回帰分析 攻略マニュアル

学部レベルの単回帰・重回帰から、推定・検定まで

目次

序 回帰分析の全体像 **4**

序.1 4段階の地図 4

序.2 使う道具：3つの公式 4

序.3 読み方のガイド 5

序.4 この教材の立ち位置 6

1 回帰分析で何をしたいのか **7**

1.1 回帰分析を一言でいうと 7

1.2 出発点：条件付き期待値 8

1.3 誤差項 u 9

1.4 この仮定が言っていること、言っていないこと 9

2 母集団で β は何に等しいか（識別） **11**

2.1 2本の条件を作る 11

2.2 連立方程式を解く 12

3 OLS推定量 **14**

3.1 考え方1：標本対応物 14

3.2 考え方2：残差の2乗和を最小にする 14

3.3 2つの考え方が同じ答えになる 15

3.4 傾きの直感：平均の点からの傾きの加重平均 16

3.5 計算例と図 17

4 当てはめ値・残差・決定係数 20

4.1 定義 20

4.2 残差の2つの性質 21

4.3 決定係数 R^2 22

5 重回帰 24

5.1 モデルと OLS 24

5.2 係数の解釈：「他を一定にして」 25

5.3 「取り除く」の正確な意味 25

5.4 5人のデータで確かめる 26

6 欠落変数バイアス 29

6.1 設定 29

6.2 OLS は何を推定してしまうか 30

6.3 バイアスの向き 30

6.4 標本の中でぴったり成り立つ関係 31

6.5 内生変数 31

7 OLSの統計的性質：不偏性 33

7.1 推定量は確率変数である 33

7.2 古典的仮定 33

7.3 不偏性の証明（単回帰） 33

7.4 標本をとり直すとはどういうことか：32通りの実験 35

8 推定の精度：分散と標準誤差 37

8.1 均一分散のもとでの分散 37

8.2 標準誤差 38

8.3 不均一分散と頑健な標準誤差 39

8.4 OLSは最良か：ガウス＝マルコフ定理（BLUE） 40

9 仮説検定と信頼区間 44

9.1 検定の考え方 44

9.2 t検定 44
9.3 p値 46
9.4 信頼区間 46
9.5 複数の制約の検定：F検定 47
9.6 回帰表の読み方 49

10 ダミー変数と対数 **51**
10.1 ダミー変数 51
10.2 対数を使ったモデル 52

終 まとめと練習問題 **54**
終.1 用語・記号の整理 54
終.2 学部と大学院のちがい 55
終.3 練習問題 55
終.4 解答 56

回帰分析の全体像

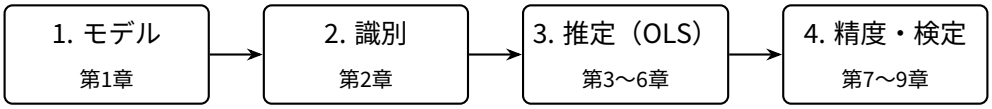
> **この章の問い**　回帰分析は、どんな順番で、何を目指して進むのでしょうか。
> **この章で分かること**　全10章の地図、使う数学の道具（期待値・共分散・偏差の和）、そして本マニュアルが「学部の標準的な内容」に基づくことです。

序.1　4段階の地図

回帰分析は、次の4段階で進みます。本マニュアルもこの順番で並んでいます。

1. **モデルを立てる**（第1章）：知りたい関係を $y = \beta_1 + \beta_2 x + u$ のような式で表します。
2. **母集団で β が何に等しいかを考える**（第2章）：仮定のもとで、β を「観測できる量」で表します。これを**識別**といいます。
3. **データから推定する**（第3〜6章）：最小二乗法（OLS）で $\hat{\beta}$ を計算し、当てはまりの良さ・重回帰・欠落変数バイアスを学びます。
4. **精度を評価し、検定する**（第7〜9章）：不偏性、標準誤差、t検定、信頼区間を学びます。

図 序.1　回帰分析の4段階



図の読み方

1. 左から右へ進みます。「何を知りたいか」（モデル）→「理論上は何で決まるか」（識別）→「データでどう計算するか」（推定）→「どれくらい信用できるか」（検定）という流れです。
2. 多くの教科書は第3段階から始めますが、第2段階（識別）を先に置くと、「OLS はなぜ意味のある推定量なのか」がはっきり分かります。

序.2　使う道具：3つの公式

本文では、次の3つの事実を何度も使います。ここで確認しておきましょう。

要点：道具箱

1. **共分散と分散**：$\mathrm{Cov}(x, y) = \mathrm{E}(xy) - \mathrm{E}(x)\mathrm{E}(y)$、$\mathrm{Var}(x) = \mathrm{E}(x^2) - (\mathrm{E}x)^2$。共分散の性質は「定数は消え、定数倍は外に出て、和は分けられる」です。
2. **条件付き期待値の性質**：x の関数は条件付き期待値の外に出せます。$\mathrm{E}[g(x)\,u \mid x] = g(x)\,\mathrm{E}(u \mid x)$。また繰り返し期待値の法則 $\mathrm{E}[\mathrm{E}(a \mid x)] = \mathrm{E}(a)$ が成り立ちます。
3. **偏差の和は0**：標本平均 $\bar{y}$ について $\sum_{i=1}^{N}(y_i - \bar{y}) = 0$。

補足：偏差の和が0になる理由

$\sum_{i=1}^{N}(y_i - \bar{y}) = \sum_{i=1}^{N} y_i - N\bar{y} = N\bar{y} - N\bar{y} = 0$ です。第3章以降で何度も使うので、見かけたら「また偏差の和が0だ」と思い出してください。

序.3　読み方のガイド

回帰分析を「なんとなく」しか理解できていないと感じる原因の多くは、式の記号と、その式が表している**絵**が結びついていないことにあります。そこで、この教材では各章を次の順番で作っています。

1. 章の冒頭の枠で、「何を知りたいのか」という問いを確認します。
2. 本文で、まず言葉と図で直感をつかみ、そのあと一般的な式で書き直します。
3. 「数式の読み方」の枠で、式の各項が何を意味するかを1つずつ確認します。
4. 「計算例」で、小さなデータ（5人分）を使って自分の手で同じ結果を出します。

初めて読むときは、本文・図・計算例・要点の枠を順に読めば本筋が分かります。緑の枠（発展）は、厳密な証明や大学院につながる内容なので、飛ばしても構いません。2周目で式の導出を1行ずつ追うと、理解が一段深まります。

補足：全章で使う5人のデータ

計算例の多くは、次の小さなデータを共通して使います（教育年数と賃金のような関係を想像してください。数値は説明用の仮想データです）。

i	1	2	3	4	5
x_i	1	2	3	4	5
y_i	2	4	5	4	5

同じデータで何度も計算するので、章をまたいで結果（$\hat{\beta}_2 = 0.6$ など）がつながっていく様子を確認できます。

序.4　この教材の立ち位置

本マニュアルは、Wooldridge『入門計量経済学』などの標準的な学部教科書の内容に基づきます。残差の分解（SST・SSE・SSR）、Frisch–Waugh–Lovell の定理、古典的仮定 MLR.1–6、ガウス＝マルコフ定理（BLUE）、t分布、p値、F検定といった用語は、ここで1から定義して導出します。この教材だけで読み進められます。

発展：大学院との関係

大学院の計量経済学では、行列表記・線形射影・不均一分散を許す漸近理論（一致性、漸近正規性、頑健な標準誤差）が中心になります。本マニュアルの各章末の「大学院とのつながり」で、その橋渡しをします。

次の章へ　まず、回帰分析が「何を」知ろうとしているのかを、条件付き期待値から出発して確認します。

第1章

回帰分析で何をしたいのか

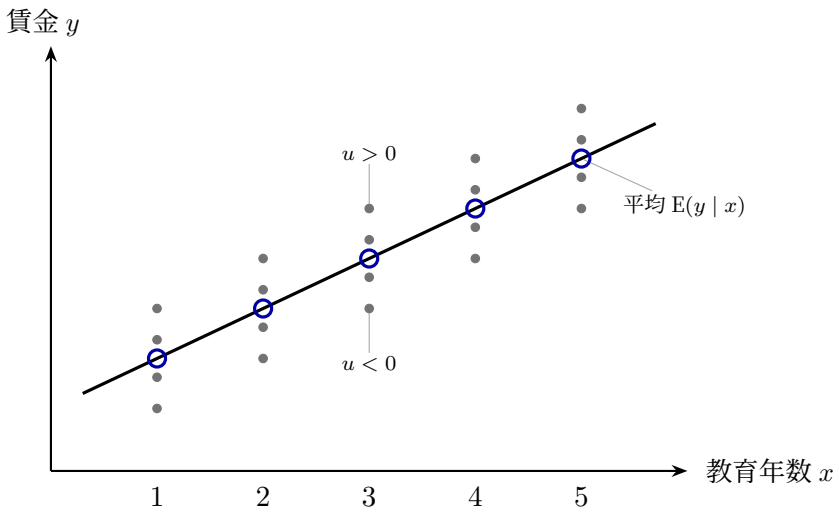
ここまでの流れ　序章で、回帰分析は「モデル→識別→推定→検定」の順に進むことを確認しました。
この章の問い　「教育が1年長いと賃金はどれだけ高いか」という問いは、数式ではどう表せるのでしょうか。
この章で分かること　条件付き期待値 $\mathrm{E}(y \mid x)$ と誤差項 u を使って、単回帰モデル $y = \beta_1 + \beta_2 x + u$ を作る考え方と、その限界です。

