	0.316	5.684
22						1	2.5				4.0		0.350	2.117	5.883	−0.129	8.129	1.163	6.837
23						1	3.0				5.0		0.600	2.535	7.465	0.575	9.425	1.748	8.252
24						1	3.5				6.0		1.017	2.791	9.209	1.121	10.879	2.153	9.847
25						1	4.0				7.0		1.600	2.974	11.026	1.549	12.451	2.450	11.550
26						1	4.5				8.0		2.350	3.121	12.879	1.892	14.108	2.680	13.320
27						1	5.0				9.0		3.267	3.248	14.752	2.174	15.826	2.869	15.131
28						1	6.0				11.0		5.600	3.469	18.531	2.620	19.380	3.176	18.824
29						1	7.0				13.0		8.600	3.667	22.333	2.970	23.030	3.429	22.571
30						1	8.0				15.0		12.267	3.854	26.146	3.264	26.736	3.654	26.346
31						1	9.0				17.0		16.600	4.034	29.966	3.523	30.477	3.861	30.139
32						1	10.0				19.0		21.600	4.209	33.791	3.760	34.240	4.058	33.942

図 2.11　図作成のための計算シート

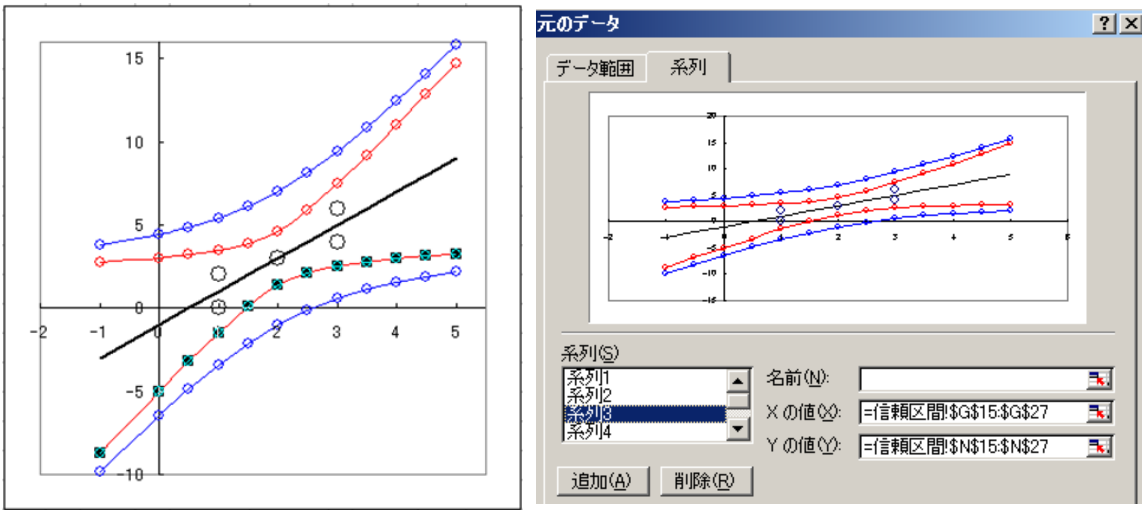


図 2.12　Excel の散布図【元のデータ／系列／系列 3】を用いた 95%信頼区間の表示

Excel のグラフ･ウイザードの【散布図】を選びデータの範囲を【系列】によって X 軸とする x の範囲，Y 軸に y の範囲を【系列】を用いて逐次追加する．測定データは，X 軸に F9:G13，Y 軸に D9:D13 を設定する．回帰直線の下側 95%信頼区間は，【系列 3】で軸に G15:G27，Y 軸に N15:N27 を設定し

ている(X 軸の範囲を-2 から 10 としている).

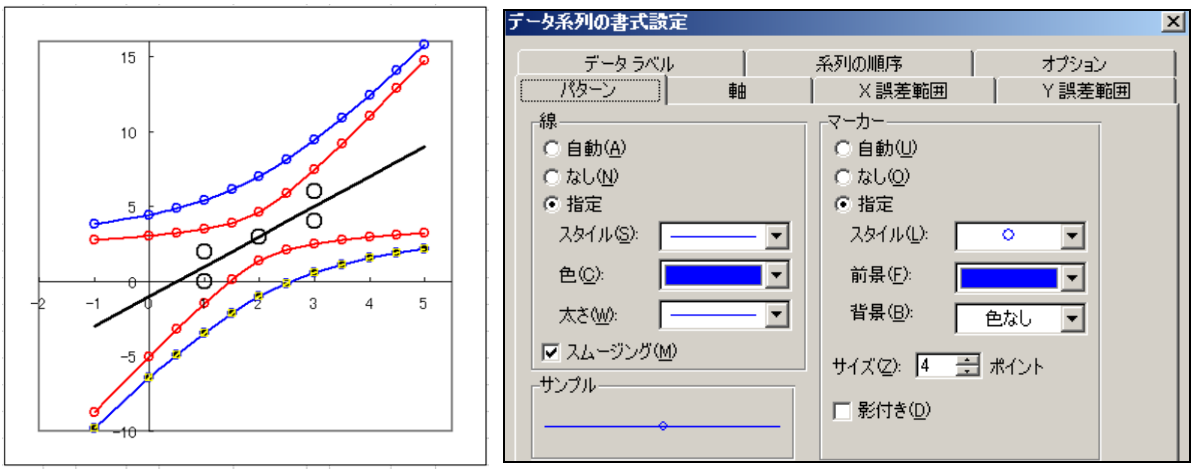


図 2.13 データ系列の書式設定の例示 1

Excel の散布図【元のデータ／系列】で作成しただけでは，ディフォルトのマーカと色の設定なので，【データ系列の書式設定／パターン／線】で好みの書式に設定する．曲線を滑らかにするために，【スムージング】を設定するとよい．

【マーカ】の機能を使い，オープンサークル○を設定するために，スタイルを○，前景を【青】，背景を【色なし】，サイズを【8 ポイント】と設定している．背景を【青】とすると●のように自由に変更できる..

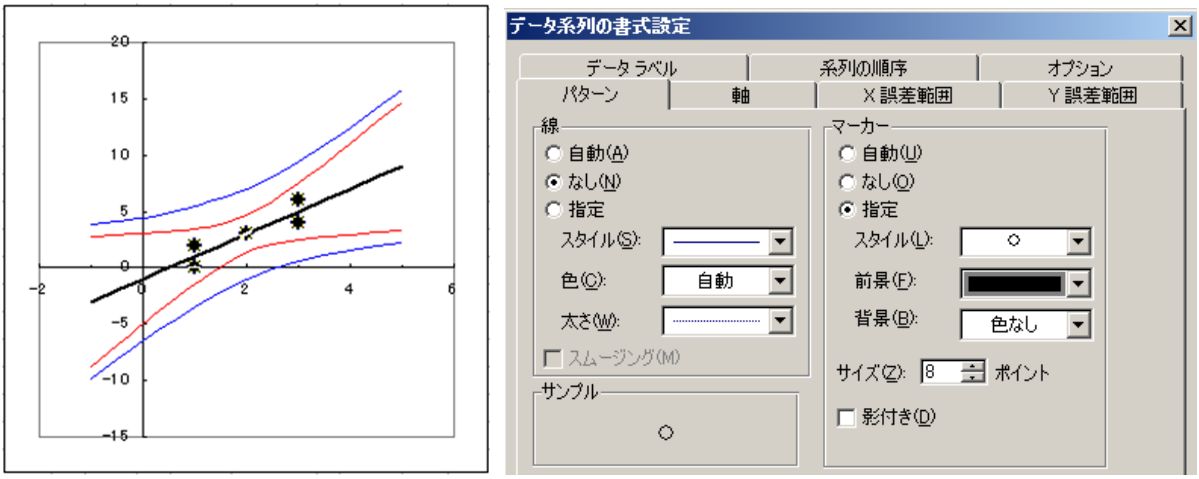


図 2.14 データ系列の書式設定の例示 2

図 2.6 に示した逆推定の 95%信頼区間を散布図に $\hat{x}_{L95}$， $\hat{x}_{U95}$ を上書きするためには【系列】を追加し，次のような(x, y)の 2 点（1.054，3.50），（6.150，3.50）を用いて【データ系列の書式設定／パターン／線】で線とマーカを設定する．

逆推定 y_0 =	3.500	x_0^ =	2.250		
回帰直線の95%CL					y
a =	−0.399	x_{L95}^ =	1.054		3.50
b =	0.444	x_{U95}^ =	6.151		3.50
c =	−0.062				

2.7. 統計ソフトJMPによる検証

Excel での統計計算結果が正しいか，常に他の統計ソフトの結果と照合することが必要である．逆推定を標準的にサポートしている統計ソフト JMP によって，結果の照合をしてみよう．

JMP で計算した回帰パラメータの推定値を次に示す．表 2.1 の Excel の行列計算の結果， と一致し，x のパラメータ $\hat{\beta}_1 = 2$ の p 値も 0.0405 と一致した．

JMP

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	−1	1.264911	−0.79	0.4869
x	2	0.57735	3.46	0.0405*

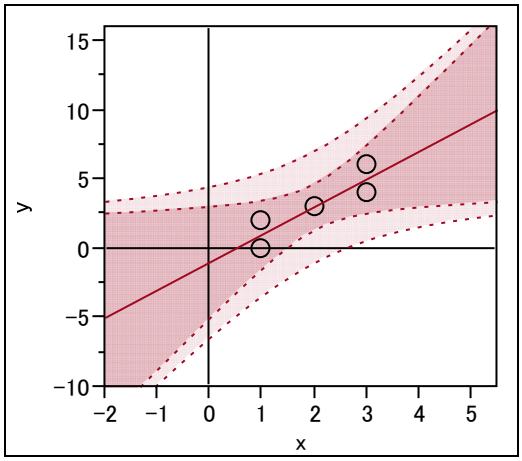
回帰直線の 95%信頼区間を JMP で求めた結果を次に示す．回帰直線の 95%信頼区間は JMP では，平均 y の 95%下側（上側）95%となっていて，Excel で計算した $\hat{y}_{L95}(x=3) = 2.535$，$\hat{y}_{U95}(x=3) = 7.4651$，個別データの 95%信頼区間，$\hat{y}_{L95I}(x=3) = 0.5750$，$\hat{y}_{L95I}(x=3) = 9.4250$ と Excel の行列計算でも一致した．

JMP

	x	y	予測値 y	残差 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	1	0	1	−1	−1.4651	3.4651	−3.4250	5.4250
2	1	2	1	1	−1.4651	3.4651	−3.4250	5.4250
3	2	3	3	0	1.3566	4.6434	−1.0255	7.0255
4	3	4	5	−1	2.5349	7.4651	0.5750	9.4250
5	3	6	5	1	2.5349	7.4651	0.5750	9.4250

JMP で回帰直線の 95%信頼区間，個別データの信頼区間を表示された図を次に示す．Excel で作図したものと一致している．

JMP



Excel

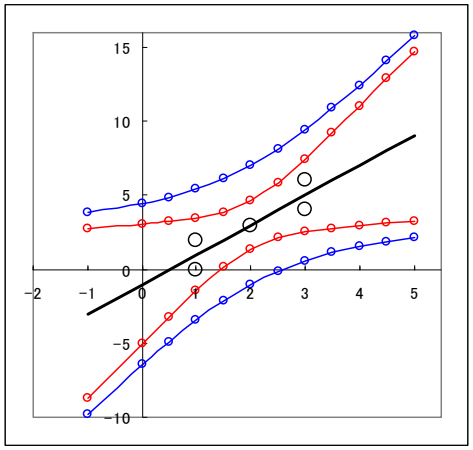


図 2.15 JMP と Excel による 95%信頼区間の図示

個別データの逆推定の 95%信頼区間は，JMP を用いて計算できるようになっていて，Excel での計算結果と一致したことにより，Excel での計算方法が検証されたことになる．

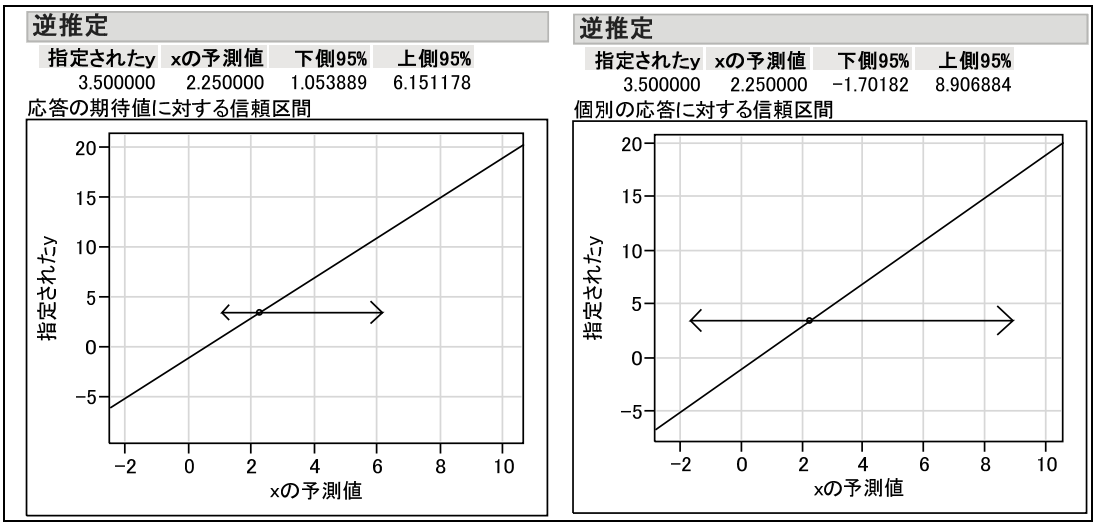


図 2.16　JMP による逆指定

JMP では，回帰直線の 95%信頼区間，個別データの 95%信頼区間の逆推定を計算できるようになっている．次に示すように【応答の期待値に対する信頼区間】から，$\hat{x}_{L95}(y_0 = 3.5) = 1.054$，$\hat{x}_{U95}(y_{L95} = 3.5) = 6.151$ が得られ，【個別の期待値に対する信頼区間】から $\hat{x}_{L95I}(y_{U95} = 3.5) = -1.702$，$\hat{x}_{U95I}(y_{L95} = 3.5) = 8.907$ と Excel の結果と一致した．

逆推定 y_0 =	3.500	x_0^ =	2.250
回帰直線の95%CL			
a =	−0.399	x_{L95R}^ =	1.054
b =	0.444	x_{U95}^ =	6.151
c =	−0.062		

個別データの95%CL			
a' =	0.934	x_{L95I}^ =	−1.702
b =	0.444	x_{U95I}^ =	8.907
c =	−0.062		

図 2.17　Excel による逆推定（再掲）

JMP では，m 回の繰返し測定が行なわれたときの，95%信頼区間を求める機能は提供されていない．繰返し測定された検体の平均値について，その 95%信頼区間については，他に参照できるものが存在しない．このような場合には，理論的な考え方を丁寧に説明し，計算式の導出方法，数値例を示す必要がある．

2.8. 補足

正規方程式

回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ は，$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ で求められると，天下り的に示したので，補足をする．回帰式を，

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\ \ i = 1, 2, \ldots, n \tag{2.37}$$

としたときに，β_0 と β_1 の最小 2 乗推定量は，

$$Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{2.38}$$

を β_0 と β_1 で微分した式を 0 と置くと，式(4.20)が得られる．

$$\begin{aligned} \frac{\partial Q}{\partial \hat{\beta}_0} &= -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \\ \frac{\partial Q}{\partial \hat{\beta}_1} &= -2\sum_{i=1}^{n}\left\{\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i\right\} = 0 \end{aligned} \tag{2.39}$$

である．これらの微分した式を分解すると，

$$\begin{aligned} &\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1\sum_{i=1}^{n} x_i = 0 \\ &\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0\sum_{i=1}^{n} x_i - \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = 0 \end{aligned} \tag{2.40}$$

となり $\hat{\beta}_0$ と $\hat{\beta}_0$ が，含まれない項を右辺に移して整理すると，

$$\begin{aligned} &n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ &\hat{\beta}_0\sum_{i=1}^{n} x_i + \hat{\beta}_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{aligned} \tag{2.41}$$

正規方程式が得られる．ところで，$\boldsymbol{X}^T\boldsymbol{X}$ は，

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \tag{2.42}$$

であり，$\boldsymbol{X}^T\boldsymbol{y}$ は，

$$\boldsymbol{X}^T\boldsymbol{y}=\begin{bmatrix}\sum_{i=1}^{n}y_i\\ \sum_{i=1}^{n}x_iy_i\end{bmatrix} \tag{2.43}$$

なので，$\boldsymbol{X}^T\boldsymbol{X}$ に $\hat{\boldsymbol{\beta}}$ を掛けると

$$(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\beta}=\begin{bmatrix}n & \sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}x_i & \sum_{i=1}^{n}x_i^2\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_0\end{bmatrix}=\begin{bmatrix}n\hat{\beta}_0+\hat{\beta}_1\sum_{i=1}^{n}x_i\\ \hat{\beta}_0\sum_{i=1}^{n}x_i+\hat{\beta}_1\sum_{i=1}^{n}x_i^2\end{bmatrix} \tag{2.44}$$

のように式(2.41)正規方程式の左辺となり，$\boldsymbol{X}^T\boldsymbol{y}$ が右辺となるので

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\boldsymbol{\beta}}=\boldsymbol{X}^T\boldsymbol{y} \tag{2.45}$$

が成り立つ．逆行列 $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ を両辺にかけて

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{2.46}$$

が得られる．そして，得られた $\hat{\boldsymbol{\beta}}$ を用いて，$\boldsymbol{y}$ の推定値は，

$$\hat{\boldsymbol{y}}=\boldsymbol{X}\hat{\boldsymbol{\beta}} \tag{2.47}$$

となる．これを機会に，行列での計算に慣れ親しんでもらいたい．

回帰パラメータの分散

回帰パラメータの分散共分散を $\boldsymbol{\Sigma}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}_e^2$ で与えたのであるが，詳しく展開してみよう．一般的には展開することが面倒ではあるが，2×2 の場合には，行列 $\boldsymbol{A}$ を

$$\boldsymbol{A}=\begin{bmatrix}a & b\\ c & d\end{bmatrix} \tag{2.48}$$

で与えたときに，その逆行列は，次のように $\boldsymbol{A}$ の行列式 $|A|=D=(ad-bc)$ の逆数を行列 $\boldsymbol{A}$ の余因子行列に掛けることによって求められる．

$$\boldsymbol{A}^{-1}=\frac{1}{(ad-bc)}\begin{bmatrix}d & -c\\ -b & a\end{bmatrix} \tag{2.49}$$

デザイン行列の $(\boldsymbol{X}^T\boldsymbol{X})$ は，次のような 2×2 の行列になる．

$$(\boldsymbol{X}^T\boldsymbol{X})=\begin{bmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_n\end{bmatrix}\begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_n\end{bmatrix}=\begin{bmatrix}n & \sum x_i\\ \sum x_i & \sum x_i^2\end{bmatrix} \tag{2.50}$$

$(\boldsymbol{X}^T\boldsymbol{X})$ の行列式 D は，

$$D=n\sum x_i^2-\left(\sum x_i\right)^2=n\sum(x_i-\bar{x})^2 \tag{2.51}$$

なので，逆行列は，

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}=\frac{1}{D}\begin{bmatrix}\sum x_i^2 & -\sum x_i\\ -\sum x_i & n\end{bmatrix} \tag{2.52}$$

となり，回帰パラメータの分散は，逆行列に誤差分散 $\hat{\sigma}_e^2$ を掛けたセルから，

$$Var(\hat{\beta}_0) = \frac{\sum x_i^2}{D}\hat{\sigma}_e^2 = \frac{\sum x_i^2}{n\sum(x_i - \overline{x})^2}\hat{\sigma}_e^2 = \left\{\frac{1}{n} + \frac{\overline{x}^2}{\sum(x_i - \overline{x})^2}\right\}\hat{\sigma}_e^2 \tag{2.53}$$

$$Var(\hat{\beta}_1) = \frac{n}{D}\hat{\sigma}_e^2 = \frac{1}{\sum(x_i - \overline{x})^2}\hat{\sigma}_e^2 \tag{2.54}$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sum x_i}{D}\hat{\sigma}_e^2 = -\frac{\overline{x}}{\sum(x_i - \overline{x})^2}\hat{\sigma}_e^2 \tag{2.55}$$

が求められる．

回帰直線の 95%信頼区間の幅が最小になる点

回帰式の推定値の分散は，ある x_0 に対して，次のようになる．

$$\begin{aligned} Var(\hat{y}_0) &= Var(\hat{\beta}_0) + 2x_0 Cov(\hat{\beta}_0, \hat{\beta}_1) + x_0^2 Var(\hat{\beta}_1) \\ &= \frac{\sum x_i^2 - 2x_0\sum x_i + nx_0^2}{D}\hat{\sigma}_e^2 \end{aligned} \tag{2.56}$$

回帰式の 95%信頼区間の幅が最小になる x は，分散の分母を x_0 について偏微分し，0 おいて x_0 について解くことによって求められる．

$$\frac{\partial\left(\sum x_i^2 - 2x_0\sum x_i + nx_0^2\right)}{\partial x_0} = -2\sum x_i + 2nx_0 = 0 \tag{2.57}$$

$$x_0 = \frac{\sum x_i}{n} = \overline{x}. \tag{2.58}$$

のように，算術平均 $\overline{x}$. の点であることがわかる．

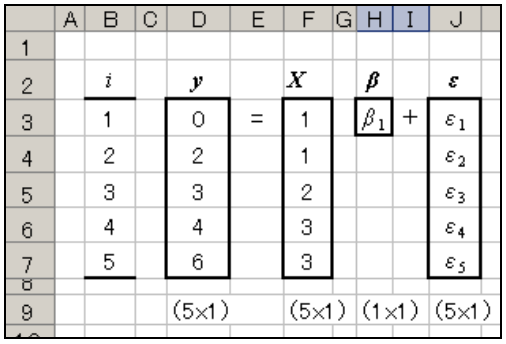
3. 原点を通る回帰式

3.1. 回帰分析の行列による表記

第 2 章で用いた 5 個のデータ $(x, y) = (1, 0), (1, 2), (2, 3), (3, 4), (3, 6)$ について，切片項 β_0 がない次の式に，

$$y_i = \beta_1 x_i + \varepsilon_i \ , \ i = 1, 2, ..., 5 \tag{3.1}$$

最小 2 乗法によって回帰分析を行ない，回帰パラメータ β_1 を推定したい．行列で現すと，2.1 節で示したと同様な一般式 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ となる．Excel シート上で数値を入れて示そう．



	A	B	C	D	E	F	G	H	I	J
1										
2		i		$\boldsymbol{y}$		$\boldsymbol{X}$		$\boldsymbol{\beta}$		$\boldsymbol{\varepsilon}$
3		1		0	=	1		β_1	+	ε_1
4		2		2		1				ε_2
5		3		3		2				ε_3
6		4		4		3				ε_4
7		5		6		3				ε_5
8										
9				(5×1)		(5×1)		(1×1)		(5×1)

図 3.1 Excel シート上での行列表現

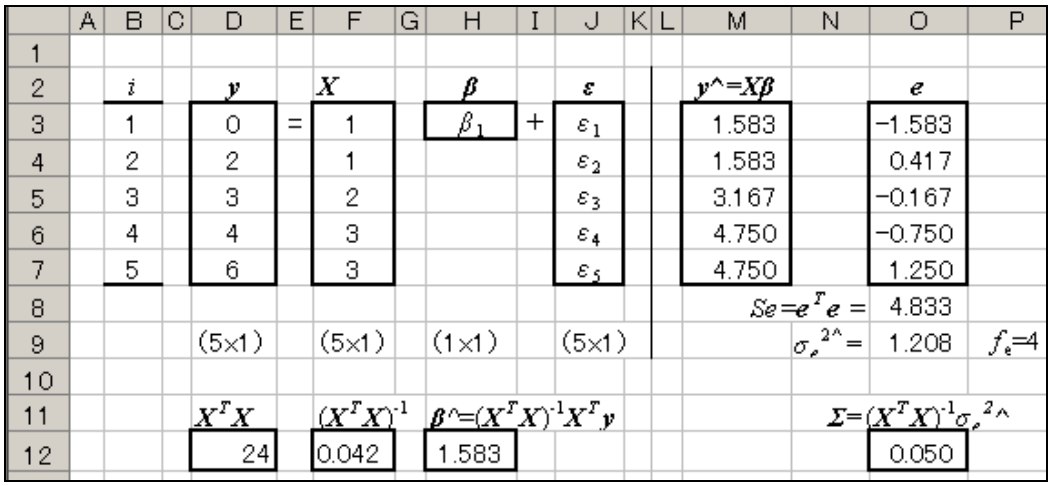
行列 $\boldsymbol{y}$ は大きさ(5×1)，$\boldsymbol{X}$ も(5×1)，$\boldsymbol{\beta}$ は(1×1)，$\boldsymbol{\varepsilon}$ は(5×1)の長方形(矩形)の行列となる．回帰パラメータ $\boldsymbol{\beta}$ の最小 2 乗推定量 $\hat{\boldsymbol{\beta}}$ も 2.1 節で示したと同様に，次式で与えられ，

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} \tag{3.2}$$

傾き $\hat{\beta}_1$ の分散 $Var(\hat{\beta}_1)$ は，(1×1)の分散共分散行列 $\boldsymbol{\Sigma}$ から，

$$\boldsymbol{\Sigma} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \hat{\sigma}_e^2 \tag{3.3}$$

求めることができる．Excel シートで，これらの計算過程を次に示す．



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2		i		$\boldsymbol{y}$		$\boldsymbol{X}$		$\boldsymbol{\beta}$		$\boldsymbol{\varepsilon}$			$\boldsymbol{y}$^=$\boldsymbol{X\beta}$		$\boldsymbol{e}$	
3		1		0	=	1		β_1	+	ε_1			1.583		−1.583	
4		2		2		1				ε_2			1.583		0.417	
5		3		3		2				ε_3			3.167		−0.167	
6		4		4		3				ε_4			4.750		−0.750	
7		5		6		3				ε_5			4.750		1.250	
8														$Se=\boldsymbol{e}^T\boldsymbol{e}=$	4.833	
9				(5×1)		(5×1)		(1×1)		(5×1)				σ_e^{2}^=	1.208	f_e=4
10																
11				$\boldsymbol{X}^T\boldsymbol{X}$		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		$\boldsymbol{\beta}$^=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$							$\boldsymbol{\Sigma}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e^{2}$^	
12				24		0.042		1.583							0.050	

図 3.2 原点を通る回帰直線のパラメータ推定

回帰パラメータに対する推定と検定は，次のようになる．t 統計量に対する p 値は，自由度 $f_e = n - p = 5 - 1 = 4$ の t 分布関数で計算する．

表 3.1 傾きのみの回帰パラメータ

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k$^	$Var(\boldsymbol{\beta}_k$^)	SE	t	p
1	x_1:傾き	β_1	1.583	0.050	0.224	7.056	0.0021

各種の 95%信頼区間の推定

回帰直線の 95%信頼区間は，$\hat{y}_i$ の分散 $Var(\hat{y}_i)$ に自由度が $f_e = n - p = 4$ の t 分布の両側 5%点の確率を掛けて $\hat{y}_i$ にプラス・マイナスした，2.4 節と同様に，

$$\hat{y}_i \text{ の 95\%信頼区間} = \hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)} \tag{3.4}$$

として求められる．$Var(\hat{y}_i)$ は，$\hat{y}_i = \hat{\beta}_1 x_i$ なので，次のように

$$Var(\hat{y}_i) = Var(\hat{\beta}_1 x_i) = x_i^2 Var(\hat{\beta}_1) \tag{3.5}$$

として，求められる．行列（ベクトル）$\boldsymbol{x}_i$ を用いると，

$$Var(\hat{y}_i) = \boldsymbol{x}_i \boldsymbol{\Sigma} \boldsymbol{x}_i^T \tag{3.6}$$

となり，$\boldsymbol{x}_5$ の場合，

$\boldsymbol{x}$	$\boldsymbol{\Sigma}$	$\boldsymbol{x}^T$	=	$Var(y_5$^)
3	0.050	3		0.453

として計算される．個別データについての 95%信頼区間も 2.4 節と同様に，

$$\text{個々のデータの95\%信頼区間} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + \hat{\sigma}_e^2} \tag{3.7}$$

複数データの平均について 95%信頼区間は，$Var(\hat{y}_i)$ に，誤差分散 σ_e^2 / m を加えたものを用いればよい．したがって，その 95%信頼区間は，

$$m\text{個のデータの95\%信頼区間} = \hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i) + \frac{\hat{\sigma}_e^2}{m}} \tag{3.8}$$

となる．

第 2 章で用いた Excel シートの行列計算を一部変更することにより，原点を通る回帰式と関連する計算が行なえる．図 3.4 に示すようにデザイン行列 $\boldsymbol{X}$ の 1 列目を全て 0 とし，(2×2)の $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ は(1×1)に縮小し，$\boldsymbol{\Sigma}$ も同様に(1×1)に，(2×1)の $\hat{\boldsymbol{\beta}}$ は(1×1)に縮小する．

回帰直線の 95%信頼区間は，図 3.4 に示すように原点を通る直線となる．これは，$\hat{y}_i$ の推定値が

$$\sqrt{Var(\hat{y}_i)} = \sqrt{x_i^2 Var(\hat{\beta}_1)} = x_i\sqrt{Var(\hat{\beta}_1)} \tag{3.9}$$

のように，x_i に比例すると仮定したためである．

個別データの 95%信頼区間は，$Var(\hat{y}_i) = x_i^2 Var(\hat{\beta}_1) + \hat{\sigma}_e^2$ なので，x_i が 0 の場合にも誤差分散

$\hat{\sigma}_e^2$ が残るために 0 とはならない.

	$(X^TX)^{-1}$			$t_{0.05}$=	2.7764	f_e= 4						
	0	0		Σ=	0.0000	0.0000						
	0	0.0417			0.0000	0.0503						
				$\sigma_e{}^2$ ^=	1.2083							
										m=	3	
y	X		β^		y^	$Var(y_i\hat{})$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
0	0	1	0	=	1.58	0.050	0.960	2.206	−1.532	4.698	−0.286	3.452
2	0	1	1.583		1.58	0.050	0.960	2.206	−1.532	4.698	−0.286	3.452
3	0	2			3.17	0.201	1.921	4.413	−0.130	6.463	1.009	5.325
4	0	3			4.75	0.453	2.881	6.619	1.171	8.329	2.181	7.319
6	0	3			4.75	0.453	2.881	6.619	1.171	8.329	2.181	7.319

図 3.3 原点を通る回帰直線

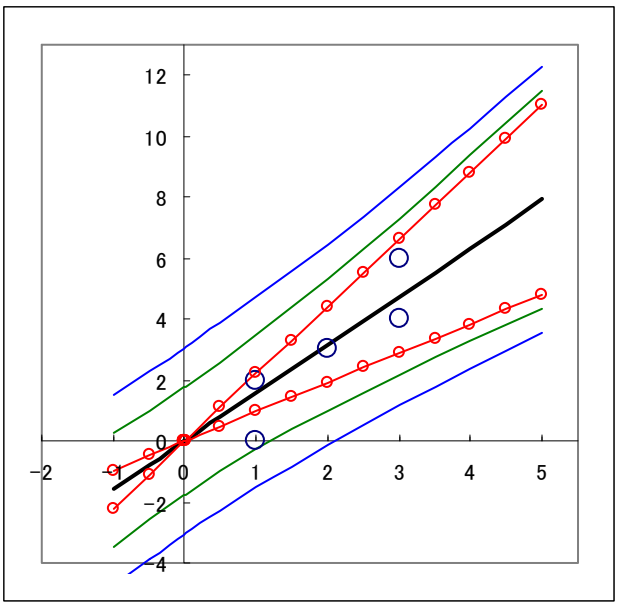


図 3.4 原点を通る回帰直線と各種の 95%信頼区間

3.2. 各種の逆推定

回帰直線

回帰分析における逆推定とは，ある測定値 y_0 が得られたときに，あらかじめ求められた回帰直線上の $\hat{x}_0$ を推定することである．回帰直線が $y_0 = \hat{\beta}_1\hat{x}_0$ の場合に $\hat{x}_0$ について解けば，

$$\hat{x}_0 = \frac{y_0}{\hat{\beta}_1} \tag{3.10}$$

となり，$y_0 = 3.5$ の場合であれば，傾き $\hat{\beta}_1 = 1.5833$ なので，

$$\hat{x}_0 = \frac{y_0}{\hat{\beta}_1} = \frac{3.5}{1.5833} = 2.2106 \tag{3.11}$$

となる．この 95%信頼区間は，2.5 節で用いた 95%信頼区間の式

$$(\hat{\beta}_0 + \hat{\beta}_1\hat{x}_{L95}) + t_\alpha\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1\hat{x}_{L95})} = y_0 \tag{3.12}$$

の，$\hat{\beta}_0$ に関連する項を 0 とすることにより計算できる.

$$
\begin{aligned}
a_o &= Var(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} = 0 - \frac{(y_0 - 0)^2}{t_\alpha^2} \\
&= -\frac{y_0^2}{t_\alpha^2} = -\frac{3.5^2}{2.776^2} = -1.589 \\
b_o &= 2Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2 \times 0 + \frac{2(y_0 - 0)\hat{\beta}_1}{t_\alpha^2} \\
&= \frac{2y_0\hat{\beta}_1}{t_\alpha^2} = \frac{2 \times 3.5 \times 1.583}{2.776^2} = 1.438 \\
c_o &= Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^{\,2}}{t_\alpha^2} = \frac{1.583^2}{2.776^2} = -0.275
\end{aligned}
\tag{3.13}
$$

ただし，$a_o + b_o\hat{x}_{L95} + c_o\hat{x}_{L95}^2 = 0$

最終的には, 2.5 節と同様に，$\hat{x}_{L95}$ に関する 2 次式を解くと次の結果が得られる.

回帰直線の 95%信頼区間=(1.586,3.645)

個別データ

個別データに関しては, 次に示すように $\hat{x}_{L95}$ の項が含まれないので, a_o の項のみが a_o' となり, b_o および c_o は，回帰式の 95%信頼区間の場合と同じである.

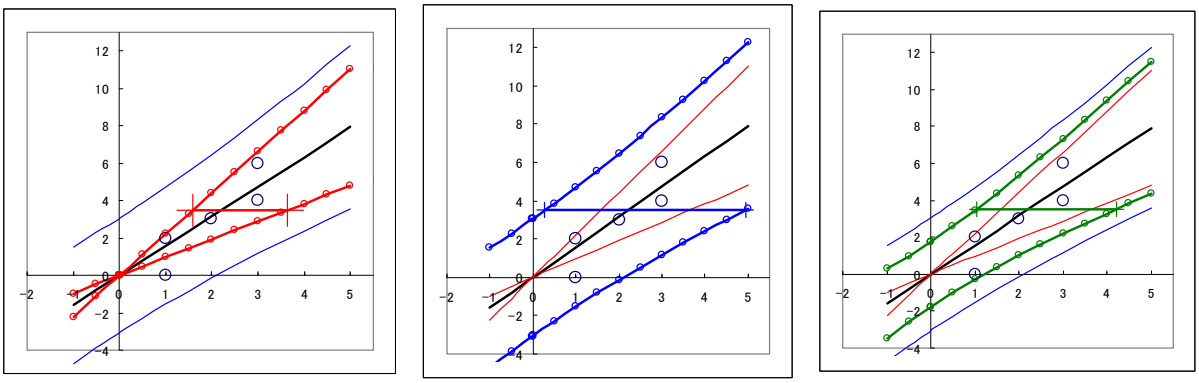
$$
\begin{aligned}
&(\hat{\beta}_1\hat{x}_{L95}) + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95}) + \hat{\sigma}_e^2} = y_0 \\
&a_o' = a_o + \hat{\sigma}_e^2 \\
&\quad = -1.589 + 1.208 = -0.381
\end{aligned}
\tag{3.14}
$$

複数データの平均値

複数データに関しても，個別データと同様に $\hat{x}_{L95}$ の項が含まれないので，a_o の項のみが a'' となる.

$$
\begin{aligned}
&(\hat{\beta}_1\hat{x}_{L95}) + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95}) + \frac{\hat{\sigma}_e^2}{m}} = y_0 \\
&a_o'' = a_o + \frac{\hat{\sigma}_e^2}{m} \\
&\quad = -1.589 + \frac{1.208}{3} = -1.186
\end{aligned}
\tag{3.15}
$$

これらの逆推定結果を散布図上に落とし込む.



逆推定 y_0 = 3.500　x_0^ = 2.211

回帰直線の95%CL				個別データの95%CL				m 個の平均の95%CL			
a_o =	−1.589	x_{L95R}^ =	1.586	a'_o =	−0.381	x_{L95I}^ =	0.280	a''_o =	−1.186	x_{L95m}^ =	1.027
b_o =	1.438	x_{U95R}^ =	3.645	b_o =	1.438	x_{U95I}^ =	4.951	b_o =	1.438	x_{U95m}^ =	4.204
c_o =	−0.275			c_o =	−0.275			c_o =	−0.275		

図 3.5　原点を通る回帰直線の各種の逆推定

3.3. 統計ソフトJMPによる検証

統計ソフトの回帰分析は，切片をディフォルトで含めていて，JMP も同様である．そのために【モデルのあてはめ】で【切片を含めない】との明示的な指定を行い，次の結果を得る．Excel の表 3.1 および図 3.4 の結果と一致してた．

JMP

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
x	1.5833333	0.224382	7.06	0.0021*

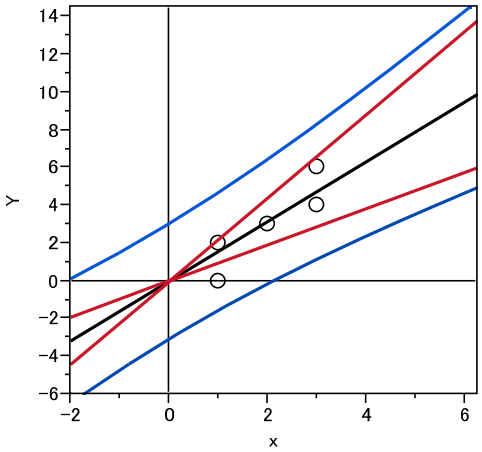
回帰直線の 95%信頼区間，個別データの信頼区間共に Excel の結果と一致した．

JMP

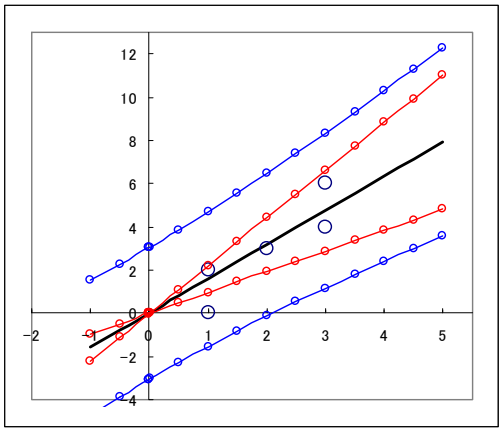
	x	y	予測式 y	平均 yの下側95%	平均 yの上側95%	個別 yの下側95%	個別 yの上側95%
1	1	0	1.583	0.960	2.206	-1.532	4.698
2	1	2	1.583	0.960	2.206	-1.532	4.698
3	2	3	3.167	1.921	4.413	-0.130	6.463
4	3	4	4.750	2.881	6.619	1.171	8.329
5	3	6	4.750	2.881	6.619	1.171	8.329

JMP の【重ねあわせプロット】により，回帰式の 95%信頼区間，個別データの 95%信頼区間を上書きした結果を次に示す．Excel で作成した図と一致している．

JMP



Excel



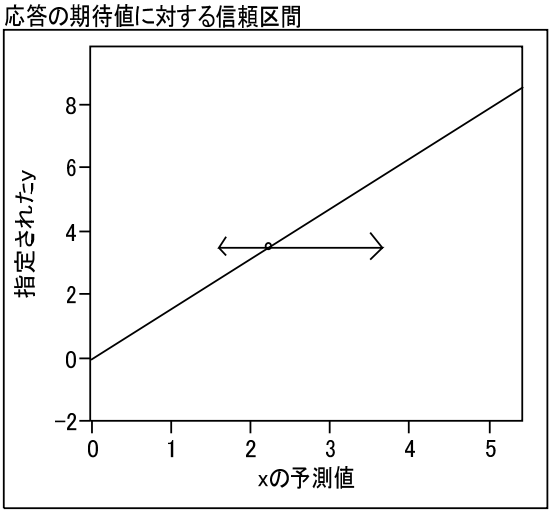
逆推定の結果も次に示すように，$y_0 = 3.5$ に対する逆推定値は，$x_0 = 2.211$，回帰直線の 95%信頼区間は(1.586, 3.645)，個別データでは(0.280, 4.951)と図 3.5 の結果と一致した.

JMP

逆推定

指定されたy	xの予測値	下側95%	上側95%
3.500000	2.210526	1.586354	3.644507

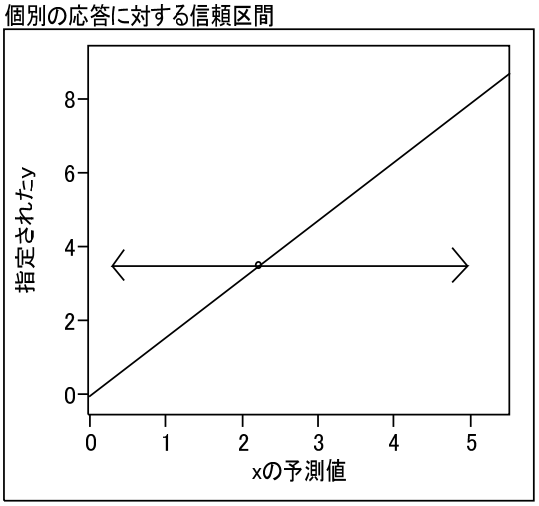
応答の期待値に対する信頼区間



逆推定

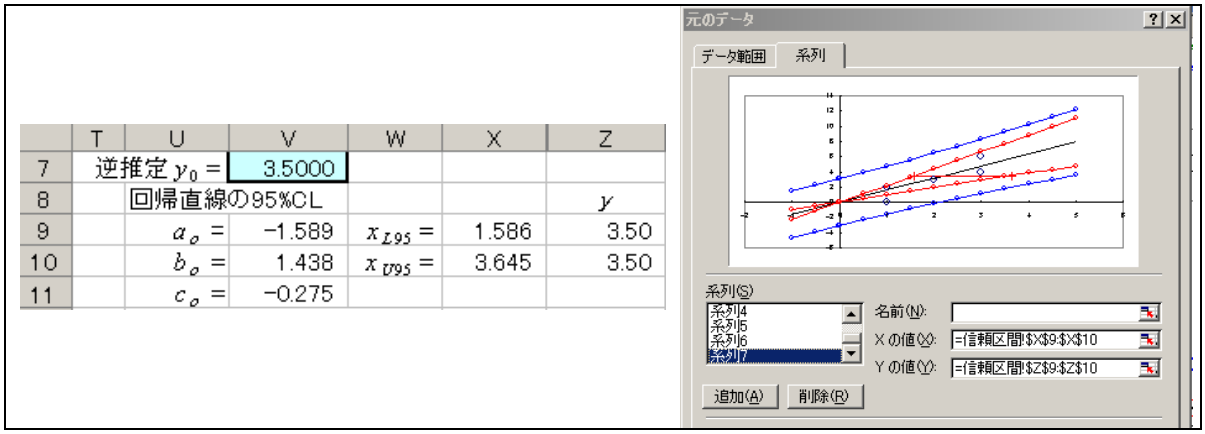
指定されたy	xの予測値	下側95%	上側95%
3.500000	2.210526	0.2798149	4.951046

個別の応答に対する信頼区間



3.4. Excel による逆推定の95%信頼幅の作図法

Excel で 95%信頼区間の作図法については，2.6 節で詳しく述べたが，逆推定の 95%信頼区間の作図法については割愛したので，この章の事例で作図法について解説する．



	T	U	V	W	X	Z
7	逆推定 y_0 =		3.5000			
8		回帰直線の95%CL				y
9		a_o =	−1.589	x_{L95} =	1.586	3.50
10		b_o =	1.438	x_{U95} =	3.645	3.50
11		c_o =	−0.275			

図 3.6　Excel シート上での逆推定の計算結果と【系列】での設定

Excel シート上で，X 軸および Y 軸の 2 つの座標を並列させ，【系列】で新たな系列を【追加】し，それぞれの範囲を設定する．

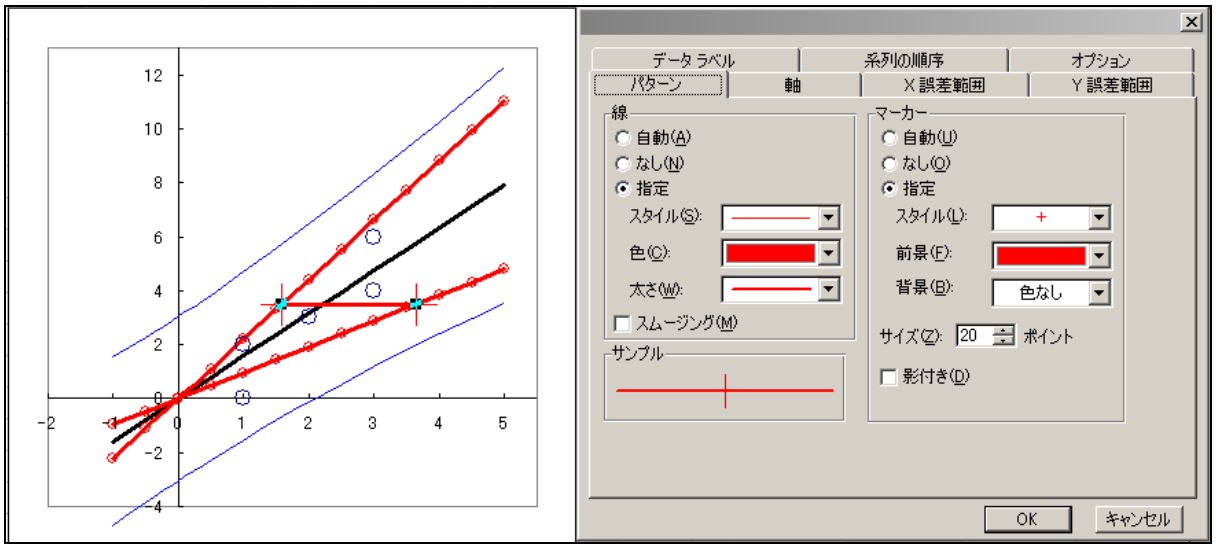


図 3.7　データ系列の書式設定による逆推定の 95%信頼区間の表示

【データ系列の書式設定／パターン】で線の色などの設定，【マーカ】で範囲のスタイルを設定する．

3.5. 補足

正規方程式

回帰式を,

$$y_i = \beta_1 x_i + \varepsilon_i \ , \ i = 1, 2, ..., n \tag{3.16}$$

としたときに，β_1 の最小 2 乗推定量は,

$$Q = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)^2 \tag{3.17}$$

となり，$\hat{\beta}_1$ で微分した式を 0 と置くと，次式が得られる.

$$\frac{\partial Q}{\partial \beta_1} = -2 \sum_{i=1}^{n} \left\{ \left(y_i - \hat{\beta}_1 x_i \right) x_i \right\} = 0 \tag{3.18}$$

である. これを展開し,

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{3.19}$$

整理すると,

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{3.20}$$

となる. ところで，$\boldsymbol{X}^T \boldsymbol{X}$ は,

$$\boldsymbol{X}^T \boldsymbol{X} = \left[\sum_{i=1}^{n} x_i^2 \right] \tag{3.21}$$

であり，$\boldsymbol{X}^T \boldsymbol{y}$ は,

$$\boldsymbol{X}^T \boldsymbol{y} = \left[\sum_{i=1}^{n} x_i y_i \right] \tag{3.22}$$

であるで，$\boldsymbol{X}^T \boldsymbol{X}$ に $\hat{\boldsymbol{\beta}}$ を掛けると正規方程式の左辺となり，$\boldsymbol{X}^T \boldsymbol{y}$ が右辺となるので行列表記で

$$(\boldsymbol{X}^T \boldsymbol{X}) \hat{\boldsymbol{\beta}} = \boldsymbol{X}^T \boldsymbol{y} \tag{3.23}$$

となる. 逆行列 $(\boldsymbol{X}^T \boldsymbol{X})^{-1}$ を両辺にかけて

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y} \tag{3.24}$$

が得られる. そして，得られた $\hat{\boldsymbol{\beta}}$ を用いて，$\boldsymbol{y}$ の推定値は,

$$\hat{\boldsymbol{y}} = \boldsymbol{X} \hat{\boldsymbol{\beta}} \tag{3.25}$$

となる.