1.1　回帰分析を一言でいうと

回帰分析を一言でいえば、「**x が違う人たちの y の平均の違いを、直線で要約する方法**」です。

具体的に考えましょう。ある集団の人々について、教育年数 x と賃金 y のデータがあるとします。散布図を描くと、点は直線上にぴったり並ぶのではなく、雲のように散らばります。それでも、教育年数ごとに平均をとると、教育年数が長い層ほど平均賃金が高い、という傾向が見えてきます。回帰分析は、この「各 x での平均」を結んだ線を、1本の直線でうまく表そうとします。

図 1.1　回帰分析のイメージ（模式図）



説明のための模式図。直線は $\mathrm{E}(y \mid x) = 1 + 0.8x$ とし、各 x に4人ずつ、平均からのずれが $-0.8, -0.3, +0.3, +0.8$ の人を置いた（各 x でずれの平均は0）。

図の読み方

1. 灰色の点が1人1人のデータです。同じ教育年数でも賃金はばらつきます。
2. 青い輪が、各教育年数での平均（条件付き期待値）です。黒い直線は、この平均を結んだ線で、これが回帰直線に当たります。
3. 各点と青い輪の縦のずれが誤差項 u です。上にずれた人は $u > 0$、下にずれた人は $u < 0$ で、同じ x の中で平均すれば u は0になります。
4. 傾きは、「教育が1年長い層は、平均賃金が直線の傾きの分だけ高い」ことを表します。

要点：回帰分析の2つの使い道

- **予測**：x が分かっている人の y を、平均的にどれくらいと見込むか。直線をそのまま使えば足ります。
- **因果効果の推定**：x を変えたら y がどれだけ変わるか。こちらは第6章で見るように、強い仮定が必要です。

同じ回帰直線でも、使い道によって求められる仮定の強さが違うことを意識しておくと、後の章が分かりやすくなります。

補足：「回帰」という名前

「回帰（regression）」という名前は、19世紀にゴルトンが親と子の身長の関係を調べたとき、背の高い親の子は平均して親ほど高くならず、平均に「戻る」傾向があると指摘したことに由来します。現在の「回帰分析」は、この意味に限らず、x と y の平均的な関係を調べる方法全般を指します。

1.2 出発点：条件付き期待値

「教育年数 x が1年長い人は、平均して賃金 y がどれだけ高いか」を知りたいとします。同じ教育年数の人でも賃金はばらばらなので、1人1人の賃金を x でぴったり当てることはできません。そこで、「教育年数ごとの平均賃金」を考えます。これが条件付き期待値 $\mathrm{E}(y \mid x)$ です。

$\mathrm{E}(y \mid x)$ は x の関数ですが、どんな形かは分かりません。そこで、最も簡単な直線を仮定します。

$$\mathrm{E}(y \mid x) = \beta_1 + \beta_2 x. \tag{1.1}$$

数式の読み方：直線の2つの係数

- β_1（切片）：$x = 0$ の人の平均的な y です。
- β_2（傾き）：x が1単位増えたときの、y の**平均**の変化です。これが私たちの知りたいものです。

1.3 誤差項 u

1人1人の y は平均からずれます。そのずれを $u = y - \mathrm{E}(y \mid x)$ と書くと、次のモデルになります。

$$y = \beta_1 + \beta_2 x + u, \qquad \mathrm{E}(u \mid x) = 0. \tag{1.2}$$

説明変数が1つなので、**単回帰モデル**といいます。

数式の読み方：誤差項に含まれるもの

- u には、能力・家庭環境・運など、x 以外で y に影響するすべてのものが含まれます。
- $\mathrm{E}(u \mid x) = 0$ は、u を「y と条件付き期待値の差」と定義すれば自動的に成り立ちます。確かめてみましょう。

$$\begin{aligned} \mathrm{E}(u \mid x) &= \mathrm{E}\big(y - \mathrm{E}(y \mid x) \,\big|\, x\big) && (u \text{ の定義を代入}) \\ &= \mathrm{E}(y \mid x) - \mathrm{E}(y \mid x) && (\mathrm{E}(y \mid x) \text{ は } x \text{ の関数なので外に出る}) \\ &= 0. \end{aligned}$$

1.4 この仮定が言っていること、言っていないこと

要点：2つの注意

1. 式(1.1)の「$\mathrm{E}(y \mid x)$ が本当に直線である」ことは**仮定**です。現実の関係が曲がっていれば、この仮定は誤りです。
2. β_2 を「教育の**因果**効果」と解釈したいなら、u の中身（能力など）が x と無関係であることが必要です。これは強い仮定で、成り立たないと第6章の欠落変数バイアスが生じます。

発展：大学院では

大学院では、$\mathrm{E}(y \mid x)$ が直線でない場合でも意味を持つ**線形射影** $L(y \mid x)$ を推定対象として区別します。また、$\mathrm{E}(u \mid x) = 0$（外生性条件）と、より弱い $\mathrm{E}(xu) = 0$（直交条件）を区別します。後者は第2章で出てきます。

要点

- 回帰分析は「x ごとの y の平均（条件付き期待値）」を直線で表す試みです。
- 誤差項 u は「y のうち x で説明されない部分」で、$\mathrm{E}(u \mid x) = 0$ を満たします。
- 直線を仮定すること自体が1つの仮定で、因果的解釈にはさらに強い仮定が必要です。

次の章へ　モデルが立ちました。次は「もし母集団全体が分かっていたら、β_1, β_2 を計算できるか」を考えます。

大学院とのつながり　この「母集団での決まり方」の議論が、大学院で**識別**と呼ばれるものです。

第2章

母集団で β は何に等しいか（識別）

ここまでの流れ　$y = \beta_1 + \beta_2 x + u$、$\mathrm{E}(u \mid x) = 0$ というモデルを立てました。
この章の問い　母集団全体が分かっていれば、β_1, β_2 はただ1つに決まるでしょうか。
この章で分かること　$\beta_2 = \mathrm{Cov}(x, y)/\mathrm{Var}(x)$ という結果と、その直感です。これが次章の OLS の出発点になります。

2.1　2本の条件を作る

以下、$\mathrm{E}(x^2)$ と $\mathrm{E}(y^2)$ は有限であると仮定します（そうでないと $\mathrm{E}(xu)$ などの期待値が定義できません）。このとき、$\mathrm{E}(u \mid x) = 0$ からは次の2つの条件が導かれます。

$$\text{(i) } \mathrm{E}(u) = 0, \qquad \text{(ii) } \mathrm{E}(xu) = 0. \tag{2.1}$$

数式の読み方：2つの条件の意味

- (i)：誤差は平均すると0です。
- (ii)：誤差と x は（直線的に）無関係です。
- 以下では $\mathrm{E}(x^2) < \infty$、$\mathrm{E}(y^2) < \infty$（期待値や共分散が存在すること）を前提とします。

導出は次の2行です。

$$\mathrm{E}(u) = \mathrm{E}\big[\mathrm{E}(u \mid x)\big] = \mathrm{E}[0] = 0 \qquad \text{（繰り返し期待値の法則）}$$
$$\mathrm{E}(xu) = \mathrm{E}\big[\mathrm{E}(xu \mid x)\big] = \mathrm{E}\big[x\,\mathrm{E}(u \mid x)\big] = \mathrm{E}[x \cdot 0] = 0 \qquad \text{（x は条件付き期待値の外に出せる）}$$

ここに $u = y - \beta_1 - \beta_2 x$ を代入すると、未知数 β_1, β_2 についての2本の式が得られます。

$$\text{(i)} \quad \mathrm{E}(y) - \beta_1 - \beta_2 \mathrm{E}(x) = 0, \tag{2.2}$$
$$\text{(ii)} \quad \mathrm{E}(xy) - \beta_1 \mathrm{E}(x) - \beta_2 \mathrm{E}(x^2) = 0. \tag{2.3}$$

2.2 連立方程式を解く

$$
\begin{aligned}
\beta_1 &= \mathrm{E}(y) - \beta_2 \mathrm{E}(x) && \text{（式(2.2)を } \beta_1 \text{ について解く）} \\
\mathrm{E}(xy) - [\mathrm{E}(y) - \beta_2 \mathrm{E}(x)]\mathrm{E}(x) - \beta_2 \mathrm{E}(x^2) &= 0 && \text{（これを式(2.3)に代入）} \\
\mathrm{E}(xy) - \mathrm{E}(x)\mathrm{E}(y) &= \beta_2 \left[\mathrm{E}(x^2) - (\mathrm{E}x)^2\right] && \text{（} \beta_2 \text{ を含む項を右辺へ）} \\
\mathrm{Cov}(x, y) &= \beta_2 \mathrm{Var}(x) && \text{（共分散・分散の公式）}
\end{aligned}
$$

したがって次の結果が得られます。

$$
\beta_2 = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)}, \qquad \beta_1 = \mathrm{E}(y) - \beta_2 \mathrm{E}(x). \tag{2.4}
$$

ただし $\mathrm{Var}(x) > 0$、つまり x が定数でないことが必要です。これが単回帰版の**ランク条件**です。

数式の読み方：式(2.4)の読み方

- $\mathrm{Cov}(x, y)$：x と y が一緒に動く度合いです。
- $\mathrm{Var}(x)$：x 自身の動きの大きさです。
- その比 β_2 は、「x が1動くと y がどれだけ動くか」という比率になります。
- $\mathrm{Var}(x) = 0$ なら x は全員同じ値で、傾きは決まりません（0で割れません）。

補足：「識別」とは何か、なぜ先に考えるのか

識別とは、「母集団全体のデータが手に入ったと仮定したとき、知りたい β がただ1つの値に決まるか」という問いです。これを先に確認するのには理由があります。

- 識別されないなら、データがどれだけ増えても β は分かりません。推定の技術以前の問題だからです。
- 逆に識別されるなら、あとは「母集団の量を標本で置き換える」だけで推定量が作れます。次章の OLS がその例です。
- 例えば $\mathrm{Var}(x) = 0$（全員が同じ教育年数）なら、教育の効果は全く観察できません。式(2.4)の分母が0になるのは、この状況を表しています。
- また、$\mathrm{E}(xu) = 0$ が成り立たない（x が誤差と相関している）と、式(2.4)は β_2 ではなく別のものに等しくなります。第6章で見る欠落変数バイアスは、この例です。

計算例：仮想の母集団での確認

$x=1$（大卒）の人が半分、$x=0$（非大卒）の人が半分いる仮想の母集団で、平均月給が非大卒28、大卒34（単位は万円）だとします。このとき $\mathrm{E}(x)=0.5$、$\mathrm{E}(y)=31$、$\mathrm{E}(xy)=0.5\times 34=17$ なので

$$\mathrm{Cov}(x,y)=17-0.5\times 31=1.5,\qquad \mathrm{Var}(x)=0.5-0.25=0.25.$$

よって $\beta_2=1.5/0.25=6$、$\beta_1=31-6\times 0.5=28$ です。これは「非大卒の平均28、大卒との差6」と一致します。

発展：行列で書くと

$\mathbf{x}_i=(1,x_i)$ とすると、式(2.1)の2本は $\mathrm{E}[\mathbf{x}_i^\top(y_i-\mathbf{x}_i\beta)]=0$ とまとめられ、

$$\beta=[\mathrm{E}(\mathbf{x}_i^\top\mathbf{x}_i)]^{-1}\mathrm{E}(\mathbf{x}_i^\top y_i)$$

と解けます。単回帰で計算すると式(2.4)になります。「母集団の量で β がただ1つに決まる」ことを、大学院では**点識別される**といいます。

要点

- $\mathrm{E}(u\mid x)=0$ から $\mathrm{E}(u)=0$ と $\mathrm{E}(xu)=0$ の2本の条件が出て、β_1,β_2 が決まります。
- $\beta_2=\mathrm{Cov}(x,y)/\mathrm{Var}(x)$、$\beta_1=\mathrm{E}(y)-\beta_2\mathrm{E}(x)$。
- 必要なのは $\mathrm{Var}(x)>0$（x が定数でないこと）だけです。

次の章へ　母集団の E、Cov、Var は分かりません。次章では、これらを標本の平均で置き換えて OLS 推定量を作ります。

大学院とのつながり　「識別結果の標本対応物が推定量になる」という作り方は、大学院でも同じです。

第3章

OLS推定量

ここまでの流れ　母集団では $\beta_2 = \mathrm{Cov}(x,y)/\mathrm{Var}(x)$ と決まることが分かりました。
この章の問い　データから β を推定するには、どう計算すればよいでしょうか。
この章で分かること　「標本対応物」と「残差2乗和の最小化」という2つの考え方が、同じ式 $\hat{\beta}_2 = \sum(x_i-\bar{x})(y_i-\bar{y})/\sum(x_i-\bar{x})^2$ になることです。

3.1　考え方1：標本対応物

母集団の期待値・分散・共分散は分からないので、標本の平均で置き換えます。

$$\hat{\beta}_2 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(x_i-\bar{x})^2}, \qquad \hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}. \tag{3.1}$$

数式の読み方：式(3.1)の読み方

- 分子は標本共分散（の N 倍）、分母は標本分散（の N 倍）です。$\frac{1}{N}$ は分子と分母で打ち消し合います。
- 第2章の $\mathrm{Cov}(x,y)/\mathrm{Var}(x)$ の、期待値を標本平均に置き換えたものです。

3.2　考え方2：残差の2乗和を最小にする

散布図に直線 $b_1 + b_2 x$ を引くと、各点の縦のずれ（残差）は $y_i - b_1 - b_2 x_i$ です。ずれには正と負があり、そのまま足すと打ち消し合うので、2乗してから足して最小にします。

$$(\hat{\beta}_1, \hat{\beta}_2) = \arg\min_{b_1,b_2} S(b_1,b_2), \qquad S(b_1,b_2) = \sum_{i=1}^{N}(y_i - b_1 - b_2 x_i)^2. \tag{3.2}$$

「2乗（square）を最小（least）にする」ので**最小二乗法**（OLS: Ordinary Least Squares）といいます。2乗を使う理由は次のとおりです。

- 正負が打ち消し合いません。
- 微分できるので計算しやすいです（絶対値は微分しにくい）。
- 母集団でも「$\mathrm{E}(y \mid x)$ は2乗誤差を最小にする関数」という性質があり、これと対応します。

3.3 2つの考え方が同じ答えになる

S を b_1, b_2 でそれぞれ偏微分して0とおきます（最小点では傾きが0）。

b_1 での偏微分

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{N}(y_i - b_1 - b_2 x_i) = 0 \qquad \text{（合成関数の微分：内側の微分は } -1\text{）}$$

$$\sum_{i=1}^{N} y_i = Nb_1 + b_2\sum_{i=1}^{N} x_i \qquad \text{（整理）}$$

$$\bar{y} = b_1 + b_2\bar{x} \qquad \text{（両辺を } N \text{ で割る）}$$

よって $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}$ です。回帰直線は必ず平均の点 $(\bar{x}, \bar{y})$ を通ります。

b_2 での偏微分

$$\frac{\partial S}{\partial b_2} = -2\sum_{i=1}^{N} x_i(y_i - b_1 - b_2 x_i) = 0 \qquad \text{（内側の微分は } -x_i\text{）}$$

$$\sum_{i=1}^{N} x_i\big[(y_i - \bar{y}) - b_2(x_i - \bar{x})\big] = 0 \qquad \text{（}b_1 = \bar{y} - b_2\bar{x} \text{ を代入）}$$

$$\sum_{i=1}^{N} x_i(y_i - \bar{y}) = b_2\sum_{i=1}^{N} x_i(x_i - \bar{x}) \qquad \text{（移項）}$$

$$\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = b_2\sum_{i=1}^{N}(x_i - \bar{x})^2 \qquad \text{（下の注を参照）}$$

最後の行は、偏差の和が0（$\sum(y_i - \bar{y}) = 0$）なので $\sum \bar{x}(y_i - \bar{y}) = 0$ であり、これを左辺から引いても値が変わらないこと、右辺も同様であることを使っています。したがって

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

となり、考え方1と同じ式になりました。

補足：つまずきポイント

ステップの途中で使った「偏差の和は0」は、このあとも何度も使う重要な事実です。見かけたら、序章の道具箱を思い出してください。

3.4 傾きの直感：平均の点からの傾きの加重平均

式(3.1)の意味を、もう少し直感的に理解しましょう。$x_i \neq \bar{x}$ の人について、分子と分母を書き直すと

$$\hat{\beta}_2 = \sum_i \underbrace{\frac{(x_i - \bar{x})^2}{S_{xx}}}_{\text{重み } \omega_i} \cdot \underbrace{\frac{y_i - \bar{y}}{x_i - \bar{x}}}_{\text{人 } i \text{ の傾き } s_i}, \qquad S_{xx} = \sum_j (x_j - \bar{x})^2. \tag{3.3}$$

各項を S_{xx} で割って足しているだけで、$\frac{(x_i-\bar{x})^2}{S_{xx}} \cdot \frac{y_i-\bar{y}}{x_i-\bar{x}} = \frac{(x_i-\bar{x})(y_i-\bar{y})}{S_{xx}}$ だから、式(3.1)と同じです。

数式の読み方：式の読み方

- $s_i = \dfrac{y_i - \bar{y}}{x_i - \bar{x}}$ は、平均の点 $(\bar{x}, \bar{y})$ と人 i の点を結んだ直線の傾きです。人 i が「自分の点を見る限り、傾きはこれだ」と主張している値です。
- $\omega_i = (x_i - \bar{x})^2 / S_{xx}$ は重みで、$\omega_i \geq 0$ かつ $\sum \omega_i = 1$ です。
- したがって OLS の傾きは、「各人の傾きの加重平均」です。x が平均から離れた人ほど重みが大きくなります。
- 理由：x が平均に近い人は、x の違いが小さいため、y の違いが誤差 u によるものか x によるものか区別しにくく、情報が少ないからです。逆に x が平均から大きく離れた人は、傾きについて多くの情報を持ちます。

計算例：5人のデータ

i	1	2	3	4	5
$x_i - \bar{x}$	-2	-1	0	1	2
$y_i - \bar{y}$	-2	0	1	0	1
傾き s_i	1	0	（定まらない）	0	0.5
重み ω_i	$4/10$	$1/10$	0	$1/10$	$4/10$