4. べき乗を含む回帰分析

4.1. 2次式のあてはめ

回帰直線(1 次式)のあてはまりが悪い場合に, 2 次式

$$y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon_i\,,\ i = 1,2,\ldots,5 \tag{4.1}$$

を用いて曲線をあてはめることが可能なのか検討したくなる. これは, 2 乗の項が入っても線形回帰(重回帰)で定式化でき, 手軽に計算できることによる. 式(4.1)を

$$y_i = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i\,,\ \varepsilon_i \sim N(0,\sigma^2)\,,\quad i = 1,2,\ldots,5 \tag{4.2}$$

$$y_i = \sum_{k=0}^{2} \beta_k x_{ki} + \varepsilon_i\,,\ i = 1,2,\ldots,5 \tag{4.3}$$

のように書き直す. ただし, 全ての i について $x_{0i} = 1$, $x_{1i} = x$, $x_{2i} = x^2$ とする. このように定式化することにより, これまでと同じ形式の行列で現すことができ,

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{4.4}$$

$\hat{\boldsymbol{\beta}}$ を求める方法についても, これまでと全く同一である.

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{4.5}$$

Excel での計算は, デザイン行列 $\boldsymbol{X}$ の大きさが変わるので, 関連する行列計算のサイズを変更する. なお, データは $y_4 = 8$, $y_5 = 10$ のように 2 次式のあてはめが妥当となるように変更してある.

i	x	$\boldsymbol{y}$	=	$\boldsymbol{X}$	x	x^2	$\boldsymbol{\beta}$	+	$\boldsymbol{\varepsilon}$	$\boldsymbol{y}$^=$\boldsymbol{X\beta}$	$\boldsymbol{e}$
1	1	0		1	1	1	β_0		ε_1	1	−1
2	1	2		1	1	1	β_1		ε_2	1	1
3	2	3		1	2	4	β_2		ε_3	3	−0
4	3	8		1	3	9			ε_4	9	−1
5	3	10		1	3	9			ε_5	9	1
		(5x1)		(5x3)			(3x1)		(5x1)	$S_e = \boldsymbol{e}^T\boldsymbol{e} =$	4
										$\sigma_e^{2\wedge} =$	2

$\boldsymbol{X}^T\boldsymbol{X}$			$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$			$\boldsymbol{\beta}$^=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$	$\boldsymbol{\Sigma}$=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e^{2\wedge}$		
5	10	24	14	−16.5	4	3	28	−33	8
10	24	64	−16.5	20.25	−5	−4	−33	40.5	−10
24	64	180	4	−5	1.25	2	8	−10	2.5

図 4.1 2 次式の Excel の行列関数による回帰分析

回帰式の 95%信頼区間

回帰曲線および 95%信頼区間の計算は, 分散を $Var(\hat{y}_5) = \boldsymbol{x}_5\boldsymbol{\Sigma}\boldsymbol{x}_5^T$ として次のように計算する. この式は, これまでの回帰直線の式と同じで, ただ行列の大きさが異なるだけで, Excel での行列計算式も同じである.

$\boldsymbol{x}_5$	$\Sigma=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e{}^2$ ^			$\boldsymbol{x}_5{}^T$	=	$Var(y_5\hat{})$
1 3 9	28	−33	8	1		1
	−33	40.5	−10	3		
	8	−10	2.5	9		
(1x3)	(3x3)			(3x1)		(1x1)

図 4.2　2 次式の回帰直線上の分散の計算例

分散が計算できれば，これまでと同様に回帰直線の 95%信頼区間は $\hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)}$，個別データ：$\hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)+\hat{\sigma}_e^2}$，繰返データ：$\hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)+\hat{\sigma}_e^2/m}$ で計算できる．ただしパラメータ数が 3 となるので，自由度 2 の $t_{0.05}=4.3027$ を用いる．繰返が $m=3$ とした場合，$\hat{y}_5 \pm t_{0.05}\sqrt{Var(\hat{y}_5)+\hat{\sigma}_e^2/m}=9\pm 4.3027\sqrt{1.0+2.0/3}=(3.445,14.555)$ となる．信頼区間の分散には x の 4 乗の項が入っているので，両サイドの発散が著しい．

i	y	$\boldsymbol{X}$	x	x^2	β ^	=	y ^	$Var(y_i\hat{})$	L95%回帰	U95%回帰	L95%個別	U95%個別	L95%m	U95%m
													$m=$	3
1	0	1	1	1	3		1	1.000	−3.303	5.303	−6.452	8.452	−4.555	6.555
2	2	1	1	1	−4		1	1.000	−3.303	5.303	−6.452	8.452	−4.555	6.555
3	3	1	2	4	2		3	2.000	−3.085	9.085	−5.605	11.605	−4.026	10.026
4	8	1	3	9			9	1.000	4.697	13.303	1.548	16.452	3.445	14.555
5	10	1	3	9			9	1.000	4.697	13.303	1.548	16.452	**3.445**	**14.555**

図 4.3　各種の 95%信頼区間の計算

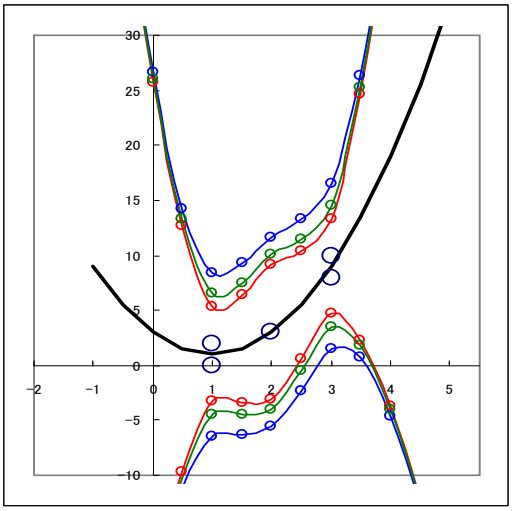


図 4.4　2 次式のあてはめに対する各種の 95%信頼区間

逆推定

検量線は，単調増加することが望ましいのであるが，2 次式を適用した場合には，回帰曲線の 95%信頼区間は，図 4.2 の計算例から分かるように，分散の計算に x^2 の項を含んでいるので，その 2 次形式の結果は，次式のように 4 次式となり，それを解くのは煩雑で解の求め方を定式化することは困難である．

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} + \hat{\beta}_2 \hat{x}_{L95}^2) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} + \hat{\beta}_2 \hat{x}_{L95}^2)} = y_0 \tag{4.6}$$

$$Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95} + \hat{\beta}_2 \hat{x}_{L95}^2) - \frac{(y_0 - \hat{\beta}_0 - \hat{\beta}_1 \hat{x}_{L95} - \hat{\beta}_2 \hat{x}_{L95}^2)^2}{t_\alpha^2} = 0 \tag{4.7}$$

したがって，一般的に単調増加となる検量線の回帰式として，2 次式の使用は，全く薦められない．次節に示すように，2 次式の 1 次の項を除くことにより，x^2 に関して回帰直線をあてはめることを薦める．

4.2. 2次式から1次の項を除いたモデル

回帰直線（1 次式）のあてはまりが悪い場合に，2 次式の 1 次の項を除くことにより，単調に増加する曲線をあてはめることが可能となる．

$$y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i,\ \varepsilon_i \sim N(0, \sigma^2),\ i = 1,2,...,5 \tag{4.8}$$

これは，X 軸の目盛りを x^2 とした場合には，2 次曲線ではなく，x^2 に関して直線のあてはめとなり，x^2 に関して逆推定も可能となる．X 軸の目盛りを x とした場合にした場合に見かけ上，2 次曲線となるが，その場合の逆推定は，X 軸の目盛りを x^2 とした場合で求めた範囲について平方根を取ればよい．もちろん範囲が負となっている場合は，計算不能であり便宜的に 0 とみなすことになる．

図 4.5 に示すように，デザイン行列 $\boldsymbol{X}$ に関してみれば，x^2 を計算した結果となってはいるが数値上は，x^2 に関して一般的な回帰式となっていて，次に示すように第 2 章で示した計算手順とまったく同じとなる．

i	x	$\boldsymbol{y}$		$\boldsymbol{X}$	x^2		$\boldsymbol{\beta}$		$\boldsymbol{\varepsilon}$		$\boldsymbol{y}$^=$\boldsymbol{X\beta}$	$\boldsymbol{e}$
1	1	0	=	1	1		β_0	+	ε_1		0.753	−0.753
2	1	2		1	1		β_1		ε_2		0.753	1.247
3	2	3		1	4				ε_3		3.790	−0.790
4	3	8		1	9				ε_4		8.852	−0.852
5	3	10		1	9				ε_5		8.852	1.148
											$Se = \boldsymbol{e}^T\boldsymbol{e} =$	4.790
		(5x1)		(5x2)			(2x1)		(5x1)		$\sigma_e^{2\wedge} =$	1.597

$\boldsymbol{X}^T\boldsymbol{X}$		$(\boldsymbol{X}^T\boldsymbol{X})^{-1}$		$\boldsymbol{\beta}$^=$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$	$\boldsymbol{\Sigma}$	
5	24	0.556	−0.074	−0.259	0.887	−0.118
24	180	−0.074	0.015	1.012	−0.118	0.025

図 4.5　x^2 に関しての回帰式の行列計算

各種の 95%信頼区間

回帰式および 95%信頼区間の計算は，x^2 に関して行う．各種の 95%信頼区間は，Excel 上では行列 $Var(\hat{y}_i) = Var(\boldsymbol{x}_i \hat{\boldsymbol{\beta}}) = \boldsymbol{x}_i \boldsymbol{\Sigma} \boldsymbol{x}_i^T$ で計算するが，これを展開すれば，

$$Var(\hat{y}_i) = Var(\beta_0 + \beta_1 x_i^2) = Var(\beta_0) + 2(x_i^2)Cov(\beta_0, \beta_1) + (x_i^2)^2 Var(\beta_1) \tag{4.9}$$

となる. 回帰直線の 95%信頼区間：$\hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)}$，個別データ：$\hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)+\hat{\sigma}_e^2}$，繰返データ：$\hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)+\hat{\sigma}_e^2/m}$ と同じ計算方式を適用する．ただしパラメータ数は 2 となるので，自由度 $n-2$ の t 分布を使う．

$n=$ 5, $p=$ 2, $x^{\lambda}=$ 2

$(X^TX)^{-1}$: 0.5556 −0.0741 / −0.0741 0.0154

$t_{0.05}=$ 3.1824　$f_e=$ 3

$\Sigma=$ 0.8871 −0.1183 / −0.1183 0.0246

$\sigma_e{}^2\hat{}=$ 1.5967

$m=$ 3

i	x	y	X		$\hat{\beta}$		$y\hat{}$	$Var(y_i\hat{})$	L95%回帰	U95%回帰	L95%個別	U95%個別	L95%m	U95%m
1	1	0	1	1	−0.259	=	0.753	0.675	−1.862	3.368	−4.044	5.550	−2.744	4.250
2	1	2	1	1	1.012		0.753	0.675	−1.862	3.368	−4.044	5.550	−2.744	4.250
3	2	3	1	4			3.790	0.335	1.948	5.632	−0.633	8.213	0.826	6.754
4	3	8	1	9			8.852	0.754	6.088	11.615	3.973	13.731	5.243	12.461
5	3	10	1	9			8.852	0.754	6.088	11.615	3.973	13.731	5.243	12.461

図 4.6

第 5 番目のデータについて回帰直線の 95%信頼区間を計算してみよう．

$$Var(\hat{y}_5)=\begin{bmatrix}1 & 9\end{bmatrix}\begin{bmatrix}0.887 & -0.118\\ 0.118 & 0.025\end{bmatrix}\begin{bmatrix}1\\ 9\end{bmatrix}=0.754$$

$$\hat{y}_5 \pm t_{0.05}\sqrt{Var(\hat{y}_5)}=8.852\pm 3.182\sqrt{0.754}$$
$$=(6.088,\ 11.615\ )$$

図 4.6 の右は, X 軸に x を取ったもので，y 切片が −0.259 の単調な増加曲線となっているが, 左の X 軸を x^2 とした場合には, 直線となっている. 逆推定は, X 軸を x^2 とした場合について定式化し, その範囲について平方根を取って, x に関する信頼区間とする.

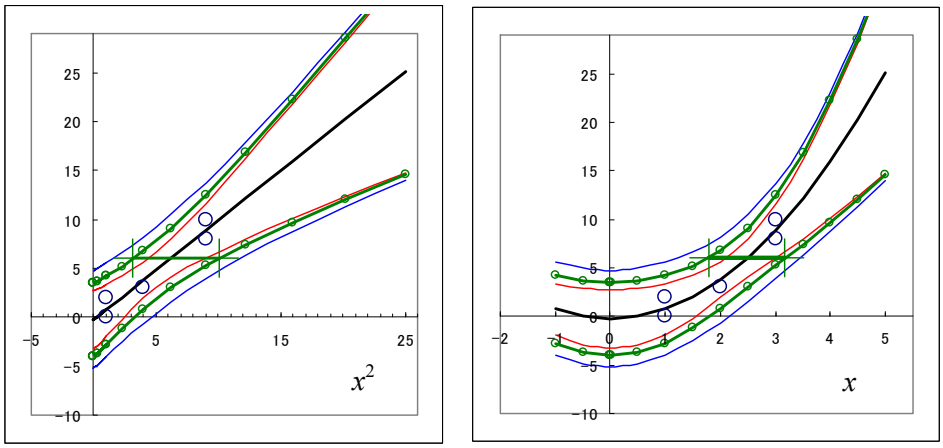


図 4.7　各種 95%信頼区間と繰返データの 3 個の平均についての逆推定

逆推定も，x^2 について第 2 章で示した 2 次式の根の公式を用いて計算し, 計算された 95%信頼区間の下限と上限について次のように平方根を計算する. 繰返 $m=3$ の平均が $y_0=6.0$ とした場合の $\hat{x}_{L95}^2$ および $\hat{x}_{L95}$ は, 次のように計算される.

$$a'' = Var(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} + \frac{\hat{\sigma}_e^2}{m} = 0.887 - \frac{(6-(-0.259))^2}{3.182^2} + \frac{1.597}{3} = -2.449$$

$$b = 2Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2(y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2\times(-0.118) + \frac{2\times(6-(-0.259))\times 1.1012}{3.182^2} = 1.015$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^{\,2}}{t_\alpha^2} = 0.0246 - \frac{1.012^2}{3.182^2} = -0.0765$$

$$(\hat{x}_{L95}^2, \hat{x}_{U95}^2) = \frac{1.015 \pm \sqrt{1.015^2 - 4\times(-2.449)\times(-0.0765)}}{2\times(-0.0765)} = (3.173, 10.083)$$

$$(\hat{x}_{L95}, \hat{x}_{U95}) = (\sqrt{3.173}, \sqrt{10.083}) = (1.781,\ 3.175)$$

逆推定 y_0 =	6.000			
回帰直線の95%CL			x^2に関して	xに関して
a =	−2.981	x_{L95}=	4.395	2.096
b =	1.015	x_{U95}=	8.861	2.977
c =	−0.0765			
個々のデータの95%CL				
a' =	−1.385	x_{L95I}=	1.544	1.243
b =	1.015	x_{U95I}=	11.712	3.422
c =	−0.0765			
m個の平均の95%CL				
a'' =	−2.449	x_{L95m} =	3.173	1.781
b =	1.015	x_{U95m} =	10.083	3.175
c =	−0.0765			

図 4.8　Excel による各種の逆推定の結果

この方法は，x^2 だけでなく，x^λ のように任意のべき乗にしても適用可能である．詳しくは，8.1 節を参照のこと．

5. 分散が均一とはみなせない場合

5.1. 回帰分析の前提条件

これまで示してきた回帰分析を適用するための前提条件は，回帰式をあてはめて 95%信頼区間などの推定を行なう場合に，

1）誤差 ε_i の期待値が 0
2）誤差分散が x_i によらず一定（均一）
3）誤差 ε_i と ε_j $(i \neq j)$ が互いに独立
4）誤差 ε_i は正規分布に従う

などの仮定を必要としている．実際のデータで x_i の範囲が広い場合に，これらの条件を満たすことは稀である．また，測定値 y_i が大きくなるに従い，分散も大きくなることもしばしば経験する．回帰直線を引くだけなら，誤差分散が x_i によらず一定とみなせない場合でも，極端な飛び離れ値がなければ回帰直線のあてはめに疑問が生ずることは少ない．

回帰直線の 95%信頼区間，あるいは個別データの 95%信頼区間を描いてみると，x_i が小さくなるにつれ幅が拡大することに気がつく．誤差分散が x_i によらず一定と仮定しているにもかかわらず，個別データの 95%信頼区間はかなり大きい．これは，推定した回帰直線の 95%信頼区間がデータの平均値 $\bar{x}$ を中心に揺れ動く範囲を示しており，回帰の分散に誤差分散 $\hat{\sigma}_e^2$ を一律に加算しているためである．

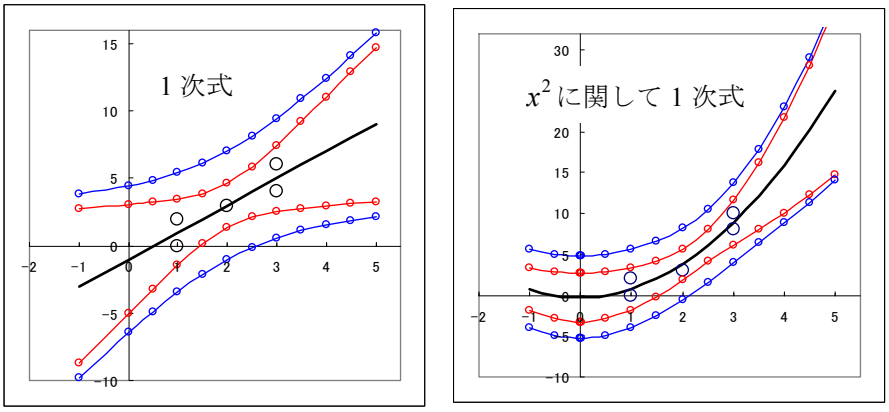


図 5.1　回帰および個別データの 95%信頼区間の例

図 5.2 に示すように，原点を通るとみなした回帰直線の場合は，回帰直線の 95%信頼区間の幅は，x_i が小さくなるにつれ小さくなるが，個別データの場合は広い幅のままである．これは，回帰直線の 95%信頼区間に x_i の大きさに関係なく誤差分散 $\hat{\sigma}_e^2$ を一律に加算しているためである．

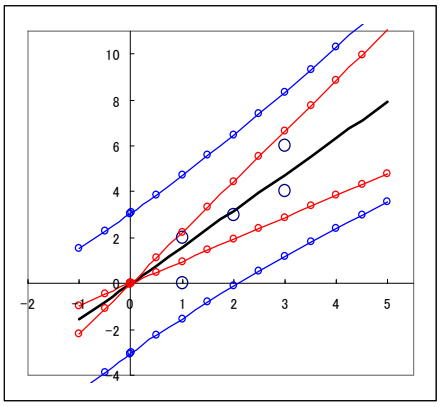


図 5.2　原点を通る回帰および個別データの 95%信頼区間の例

回帰直線をあてはめ x の広い範囲で，ある x_0 を与えた場合の $\hat{y}_0$ を予測し，個別データの 95%信頼区間も評価したい，あるいは，ある y_0 を与えて $\hat{x}_0$ を予測し，個別データの x 方向の 95%信頼区間も評価したい．誤差分散 $\hat{\sigma}_e^2$ が，x_i が小さい場合には小さく，x_i が大きい場合には大きく，均一とはみなせないような場合に，これまで示してきた誤差分散 $\hat{\sigma}_e^2$ が x の大きさに限らず一定と仮定した回帰分析を適用した場合には，適切な 95%信頼区間を算出することができない．

5.2. 分散が x^2 または x に比例する場合

現実に経験するさまざまな事例を Excel の乱数を用いて発生させ，概観してみよう．これは，現実の問題に対して適切な回帰モデルを選択する際の手引きとなることを期待している．また，次章以後で取り上げるさまざまな回帰分析の応用について，回帰パラメータをあらかじめ設定した回帰モデルにに，各種の誤差を載せたときに，現実のどのような散布図になるか外観しておくことが，適切な回帰モデルの選択の助けになる．図 5.3 に Excel で正規乱数を 100 個発生させ，あらかじめ設定した共通の回帰パラメータに対し，各種の分散構造にしたがったデータ y を計算し，散布図としたものである．

注）Excel には正規乱数を発生させる関数がないので，一様乱数発生の Rand 関数を，標準正規分布のパーセント点を求める Normsinv 関数に入れ子にすることにより標準正規乱数の発生ができる．Excel の使い方については，5.5 節に示す．

分散	回帰式	誤差構造
1）分散が一定：	$y_i = 10 + 2.5x_i + \varepsilon_i$，	$\varepsilon_i \sim N(0, 2.5^2)$
2）分散がおおよそ x_i^2 に比例：	$y_i' = 10 + 2.5x_i + \varepsilon_i'$，	$\varepsilon_i' \sim N(0, (4 + x_i^2) \times 0.5^2)$
3）分散がおおよそ x_i に比例：	$y_i = 10 + 2.5x_i + \varepsilon_i''$，	$\varepsilon_i'' \sim N(0, (1 + x_i) \times 1.5^2)$

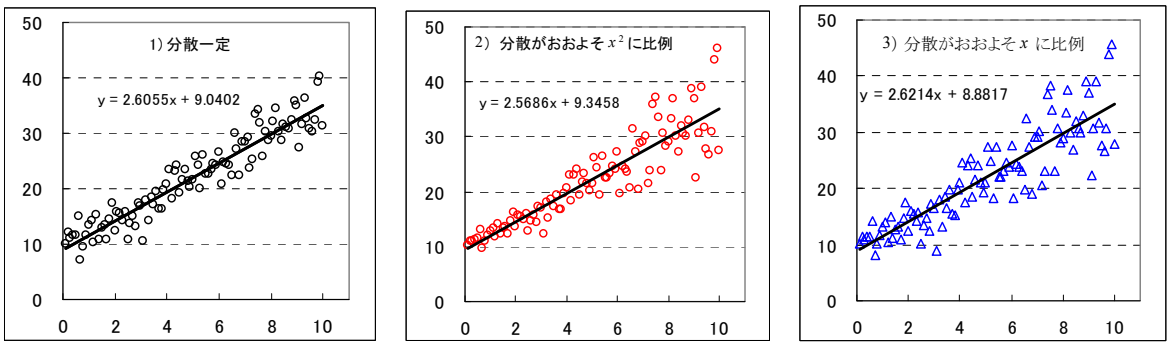


図 5.3　誤差構造が異なるデータに対する回帰直線のあてはめ（1）

1）の散布図は，分散が x_i によらず一定の場合で，x_i は，0.1から10までの範囲である．このような帯状の散布図は現実には稀である．回帰分析が仮定している「2)誤差分散が x_i によらず一定，4)誤差 ε_i は正規分布に従う」を完全に満たしている．もちろん，1)および 3)の条件も満たしている．個別データの 95%信頼区間および残差プロットを次に示す．散布図上のデータは，ほとんどその範囲に含まれている．X 軸を予測値とした残差プロットから回帰分析の 4 条件を全て満たしていることが読み取れる．

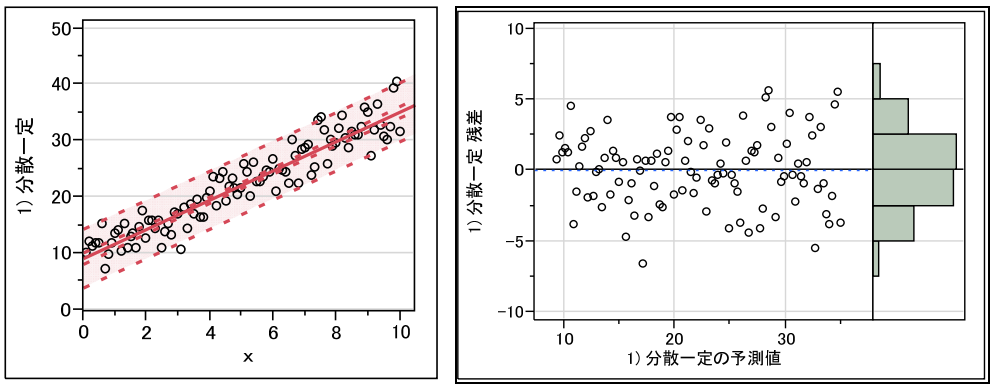


図 5.4　1）分散が一定の場合の残差プロット

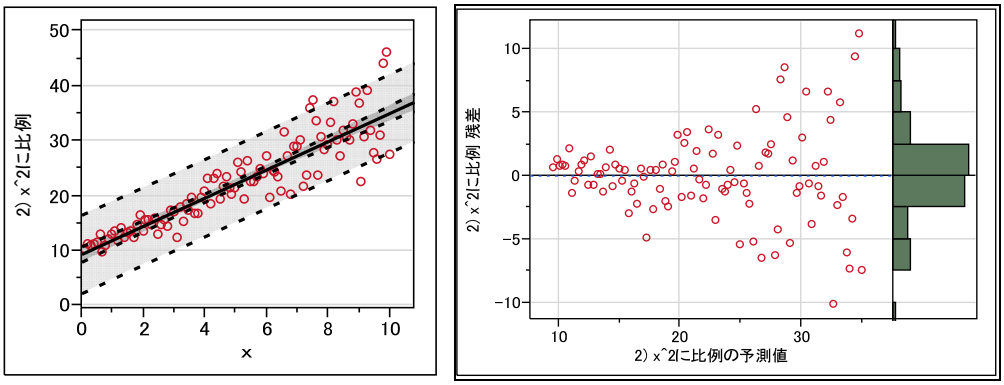


図 5.5　2）分散がおおよそ x^2 に比例する場合の残差プロット

2）の散布図は，x が大きくなるにつれて y の変動も増大している．よく目にする散布図である．このようなデータで学会などでプレゼンをすると，「分散を考慮した回帰分析をしたのか」との質問を

受ける．通常の回帰分析を適用し，残差プロットをみれば扇型となり，何らかの変数変換を考え，分散を一定にして解析する必要性が指摘される．

扇型の残差プロットとなるような場合には，第 6 章で詳しく説明するが，何らかの重み付け回帰を行なう必要がある．図 5.6 に分散が $(4+x^2)$ に比例という重み付き回帰分析による個別データの 95%信頼区間を示す．この 95%信頼区間は，x が大きくなるにつれて広がり，広い範囲で全データを均等に包み込んでいる．

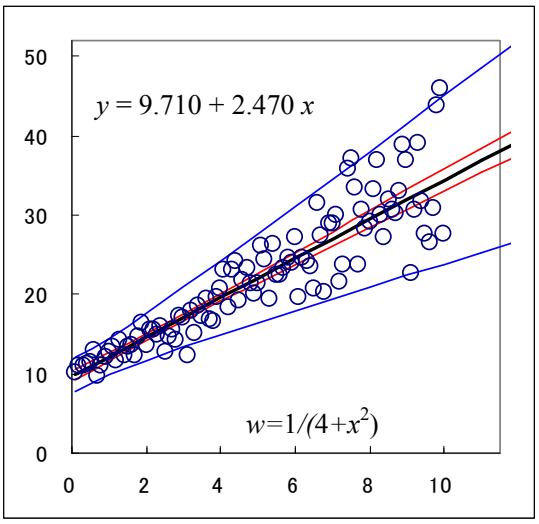


図 5.6　重み付き回帰での個別データの 95%信頼区間の例示

3）の散布図は，2)と同様に見えるが，x_i が小さいときの変動に比べて，x_i が大きくなったときの変動が 2)に比べ相対的に小さい．通常の回帰分析を適用していいように思われるが，x_i が 8～10 あたりの変動の大きさは 1～2 あたりに比べると無視できないくらい大きい．

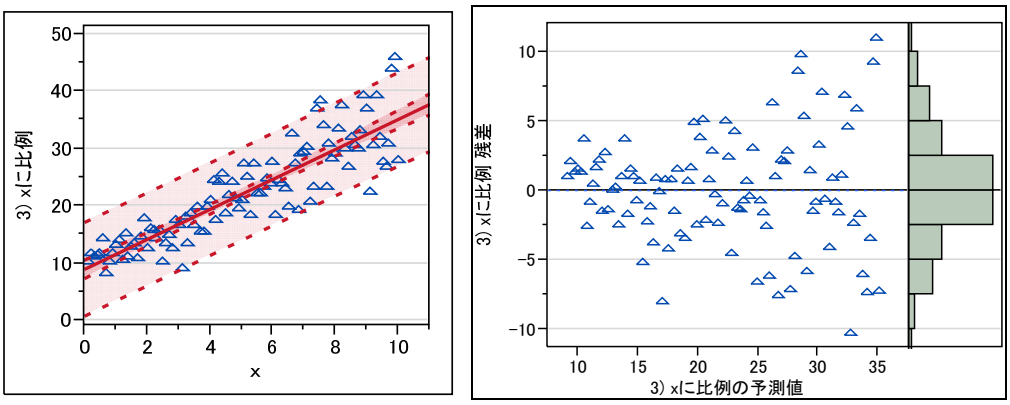


図 5.7　3）分散がおおよそ x に比例する場合の残差プロット

5.3. 分散が x^2 に比例するか x に比例するかの判断

「分散が x^2 に比例するか x に比例するか」との表現に本書は統一しようとしている．しかしながら，実際のデータで，どちらの重みを選択したらよいのかを判断する際には，分散ではなく平方根をとった「標準偏差（SD）が x に比例するか $\sqrt{x}$ に比例するか」かを用いることが，視覚的に優れている．

図 5.5 は，SD が x に比例する場合，図 5.7 は SD が $\sqrt{x}$ に比例する場合であるが，これらの散布図からは，ある x での SD を読み取ることはできない．そこで，$x = 0.5$，1，2，4，6，10 に固定して正規乱数で発生させた 100 個のデータについての散布図を次に示す．

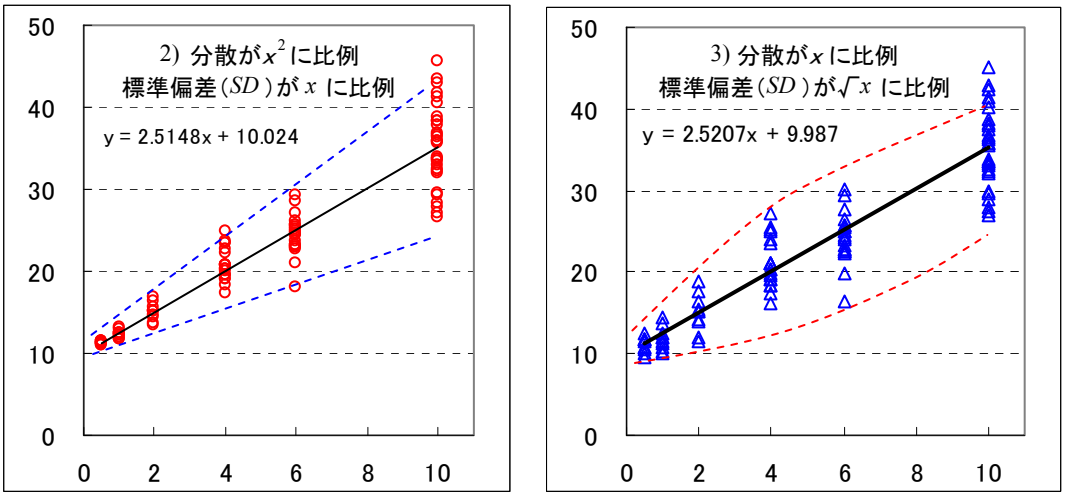


図 5.8　標準偏差（SD）が x に比例するか $\sqrt{x}$ に比例するかの判断

「SD が x に比例する場合」は，おおよそのデータの範囲が，直線的に増加することが観察される．「SD が $\sqrt{x}$ に比例する場合」は，おおよそのデータの範囲が回帰直線に対してルート状の曲線となっている．図 5.3 の散布図についても，データの範囲が直線，または，ルート曲線かあてはめてみて，どちらがデータの範囲を適切に示しているで，おおよその判断をするとよい．

現実のデータでは，データの範囲が小さいために判断が付きにくい場合もある．その場合は，どちらかの重みを選択して重み付き回帰分析を行い，残差プロット上で，データの範囲が直線なのかルート状の曲線なのかで判断する．

データの範囲が，回帰直線に対して 2 次曲線となるようなことは，現実には起こりにくく，そのような場合には，データを大きく発散させるような何らかの原因が潜んでいる．もちろん，分散が x^3 に比例する，さらに分散が x^4 にするような場合に対して，それらの逆数を重みとして，回帰直線，各種の 95%信頼区間を求めて図示することはたやすいのであるが，ある y_0 に対する逆推定 $\hat{x}_0$ をもとめ，個別データの 95%信頼区間を 2 次式の根の公式を使って計算ができなくなってしまう．9.4 節に分散が $x^{4.89}$ に比例する事例を示す．

5.4. 直線をあてはめることが困難な場合

一般的には，得られた散布図のデータに直線を統計的にあてはめることがためらわれる場合に，2 次式をあてはめて，2 次の項が統計的に有意ならば，直線のあてはめに問題があると判断している．検量線の課題に対して直線でなければ，2 次式をあてはめることが選択肢の一つであるが，第 4.1 節で，95%信頼区間の計算式が 4 次式となり，逆推定が一意に定まらないと理由で「推奨できない」とした．それに代わる方法として，2 次式の 1 次項を除いた x^2 に関して 1 次式となる $y=\beta_0+\beta_1 x^2$ を用いることを推奨した．

しかしながら，現実の問題は多様であり，x の 2 乗ではなく，x の λ 乗を用いた $y=\beta_0+\beta_1 x^{\lambda}$ も視野に入れることも示唆した．現実のデータで x の λ を推定する方法は，これまで示してきた線形最小 2 乗法の考え方でなく，非線形最小 2 乗法，あるいは，最尤法を用いる必要性がある．この問題は，Excel のソルバーを使うことにより手軽に推定できる．この方法については，8.2 節で概説する．

ここでは，これまで経験してきたが便宜的な対応をせざるを得なかった問題を定式化するためには，正規乱数を用いて人工的に幾つかの λ 乗を設定し，さらに分散が x の関数として与えられたデータを 100 組発生させた散布図を次に示す．

べき	分散	回帰式	誤差構造
4) $\lambda=0.5$，	$(x_i^{0.5})^2$ に比例：	$y_i=10+9x_i^{0.5}+\varepsilon_i$，	$\varepsilon_i \sim N(0,\,(5+x_i^{0.5\times 2})\times 0.9^2)$
5) $\lambda=1.5$，	$(x_i^{1.5})$ に比例：	$y_i=10+0.8x_i^{1.5}+\varepsilon_i'$，	$\varepsilon_i' \sim N\left(0,\,(6+x_i^{1.5})\times 0.8^2\right)$
6) $\lambda=1.5$，	$(x_i^{1.5})^2$ に比例：	$y_i=10+0.8x_i^{1.5}+\varepsilon_i''$，	$\varepsilon_i'' \sim N(0,\,(100+x_i^{1.5\times 2})\times 0.2^2))$

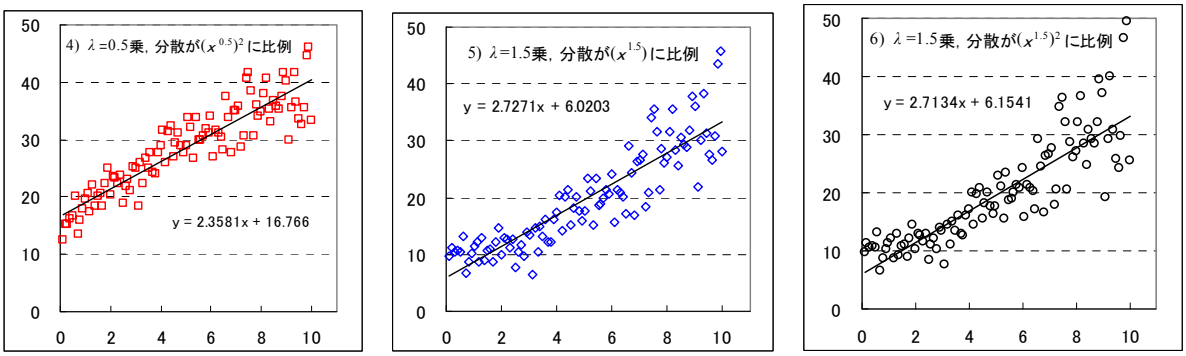


図 5.9　誤差構造が異なるデータに対する回帰直線のあてはめ（2）

4）べきが $\lambda=0.5$ で分散がおおよそ $(x_i^{0.5})^2$ に比例する場合

図 5.10 に通常の回帰分析を適用し，個別データの 95%信頼区間をを示す．回帰直線がうまくあてはまっているようだが，子細に見ると，x が小さい場合は散布図上の点は回帰直線の下側，中ほどは上側，x が大きい場合は上下均一となる．残差プロットでも直線のあてはめが苦しいことが確認できる．

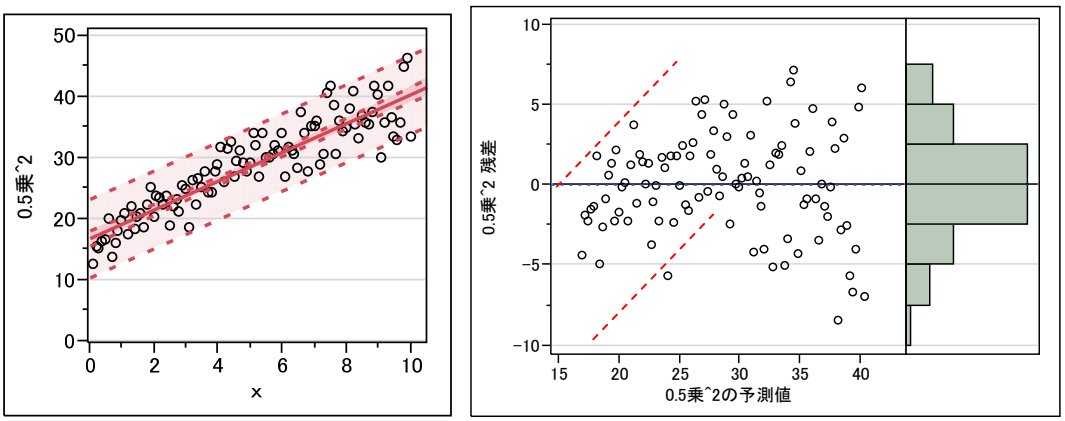


図 5.10　4）上に凸のデータに対する直線のあてはめ

5）べきが $\lambda = 1.5$ で分散がおおよそ $x_i^{1.5}$ に比例

図 5.11 に通常の回帰分析を適用し，個別データの 95%信頼区間を示す．回帰直線がうまくあてはまっているようだが，子細に見ると，x が小さい場合は散布図上の点は回帰直線の上側，中ほどは下側，x が大きい場合は上下均一となる．残差プロットでも直線のあてはめが苦しいことが確認できる．

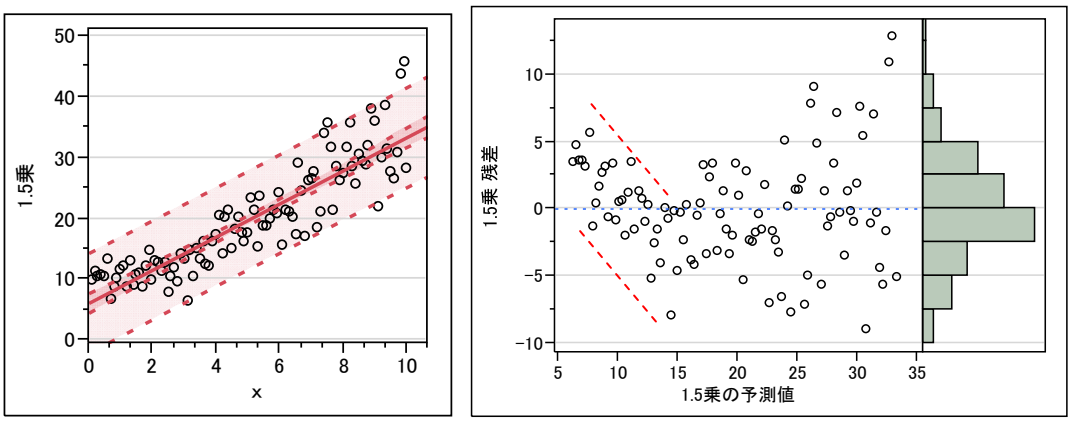


図 5.11　5）下に凹のデータに対する直線のあてはめ 1

6）べきが $\lambda = 1.5$ で分散がおおよそ $(x_i^{1.5})^2$ に比例

図 5.12 は，明らかに下に凹んでいて通常の回帰分析を適用することができないと．はし，個別データの 95%信頼区間を示す．回帰直線がうまくあてはまっているようだが，子細に見ると，x が小さい場合は散布図上の点は回帰直線の上側，中ほどは下側，x が大きい場合は上下均一となる．残差プロットでも直線のあてはめが苦しいことが確認できる．

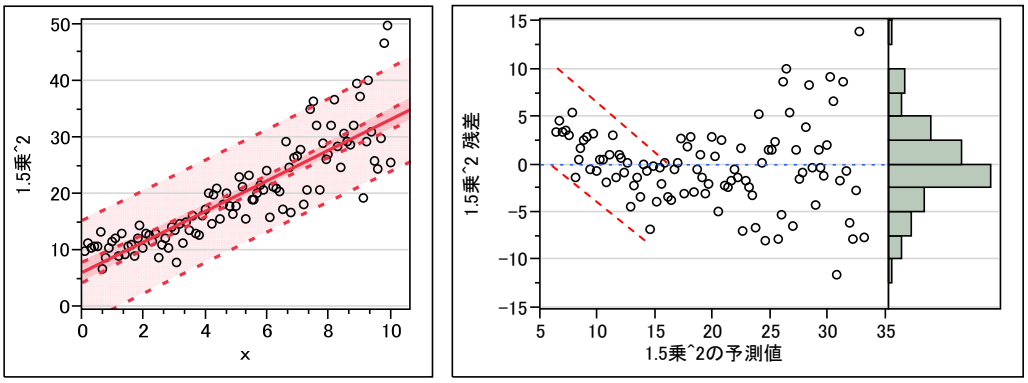


図 5.12　6）下に凹のデータに対する直線のあてはめ 2

下に凹の $x_i^{1.5}$ に関して一次式：$y_i = \beta_0 + \beta_1 x_i^{1.5} + \varepsilon_i$，重み：$w_i = 1/(2^2 + x_i^{1.5\times 2})$ とする回帰分析を行い，個別データの 95%信頼区間を散布図上に上書きしてみよう．重みの分母の定数項 100 は，分散が 0.2^2 となるような正規乱数を用いてデータを発生させたときに設定したものである．これは，$x_i = 0$ の場合の分散を $0.2^2 \times 100 = 2^2$ と見積もったことになる．

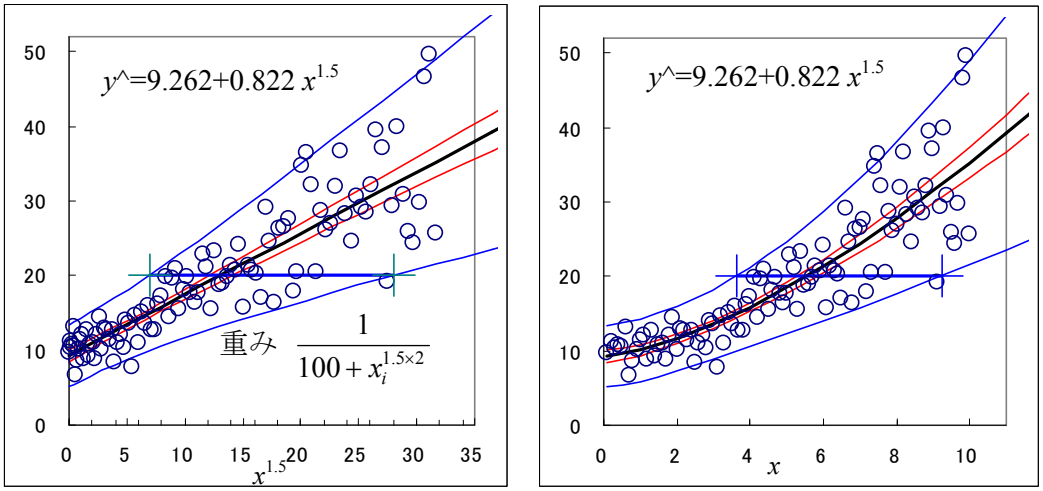


図 5.13　$x^{1.5}$ の関する回帰直線（左）と X 軸を x とした回帰曲線の表示（右）

5.5. 各種の誤差構造を持つデータの生成

回帰式が　$y_i = 10 + 2.5x_i + \varepsilon_i$，$\varepsilon_i$ が正規分布　$\varepsilon_i \sim N(0, (4 + x_i^2)\times 0.5^2)$　に従い，x_i が　0.1　から 0.1 刻みに 10 まで増加する 100 個を作成する手順を示す．分散は，$(4 + x_i^2)\times 0.5^2$ のように x_i によって変化する．正規分布に従う擬似乱数をコンピュータ内で発生させるために，乱数を発生する関数を用いる．この誤差構造は，$x = 0$ の場合に，分散が 0 ではなく $4\times 0.5^2 = 1.0$ となることを意味している．

Excel には，0～1 の範囲の一様乱数を発生させる Rand 関数があるので，これを標準正規分布の逆関数 Norsinv（Normal Standard Inverse）関数に入れ子にすることにより，正規乱数を得ることがで

きる.