加重平均は $0.4 \times 1 + 0.1 \times 0 + 0.1 \times 0 + 0.4 \times 0.5 = 0.6$ となり、$\hat{\beta}_2 = 0.6$ と一致します。端の2人（$i = 1, 5$）で重みの80%を占め、$i = 3$（平均と同じ x）は重み0です。

補足：外れ値への注意

x が平均から大きく離れた1人のデータが、傾きに大きな影響を与えうることが、この式から分かります。そのような点を **てこ比の高い点** と呼びます。回帰分析の結果に違和感があるときは、散布図を描いて、端にある点を確認する習慣をつけましょう。

要点：手計算の手順（単回帰）

1. $\bar{x}$ と $\bar{y}$ を求める。
2. 偏差 $x_i - \bar{x}$、$y_i - \bar{y}$ を表にする（合計が0かを確認）。
3. 積 $(x_i - \bar{x})(y_i - \bar{y})$ の合計と、$(x_i - \bar{x})^2$ の合計を求める。
4. 傾き $\hat{\beta}_2 =$（積の合計）/（2乗の合計）。
5. 切片 $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$。

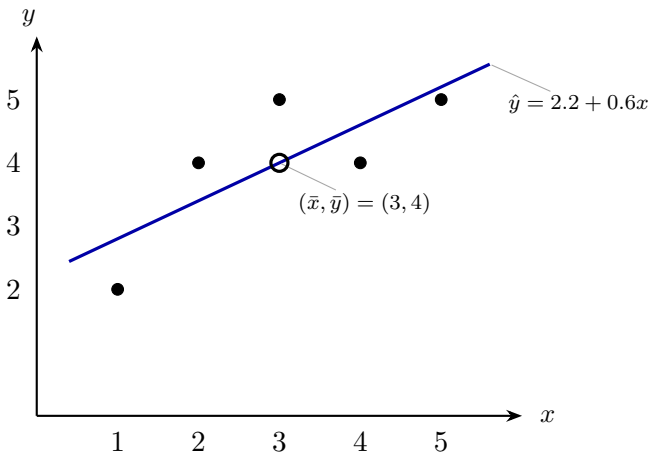
3.5 計算例と図

計算例：5人のデータ

i	1	2	3	4	5	合計/平均
x_i	1	2	3	4	5	平均 3
y_i	2	4	5	4	5	平均 4
$x_i - \bar{x}$	-2	-1	0	1	2	0
$y_i - \bar{y}$	-2	0	1	0	1	0
$(x_i - \bar{x})(y_i - \bar{y})$	4	0	0	0	2	6
$(x_i - \bar{x})^2$	4	1	0	1	4	10

$\hat{\beta}_2 = 6/10 = 0.6$、$\hat{\beta}_1 = 4 - 0.6 \times 3 = 2.2$。回帰直線は $\hat{y} = 2.2 + 0.6x$ です。

図 3.1　5人のデータと回帰直線

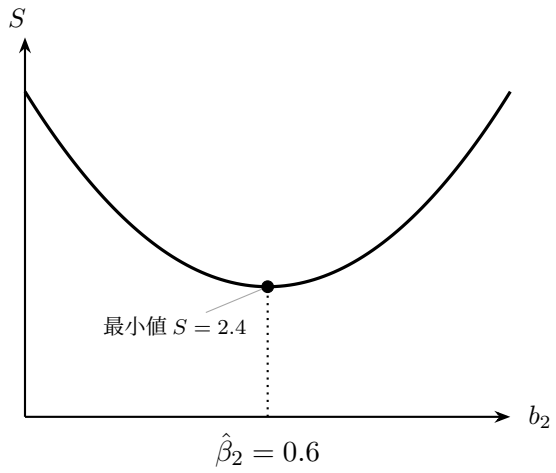


直線は $\hat{y} = 2.2 + 0.6x$。黒い点が観測値、白抜きの丸が平均の点 $(\bar{x}, \bar{y})$。

図の読み方

1. 黒い点が5人の観測値、青い直線が OLS で引いた回帰直線です。
2. 白抜きの丸は平均の点 $(\bar{x}, \bar{y}) = (3, 4)$ で、直線はここを必ず通ります（$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$ の意味）。
3. 傾き0.6は「x が1増えると y の平均的な値が0.6増える」ことを表します。

図 3.2　残差2乗和 S は b_2 の2次関数



5人のデータで $b_1 = \bar{y} - b_2 \bar{x}$ を代入した関数 $S(b_2) = \sum[(y_i - \bar{y}) - b_2(x_i - \bar{x})]^2 = 6 - 12b_2 + 10b_2^2$ を描いた。最小点は $b_2 = 0.6$、そのとき $S = 6 - 7.2 + 3.6 = 2.4$。

図の読み方

1. 横軸は傾きの候補 b_2、縦軸は残差の2乗和 S です。
2. 曲線は下に凸の放物線で、最小点がちょうど1つあります。そこが OLS の傾き $\hat{\beta}_2 = 0.6$ です。
3. 微分して0とおくとは、この谷底（接線の傾きが0の点）を探すことです。

要点

- OLS の傾きは $\hat{\beta}_2 = \sum(x_i - \bar{x})(y_i - \bar{y}) / \sum(x_i - \bar{x})^2$、切片は $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$ です。
- 「標本対応物」と「残差2乗和の最小化」は同じ答えになります。
- 回帰直線は必ず平均の点 $(\bar{x}, \bar{y})$ を通ります。

次の章へ　直線を引いたあとに残る「ずれ」＝残差と、当てはまりの良さを表す決定係数を調べます。

大学院とのつながり　行列で書くと $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ の1つの式にまとまります（第5章）。

第4章

当てはめ値・残差・決定係数

ここまでの流れ　OLS で回帰直線 $\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x$ が引けました。
この章の問い　直線はデータにどれくらい当てはまっているのでしょうか。
この章で分かること　当てはめ値・残差の定義、残差が必ず満たす2つの性質、決定係数 R^2 の意味と限界です。

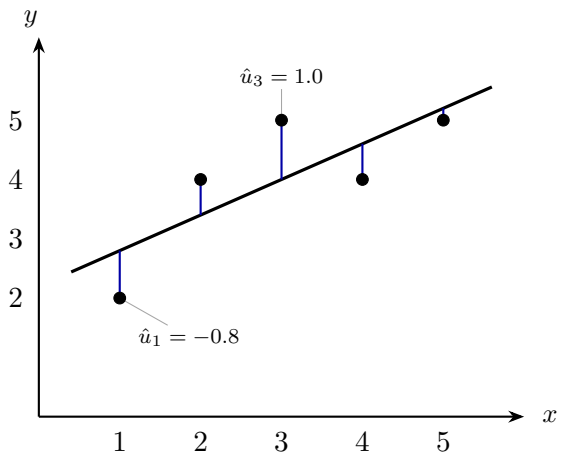
4.1 定義

- **当てはめ値**（予測値）：$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$。回帰直線上の値です。
- **残差**：$\hat{u}_i = y_i - \hat{y}_i$。実際の値と直線の縦のずれです。

計算例：5人のデータ

i	1	2	3	4	5
$\hat{y}_i = 2.2 + 0.6x_i$	2.8	3.4	4.0	4.6	5.2
$\hat{u}_i = y_i - \hat{y}_i$	-0.8	0.6	1.0	-0.6	-0.2

図 4.1　残差は点と直線の縦のずれ



黒い直線は $\hat{y} = 2.2 + 0.6x$、青い縦線が各点の残差。点が直線より上なら残差は正、下なら負。

図の読み方

1. 青い縦線の長さ（符号つき）が残差 $\hat{u}_i$ です。OLS はこの縦線の長さの2乗の和が最小になる直線を選んでいます。

2. 点が直線の上にあれば $\hat{u}_i > 0$、下にあれば $\hat{u}_i < 0$ です。

補足：誤差と残差は別物

誤差 $u_i = y_i - \beta_1 - \beta_2 x_i$ は、真の（未知の）β からのずれで、観測できません。残差 $\hat{u}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$ は推定した $\hat{\beta}$ からのずれで、データから計算できます。残差は誤差の「推定値」です。

4.2 残差の2つの性質

要点

$$\sum_{i=1}^{N} \hat{u}_i = 0, \qquad \sum_{i=1}^{N} x_i \hat{u}_i = 0.$$

これは第3章の偏微分の式（OLS の1階条件）そのものです。

計算例：確認

$-0.8 + 0.6 + 1.0 - 0.6 - 0.2 = 0$。また $1(-0.8) + 2(0.6) + 3(1.0) + 4(-0.6) + 5(-0.2) = -0.8 + 1.2 + 3.0 - 2.4 - 1.0 = 0$。いずれも成り立っています。

数式の読み方：意味

- この2本は、母集団の $\mathrm{E}(u) = 0$、$\mathrm{E}(xu) = 0$ の標本版です。OLS は「データの中でこの2本が成り立つように」$\hat{\beta}$ を選んでいる、と言うこともできます。

- 行列で書けば $X^\top \hat{U} = 0$ です。

- 直感：$\sum \hat{u}_i = 0$ は「直線は点の真ん中を通っていて、上にずれた分と下にずれた分が釣り合う」こと。$\sum x_i \hat{u}_i = 0$ は「残差の中に、x と一緒に動く傾向がもう残っていない」ことです。もし残っていれば（例えば x が大きい人ほど残差が正なら）、直線を少し傾けるだけで残差2乗和をさらに減らせるので、その直線は最小ではなかったことになります。