まず C 列の 4 行目から 13 行目まで=Rand()関数で埋めると 0～1 の範囲の値が発生される. それを{=Normsinv(C4:C13)}として引数として用いると標準正規乱数が得られる. この値が, 平均 0, 標準偏差が 1 の正規分布に従うことになる. ファンクションキー F9 を押したり, 何か操作をするたびに関数の値は変化する. ここに示すのは, ある時点のスナップショットであり, 同じ乱数のセットを再現することはできない.

	A	B	C	D	E
1					
2			正規乱数の発生方法		
3		i	Rand()	Normsinv(C列)	Normsdist(D列)
4		1	0.462	−0.097	0.462
5		2	0.728	0.606	0.728
6		3	0.559	0.148	0.559
7		4	0.594	0.238	0.594
8		5	0.552	0.131	0.552
9		6	0.923	1.427	0.923
10		7	0.030	−1.885	0.030
11		8	0.173	−0.941	0.173
12		9	0.399	−0.256	0.399
13		10	0.620	0.304	0.620
14		11	0.708	0.548	0.708

図 5.14　Excel による標準正規乱数の発生

E 列で Normsdist(Normal Standard Distribution)関数を用いて D 列の標準正規乱数を{=Normsdist(D3:D14)} を引数として引用すると元の一様乱数に戻る.

ある指定した標準偏差 σ_e (分散は σ_e^2)に従う正規乱数を得るためには, 標準正規分布に従う乱数に σ_e を掛ければよい. 分散が $(4+x_i^2)\times\sigma_e^2$ となるような正規乱数を発生させるためには, 次の Excel シートに示すように σ_e に $\sqrt{4+x_i^2}$ を掛ければよい.

このような構造の正規乱数は, x_i が 0 に近づくときに, 分散が $4\sigma_e^2$ となる. さらに, x_i が大きくなれば, 近似的に分散が $x_i^2\sigma_e^2$ となる.

回帰式 $\hat{y}_i = 10 + 2.5x_i$ に, ある x_i について $\hat{y}_i$ を計算し, 分散が $(4+x_i^2)\times\sigma_e^2$ に従う正規乱数を加えればよい. 分散を $\sigma_e^2 = 0.5^2$ とした場合に Excel のシートを次に示す.

	A	B	C	D	E	F	G	H	I	J
1										
2			分散がおおよそx^2比例するデータ							
3			一様乱数	正規乱数	分散=0.5^2			e_i	10+2x_i	$y\hat{}_i$ +e_i
4		i	Rand	Nornsinv	D列×0.5	x_i	$\sqrt{(4+x_i^2)}$	G列×E列	$y\hat{}_i$	y_i
5		1	0.462	−0.097	−0.048	0.0	2.000	−0.097	10.0	9.9
6		2	0.728	0.606	0.303	1.0	2.236	0.678	12.0	12.7
7		3	0.559	0.148	0.074	2.0	2.828	0.209	14.0	14.2
8		4	0.594	0.238	0.119	3.0	3.606	0.429	16.0	16.4
9		5	0.552	0.131	0.066	4.0	4.472	0.294	18.0	18.3
10		6	0.923	1.427	0.713	5.0	5.385	3.841	20.0	23.8
11		7	0.030	−1.885	−0.943	6.0	6.325	−5.961	22.0	16.0
12		8	0.173	−0.941	−0.471	7.0	7.280	−3.426	24.0	20.6
13		9	0.399	−0.256	−0.128	8.0	8.246	−1.055	26.0	24.9
14		10	0.620	0.304	0.152	9.0	9.220	1.404	28.0	29.4
15		11	0.708	0.548	0.274	10.0	10.198	2.796	30.0	32.8

図 5.15 分散がおおよそ x^2 比例するデータ

5.2 節で例示した 1)から 3)の散布図は，回帰式は $\hat{y}_i = 10 + 2.5x_i$ と共通であるが，

1）分散一定: $\varepsilon_i \sim N(0, 2.5^2)$,

2）分散がおおよそ x_i^2 に比例: $\varepsilon_i' \sim N(0, (4+x_i^2)\times 0.5^2)$,

3）分散がおおよそ x_i に比例: $\varepsilon_i'' \sim N(0, (1+x_i)\times 1.5^2)$

誤差を x の関数であたえた結果である．y_i のおおよその最小値と最大値が同じになるように，それぞれの分散を調整している．

	A	B	C	D	E	F	G	H
1								
2				回帰 切片	β_0	10	10	10
3				回帰 冪パラ	β_1	2.5	2.5	2.5
4								
5				切片分散	ω_0	0	4	1
6				分散	σ_e^2	6.25	0.25	2.25
7					σ_e	2.5	0.5	1.5
8								
9						1)	2)	3)
10		i	x	一様乱数 Rand	正規乱数 Nornsinv	分散一定	おおよそ x^2に比例	おおよそ xに比例
11		1	0.1	0.462	−0.097	10.0	10.2	10.1
12		2	0.2	0.728	0.606	12.0	11.1	11.5
13		3	0.3	0.559	0.148	11.1	10.9	11.0
14		4	0.4	0.594	0.238	11.6	11.2	11.4
15		5	0.5	0.552	0.131	11.6	11.4	11.5
105		95	9.5	0.108	−1.239	30.7	27.7	27.7
106		96	9.6	0.065	−1.516	30.2	26.6	26.6
107		97	9.7	0.243	−0.696	32.5	30.8	30.8
108		98	9.8	0.970	1.879	39.2	43.9	43.8
109		99	9.9	0.987	2.237	40.3	46.0	45.8
110		100	10	0.073	−1.455	31.4	27.6	27.8

図 5.16　直線関係であるが分散の構造が異なるデータ

これらの 100 個のデータについて散布図を描き,【近似曲線の追加／直線近似／グラフに数式を表示】として作成したのが, 図 5.3 である. これらの 3 組の第 100 番目のデータは, 次のように計算されている.

1) 分散一定: $y_{100} = 10 + 2.5 \times 10 + \sqrt{2.5^2} \times (-1.455) = 31.4$,

2) 分散がおおよそ x_i^2 に比例: $y'_{100} = 10 + 2.5 \times 10 + \sqrt{4 + 10^2} \times \sqrt{0.5^2} \times (-1.455) = 27.6$

3) 分散がおおよそ x_i に比例: $y''_{100} = 10 + 2.5 \times 10 + \sqrt{1 + 10} \times \sqrt{1.5^2} \times (-1.455) = 27.8$

直線ではなく, わずかに凸になったり凹んだりする場合は, x^λ に関して 1 次式となる $y = \beta_0 + \beta_1 x^\lambda$ で, $\lambda = 0.5$ とすれば凸になり, $\lambda = 1.5$ とすれば凹む. これらの式で, 分散がおおよそ $x^{2\lambda}$ に比例する場合, 分散がおおよそ x^λ に比例する場合のデータを正規乱数を使って発生させる.

4) $y_i = 10 + 9x_i^{0.5} + \varepsilon_i$, $\varepsilon_i \sim N(0, (5 + x_i^{0.5\times 2}) \times 0.9^2)$,

5) $y'_i = 10 + 0.8x_i^{1.5} + \varepsilon'_i$, $\varepsilon'_i \sim N(0, (6 + x_i^{1.5}) \times 0.8^2)$,

6) $y''_i = 10 + 0.8x_i^{1.5} + \varepsilon''_i$, $\varepsilon''_i \sim N(0, (100 + x_i^{1.5\times 2}) \times 0.2^2)$

	A	B	C	D	E	I	J	K	L	M
1										
2				回帰 切片	β_0			10	10	10
3				回帰 冪パラ	β_1			9	0.8	0.8
4						回帰 冪		0.5	1.5	1.5
5				切片分散	ω_0			5	6	100
6				分散	σ_e^2			0.81	0.64	0.04
7					σ_e			0.9	0.8	0.2
8										
9								4)	5)	6)
10		i	x	一様乱数 Rand	正規乱数 Nornsinv			$\lambda = 0.5$乗 分散: $(x^{0.5})^2$	$\lambda = 1.5$乗 分散: $(x^{1.5})$	$\lambda = 1.5$乗 分散: $(x^{1.5})^2$
11		1	0.1	0.462	−0.097			12.6	9.8	9.8
12		2	0.2	0.728	0.606			15.3	11.3	11.3
13		3	0.3	0.559	0.148			15.2	10.4	10.4
14		4	0.4	0.594	0.238			16.2	10.7	10.7
15		5	0.5	0.552	0.131			16.6	10.5	10.5
106		96	9.6	0.065	−1.516			32.7	26.5	24.3
107		97	9.7	0.243	−0.696			35.6	30.8	29.7
108		98	9.8	0.970	1.879			44.7	43.6	46.7
109		99	9.9	0.987	2.237			46.1	45.8	49.6
110		100	10	0.073	−1.455			33.4	28.2	25.6

図 5.17 直線関係でなく分散の構造も異なるデータ

これらの 100 個のデータについて散布図を描き,【近似曲線の追加／直線近似／グラフに数式を表示】として作成したのが, 図 5.9 である. これらの 3 組の第 100 番目のデータは,

4) $y_{100} = 10 + 9 \times 10^{0.5} + \sqrt{5 + 10^{0.5\times 2}} \times 0.9^2 \times (-1.455) = 33.4$,

5) $y'_{100} = 10 + 0.8 \times 10^{1.5} + \sqrt{6 + 10^{1.5\times 2}} \times \sqrt{0.8^2} \times (-1.455) = 28.2$

6) $y''_{100} = 10 + 0.8 \times 10^{1.5} + \sqrt{100 + 10^{1.5\times 2}} \times \sqrt{0.2^2} \times (-1.455) = 25.6$

となる.

6. 分散が x の関数で現される場合

6.1. 分散が x^2 に比例する回帰直線

回帰分析の行列による表記

第 2 章では，5 個のデータ（x, y）＝（1, 0），（1, 2），（2, 3），（3, 4），（3, 6）について

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\ \ i = 1,2,...,5\,,\ \ \varepsilon_i \sim N(0, \sigma_e^2) \tag{6.1}$$

x の大きさに関係なく，誤差 ε_i が互いに独立に平均が 0，分散が σ_e^2 の正規分布に従うと仮定した．

分散が x^2 に比例する例として，新たに 5 個のデータ（x, y）＝（1, 10），（1, 12），（2, 21），（3, 30），（3, 37）を用いる．このデータでは，$x = 1$ の場合に 2 つのデータの範囲が 2 であるのに対し，$x = 3$ の場合は 7 と大きくなり，分散が一定とみなすことが困難である．分散が x^2 に比例する場合には，式（6.1）の ε_i の分散が，次式のように，

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\ \ i = 1,2,...,5\,,\ \ \varepsilon_i \sim N(0, x_i^2\sigma_e^2) \tag{6.2}$$

誤差 ε_i が互いに独立に平均 0，分散が $x_i^2\sigma_e^2$ の正規分布に従うと定式化できる．誤差 ε_i のそれぞれの分散は，$x_i = 1$ の場合：$1\times\sigma_e^2$，$x_i = 2$ の場合：$2^2\sigma_e^2 = 4\sigma_e^2$，$x_i = 3$ の場合：$3^2\sigma_e^2 = 9\sigma_e^2$ のように x の大きさに応じて変化する．

なお，「分散が x^2 に比例」することは，「標準偏差が x に比例」することと同じである．「標準偏差が x に比例」との表現は，標準偏差を x で割った変動係数が一定になるので，一般的に使われているが，「分散が x に比例」する場合に「標準偏差が $\sqrt{x}$ に比例」という表現になり，本書では，「分散が，．．．」という表現に統一した．

i	$\boldsymbol{X}$	(x_i)	$\boldsymbol{\varepsilon}$	$x_i{}^2\times e_i$
1	1	1	ε_1	$1\times e_1$
2	1	1	ε_2	$1\times e_2$
3	1	2	ε_3	$4\times e_3$
4	1	3	ε_4	$9\times e_4$
5	1	3	ε_5	$9\times e_5$

重み付き回帰分析

誤差分散が一定でないと最小 2 乗法が適用できないので，誤差項を x_i で割ることにより，誤差分散を次式のように σ_e^2 と一定にすることができる．

$$y_i = \beta_0 + \beta_1 x_i + \frac{\varepsilon_i}{x_i}\,,\ \ i = 1,2,...,5\,,\ \ \frac{\varepsilon_i}{x_i} \sim N(0, \sigma_e^2) \tag{6.3}$$

最小にしたい誤差平方和 S_e は，

$$S_e = \sum_{i=1}^{5}\left(\frac{y_i - \hat{y}_i}{x_i}\right)^2 = \sum_{i=1}^{5}\left(\frac{y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i}{x_i}\right)^2 = \sum_{i=1}^{5}\left(\frac{\hat{\varepsilon}_i}{x_i}\right)^2 = \sum_{i=1}^{5} e_i^2 \tag{6.4}$$

と定式化できる. 行列計算で S_e を計算するためには, x_i の 2 乗の逆数を対角要素となるような行列 $\boldsymbol{V}^{-1}$ を挟み込み, 次に示すように $\boldsymbol{\varepsilon}$ のベクトルの 2 次形式をとればよい.

$$\underset{(1\times5)}{\boldsymbol{\varepsilon}^T}\ \underset{(5\times5)}{\boldsymbol{V}^{-1}}\ \underset{(5\times1)}{\boldsymbol{\varepsilon}}$$

$$\boldsymbol{\varepsilon}^T = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 & \varepsilon_5 \end{bmatrix}, \quad \boldsymbol{V}^{-1} = \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/9 & 0 \\ 0 & 0 & 0 & 0 & 1/9 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

図 6.1　誤差分散が x^2 に比例する場合の誤差平方和の計算

正規方程式を解くと, $\boldsymbol{V}^{-1}$ を挟み込んだ次式が得られる. 詳細は本節 の「正規方程式の誘導」を参照のこと.

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\,\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{6.5}$$

さらに $\hat{\boldsymbol{\beta}}$ について解くために両辺に $(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ を掛けて, 重み付き回帰式

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{6.6}$$

が得られる. Excel の標準的な関数で計算する際に大きさ ($n\times n$) の行列 $\boldsymbol{V}^{-1}$ を必要とするのであるが, 次に示すように求めることができる. 注) 実際の計算には分数ではなく実数で与える.

$$\underset{(2\times5)}{\boldsymbol{X}^T}\ \underset{(5\times5)}{\boldsymbol{V}^{-1}}\ \underset{(5\times2)}{\boldsymbol{X}} = \underset{(2\times2)}{\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}}$$

$$\boldsymbol{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 3 \end{bmatrix}, \quad \boldsymbol{V}^{-1} = \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/9 & 0 \\ 0 & 0 & 0 & 0 & 1/9 \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}, \quad \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} = \begin{bmatrix} 2.472 & 3.167 \\ 3.167 & 5 \end{bmatrix}$$

$$\underset{(2\times1)}{(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}}\ \underset{(2\times5)}{\boldsymbol{X}^T}\ \underset{(5\times5)}{\boldsymbol{V}^{-1}}\ \underset{(5\times1)}{\boldsymbol{y}} = \underset{(2\times1)}{\boldsymbol{\beta}\hat{}}$$

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} = \begin{bmatrix} 2.143 & -1.357 \\ -1.357 & 1.060 \end{bmatrix}, \quad \boldsymbol{X}^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 3 \end{bmatrix}, \quad \boldsymbol{V}^{-1} = \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 \\ 0 & 0 & 0 & 1/9 & 0 \\ 0 & 0 & 0 & 0 & 1/9 \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} 10 \\ 12 \\ 21 \\ 30 \\ 37 \end{bmatrix}, \quad \boldsymbol{\beta}\hat{} = \begin{bmatrix} -0.071 \\ 11.012 \end{bmatrix}$$

図 6.2　誤差分散が x^2 に比例する場合の回帰パラメータの推定

推定された $\hat{\boldsymbol{\beta}}$ を用いて推定値 $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ を計算し, 誤差平方和 S_e を次式で求める.

$$S_e = (\boldsymbol{y} - \hat{\boldsymbol{y}})^T \boldsymbol{V}^{-1} (\boldsymbol{y} - \hat{\boldsymbol{y}}) = \hat{\boldsymbol{\varepsilon}}^T \boldsymbol{V}^{-1} \hat{\boldsymbol{\varepsilon}} = \boldsymbol{e}^T \boldsymbol{e} \tag{6.7}$$

誤差平方和は，行列計算を使わなくても，それぞれの誤差 $\hat{\varepsilon}_i = y_i - \hat{y}_i$ を求め，e_i を x_i で割った e_i' の平方から，それらの和によって平方和 S_e を求めることもできる．誤差分散の推定値 $\hat{\sigma}_e^2$ は，平方和 S_e を自由度 $f_e = 5 - 2 = 3$ で割って求める．

$\boldsymbol{X}$		$\boldsymbol{\beta}$^	=	$\boldsymbol{y}$^	$\boldsymbol{y}$	$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{y}$^	$e_i = \varepsilon_i / x_i$	$\boldsymbol{\Sigma} = (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1} \sigma_e^{2}$^	
1	1	−0.07		10.94	10	0.940	0.940	3.586	−2.271
1	1	11.01		10.94	12	−1.060	−1.060	−2.271	1.773
1	2			21.95	21	0.952	0.476		
1	3			32.96	30	2.964	0.988		
1	3			32.96	37	−4.036	−1.345		
						$S_e = \Sigma e_i^2$	5.020		
(5x2)		(2x1)		(5x1)	(5x1)	σ_e^{2}^ =	1.673		

図 6.3　誤差分散が x^2 に比例する場合の誤差分散と分散共分散行列

誤差分散の推定値 $\hat{\sigma}_e^2 = S_e / f_e$ は，これまでの誤差分散とは明らかに異なり，推定された $\hat{\sigma}_e^2$ は，$x = 1$ の場合での分散の推定値であり，$x = 2$ の場合は $4\hat{\sigma}_e^2$ となる．回帰パラメータの分散は，分散共分散行列 $\boldsymbol{\Sigma}$ の対角要素で求められているので，その平方根で SE を求め t 統計量および p 値を計算した結果を次に示す．

表 6.1　誤差分散が x^2 に比例する場合の回帰パラメータの推定

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k$^	$Var(\beta_k$^$)$	SE	t	p
0	x_0:切片	β_0	−0.071	3.586	1.894	−0.038	0.972
1	x_1:傾き	β_1	11.012	1.773	1.331	8.270	0.004

各種の 95%信頼区間

回帰直線の 95%信頼区間は，$\hat{y}_i$ の分散 $Var(\hat{y}_i)$ に自由度が $f_e = n - p = 3$ の t 分布の両側 5%点 3.182 を掛けて $\hat{y}_i$ にプラス・マイナスした，

$$\hat{y}_i \text{ の 95\%信頼区間} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)} \tag{6.8}$$

から求める．$Var(\hat{y}_5)$ は，5 番目のベクトル $\boldsymbol{x}_5$ としたときに

$$Var(\hat{y}_5) = Var(\boldsymbol{x}_5 \hat{\boldsymbol{\beta}}) = \boldsymbol{x}_5 \boldsymbol{\Sigma} \boldsymbol{x}_5^T$$

で計算する．実際に計算すると次のようになる．

$\boldsymbol{x}_5$		$\boldsymbol{\Sigma} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \sigma_e^{2}$^		$\boldsymbol{x}_5^T$	=	$Var(y_5$^$)$
1	3	3.586	−2.271	1		5.916
		−2.271	1.773	3		
(2x1)		(2x2)		(2x1)		(1x1)

図 6.4　誤差分散が x^2 に比例する場合の回帰推定値の分散

これを用いて回帰直線の 95%信頼区間は,

$$\hat{y}_5 \text{の}(L95R,\ U95R) = 32.964 \pm 3.182\sqrt{5.916} = (25.224\ ,40.705) \tag{6.9}$$

となる.

個別データの 95%信頼区間は, 回帰直線上の $\hat{y}_i$ の分散に, 誤差 $y_i - \hat{y}_i = x_i e_i$ の分散を加える. た分散を求めればよい,

$$Var(\hat{y}_i + x_i e_i) = Var(\hat{y}_i) + x_i^2 \hat{\sigma}_e^2 \tag{6.10}$$

第 5 番目について計算すると

$$Var(\hat{y}_5) + x_5^2 \hat{\sigma}_e^2 = 5.916 + 3^2 \times 1.673 = 20.976 \tag{6.11}$$

$$\hat{y}_5 \text{の個別}(L95I,\ U95I) = 32.964 \pm 3.182\sqrt{20.976} = (18.389\ ,47.540\ )$$

複数データの平均に対する分散 $Var\ (\overline{y}_{i.})$ は, データ数 m で誤差分散を除して, 次のようになる.

$$Var(\overline{y}_{i.}) = Var(\hat{y}_i) + x_i^2 \frac{\hat{\sigma}_e^2}{m} \tag{6.12}$$

第 5 番目について計算すると

$$Var(\hat{y}_5) + x_5^2 \frac{\hat{\sigma}_e^2}{m} = 5.916 + 3^2 \times \frac{1.673}{3} = 10.936 \tag{6.13}$$

$\hat{y}_5$ の 3 個のデータの平均(L95m3, U95 m3)

$$= 32.964 \pm 3.182\sqrt{10.936} = (22.440, 43.489)$$

これらの 95%信頼区間についての計算結果を次にまとめて示す.

表 6.2　誤差分散が x^2 に比例する場合の各種 95%信頼区間の推定

								$m=$	3
i	x	y	$Var(y_i\hat{})$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
1	1	10	0.8167	8.064	13.817	5.919	15.962	7.209	14.672
2	1	12	0.8167	8.064	13.817	5.919	15.962	7.209	14.672
3	2	21	1.5936	17.935	25.970	12.791	31.114	15.729	28.176
4	3	30	5.9162	25.224	40.705	18.389	47.540	22.440	43.489
5	3	37	5.9162	25.224	40.705	18.389	47.540	22.440	43.489

図 6.5 は, x について -1 から 0.5 刻みに 5 まで変化させ, それらの 95%信頼区間を計算した結果を用いて作図したものである. 重み付き回帰分析の回帰の 95%信頼区間の形状は, 重みなしの場合と異なる. 重みなしの場合, 回帰の 95%信頼区間の幅が最も狭くなるのは, x_i の算術平均であるが, 分散が x^2 に比例する場合, $w_i = 1/x_i^2$ を重みとした平均になるので (式の誘導は本節の補足を参照),

$$\mu_{(w=1/x^2)} = \frac{\sum_{i=1}^{5} w_i x_i}{\sum_{i=1}^{5} w_i} = \frac{\sum_{i=1}^{5} \frac{1}{x_i^2} x_i}{\sum_{i=1}^{5} \frac{1}{x_i^2}} = \frac{3.167}{2.472} = 1.281$$

重み付き平均 1.281 で幅が最小となる．

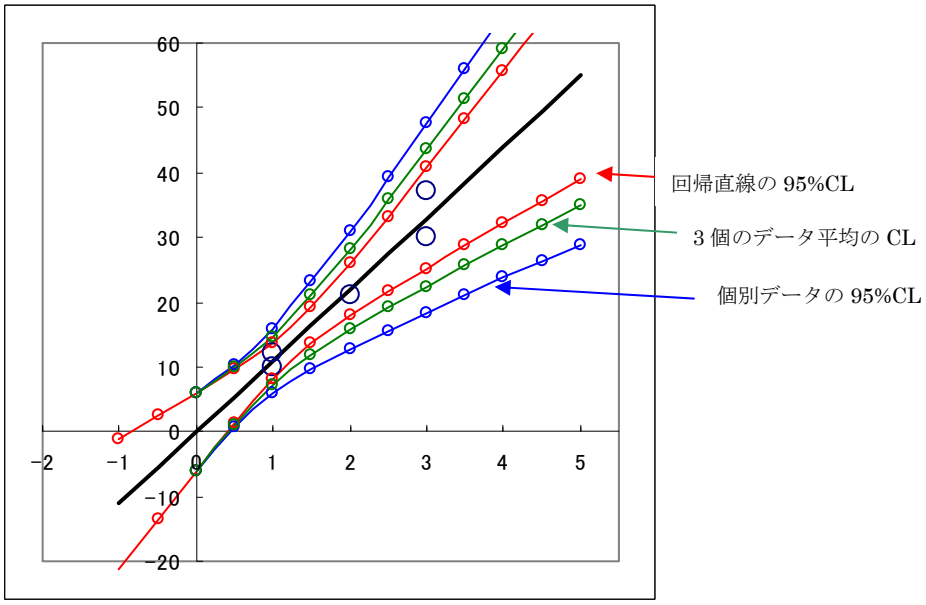


図 6.5　誤差分散が x^2 に比例する場合の各種の 95%信頼区間

個別データの 95%信頼区間は，$x \geq 0$ の場合にのみ図示してある．これは，分散が x^2 に比例する場合には，x が負の場合でも分散は正となるが，一般的には $x \geq 0$ の場合に限定される．これは，分散が x^2 に比例するような場合に x が負の場合には，分散が定義できなためである．

個別データの 95%信頼区間は，回帰式の分散に $x_i^2 \hat{\sigma}_e^2$ を加えているので，$x = 0$ の場合に個別データの分散は 0 となり，95%信頼区間は，回帰の 95%信頼区間に一致する．個別データの 95%信頼区間の幅が最小になるのは，$x_{0(I)} = 0.659$ の点となる．

$$x_{0(I)} = \frac{\sum w_i x_i}{\sum w_i + D} = \frac{3.167}{2.472 + (2.472 \times 5 - 3.157^2)} = 0.659$$

ただし，$D = \left| \boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X} \right| = \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{i=1}^{n} w_i x_i \right)^2$

複数データの平均の 95%信頼区間が最小になるのは，

$$x_{0(m)} = \frac{m \sum w_i x_i}{m \sum w_i + D} = \frac{3 \times 3.167}{3 \times 2.472 + (2.472 \times 5 - 3.157^2)} = 0.974$$

の点とらり，個別データの場合に比べ大きくなる．

各種の逆推定の 95%信頼区間

回帰直線の逆推定

回帰直線の 95%信頼区間について与えられた y_0 に対する逆推定値の下限は，重みなしの場合に示した 式（2.26）と同じであり，

$$(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95})+t_\alpha\sqrt{Var(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95})}=y_0 \tag{6.14}$$

この式を $\hat{x}_{L95}$ について整理すると，

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

となる．$y_0=25.0$ の場合には，$t_\alpha=t(3,0.05)=3.182$ として，

$$a=Var(\hat{\beta}_0)-\frac{(y_0-\hat{\beta}_0)^2}{t_\alpha^2}=3.586\ -\frac{(25-(-0.071))^2}{3.182^2}=-58.478$$

$$b=2Cov(\hat{\beta}_0,\hat{\beta}_1)+\frac{2(y_0-\hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2}=2\times(-2.271)+\frac{2\times\{25-(-0.071)\}\times 11.012}{3.182^2}=49.98$$

$$c=Var(\hat{\beta}_1)-\frac{\hat{\beta}_1^{\,2}}{t_\alpha^2}=1.773-\frac{11.012^2}{3.182^2}=-10.200$$

この a，b，c を用いて 2 次式の根の公式から，図 6.6 に示すように 95%信頼区間は（1.931，2.968）が得られる．

個別データの逆推定の 95%信頼区間は，回帰直線の分散に個別データの分散が加わり，$\hat{x}_{L95\%}^2$ の項の係数 c にを次のように c' として 2 次式の根の公式を適用すればよい．

$$(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95I})+t_\alpha\sqrt{Var(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95I})+\hat{x}_{L95I}^2\hat{\sigma}_e^2}=y_0 \tag{6.15}$$

$$c'=c+\hat{\sigma}_e^2=-10.200+1.673=-8.527$$

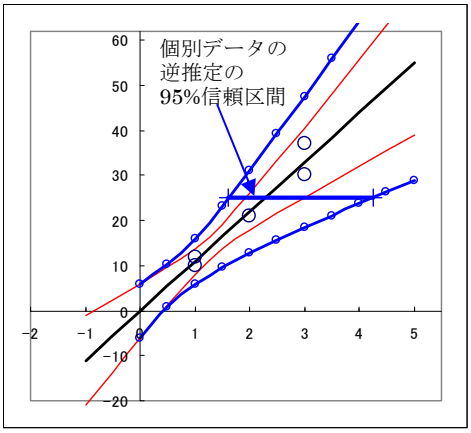
図 6.6 に示すように，個別データの 95%信頼区間（1.615，4.246）が得られる．

複数データの平均値の 95%信頼区間は，

$$(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95I})+t_\alpha\sqrt{Var(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95I})+\hat{x}_{L95I}^2\frac{\hat{\sigma}_e^2}{m}}=y_0 \tag{6.16}$$

$$c''=c+\frac{\hat{\sigma}_e^2}{m}=-10.200+\frac{1.673}{3}=-9.642$$

であり， $m=3$ の場合の 95%信頼区間は，図 6.6 に示すように（1.784，3.399）が得られる．



逆推定 y_0 = 25.0　x_0^= 2.277

回帰直線の95%CL

a = −58.478　x_{L95R}^ = 1.931

b = 49.977　x_{U95R}^ = 2.968

c = −10.200

個別データの95%CL

a = −58.478　x_{L95I}^ = 1.615

b = 49.977　x_{U95I}^ = 4.246

c' = −8.527

複数データ(m=3)の平均の95%CL

a = −58.478　x_{L95m}^ = 1.784

b = 49.977　x_{U95m}^ = 3.399

c'' = −9.642

図 6.6　$y_0 = 25$ における個別データの逆推定の 95%信頼区間

統計ソフトによる検証

JMP を用いて，分散が x^2 に比例する場合について回帰分析を行なう．データとして重み w を計算式 $w = 1/x_i^2$ で次のように与える．

データ

	y	X0	X1	w
1	10	1	1	1.000
2	12	1	1	1.000
3	21	1	2	0.250
4	30	1	3	0.111
5	37	1	3	0.111

【モデルのあてはめ】で解析すると，誤差分散は，次に示す分散分析表の誤差の平均平方から 1.673 となり，回帰パラメータの推定値は，$\hat{\beta}_0 = -0.0714$，$\hat{\beta}_1 = -11.0119$ と Excel の結果と一致した．

解析モデル　y = x0　x1 / 重み= w,　切片なし

要因	自由度	平方和	平均平方
モデル	1	114.4496	114.450
誤差	3	5.0198	1.673
全体(修正済み)	4	119.4694	

項	推定値	標準誤差	t値	p値(Prob>|t|)
X0	−0.0714	1.8936	−0.0377	0.9723
X1	11.0119	1.3315	8.2703	0.0037*

回帰直線および個別データの 95%信頼区間を【列の保存】でデータファイルに出力した結果を次に示す．

各種の 95%信頼区間（個別 y の 95% は，Excel の結果と一致しない）

	y	X0	X1	w	予測式 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	10	1	1	1.000	10.904	7.443	14.365	5.960	15.847
2	12	1	1	1.000	10.904	7.443	14.365	5.960	15.847
3	21	1	2	0.250	21.962	18.170	25.753	16.781	27.142
4	30	1	3	0.111	33.019	25.926	40.112	25.096	40.942
5	37	1	3	0.111	33.019	25.926	40.112	25.096	40.942

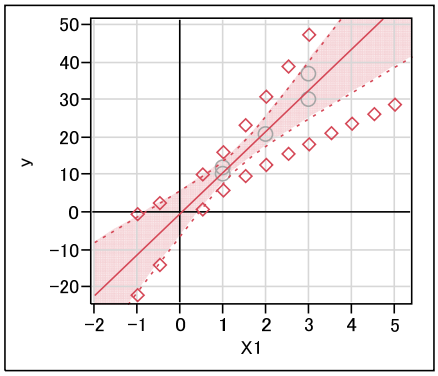
回帰直線の 95%信頼区間は【平均 y の下側，上側】となっていて，表 6.2 に示した Excel の結果と一致するが，個別データの 95%信頼区間の推定は，Excel の結果と一致しない．JMP では，次式のように，

$$
\begin{aligned}
(\hat{x}_{L95I}, \hat{x}_{U95I}) &= \hat{y}_i \pm t_{0.05}\sqrt{(1+\boldsymbol{x}_i(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i^T)\hat{\sigma}_e^2} \\
&= \hat{y}_i \pm t_{0.05}\sqrt{Var(\hat{y}_i)+\hat{\sigma}_e^2}
\end{aligned}
\tag{6.17}
$$

個別データの誤差を $x_i^2\hat{\sigma}_e^2$ とすべきなのであるが，$x=1$ の場合の誤差 $\hat{\sigma}_e^2$ を用いており，$x=1$ 以外の場合は不一致となる．これは，JMP の統計計算の母体となっている SAS 本体の回帰分析 REG プロシジャでも，次のように計算式が示されていて JMP と整合している．これは，重みを x の関数ではなく，計算された結果としての数値とみなして対応しているため，結果として x による誤差の変化が認識されていないためであろう．

このように，JMP などの統計ソフトと Excel の結果が一致できないような場合に，統計ソフトでの計算に対し批判的吟味が必要となる．なお，JMP の【2 変量の関係】による個別データの 95%信頼区間は，数値としての出力はないが，図では，x が大きくなるにつれ Excel と同様の扇型 95%信頼区間が表示されるので，それぞれの x_i での重みを考慮しているように思われる．ただし，用いている数式が不明であるので，これ以上の論評はしない．

JMP の【2 変量の関係】による個別データの 95%信頼区間



与えられた $y_0=25$ に対する回帰直線の逆推定と 95%信頼区間は，次のように一致した．しかしながら，JMP の場合 式（6.17）を用いているため，式（6.15）で求めた個別データの 95%信頼区間とは異なる．

JMP

指定されたy	X1の予測値	下側95%	上側95%
25.0000	2.2768	1.9315	2.9682

応答の期待値に対する信頼区間

指定されたy	X1の予測値	下側95%	上側95%
25.0000	2.2768	1.792022	3.107672

個別の応答に対する信頼区間

Excel

逆推定 y_0 =	25.0	x_0^=	2.277
回帰直線の95%CL			
a =	−58.478	$x_{L\,95R}$ =	1.931
b =	49.977	$x_{U\,95R}$ =	2.968
c =	−10.200		

個別データの95%CL			
a =	−58.478	$x_{L\,95I}$^ =	1.615
b =	49.977	$x_{U\,95I}$^ =	4.246
c' =	−8.527		

正規方程式の誘導

回帰係数の推定値 $\hat{\boldsymbol{\beta}}$ は，$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$ で求められると，天下り的に示したので，補足をする．回帰式を，

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\ \ i = 1,2,\ldots,n\,,\ \ \varepsilon_i \sim N(0, x_i^2\sigma_e^2) \tag{6.18}$$

としたときに，重みを $w_i = 1/x_i^2$ としたときの β_0 と β_1 の最小 2 乗推定量は，

$$Q = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} w_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \sum_{i=1}^{n} w_i \varepsilon_i^2 \tag{6.19}$$

式(6.19)を $\hat{\beta}_0$ と $\hat{\beta}_1$ で微分した式を 0 と置くと，式(6.20)が得られる．

$$\begin{aligned} \frac{\partial Q}{\partial \hat{\beta}_0} &= -2\sum_{i=1}^{n} w_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \\ \frac{\partial Q}{\partial \hat{\beta}_1} &= -2\sum_{i=1}^{n} \left\{ w_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i \right\} = 0 \end{aligned} \tag{6.20}$$

である．これらの微分した式を分解すると，

$$\begin{aligned} &\sum_{i=1}^{n} w_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} w_i - \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i = 0 \\ &\sum_{i=1}^{n} w_i x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} w_i x_i - \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i^2 = 0 \end{aligned} \tag{6.21}$$

となり $\hat{\beta}_0$ と $\hat{\beta}_0$ が，含まれない項を右辺に移して整理すると，

$$\begin{aligned} &\hat{\beta}_0 \sum_{i=1}^{n} w_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i = \sum_{i=1}^{n} w_i y_i \\ &\hat{\beta}_0 \sum_{i=1}^{n} w_i x_i + \hat{\beta}_1 \sum_{i=1}^{n} w_i x_i^2 = \sum_{i=1}^{n} w_i x_i y_i \end{aligned} \tag{6.22}$$

正規方程式が得られる．ところで，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}$ は，

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} = \begin{bmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} w_i x_i \\ \sum_{i=1}^{n} w_i x_i & \sum_{i=1}^{n} w_i x_i^2 \end{bmatrix} \tag{6.23}$$

であり，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$ は，

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} = \begin{bmatrix} \sum_{i=1}^{n} w_i y_i \\ \sum_{i=1}^{n} w_i x_i y_i \end{bmatrix} \tag{6.24}$$

であるで，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}$ に $\hat{\boldsymbol{\beta}}$ を掛けると正規方程式(6.22)の左辺となり，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$ が右辺となるので行列表記で

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{6.25}$$

となる．逆行列 $(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ を両辺にかけて

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{6.26}$$

が得られる．重みが全て 1 の場合は，$\boldsymbol{V}^{-1}$ が消えて，重みなしの回帰の一般式，

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{6.27}$$

となる．

誤差分散が最小となる x_0

重みなしの回帰分析の場合，誤差分散が最小となるのは，x_0 が算術平均 $\bar{x}$. であった．これと同様な手順で重み付きの回帰分析の誤差分散が最小となる x_0 を求めてみよう．

$$Var(\hat{y}_i) = Var(\boldsymbol{x}_i\hat{\boldsymbol{\beta}}) = \boldsymbol{x}_i\boldsymbol{\Sigma}\boldsymbol{x}_i^T \tag{6.28}$$

である．ここで，$\boldsymbol{\Sigma}$ は，

$$\boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\hat{\sigma}_e^2 \tag{6.29}$$

であり，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}$ は，

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} = \begin{bmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} w_i x_i \\ \sum_{i=1}^{n} w_i x_i & \sum_{i=1}^{n} w_i x_i^2 \end{bmatrix} \tag{6.30}$$

であるので，その逆行列 $(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ は，

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} = \frac{1}{D}\begin{bmatrix} \sum_{i=1}^{n} w_i x_i^2 & -\sum_{i=1}^{n} w_i x_i \\ -\sum_{i=1}^{n} w_i x_i & \sum_{i=1}^{n} w_i \end{bmatrix} \tag{6.31}$$

ただし，$D = |\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}| = \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{i=1}^{n} w_i x_i\right)^2$ である．

$$Var(\hat{\beta}_0) = \frac{\sum w_i x_i^2}{D}\hat{\sigma}_e^2 \tag{6.32}$$

$$Var(\hat{\beta}_1) = \frac{\sum w_i}{D}\hat{\sigma}_e^2 \tag{6.33}$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sum w_i x_i}{D}\hat{\sigma}_e^2 \tag{6.34}$$

ある x_0 に対する $\hat{y}_0$ の誤差分散 $Var(\hat{y}_0)$ は，

$$\begin{aligned} Var(\hat{y}_0) &= Var(\hat{\beta}_0) + 2x_0 Cov(\hat{\beta}_0, \hat{\beta}_1) + x_0^2 Var(\hat{\beta}_1) \\ &= \frac{\sum w_i x_i^2 - 2x_0 \sum w_i x_i + x_0^2 \sum w_i}{D}\hat{\sigma}_e^2 \end{aligned} \tag{6.35}$$

なので， x_0 について偏微分では，D および $\hat{\sigma}_e^2$ には x_0 が含まれていないので，分母を x_0 について偏微分し，0 とおいて x_0 について解くと

$$\frac{\partial\left(\sum w_i x_i^2 - 2x_0 \sum w_i x_i + x_0^2 \sum w_i\right)}{\partial x_0} = -2\sum w_i x_i + 2x_0 \sum w_i = 0 \tag{6.36}$$

$$x_0 \sum w_i = \sum w_i x_i \tag{6.37}$$

$$x_0 = \frac{\sum w_i x_i}{\sum w_i} \tag{6.38}$$

となり，重み付き平均のときに分散が最小になる．

個別データの分散が最小になる $x_{0(I)}$

$$\begin{aligned} Var(y_0 + x_{0(I)} e_i) &= Var(y_0) + x_{0(I)}^2 \hat{\sigma}_e^2 \\ &= \frac{\sum w_i x_i^2 - 2x_{0(I)} \sum w_i x_i + x_{0(I)}^2 \sum w_i}{D}\hat{\sigma}_e^2 + x_{0(I)}^2 \hat{\sigma}_e^2 \end{aligned} \tag{6.39}$$

分母を $x_{0(I)}$ について偏微分し，0 おいて $x_{0(I)}$ について解くと

$$-2\sum w_i x_i + 2x_{0(I)} \sum w_i + 2Dx_{0(I)} = 0 \tag{6.40}$$

$$x_{0(I)} \sum w_i + Dx_{0(I)} = \sum w_i x_i \tag{6.41}$$

$$x_{0(I)} = \frac{\sum w_i x_i}{\sum w_i + D} \tag{6.42}$$

が得られる．

複数データの分散が最小になる $x_{0(m)}$

$$
\begin{aligned}
Var(y_0 + x_{0(m)}e_i) &= Var(y_0) + x_{0(m)}^2 \frac{\hat{\sigma}_e^2}{m} \\
&= \frac{\sum w_i x_i^2 - 2x_{0(m)} \sum w_i x_i + x_{0(m)}^2 \sum w_i}{D} \hat{\sigma}_e^2 + x_{0(m)}^2 \frac{\hat{\sigma}_e^2}{m} \qquad (6.43) \\
&= \frac{m\left(\sum w_i x_i^2 - 2x_{0(m)} \sum w_i x_i + x_{0(m)}^2 \sum w_i\right) + Dx_{0(m)}^2}{mD} \hat{\sigma}_e^2
\end{aligned}
$$

分母を $x_{0(m)}$ について偏微分し，0 おいて $x_{0(m)}$ について解くと

$$
-2m\sum w_i x_i + 2x_{0(m)} m \sum w_i + 2Dx_{0(m)} = 0
$$

$$
x_{0(m)} m \sum w_i + Dx_{0(m)} = m \sum w_i x_i \qquad (6.44)
$$

$$
x_{0(m)} = \frac{m\sum w_i x_i}{m\sum w_i + D} \qquad (6.45)
$$

が得られる.

6.2. 分散が x に比例する場合

前節で用いた 5 個のデータ（x, y）＝(1, 10)，(1, 12)，(2, 21)，(3, 30)，(3, 37)を用いて，誤差分散 σ_e^2 が x_i に比例して大きくなるとする. 前節の，誤差分散が x_i^2 に比例する場合に比べ，x_i が大きくなっても測定値のばらつきが相対的に小さくなる. ,

$$
y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\ i = 1,2,...,5\,,\ \varepsilon_i \sim N(0, x_i\sigma_e^2) \qquad (6.46)
$$

誤差 ε_i が互いに独立に平均 0，分散が $x_i\sigma_e^2$ の正規分布に従うことになる. これは，誤差 ε_i が，$x_i = 1$ の場合: $(1\times\sigma_e)^2 = \sigma_e^2$，$x_i = 2$ の場合: $2\sigma_e^2$，$x_i = 3$ の場合: $3\sigma_e^2$ となることを意味する.

i	X		ε	$x_i e_i$
1	1	1	ε_1	$1\times\sigma_e^2$
2	1	1	ε_2	$1\times\sigma_e^2$
3	1	2	ε_3	$2\times\sigma_e^2$
4	1	3	ε_4	$3\times\sigma_e^2$
5	1	3	ε_5	$3\times\sigma_e^2$

図 6.7　誤差構造

それぞれの ε_i に $\sqrt{x_i}$ の逆数を掛ければ,

$$
y_i = \beta_0 + \beta_1 x_i + \frac{\varepsilon_i}{\sqrt{x_i}}\,,\ i = 1,2,...,5\,,\ \frac{\varepsilon_i}{\sqrt{x_i}} \sim N(0, \sigma_e^2) \qquad (6.47)
$$

と，誤差分散を一定化できる. 最小にしたい誤差平方和 S_e は,

$$S_e = \sum_{i=1}^{5}\left(\frac{y_i - \hat{y}_i}{\sqrt{x_i}}\right)^2 = \sum_{i=1}^{5}\frac{\varepsilon_i^2}{x_i} = \sum_{i=1}^{5} e_i^2 \tag{6.48}$$

となるので，$1/x_i$ を対角要素となるような行列 $\boldsymbol{V}^{-1}$ を定義し，前節と同様な行列計算を行なう．

$\boldsymbol{X}^T$ (2x5)

1	1	1	1	1
1	1	2	3	3

$\boldsymbol{V}^{-1}$ (5x5)

1/1	0	0	0	0
0	1/1	0	0	0
0	0	1/2	0	0
0	0	0	1/3	0
0	0	0	0	1/3

$\boldsymbol{X}$ (5x2)

1	1
1	1
1	2
1	3
1	3

$= \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}$ (2x2)

3.167	5.000
5.000	10

$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ (2x2)

1.500	−0.750
−0.750	0.475

$\boldsymbol{X}^T$ (2x5)

1	1	1	1	1
1	1	2	3	3

$\boldsymbol{V}^{-1}$ (5x5)

1/1	0	0	0	0
0	1/1	0	0	0
0	0	1/2	0	0
0	0	0	1/3	0
0	0	0	0	1/3

$\boldsymbol{y}$ (5x1)

10
12
21
30
37

$= \boldsymbol{\beta}$^ (2x1)

−0.250
11.125

図 6.8　回帰パラメータの行列計算

$\boldsymbol{X}$ (5x2)

1	1
1	1
1	2
1	3
1	3

$\boldsymbol{\beta}$^ (2x1)

−0.250
11.125

$=$ $\boldsymbol{y}$^

10.88
10.88
22.00
33.13
33.13

$\boldsymbol{y}$ (5x1)

10
12
21
30
37

$\boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{y}$^

0.875
−1.125
1.000
3.125
−3.875

$e_i = \varepsilon_i/\sqrt{x_i}$

0.875
−1.125
0.500
1.042
−1.292

$S_e = \boldsymbol{e}^T\boldsymbol{e} =$ 5.035

$\sigma_e^{2\wedge} =$ 1.678

$\boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\sigma_e^2$

2.517	−1.259
−1.259	0.797

図 6.9　分散共分散行列の計算

回帰直線の 95%信頼区間は，$\hat{y}_i$ の分散 $Var(\hat{y}_i)$ に自由度が $f_e = n - p = 3$ の t 分布の両側 5%点，を掛けて $\hat{y}_i$ にプラス・マイナスした，

$$\hat{y}_i \text{ の 95\%信頼区間} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)} \tag{6.49}$$

で求められる．

個別データについての 95%信頼区間は，回帰直線上の $\hat{y}_i$ から誤差 $y_i - \hat{y}_i = x_i e_i$ を加えた分散から同様に求めればよい．

$$Var(\hat{y}_i) = Var(\hat{y}_i + \sqrt{x_i}e_i) = Var(\hat{y}_i) + x_i\hat{\sigma}_e^2 \tag{6.50}$$

複数データの平均の分散 $Var\ (\overline{y}_i)$ は

$$Var(\bar{y}_i) = Var(\hat{y}_i) + x_i \frac{\hat{\sigma}_e^2}{m} \tag{6.51}$$

で求めることができる

表 6.3　各種の 95%信頼区間

									$m=$	3
i	x	y	y^	$Var(y_i\hat{})$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
1	1	10	10.88	0.7972	8.03	13.72	5.87	15.88	7.17	14.58
2	1	12	10.88	0.7972	8.03	13.72	5.87	15.88	7.17	14.58
3	2	21	22.00	0.6713	19.39	24.61	15.61	28.39	17.74	26.26
4	3	30	33.13	2.1398	28.47	37.78	24.60	41.65	26.91	39.34
5	3	37	33.13	2.1398	28.47	37.78	24.60	41.65	26.91	39.34

各種の逆推定

逆推定の点推定値は，$y_0 = 25.0$ の場合，

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1} = \frac{25-(0.250)}{11.125} = 2.270$$

となり，回帰直線の 95%信頼区間についての与えられた y_0 に対する逆推定値の下限は，

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} = y_0 \tag{6.52}$$

を $\hat{x}_{L95}$ について整理すると，式（2.26）と同じ重みなしの場合と同じ 2 式式となり，

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

となり，$y_0 = 25.0$ の場合には，$t_\alpha = t(3, 0.05) = 3.182$ として，

$$a = Var(\hat{\beta}_0) - \frac{(y_0 - \hat{\beta}_0)^2}{t_\alpha^2} = 2.517 - \frac{(25-0.250)^2}{3.182^2} = 2.031$$

$$b = 2 \times Cov(\hat{\beta}_0, \hat{\beta}_1) + \frac{2 \times (y_0 - \hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} = 2 \times (-1.259) + \frac{2 \times (25 - (-0.250)) \times 11.125}{3.182^2} = 52.954$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^2}{t_\alpha^2} = 0.797 - \frac{11.125}{3.182^2}$$

この a，b，c を用いて 2 次式の根の公式から，95%信頼区間は

$$(L95R, U95R) = (2.031,\ 2.604)$$

が得られる．

個別データの逆推定の 95%信頼区間は，$\hat{x}_{L95}$ の項に $\hat{\sigma}_e^2$ の係数が増えるので，b を次のように b' として 2 次式の根の公式を適用する．

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + \hat{x}_{L95}\hat{\sigma}_e^2} = y_0$$

$$b' = b + \hat{\sigma}_e^2 = 52.954 + 1.678 = 54.632$$

$$(L95I, U95I) = (1.737,\ 3.045)$$

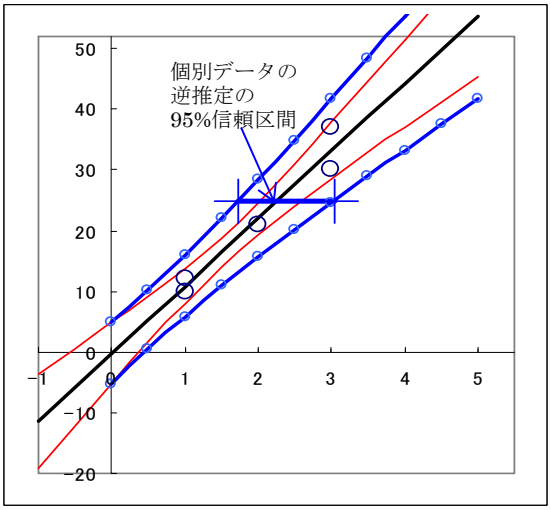
複数データの平均値の 95%信頼区間は，b を次のように b'' として 2 次式の根の公式を適用す

る．

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + \hat{x}_{L95} \frac{\hat{\sigma}_e^2}{m}} = y_0$$

$$b'' = b + \frac{\hat{\sigma}_e^2}{m} = 52.954 + \frac{1.678}{3} = 53.514$$

$$(L95m, U95m) = (1.899,\ 2.785)$$



逆推定 y_0 = 25.0　　x_0^ = 2.270

回帰直線の95%CL

a =	−60.433	x_{L95R}^ =	2.031
b =	52.954	x_{U95R}^ =	2.604
c =	−11.423		

個別データの95%CL

a =	−60.433	x_{L95I}^ =	1.737
b' =	54.632	x_{U95I}^ =	3.045
c =	−11.423		

数個の平均の95%CL

a =	−60.433	x_{L95m}^ =	1.899
b'' =	53.513	x_{U95m}^ =	2.785
c =	−11.423		
m =	3		

図 6.10　個別データの逆推定の 95%信頼区間

6.3. 分散がおおよそ x^2 に比例

小さい x に対してある測定値 y の分散には各種のノイズが発生するために，分散が x^2 に比例するとの仮定が，現実的でない場合がある．第 5 章の 5.2 節で，分散がおおよ x^2 に比例する場合について例示した．図 6.11 の右側は図 5.3 の再掲であり，左側は，分散がよ x^2 に比例する場合で，小さい x に対し Y 軸方向の変動が右側に比べ小さくなっている．

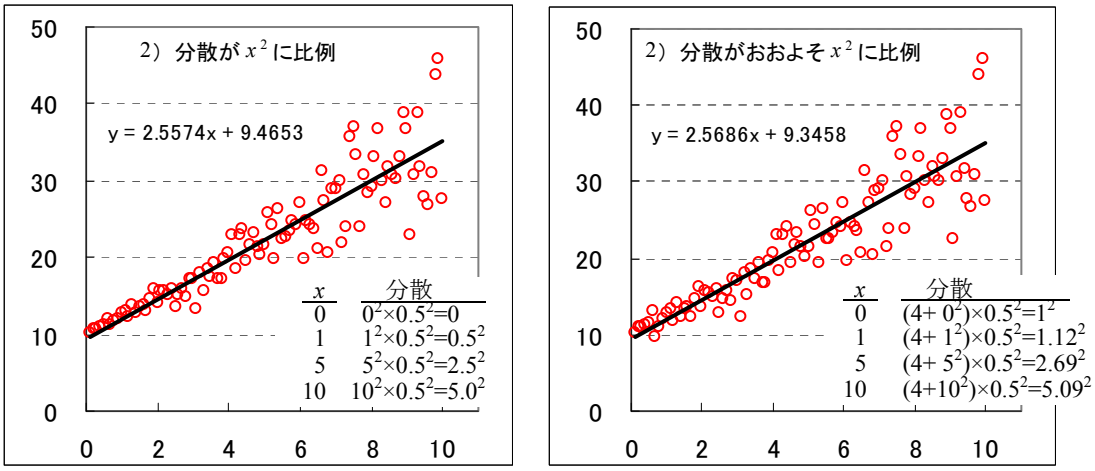


図 6.11　小さめの x にノイズが入る場合

測定値 y にノイズが入るような場合，これを考慮した重みをデータから設定する必要がある．誤差分散が x^2 に比例する場合に，ノイズを $x=0$ における誤差分散が σ_0^2 と推定された場合に，$x=1$ の場合の誤差分散を σ_e^2 としたときに，その比 $v_0=\sigma_0^2/\sigma_e^2$ を求め，ある x に対する誤差分散を $(v_0+x_i^2)\sigma_e^2$ とする．

最小 2 乗法による回帰分析では，誤差項は一つに限られているので，次式で示すように，ε_i が，平均 0，分散が $(v_0+x_i^2)\sigma_e^2$ となる正規分布に従うとする．

$$y_i=\beta_0+\beta_1 x_i+\varepsilon_i,\ i=1,2,\ldots,n,\ \varepsilon_i\sim N(0,(v_0+x_i^2)\sigma_e^2) \tag{6.53}$$

このようにすることにより，$x=0$ の場合の分散が $\sigma_0^2=v_0\sigma_e^2$，x が 0 以上の場合に $x_i^2\sigma_e^2$ が加わり，誤差分散が $(\sigma_0^2+x_i^2)\sigma_e^2$ と同等になる．誤差分散 σ_e^2 は，重み付き回帰分析で推定するのであるが，$x=0$ の場合の分散が，何らかの方法によって推定されたとしても，$v_0=\sigma_0^2/\sigma_e^2$ として計算し，定数としてあらかじめ定めなければならない．

誤差 ε_i をこの分散の平方根で割ることにより，次式で示すように重み付き最小 2 乗法の適用が可能となる．

$$y_i=\beta_0+\beta_1 x_i+\frac{\varepsilon_i}{\sqrt{v_0+x_i^2}},\ i=1,2,\ldots,n,\ \frac{\varepsilon_i}{\sqrt{v_0+x_i^2}}\sim N(0,\sigma_e^2) \tag{6.54}$$

誤差平方和 S_e は，次式のようになり，この分散の平方根で割ることにより，次式で示すように重みを $(v_0+x_i^2)$ とする最小 2 乗法の適用が可能となる．

$$S_e=\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{\sqrt{v_0+x_i^2}}\right)^2=\sum_{i=1}^{n}\left(\frac{\hat{\varepsilon}_i}{\sqrt{v_0+x_i^2}}\right)^2=\sum_{i=1}^{n}e_i^2 \tag{6.55}$$

各種の 95%信頼区間は，分散が x_i^2 に比例する場合と同様に，重みに ω_0 を加えた次式となる，

回帰直線：　$=\hat{y}_i\pm t(f_e,0.05)\sqrt{Var(\hat{y}_i)}$

個別データ：$=\hat{y}_i\pm t(f_e,0.05)\sqrt{Var(\hat{y}_i)+(v_0+x_i^2)\hat{\sigma}_e^2}$　(6.56)

$$\text{複数データ}: = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + (v_0 + x_i^2)\frac{\hat{\sigma}_e^2}{m}}$$

回帰直線に対する逆推定の 95%信頼区間は，これまでと同様に，

$$\text{回帰直線}: \ (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) \pm t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} = y_0 \tag{6.57}$$

であり，これを整理すると $\hat{x}_{L95}$ に関する 2 次式となり，2 次式の解の公式から，$\hat{x}_{L95R}$ および $\hat{x}_{U95R}$ が計算できる．個別データおよび複数データの平均の逆推定の 95%信頼区間は，2 次式の 2 つのパラメータ a および c が以下のように変わり，$\hat{x}_{L95I}$ および $\hat{x}_{U95I}$，$\hat{x}_{L95m}$ および $\hat{x}_{U95m}$ を求めることができる．

$$\text{個別データ}: (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) \pm t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + (v_0 + \hat{x}_{L95}^2)\hat{\sigma}_e^2} = y_0 \tag{6.58}$$

$$\begin{aligned} a' &= a + v_0 \hat{\sigma}_e^2 \\ c' &= c + \hat{\sigma}_e^2 \end{aligned} \tag{6.59}$$

$$\text{複数データ}: (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95m}) \pm t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95m}) + (v_0 + \hat{x}_{L95m}^2)\frac{\hat{\sigma}_e^2}{m}} = y_0 \tag{6.60}$$

$$\begin{aligned} a'' &= a + v_0 \frac{\hat{\sigma}_e^2}{m} \\ c'' &= c + \frac{\hat{\sigma}_e^2}{m} \end{aligned} \tag{6.61}$$

Excel による行列計算

これまでと同様に 5 個のデータ （x, y）＝（1, 10），（1, 12），（2, 21），（3, 30），（3, 37）が得られたとする．誤差 ε_i が互いに独立に平均 0，分散が $(v_0 + x_i^2)\sigma_e^2$ の正規分布に従うことにする．それぞれ x_i における分散は，$v_0 = 1$ としたときに，データの誤差構造は，次のようになる．

i	X	(x_i)	ε	$(v_0+x_i^2)\,e_i$	$v_0=1$
1	1	1	ε_1	$(v_0+1^2)\,e_1$	$2\,e_1$
2	1	1	ε_2	$(v_0+1^2)\,e_2$	$2\,e_2$
3	1	2	ε_3	$(v_0+2^2)\,e_3$	$5\,e_3$
4	1	3	ε_4	$(v_0+3^2)\,e_4$	$10\,e_4$
5	1	3	ε_5	$(v_0+3^2)\,e_5$	$10\,e_5$

誤差平方和 S_e は，$(v_0 + x_i^2)$ の逆数を対角要素となるような行列 V^{-1} を挟み込めば，行列計算で誤差平方和 S_e が求まる．

$$\boldsymbol{\varepsilon}^T \quad \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 & \varepsilon_5 \end{bmatrix} \qquad \boldsymbol{V}^{-1} \quad \begin{bmatrix} 1/(1+1) & 0 & 0 & 0 & 0 \\ 0 & 1/(1+4) & 0 & 0 & 0 \\ 0 & 0 & 1/(1+4) & 0 & 0 \\ 0 & 0 & 0 & 1/(1+9) & 0 \\ 0 & 0 & 0 & 0 & 1/(1+9) \end{bmatrix} \qquad \boldsymbol{\varepsilon} \quad \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

(1×5) (5×5) (5×1)

図 6.12　誤差分散が $(1+x^2)$ に比例する場合の誤差平方和の計算

さらに $\boldsymbol{\beta}$ について解くと，これまでと同様である．

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{6.62}$$

$$\boldsymbol{X}^T \quad \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 3 \end{bmatrix} \qquad \boldsymbol{V}^{-1} \quad \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix} \qquad \boldsymbol{X} \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \quad = \quad \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} \quad \begin{bmatrix} 1.400 & 2.000 \\ 2.000 & 4 \end{bmatrix}$$

(2x5) (5x5) (5x2) (2x2)

図 6.13　誤差分散が $(1+x^2)$ に比例する場合の誤差分散と分散共分散行列

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \quad \begin{bmatrix} 3.462 & -1.923 \\ -1.923 & 1.346 \end{bmatrix} \qquad \boldsymbol{X}^T \quad \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 3 \end{bmatrix} \qquad \boldsymbol{V}^{-1} \quad \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \end{bmatrix} \qquad \boldsymbol{y} \quad \begin{bmatrix} 10 \\ 12 \\ 21 \\ 30 \\ 37 \end{bmatrix} \quad = \quad \boldsymbol{\beta}^\wedge \quad \begin{bmatrix} -0.154 \\ 11.058 \end{bmatrix}$$

(2x2) (2x5) (5x5) (5x1) (2x1)

図 6.14　誤差分散が $(1+x^2)$ に比例する場合の回帰パラメータの推定

推定値の分散共分散行列は，重みなしの場合と同様に次のように計算される．

$$\boldsymbol{X} \quad (x_i) \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} \qquad \boldsymbol{\beta}^\wedge \quad \begin{bmatrix} -0.154 \\ 11.058 \end{bmatrix} \quad = \quad \boldsymbol{y}^\wedge \quad \begin{bmatrix} 10.904 \\ 10.904 \\ 21.962 \\ 33.019 \\ 33.019 \end{bmatrix}$$

$$\boldsymbol{y} \quad \begin{bmatrix} 10 \\ 12 \\ 21 \\ 30 \\ 37 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \boldsymbol{y} - \boldsymbol{y}^\wedge \quad \begin{bmatrix} 0.904 \\ -1.096 \\ 0.962 \\ 3.019 \\ -3.981 \end{bmatrix} \qquad e_i = \varepsilon_i / \sqrt{(1+x_i^2)} \quad \begin{bmatrix} 0.639 \\ -0.775 \\ 0.430 \\ 0.955 \\ -1.259 \end{bmatrix} \qquad \boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e^{2\wedge} \quad \begin{bmatrix} 4.258 & -2.366 \\ -2.366 & 1.656 \end{bmatrix}$$

$$S_e = \Sigma e_i^2 \quad 3.690$$

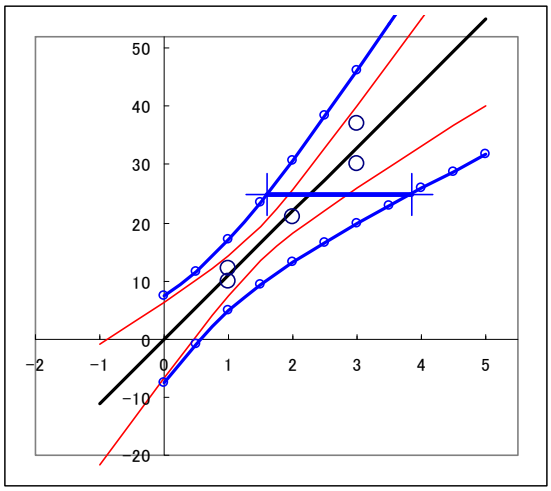
$$\sigma_e^{2\wedge} = \quad 1.230$$

(5x2) (2x1) (5x1)

図 6.15　誤差分散が $(1+x^2)$ に比例する場合の誤差分散と分散共分散行列

表 6.4 回帰パラメータ

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k$^	$Var(\beta_k$^$)$	SE	t	p
0	x_0:切片	β_0	−0.154	4.258	2.064	−0.075	0.9453
1	x_1:傾き	β_1	11.058	1.656	1.287	8.593	0.0033



逆推定 y_0 = 25.0　x_0^= 2.275

回帰直線の95%CL

a = −58.214　x_{L95R}^= 1.944

b = 50.195　x_{U95R}^= 2.874

c = −10.417

個々のデータの95%CL

a' = −56.984　x_{L95I}^= 1.609

b = 50.195　x_{U95I}^= 3.855

c' = −9.187

m 個の平均の95%CL

a'' = −57.804　x_{L95m}^= 1.791

b = 50.195　x_{U95m}^= 3.225

c'' = −10.007

図 6.16　誤差分散が $(1+x^2)$ に比例する場合の個別データの逆推定

6.4. 分散がおおよそ x に比例する場合

前節では，「分散がおおよそ x^2 に比例する場合」を取り上げた．本節では，分散が x に比例するがノイズが入る場合について取り上げる．最小 2 乗法による回帰分析では，誤差項は一つに限られているので，次式で示すように，ε_i が，平均 0，分散が $(v_0+x_i)\sigma_e^2$ となる正規分布に従うとする．なお，前節では，分散が $(v_0+x_i^2)\sigma_e^2$ と x_i の 2 乗の項となっていたが，本節では x_i のみである．

$$y_i=\beta_0+\beta_1 x_i+\varepsilon_i,\ i=1,2,\ldots,n,\ \varepsilon_i\sim N(0,(v_0+x_i)\sigma_e^2) \tag{6.63}$$

誤差 ε_i を分散の平方根で割ることにより，次式で示すように重み付き最小 2 乗法の適用が可能となる．

$$y_i=\beta_0+\beta_1 x_i+\frac{\varepsilon_i}{\sqrt{v_0+x_i}},\ i=1,2,\ldots,n,\ \frac{\varepsilon_i}{\sqrt{v_0+x_i}}\sim N(0,\sigma_e^2) \tag{6.64}$$

誤差平方和 S_e は，次式のようになり，をこの分散の平方根で割ることにより，次式で示すように重みを $(\omega_0+x_i^2)$ とする最小 2 乗法の適用が可能となる．

$$S_e=\sum_{i=1}^{n}\left(\frac{y_i-\hat{y}_i}{\sqrt{v_0+x_i}}\right)^2=\sum_{i=1}^{n}\left(\frac{\hat{\varepsilon}_i}{\sqrt{v_0+x_i}}\right)^2=\sum_{i=1}^{n}e_i^2 \tag{6.65}$$

各種の 95%信頼区間は，分散が x_i に比例する場合と同様に，重みに v_0 を加えた次式となる，

$$\text{回帰直線：}\ =\hat{y}_i\pm t(f_e,0.05)\sqrt{Var(\hat{y}_i)} \tag{6.66}$$

$$\text{個別データ：}=\hat{y}_i\pm t(f_e,0.05)\sqrt{Var(\hat{y}_i)+(v_0+x_i)\hat{\sigma}_e^2} \tag{6.67}$$

$$\text{複数データ}: = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + (v_0 + x_i)\frac{\hat{\sigma}_e^2}{m}} \tag{6.68}$$

回帰直線に対する逆推定の 95%信頼区間は，これまでと同様に，

$$\text{回帰直線}: \ (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t(f_e, 0.05)\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})} = y_0 \tag{6.69}$$

であり，前節と同様に，2 次式の解の公式から，$\hat{x}_{L95R}$ および $\hat{x}_{U95R}$ が計算できる．個別データおよび複数データの平均の逆推定の 95%信頼区間は，2 次式の 2 つのパラメータ a および b が以下のように変えて，$\hat{x}_{L95I}$ および $\hat{x}_{U95I}$，$\hat{x}_{L95m}$ および $\hat{x}_{U95m}$ を求めることができる．

$$\text{個別データ}: (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t(f_e, 0.05)\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + (v_0 + \hat{x}_{L95})\hat{\sigma}_e^2} = y_0 \tag{6.70}$$

$$a' = a + v_0 \hat{\sigma}_e^2$$
$$b' = b + \hat{\sigma}_e^2$$

$$\text{複数データ}: (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + t(f_e, 0.05)\sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) + (v_0 + \hat{x}_{L95})\frac{\hat{\sigma}_e^2}{m}} = y_0 \tag{6.71}$$

$$a'' = a + v_0 \frac{\hat{\sigma}_e^2}{m}$$
$$b'' = b + \frac{\hat{\sigma}_e^2}{m} \tag{6.72}$$

Excel による行列計算

前節と同字データ（x, y）＝（1, 10），（1, 12），（2, 21），（3, 30），（3, 37）に対して Excel の行列計算の結果を示す．

X^T (2x5)

1	1	1	1	1
1	1	2	3	3

V^{-1} (5x5)

0.5	0	0	0	0
0	0.5	0	0	0
0	0	0.33	0	0
0	0	0	0.25	0
0	0	0	0	0.25

X (5x2)

1	1
1	1
1	2
1	3
1	3

= $X^T V^{-1} X$ (2x2)

1.833	3.167
3.167	7

図 6.17　誤差分散が $(1+x)$ に比例する場合の誤差分散と分散共分散行列

$(X^T V^{-1} X)^{-1}$ (2x2)

3.462	−1.923
−1.923	1.346

X^T (2x5)

1	1	1	1	1
1	1	2	3	3

V^{-1} (5x5)

0.5	0	0	0	0
0	0.5	0	0	0
0	0	0.33	0	0
0	0	0	0.25	0
0	0	0	0	0.25

y (5x1)

10
12
21
30
37

= β^ (2x1)

−24.423
34.471

図 6.18　誤差分散が $(1+x)$ に比例する場合の回帰パラメータの推定

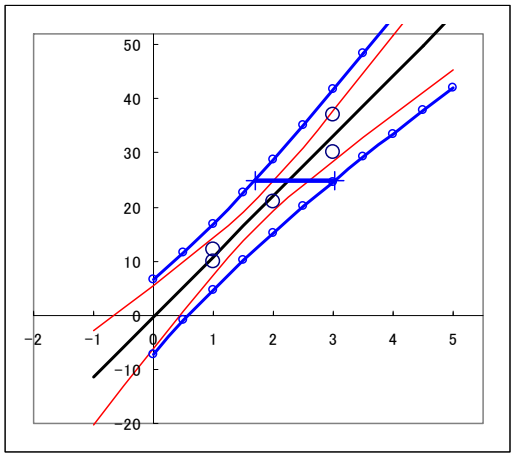
推定値の分散共分散行列は，重みなしの場合と同様に次のように計算される．

X	(x_i)	β^	=	y^	y	$\varepsilon = y - y$^	$e_i = \varepsilon_i / \sqrt{(1+x_i^2)}$	$\Sigma = (X^T X)^{-1} \sigma_e^{2}$^	
1	1	−0.154		10.904	10	0.904	0.639	4.258	−2.366
1	1	11.058		10.904	12	−1.096	−0.775	−2.366	1.656
1	2			21.962	21	0.962	0.430		
1	3			33.019	30	3.019	0.955		
1	3			33.019	37	−3.981	−1.259		
						$S_e = \Sigma e_i^2$	3.690		
(5x2)		(2x1)			(5x1)	σ_e^{2}^ =	1.230		

図 6.19　誤差分散が $(1+x)$ に比例する場合の誤差分散と分散共分散行列

表 6.5 回帰パラメータ

k	x	β_k	β_k^	$Var(\beta_k\hat{})$	SE	t	p
0	x_0:切片	β_0	−0.333	3.371	1.836	−0.182	0.868
1	x_1:傾き	β_1	11.167	0.904	0.951	11.742	0.001



逆推定 y_0 = 25.0　　x_0^= 2.269

回帰直線の95%CL

a = −59.996　x_{L95R}^= 2.022

b = 52.738　x_{U95R}^= 2.602

c = −11.407

個々のデータの95%CL

a' = −58.762　x_{L95I}^= 1.699

b' = 53.972　x_{U95I}^= 3.033

c = −11.407

m 個の平均の95%CL

a'' = −59.585　x_{L95m}^= 1.878

b'' = 53.150　x_{U95m}^= 2.781

図 6.20　誤差分散が $(1+x)$ に比例する場合の各種の 95%信頼区間

6.5. 各種の重みに対する逆推定の95%信頼区間の計算式

これまで示してきた各種の重みに対して，回帰直線の逆推定の 95%信頼区間は，どのような重みに対しても，推定された分散共分散行列に重みが考慮されているために，2 次式の根を求めるためのパラメータ（a, b, c）は変化しない．

個別データの場合には，重みが 1 の場合には，a に誤差 $\hat{\sigma}_e^2$ が加わり．x の項が入っていれば，b に誤差 $\hat{\sigma}_e^2$ が，x^2 の項が入っていれば c に誤差 $\hat{\sigma}_e^2$ が加わる．複数データの場合には，$\hat{\sigma}_e^2$ ではなく $\hat{\sigma}_e^2/m$ におきかえる．切片の分散に v_0 を含めた場合には，a に誤差 $v_0\hat{\sigma}_e^2$ が加わる．

表 6.6 各種の重みに対する逆推定のための 2 次式のパラメータ

			逆推定のための 2 次式のパラメータ		
	重み	分散	切片 a	$b\,\hat{x}_{L95}$	$c\,\hat{x}_{L95}^2$
回帰直線	1	$Var(\hat{y})$	a	b	c
	$1/x$	$Var(\hat{y})$	a	b	c
	$1/x^2$	$Var(\hat{y})$	a	b	c
	$1/(v_0+x)$	$Var(\hat{y})$	a	b	c
	$1/(v_0+x^2)$	$Var(\hat{y})$	a	b	c
個別データ	1	$Var(\hat{y}+\hat{\sigma}_e^2)$	$a+\hat{\sigma}_e^2$	b	c
	$1/x$	$Var(\hat{y}+x\hat{\sigma}_e^2)$	a	$b+\hat{\sigma}_e^2$	c
	$1/x^2$	$Var(\hat{y}+x^2\hat{\sigma}_e^2)$	a	b	$c+\hat{\sigma}_e^2$
	$1/(v_0+x)$	$Var(\hat{y}+(v_0+x)\hat{\sigma}_e^2)$	$a+v_0\hat{\sigma}_e^2$	$b+\hat{\sigma}_e^2$	c
	$1/(v_0+x^2)$	$Var(\hat{y}+(v_0+x^2)\hat{\sigma}_e^2)$	$a+v_0\hat{\sigma}_e^2$	b	$c+\hat{\sigma}_e^2$
複数 m 平均	1	$Var(\hat{y}+\hat{\sigma}_e^2/m)$	$a+\hat{\sigma}_e^2/m$	b	c
	$1/x$	$Var(\hat{y}+x\hat{\sigma}_e^2/m)$	a	$b+\hat{\sigma}_e^2/m$	c
	$1/x^2$	$Var(\hat{y}+x^2\hat{\sigma}_e^2/m)$	a	b	$c+\hat{\sigma}_e^2/m$
	$1/(v_0+x)$	$Var(\hat{y}+(v_0+x)\hat{\sigma}_e^2/m)$	$a+v_0\hat{\sigma}_e^2/m$	$b+\hat{\sigma}_e^2/m$	c
	$1/(v_0+x^2)$	$Var(\hat{y}+(v_0+x^2)\hat{\sigma}_e^2/m)$	$a+v_0\hat{\sigma}_e^2/m$	b	$c+\sigma_e^2/m$

7. 分散が x の関数で原点を通る回帰直線

7.1. 分散が x^2 に比例し原点を通る場合

回帰分析の行列による表記

次の 5 個のデータ $(x, y) = (1, 10)$, $(1, 12)$, $(2, 21)$, $(3, 30)$, $(3, 37)$ が得られたとき, 標準偏差を σ_e としたときに, 誤差 ε_i が $x_i\sigma_e$ のように比例的に大きくなり, 原点を通る重み付き回帰分析を考える.

$$y_i = \beta_1 x_i + \varepsilon_i\ ,\ \ i = 1,2,...,5\ ,\ \ \varepsilon_i \sim N(0, (x_i\sigma_e)^2) \tag{7.1}$$

それぞれの ε_i に x_i の逆数を掛ければ,

$$y_i = \beta_1 x_i + \frac{\varepsilon_i}{x_i}\ ,\ \ i = 1,2,...,5\ ,\ \ \frac{\varepsilon_i}{x_i} \sim N(0, \sigma_e^2) \tag{7.2}$$

となり, 誤差分散を σ_e^2 と一定化できる. 最小にしたい誤差平方和 S_e は,

$$S_e = \sum_{i=1}^{5} \frac{1}{x_i^2}\varepsilon_i^2$$

であるので, x_i の 2 乗の逆数を対角要素となるような行列 $\boldsymbol{V}^{-1}$ を挟み込めば, 行列計算で誤差平方和 S_e が求まる.

$$\underset{(1\times5)}{\overset{\boldsymbol{\varepsilon}^T}{\begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \varepsilon_4 & \varepsilon_5 \end{bmatrix}}}\ \underset{(5\times5)}{\overset{\boldsymbol{V}^{-1}}{\begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/2^2 & 0 & 0 \\ 0 & 0 & 0 & 1/3^2 & 0 \\ 0 & 0 & 0 & 0 & 1/3^2 \end{bmatrix}}}\ \underset{(5\times1)}{\overset{\boldsymbol{\varepsilon}}{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}}}$$

正規方程式を解くと, $\boldsymbol{V}^{-1}$ を挟み込んだ次式が得られる.

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}\,\boldsymbol{\beta} = \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{7.3}$$

さらに $\boldsymbol{\beta}$ について解くために $(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ を掛けて,

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{7.4}$$

が得られる. パラメータ $\boldsymbol{\beta}$ の推定値 $\hat{\boldsymbol{\beta}}$ は, 次に示すように求めることができる. 注) 実際の計算には分数ではなく実数で与える.

$$\underset{(1\times5)}{\boldsymbol{X}^T}\ \underset{(5\times5)}{\boldsymbol{V}^{-1}}\ \underset{(5\times1)}{\boldsymbol{X}} = \underset{(1\times1)}{\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}}$$

$$\begin{bmatrix} 1 & 1 & 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/2^2 & 0 & 0 \\ 0 & 0 & 0 & 1/3^2 & 0 \\ 0 & 0 & 0 & 0 & 1/3^2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 5.000 \end{bmatrix}$$

$$\underset{(1\times1)}{(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}}\ \underset{(1\times5)}{\boldsymbol{X}^T}\ \underset{(5\times5)}{\boldsymbol{V}^{-1}}\ \underset{(5\times1)}{\boldsymbol{y}} = \underset{(1\times1)}{\boldsymbol{\beta}\hat{}}$$

$$\begin{bmatrix} 0.200 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/2^2 & 0 & 0 \\ 0 & 0 & 0 & 1/3^2 & 0 \\ 0 & 0 & 0 & 0 & 1/3^2 \end{bmatrix} \begin{bmatrix} 10 \\ 12 \\ 21 \\ 30 \\ 37 \end{bmatrix} = \begin{bmatrix} 10.967 \end{bmatrix}$$

図 7.1　原点を通る重み x^2 の回帰のパラメータ推定

推定値の分散共分散行列は，重みなしの場合と同様に次のように計算される．

$$\underset{(5\times1)}{\boldsymbol{X}}\ \underset{(2\times1)}{\boldsymbol{\beta}\hat{}} = \underset{(2\times1)}{\boldsymbol{y}\hat{}}\ \underset{(5\times1)}{\boldsymbol{y}}\ \underset{(5\times1)}{\boldsymbol{\varepsilon}\hat{}=\boldsymbol{y}\hat{}-\boldsymbol{y}}\ \underset{(5\times1)}{\boldsymbol{e}=\boldsymbol{\varepsilon}\hat{}/x_i}\ \underset{(2\times2)}{\boldsymbol{\Sigma}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e^{2\hat{}}}$$

$$\begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} \begin{bmatrix} 10.967 \end{bmatrix} = \begin{bmatrix} 10.97 \\ 10.97 \\ 21.93 \\ 32.90 \\ 32.90 \end{bmatrix} \begin{bmatrix} 10 \\ 12 \\ 21 \\ 30 \\ 37 \end{bmatrix} \begin{bmatrix} 0.967 \\ -1.033 \\ 0.933 \\ 2.900 \\ -4.100 \end{bmatrix} \begin{bmatrix} 0.967 \\ -1.033 \\ 0.467 \\ 0.967 \\ -1.367 \end{bmatrix} \begin{bmatrix} 0.251 \end{bmatrix}$$

$$S_e = \boldsymbol{e}^T\boldsymbol{e} = 5.022$$

$$\boldsymbol{\sigma}_e^{2\hat{}} = 1.256$$

図 7.2 原点を通る重み x^2 の回帰の分散共分散行列

表 7.1　原点を通る x^2 の重み付き回帰パラメータの t 検定

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k\hat{}$	$Var(\beta_k\hat{})$	SE	t	p
1	x_1:傾き	β_1	10.967	0.251	0.501	21.885	0.000

各種の 95%信頼区間

回帰直線の 95%信頼区間は，$\hat{y}_i$ の分散 $Var(\hat{y}_i)$ に自由度が $f_e = n - p = 4$ の t 分布の両側 5%点を掛けて $\hat{y}_i$ にプラス・マイナスした，

$$\hat{y}_i \text{ の 95\%信頼区間} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)} \tag{7.5}$$

個別データの 95%信頼区間は，回帰直線上の $\hat{y}_i$ から誤差分散 $(x_i e_i)^2$ を加えればよい．

$$Var(\hat{y}_i + x_i e_i) = Var(\hat{y}_i) + x_i^2 \hat{\sigma}_e^2 \tag{7.6}$$

複数データの平均に対する分散 $Var\ (\bar{y}_{i\bullet})$ は，データ数 m で誤差分散を除して，次のようになる．

$$Var(\bar{y}_{i\bullet}) = Var(\hat{y}_i) + x_i^2 \frac{\sigma_e^2}{m} \tag{7.7}$$

表 7.2 原点を通る x^2 の重み付き回帰の各種 95%信頼区間

$m=$ 3

i	x	y	$Var(y_i\hat{})$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
1	1	10	0.251	9.575	12.358	7.559	14.375	8.695	13.239
2	1	12	0.251	9.575	12.358	7.559	14.375	8.695	13.239
3	2	21	1.004	19.151	24.716	15.117	28.749	17.389	26.477
4	3	30	2.260	28.726	37.074	22.676	43.124	26.084	39.716
5	3	37	2.260	28.726	37.074	22.676	43.124	26.084	39.716

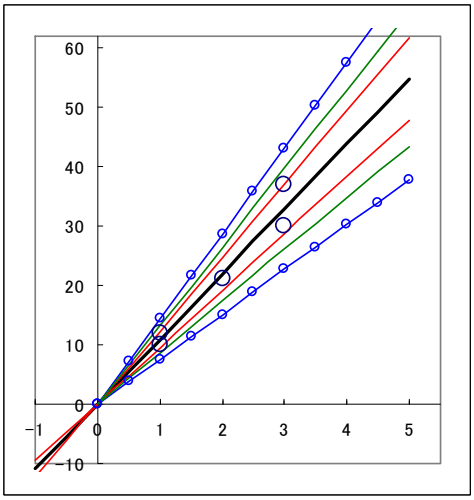


図 7.3　原点を通る x^2 の重み付き回帰の各種の 95%信頼区間

ここでの回帰直線の各種の 95%信頼区間は，全て原点を通る直線となっている．回帰式の 95%信頼区間は，3.1 節と同様に，$\hat{y}_i$ の推定値の SE が，

$$\sqrt{Var(\hat{y}_i)} = \sqrt{x_i^2 Var(\hat{\beta}_1)} = x_i \sqrt{Var(\hat{\beta}_1)}$$

と x_i に比例するためである．個別データの場合は，回帰の 95%信頼区間は，

$$\sqrt{Var(\hat{y}_i) + x_i^2 \hat{\sigma}_e^2} = \sqrt{x_i^2 Var(\hat{\beta}_1) + x_i^2 \hat{\sigma}_e^2} = x_i \sqrt{Var(\hat{\beta}_1) + \hat{\sigma}_e^2}$$

と x_i に比例するためである．複数データの場合も同様に x_i に比例する．

各種の逆推定の 95%信頼区間

逆推定　回帰

$$\hat{\beta}_1 \hat{x}_{L95} + t_\alpha \sqrt{Var(\hat{\beta}_1 \hat{x}_{L95})} = y_0 \tag{7.8}$$

2 次式　：

$$a = -\frac{(y_0)^2}{t_\alpha^2} \tag{7.9}$$

$$b = \frac{2y_0\hat{\beta}_1}{t_\alpha^2} \tag{7.10}$$

$$c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1^{\,2}}{t_\alpha^2} \tag{7.11}$$

個別データ

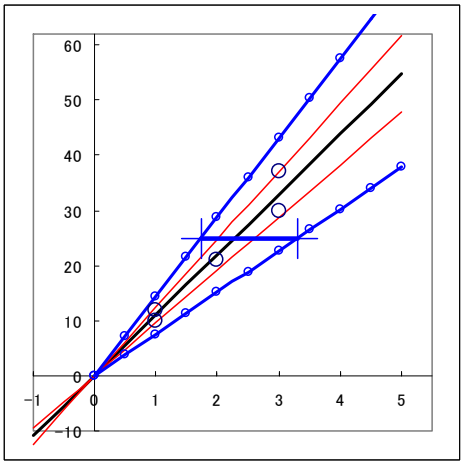
$$\hat{\beta}_1\hat{x}_{L95\mathrm{I}} + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95\mathrm{I}}) + \hat{x}_{L95\mathrm{I}}^2\hat{\sigma}_e^2} = y_0$$

$$c' = c + \hat{\sigma}_e^2$$

m 個のデータ

$$(\hat{\beta}_1\hat{x}_{L95\mathrm{I}}) + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95\mathrm{I}}) + \hat{x}_{L95\mathrm{I}}^2\frac{\hat{\sigma}_e^2}{m}} = y_0$$

$$c'' = c + \frac{\hat{\sigma}_e^2}{m}$$



逆推定 y_0 = 25.0　x_0^ = 2.280

回帰直線の95%CL

a = −81.078　x_{L95R} = 2.023

b = 71.132　x_{U95R} = 2.611

c = −15.351

個々のデータの95%CL

a = −81.078　x_{L95I} = 1.739

b = 71.132　x_{U95I} = 3.307

c' = −14.095

m 個の平均の95%CL

a = −81.078　x_{L95m} = 1.888

b = 71.132　x_{U95m} = 2.875

c'' = −14.932

図 7.4　$y_0 = 25$ における個別データの逆推定

7.2. 分散が x に比例し原点を通る直線

新たに 5 個のデータ $(x, y) = (1, 10), (1, 12), (2, 21), (3, 30), (3, 37)$ が得られたとき，誤差 ε_i が $x_i\sigma_e$ のように比例的に大きくなるとする．

$$y_i = \beta_1 x_i + \varepsilon_i\,,\ \ i = 1,2,\ldots,5\,,\ \ \varepsilon_i \sim N(0, x_i\sigma_e^2) \tag{7.12}$$

誤差 ε_i が互いに独立に平均 0，分散が $x_i\sigma_e^2$ の正規分布に従うことになる．これは，誤差 ε_i が，$x_i = 1$ の場合：$(1\times\sigma_e)^2 = \sigma_e^2$，$x_i = 2$ の場合：$2\sigma_e^2$，$x_i = 3$ の場合：$3\sigma_e^2$ となることを意味する．

それぞれの ε_i に $\sqrt{x_i}$ の逆数を掛ければ,

$$y_i = \beta_1 x_i + \frac{\varepsilon_i}{\sqrt{x_i}}\ ,\ \ i = 1,2,...,5\ ,\ \ \frac{\varepsilon_i}{\sqrt{x_i}} \sim N(0, \sigma_e^2) \tag{7.13}$$