4.3 決定係数 R^2

y の全体のばらつきは、回帰直線で説明できる部分と残差の部分に分解できます。

$$\underbrace{\sum_{i=1}^{N}(y_i-\bar{y})^2}_{\text{SST（全変動）}} = \underbrace{\sum_{i=1}^{N}(\hat{y}_i-\bar{y})^2}_{\text{SSE（説明された変動）}} + \underbrace{\sum_{i=1}^{N}\hat{u}_i^2}_{\text{SSR（残差変動）}} . \tag{4.1}$$

$$\begin{aligned}\sum(y_i-\bar{y})^2 &= \sum\left[(\hat{y}_i-\bar{y})+\hat{u}_i\right]^2 && (y_i=\hat{y}_i+\hat{u}_i \text{ を代入})\\ &= \sum(\hat{y}_i-\bar{y})^2+\sum\hat{u}_i^2+2\sum(\hat{y}_i-\bar{y})\hat{u}_i && (\text{展開})\end{aligned}$$

交差項が0になることは、$\hat{y}_i=\hat{\beta}_1+\hat{\beta}_2 x_i$ が x_i の1次式であることと、残差の2つの性質から示せます。

$$\begin{aligned}\sum(\hat{y}_i-\bar{y})\hat{u}_i &= \sum(\hat{\beta}_1-\bar{y}+\hat{\beta}_2 x_i)\hat{u}_i && (\hat{y}_i \text{ を代入})\\ &= (\hat{\beta}_1-\bar{y})\underbrace{\sum\hat{u}_i}_{=0}+\hat{\beta}_2\underbrace{\sum x_i\hat{u}_i}_{=0}=0. && (\text{定数は和の外に出す})\end{aligned}$$

よって式(4.1)が成り立ちます。決定係数は

$$R^2=\frac{\text{SSE}}{\text{SST}}=1-\frac{\text{SSR}}{\text{SST}}. \tag{4.2}$$

R^2 は0以上1以下で、「y のばらつきのうち、x で説明できた割合」です。

計算例：5人のデータ

SST $=4+0+1+0+1=6$、SSE $=(-1.2)^2+(-0.6)^2+0^2+0.6^2+1.2^2=3.6$、SSR $=0.64+0.36+1+0.36+0.04=2.4$。$3.6+2.4=6$ で分解が成り立ち、$R^2=3.6/6=0.6$ です。

補足：R^2 が低くても慌てない

R^2 が低いことは、推定が間違っていることを意味しません。賃金は教育以外にも多くの要因で決まるので、R^2 が0.2程度でも、教育の係数 β_2 は正しく推定できていることがあります。計量経済学で重要なのは、多くの場合 R^2 ではなく係数の推定値とその精度です。

要点

- 残差 $\hat{u}_i = y_i - \hat{y}_i$ は誤差 u_i の推定値で、観測できます。
- 残差は必ず $\sum \hat{u}_i = 0$、$\sum x_i \hat{u}_i = 0$ を満たします。
- $R^2 = 1 - \text{SSR}/\text{SST}$ は「x で説明できた割合」で、係数の信頼性とは別の話です。

次の章へ　説明変数が複数になると、係数の意味が「他を一定にして」に変わります。

大学院とのつながり　大学院でも R^2 は定義されますが、その後はほとんど使われません。

第5章

重回帰

ここまでの流れ　説明変数が1つの単回帰で、OLS の計算と当てはまりの測り方を学びました。
この章の問い　説明変数が複数あるとき、各係数は何を意味するのでしょうか。
この章で分かること　「他を一定にして」の正確な意味（Frisch–Waugh–Lovell の定理）です。

5.1　モデルと OLS

説明変数が複数ある場合のモデルは次のとおりです。

$$y = \beta_1 + \beta_2 x_2 + \cdots + \beta_K x_K + u = \mathbf{x}\beta + u, \qquad \mathbf{x} = (1, x_2, \ldots, x_K). \tag{5.1}$$

OLS 推定量は、単回帰と同じく残差2乗和の最小化で求まり、行列で次のように書けます。

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y. \tag{5.2}$$

数式の読み方：式(5.2)の読み方

- X は各行が1人分の $(1, x_{2i}, \ldots, x_{Ki})$ の行列、Y は y_i を縦に並べたベクトルです。
- 単回帰のような簡単な分数の公式はなく、行列で書くのが基本です。これが大学院で行列が必要になる最大の理由です。
- $X^\top X$ に逆行列が存在すること（完全多重共線性がないこと）が必要です。

発展：式(5.2)の導出（正規方程式）

残差2乗和を行列で書くと $S(b) = (Y - Xb)^\top (Y - Xb)$ です。展開すると

$$
\begin{aligned}
S(b) &= Y^\top Y - 2b^\top X^\top Y + b^\top X^\top X b && (b^\top X^\top Y \text{ はスカラーなので転置しても等しい}) \\
\frac{\partial S}{\partial b} &= -2X^\top Y + 2X^\top X b && (\text{ベクトルの微分公式}) \\
&= -2X^\top (Y - Xb) = 0. && (\text{括り出す})
\end{aligned}
$$

これを**正規方程式**といい、$X^\top X b = X^\top Y$ と書き直せます。X の列が1次独立（完全多重共線性がない）なら $X^\top X$ は正定値で逆行列を持つので、$\hat{\beta} = (X^\top X)^{-1} X^\top Y$ が得られます。2階微分 $2X^\top X$ が正定値なので、これは最小点です。また正規方程式は $X^\top (Y - X\hat{\beta}) = X^\top \hat{U} = 0$、つまり第4章の残差の性質そのものです。

5.2 係数の解釈：「他を一定にして」

要点

重回帰の係数 β_j は、「他の説明変数をすべて一定に保ったまま、x_j だけを1単位増やしたときの $\mathrm{E}(y \mid x)$ の変化」です。

$$\beta_j = \frac{\partial \, \mathrm{E}(y \mid x)}{\partial x_j}.$$

計算例：賃金の例

賃金 $= \beta_1 + \beta_2 \times$教育年数$+\beta_3 \times$経験年数$+u$ で $\hat{\beta}_2 = 0.08$ なら、「経験年数が同じ人どうしで比べると、教育が1年長い人は平均して賃金が0.08（単位による）高い」と読みます。

なぜ単回帰と重回帰で係数が変わるのか　教育年数が長い人ほど就職が遅く、経験年数が短い傾向があるとします。賃金を教育だけで単回帰すると、「教育が長い→経験が短い→賃金が低い」という経路も教育の係数に混ざってしまいます。重回帰で経験年数も入れると、その経路が取り除かれ、「経験が同じ人どうしでの教育の差」だけを比べることになります。

5.3 「取り除く」の正確な意味

重回帰の $\hat{\beta}_2$ は、次の2段階の計算で得られる値とぴったり一致します。これが **Frisch–Waugh–Lovell の定理**です。

1. x_2 を、他の説明変数（定数・x_3 など）に回帰し、その残差 $\tilde{x}_2$ を作ります。$\tilde{x}_2$ は「x_2 のうち、他の変数では説明できない部分」です。

2. y を $\tilde{x}_2$ に単回帰した傾き $\sum \tilde{x}_{2i} y_i / \sum \tilde{x}_{2i}^2$ が $\hat{\beta}_2$ です。

つまり重回帰の係数は、「他の変数の影響を取り除いた x_2 の変動だけを使った単回帰の係数」です。「他を一定にして」の正体はこれです。

発展：定理の証明の骨組み

$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_K x_{Ki} + \hat{u}_i$ の両辺に $\tilde{x}_{2i}$ を掛けて足します。

$$\sum \tilde{x}_{2i} y_i = \hat{\beta}_1 \sum \tilde{x}_{2i} + \hat{\beta}_2 \sum \tilde{x}_{2i} x_{2i} + \sum_{j \geq 3} \hat{\beta}_j \sum \tilde{x}_{2i} x_{ji} + \sum \tilde{x}_{2i} \hat{u}_i .$$

- $\tilde{x}_2$ は「定数・$x_3, \ldots$」への回帰の残差なので、定数や x_j（$j \geq 3$）との積和は0です（第4章の残差の性質）。
- $x_2 = \hat{x}_2 + \tilde{x}_2$ で $\sum \tilde{x}_2 \hat{x}_2 = 0$ なので、$\sum \tilde{x}_2 x_2 = \sum \tilde{x}_2^2$ です。
- $\hat{u}$ は全説明変数と直交し、$\tilde{x}_2$ はそれらの1次結合なので $\sum \tilde{x}_2 \hat{u} = 0$ です。

残るのは $\sum \tilde{x}_{2i} y_i = \hat{\beta}_2 \sum \tilde{x}_{2i}^2$ だけで、これを解けば定理の式になります。

5.4 5人のデータで確かめる

第3章の y（賃金）に、経験年数 x_3 を加えた仮想データで、ここまでの内容を手で確かめましょう。x_2 を教育年数とします。

i	1	2	3	4	5
教育 x_{2i}	1	2	3	4	5
経験 x_{3i}	4	3	3	1	1
賃金 y_i	2	4	5	4	5

教育が長い人ほど経験が短い（負の相関）という設定で、これは本章冒頭の例と同じ構造です。

計算例：単回帰と重回帰

(1) 教育だけの単回帰：第3章より $\hat{\beta}_2 = 0.6$。

(2) 経験も入れた重回帰：OLS の解は $\hat{\beta} = (-2.6,\ 1.4,\ 1.0)$ です。連立方程式を解く代わりに、正規方程式（3本の条件）を満たすことで確かめます。当てはめ値は $\hat{y} = -2.6 + 1.4x_2 + 1.0x_3 = (2.8,\ 3.2,\ 4.6,\ 4.0,\ 5.4)$、残差は $\hat{u} = (-0.8,\ 0.8,\ 0.4,\ 0,\ -0.4)$ です。

$$\begin{aligned}\sum \hat{u}_i &= -0.8 + 0.8 + 0.4 + 0 - 0.4 = 0,\\ \sum x_{2i}\hat{u}_i &= -0.8 + 1.6 + 1.2 + 0 - 2.0 = 0,\\ \sum x_{3i}\hat{u}_i &= -3.2 + 2.4 + 1.2 + 0 - 0.4 = 0.\end{aligned}$$

3つとも0なので、これが OLS 解です。$\mathrm{SSR} = 0.64 + 0.64 + 0.16 + 0 + 0.16 = 1.6$、$R^2 = 1 - 1.6/6 \approx 0.733$（教育だけの $R^2 = 0.6$ より上がりました）。

(3) Frisch–Waugh–Lovell の確認：x_2 を定数と x_3 に回帰すると $\hat{x}_2 = \frac{17}{3} - \frac{10}{9}x_3$ で、残差は

$$\tilde{x}_2 = \frac{1}{9}(-2,\ -3,\ 6,\ -5,\ 4).$$

確かめると $\sum \tilde{x}_{2i} = 0$、$\sum \tilde{x}_{2i}x_{3i} = \frac{1}{9}(-8 - 9 + 18 - 5 + 4) = 0$。そして

$$\sum \tilde{x}_{2i}y_i = \frac{14}{9}, \qquad \sum \tilde{x}_{2i}^2 = \frac{10}{9} \quad \Longrightarrow \quad \frac{14/9}{10/9} = 1.4.$$

これは重回帰の $\hat{\beta}_2 = 1.4$ とぴったり一致します。

数式の読み方：結果の読み方

- 教育の係数は、単回帰の0.6から重回帰の1.4へ大きく変わりました。
- 理由：教育が長い人ほど経験が短く（x_3 を x_2 に回帰した傾きは -0.8）、経験は賃金を上げる（$\hat{\beta}_3 = 1.0$）ので、単回帰の0.6には「教育が長い→経験が短い→賃金が低い」という経路が含まれ、教育の効果が小さく見えていました。
- 重回帰の1.4は、「経験が同じ人どうしで比べたときの、教育1年あたりの賃金差」です。
- この「0.6と1.4の差」を式で説明するのが、次章の欠落変数バイアスです。

要点

- 重回帰でも OLS は残差2乗和の最小化で、$\hat{\beta} = (X^\top X)^{-1} X^\top Y$ です。
- β_j は「他を一定にしたときの x_j の部分効果」です。
- 「他を一定にする」とは、x_j から他の変数で説明できる部分を取り除くことです。

次の章へ　重要な変数を入れ忘れると何が起きるか。欠落変数バイアスを調べます。

大学院とのつながり　大学院でも「部分効果 $\partial \mathrm{E}[y \mid x]/\partial x_j$」として同じ解釈が使われます。

第6章

欠落変数バイアス

ここまでの流れ　重回帰の係数は「他を一定にした」効果を表すことが分かりました。
この章の問い　本当は影響している変数をモデルに入れ忘れると、OLS は何を推定してしまうのでしょうか。
この章で分かること　バイアスの式 $\gamma \,\mathrm{Cov}(x,q)/\mathrm{Var}(x)$、その符号の読み方、内生変数という言葉です。

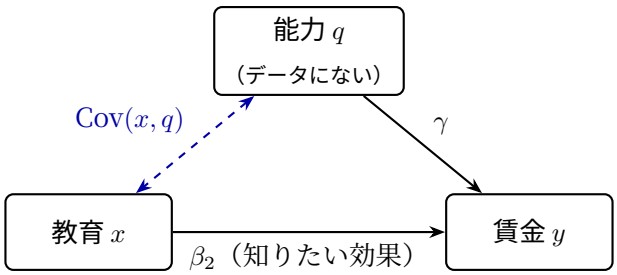
6.1　設定

本当のモデルが

$$y = \beta_1 + \beta_2 x + \gamma q + v, \qquad \mathrm{E}(v \mid x, q) = 0 \tag{6.1}$$

だとします。x＝教育年数、q＝能力、y＝賃金です。ところが能力 q はデータにないので、y を x だけに単回帰します。このとき誤差は $u = \gamma q + v$ となり、能力が誤差の中に入り込みます。

図 6.1　欠落変数の構図



図の読み方

1. 実線の矢印は因果です。教育 x が賃金 y に与える効果 β_2 が知りたいもので、能力 q も賃金に効きます（γ）。
2. 青い破線は、能力と教育が相関していること（$\mathrm{Cov}(x,q) \neq 0$）を表します。
3. q を入れないと、「q 経由の関係」まで教育の効果として読み込んでしまいます。

6.2 OLS は何を推定してしまうか

第2章より、単回帰の OLS が（標本が大きいとき）推定しているのは $\mathrm{Cov}(x,y)/\mathrm{Var}(x)$ です。ここに本当のモデルを代入します。共分散の性質（定数は消え、定数倍は外に出て、和は分けられる）を使うと

$$\begin{aligned}\mathrm{Cov}(x,y) &= \mathrm{Cov}(x, \beta_1 + \beta_2 x + \gamma q + v) && \text{（本当のモデルを代入）}\\ &= \beta_2 \mathrm{Var}(x) + \gamma \mathrm{Cov}(x,q) + \mathrm{Cov}(x,v) && \text{（共分散の性質）}\\ &= \beta_2 \mathrm{Var}(x) + \gamma \mathrm{Cov}(x,q) && \text{（}\mathrm{E}(v \mid x,q)=0 \text{ より } \mathrm{Cov}(x,v)=0\text{）}\end{aligned}$$

両辺を $\mathrm{Var}(x)$ で割ると

$$\frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)} = \beta_2 + \underbrace{\gamma \frac{\mathrm{Cov}(x,q)}{\mathrm{Var}(x)}}_{\text{欠落変数バイアス}}. \tag{6.2}$$

数式の読み方：式(6.2)の読み方

- 第1項 β_2 は本当の効果、第2項がバイアスです。
- バイアスは「欠落変数が y に与える効果 γ」×「欠落変数と x の関係の強さ $\mathrm{Cov}(x,q)/\mathrm{Var}(x)$」です。
- $\mathrm{Cov}(x,q)/\mathrm{Var}(x)$ は、q を x に単回帰したときの傾きに当たります。

6.3 バイアスの向き

バイアスの符号は「γ の符号 $\times$ $\mathrm{Cov}(x,q)$ の符号」で決まります。

	$\mathrm{Cov}(x,q)>0$	$\mathrm{Cov}(x,q)<0$
$\gamma>0$	過大評価（上向き）	過小評価（下向き）
$\gamma<0$	過小評価（下向き）	過大評価（上向き）

計算例：教育と能力

能力が高い人ほど賃金が高く（$\gamma>0$）、能力が高い人ほど教育も長い（$\mathrm{Cov}(x,q)>0$）なら、バイアスは正です。教育だけで回帰すると、能力の効果まで教育の効果と取り違えて、教育の効果を過大評価します。

バイアスがなくなるのは、$\gamma=0$（q が y に影響しない）か、$\mathrm{Cov}(x,q)=0$（q と x が無相関）のどちらかのときだけです。

6.4 標本の中でぴったり成り立つ関係

式(6.2)は母集団での関係でしたが、これと対応する関係は、手元のデータの OLS の計算結果の間でも**厳密に**成り立ちます。これを理解すると、バイアスの正体がはっきりします。

説明変数が x_2（教育）と x_3（経験）で、x_3 を落として y を x_2 だけに回帰した傾きを $\tilde{\beta}_2$、x_3 を含めた重回帰の係数を $\hat{\beta}_2, \hat{\beta}_3$、$x_3$ を x_2 に回帰した傾きを $\hat{\delta}$ とします。このとき次の式が成り立ちます。

$$\tilde{\beta}_2 = \hat{\beta}_2 + \hat{\beta}_3 \hat{\delta}. \tag{6.3}$$

$$\begin{aligned}
\tilde{\beta}_2 &= \frac{\sum (x_{2i} - \bar{x}_2) y_i}{\sum (x_{2i} - \bar{x}_2)^2} && \text{（単回帰の傾き）} \\
&= \frac{\sum (x_{2i} - \bar{x}_2)(\hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i)}{\sum (x_{2i} - \bar{x}_2)^2} && \text{（当てはめ式を代入）} \\
&= \hat{\beta}_2 + \hat{\beta}_3 \frac{\sum (x_{2i} - \bar{x}_2) x_{3i}}{\sum (x_{2i} - \bar{x}_2)^2} + \frac{\sum (x_{2i} - \bar{x}_2) \hat{u}_i}{\sum (x_{2i} - \bar{x}_2)^2} && \text{（各項に分解）} \\
&= \hat{\beta}_2 + \hat{\beta}_3 \hat{\delta} + 0. && \text{（第2項の分数は } \hat{\delta}\text{、第3項は0）}
\end{aligned}$$