誤差分散を $x_i\sigma_e^2$ と一定化できる. 最小にしたい誤差平方和 S_e は,

$$S_e = \sum_{i=1}^{5} \frac{1}{x_i}\varepsilon_i^2$$

であるので, x_i の逆数を対角要素となるような行列 $\boldsymbol{V}^{-1}$ を挟み込めば, 行列計算で誤差平方和 S_e が求まる. パラメータ $\boldsymbol{\beta}$ の推定値 $\hat{\boldsymbol{\beta}}$ は, $\boldsymbol{V}^{-1}$ を置き換え前節と同様に計算することができる.

$$\underset{(1\times5)}{\boldsymbol{X}^T} \quad \underset{(5\times5)}{\boldsymbol{V}^{-1}} \quad \underset{(5\times1)}{\boldsymbol{X}} = \underset{(1\times1)}{\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}}$$

$$\begin{bmatrix} 1 & 1 & 2 & 3 & 3 \end{bmatrix} \quad \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/3 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \end{bmatrix}$$

$$\underset{(1\times1)}{(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}} \ \underset{(1\times5)}{\boldsymbol{X}^T} \quad \underset{(5\times5)}{\boldsymbol{V}^{-1}} \quad \underset{(5\times1)}{\boldsymbol{y}} = \underset{(1\times1)}{\boldsymbol{\beta}\hat{}}$$

$$\begin{bmatrix} 0.100 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 2 & 3 & 3 \end{bmatrix} \quad \begin{bmatrix} 1/1 & 0 & 0 & 0 & 0 \\ 0 & 1/1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/3 \end{bmatrix} \quad \begin{bmatrix} 10 \\ 12 \\ 21 \\ 30 \\ 37 \end{bmatrix} = \begin{bmatrix} 11.000 \end{bmatrix}$$

図 7.5　原点を通る重み x の回帰のパラメータ推定

$\boldsymbol{X}$	$\boldsymbol{\beta}\hat{}$	=	$\boldsymbol{y}\hat{}$	$\boldsymbol{y}$	$\boldsymbol{\varepsilon}=\boldsymbol{y}\hat{}-\boldsymbol{y}$	$\boldsymbol{e}=\boldsymbol{\varepsilon}/x_i$	$\boldsymbol{\Sigma}=(\boldsymbol{X}^T\boldsymbol{X})^{-1}\sigma_e{}^{2\hat{}}$
1	11.000		11.00	10	1.000	1.000	0.126
1			11.00	12	−1.000	−1.000	
2			22.00	21	1.000	0.500	
3			33.00	30	3.000	1.000	
3			33.00	37	−4.000	−1.333	
					$S_e=\boldsymbol{e}^T\boldsymbol{e}=$	5.028	
(5x1)	(2x1)		(5x1)		$\sigma_e{}^{2\hat{}}=$	1.257	

図 7.6　原点を通る重み x の回帰の分散共分散行列

回帰直線の 95%信頼区間は, $\hat{y}_i$ の分散 $Var(\hat{y}_i)$ に自由度が $f_e = n - p = 4$ の t 分布の両側 5%点を掛けて $\hat{y}_i$ にプラス・マイナスした,

$$\hat{y}_i \text{ の 95\%信頼区間} = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)} \tag{7.14}$$

個別データについての 95%信頼区間は，回帰直線上の $\hat{y}_i$ から誤差 $\sqrt{x_i}e_i$ を加えた分散を求めればよい．この場合，$\hat{y}_i$ と e_i の共分散は 0 であり，誤差 e_i の分散は $\hat{\sigma}_e^2$ なので，

$$Var(\hat{y}_i + \sqrt{x_i}e_i) = Var(\hat{y}_i) + x_i\hat{\sigma}_e^2 \tag{7.15}$$

となる．複数データの平均の分散 $Var\ (\overline{y}_{i\bullet})$ は

$$Var(\overline{y}_{i\bullet}) = Var(\hat{y}_i) + x_i\frac{\sigma_e^2}{m} \tag{7.16}$$

表 7.3　原点を通る x の重み付き回帰の各種 95%信頼区間

								$m=$	3
i	x	y	$Var(y_i\hat{})$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
1	1	10	0.1257	10.016	11.984	7.735	14.265	8.951	13.049
2	1	12	0.1257	10.016	11.984	7.735	14.265	8.951	13.049
3	2	21	0.5028	20.031	23.969	17.178	26.822	18.785	25.215
4	3	30	1.1313	30.047	35.953	26.853	39.147	28.709	37.291
5	3	37	1.1313	30.047	35.953	26.853	39.147	28.709	37.291

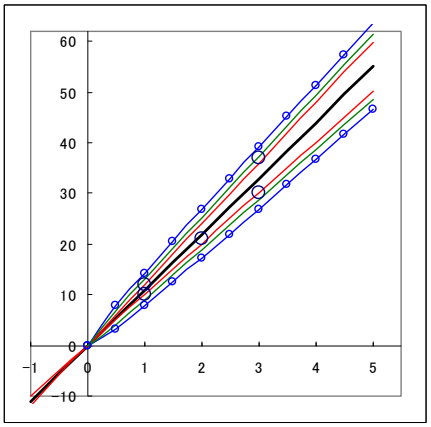


図 7.7

回帰式の 95%信頼区間が直線になるのは，3.1 節と同様に，回帰の推定値 $\hat{y}_i$ の SE が，傾き $\hat{\beta}_1$ の SE に x_i を掛けているためである．

$$\hat{y}_i \text{ の SE} = \sqrt{Var(\hat{y}_i)} = \sqrt{x_i^2 Var(\hat{\beta}_1)} = x_i\sqrt{Var(\hat{\beta}_1)}$$

個別データの 95%信頼区間が紡錘型となるのは，$\hat{y}_i$ の SE に x_i の平方根となる成分が入っているためである．

$$\text{個別データの } SE = \sqrt{Var(\hat{y}_i) + x_i\hat{\sigma}_e^2} = \sqrt{x_i^2 Var(\hat{\beta}_1) + x_i\hat{\sigma}_e^2} = \sqrt{x_i}\sqrt{x_i Var(\hat{\beta}_1) + \hat{\sigma}_e^2}$$

各種の逆推定の 95%信頼区間

逆推定 回帰

$$\hat{\beta}_1\hat{x}_{L95} + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95})} = y_0 \tag{7.17}$$

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

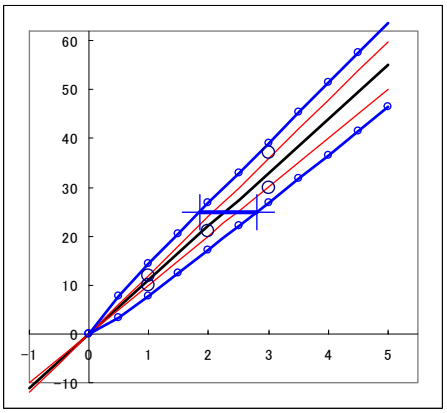
$$a = -\frac{(y_0)^2}{t_\alpha^2}, \quad b = \frac{2y_0\hat{\beta}_1}{t_\alpha^2}, \quad c = Var(\hat{\beta}_1) - \frac{\hat{\beta}_1{}^2}{t_\alpha^2} \tag{7.18}$$

個別データ

$$\begin{aligned} &\hat{\beta}_1\hat{x}_{L95} + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95}) + \hat{x}_{L95}\hat{\sigma}_e^2} = y_0 \\ &b' = b + \hat{\sigma}_e^2 \end{aligned} \tag{7.19}$$

複数データ

$$\begin{aligned} &\hat{\beta}_1\hat{x}_{L95} + t_\alpha\sqrt{Var(\hat{\beta}_1\hat{x}_{L95}) + \hat{x}_{L95}\frac{\hat{\sigma}_e^2}{m}} = y_0 \\ &b'' = b + \frac{\hat{\sigma}_e^2}{m} \end{aligned} \tag{7.20}$$



逆推定 y_0 = 25.0　　x_0^=

回帰直線の95%CL

a =	−81.078	x_{L95} =	2.086
b =	71.348	x_{U95} =	2.496
c =	−15.571		

個々のデータの95%CL

a =	−81.078	x_{L95I} =	1.853
b' =	72.605	x_{U95I} =	2.810
c =	−15.571		

m 個の平均の95%CL

a =	−81.078	x_{L95m} =	1.982
b'' =	71.767	x_{U95m} =	2.627
c =	−15.571		

図 7.8　$y_0 = 25$ における個別データの逆推定

8.　曲線となる検量線

8.1.　データの例示

第 4 章で直線があてはまりそうもない場合に，2 次式をあてはめることは，各種の 95%信頼区間が 4 次式となり，逆推定値の 95%信頼区間が容易に求められなくなるので，$y=\beta_0+\beta_1 x^2$ のように，x^2 に関して 1 次式となるようなあてはめを推奨したのであるが，1 次の項がなくなるために，さまざまな現実のデータに対して適応性が欠けてしまう．そこで，$y=\beta_0+\beta_1 x^\lambda$ のように，x^λ に関して 1 次式となるようなあてはめに拡張する．

重みに関しては，これまで，x^2 または x のどちらかを用いる場合について例示をしてきたのであるが，$y=\beta_0+\beta_1 x^\lambda$ の場合には，重みは $x^{2\lambda}$ または x^λ のどちらかを使うことと置き換えればよい．小さな x^λ に対してノイズが入る場合は，6.3 節，6.4 節で示したように $(v_0+x^{2\lambda})$ または (v_0+x^λ) を重みとすることにより，各種の信頼区間の推定を容易に行なえる．

第 5.4 節で，直線のあてはめが現実的ではないデータを用いて，回帰直線のあてはめ例を示した．

表 8.1　直線があてはまらない 100 個のデータ（6）

x	y	x	y	x	y	x	y	x	y
0.1	9.8	2.1	12.9	4.1	19.9	6.1	15.7	8.1	32.0
0.2	11.3	2.2	12.7	4.2	14.5	6.2	21.3	8.2	36.7
0.3	10.4	2.3	11.5	4.3	19.7	6.3	20.8	8.3	28.3
0.4	10.7	2.4	12.8	4.4	20.9	6.4	20.2	8.4	24.7
0.5	10.5	2.5	8.5	4.5	15.5	6.5	17.1	8.5	30.7
0.6	13.2	2.6	11.0	4.6	18.1	6.6	29.2	8.6	29.1
0.7	6.7	2.7	12.1	4.7	19.9	6.7	24.5	8.7	28.5
0.8	8.7	2.8	10.3	4.8	17.7	6.8	16.5	8.8	32.1
0.9	10.2	2.9	14.0	4.9	16.3	6.9	26.4	8.9	39.4
1.0	11.4	3.0	13.5	5.0	17.7	7.0	26.6	9.0	37.1
1.1	12.0	3.1	7.7	5.1	22.9	7.1	27.7	9.1	19.3
1.2	8.8	3.2	14.6	5.2	21.1	7.2	17.9	9.2	29.3
1.3	12.8	3.3	11.1	5.3	15.5	7.3	20.6	9.3	40.0
1.4	9.2	3.4	15.0	5.4	23.3	7.4	34.8	9.4	30.9
1.5	10.7	3.5	13.5	5.5	18.8	7.5	36.4	9.5	25.8
1.6	11.0	3.6	16.0	5.6	18.9	7.6	32.1	9.6	24.3
1.7	8.9	3.7	12.8	5.7	19.9	7.7	20.6	9.7	29.7
1.8	12.0	3.8	12.7	5.8	21.4	7.8	28.8	9.8	46.7
1.9	14.4	3.9	16.1	5.9	20.7	7.9	26.0	9.9	49.6
2.0	10.2	4.0	17.2	6.0	24.1	8.0	27.0	10.0	25.6

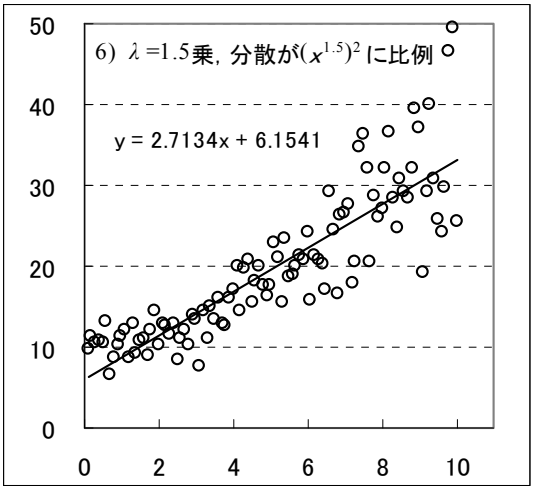


図 8.1　直線のあてはめが現実的でない（図 5.9 右端 再掲）

この例に対して，回帰式を $y=\beta_0+\beta_1 x^{1.5}$ とし，重みを $1/(100+x_i^{2\times1.5})$ とした回帰分析の結果を図 5.13 として示した．

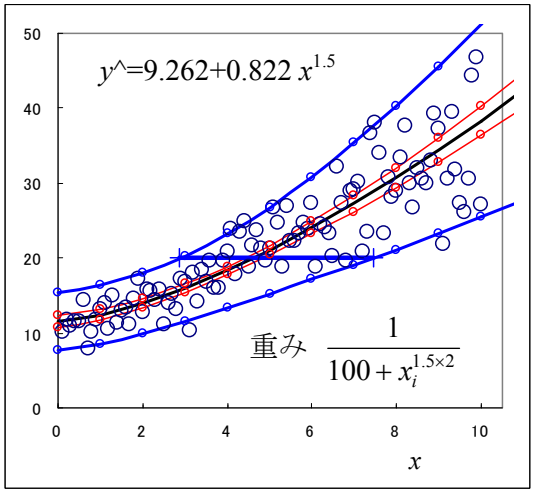


図 8.2　曲線となる重み付き回帰（図 5.13 右 再掲）

Excel による重みつき回帰の実用的な解析手順

データ数 $n=100$ と多くなると行列関数を用いた重み付き回帰では大きさ 100×100 の逆行列 $\boldsymbol{V}^{-1}$ を Excel シート上に手作業で展開するのは手間がかかり，操作ミスを起こしやすくなる．逆行列 $\boldsymbol{V}^{-1}$ は対角要素に重み w_i，他の要素は全て 0 なので，対角要素を全て左詰めにして大きさ 100×1 の行列にする．これは，Excel にはベクトルを対角化する関数がないので，便宜的な対応をとする．

$\boldsymbol{V}^{-1}$ を用いるのは，回帰パラメータ $\boldsymbol{\beta}$ の推定，

$$\hat{\boldsymbol{\beta}}=(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \tag{8.1}$$

する際に用いられ，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X}$ は，

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} = \begin{bmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} w_i x_i \\ \sum_{i=1}^{n} w_i x_i & \sum_{i=1}^{n} w_i x_i^2 \end{bmatrix} \tag{8.2}$$

なので，大きさ100×2のデザイン行列$\boldsymbol{X}$の行方向の 2 つの要素に，対応する重みw_iを掛けた大きさ100×2の行列$[\boldsymbol{wX}]$を求めて，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} = \boldsymbol{X}^T[\boldsymbol{wX}]$と置き換える．$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$は，

$$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} = \begin{bmatrix} \sum_{i=1}^{n} w_i y_i \\ \sum_{i=1}^{n} w_i x_i y_i \end{bmatrix} \tag{8.3}$$

なので，大きさ100×1の$\boldsymbol{y}$の行方向の要素に，対応する重みw_iを掛けた大きさ100×1の行列$[\boldsymbol{wy}]$を求めて，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} = \boldsymbol{X}^T[\boldsymbol{wy}]$ と置き換える．これらの計算方法により，手作業による対角化をする必要がなくなる．

用いる回帰式

用いる回帰式は，4.2 節で取り上げた

$$y = \beta_0 + \beta_1 x^{\lambda} \tag{8.4}$$

x^{λ}に関して 1 次式とする．べきλは 0.5～2 程度が現実的な選択範囲である．ここでは，$\lambda = 1.5$の場合について例示する．表 8.1 に示した 100 個のデータを(x_i, y_i)とし，解析に用いるデザイン行列$\boldsymbol{X}$には，切片を求めるために 1 を，傾きを求めるためにx_i^{λ}を計算する．重みは，$w_i = 1/(v_0 + x_i^2) = 1/(100 + x_i^2)$とする．

表 8.2 べき乗（λ=1.5）に関して線形モデル

			$\boldsymbol{X}$		
i	x_i	y_i	切片	x_i^{λ}	$\boldsymbol{w}$
1	0.1	9.8	1	0.03	0.010
2	0.2	11.3	1	0.09	0.010
3	0.3	10.4	1	0.16	0.010
4	0.4	10.7	1	0.25	0.010
5	0.5	10.5	1	0.35	0.010
99	9.9	49.6	1	31.15	0.001
100	10.0	25.6	1	31.62	0.001

パラメータ$\boldsymbol{\beta}$を推定のために$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X} = \boldsymbol{X}^T[\boldsymbol{wX}]$，その逆行列$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$，$\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} = \boldsymbol{X}^T[\boldsymbol{wy}]$を計算する．Excel では，要素ごとの掛け算は，（$\boldsymbol{w}$の範囲）*（$\boldsymbol{X}$の範囲）のように加減算と同様の手順で計算できる．これらを変更すれば，これまでの計算手順と同様である．

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \quad \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \quad = \quad \overline{\boldsymbol{\beta}\hat{}}$$

$$\begin{bmatrix} 3.850 & -0.274 \\ -0.274 & 0.040 \end{bmatrix} \begin{bmatrix} 7.58 \\ 72.79 \end{bmatrix} = \begin{bmatrix} 9.260 \\ 0.822 \end{bmatrix}$$

回帰パラメータ $\hat{\boldsymbol{\beta}}$ が求まれば，x_i^{λ} に関して一次式の推定値 $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$ を求め，誤差 $\hat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \hat{\boldsymbol{y}}$，重みの平方根を掛けた $e_i = \sqrt{w_i}\hat{\varepsilon}_i$ を求める．

表 8.3 べき乗（λ=1.5）に関しての回帰分析

i	x_i	y_i	$\boldsymbol{X}$ 切片	$\boldsymbol{X}$ x_i^{λ}	$\overline{\boldsymbol{\beta}\hat{}}$		$\boldsymbol{y}\hat{}$	$\boldsymbol{\varepsilon}\hat{}$	$\boldsymbol{w}$	$\boldsymbol{e}$
1	0.1	9.8	1	0.03	9.260	=	9.3	0.514	0.010	0.051
2	0.2	11.3	1	0.09	0.822		9.3	1.966	0.010	0.197
3	0.3	10.4	1	0.16			9.4	1.005	0.010	0.100
4	0.4	10.7	1	0.25			9.5	1.232	0.010	0.123
5	0.5	10.5	1	0.35			9.6	0.949	0.010	0.095
99	9.9	49.6	1	31.15			34.9	14.728	0.001	0.450
100	10.0	25.6	1	31.62			35.3	−9.661	0.001	−0.291

誤差分散は，$\hat{\sigma}_e^2 = \boldsymbol{e}^T\boldsymbol{e}/f_e = \boldsymbol{e}^T\boldsymbol{e}/(100-2) = 0.0426$ となり，分散共分散行列 $\boldsymbol{\Sigma} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\hat{\sigma}_e^2$ を計算する．

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0.164 & -0.012 \\ -0.012 & 0.002 \end{bmatrix}$$

$$\sigma_e{}^2\hat{} = \begin{bmatrix} 0.0426 \end{bmatrix}$$

回帰パラメータ $\hat{\beta}_k$ の誤差分散は，$\boldsymbol{\Sigma}$ の対角要素に計算されているので，それを用いて，次のように t 統計量，p 値を求めることができる．

表 8.4 べき乗（λ=1.5）に関しての回帰パラメータ

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k\hat{}$	$Var(\beta_k\hat{})$	SE	t	p
0	x_0:切片	β_0	9.260	0.164	0.405	22.865	0.000
1	x_1:傾き	β_1	0.822	0.002	0.041	19.956	0.000

各種の 95%信頼区間のグラフ表示のための計算を別途おこなう．それぞれの推定値は，次式で求める．

$\hat{y}_i$ の推定値：　$\hat{y}_i = \boldsymbol{x}_i\hat{\boldsymbol{\beta}}$，

回帰の分散：　$Var(\hat{y}_i) = \boldsymbol{x}_i\boldsymbol{\Sigma}\boldsymbol{x}_i^T$，

回帰の 95%CL：　$(L95R, U95R) = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i)}$，

個別データの 95%CL：$(L95I, U95I) = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + (v_0 + x_i^{2\lambda})\hat{\sigma}_e^2}$，

複数データの 95%CL：$(L95m, U95m) = \hat{y}_i \pm t(f_e, 0.05)\sqrt{Var(\hat{y}_i) + (v_0 + x_i^{2\lambda})\dfrac{\hat{\sigma}_e^2}{m}}$，

ただし，$\boldsymbol{x}_i = [1 \quad x_i^{\lambda}]$.

表 8.5　各種の 95%信頼区間

x_i	切片	x_i^{λ}	y^	$Var(y_i$^$)$	$L95R$	$U95R$	$L95I$	$U95I$	$L95m$	$U95m$
0.1	1	0.03	9.3	0.163	8.484	10.088	5.112	13.460	6.789	11.783
1.0	1	1.00	10.1	0.142	9.333	10.831	5.898	14.266	7.590	12.574
2.0	1	2.83	11.6	0.112	10.923	12.248	7.278	15.894	9.040	14.131
3.0	1	5.20	13.5	0.089	12.942	14.123	8.879	18.186	10.803	16.262
4.0	1	8.00	15.8	0.086	15.256	16.419	10.560	21.115	12.754	18.922
5.0	1	11.18	18.5	0.115	17.779	19.126	12.272	24.634	14.842	22.063
6.0	1	14.70	21.3	0.188	20.485	22.204	14.012	28.676	17.053	25.635
7.0	1	18.52	24.5	0.314	23.376	25.600	15.795	33.180	19.388	29.588
8.0	1	22.63	27.9	0.505	26.454	29.275	17.634	38.095	21.847	33.883
9.0	1	27.00	31.5	0.771	29.717	33.203	19.538	43.382	24.431	38.488
10.0	1	31.62	35.3	1.123	33.157	37.364	21.514	49.008	27.140	43.381
11.0	1	36.48	39.3	1.572	36.769	41.745	23.564	54.950	29.971	48.542
12.0	1	41.57	43.4	2.127	40.545	46.333	25.689	61.189	32.922	53.956

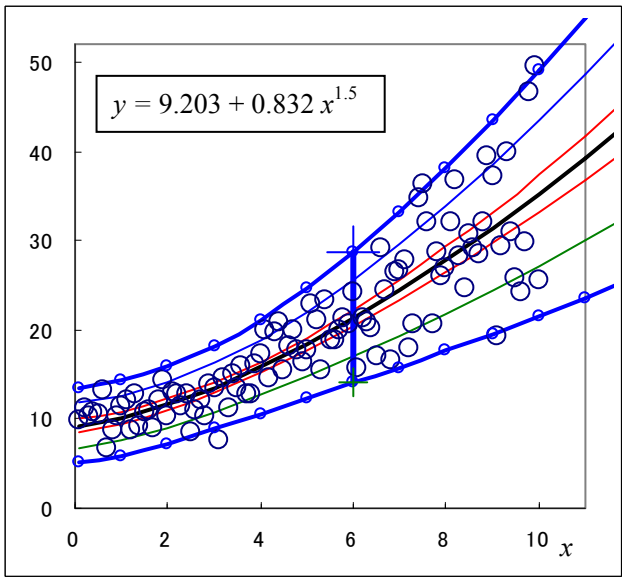


図 8.3　べきが 1.5 の回帰曲線に対する個別データの 95%信頼区間

各種の逆推定

べきλに関して 1 次式$y = \beta_0 + \beta_1 x^{\lambda}$の場合は，表 8.6 に示すように$x$を$x^{\lambda}$に置き換えればよい.

表 8.6　逆推定のための 2 次式のパラメータ

	重み	分散	逆推定のための 2 次式のパラメータ 切片 a	$b\,\hat{x}_{L95}^{\lambda}$	$c\,\hat{x}_{L95}^{2\lambda}$
回帰直線	$1/(v_0+x_i^{2\lambda})$	$Var(\hat{y}_i)$	a	b	c
個別データ	$1/(v_0+x_i^{2\lambda})$	$Var(\hat{y}_i+(v_0+x_i^{2\lambda})\hat{\sigma}_e^2)$	$a+v_0\hat{\sigma}_e^2$	b	$c+\hat{\sigma}_e^2$
複数 m 平均	$1/(v_0+x_i^{2\lambda})$	$Var(\hat{y}_i+((v_0+x_i^{2\lambda})\hat{\sigma}_e^2/m)$	$a+v_0\hat{\sigma}_e^2/m$	b	$c+\sigma_e^2/m$

表 8.6 に示した逆推定のための 2 次式のパラメータには，一定の規則性があり，Excel で自動的に計算させることができる．回帰直線の 95%信頼区間についての与えられた y_0 に対する逆推定値の下限は，

$$(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95}^{\lambda})+t_\alpha\sqrt{Var(\hat{\beta}_0+\hat{\beta}_1\hat{x}_{L95}^{\lambda})}=y_0 \tag{8.5}$$

を $\hat{x}_{L95}^{\lambda}$ について整理すると，これまでと同様の 2 次式のパラメータが得られる．

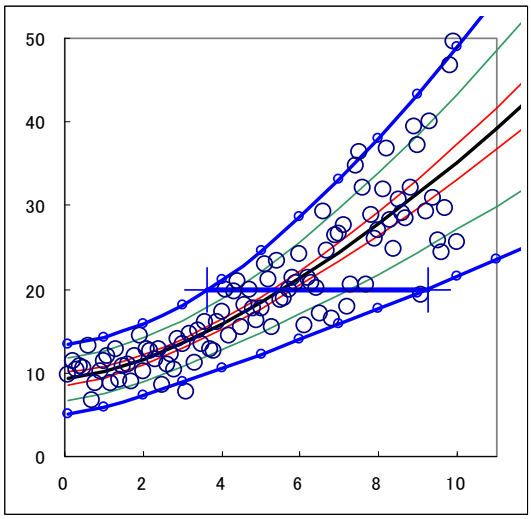
$$\text{2 次式：}\quad a+b\hat{x}_{L95}^{\lambda}+c\hat{x}_{L95}^{2\lambda}=0 \tag{8.6}$$

$$a=Var(\hat{\beta}_0)-\frac{(y_0-\hat{\beta}_0)^2}{t_\alpha^2} \tag{8.7}$$

$$b=2\times Cov(\hat{\beta}_0,\hat{\beta}_1)+\frac{2\times(y_0-\hat{\beta}_0)\hat{\beta}_1}{t_\alpha^2} \tag{8.8}$$

$$c=Var(\hat{\beta}_1)-\frac{\hat{\beta}_1^2}{t_\alpha^2} \tag{8.9}$$

2 次式の解の計算は，x^{λ} に関して行なわれるので，元の x に戻すためには $1/\lambda$ 乗すればよい．



			x^{λ}	x
逆推定 y_0 =	20.0	x_0^ =	13.062	5.546
回帰直線の95%CL				
a =	−29.126	x_{L95R}^ =	12.186	5.296
b =	4.461	x_{U95R}^ =	14.062	5.826
c =	−0.170			
個々のデータの95%CL			x^{λ}	x
a' =	−24.866	x_{L95I}^ =	6.954	3.643
b' =	4.461	x_{U95I}^ =	28.074	9.237
c' =	−0.127			

図 8.4　元のスケールでの逆推定

2 次式の解を求めるためのパラメータの変更方法

個別データの場合には，重み (1, $1/x$, $1/x^2$) に応じてパラメータをそれぞれ ($a+\hat{\sigma}_e^2$, $b+\hat{\sigma}_e^2$, $c+\hat{\sigma}_e^2$) とする．ノイズ v_0 がある場合には，パラメータ a を $a+v_0\hat{\sigma}_e^2$ とする．重みとして

$1/(v_0+x_i^2)=1/(100+x_i^2)$ としたので

$$a'=a+v_0\hat{\sigma}_e^2=-29.126+100\times 0.0426=-24.866$$

$$c'=c+\hat{\sigma}_e^2=-0.170+0.0426=-0.127$$

となる．

8.2. 回帰式のべき λ の最尤法による推定

測定されたデータからべき λ の推定をしたい．これまで，式 $y=\beta_0+\beta_1 x^\lambda$ が x^λ に関して 1 次式となり，様々な推定が行なえることを示してきた．

本書では，x^λ に関しての 1 次式を推奨してきたのであるが，ある物質の測定系について妥当なべき λ を推定する必要がある．もちろん，異なるデータセットで何回もべき λ を計算し，その平均を求めて，あらかじめ定めておくべきである．

ある $n=20$ データセットが得られたときに，

$$y_i=\beta_0+\beta_1 x_i^\lambda+\varepsilon_i,\quad i=1,2,...,n,\quad \varepsilon_i\sim N(0,\sigma_e^2) \tag{8.10}$$

となる回帰式をあてはめたい．べき λ を既知，例えば $\lambda=1.3$ であることが分かっていれば，通常の回帰分析の手順で推定値 $\hat{\beta}_0$，$\hat{\beta}_1$，および，$\hat{\sigma}_e^2$ を求めることはたやすい．切片がない $y=\beta_1 x^\lambda$ ならば，両辺に対数を取ることにより 1 次式となり，$\hat{\beta}_1$，$\hat{\lambda}$，および，$\hat{\sigma}_e^2$ を推定することができるが，切片 β_0 が入っているために容易ではない．

最尤法を用いれば，このような分布に関するパラメータを推定することができるが，反復計算を必要とするので，この方法はまったく普及してない．Excel の標準機能（アドインする必要がある）として備わっているソルバーによって，最尤法による各種のパラメータの推定が容易に行なえるようになった．

本書では，これまで重み付き最小 2 乗法を用いて，各種の誤差モデルに対して未知パラメータ β_0 や β_1 の推定値を求めてきたが，分散は $\hat{y}_i$ を求めてから別途計算していた．最尤法は，分散も未知パラメータとして β_0，β_1，λ も含めて同時にこれらのパラメータを推定する方法である．

最尤法は，誤差が正規分布するという事前情報を用いて，第 1 ステップとして，y_i が推定値 $\hat{y}_i$ を「平均」，分散 σ_e^2 の正規分布の確率密度 L_i を求める．Excel の関数では L_i=Normdist（y_i，$\hat{y}_i$，σ_e^2, false)とすればよい．第 2 ステップとして，全ての i に対して L_i の積，または ln（L_i）の和が，最大となるようにソルバーで β_0，β_1，λ，σ_e^2 を変化させ，最大値を求める．なお，尤度を最大にするようなパラメータを最尤推定量という．

式（8.10）での尤度の積 L は，適当な β_0'，β_1'，λ'，$\sigma_e'^2$ を与えて，$\hat{y}_i' = \beta_0' + \beta_1' x_i^{\lambda'}$ を計算し，

$$L = \Pi L_i = \Pi \mathrm{Normdist}(y_i,\ \hat{y}_i',\ \sigma_e'^2,\ \mathrm{false}) \tag{8.11}$$

または，

$$\ln L = \Sigma \ln L_i = \Pi \ln(\mathrm{Normdist}(y_i,\ \hat{y}_i',\ \sigma_e'^2,\ \mathrm{false})) \tag{8.12}$$

L が最大となるように β_0'，β_1'，λ'，$\sigma_e'^2$ を少しずつ変化させる．膨大な計算量となるので，一昔の計算機では，計算不能であったが，最近では，普通の PC でも 10 秒ほどで計算できるようになっている．

数値例が示されている文献がないので，自己検証できるように，あらかじめ，4 個の β_0，β_1，λ，σ_e^2 のパラメータを与えて，n 組のデータセットを作成する．あらかじめ Excel 上に設定した x_i について，$y_i' = \beta_0' + \beta_1' x_i^{\lambda'}$ を計算する．一様乱数の Rand()を入れ子 Normsinv(Rand())にした標準正規乱数を発生させ，与えられた分散 σ_e^2 の平方根を掛ける．D26 セルでの Excel の計算式は次のようである．

= $C$17+$C$18*$C26^$E$17 + NORMSINV(RAND())*SQRT($E$18)

β_0，　β_1，　x_i　λ　標準正規乱数　$\sqrt{\sigma_e^2}$

データ発生のためのパラメータ設定

	A	B	C	D	E
16		データ発生のためのパラメータ，変更可能			
17		β_0 =	20	λ =	1.5
18		β_1 =	1.5	σ_e^2 =	100
19		操作：F9キーを押すと新たなデータが			
20		発生され散布図の y も同時に変わる．			

$y = \beta_0 + \beta_1 x^{\lambda} + \varepsilon$

	A	B	C	D
25		i	x	y
26		1	1.0	24.0
27		2	2.0	21.5
28		3	3.0	29.2
29		4	4.0	22.0
30		5	5.0	44.0
31		6	6.0	21.2
32		7	7.0	34.9
33		8	8.0	59.5
34		9	9.0	68.6
35		10	10.0	42.1
36		11	11.0	60.1
37		12	12.0	92.5
38		13	13.0	75.5
39		14	14.0	124.4
40		15	15.0	95.6
41		16	16.0	109.8
42		17	17.0	105.2
43		18	18.0	124.1
44		19	19.0	159.7
45		20	20.0	157.9

乱数によって発生された y は，F9 キーを押すことより変化する．ここに示したのは，ある時点でのスナップショットであり，再現することは Excel では不可能である．そのために，このデータ全体をコピーし，【形式を選択して貼り付け／値のみ】で次に示すように別のセルに固定化する．

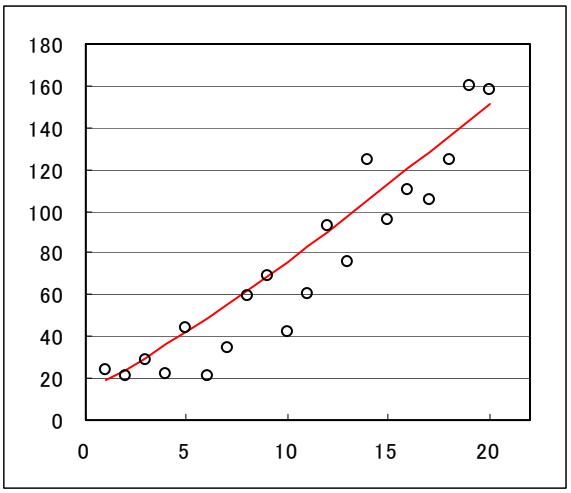


図 8.5 発生したデータとおおよそのあてはめ曲線

	C	H	I	J	K	L
				偏差	正規分布	
		yの値のみ	推定値	（誤差）	確率密度	対数尤度
25	x	コピー	y^	e=y-y^	尤度 L	ln L
26	1.0	24.0	18.9	22.3	0.0012	-6.76
27	2.0	21.5	24.2	-4.0	0.0435	-3.14
28	3.0	29.2	29.9	11.1	0.0193	-3.95
29	4.0	22.0	35.9	-8.1	0.0299	-3.51
30	5.0	44.0	42.2	2.0	0.0474	-3.05
31	6.0	21.2	48.6	-7.8	0.0308	-3.48
32	7.0	34.9	55.3	8.0	0.0301	-3.50
33	8.0	59.5	62.1	-32.7	0.0000	-11.03
34	9.0	68.6	69.0	6.0	0.0372	-3.29
35	10.0	42.1	76.0	-13.9	0.0115	-4.47
36	11.0	60.1	83.1	-6.9	0.0341	-3.38
37	12.0	92.5	90.4	-26.1	0.0003	-8.13
38	13.0	75.5	97.7	-21.2	0.0017	-6.40
39	14.0	124.4	105.1	-18.5	0.0037	-5.60
40	15.0	95.6	112.6	-23.1	0.0009	-7.04
41	16.0	109.8	120.2	-6.7	0.0348	-3.36
42	17.0	105.2	127.9	8.8	0.0273	-3.60
43	18.0	124.1	135.6	-4.1	0.0431	-3.14
44	19.0	159.7	143.4	6.4	0.0359	-3.33
45	20.0	157.9	151.3	5.7	0.0382	-3.27
46			Σe =	-102.7		

図 8.6 データ y の固定化，推定値 $\hat{y}$ にたいする確率密度（尤度）の計算

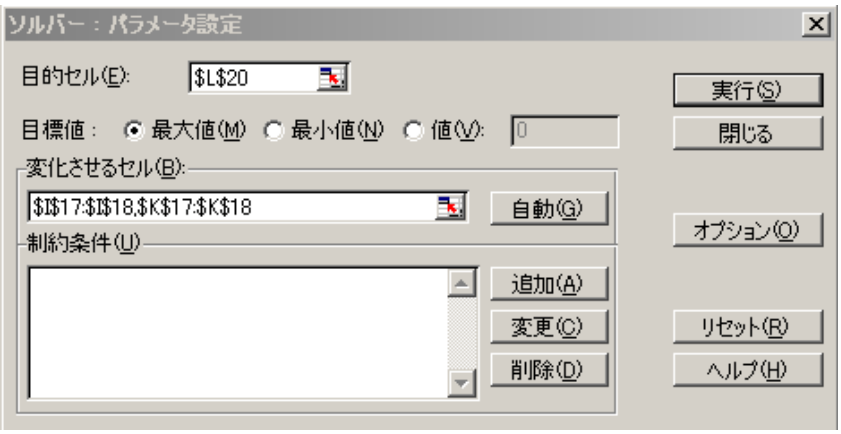
推定値 $\hat{y}_i$ は，次に示すような別のセルに用意した推定値の初期値 $\hat{\beta}_0$，$\hat{\beta}_1$，$\hat{\lambda}$ を用いた結果である，正規乱数を発生させ，誤差 ε_i を加えた y_i との差 $e_i = y_i - \hat{y}_i$ が計算されている．これらを全て加えた結果は，0 ではなく －102.7 となっていて，誤差を含んだ実際のデータセットに対しては，適切な推定値となっていない．

H	I	J	K	L	M
ここのデータを変化させ尤度の和を最大化する					
$\hat{\beta}_0$ =	20.000	$\hat{\lambda}$ =	1.500		
$\hat{\beta}_1$ =	1.500	$\tilde{\sigma}_e^2$ =	100.000	117.647	不偏分散
	20個の対数尤度の和 $\Sigma \ln L$ =			-89.502	ここを最大化

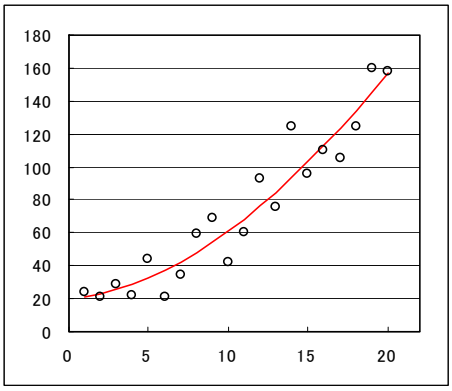
図 8.7　推定値 $\hat{y}$ の計算のための暫定的なパラメータ設定

Excel の標準のアドオンのソルバーの操作画面は，次のようである．【目的セル】に最大化したい対数尤度の和のセルアドレス，【変化させるセル】に，$\hat{\beta}_0$，$\hat{\beta}_1$，$\hat{\lambda}$，および，$\hat{\sigma}_e^2$ のセルアドレスをセットする．目標値が【最大値】に選択されていることを確認し，【実行】する．

Excel のソルバーの設定画面



【実行】すると次のようにパラメータおよび尤度の和が変化する．



H	I	J	K	L	M
ここのデータを変化させ尤度の和を最大化する					
$\hat{\beta}_0$ =	20.604	$\hat{\lambda}$ =	1.750		
$\hat{\beta}_1$ =	0.720	$\tilde{\sigma}_e^2$ =	163.280	192.095	不偏分散
	20個の対数尤度の和 $\Sigma \ln L$ =			-79.329	ここを最大化

図 8.8　Excel のソルバーで対数尤度の和を最大化した結果

得られた推定値は，$\hat{\beta}_0 = 20.604$，$\hat{\beta}_1 = 0.720$，$\hat{\lambda} = 1.750$，および，$\tilde{\sigma}_e^2 = 163.280$ となる．データの散布図上に，回帰曲線 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i^{\hat{\lambda}}$ を上書きした結果を次に示す．分散の推定値を $\tilde{\sigma}_e^2$ としたのは，最尤法で推定した場合に，誤差平方和をデータ数 $n = 20$ で割った結果となり，

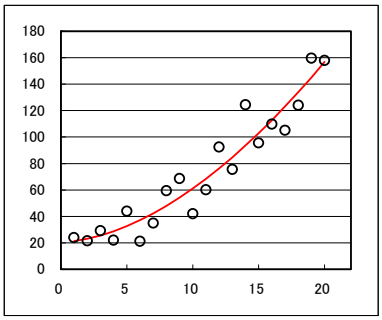
推定された分散が過小評価となるためである．不偏分散とするためには，パラメータ数は $p=3$ なので，　$\hat{\sigma}_e^2=\tilde{\sigma}_e^2\times n/(n-p)=163.280\times 20/(20-3)=192.095$ となる．

計算に用いた Excel シート全体

冪 λ などの推定

目的：測定されたデータ（x, y）から
$y=\beta_0+\beta_1 x^{\lambda}$ のパラメータ
β_0, β_1, λ および誤差分散 σ_e^2 を同時推定する．

方法：あらかじめ設定した4パラメータを用い，
正規乱数を用いてシミュレーションデータを n 組作成する．
その n 組のデータのみから 4パラメータを推定する．



データ発生のためのパラメータ，変更可能

β_0 =	20	λ =	1.5
β_1 =	1.5	σ_e^2 =	100

操作：F9キーを押すと新たなデータが
発生され散布図の y も同時に変わる．

ここのデータを変化させ尤度の和を最大化する

$\hat{\beta}_0$ =	20.604	$\hat{\lambda}$ =	1.750
$\hat{\beta}_1$ =	0.720	$\tilde{\sigma}_e^2$ =	163.280

192.095　不偏分散

20個の対数尤度の和 $\Sigma \ln L$ = −79.329　ここを最大化

$y=\beta_0+\beta_1 x^{\lambda}+\varepsilon$　F9キーで y を変えたら

i	x	y	手動で値のみをコピー
1	1.0	25.7	====>
2	2.0	22.3	====>
3	3.0	19.1	:
4	4.0	24.8	:
5	5.0	33.0	
6	6.0	44.2	
7	7.0	59.5	
8	8.0	61.3	
9	9.0	49.6	
10	10.0	68.1	
11	11.0	74.3	
12	12.0	73.3	
13	13.0	82.0	
14	14.0	80.6	
15	15.0	82.0	
16	16.0	127.9	
17	17.0	110.6	
18	18.0	127.2	
19	19.0	142.4	
20	20.0	168.0	====>

ソルバー起動後に変化する

散布図上の○印と曲線

y の値のみコピー	推定値 y^	偏差（誤差） $e=y-y$^	正規分布確率密度尤度 L	対数尤度 $\ln L$
24.0	21.3	2.7	0.0305	−3.49
21.5	23.0	−1.5	0.0310	−3.47
29.2	25.5	3.7	0.0300	−3.51
22.0	28.7	−6.7	0.0272	−3.61
44.0	32.6	11.4	0.0210	−3.86
21.2	37.2	−16.0	0.0143	−4.25
34.9	42.3	−7.4	0.0264	−3.63
59.5	48.0	11.5	0.0208	−3.87
68.6	54.3	14.3	0.0167	−4.09
42.1	61.1	−19.0	0.0103	−4.57
60.1	68.4	−8.3	0.0252	−3.68
92.5	76.3	16.2	0.0140	−4.27
75.5	84.7	−9.2	0.0241	−3.73
124.4	93.6	30.8	0.0017	−6.38
95.6	102.9	−7.3	0.0265	−3.63
109.8	112.8	−3.0	0.0304	−3.49
105.2	123.1	−17.9	0.0117	−4.45
124.1	133.9	−9.8	0.0233	−3.76
159.7	145.1	14.6	0.0163	−4.12
157.9	156.8	1.1	0.0311	−3.47
	Σe =	0.0		

誤差平方和 S_e = 3264.164
不偏分散 192.010

図 8.9　計算に用いた Excel シート全体

8.3. 重みに用いる x のべき乗の最尤法による推定

これまで，分散が x^2 に比例するか，x に比例するかの 2 通りの場合について示したきた．どちらを選択するのかについては，これまで示してこなかった．前節では回帰曲線 $y_i = \beta_0 + \beta_1 x_i^{\lambda}$ のべき λ の推定について示した. 分散を x の関数で現す場合に, べき λ を含めて, $x^{2\lambda}$ あるいは x^{λ} どちらを使うのかの判断が必要となる.

次式で示す 1 次回帰式で，分散が x^2 に比例するモデルを取り上げ，前節で示した正規乱数を用いたデータを発生させ，そのデータから分散が x^2 に比例するデータであることの推測を最尤法で行うことにする.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\quad i = 1,2,...,5\,,\quad \varepsilon_i \sim N(0, x_i^2 \sigma_e^2) \tag{8.13}$$

データは, 分散が x^2 に比例するとし場合に，次に示すように $\delta = 2$，$x^{\delta} = x^2$ で正規乱数を発生させた.