計算例：第5章のデータ

$\hat{\beta}_2 = 1.4$、$\hat{\beta}_3 = 1.0$、$\hat{\delta} = -0.8$ なので、$1.4 + 1.0 \times (-0.8) = 0.6 = \tilde{\beta}_2$。単回帰の0.6が、重回帰の1.4から 0.8 低くなっている理由が、数式で説明できました。

数式の読み方：式(6.3)の読み方

- 単回帰の傾き＝「本当の（他を一定にした）効果 $\hat{\beta}_2$」＋「経験を落としたことで混ざった分 $\hat{\beta}_3\hat{\delta}$」。
- $\hat{\beta}_3$ は欠落した変数の効果（母集団の式の γ に対応）、$\hat{\delta}$ は欠落した変数と x_2 の関係の強さ（$\mathrm{Cov}(x,q)/\mathrm{Var}(x)$ に対応）です。
- 符号は $\hat{\beta}_3$ と $\hat{\delta}$ の符号の積で決まります。上の例では $1.0 > 0$ と $-0.8 < 0$ で積が負なので、単回帰は重回帰より低くなります。

6.5 内生変数

欠落変数があるとき、$\mathrm{Cov}(x, u) = \mathrm{Cov}(x, \gamma q + v) = \gamma \mathrm{Cov}(x, q) \neq 0$ となり、x は誤差と相関します。このような説明変数を**内生変数**といいます。第2章の直交条件 $\mathrm{E}(xu) = 0$ が崩れているので、OLS は β_2 を識別できません。

発展：解決策

- **代理変数**：能力の代わりに、能力と関連する変数（例：IQ）をモデルに入れます。
- **操作変数法**：x と相関するが誤差とは無相関な変数を使います。大学院で学ぶ重要な手法です。

計算例：練習：向きの判断

y＝農家の収穫量、x＝肥料の投入量、q＝土地のやせ具合とします。やせた土地ほど収穫が少なく（$\gamma < 0$）、やせた土地ほど肥料を多く使う（$\mathrm{Cov}(x, q) > 0$）なら、バイアスは $\gamma \mathrm{Cov}/\mathrm{Var} < 0$ なので肥料の効果を過小評価します。

要点

- 欠落変数があると、OLS は $\beta_2 + \gamma \mathrm{Cov}(x, q)/\mathrm{Var}(x)$ を推定します。
- バイアスの符号は γ と $\mathrm{Cov}(x, q)$ の符号の積で決まります。
- x が誤差と相関する（内生変数）とき、OLS は因果効果を識別できません。

次の章へ　バイアスがない場合（$\mathrm{E}(u \mid x) = 0$）に、OLS は「平均的には当たる」ことを証明します（不偏性）。

大学院とのつながり　大学院では、この式の行列版 $\operatorname{plim} \hat{\beta} = \beta + A^{-1}\mathrm{E}(\mathbf{x}^\top q)\gamma$ を学びます。$A^{-1} \leftrightarrow 1/\mathrm{Var}(x)$、$\mathrm{E}(\mathbf{x}^\top q) \leftrightarrow \mathrm{Cov}(x, q)$ と対応します。

第7章

OLSの統計的性質：不偏性

> **ここまでの流れ**　$\mathrm{E}(u \mid x) = 0$ が崩れると OLS は偏る（欠落変数バイアス）ことを見ました。
> **この章の問い**　$\mathrm{E}(u \mid x) = 0$ が成り立つとき、OLS は本当に真の値を当てるのでしょうか。
> **この章で分かること**　古典的仮定と、$\mathrm{E}(\hat{\beta}_2) = \beta_2$（不偏性）の証明です。

7.1 推定量は確率変数である

$\hat{\beta}_2$ はデータから計算されるので、どの標本が得られたかによって値が変わります。同じ母集団から別の5人を選べば、違う $\hat{\beta}_2$ が得られます。つまり $\hat{\beta}_2$ も確率変数で、分布（平均や分散）を持ちます。「推定量の性質を調べる」とは、この分布を調べることです。

7.2 古典的仮定

学部の教科書は、次の仮定を置きます。MLR は multiple linear regression の略で、番号は仮定の通し番号です。

仮定	内容	大学院での扱い
MLR.1	線形モデル $y = \mathbf{x}\beta + u$	モデルの設定
MLR.2	無作為抽出（各人は独立で同じ分布）	IID の仮定
MLR.3	完全多重共線性なし（単回帰なら x が全員同じ値でない）	ランク条件
MLR.4	ゼロ条件付き平均 $\mathrm{E}(u \mid x) = 0$	外生性条件
MLR.5	均一分散 $\mathrm{Var}(u \mid x) = \sigma^2$	置かない
MLR.6	$u \sim \mathrm{Normal}(0, \sigma^2)$ で x と独立	置かない（漸近理論で代用）

7.3 不偏性の証明（単回帰）

要点：定理

MLR.1–4 のもとで、$\mathrm{E}(\hat{\beta}_2) = \beta_2$（不偏）。

前提　MLR.3 より $S_{xx} = \sum (x_i - \bar{x})^2 > 0$ なので、分母は0になりません。

ステップ1：分子を書き換える　$\bar{y}$ は i によらない定数なので $\sum(x_i-\bar{x})\bar{y}=\bar{y}\sum(x_i-\bar{x})=0$ です。よって

$$\sum(x_i-\bar{x})(y_i-\bar{y})=\sum(x_i-\bar{x})y_i-\bar{y}\sum(x_i-\bar{x})=\sum(x_i-\bar{x})y_i.$$

ステップ2：モデル $y_i=\beta_1+\beta_2x_i+u_i$ を代入する　途中で使う等式 $\sum(x_i-\bar{x})x_i=\sum(x_i-\bar{x})^2$ は、$\sum(x_i-\bar{x})^2=\sum(x_i-\bar{x})x_i-\bar{x}\sum(x_i-\bar{x})$ で最後の項が偏差の和が0により消えるために成り立ちます。下の第2項の等式は、$\sum(x_i-\bar{x})x_i=\sum(x_i-\bar{x})x_i-\bar{x}\sum(x_i-\bar{x})=\sum(x_i-\bar{x})^2$（引いた項が0）から出ます。

$$\sum(x_i-\bar{x})y_i=\beta_1\underbrace{\sum(x_i-\bar{x})}_{=0}+\beta_2\underbrace{\sum(x_i-\bar{x})x_i}_{=\sum(x_i-\bar{x})^2}+\sum(x_i-\bar{x})u_i.$$

ステップ3：$\hat{\beta}_2$ を「真の値＋誤差の部分」に分解する　分母 $\sum(x_i-\bar{x})^2$ で割ると

$$\hat{\beta}_2=\beta_2+\frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}. \tag{7.1}$$

推定量＝真の値＋（誤差 u_i の重み付き和）という形になりました。

ステップ4：x を固定して期待値をとる　$x_1,\dots,x_N$ の値を全部知っている（条件付ける）と、分数の中の x の部分はすべて定数です。x の関数は外に出せるので

$$\mathrm{E}(\hat{\beta}_2\mid x_1,\dots,x_N)=\beta_2+\frac{\sum(x_i-\bar{x})\,\mathrm{E}(u_i\mid x_1,\dots,x_N)}{\sum(x_i-\bar{x})^2}.$$

MLR.2（各人は独立）より $\mathrm{E}(u_i\mid x_1,\dots,x_N)=\mathrm{E}(u_i\mid x_i)$、MLR.4 よりこれは0です。よって $\mathrm{E}(\hat{\beta}_2\mid x_1,\dots,x_N)=\beta_2$。

ステップ5：繰り返し期待値の法則

$$\mathrm{E}(\hat{\beta}_2)=\mathrm{E}\big[\mathrm{E}(\hat{\beta}_2\mid x_1,\dots,x_N)\big]=\mathrm{E}[\beta_2]=\beta_2. \qquad \blacksquare$$

数式の読み方：意味

- 何度も標本をとり直して $\hat{\beta}_2$ を計算すると、その平均は真の β_2 に一致します。「平均的には当たる」ということです。
- 1回の推定で当たる保証ではありません。1回のぶれの大きさは次章の分散で測ります。

補足：つまずきポイント

ステップ4で $\mathrm{E}(u \mid x) = 0$（MLR.4）が使われています。欠落変数がある場合のように $\mathrm{E}(u \mid x) \neq 0$ だと、OLS は不偏ではなくなります。また、ステップ4の最初の等号は MLR.2、最後の等号が MLR.4 です。

7.4 標本をとり直すとはどういうことか：32通りの実験

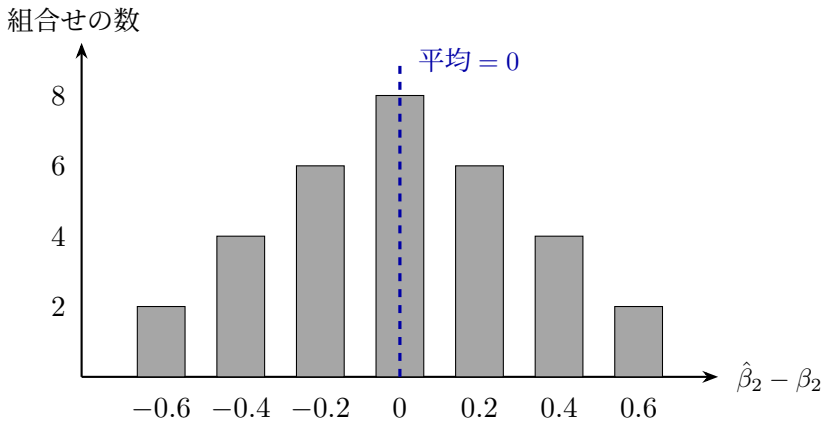
「$\hat{\beta}_2$ は確率変数」「平均的には当たる」という言い方は、最初は抽象的に感じられます。そこで、すべての場合を数え上げられる小さな世界で確かめましょう。