分散が x^2 に比例する場合のパラメータ設定

データ発生のためのパラメータ, 変更可能

β_0 =	10	λ =	1
β_1 =	1.5	σ_e^2 =	0.4

重み

v_0 =	0	δ =	2

発生した 20 個のデータ

i	x	y
1	1.0	12.3
2	2.0	12.9
3	3.0	14.2
4	4.0	15.1
5	5.0	17.8
6	6.0	22.2
7	7.0	21.1
8	8.0	18.6
9	9.0	26.7
10	10.0	25.1
11	11.0	30.5
12	12.0	32.0
13	13.0	29.5
14	14.0	23.0
15	15.0	35.7
16	16.0	34.3
17	17.0	27.1
18	18.0	50.1
19	19.0	40.4
20	20.0	37.2

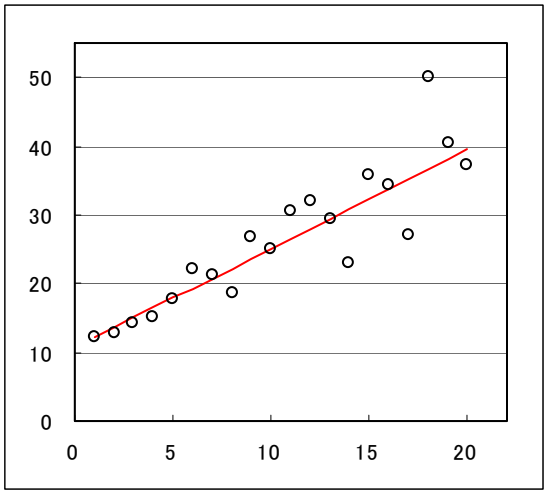


図 8.10 発生したデータとおおよそのあてはめ曲線

分散が x^{δ} に比例するとしたときに，

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i\,,\ \ i = 1,2,\ldots,5\,,\ \ \varepsilon_i \sim N(0, x_i^{\delta}\sigma_e^2) \tag{8.14}$$

δ を最尤法によって推定するためには，誤差 ε_i が，分散が $x_i^{\delta}\sigma_e^2$ の正規分布に従うとして，確率密度を計算すればよい．もちろん，y_i が，平均 $\hat{y}_i$，分散 $x_i^{\delta}\sigma_e^2$ の正規分布に従うとしても同じである．

パラメータの初期値

ここのデータを変化させ尤度の和を最大			
$\hat{\beta}_0$ =	10.000	$\hat{\lambda}$ =	1.000
$\hat{\beta}_1$ =	1.500	$\tilde{\sigma}_e^2$ =	0.400
重み			
w_0=	0	δ =	2.000

Excel シート上の正規分布の確率密度（尤度）の初期値

x	yの値のみコピー	推定値 y^	残差 e=y-y^	分散 $x^{\delta}\tilde{\sigma}_e^2$	標準偏差 SD	正規分布 確率密度 尤度 L	対数尤度 $\ln L$
1.0	12.3	12.1	0.2	0.1	0.2	1.2406	0.22
2.0	12.9	13.6	−0.7	0.3	0.5	0.3442	−1.07
3.0	14.2	15.0	−0.8	0.7	0.8	0.3003	−1.20
4.0	15.1	16.5	−1.4	1.4	1.2	0.1752	−1.74
5.0	17.8	17.9	−0.1	2.3	1.5	0.2612	−1.34
6.0	22.2	19.3	2.9	3.6	1.9	0.0673	−2.70
7.0	21.1	20.8	0.3	5.1	2.3	0.1746	−1.75
8.0	18.6	22.2	−3.6	7.0	2.6	0.0587	−2.84
9.0	26.7	23.7	3.0	9.2	3.0	0.0802	−2.52
10.0	25.1	25.1	0.0	11.8	3.4	0.1159	−2.15
11.0	30.5	26.6	3.9	14.8	3.8	0.0616	−2.79
12.0	32.0	28.0	4.0	18.2	4.3	0.0606	−2.80
13.0	29.5	29.5	0.0	21.9	4.7	0.0852	−2.46
14.0	23.0	30.9	−7.9	26.1	5.1	0.0235	−3.75
15.0	35.7	32.4	3.3	30.7	5.5	0.0601	−2.81
16.0	34.3	33.8	0.5	35.7	6.0	0.0665	−2.71
17.0	27.1	35.3	−8.2	41.2	6.4	0.0277	−3.59
18.0	50.1	36.7	13.4	47.1	6.9	0.0086	−4.75
19.0	40.4	38.1	2.3	53.5	7.3	0.0520	−2.96
20.0	37.2	39.6	−2.4	60.3	7.8	0.0490	−3.02
		Σ e =	8.7				

パラメータの初期値　　Excel シート上の正規分布の確率密度（尤度）の最大化

ここのデータを変化させ尤度の和を最大化す			
$\hat{\beta}_0$ =	10.665	$\hat{\lambda}$ =	1.000
$\hat{\beta}_1$ =	1.446	$\tilde{\sigma}_e^2$ =	0.053
重み			
v_0=	0	δ=	2.349
20個の対数尤度の和 $\Sigma \ln L$ =			−48.731

x	yの値のみコピー	推定値 y^	残差 e=y-y^	分散 $x^{\delta}\tilde{\sigma}_e^2$	標準偏差 SD	正規分布 確率密度尤度 L	対数尤度 $\ln L$
1.0	12.3	12.1	0.2	0.1	0.2	1.2406	0.22
2.0	12.9	13.6	−0.7	0.3	0.5	0.3442	−1.07
3.0	14.2	15.0	−0.8	0.7	0.8	0.3003	−1.20
4.0	15.1	16.5	−1.4	1.4	1.2	0.1752	−1.74
5.0	17.8	17.9	−0.1	2.3	1.5	0.2612	−1.34
6.0	22.2	19.3	2.9	3.6	1.9	0.0673	−2.70
7.0	21.1	20.8	0.3	5.1	2.3	0.1746	−1.75
8.0	18.6	22.2	−3.6	7.0	2.6	0.0587	−2.84
9.0	26.7	23.7	3.0	9.2	3.0	0.0802	−2.52
10.0	25.1	25.1	0.0	11.8	3.4	0.1159	−2.15
11.0	30.5	26.6	3.9	14.8	3.8	0.0616	−2.79
12.0	32.0	28.0	4.0	18.2	4.3	0.0606	−2.80
13.0	29.5	29.5	0.0	21.9	4.7	0.0852	−2.46
14.0	23.0	30.9	−7.9	26.1	5.1	0.0235	−3.75
15.0	35.7	32.4	3.3	30.7	5.5	0.0601	−2.81
16.0	34.3	33.8	0.5	35.7	6.0	0.0665	−2.71
17.0	27.1	35.3	−8.2	41.2	6.4	0.0277	−3.59
18.0	50.1	36.7	13.4	47.1	6.9	0.0086	−4.75
19.0	40.4	38.1	2.3	53.5	7.3	0.0520	−2.96
20.0	37.2	39.6	−2.4	60.3	7.8	0.0490	−3.02

推定されたのは，$\delta = 2.349$ で $\sigma_e^2 = 0.053$ なので分散が $0.053x_i^{2.349}$ の正規分布に従っていると推論される．実際に $x_i^{2.349}$ を重みに使うことは，様々な信頼区間の推定ができななくなるので，x_i^2 を使うことが現実的である．

9. 文献に対する批判的吟味

［1］日本規格協会（2001），標準物質を用いた校正（検量線が直線の場合）JIS Z 8461.
［2］Draper N.R., Smith N. (1998), Applied Regression Analysis, 3rd ed, 83-86, 226-229, Wiley.
［3］Snedecor G.W., Cochran W.G.(1967), 畑村又好，奥野忠一，津村善郎 共訳（1972），統計的方法，原著第 6 版，144-155，岩波書店.
［4］芳賀敏郎，野澤昌弘，岸本淳司（1996），SAS による回帰分析，181-191，東大出版会.
［5］Taylor J.R. (1997), 林茂雄，馬場涼 訳（2000），計測における誤差解析入門，204-213，東京化学同人.
［6］竹内啓（1963），数理統計学，312-315，東洋経済新報社.
［7］Miller J.N., Miller J.C.（2000），Statistics and Chemometrics for Analytical Chemistry 4thed. 宗森信，佐藤寿邦 訳（2004），データの取りまとめ方-化学分析のための統計学とケモメトリックス－，142-173，共立出版.

逆推定について

回帰分析での逆推定について，最近の統計の教科書では見出すことができない．生物統計に関する名著[3]で，個別データの逆推定の 95%信頼区間が例示されている．回帰分析での逆推定の考え方，95%信頼区間の導出法については，[6]が参考になった．回帰分析の名著[2]にも簡単な記述がある（第 1 版の日本語訳「応用回帰分析」には記述がない）．[7]は分析化学者のための代表的な教科書であり，逆推定の 95%信頼区間の算出に近似的な誤差分散を用いた方法が示されている．

重み付き回帰について

重み付き回帰の数理については，[2]が参考になった．重みをどのように設定するかについて，分散の推定に x の関数を用いる例示が[2]にあり，勇気付けられた．重みとして x_i^2 の逆数を使う場合に，回帰式の右辺および左辺を x_i^2 で割って，誤差が均一になるような「変数変換」し，通常の回帰分析を適用する例示が[1]にある．重みとして y の推定値 $\hat{y}$ を使う例示が[4]にあり，反復計算の結果も示されている．[5]は演習問題形式で重み付き回帰を論じ，誤差 σ_i がわからないかぎり適用できないとしている．[7]は x_i における繰返し測定から標準偏差 SD_i が求められていて，重みとして分散 $1/SD_i^2$ の逆数を用いている．

直線があてはまらない場合

[7]に直線があてはまらない事例が例題として取り上げられている．

批判的吟味

[1]について $1/x^2$ を重みとした回帰分析結果を示し，逆推定の 95%信頼区間を追加し，さらに，対数尤度を用いて重みとして $1/x^2$ が適していることを示した．

［2］の重み付き回帰は，x の 2 次式の逆数を用いているが，個別データの 95%信頼区間から，不適切であることを示し，重みとして$1/(0.02689x^2)$ が適していることを追加し，重みとして$1/x^2$ を用いても同じ結果が得られることを示した．

［3］には，重みは考慮されていないので，重みを$1/x$ とした場合，$1/x^2$ の場合について例示をし対数尤度から，$1/x^2$ が優れていることを追加した．

［4］の y の推定値　$1/\hat{y}$ を重みにする方法は，原点付近の分散を過大評価させていることを個別データの 95%信頼区間を示し，$1/\hat{y}^2$ の方が優れていることを例示した．重みとして y の推定値を使うことが適しているのか，x の関数を使うのが適しているのかについて考察を加え，回帰分析においては，分散の推定において x の関数を使うことの優位性を論じた．

［6］は基本的な考え方が示されているだけなので，本文中で用いた図を示しつつ逆推定の構造について追加した．

［7］で個別データの逆推定値 $\hat{x}_0$ の 95%信頼区間を $SE(\hat{x}_0)$ を近似式で求め，t 分布の両側 5%点を掛けて算出している．近似式はあたえられたとしているので，その導出の考え方を示し，重み付き回帰の場合への応用について示した．

標準偏差 $\hat{\sigma}_i$ を分散 $\hat{\sigma}_i^2$ について x^2 の 1 次式から分散 $\hat{\sigma}_i'^2$ を推定し，その逆数を重みとした解析を追加し，その結果として個別データの 95%信頼区間が求められることをしめした．

直線のあてはめが困難な場合に，x^λ の 1 次式について，λ を変化させたときに誤差分散 $\hat{\sigma}_e^2$ が最も小さくなる λ の推定法を示し，その場合の個別データの逆推定の 95%信頼区間を算出した結果を示した．

9.1. ［付録 1］ 日本規格協会(2001)

［1］日本規格協会(2001)，標準物質を用いた校正(検量線が直線の場合)JIS Z 8461.

重み付き回帰分析については，JIS 規格で制定されたいる．数値例，計算手順などが詳細に述べられている．ただし，逆推定の 95%信頼区間は例示がないので追加解析の結果を示した.

序文（原著の引用）

この規格は，1996 年に第 1 版として制定された ISO 11095，Linear calibration using reference materials を翻訳し，技術的内容及び規格票の様式を変更することなく作成され，2001 年に制定された日本工業規格である。（中略）

比例する残差標準偏差を仮定した校正関数（原著の引用）

6.4　比例する残差標準偏差を仮定した校正関数（検量線）の推定と，校正関数(検量線)及び残差のプロット

6.4.1 **モデル**　ステップ 6.2.1 に示したモデルに代わるもう一つのモデルは，校正関数（検量線）は直線であるが, RM 採択値の増加につれて残差標準偏差も増加するモデルである。これは，次の式で表される（高橋注，RM：標準物質)。

$y_{nk} = \gamma_0 + \gamma_1 x_n + \eta_{nk}$,

ここに

x_n ：n 番目の RM の採択値（$n = 1,2,...,N$）

y_{nk} ：n 番目の RM の k 番目の測定値（$k = 1,2,...,K$）

$\gamma_0 + \gamma_1 x_n$ ：n 番目の RM の測定値の期待値

η_{nk} ：y_{nk} の n 番目の RM の測定値の期待値からの偏差（これらの偏差は独立で，平均 0，分散が x_n^2 に比例する正規分布に従うものと仮定する。)，すなわち，

$\mathrm{var}(\eta_{nk}) = \mathrm{var}(y_{nk}) = x_n^2 \tau^2$

γ_0，γ_1 及び τ^2 ：校正（検量線作成）実験で得られたデータから推定する三つのパラメータ

$\eta_{nk} \sim N(0, x_n^2 \tau^2)$ ：校正関数（検量線）の切片と傾き

τ^2 ：測定系の相対制度の尺度（高橋注，x_n =1 の場合の分散となる）

このモデルは，6.2.1 のモデルと同等なモデル,すなわち，誤差分散が一定であるモデルへ変換することができる．この変換を行うため，次の式の両辺を x_n で割る。

（高橋注，6.2.1 のモデル：$y_{nk} = \beta_0 + \beta_1 x_n + \varepsilon_{nk}$）

$y_{nk} = \gamma_0 + \gamma_1 x_n + \eta_{nk}$,

したがって．次の式が得られる。

$$\frac{y_{nk}}{x_n} = \frac{1}{x_n}\gamma_0 + \gamma_1 + \frac{1}{x_n}\eta_{nk}$$

これは，次式と同等である。

$$z_{nk} = \gamma_1 + \gamma_0 w_n + \varepsilon_{nk}$$

ここに,

$$z_{nk} = y_{nk} / x_n$$

$$w_n = 1 / x_n$$

$$\varepsilon_{nk} = \eta_{nk} / x_n$$

6.2 に示した式の各項に正しく代入すれば, この新しいモデルを同じ方法で解析することができる。（中略）

基本的方法（原著の引用）

9.2　基本的方法

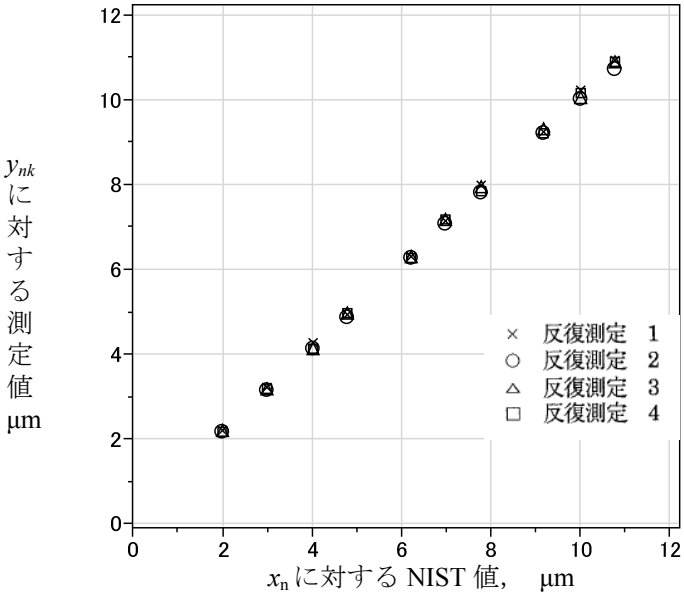
9.2.1 **背景とデータ**　0.5 μm から 12 μm の範囲にある集積回路フォトマスクの線間隔は, 光学的撮像システム（測定用アタッチメントをもつ光学顕微鏡）によって測定する。このようなシステムは，国立標準技術研究所（NIST: NationalInstitute of Standards and Technology）が提供する標準物質 SRM-474 を用いて校正（検量線作成）することができる。SRM-474 は各種機能のほかに,間隔が 0.5 μm から 12 μm の範囲で，ランダムに配列された直線を 10 個もっている。ここでは，光学的撮像システムに対して行われた校正（検量線作成）実験について述べる。標準の線間隔 10 個は，それぞれ 4 回測定された。測定値の間の独立性を保証するため，2 週間間隔をあけて反復測定が行われた。表 3 のデータは，NIST が採択値を規定する線間隔 10 個（N=10）のそれぞれについて反復測定された値 4 個（K=4）を示す。

原著・表 3　線間隔の校正（検量線作成）実験　（単位 μm）

NITS 値	測定値			
	反復測定 1	反復測定 2	反復測定 3	反復測定 4
x_n	y_{n1}	y_{n2}	y_{n3}	y_{n4}
6.19	6.31	6.27	6.31	6.28
9.17	9.27	9.21	9.34	9.23
1.99	2.21	2.19	2.22	2.20
7.77	8.00	7.81	7.95	7.84
4.00	4.27	4.15	4.15	4.15
10.77	10.93	10.73	10.92	10.89
4.78	4.95	4.87	5.00	5.00
2.99	3.24	3.17	3.21	3.21
6.98	7.14	7.07	7.18	7.20
9.98	10.23	10.02	10.07	10.17

9.2.2　データのプロット

図 6 に示す実験において収集されたデータのプロットでは，校正（検量線作成）実験において明白な異常値が現れたことも，システムが異常な挙動を行ったことも特定されない。これは，校正関数（検量線）の直線性を裏付けるが，各 NIST 値が増加すると，その NIST 値に対するデータの広がりが拡大するところから，残差標準偏差一定の仮定に関して疑いが生ずる。



原著・図 6　線間隔校正実験において収集されたデータ

（高橋の追試）

このデータに対して通常の回帰分析を適用すると，測定誤差が小さいために，回帰直線の95%信頼区間，および，個別データの 95%信頼区間は回帰直線とほとんど重なっている．この規格は，標準偏差が x に比例する（分散が x^2 に比例する）場合について示されている．ただし，重み付回帰ではなく「線形化」の方法による一般的な回帰分析の問題として定式化している．なお，回帰直線および個別データの 95%信頼区間，分散が x に比例する場合は，この規格に含まれていない．

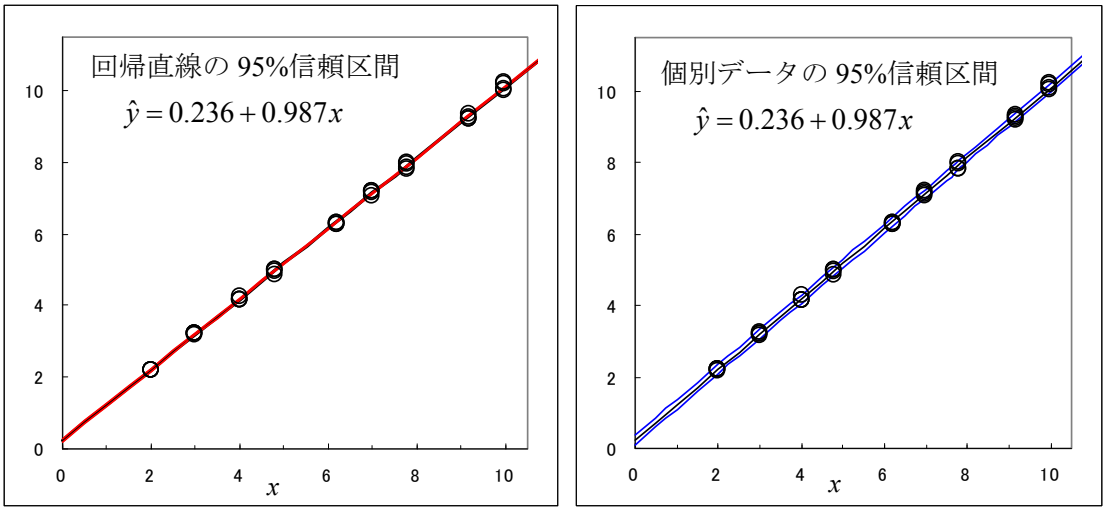


図 9.1　線間隔の校正実験での回帰直線および 95%信頼区間

残差プロットを図 9.2 に示す．反復測定 2 の○印は，他の反復測定よりも一貫して小さいことが読み取れる．反復に伴う系統的な誤差が示唆されるが，ここでは無視して先に進める．残差は，x が大きくなるにつれ扇型に増大し，x に関して残差が均一とはみなせない．

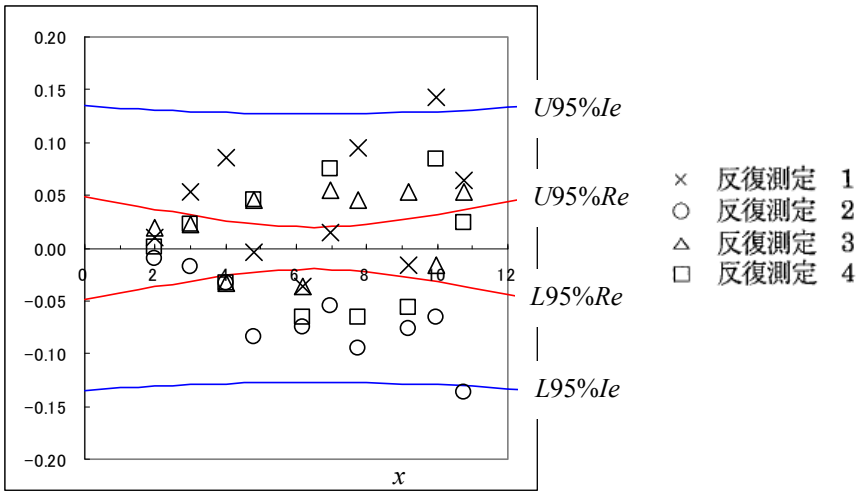


図 9.2　残差プロット（$\ln L = 55.447$）

残差プロットに個別データの 95%信頼区間から $\hat{y}$ の推定値を引いた相対的な 95%信頼区間（*L*95%*Ie*，*U*95%*Ie*）を上書きして見ると，x が小さくなるにつれ扇方の残差プロットとの乖離している．したがって，通常の回帰分析の適用することはできない．

JIS 規格では，残差の不均衡を是正するために表 9.1 に示すように回帰式の両辺を x_n で割ったデータにし，誤差分散を均一化する工夫をしている．

表 9.1　回帰式の両辺を x_n で割ったデータ

NITS 値		測定値			
		反復測定 1	反復測定 2	反復測定 3	反復測定 4
x_n	$1/x_n$	y_{n1}/x_n	y_{n2}/x_n	y_{n3}/x_n	y_{n4}/x_n
6.19	0.162	1.019	1.013	1.019	1.015
9.17	0.109	1.011	1.004	1.019	1.007
1.99	0.503	1.111	1.101	1.116	1.106
7.77	0.129	1.030	1.005	1.023	1.009
4.00	0.250	1.068	1.038	1.038	1.038
10.77	0.093	1.015	0.996	1.014	1.011
4.78	0.209	1.036	1.019	1.046	1.046
2.99	0.334	1.084	1.060	1.074	1.074
6.98	0.143	1.023	1.013	1.029	1.032
9.98	0.100	1.025	1.004	1.009	1.019

測定値 y_{nk} の変わりに，$z_{nk} = y_{nk}/x_n$ を測定値とみなし，x_n の変わりに $w_n = 1/x_n$ とする通常の回帰分析

$$z_{nk} = \gamma_1 + \gamma_0 w_n + \varepsilon_{nk}$$

$$\varepsilon_{nk} \sim N(0, \tau^2)$$

を行なうことにより，「重み付き回帰」の代わりとなることを推奨している．実際にこのデータで回帰分析を行なうと，次のような回帰パラメータが推定される．

表 9.2　回帰パラメータの推定

k	w	γ_k	γ_k^	$Var(\gamma_k\hat{})$	SE	t	p
1	切片	γ_1	0.9851	0.00001	0.00288	342.53	0.0000
0	傾き	γ_0	0.2469	0.00015	0.01211	20.40	0.0000

ここで推定されたパラメータ切片が $\gamma_1 = 0.985$，傾きが $\gamma_0 = 0.247$ となる．JIS 規格では，この回帰式を用いて未知の y_0 に対する $\hat{x}_0$ の推定の推定を行なっている．図 9.3（右）の残差プロットから誤差が均一化されていることが確かめられる．

回帰直線の 95%信頼区間　　個別データの 95%信頼区間　　残差プロット

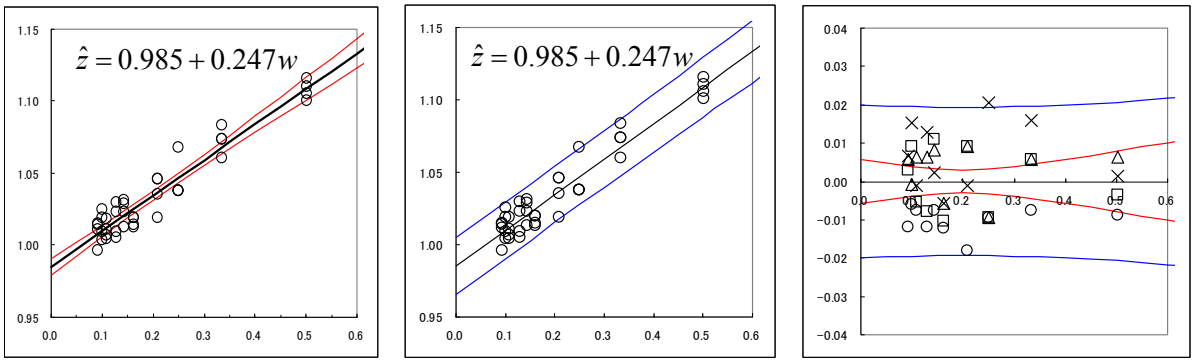


図 9.3　分散を一定化した回帰式および残差プロット

この切片と傾きを入れ替えて，

$$\hat{y}_{nk} = 0.247 + 0.985_1 x_n, \quad n = 1,2,\ldots,N, \quad k = 1,2,\ldots,K \text{（反復測定回数）} \qquad (9.1)$$

が得られる．JIS 規格では，この回帰式を用いて未知の y_0 に対する $\hat{x}_0$ の推定の推定を行なっている．

未知検体の 95%信頼区間の例示

JIS には、未知の y_0 に対する $\hat{x}_0$ の推定式は示されているが，その 95%信頼区間の求め方は示されていないので，　Excel を用いて 6.1 節で示した方法で重み付き回帰を行い個別データおよび複数データの 95%信頼区間を計算した結果を示す．

計算に用いるデータは，表 9.3 に示すように 4 反復×10 = 40 行のデータとし，40×2 のデザイン行列 $\boldsymbol{X}$，40×1 の測定データ行列 $\boldsymbol{y}$，$1/x_i^2$ を対角要素とする 40×40 の重み行列 $\boldsymbol{V}$ としたときに，2×1 の係数行列の推定値 $\hat{\boldsymbol{\beta}}$（JIS では γ_0，γ_1）は，次のように

$$(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \quad \boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y} \quad = \quad \boldsymbol{\beta}\hat{}$$

$$\begin{bmatrix} 1.649 & -0.335 \\ -0.335 & 0.093 \end{bmatrix} \begin{bmatrix} 8.564 \\ 41.412 \end{bmatrix} = \begin{bmatrix} 0.247 \\ 0.985 \end{bmatrix}$$

として JIS の計算結果と同様に求められる．

表 9.3　40 行にした線間隔の校正実験データ

反復測定 k	n	i	x_i	y_i
1	1	1	6.19	6.31
	2	2	9.17	9.27
	:	:	:	:
	10	10	9.98	10.23
2	1	11	6.19	6.27
	2	12	9.17	9.21
:	:	:		
:	:	:		
4	10	40	9.98	10.17

推定値 $\hat{y}_i = \hat{\beta}_0 \times 1 + \hat{\beta}_1 x_i$ の分散 $Var(\hat{y}_i)$ は，2×2 の分散共分散行列 $\boldsymbol{\Sigma}$ を用いて，次のように計算できる．

$$Vay(y_i\hat{}) = \boldsymbol{x}_i \begin{bmatrix} 1 & x_i \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} 0.00015 & -0.00003 \\ -0.00003 & 0.00001 \end{bmatrix} \boldsymbol{x}_i^T \begin{bmatrix} 1 \\ x_i \end{bmatrix}$$

推定された回帰直線との個別データの 95%信頼区間，残差プロット上での回帰および個別データの 95%信頼区間の上書き結果を図 9.4 に示す．

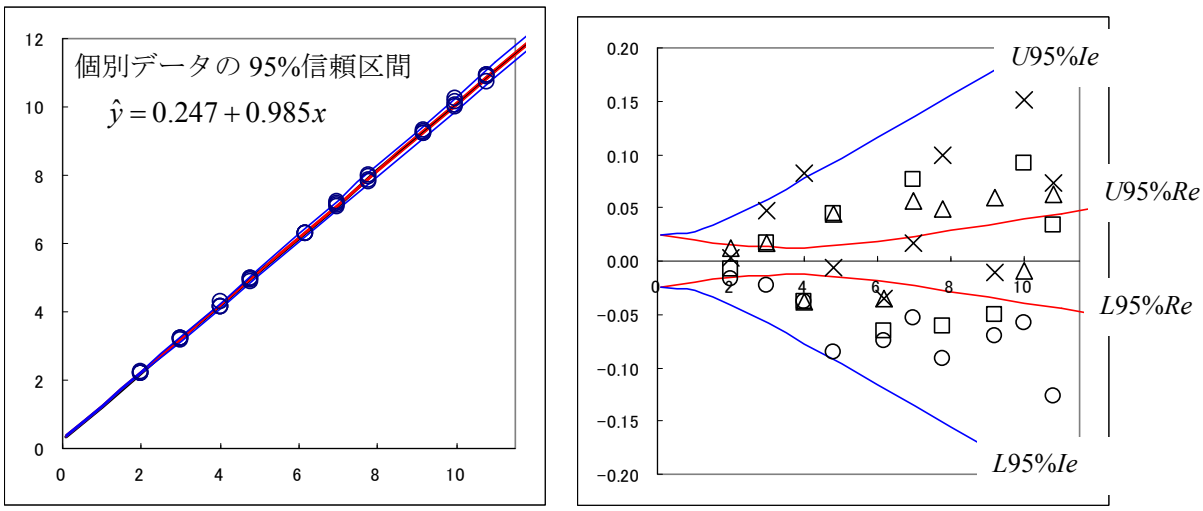


図 9.4　重みつき回帰式および残差プロット（$\ln L = 61.037$）

未知の y_0 に対する $\hat{x}_0$ は，

$$\hat{x}_0 = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1}$$

個別データの逆推定の 95%信頼区間は，回帰直線の分散に個別データの分散を加えた得た次の

$$(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95I}) + t_\alpha \sqrt{Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95I}) + \hat{x}_{L95I}^2 \hat{\sigma}_e^2} = y_0$$

下限 $\hat{x}_{L95I}$ について解いた 2 次式の 2 つの解から上限 $\hat{x}_{U95I}$ 求められることは，これまで示してきた．ここでは，$y_0 = 3$ および $y_0 = 10$ に対して Excel で計算した結果のみを示す．

逆推定 y_0 =	3.0	x_0^ =	2.795	逆推定 y_0 =	10.0	x_0^ =	9.900
個々のデータの95%CL			x	**個々のデータの95%CL**			x
a' =	−1.849	x_{L95I}^ =	2.740	a' =	−23.211	x_{L95I}^ =	9.708
b' =	1.324	x_{U95I}^ =	2.851	b' =	4.689	x_{U95I}^ =	10.100
c' =	−0.237			c' =	−0.237		

Excel の計算シートは，任意のべき乗の逆数 $(1/x^{\lambda})$ について計算できるように設定してあるので，べき λ を（0，0.5，1，2，4）と変化させ，残差プロット上で 95%信頼区間の形状をどのように変化するか調べてみよう．

$1/x^0 = 1$　　$1/x^{0.5}$　　$1/x^1$　　$1/x^2$　　$1/x^3$

$\ln L = 55.447$　　$\ln L = 57.994$　　$\ln L = 59.853$　　$\ln L = 61.037$　　$\ln L = 58.249$

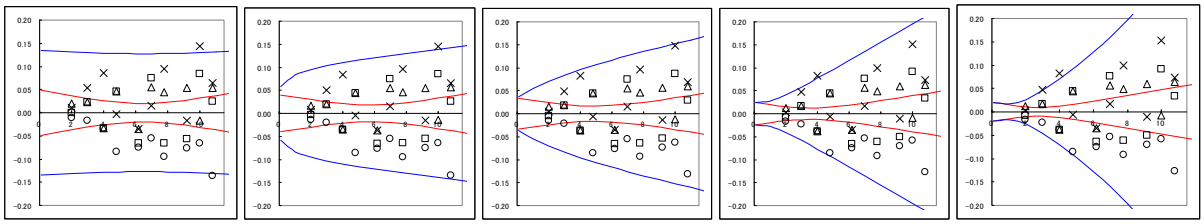


図 9.5　各種の重みに対する残差プロット上の 95%信頼区間

何がしかの $(1/x^{\lambda})$ の重みを与えると x が小さい場合には，個別データの 95%信頼区間は，回帰直線の 95%信頼区間に収束し，紡錘型から $\lambda = 2$ で直線的になり，さらに λ を大きくすると，x が大きくなると過度に発散する．どのような重みを選択するのかは，視覚的には残差プロット上に示された点を x の大きさによらず均一に包み込むような重み $1/x^1$ の選択も考えられる．統計的には，対数尤度が最大になるような重み $1/x^2$ が最適と判断される．重みを変化させ対数尤度を計算した結果を次のグラフに示す．

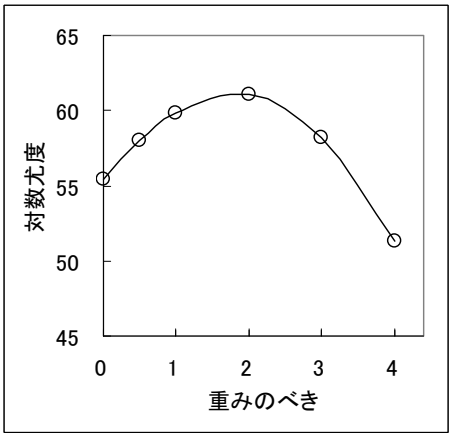


図 9.6　重みのべきを変化させた場合の対数尤度

この結果から重みが $1/x^2$ 付近で対数尤度が $\ln L = 61.037$ と最大となっている．対数尤度は，

測定値 y_{nk} が，推定値 $\hat{y}_n$ を平均値とし，標準偏差 $\sqrt{x_n^2\sigma_e^2}$ に従うとして，Excel の Normdist 関数で確率密度を計算し，その対数を取り，$n=40$ データについて加えて求める．

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		i		y_i		切片	x_i^{λ}		β^		y^		$\ln Li$
3		1		6.31		1	6.19		0.2469	=	6.34		1.7430
4		2		9.27		1	9.17		0.9851		9.28		1.5217
5		3		2.21		1	1.99				2.21		3.0472
6		4		8.00		1	7.77				7.90		0.7901
7		5		4.27		1	4.00				4.19		−0.0355
8													
9		40		10.17		1	9.98				10.08		0.9731
10							σ_e^2^	=	0.00008886				
11										合計	$\ln L$	=	61.0369
12													
13		1.7430		=LN(NORMDIST(D3, K3, SQRT((G3^2)*I10), FALSE))									

図 9.7　対数尤度の計算方法

9.2. ［付録 2］Draperら（1998）

［2］ Draper.N.R., Smith.N. (1998), Applied Regression Analysis, 3rd ed, 83-86, 226-229, Wiley.

逆推定の 95%信頼区間の推定について 3.2 節に解説がある．式の導出の考え方が示されているが，事例は示されていない．なお，この本の第 1 版（1966）は，「応用回帰分析」として中村慶一によって翻訳されて森北出版より 1968 年に出版されているが，この問題は含まれていない．

重みつき最小 2 乗法の回帰分析への適用した数少ない事例が，9.4 節 p226 に示されている．さらに，「重みは分散の逆数，重みは x の関数で与える」との明確に示されている貴重な文献である．

INVERSE REGRESSION（原著の引用）

3.2. INVERSE REGRESSION (STRAIGHT LINE CASE)

Suppose we have fitted a straight line $\hat{Y} = b_0 + b_1 X$ to a set of data (X_i , Y_i), i = 1, 2, ... , n, and now, for a specified value of Y, say Y_0, we wish to obtain a predicted value $\hat{X}_0$, the corresponding value of X, as well as some sort of interval confidence statement for X surrounding lo. A practical example of such a problem is the following: （中略）

To obtain the values of X_L and X_U we can proceed as follows. In the figure, X_L is the X-coordinate of the point of intersection of the line

$$Y = Y_0 \qquad \text{(i.e.} \quad Y = b_0 + b_1 \hat{X}_0 \text{)} \tag{3.2.1}$$

and the curve

$$Y = Y_{X_L} - ts\left\{\left\{\frac{1}{n} + \frac{(X_L - \bar{X})^2}{S_{XX}}\right\}\right.^{1/2} \tag{3.2.2}$$

where $S_{XX} = \sum (X_i - \bar{X})^2$, $Y_{X_L} = b_0 + b_1 X_L$, $t = t(\nu, 1 - \alpha/2)$ is the usual t percentage point, and ν is the number of degrees of freedom of s^2 . Setting Eqs. (3.2.1) and (3.2.2) equal, canceling a b_0 , rearranging to leave the square root alone on one side of the equation, and squaring both sides to get rid of the square root leads to a quadratic equation

$$PX_L^2 + 2QX_L + R = 0 \tag{3.2.3}$$

in X_L , where

$$\begin{aligned} P &= b_1^2 - t^2 s^2 / S_{XX} , \\ Q &= t^2 s^2 \bar{X} / S_{XX} - b_1^2 \hat{X}_0 , \\ R &= b_1^2 \hat{X}_0^2 - t^2 s^2 / n - t^2 s^2 \bar{X}^2 / S_{XX} . \end{aligned} \tag{3.2.4}$$

（高橋のコメント）

この節では，逆推定の 95%信頼区間の導出方法が示されている．2.5 節 逆推定の定式化もこの Draper と同様の導出方法を用いている．だだし，この節では，導出の過程が省略されているために，個別データの場合，重みがある場合などの拡張する場合に見通しが悪い．（3.2.2）式に対応するのは，式（2.26）で，

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

とし，$Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95})$ は，

$$Var(\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{L95}) = Var(\hat{\beta}_0) + 2Cov(\hat{\beta}_0, \hat{\beta}_1)\hat{x}_{L95} + Var(\hat{\beta}_1)\hat{x}_{L95}^2$$

のように，推定したい $\hat{x}_{L95}$ についての 2 次式に展開し，このままの形式で定式化した．これは, 実際の計算で Excel の行列計算によって回帰パラメータの分散共分散行列を直接計算し，これを用いることに対応している．（コメント終り）

WEIGHTED LEAST SQUARES（原著の引用）

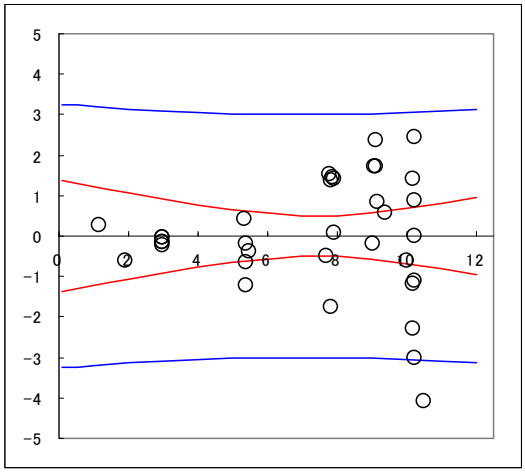
9.4　A NUMERICAL EXAMPLE OF WEIGHTED LEAST SQUARES

The data in Table 9.1, which have been rearranged in an order convenient for purposes of analysis, consist of 35 observations (X_i, Y_i,) with a number of sets that are either exact repeats at the same X-value or approximate repeats. These are indicated by the groupings. A fit of the data by (ordinary) least squares produces the fitted model.

原著・Table 9.1　Data for Weighted Least Squares Example

i	x_i	y_i	w_i	i	x_i	y_i	w_i
1	1.15	0.99	1.24	18	7.91	9.81	0.80
2	1.90	0.98	2.18	19	7.94	8.50	0.78
3	3.00	2.60	7.85	20	9.03	9.47	0.47
4	3.00	2.67	7.85	21	9.07	11.45	0.47
5	3.00	2.66	7.85	22	9.11	12.14	0.46
6	3.00	2.78	7.85	23	9.14	11.50	0.45
7	3.00	2.80	7.85	24	9.16	10.65	0.45
8	5.34	5.92	7.44	25	9.37	10.64	0.41
9	5.38	5.35	6.99	26	10.03	10.19	0.33
10	5.40	4.33	6.79	27	10.17	9.78	0.31
11	5.40	4.89	6.79	28	10.18	12.39	0.31
12	5.45	5.21	6.31	29	10.18	8.68	0.31
13	7.70	7.68	0.89	30	10.22	11.03	0.31
14	7.80	9.81	0.84	31	10.22	8.00	0.31
15	7.81	6.52	0.84	32	10.22	11.90	0.31
16	7.85	9.71	0.82	33	10.23	13.46	0.31
17	7.87	9.82	0.81	34	10.23	9.93	0.31
				35	10.50	7.25	0.28

$\hat{Y} = -0.5790 + 1.1354X$ and the residuals plots in Figure 9.1. A clear indication that the observations have unequal variances is seen. The overall plot of residuals (not shown) is somewhat skewed toward negative values, also. None of the usual (ordinary) least squares analyses are appropriate and it seems sensible to apply generalized least squares.



原著・Figure 9.1　重みなしの回帰分析の残差プロット（95%信頼区間は高橋の追加）

We assume (until contrary indications appear) that the Y_i are independent so that **V** has the diagonal pattern with different variances given earlier. We now need to obtain information on the variance pattern. For each of the sets of repeats or near repeats we evaluate the average X-value, X_j say, and the pure error mean square s_{ej}^2. These are:

$\bar{X}$	3	5.4	7.8	9.1	10.2
s_{ej}^2	0.0072	0.3440	1.7404	0.8683	3.8964

A plot of these suggests a quadratic relationship, which we estimate by least squares as

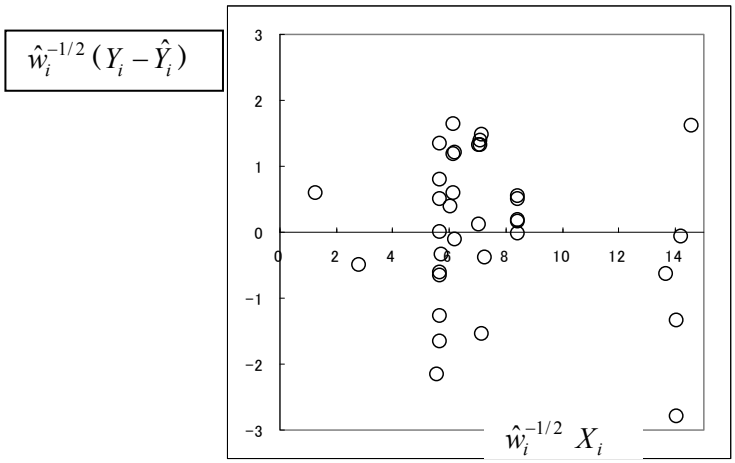
$$\hat{s}_e^2 = 1.5329 - 0.7334\bar{X} + 0.0883\bar{X}^2$$

We can now substitute each individual X_i, into this equation, estimate s_{ej}^2, i = 1, 2, …, 35, and invert these values to give the estimated weights $\hat{w}_i$ shown in the table. The matrix **P** of our text is diagonal with entries $\hat{w}_i^{-1/2}$. Using these weights leads to the weighted least squares prediction equation $\hat{Y} = -0.8891 + 1.1648X$ and an analysis of variance table as follows:

Source	df	SS	MS
$b_1 \mid b_0$	1	496.96	496.96
Residual	33	42.66	1.29
Total, corrected	34	539.62	

The appropriate“observations”and“fitted values”are now $\hat{w}_i^{-1/2}Y_i$ and $\hat{w}_i^{-1/2}\hat{Y}$ and the“residuals” to be examined are $\hat{w}_i^{-1/2}(Y_i - \hat{Y}_i)$, notice. An overall plot of residuals still shows some skewness but the pattern is slightly better behaved. The residuals plots in Figure 9.2 reveal that the vertical spread of residuals is now roughly the same at the two main levels of the transformed response. (At lower levels there are only two observations so that there is not much of an estimate of the spread there.) The employment of weighted least squares here appears to be justified and useful.

A weighted least squares program exists in most computing systems, but some do not have a generalized least squares program.



原著・Figure 9.2　重み付き回帰分析の残差プロット

（高橋の追加）

通常の回帰分析を行なうと，その残差プロットは次のように扇方となり，何らかの重み付き回帰の必要性が示唆される．

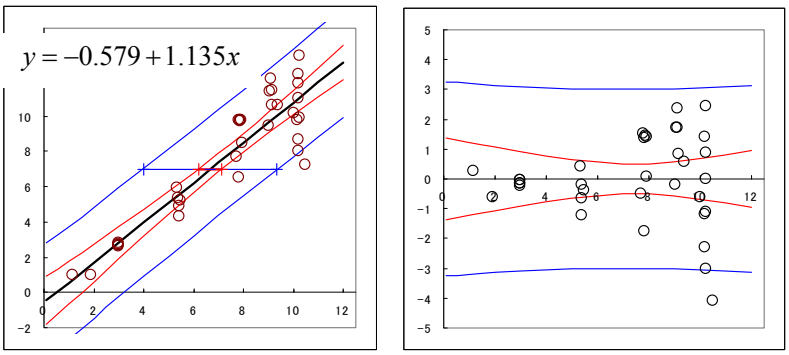


図 9.8 重みなしの回帰分析の結果（95%信頼区間は高橋の追加 ln L = -61.827）

Draper らは，分散を $\bar{X}$ の 2 次式をデータから推定し，その逆数を重みとして用いている．この 2 次式での分散の推定は，図 9.9 に示すように X が小さくなると大きくなってしまい，実際のデータに適合しない．

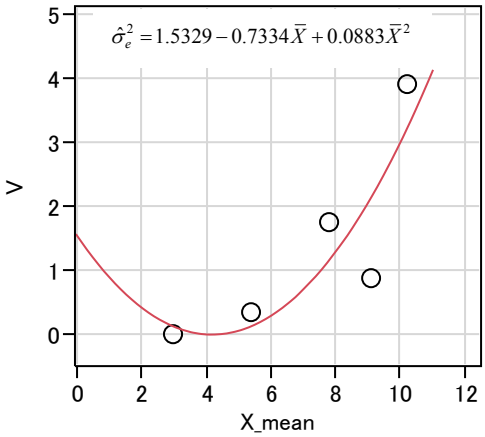


図 9.9　X の 2 次式を用いた分散の推定

実際に 2 次式で与えられた分散の逆数を使った重み付き回帰の結果は，図 9.10 に示すよ

うに，個別データの 95%信頼区間は原点付近で拡大し，扇型の残差プロットを反映していない．

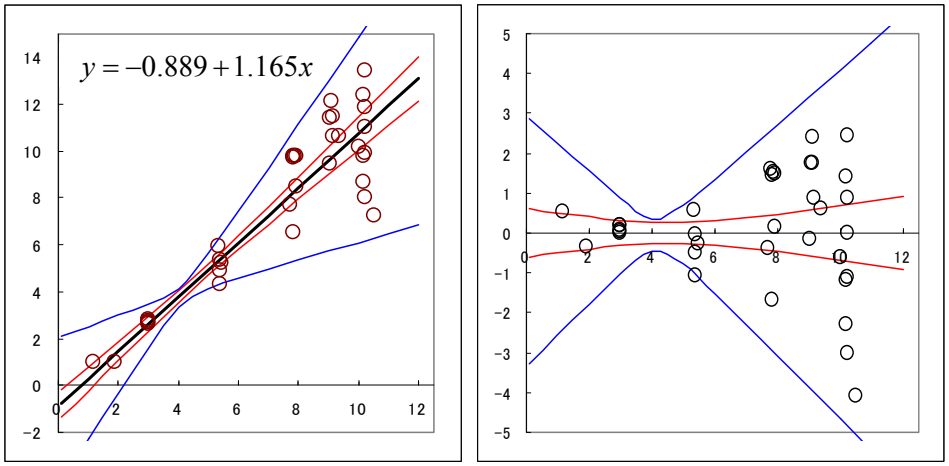


図 9.10　2 次式の分散の逆数を重みにした回帰（$\ln L$ = -51.648）

X に関して分散の推定値が単調に増加することが経験的に知られており，このためには，2 次式の中の X の一次項を除いて $\bar{X}^2$ 関して 1 次式とすることにより実現できる．

$$s_{ej}^2 = \beta_0 + \beta_1 \bar{X}^2$$

1 次式での切片は $\beta_0 = -0.5555$ と負となるので，切片なしの X^2 に関する 1 次式をあてはめると $\beta_1 = 0.02658$ を得る．

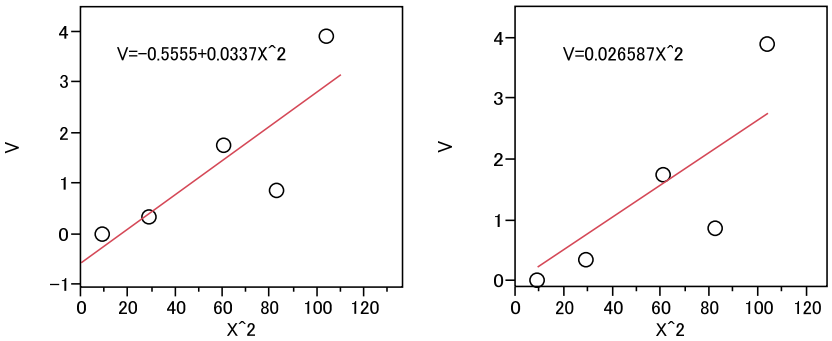


図 9.11　推定分散に $\bar{X}^2$ に関して 1 次式のあてはめ

分散 $V' = s_{ej}'^2 = 0.02689\bar{X}^2$　の逆数を重みとする回帰分析を行い次の結果を得る．

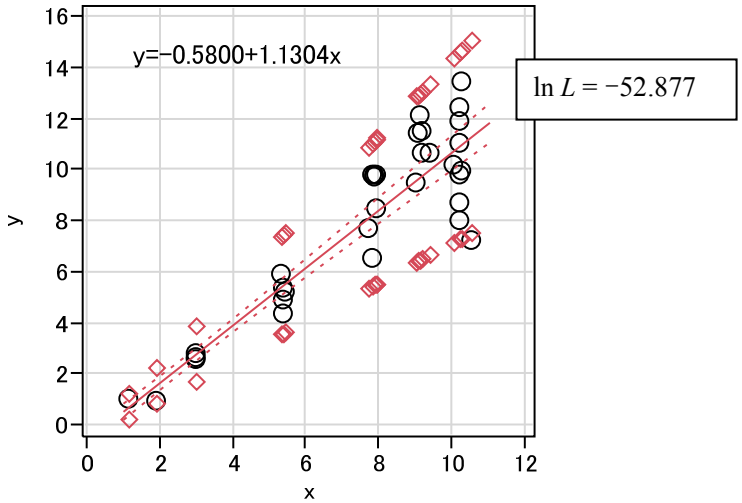


図 9.12　$0.02689\bar{X}^2$ の逆数を重みにした 1 次式のあてはめ

Draper らは，得られたデータである X について重複したいくつかの点から分散を計算し，X の 2 次式から分散を推定し，その逆数をそれぞれ X に対する重みとした．この考え方は，重み付き回帰分析の「重み」をどのように与えたらよのかについて示唆に富んだ切り口である．

重み付き回帰分析において，どのように「重み」を与えるか百家争鳴となりがちであるが，Draper らの「重みは分散の逆数，ただし X の関数で分散を推定する」考え方は，正論である．しかしながら，一部のデータによって分散を X の関数として推論することは便宜的といわざるを得ない．原則は，得られたデータを全て用いて推定し，恣意性が入らないようにする必要があるが，その時代その時代の限られた計算手段のでの対応が現実的であるので便宜的ではるが最善の方法を選択せざるを得ない．

重みがおおよその分散の推定値 $0.02689\,X^2$ の逆数に従うということは，分散が X^2 の逆数に比例するということと同じである．実際にそれぞれの重みで回帰分析を行なうと，分散分析表の平方和は異なるが，F 値は一致し，推定されたパラメータ，SE は完全に一致する．

重み ＝ $1/(0.02689x^2)$

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	674.7146	674.7146	606.7450
誤差	33	36.6968	1.1120	p値(Prob>F)
全体(修正済み)	34	711.4113		<.0001*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	−0.5800	0.1900	−3.05	0.0045*
x	1.1304	0.0459	24.63	<.0001*

重み ＝ $1/x^2$

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	18.1431	18.1431	606.7450
誤差	33	0.9868	0.0299	p値(Prob>F)
全体(修正済み)	34	19.1299		<.0001*

パラメータ推定値

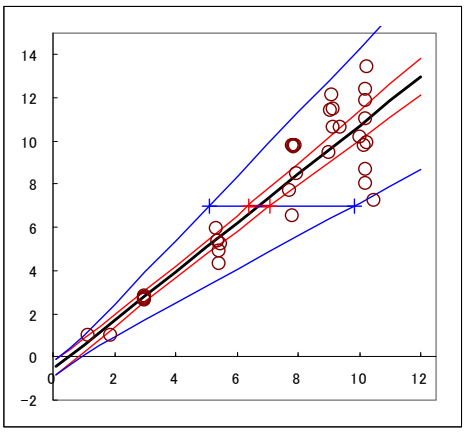
項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	−0.5800	0.1900	−3.05	0.0045*
x	1.1304	0.0459	24.63	<.0001*

図 9.13　重みを簡単化した回帰分析の分散分析表およびパラメータの推定

分散分析表の平均平方は，重み ＝ $1/(0.02689x^2)$ の場合 1.1120 であり，重み ＝ $1/x^2$ の場

合 0.0299 と異なる．重み ＝ $1/(0.02689x^2)$ の場合の平均平方は，$x^2 = 1/0.02689 = 37.19$ の場合の分散となっている．重み ＝ $1/x^2$ の場合の平均平方 0.0299 は $x^2 = 1$ の場合であり，$x^2 = 37.19$ の場合に換算するためには，37.19 倍すると，$0.0299 \times 37.19 = 1.1120$ と同じ平均平方が得られる．

どちらの重みでも本質的に同じなので，推定された平均平方は，$x^2 = 1$ の場合の分散の推定値となっていると明示的に言えるので「重み ＝ $1/x^2$」を使うことを推奨する．Excel の計算シートを用いて個別データの逆推定および 95%信頼区間の推定結果を図 9.14 に示す．



				x^{λ}
逆推定 y_0 =	7.0		x_0^ =	6.706
回帰直線の95%CL				
a =	−13.845		$x_{L\,95R}$^ =	6.365
b =	4.127		$x_{U\,95R}$^ =	7.095
c =	−0.307			
個々のデータの95%CL				x
a' =	−13.845		$x_{L\,95I}$^ =	5.097
b' =	4.127		$x_{U\,95I}$^ =	9.818
c' =	−0.277			
m 個の平均の95%CL				x^{λ}
a'' =	−13.845		$x_{L\,95m}$^ =	5.648
b'' =	4.127		$x_{U\,95m}$^ =	8.264
c'' =	−0.297			

図 9.14　Excel による逆推定

9.3.［付録 3］Snedecorら（1967），畑村ら共訳（1972）

［3］Snedecor G.W., Cochran W.G.(1967), 畑村又好, 奥野忠一, 津村善郎 共訳（1972）, 統計的方法, 原著第 6 版, 144-155, 岩波書店.

回帰直線の逆推定の 95%信頼区間について事例が示されている書物はなかなか見出すことができない．唯一，畑村，奥野，津村共訳（1972），「統計的方法，原著第 6 版」の第 6 章の「りんごの収量にたいする虫害果の百分率の回帰」に個別データの逆推定のおける 95%信頼区間の例示を見出すことができた．この成書は，多くの具体的な事例に対した広範な統計的方法を論じた名著である（原著は第 8 版（1989）となっている）.

X が 標本ごとに変わる場合（原著の引用）

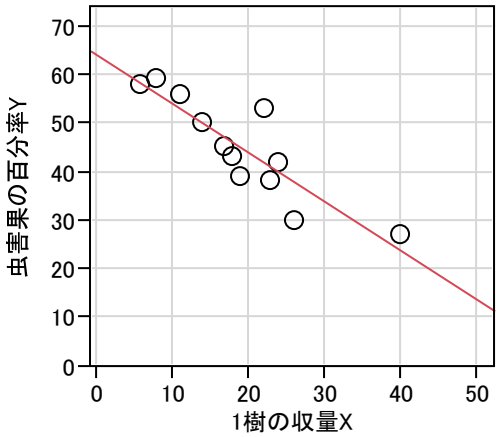
6.9　*X* が 標本ごとに変わる場合

ふつう, コドリンガの幼虫による虫害果の百分率は, できのわるいリンゴの木ほど大きい, と考えられている. 蛾の飛来する密度が一様になろうとする傾向のあるのは明らかであって, そのため特定の果実がおそわれる機会は,木に果実が少なければ増大するのである．表 6.9.1 のデータは，この現象を解明するためのものである．12 本の木にはすべて，砒鉛を萼（がく）に散布し，ついで 100 ガロンあたり，3 ポンドの砒酸マンガンと 1 クォートの魚油を，5 回にわたり全面散布した.

原著・表 6.9.1　虫害果の百分率

樹木番号	1樹の収容（100果）	虫害果の百分率
	X	*Y*
1	8	59
2	6	58
3	11	56
4	22	53
5	14	50
6	17	45
7	18	43
8	24	42
9	19	39
10	23	38
11	26	30
12	40	27

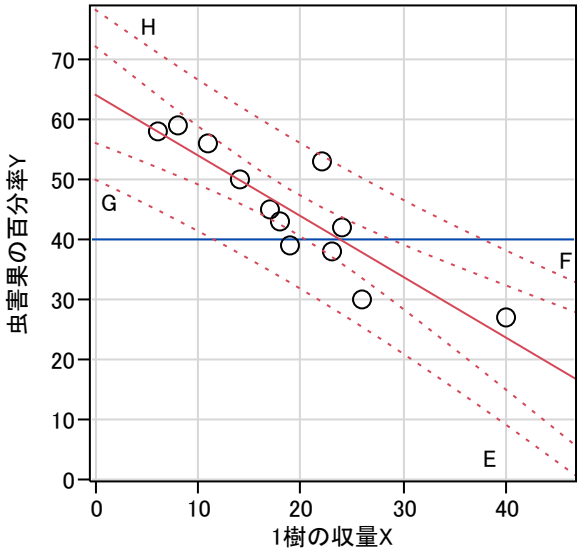
図 6.9.1 にはっきり示されているように，樹上のリンゴの数が増加するにつれて，虫害果の百分率が減少するという傾向は顕著である．この特定の木の群では，2 変量の関係はふつうの場合よりずっと緊密である．（中略）



原著・図 6.9.1　りんごの木における収量にたいする虫害果百分率の標本回帰

6.14　Y からの X の予測，線型校正

（前略）　虫害果が 40%である新しい木の収量の 95%信頼限界を求めることにしよう．図 6.9.1 にもどる．Y=40 のところで水平直線をひく．2 つの信頼限界は,この直線が信頼曲線 GE および HF にであう点での X の値である．目分量によれば，X=12 および X=38 である．X の点推定値 $\hat{X}$ は, もちろん, あてはめられた回帰直線にこの水平直線があうところでの X の値, 24 である.



原著・図 6.11.1　μ の信頼幅

数値解としては，あてはめた直線は $\bar{Y}+bx$，ただし，$\bar{Y}=45$，$b=-1.013$ である．したがって．$Y=40$ のときの x の値はつぎのように推定される．

$$\hat{x}=(Y-\bar{Y})/b=-(40-45)/1.031=4.936$$

$$\hat{X}=\bar{X}+\hat{x}=19+4.936=23.936 \quad \text{（100 果単位）}$$

x についての 95%信頼限界を求めるには，x が与えられたときの Y の信頼限界からはじめる．

$$Y = \bar{Y} + bx \pm t\, s_{y\cdot x}\sqrt{1 + \frac{1}{n} + \frac{x^2}{\sum x^2}}$$

ここに t は $(n-2)d.f.$ での 5%水準である．この式は Y が与えられたときの 2 次方程式としてとける．多少の操作ののち，2 つの根はつぎに示す形であらわされる．そしてこれがもっとも計算に楽のようである．

$$x = \frac{\hat{x} \pm \dfrac{t\, s_{y\cdot x}}{b}\sqrt{\dfrac{n+1}{n}(1-c^2) + \dfrac{\hat{x}^2}{\sum x^2}}}{1-c^2}$$

ここに,

$$c^2 = \frac{t\, s_b^2}{b^2} = \frac{1}{\sum x^2}\left(\frac{t\, s_{y\cdot x}}{b}\right)^2$$

この例では，$n = 12$，$t = 2.228(10d.f.)$，$s_{y\cdot x} = 5.233$，$\sum x^2 = 924$，$b = -1.013$，$\hat{x} = 4.936$ である．したがってつぎの値がえられる．

$$c^2 = \frac{1}{942} \times \left(\frac{2.228 \times 5.233}{-1.013}\right)^2 = \frac{1}{942} \times (-11.509)^2 = 0.1434$$

これから，x についての限界はつぎのようになる．

$$x = \frac{4.936 \pm 11.509\sqrt{\dfrac{12+1}{12}(1-0.1434) + \dfrac{4.936^2}{924}}}{1-0.1434}$$

これから，x については -7.4 および $+18.9$，X については 11.6 および 37.9 がえられ，これはグラフでの推定値とよく一致する．

（高橋追加）

JMP の【モデルのあてはめ／推定値／逆推定／個別の反応に対する信頼区間】によって，次の計算結果が得られ，文献の値（11.6，37.9）と一致する．

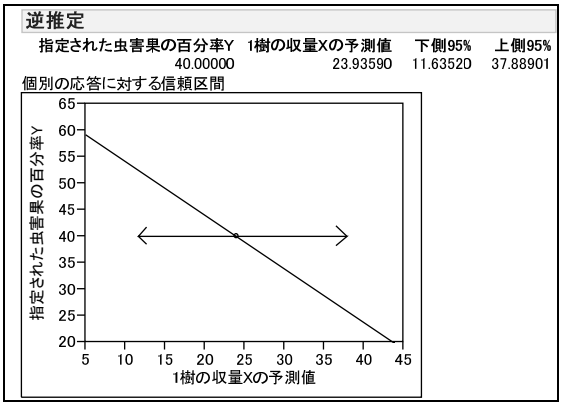


図 9.15　JMP による Y=40 における個別データの 95%信頼区間

原著・図 6.11.1 μ の信頼幅の散布図から，収量 X が小さいときに残差が小さい．そこで，重み付き回帰を行い，対数尤度を計算し，最適な重みを推定しよう．

重み=0，$\ln L$=−35.887　　重み=1/x，$\ln L$=−34.531　　重み=1/x^2，$\ln L$=−33.754

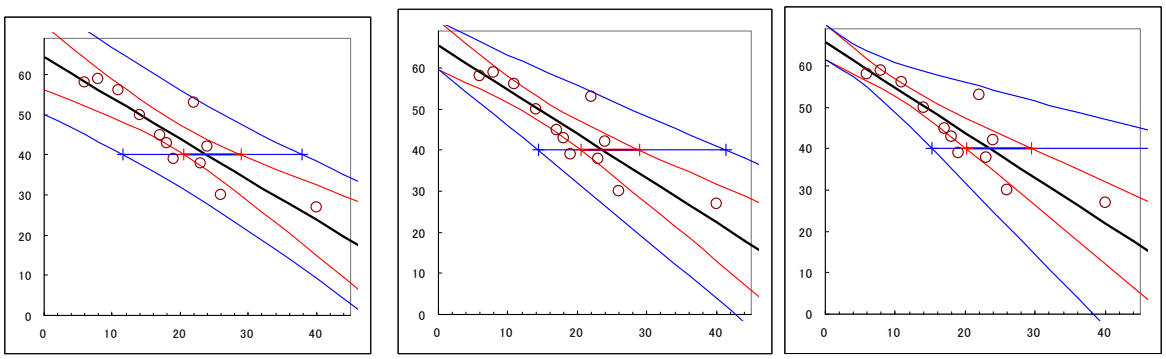


図 9.16　重みを換えた場合の対数尤度の変化

重みが 1/x^2 の場合に対数尤度 $\ln L$=−33.754 と小さくなる．なお，Excel のソルバーを用いた最尤法で，対数尤度が最大になる重みは，1/$x^{2.24}$ で，対数尤度は $\ln L$=−33.624 と大きくなっている．

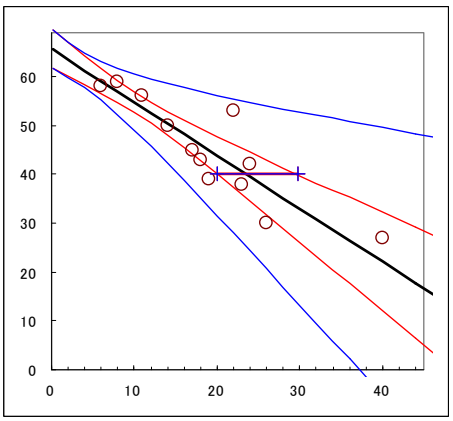


図 9.17　最適な重み 1/$x^{2.24}$ に対する回帰

9.4.［付録 4］芳賀ら(1996)

［4］芳賀敏郎, 野澤昌弘, 岸本淳司(1996), SAS による回帰分析, 181-191, 東大出版会.

第 9 章に重み付き回帰について詳しく解説されている.「9.1 いろいろの平均, 9.2 誤差が x の関数である場合, 9.3 proc reg による重み付き回帰分析, 9.4 誤差分散が η の関数である場合, 9.5 proc、nlin による解析, 9.6 曲線関係と不等分散性の関係重み付き回帰について」など, 事例とともに詳しく述べられている.

前書き（原著の引用）

9 章　前書き

等分散が成立しない場合は, 重みつき最小自乗法を用いる必要があるというのが公式的な解答である. 分散が説明変数 x の関数で表される場合は比較的簡単に解くことができる. しかし, 分散が目的変数 y の関数であるときは, 逐次計算が必要であり, 工夫を要する. 分散が等しくないことを無視して通常の回帰分析を適用すると,

i)　推定された回帰式が母回帰式の不偏推定になっている点では問題が生じない.

ii)　分散の差が大きいとき, 回帰係数の分散が大きくなり, 分散の推定値が過小評価されるために回帰係数の区間推定の信頼率が低下する.

（中略）

いろいろの平均（原著の引用）

9.1 節　いろいろの平均

ある建設会社で, 25 の建物についてその面積 x_i と建築 y_i の実績値を収集し, それから単位面積当たりの標準単価を求めようとしている. そのデータを 図表 9.1 に示す.

原著・図表 9.1　面積と建築費のデータ

i	x_i	y_i	y_i/x_i	i	x_i	y_i	y_i/x_i
1	49	120	2.449	14	180	485	2.694
2	58	138	2.379	15	183	473	2.585
3	73	185	2.534	16	199	442	2.221
4	75	192	2.560	17	199	613	3.080
5	79	183	2.316	18	206	507	2.461
6	134	296	2.209	19	212	400	1.887
7	149	314	2.107	20	213	579	2.718
8	150	415	2.767	21	224	381	1.701
9	151	228	1.510	22	248	385	1.552
10	153	425	2.778	23	256	584	2.281
11	159	235	1.478	24	258	458	1.775
12	171	408	2.386	25	271	390	1.439
13	179	445	2.486	計	4229	9281	56.355
				単価	2.195		2.254

単純に考えて，建物ごとに単価 $@_i = y_i / x_i$ を計算し，それを平均した単価 2.254 と，合計金

額 9281 を合計面積 4229 で割った単価 2.195 を求めたが，2 つの値は一致しない．どちらの計算方法が良いのだろうか．またもっと良い方法はないのだろうか．（中略）

Proc reg による重みつき回帰（原著の要約）

9.3 節　Proc reg による重みつき回帰

（前略）このデータについて，通常の回帰分析，重みを WT1=$1/x_i$ と WT2=$1/x_i^2$ とした重みつき回帰分析を実行する．ただし，いずれも原点を通るという条件を入れる．プログラムと出力を次に示す．

（結果の要約）推定された傾きは，

通常の回帰分析：　$\hat{\beta}_1 = 2.141$

重みを WT1=$1/x_i$ ：　$\hat{\beta}_1' = 2.195$　　合計金額/合計面積

重みを WT2=$1/x_i^2$ ：　$\hat{\beta}_1'' = 2.254$　　単価 $@_i$ の平均

のように対応付けられる．

（高橋追加）

重み=1，　ln L=-148.3　　重み=$1/x$，　ln L=-145.0　　重み=$1/x^2$，　ln L=--142.2　　重み=$1/x^3$，　ln L=-140.0

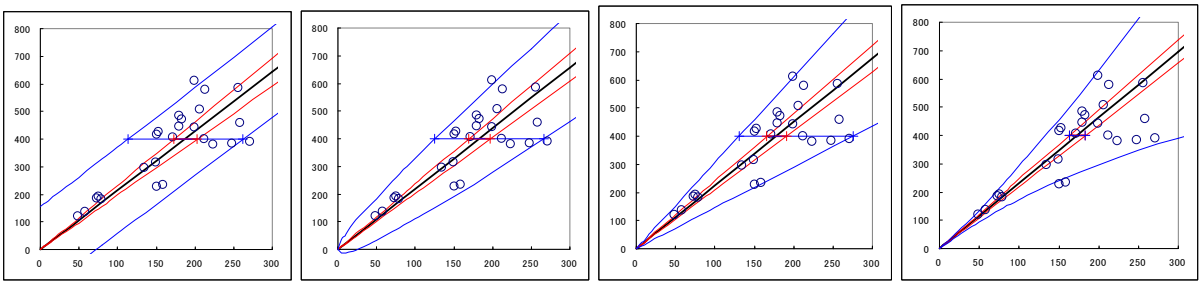


図 9.18　原点を通る各種の重み付き回帰

芳賀らは，原点を通る回帰の重みの選択に際し，各種の重み付き平均と推定されたパラメータとを対比して，重みの意味付けを試みている．どの重みを選ぶべきなのかについては示していない．重みは，分散の逆数とすることが原則であり，原点を通る場合の最適な重みを求めると重み=$1/x^{5.1492}$ の場合に ln L=-137.861 と最大になる．推定された分散は，$x^{5.1492}1.3637\times10^{-8}$ であり $x=10$ の場合 SD = 0.0439，$x=100$ の場合 SD = 16.47，$x=200$ の場合 $SD=98.10$ と推定される（高橋追加終）．

誤差分散が η の関数である場合（原著の引用）

9.4 節　誤差分散が η の関数である場合

回帰直線が原点を通らない一般の場合，

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{9.15}$$

で，ε の標準偏差 σ_i または分散 σ_i^2 が y_i の期待値 η_i に比例していると考えられる場合に，重みつき回帰分析によって，β_0，β_1 を推定する方法を考える．

原点を通る回帰式では「η_i に比例する」を「x_i に比例する」と置き換えることができたが，式 (9.15) のモデルではそれができない．β_0 が小さいときは「誤差の標準偏差または分散が x_i

に比例する」と近似的に置き換えることができるが，一般的ではない．

誤差ε_iの分散σ_i^2がy_iの期待値η_iに比例するというモデルを

$$y_i = \eta_i + \varepsilon_i$$
$$= \beta_0 + \beta_1 x_i + \varepsilon_i \qquad \varepsilon_i \sim N(0, \eta_i \sigma^2)$$

と表す．ここで，η_iは面積がx_iのときの金額y_iの期待値である．実際の金額y_iはη_iに誤差ε_iを加えたものである．（中略）

（結果の要約）Proc reg で通常の回帰分析を行い［0 次解］，$b_0^{(0)}$，$b_1^{(0)}$を得る．［1 次解］$\hat{\eta}_i^{(0)} = b_0^{(0)} + b_1^{(0)} x_i$を重みとした回帰分析を行い$b_0^{(1)}$，$b_1^{(1)}$を得る．［2 次解］$\hat{\eta}_i^{(1)} = \hat{\beta}_0^{(1)} + \hat{\beta}_1^{(1)} x_i$を重みとした回帰分析を行い$b_0^{(2)}$，$b_1^{(2)}$を得る．［3 次解］$\hat{\eta}_i^{(2)} = b_0^{(2)} + b_1^{(2)} x_i$を重みとした回帰分析を行い$b_0^{(3)}$，$b_1^{(3)}$を得る．

原著・図表 9.3 逐次計算の経過

	1次解	2次解	3次解
b_0	47.232650	44.856772	44.652586
b_1	1.915390	1.929435	1.930642

分散がη_i^2に比例する場合も同様な手順を繰り返すことによって，

$$\hat{y} = 27.334 + 2.052x$$

が得られる．

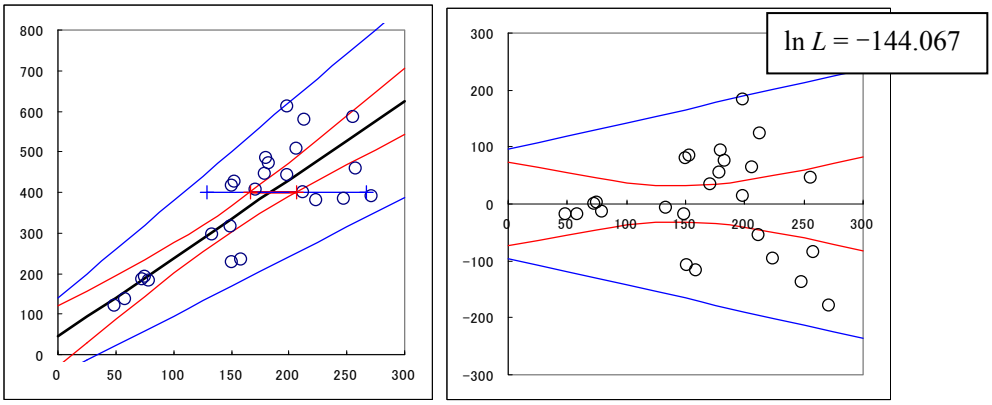
（高橋追加）

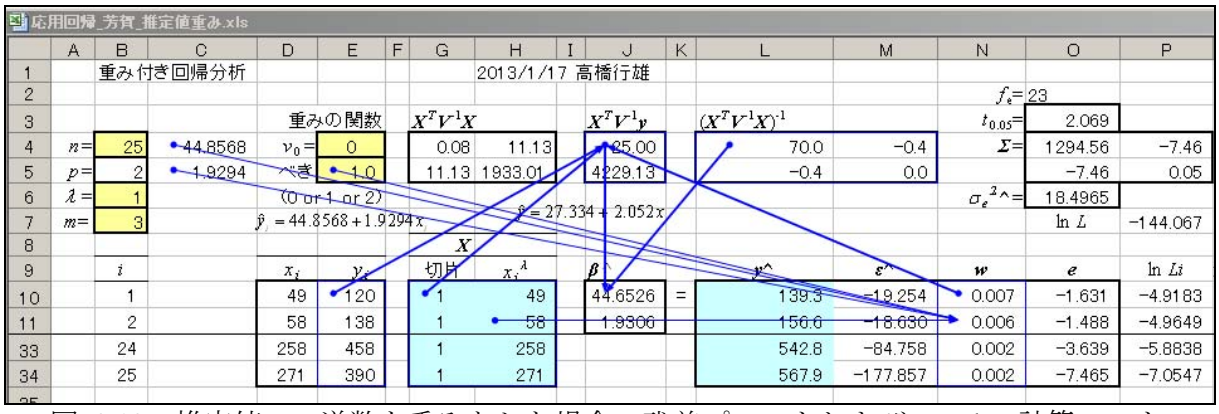
2 次解が得られたとして，重み$= 1/\hat{y}_i^{(2)} = 1/(44.8568 + 1.9294x_i)$とした重みつき回帰を行ない，パラメータとして表 9.4 に示すように 3 次解が得られる．

表 9.4　回帰の推定値の逆数を重みとした回帰による 3 次解

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k$^	$Var(\beta_k$^$)$	SE	t	p
0	x_0:切片	β_0	44.6526	1294.560	35.9800	1.241	0.2271
1	x_1:傾き	β_1	1.9306	0.053	0.2292	8.424	0.0000

2 次解　$\hat{y}_i = 44.8568 + 1.9294x_i$の逆数を用いた重み付き回帰についての残差プロットを図 9.19 に示す．回帰の推定値を重みとした場合であっても，本質的にはxについての関数で与えていることになる．このデータの散布図から推測される分散の構造は，0 に近づくと推定されるのであるが，$x = 0$の分散の推定値を回帰式の切片の 44.8568 とすることになり，矛盾が生ずる．





応用回帰_芳賀_推定値重み.xls

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		重み付き回帰分析						2013/1/17 高橋行雄								
2														f_e=	23	
3				重みの関数			$X^TV^{-1}X$			$X^TV^{-1}y$		$(X^TV^{-1}X)^{-1}$		$t_{0.05}$=	2.069	
4	n=	25	44.8568	v_0=	0		0.08	11.13		25.00		70.0	−0.4	Σ=	1294.56	−7.46
5	p=	2	1.9294	べき	1.0		11.13	1933.01		4229.13		−0.4	0.0		−7.46	0.05
6	λ=	1		(0 or 1 or 2)										σ_e^2^=	18.4965	
7	m=	3		$\hat{y}_i$ = 44.8568 + 1.9294x						$\hat{\beta}$ = 27.334 + 2.052x					ln L	−144.067
8								X								
9		i		x_i	y_i		切片	x_i^{λ}		$\hat{\beta}$		$\hat{y}$	$\hat{\varepsilon}$	w	e	ln Li
10		1		49	120		1	49		44.6526	=	139.3	−19.254	0.007	−1.631	−4.9183
11		2		58	138		1	58		1.9306		156.6	−18.630	0.006	−1.488	−4.9649
33		24		258	458		1	258				542.8	−84.758	0.002	−3.639	−5.8838
34		25		271	390		1	271				567.9	−177.857	0.002	−7.465	−7.0547

図 9.19　推定値 y の逆数を重みとした場合の残差プロットおよび Excel の計算シート

芳賀らは，9.5 節で，非線形最小 2 乗法を用いて，回帰の推定値を重みとし，平均平方が最小になる回帰式を推定している．その結果 7 回の繰返し計算で $\hat{y}_i = 44.6334 + 1.9307x_i$ が得られ，3 次解とほぼ同じと結論している．

重み $= 1/\hat{y}_i^2 = 1/(27.334 + 2.052x_i)^2$ とした場合について，に残差プロットを図 9.20 に示す．

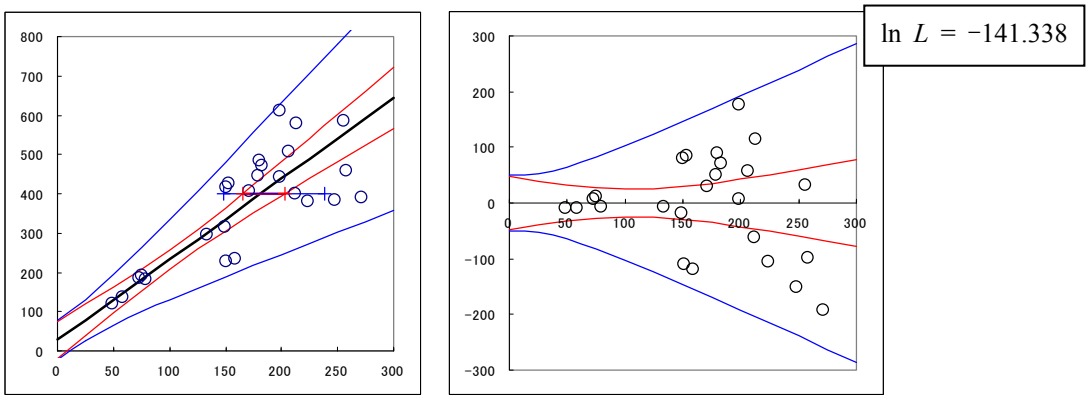


図 9.20　推定値 y の 2 乗の逆数を重みとした場合の残差プロット
個別データの逆推定の 95%信頼区間は 2 次式の解では計算できない

重みを $1/\hat{y}_i^2$ とした場合の対数尤度 ln L = -141.338 と 重みを $1/\hat{y}_i$ とした対数尤度 ln L = -144.067 を比べると，重みを $1/\hat{y}_i^2$ とした場合の方が大きい．

重みを $1/x^{\gamma}$ とし，γ を（0，1，2，4）と変化させ，対数尤度を求めると次第に大きくなる．対数尤度が最大となるのは，$\gamma = 4.89$ であることが，Excel のソルバーを用いた最尤法の結果から得られる．

$1/x^0 = 1$	$1/x^1$	$1/x^2$	$1/x^4$	$1/x^{4.89}$
$\ln L$ =-147.105	$\ln L$ =-144.064	$\ln L$ =-141.260	$\ln L$ =-137.461	$\ln L$ =-136.985

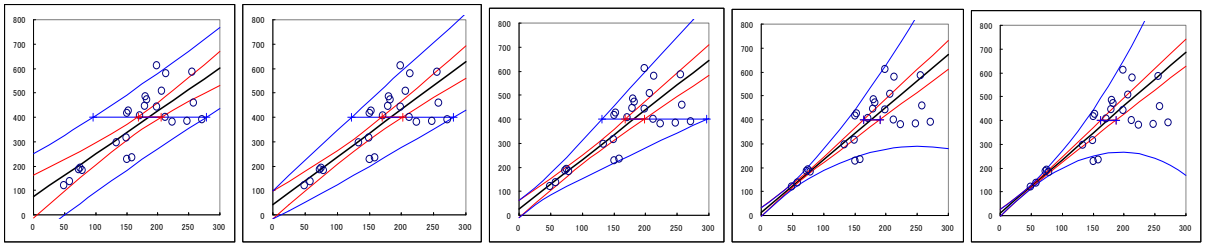


図 9.21　重みを $1/x^{\gamma}$ とし γ を（0，1，2，4，4.89）と変化させた場合の残差プロット

ロジスティック回帰における分散の推定の考え方

「回帰の推定値を重みする」考え方は，2 値反応の用量反応についてロジスティック曲線をあてはめる場合に，重み付き回帰分析を繰り返すことにより最尤解を得てきた歴史的な経験に基づくと思われる．解析の手順は次のようである．

1）　用量 x_i について（反応あり r_i，反応なし $n_i - r_i$）の例数をロジット変換 $\ln(r_i/(n_i - r_i))$ で線型化し，ロジットについて回帰分析を行なう．
2）　回帰パラメータから出現率 $\hat{p}_i$ を推定し，x_i における分散を $n_i\hat{p}_i(1-\hat{p}_i)$ 推定し，その逆数を重みとする回帰分析を行う．
3）　2）で推定された回帰パラメータを用いて，x_i における分散を推定し，その逆数を重みとする回帰分析を行う．
4）　3）を繰り返し，何らかの収束条件を用いて繰返しを終了し，その時点の回帰パラメータから得られたロジスティック曲線が，最尤法による解と一致する．

この方法は，推定された重み付き回帰パラメータを用いて，x_i における分散を推定している．したがって，重み付き回帰の原則である「重みは分散の逆数」が保たれている．

通常の回帰分析における分散の推定

芳賀らの「重み付き回帰分析」における重みは，$1/\hat{y}_i$ あるいは $1/\hat{y}_i^2$ であり，これらは，x_i における分散を推定値として $\hat{y}_i$ とする理論的な裏付けはできない．9.3 節の「りんごの木における収量にたいする虫害果百分率」では，回帰直線の傾きが負であるような場合に $x = 0$ の場合に分散が切片の大きさに依存することになり，x の増加に伴って $\hat{y}_i$ が減少するにつれ分散が減少することになり，このデータの分散の構造とは逆になっている．

回帰分析における重みに，$1/\hat{y}_i$ あるいは $1/\hat{y}_i^2$ を使うことには，慎重な態度をとるべきで

あり，6.3 節 および 6.4 節 で提示した重み $1/(v_0+x)$ あるいは $1/(v_0+x^2)$ を使うことを推奨する．

同じ x_i について複数回の測定が行なわれていれば，それから分散を推定し，その逆数を重みとすることが可能である．また，分散の推定に x に関する 1 次式があてはめられれば，重みにその 1 次式をそのまま使うことができる．

$$\hat{e}_i^2 = \hat{v}_0 + \hat{v}_1 x_i \tag{9.2}$$

ただし，この式を用いると，逆推定の 95%信頼区間の計算が煩雑になるので，両辺を $\hat{v}_1$ で割った形にするとよい．

$$\frac{\hat{e}_i^2}{\hat{v}_1} = \frac{\hat{v}_0}{\hat{v}_1} + x_i, \tag{9.3}$$

この式は，$\hat{v}_0/\hat{v}_1 = \hat{v}_0'$ と置き換えて

$$\frac{\hat{e}_i^2}{\hat{v}_1} = \hat{v}_0' + x_i, \tag{9.4}$$

とすると逆推定の 95%信頼区間の推定に際して都合がよい．

同じ x_i について 1 回の測定値しかない場合は，8.3 節 で示したように最尤法を用いることにより，分散の推定を x の関数として行なうことになる．
（高橋の追加終り）

9.5. ［付録 5］林ら(2000)

［5］Taylor J.R. (1997), 林茂雄, 馬場涼 訳(2000), 計測における誤差解析入門, 東京化学同人.

計測関連の教科書にも重み付き回帰分析が取り上げられている. 数値例がないので簡単に内容を示す.

異なる重み（原著の要約）

第 8 章　最小二乗法によるあてはめ　まとめ

直線 $y = A + Bx$ ： **異なる重み（半ページ）**

y が $y = A + Bx$ の上に載ると予想でき， y_i に異なる誤差 σ_i が含まれていれば，重み $w_i = 1/\sigma_i^2$ を導入することによって A， B の最良推定値が次のように求められる.

Σ を用いた計算式が示されている（中略）.

問題 8.9 重み付き最小二乗法

A， B の最良推定値の計算式の導出することを求めている.

また，設問の最後に「明らかに，この重み付き最小二乗法は σ_i （またはその相対値）がわかってない限り適用できない.」と注意をしている.

問題 8.19 重み付き最小二乗法の誤差

定数 A， B の誤差について誤差伝播から求められる式の導出を求めている.

問題 8.26 放射性物質の壊変

指数関数的な関数のあてはめの場合は，反応 ν_i の誤差が $\sqrt{\nu_i}$ であり，重みとして ν_i を使うこととしている.

9.6. ［付録 6］竹内啓(1963)

［6］竹内啓(1963)数理統計学, 312-315, 東洋経済新報社.

第 29 章 回帰分析(I) の IV 節 回帰直線自体についての推論 に逆推定について示唆に富む記述がある.

回帰直線自体についての推論（原著の引用）

IV　回帰直線自体についての推論

このほか推論は回帰直線自体についてもいろいろな形で行なうことができる．まず $x = x_0$ のときの回帰式の値の推定量は

$$\hat{\alpha} + \hat{\beta} x_0 \qquad [29.34]$$

となる．この推定量の分散は

$$\begin{aligned} V(\hat{\alpha} + \hat{\beta} x_0) &= V(\hat{\alpha}) + 2x_0 Cov(\hat{\alpha}, \hat{\beta}) + x_0^2 V(\hat{\beta}) \\ &= \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2} \right) \sigma^2 \end{aligned} \qquad [29.35]$$

すなわちこの値の分散は $x_0 = \bar{x}$ のとき最小となり，$\bar{x}$ から x_0 の値が離れるにつれて大きくなる.

次に $x = x_0$ のときの回帰式の値がある値に等しい，すなわち

仮説　　$\alpha + \beta x_0 = c$

を検定する.

この仮説の検定は次のようにすればよいことは容易にわかる.すなわち

$$\frac{|\hat{\alpha} + \hat{\beta} x_0 - c|}{\sqrt{V(\hat{\alpha} + \hat{\beta} x_0)}} = \frac{|\hat{\alpha} + \hat{\beta} x_0 - c|}{\sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2} \right) \sigma^2}} > t_0 \qquad [29.36]$$

のとき仮説をすてる.

またこの式を用いると，$x = x_0$ のときの回帰式 $y_0 = \alpha + \beta x_0$ の信頼区間を求めることができる．すなわち

$$\Pr \left\{ \frac{|y_0 - (\hat{\alpha} + \hat{\beta} x_0)|}{\sqrt{V(\hat{\alpha} + \hat{\beta} x_0)}} < t_\alpha \right\} = 1 - \alpha \qquad [29.37]$$

だから

$$\hat{\alpha} + \hat{\beta} x_0 - t_\alpha \sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2} \right) \sigma^2} < y_0 < \hat{\alpha} + \hat{\beta} x_0 + t_\alpha \sqrt{\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2} \right) \sigma^2} \qquad [29.38]$$

とすれば，これは y_0 に対する信頼係数 $1 - \alpha$ の信頼区間になる.

また同じ関係式から，$y_0 = c$ となるような x_0 の値に対する信頼域を作ることもできる．すなわち

$$\Pr\left\{\frac{|c-(\hat{\alpha}+\hat{\beta}x_0)|}{\sqrt{V(\hat{\alpha}+\hat{\beta}x_0)}} < t_\alpha\right\} = 1-\alpha \qquad [29.39]$$

の｛　｝内の式を x_0 について解けば，x_0 の信頼区間が得られる．｛　｝内の式を変形すると

$$(\hat{\alpha}+\hat{\beta}x_0-c)^2 < t_\alpha^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum(x-\bar{x})^2}\right)\hat{\sigma}^2 \qquad [29.40]$$

これは一般に x_0 に関する 2 次不等式となり，その解は

$$A < x_0 < B$$

という形になる場合と

$$x_0 < C \quad \text{または} \quad x_0 > D$$

となる場合とがある．前者の場合には普通の形の信頼区間が得られるが，後者の場合は普通の場合と逆に，信頼域はある区間の外ということになる．

（高橋のコメント）

回帰直線における逆推定 $y_0 = c$ となる場合の x の推定値は，

$$\hat{x}_0 = \frac{c-\hat{\alpha}}{\hat{\beta}} \quad \text{あるいは} \quad \hat{x}_0 = \frac{y_0-\hat{\alpha}}{\hat{\beta}}$$

となることの説明は容易であるが，逆推定された $\hat{x}_0$ の 95%信頼区間をどのように求められているかを説明することは，容易ではない．

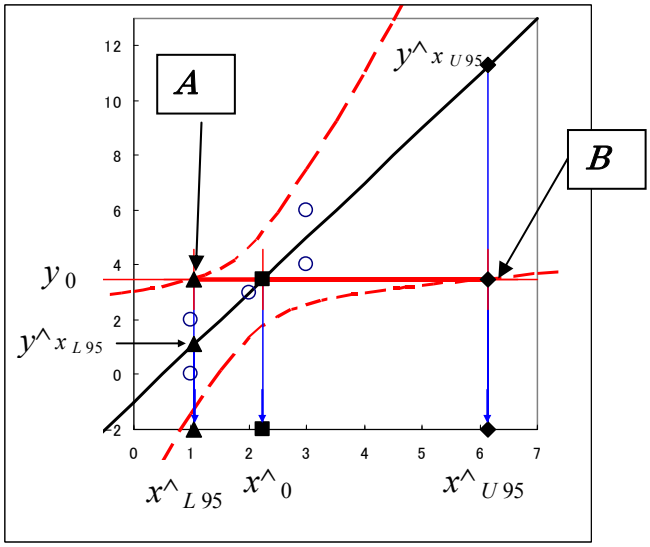


図 2.5 逆推定の模式図（再掲）

第 2 章では図 2.5 を用いて，逆推定された $\hat{x}_0$ の 95%信頼区間の下限 $\hat{x}_{L95}$ は，x について

の上側の 95%信頼区間の曲線が $y_0 = c$ を横切るときの x を求める問題とした．この解を求めるために，$\hat{x}_{L95}$ の 回帰の推定値 $\hat{y}_{x_{L95}} = \hat{\alpha} + \hat{\beta}\hat{x}_{L95}$ に 95%信頼区間幅 $t_\alpha \sqrt{Var(\hat{\alpha}_0 + \hat{\beta}\hat{x}_{L95})}$ を加えた値が，ちょうど y_0 となるような $\hat{x}_{L95}$ を求める問題とした．それは，次式，

$$(\hat{\alpha}_0 + \hat{\beta}\hat{x}_{L95}) + t_\alpha \sqrt{Var(\hat{\alpha}_0 + \hat{\beta}\hat{x}_{L95})} = y_0$$

を $\hat{x}_{L95}$ について解くと，$\hat{x}_{L95}$ に関して 2 次式となるので，2 次式の解を求めることになる．
（高橋のコメント終り）

9.7. ［付録 7］Miller ら(2000)，宗森ら訳(2004)

［7］Miller J.N., Miller J.C.(2000)，Statistics and Chemometrics for Analytical Chemistry 4thed. 宗森信，佐藤寿邦 訳(2004)，データの取りまとめ方 －化学分析のための統計学とケモメトリックスー，142-173，共立出版.

分析化学者のための統計の教科書として，1991 年に初版が翻訳され出版され 2002 年に 8 刷り，2004 年に第 2 版，2009 年に 5 刷りとなった．分析化学者のための代表的な教科書であり，個別データの逆推定値 $\hat{x}_0$ の $SE(\hat{x}_0)$ を近似式で求め，t 分布の両側 5%点を掛けて算出して 95%信頼区間を求めている．近似式はあたえられたとしているので，その導出の考え方を追加し，さらに，重み付き回帰の場合の場合についても計算式を示した．

重み付き回帰は，x_i における繰返し測定から得られた標準偏差 $\hat{\sigma}_i$ の分散の逆数を重みとしている．それに代わる方法として，分散 $\hat{\sigma}_i^2$ について x^2 の 1 次式をあてはめ，分散の推定値 $\hat{\sigma}_i'^2$ の逆数を重みとした解析を追加し，個別データの 95%信頼区間をを図示した．

直線があてはまらない事例に対して，2 次式および 3 次式をあてはめ，2 次式が妥当であるとしている．逆推定の点推定値は与えられているが，95%信頼区間は算出されていない．そこで，x^λ の 1 次式について，λ を変化させたときに誤差分散 $\hat{\sigma}_e^2$ が最も小さくなる λ の推定法を示し，その場合の個別データの逆推定の 95%信頼区間を算出した結果を示した．

濃度の計算とその偶然誤差（原著の引用）

5.6 濃度の計算とその偶然誤差

回帰線の傾斜と切片が決まれば，測定値 y に対応する濃度（x 値）を計算するのはきわめて簡単である。しかし，こうして回帰線から求めた濃度の誤差を推定する必要がある。式(5.1)を利用して与えられた y 値から x 値を計算するには，傾斜（b）と切片（a）の両方が用いられるが，前節でみたように，これらの値はともに誤差を伴う。さらに，どのような試験試料からの機器信号も偶然誤差を伴う。その結果，x 値の誤差の計算はきわめて複雑になる。たいていの研究者はつぎの近似式を使っている。

$$s_{x0} = \frac{s_{y/x}}{b}\left\{1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2}\right\}^{1/2} \qquad \text{原著（5.9）}$$

この式で，y_0 は y の実験値であって，これから濃度値 x_0 が求められる。s_{x0} は x_0 の標準偏差の推定値である。その他の記号はいままでに使ってきたものと同じである。場合によっては y_0 をえるのに，数回測定を行うことがある。m 回測定を行ったとすると，s_{x0} の式はつぎのようになる。

$$s_{x0} = \frac{s_{y/x}}{b}\left\{\frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2}\right\}^{1/2} \qquad \text{原著（5.9）}$$

m = 1 であれば，式（5.10）は式（5.9）になる。信頼限界は $x_0 \pm t_{(n-2)} s_{x0}$ で計算でき，自由度は（$n-2$）である。簡単なコンピュータープログラムはこれらの計算をすべて行えるが，たいていの，計算器は適当でないであろう。

【例題 5.6.1】
5.3 節の例題のデータを用い，蛍光強度 2.9，13.5，23.0 単位を示す溶液の x_0 と s_{x0} の値，および x_0 の信頼限界を求めよ。【例題 5.3.1】フルオレセインの標準水溶液を蛍光分光計で調べたところ，蛍光強度（任意単位）はつぎのようであった。

濃度， pg ml^{-1}	0	2	4	6	8	10	12
蛍光強度	2.1	5.0	9.0	12.6	17.3	21.0	24.7

解答

5.4 節で求めた回帰線 $y = 1.93x + 1.52$ に， y_0 の値 2.9，13.5，23.0 を代入すると， x_0 の値としてそれぞれ 0.72，6.21，11.13 pg ml^{-1} が容易にえられる。

これらの x_0 値に対応する s_{x0} 値をえるためには式（5.9）を用いるが，前節までの計算で， n = 7，b = 1.93， $s_{y/x}$ = 0.4329， $\bar{y}$ = 13.1， $\sum_i (x_i - \bar{x})^2$ = 112 であることがわかっている。これらの値を用いると y_0 値 2.9，13.5，23.0 に対して， s_{x0} 値として 0.26，0.29，0.26 がそれぞれえられる。95%信頼限界（$t_5 = 2.57$）はそれぞれ 0.72±0.67，6.21±0.62，11.13±0.67 pg ml^{-1} になる。

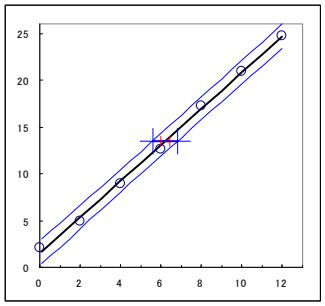
（高橋の追加 1）

この事例は，第 2 章で示した逆推定の基本であり，個別データの 95%信頼区間が示されている．

表 9.5 Miller のデータによる回帰パラメータの推定

k	$\boldsymbol{x}$	$\boldsymbol{\beta}_k$	$\boldsymbol{\beta}_k$^	$Var(\beta_k\hat{})$	SE	t	p
0	x_0:切片	β_0	1.518	0.087	0.295	5.146	0.0036
1	x_1:傾き	β_1	1.930	0.002	0.041	47.197	0.0000

図 9.12 に示すように, Excel による計算結果は，$\bar{y}$ 近辺の $y_0 = 13.5$ に対して，$\hat{x}_0 = 6.207$，個別データの 95% 信頼区間は（5.591，6.825）であり，原著の 6.21±0.62 =（5.59，6.83）と同様の結果となっている．しかし，$\bar{y}$ から離れた $y_0 = 2.9$ に対して，$\hat{x}_0 = 0.716$，個別データの 95% 信頼区間は（0.019，1.328）であり，原著の 0.72±0.67 =（0.05，1.39）と少数点 2 桁目から違いが生じている．



逆推定 y_0 =	2.9	x_0^ =	0.716
回帰直線の95%CL			
a =	−0.202	x_{L95R}^ =	0.338
b =	0.787	x_{U95R}^ =	1.062
c =	−0.562		
個々のデータの95%CL			x
a' =	−0.015	x_{L95I}^ =	0.019
b' =	0.787	x_{U95I}^ =	1.382
c' =	−0.562		

逆推定 y_0 =	13.5	x_0^ =	6.207
回帰直線の95%CL			
a =	−21.640	x_{L95R}^ =	5.989
b =	6.981	x_{U95R}^ =	6.426
c =	−0.562		
個々のデータの95%CL			x
a' =	−21.453	x_{L95I}^ =	5.591
b' =	6.981	x_{U95I}^ =	6.825
c' =	−0.562		

図 9.22　Miller のデータを用いた逆推定

Miller の式（5.9）は，切片 $\hat{\beta}_0$ の分散，傾き $\hat{\beta}_1$ の分散，$\hat{\beta}_0$ と $\hat{\beta}_1$ の共分散を用いて，$\hat{x}_0$ が次式であらわされる．

$$\hat{x}_0 = g(\hat{\beta}_0, \hat{\beta}_1) = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1}$$

この $\hat{x}_0$ の分散は，合成分散の一般的な式（9.5）から得られる．ただし，この合成分散の式は回帰直線の場合であって，個別データの分散ではない．

$$Var(\hat{x}_0) = \left(\frac{\partial g}{\partial \hat{\beta}_0}\right)^2 Var(\hat{\beta}_0) + 2\left(\frac{\partial g}{\partial \hat{\beta}_0}\frac{\partial g}{\partial \hat{\beta}_1}\right) Cov(\hat{\beta}_0, \hat{\beta}_1) + \left(\frac{\partial g}{\partial \hat{\beta}_1}\right)^2 Var(\hat{\beta}_1) \qquad (9.5)$$

個別データの分散は，式（2.19）で示したように，

$$Var(\hat{y}_i) + \hat{\sigma}_e^2 = (Var(\hat{\beta}_0) + \hat{\sigma}_e^2) + 2x_i Cov(\hat{\beta}_0, \hat{\beta}_1) + x_i^2 Var(\hat{\beta}_1) \qquad (9.6)$$

である．この式から，個別データの逆推定のための分散は，

$$Var(\hat{x}_0) = \left(\frac{\partial g}{\partial \hat{\beta}_0}\right)^2 (Var(\hat{\beta}_0) + \sigma_e^2) + 2\left(\frac{\partial g}{\partial \hat{\beta}_0}\frac{\partial g}{\partial \hat{\beta}_1}\right) Cov(\hat{\beta}_0, \hat{\beta}_1) + \left(\frac{\partial g}{\partial \hat{\beta}_1}\right)^2 Var(\hat{\beta}_1) \qquad (9.7)$$

となる．

ただし，$\dfrac{\partial g}{\partial \hat{\beta}_0} = \left(\dfrac{-1}{\hat{\beta}_1}\right)$，　$\dfrac{\partial g}{\partial \hat{\beta}_1} = \left(\dfrac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2}\right)$ ．

この式を用いて，近似的な誤差分散を計算し，Miller と同様の ±95% 信頼幅が得られる．なお，m 個の平均値の場合は $(Var(\hat{\beta}_0) + \sigma_e^2 / m)$ とすればよい．

表 9.6　個別データの近似的な 95%信頼区間

f_e =	5	
$t_{0.05}$ =	2.571	
Σ =	0.0870	−0.0100
	−0.0100	0.0017
σ_e^2^ =	0.1874	

y_0	$\frac{\partial g}{\partial \hat{\beta}_0} = \left(\frac{-1}{\hat{\beta}_1}\right)$	$\frac{\partial g}{\partial \hat{\beta}_1} = \left(\frac{-(y_0 - \hat{\beta}_0)}{\hat{\beta}_1^2}\right)$	x_0^	$Var(x_0$^$)$	s_{x0}^	±95%
2.9	−0.518	−0.371	0.72	0.070	0.265	0.680
13.5	−0.518	−3.216	6.21	0.057	0.240	0.616
23.0	−0.518	−5.765	11.13	0.069	0.263	0.677

原著　$\hat{x}_0$ ±95%　：0.72±0.67，6.21±0.62，11.13±0.67 pg ml^{-1}

（高橋の追加終）

重み付き回帰線（原著の引用）

5.10　重み付き回帰線

本節では，重み付き回帰法の利用の要点を述べる。ただし，この方法は 2 方法の比較ではなく，一成分の定量分析に用いられるものとする。どのような校正実験においても，結果の全偶然誤差は分析のいろいろな段階の誤差が組み合わされたものである（原著 2.11 節参照）。ある場合は，この全誤差は，一つあるいはそれ以上の段階が支配的な，濃度依存性でない偶然誤差かもしれない。そのような場合，校正線における y 方向の誤差はすべての点で近似的に等しい（homoscedasticity）と期待してもよく，重みなし回帰線は妥当である。他の場合，誤差は定量対象種濃度に近似的に比例する（相対誤差はほぼ一定），あるいは，y 方向の誤差は x が増えれば，比例でなく増加する。このような異精度データは重み付き回帰法によって扱うのが正しい。通常，分析者は重み付きと重みなしのどちらが適切かを経験から学ぶ。予測は難しい。二つの見かけ上似ている方法が非常に異なる誤差の振る舞いを示すような例は多数ある。重み付き回帰分析計算は，重みなしの計算よりかなり複雑で，また，多くの情報（または，より多くの仮定）を必要とする。それにもかかわらず，異精度の疑いがあるときはいつも重み付き計算をすべきである。（中略）

誤差棒のより小さい点に，より大きい重みがかかるように回帰線を計算しなければならないことは明らかである。計算線が誤差のより小さい点の近くを通る方が，大きな誤差を示す高い濃度の点の近くを通るよりも重要だからである。このようにするには，各点にそれぞれの分散 s^2 に逆比例する重みをかければよい（この論理的方法は y 方向の誤差が x に比例する場合だけでなく，すべての重み付き回帰計算に適用できる）。個々の点を（x_1，y_1），（x_2，y_2），. . . ，で表し，対応する標準偏差を s_1，s_2，. . . ，で表すと，それぞれの点の重み w_1，w_2，. . . ，は次式で与えられる[*8]。［訳注[*8]］ 式（5,13）のように重みを相対化する必然性はない。むしろ $w_i = s_i^{-2}$ とする方が単純でよい［訳注終］。

重み：　$w_i = s_i^{-2} / (\sum_i s_i^{-2} / n)$　　原著（5.13）

この式から，重みはその合計が線上の点の数に等しくなるように付けられることがわかる。こうすることによって，後の計算が簡単になる。回帰線の傾斜と切片はつぎの式で与えられる。

重み付き傾斜：　$b = \dfrac{\sum_i w_i x_i y_i - n\bar{x}_w \bar{y}_w}{\sum_i w_i x_i^2 - n\bar{x}_w^2}$　　原著（5.14）

重み付き切片：　$a = \bar{y}_w - b\bar{x}_w$　　原著（5.15）

式（5.15）で，$\bar{x}_w$ と $\bar{y}_w$ は重み付き重心の座標（$\bar{x}_w$，$\bar{y}_w$）を表し，重み付き回帰線はこの点を必ず通る。重心の座標は，$\bar{x}_w = \sum_i w_i x_i / n$，$\bar{y}_w = \sum_i w_i y_i / n$ で与えられる。

【例題 5.10.1】
つぎの校正実験データに対して，重み付けない回帰線と重み付き回帰線を計算せよ。各回帰線から，吸光度の 0.100 と 0.600 を示す試料の濃度を計算せよ。

濃度，μg ml^{-1}	0	2	4	6	8	10
吸光度	0.009	0.158	0.301	0.472	0.577	0.739
標準偏差	0.001	0.004	0.010	0.013	0.017	0.022

解答
式（5.4）と式（5.5）を適用すると，重み付けない回帰線の傾斜と切片はそれぞれ 0.0725 と 0.0133 になる。吸光度の 0.100 と 0.600 に対応する濃度はそれぞれ 1.20 と 8.09 μg ml^{-1} である。重み付き回帰線を計算するのは少し厄介である。適当な計算機プログラムがなければ，つぎのように表にして計算する。

（計算表は略）

これらの数値から，$\bar{x}_w = 1.372/6 = 0.2296$，$\bar{y}_w = 0.1561/6 = 0.0260$ がえられ，式（5.14）と式（5.15）から，b と a はそれぞれつぎのように計算される。

重み付き傾斜：　$b_w = \dfrac{0.4417 - (6 \times 0.2296 \times 0.0260)}{5.818 - 6 \times 0.2296^2} = 0.0738$

重み付き切片：　$a_w = 0.0260 - 0.0738 \times 0.2296 = 0.0091$

これら a と b の値を使って，吸光度の 0.100 と 0.600 に対応する濃度を求めると，1.23 と 8.01 μg ml^{-1} がそれぞれえられる。（解答終）

（中略）

重み付き回帰計算では，試験試料の濃度の標準偏差（$s_{x_{0w}}$）は次式で与えられる*[9]。[訳注 *9] 式（5.16）は重み付き最小二乗法の関係式と誤差伝搬の法測から導かれる式とやや異なるが，実用上，えられる数値はあまり変らない。

$$s_{x_{0w}} = \frac{s_{(y/x)w}}{b}\left\{\frac{1}{w_0} + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{b^2\left(\sum_i w_i x_i^2 - n\bar{x}_w^2\right)}\right\}^{1/2} \qquad \text{原著（5.16）}$$

この式において，$s_{(y/x)w}$ は次式で与えられる。

$$s_{(y/x)w} = \left\{\frac{\sum_i w_i (y_i - \hat{y}_i)^2}{n-2}\right\}^{1/2} \qquad \text{原著（5.17）}$$

w_0 は y_0 の値に対する重みである。式（5.16）は明らかに式（5.9）に形が似ている。この式から，点が原点に近いほど w_0 が大きく，また点が重心に近いほど $y_0 - \bar{y}_w$ が小さくなって，信頼限界は小さくなることがわかる（図 5.13：図は省略）。二つの式で最も大きい違いは，式（5.16）には $1/w_0$ 項があることである。w_0 は y が大きくなると急激に小さくなるので，この項が，予想されるように，y_0 の増加に伴って信頼限界を大きくする。

式（5.16）を上の例題のデータに適用すると，吸光度が 0.100 と 0.600 の試料に対して，計算から求められた濃度（1.23 と 8.01 μg ml^{-1}）の信頼区間はそれぞれ±0.12 と±0.72 μg ml^{-1} になる。これらの信頼区間は二つの溶液の吸光度にほぼ比例していることがわかる。さらに，二つの溶液のうちで，薄い方に対する信頼区間は重み付けない回帰計算値より小さく，一方，濃い方の試料に対しては逆になっている。

（高橋の追加 2）

重みとして x_i における分散の逆数 $w_i = s_i^{-2}$ をつかうことは，正論である．しかしながら，重みとして分散の平均値で相対化することが，後の計算が簡単となるとしている．訳者は $w_i = s_i^{-2}$ が単純でよいとしている．これは，読者が困惑する原因となる．

図 9.13 で，分散の逆数の推定値として重みを $1/(0.02689x^2)$ としても，比例関係にある $1/x^2$ としても，回帰パラメータの推定値および誤差は全く同じになることを示した．この関係は，元の分散を使わなくとも分散が比例関係にあれば結果が一致することが示唆されている．Miller は「重みとして分散の平均値で相対化する」としているが，実は元の分散に比例さえすれば，計算がしやすいような数値を選択することが可能であり，「分散の平均値で相対化」した数値を使う必然性はない．

表 9.7　各種の重み

	xi	yi	si	1/si^2	wi	wi'
1	0	0.009	0.001	1000000	5.535	100.000
2	2	0.158	0.004	62500	0.346	6.250
3	4	0.301	0.010	10000	0.055	1.000
4	6	0.472	0.013	5917	0.033	0.592
5	8	0.577	0.017	3460	0.019	0.346
6	10	0.734	0.022	2066	0.011	0.207

重みとして $w_i' = s_i^{-2}/10000$ をも用いて回帰分析をしたパラメータ推定の結果は，表 9.8 に示すように，$a_w = 0.0091$，$b_w = 0.0738$ と同じ回帰パラメータが推定される．

表 9.8　分散に比例する任意の重み w_i' を用いた回帰パラメータの推定

パラメータ推定値				
項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	0.009098	0.001071	8.497128	0.001052
xi	0.073658	0.001087	67.745196	0.000000

測定されたデータから分散を推定し，その逆数 $w_i = s_i^{-2}$ を重みにするとの手順は，ある程度の繰返し測定がある場合に適用が限定され，一般的な回帰分析への応用を阻害することになる．さらに，個別データの 95%信頼区間が関数として定義できないために，適切な逆推定が行なえない．

重み付き回帰の原則は，重みとして分散の逆数を用いるのであるが，元の分散に比例しさえすれ

ば，回帰パラメータの推定には全く影響しない．このことを活用しているのが，分散が x^2 に比例するならば，重みとして簡単化した $w_i = 1/x_i^2$ を用いるとの対応である．簡単化した重みのよさは，推定された誤差分散が，$x_i = 1$ の場合の分散となり理解しやすいとの利点もある．

実際に，この事例の標準偏差から分散の推定値を，分散が単調に増加する x^2 に関して 1 次式

$$\hat{\sigma}_i^2 = \hat{v}_0 + \hat{v}_1 x_i^2 = 0.00000365 \ + \ 0.00000471\, x_i^2 \qquad (9.8)$$

が得られる．．

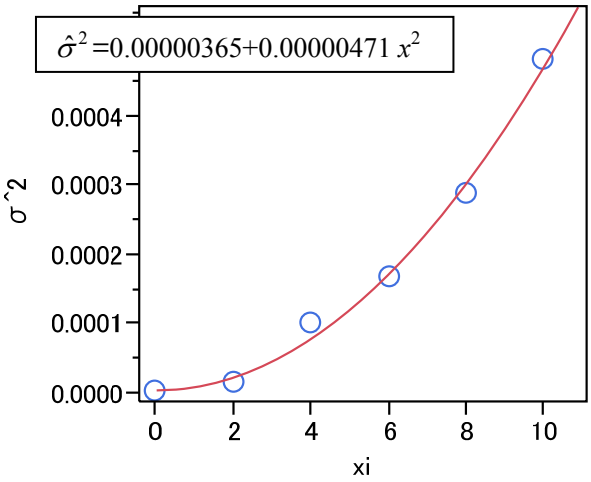


図 9.23　x^2 に関して 1 次式による分散の推定

式(9.8)を重みとした回帰パラメータを推定すると，表 9.9 に示す結果を得る

表 9.9　分散の推定式を用いた重み付き回帰

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	4676.1789	4676.18	4149.601
誤差	4	4.5076	1.13	p値(Prob>F)
全体(修正済み)	5	4680.6865		<.0001*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	0.0092767	0.00196	4.73	0.0091*
xi	0.0735419	0.001142	64.42	<.0001*

この両辺を $\hat{v}_1 = 0.00000471$ で割り，

$$\frac{\hat{\sigma}_i^2}{\hat{v}_1} = \frac{\hat{v}_0}{\hat{v}_1} + x_i^2 = \frac{0.00000365}{0.00000471} + x_i^2 = 0.775 + x_i^2 \qquad (9.9)$$

分散に比例する形に，簡単化する．これを重みとした回帰分析の結果を表 9.10 に示す．分散分析表の平方和は異なるが，F 比は同じで，パラメータの推定結果も，0.0092768，0.0735419 と表 9.9 と同じ結果となる．

表 9.10　分散に比例する推定式を用いた重み付き回帰

分散分析

要因	自由度	平方和	平均平方	F値
モデル	1	0.02204276	0.022043	4149.465
誤差	4	0.00002125	5.312e-6	p値(Prob>F)
全体(修正済み)	5	0.02206400		<.0001*

パラメータ推定値

項	推定値	標準誤差	t値	p値(Prob>|t|)
切片	0.0092768	0.00196	4.73	0.0091*
xi	0.0735419	0.001142	64.42	<.0001*

Excel による解析

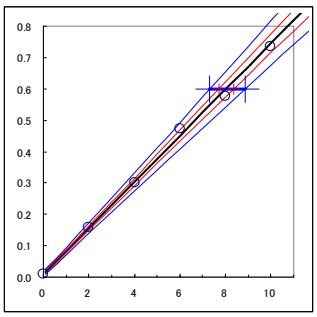
Miller の事例は，6.3 節で示した分散がおおよそ x^2 に比例する場合と同じである．重みとして

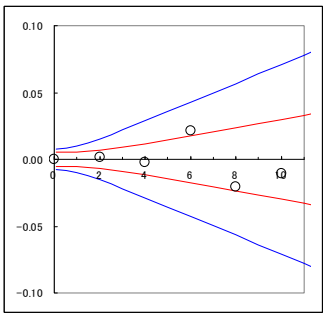
$$w_i = 1/(0.775 + x_i^2)$$

を与え，表 9.11 の結果を得る．これは，JMP による表 9.10 と同じ回帰パラメータが推定されている．

表 9.11　Excel による計算結果

			X						
i	x_i	y_i	切片	x_i^{λ}	β^		y^	ε^	w
1	0	0.009	1	0.0	0.00928	=	0.009	0.000	1.2903
2	2	0.158	1	2.0	0.07354		0.156	0.002	0.2094
3	4	0.301	1	4.0			0.303	−0.002	0.0596
4	6	0.472	1	6.0			0.451	0.021	0.0272
5	8	0.577	1	8.0			0.598	−0.021	0.0154
6	10	0.734	1	10.0			0.745	−0.011	0.0099





逆推定 y_0 =	0.6	x_0^ =	8.032
回帰直線の95%CL			
a =	−0.045	x_{L95R}^ =	7.720
b =	0.011	x_{U95R}^ =	8.372
c =	−0.001		
個々のデータの95%CL			x
a' =	−0.045	x_{L95I}^ =	7.326
b' =	0.011	x_{U95I}^ =	8.890
c' =	−0.001		

図 9.24　個別データの逆推定の 95%信頼区間

Miller の式（5.16）は，切片 $\hat{\beta}_0$ の分散，傾き $\hat{\beta}_1$ の分散，$\hat{\beta}_0$ と $\hat{\beta}_1$ の共分散を用いて，$\hat{x}_0$ が次式であらわされる場合に，

$$\hat{x}_0 = g(\hat{\beta}_0, \hat{\beta}_1) = \frac{y_0 - \hat{\beta}_0}{\hat{\beta}_1}$$

この合成分散の一般的な式（9.5）から得られる．ただし，回帰の場合であって，個別データの分散ではない．個別データの分散は，表 6.6 の各種の重みに対する逆推定のための 2 次式のパラメータから

$$Var(\hat{y}_i) + (0.775 + x_i^2)\hat{\sigma}_e^2 = (Var(\hat{\beta}_0) + 0.775\hat{\sigma}_e^2) + 2x_i Cov(\hat{\beta}_0, \hat{\beta}_1) + x_i^2 (Var(\hat{\beta}_1) + \hat{\sigma}_e^2) \quad (9.10)$$

となり，この式を用いて，次の合成分散の式が得られる．

$$Var(\hat{x}_0)=\left(\frac{\partial g}{\partial \hat{\beta}_0}\right)^2 Var(\hat{\beta}_0+0.775\sigma_e^2)+2\left(\frac{\partial g}{\partial \hat{\beta}_0}\frac{\partial g}{\partial \hat{\beta}_1}\right)Cov(\hat{\beta}_0,\hat{\beta}_1)+\left(\frac{\partial g}{\partial \hat{\beta}_1}\right)^2(Var(\hat{\beta}_1)+\sigma_e^2) \quad (9.11)$$

ただし，$\dfrac{\partial g}{\partial \hat{\beta}_0}=\left(\dfrac{-1}{\hat{\beta}_1}\right)$，　$\dfrac{\partial g}{\partial \hat{\beta}_1}=\left(\dfrac{-(y_0-\hat{\beta}_0)}{\hat{\beta}_1^2}\right)$ ．

Excel で，この式を用いて，近似的な誤差分散を計算し，Miller と同様の ±95% 信頼幅が得られる．Excel シートで計算された分散共分散行列 $\boldsymbol{\Sigma}$ から，

f_e= 4

$t_{0.05}$= 2.776

$$\boldsymbol{\Sigma}=\begin{bmatrix} 0.000003841 & -0.000000844 \\ -0.000000844 & 0.000001303 \end{bmatrix}$$

$\sigma_e{}^2$^ = 0.000005312

$Var(\hat{\beta}_0)=0.000003841$，$Cov(\hat{\beta}_0,\hat{\beta}_1)=-0.000000844$，$Var(\hat{\beta}_1)=0.000001303$，$\sigma_e^2=0.000005312$ が得られるので，これらを式（9.11）に代入して，表 9.12 の結果を得る．

表 9.12　逆推定の近似的な 95%信頼区間

y_0	$\frac{\partial g}{\partial \hat{\beta}_0}=\left(\frac{-1}{\hat{\beta}_1}\right)$	$\frac{\partial g}{\partial \hat{\beta}_1}=\left(\frac{-(y_0-\hat{\beta}_0)}{\hat{\beta}_1^2}\right)$	x_0^	$Var(x_0$^$)$	s_{x0}^	±95%
0.1	−13.598	−16.774	1.23	0.003	0.054	0.151
0.6	−13.598	−109.223	8.03	0.078	0.279	0.775

原著：y_0=(0.100, 0.600)，$\hat{x}_0$=(1.23, 8.01)，±95%=(±0.12, ±0.72)

（高橋の追加 2 終）

曲線のあてはめ（原著の引用）

5.14 曲線のあてはめ

データ変換がもたらす困難さと，校正点に曲線をあてはめる計算が容易になってきたことから，曲線回帰法が次第に利用されるようになってきた。実際に出会う曲線校正プロットは二つ，またはそれ以上の物理・化学現象が重なった結果である場合が多い，ということは重要である。たとえば，分子蛍光測定では，(a) 光学的アーチファクト（内部フィルター効果），(b) 分子間相互作用（クエンチング，エキシマー生成など），(c) 線形プロットのもとになる代数的仮定が成り立たない，ために濃度対信号のプロットは希薄溶液ではほぼ直線的であるが，高濃度で（負の）曲がりが増えてくる。(a)～(c) の効果は互いに独立であるから，実際には種々の形をした曲線が現われる。分析で，一定の予測できる形の曲線に出会うことが稀なのは，このような理由による。したがって，ある特定の問題において，校正データに当てはめるべき関数を多くの種類の曲線関数から選択するための指針はほとんどない。実際には，最も一般的な方法は x の多項式，$y=a+bx+cx^2+dx^3+\cdots$ を当てはめることである。

（中略）

例題 5．14.1

ある機器分析法によって，つぎのデータがえられた（単位は任意）。

濃度	0	1	2	3	4	5	6	7	8	9	10
信号	0.2	3.6	7.5	11.5	15.0	17.0	20.4	22.7	25.9	~~29.6~~ 27.6 が正しい	30.2

この結果に適当な多項式をあてはめ，信号が 5，16，27 の場合の濃度を求めよ。

解答の要約

1 次式のあてはめ：$y = 2.991x + 1.555$

2 次式のあてはめ：$y = 0.086 + 3.970x - 0.098x^2$

3 次式のあてはめ：$y = -0.400 + 4.170x - 0.150x^2 + 0.035x^3$

y	1 次	2 次	3 次
5	1.15	1.28	1.27
16	4.83	4.51	4.50
27	8.51	8.61	8.62

（高橋の追加 3）

1 次式をあてはめた場合の残差は上に凸であり，あてはまりは悪く，個別データ 95%信頼区間の幅も大きい．2 次式のあてはめは，系統的な誤差も消え，別データ 95%信頼区間の幅も狭くなっている．3 次式のあてはめは，表 9.13 に示すように逐次平方和の増加分が少なく F 検定でも有意ではないので，2 次式のあてはめが適当である．

表 9.13　多項式回帰に対する逐次検定

逐次(タイプ1)検定

要因	パラメータ数	自由度	逐次平方和	F値	p値(Prob>F)
x	1	1	984.0091	5736.710	<.0001*
x*x	1	1	8.2238	47.9441	0.0002*
x*x*x	1	1	0.0755	0.4403	0.5282

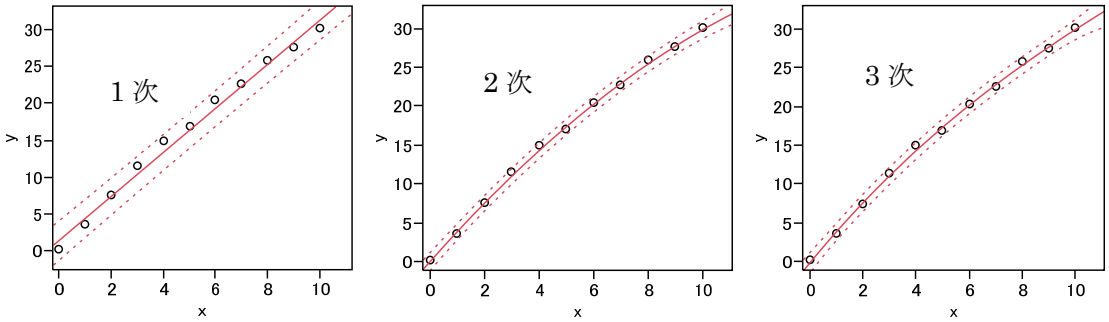


図 9.25 曲線のあてはめと個別データの 95%信頼区間

直線のあてはめが不調である場合に，多項式回帰が一般的に行なわれている．しかしながら，4.1 節で示したように，逆推定を行なう場合に 2 次式の根の公式で解くことは容易ではあるが，個別データの 95%信頼区間は 4 次式となり，計算は難儀となる．そのために，4.2 節で示したように，x^2 に関して 1 次式あるいは x^γ に関して 1 次式のあてはめが，逆推定の問題を簡単化することを示した．

この例では，曲線が上に凸なので，$0<\gamma<1$ となるような x^γ に関する 1 次式を選択すことになる．多項式のあてはめは，統計的には容易な方法であるが，x^γ に関する 1 次式といっても γ を定める必要があり，8.2 節で示した最尤法を用いて，様々な未知パラメータの推定法を示したが，ここでは，未知パラメータ γ を逐次変化させ，誤差分散が最小となるような γ を推定し，個別データの逆推定，および 95%信頼区間を求めてみよう．

図 9.26 に示すように冪 λ を変化させた場合の誤差分散が最小になるのは，$\lambda=0.82$ であることがわかる．正確には $\hat{\lambda}=0.8173$，$\hat{\sigma}_e^2=0.2677$ である．

λ	β_0^	β_1^	σ_e^2^
1.00	1.555	2.991	1.0556
0.90	0.609	3.824	0.4507
0.83	−0.122	4.545	0.2725
0.82	−0.268	4.658	**0.2679**
0.80	−0.457	4.894	0.2769
0.70	−1.265	6.265	0.7584

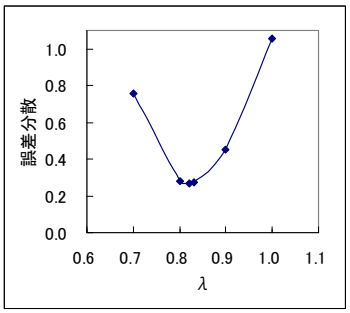
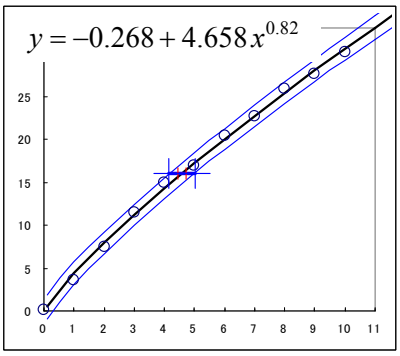


図 9.26　冪 λ を変化させた場合の誤差分散

冪 $\hat{\lambda}=0.82$ を用いた 1 次式は，図 9.27 に示すように，

$$y=-0.268+4.658x^{0.82}$$

となり，y_0 についての個別データの逆推定値，95%信頼区間も正確に求められる．



y_0	x_0^	$L95I$	$U95I$
5	1.152	0.81	1.505
16	5.583	4.165	5.008
27	8.614	8.126	9.118

原著 x_0^
1.28
4.51
8.61

図 9.27　冪 λ=0.82 の場合の逆推定と個別データの 95%信頼区間

（高橋の追加 3 終）

9.　文献に対する批判的吟味　Miller ら（2000），宗森ら訳（2004）

索　引

1次式による分散の推　129
2次形式　9
2次形式　50
2次式の1次の項　35
2次式のあてはめ　33
2次式の解の公式　14
2次式の根の公式　36
3次式のあてはめ　132
95%信頼区間の計算シート　17
95%信頼区間の幅　24
Draperら（1998）　102
Excelの行列関数　4
INVERSE REGRESSION　102
ISO 11095　94
JIS Z 8461　94
JMPによる検証　20
Miller ら（2000）　123
Minverse関数　6
Mmult関数　5
Normsdist関数　46
Normsinv関数　39
Normsinv関数　46
Proc regによる重みつき回帰　114
quadratic equation　102
Rand関数　39
Rand関数　45
Snedecorら（1967）　109
Tdist関数　8
Tinv関数　9
Transpose関数　5
t統計量　8
WEIGHTED LEAST SQUARES　103
Xの2次式を用いた分散の推定　105
アドイン　84
いろいろの平均　113
グラフ・ウイザード　18
スムージング　19
ソルバー　43
ソルバー　84
ソルバーの設定画面　87
データの取りまとめ方　123
データ系列の書式設定　19
デザイン行列　5
べきλの推定　84
べき乗の最尤法による推定　89
べき乗を含む回帰分析　33
マーカ　19
マトリックス・マルチプル関数　5
モデルのあてはめ　29
ロジスティック回帰　117
異精度データ　126
一様乱数　39
応答の期待値に対する信頼区間　21
下に凹のデータ　44
化学分析　123
回帰の95%信頼区間の幅　52
回帰パラメータ　5
回帰パラメータの標準誤差　8
回帰パラメータの分散の推定　7
回帰パラメータの分散共分散　23
回帰直線の95%信頼区間　8
回帰直線の逆推定　12
回帰直線の逆推定　15
回帰直線の傾きが負　117
回帰直線の分散　10
回帰分析の前提条件　38
各種の重み　128
簡単化した重み　129
逆行列　6
逆推定の95%信頼区間の作図法　30
逆推定の結果　30
逆推定の模式図　13
吸光度　127
曲線が上に凸　133
曲線となる検量線　78
曲線のあてはめ　131
近似式，x値の誤差の　123
近似的な誤差分散　125
近似的な誤差分散　125
近似的な誤差分散　131
繰返しm回の平均値　16
傾きのみの回帰パラメータ　26
系列　18
蛍光強度　124
計画行列　5
検量線が直線の場合　94
元のデータ　19
原点を通る回帰式　25
個別データの95%信頼区間　10
個別データの95%信頼区間の幅が　53
個別データの逆推定　15
個別データの逆推定　21
個別データの分散　10
個別の期待値に対する信頼区間　21
誤差構造　39
誤差構造　43
誤差行列　5
誤差伝搬の法測　127
誤差分散が$(1+x^2)$ に比例　66
誤差分散がηの関数　114

誤差分散が一定であるモデル 94
誤差分散が最小となるx_0 58
誤差分散の推定 7
校正実験 126
行列V^{-1} 50
行列の積 4
行列計算の基本 16
行列式 23
高濃度で（負の）曲がり 131
合成分 125
合成分散 125
合成分散 130
最尤法による推定 84
散布図 18
残差の不均衡 97
指数関数的な関数 119
宗森ら訳(2004) 123
重ねあわせプロット 29
重みつき回帰の実用的な解析手順 79
重みに用いる x のべき 89
重みのべきを変化させた場合の対数 100
重みは分散の逆数 107
重みは分散の逆数 117
重み付き回帰式 50
重み付き回帰分析 49
重み付き重心 126
小さめのxにノイズ 63
上に凸のデータ 44
図作成のための計算シート 18
数回測定 123
正規方程式 22
正規方程式 32
正規方程式の誘導 57
切片を含めない 29
扇型の残差プロッ 41
線形回帰モデルの基本 4
前景 19
相対誤差 126
多項式 131
対数尤度の計算方法 101
竹内啓(1963) 120
逐次計算 115
逐次平方和 132
虫害果の百分率 109
直線をあてはめることが困難な場合 43
直線関係でなく分散の構造も異な 48
転置行列 5
等分散が成立しない場合 113
統計ソフトによる検証 55
日本規格協会(2001) 94
任意の重み 128
背景 19
畑村ら共訳(1972) 109
反復測定 95
比例する残差標準偏差 94
非線形最小2乗法 116
標準正規分布のパーセント点 39
標準偏差SD がxに比例 42
標準偏差がに比例 49
複数データ 11
分散が x に比例し原点を通る直 74
分散が x に比例する場合 60
分散が x2 に比例し原点を通る 71
分散がx^2に比例する回帰直線 49
分散がおおよそ x に比例 67
分散がおおよそ x^2 に比例 63
分散がおおよそにx^2比例 40
分散がおおよそにx比例 41
分散が一定 40
分散に逆比例 126
分散に比例する推定式 130
分散の加法性 10
分散の構造が異なるデータ 47
分散共分散行列 8
分散共分散行列 51
分散共分散行列 131
文献に対する批判的吟味 92
芳賀ら(1996) 111
未知検体の95%信頼区間の例示 98
面積と建築費のデータ 113
林ら(2000) 119

本書の計算に用いて Excel, JMP ファイル一覧

応用回帰_01_回帰の基礎.xls	134 KB
応用回帰_02_回帰の基礎.jmp	6 KB
応用回帰_03_原点1べき乗.xls	119 KB
応用回帰_04_原点.jmp	7 KB
応用回帰_05_2次式.xls	142 KB
応用回帰_06_2次式の1次項なし.xls	91 KB
応用回帰_07_6種の誤差モデル.xls	91 KB
応用回帰_07_6種の誤差モデル.jmp	36 KB
応用回帰_08_重み1_x^2_100.xls	126 KB
応用回帰_09_3種_誤差モデルx^2_or_x.xls	60 KB
応用回帰_10_凹_1_x^2_100.xls	129 KB
応用回帰_11_重み_x^2.jmp	6 KB
応用回帰_11_重み_x^2.xls	83 KB
応用回帰_12_重み_x.xls	97 KB
応用回帰_13_重み_y0+x^2_.xls	102 KB
応用回帰_14_重み_y0+x_.xls	96 KB
応用回帰_15_重み_x^2_原点.xls	78 KB
応用回帰_16_重み_x_原点.xls	68 KB
応用回帰_17_λ1_5_重み_y0+x^2λ_100_.xls	172 KB

応用回帰_18_最尤法_20.xls	50 KB
応用回帰_19_JIS_線形化.xls	84 KB
応用回帰_20_JIS_各種重み.xls	94 KB
応用回帰_21_Draper_2次式重み.jmp	4 KB
応用回帰_21_Draper_2次式重み.xls	83 KB
応用回帰_22_Draper_各種重み.jmp	5 KB
応用回帰_22_Draper_各種重み.xls	80 KB
応用回帰_23_スネデガーりんご.JMP	5 KB
応用回帰_23_スネデガーりんご.xls	57 KB
応用回帰_24_芳賀_各種重み原点.xls	59 KB
応用回帰_25_芳賀_y^重み.xls	62 KB
応用回帰_26_芳賀_各種重み.xls	64 KB
応用回帰_27_Miller_逆推定.xls	61 KB
応用回帰_27_Miller_蛍光強度.jmp	2 KB
応用回帰_28_Miller_重み.jmp	2 KB
応用回帰_28_Miller_重みx2.xls	75 KB
応用回帰_28_Miller_分散推定.jmp	5 KB
応用回帰_29_Miller_曲線.xls	68 KB
応用回帰_29_Miller曲線.jmp	2 KB
応用回帰_30_索引.xls	39 KB

非売品，無断複製を禁ずる

第 3 回 続高橋セミナー　**応用回帰分析 1**

ー 各種の重み付き回帰における逆推定 ー

BioStat 研究所（株）

〒105-0014 東京都 港区 芝 1-12－3

2013 年 2 月 4 日　　高橋 行雄

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