設定　$x = (1, 2, 3, 4, 5)$ を固定し、誤差 u_i は互いに独立に、$+1$ か -1 を確率 $\frac{1}{2}$ ずつでとるとします。このとき $\mathrm{E}(u_i \mid x) = 0$ で、分散は $\sigma^2 = 1$ です。真の β_2 が何であっても、式(7.1)より

$$\hat{\beta}_2 - \beta_2 = \frac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2} = \frac{-2u_1 - u_2 + u_4 + 2u_5}{10}$$

です。$(u_1, \ldots, u_5)$ の組合せは $2^5 = 32$ 通りあり、どれも確率 $1/32$ で起きます。

図 7.1　$\hat{\beta}_2 - \beta_2$ の分布（32通りの数え上げ）



$u_i \in \{-1, +1\}$ の全32通りについて $(-2u_1 - u_2 + u_4 + 2u_5)/10$ を計算し、値ごとの組合せの数を数えた（-0.6 が2通り、-0.4 が4通り、-0.2 が6通り、0 が8通り、0.2 が6通り、0.4 が4通り、0.6 が2通り）。

図の読み方

1. 横軸は「推定値と真の値のずれ」、縦軸はそのずれが生じる組合せの数です。32通りの「標本」それぞれで、$\hat{\beta}_2$ は真の値より高くなったり低くなったりします。
2. 分布は0を中心に左右対称です。青い破線の位置が平均で、ちょうど0なので、「平均的には当たる」（不偏）が目で見て分かります。
3. 1回の標本では、± 0.6 のようにかなりずれることもあります。不偏性は「1回で当たる」ことを保証しません。

計算例：平均と分散の確認

平均は、左右対称なので 0。分散は

$$\frac{2(-0.6)^2 + 4(-0.4)^2 + 6(-0.2)^2 + 8 \cdot 0^2 + 6(0.2)^2 + 4(0.4)^2 + 2(0.6)^2}{32} = \frac{3.2}{32} = 0.1.$$

これは次章の公式 $\sigma^2/S_{xx} = 1/10 = 0.1$ とぴったり一致します。標準偏差は $\sqrt{0.1} \approx 0.316$ で、これが標準誤差の「本当の値」に当たります。

発展：大学院では

大学院でも、$\hat{\beta} = \beta + [\frac{1}{N}\sum \mathbf{x}_i^\top \mathbf{x}_i]^{-1} \frac{1}{N}\sum \mathbf{x}_i^\top u_i$ と分解する点は同じです。この形は、不偏性・一致性・漸近正規性すべての証明の出発点になるので、式(7.1)の形は必ず覚えてください。

要点

- 推定量 $\hat{\beta}_2$ は標本によって値が変わる確率変数です。
- MLR.1–4 のもとで $\mathrm{E}(\hat{\beta}_2) = \beta_2$（不偏）。証明の鍵は $\hat{\beta}_2 = \beta_2 + \sum(x_i - \bar{x})u_i / \sum(x_i - \bar{x})^2$ という分解です。
- $\mathrm{E}(u \mid x) = 0$ が崩れれば不偏性も崩れます。

次の章へ　「平均的に当たる」だけでなく、1回の推定がどれだけぶれるかを分散と標準誤差で測ります。

大学院とのつながり　大学院では不偏性より、標本が大きいとき真の値に近づく**一致性**を中心に学びます。

第8章

推定の精度：分散と標準誤差

ここまでの流れ　OLS は $\mathrm{E}(u \mid x) = 0$ のもとで不偏、つまり平均的には真の値を当てることが分かりました。
この章の問い　1回の推定値は、真の値からどのくらいずれうるのでしょうか。
この章で分かること　$\hat{\beta}_2$ の分散の式、標準誤差の計算、不均一分散と頑健な標準誤差です。

8.1　均一分散のもとでの分散

式(7.1)より、$S_{xx} = \sum_j (x_j - \bar{x})^2$ とおくと

$$\hat{\beta}_2 - \beta_2 = \sum_{i=1}^{N} w_i u_i, \qquad w_i = \frac{x_i - \bar{x}}{S_{xx}}. \tag{8.1}$$

$x_1, \ldots, x_N$ を条件として固定すれば、w_i は定数です。この重み w_i には、次の3つの性質があります（第7章のステップ2と同じ計算です）。

数式の読み方：重み w_i の3つの性質

(a) $\sum w_i = \dfrac{\sum (x_i - \bar{x})}{S_{xx}} = 0$　　（偏差の和は0）

(b) $\sum w_i x_i = \dfrac{\sum (x_i - \bar{x}) x_i}{S_{xx}} = \dfrac{S_{xx}}{S_{xx}} = 1$　　（$\sum (x_i - \bar{x}) x_i = \sum (x_i - \bar{x})^2$）

(c) $\sum w_i^2 = \dfrac{\sum (x_i - \bar{x})^2}{S_{xx}^2} = \dfrac{1}{S_{xx}}$　　（分子が S_{xx}）

前提　MLR.2（独立）と MLR.5（均一分散）から、x を条件として $\mathrm{Var}(u_i \mid x) = \sigma^2$、$\mathrm{Cov}(u_i, u_j \mid x) = 0$（$i \neq j$）が成り立ちます。

$$
\begin{aligned}
\mathrm{Var}(\hat{\beta}_2 \mid x) &= \mathrm{Var}\Big(\beta_2 + \sum_i w_i u_i \;\Big|\; x\Big) && \text{（式(8.1)）}\\
&= \mathrm{Var}\Big(\sum_i w_i u_i \;\Big|\; x\Big) && \text{（定数 } \beta_2 \text{ は分散に影響しない）}\\
&= \sum_i \sum_j w_i w_j \mathrm{Cov}(u_i, u_j \mid x) && \text{（和の分散は共分散の二重和）}\\
&= \sum_i w_i^2 \,\mathrm{Var}(u_i \mid x) && \text{（} i \neq j \text{ の共分散は0）}\\
&= \sigma^2 \sum_i w_i^2 = \frac{\sigma^2}{S_{xx}}. && \text{（MLR.5 と性質(c)）}
\end{aligned}
$$

要点：精密になる条件

推定が精密になる（分散が小さくなる）のは次のときです。

1. 誤差の分散 σ^2 が小さいとき（y のうち x 以外の要因によるばらつきが小さい）。
2. S_{xx} が大きいとき。(a) サンプルサイズ N が大きい、(b) x のばらつきが大きい、のどちらでも大きくなります。

$S_{xx} \approx N\mathrm{Var}(x)$ なので、分散はおよそ $\sigma^2/(N\mathrm{Var}(x))$ です。N が4倍になると分散は1/4、標準偏差は1/2になります。

8.2 標準誤差

σ^2 は未知なので、残差を使って推定します。

$$
\hat{\sigma}^2 = \frac{1}{N-2}\sum_{i=1}^{N} \hat{u}_i^2, \qquad \mathrm{se}(\hat{\beta}_2) = \sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}. \tag{8.2}
$$

$\mathrm{se}(\hat{\beta}_2)$ を**標準誤差**といい、「$\hat{\beta}_2$ の標準偏差の推定値」です。$N-2$ で割るのは、推定した係数の数（2個）だけ残差の自由度が減るためで、こうすると $\hat{\sigma}^2$ が不偏になります。N が大きければ N で割っても違いはほとんどありません。

計算例：5人のデータ

$\mathrm{SSR} = \sum \hat{u}_i^2 = 2.4$、$\hat{\sigma}^2 = 2.4/(5-2) = 0.8$、$\mathrm{se}(\hat{\beta}_2) = \sqrt{0.8/10} = \sqrt{0.08} \approx 0.283$。「傾きの推定値は0.6だが、標本によって ± 0.28 程度はぶれうる」という意味です。

発展：なぜ $N-2$ で割ると不偏になるのか

均一分散のもとで $\mathrm{E}(\hat{\sigma}^2 \mid x) = \sigma^2$ を示します。まず $\bar{y} = \beta_1 + \beta_2 \bar{x} + \bar{u}$ なので、$y_i - \bar{y} = \beta_2(x_i - \bar{x}) + (u_i - \bar{u})$ です。したがって

$$\begin{aligned} \hat{u}_i &= (y_i - \bar{y}) - \hat{\beta}_2(x_i - \bar{x}) && (\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x} \text{ を代入}) \\ &= (u_i - \bar{u}) - (\hat{\beta}_2 - \beta_2)(x_i - \bar{x}). && (\text{上の式を代入して整理}) \end{aligned}$$

両辺を2乗して足します。$\sum (x_i - \bar{x})(u_i - \bar{u}) = \sum (x_i - \bar{x}) u_i = (\hat{\beta}_2 - \beta_2) S_{xx}$（式(7.1)）なので、交差項は $-2(\hat{\beta}_2 - \beta_2)^2 S_{xx}$ となり

$$\sum \hat{u}_i^2 = \sum (u_i - \bar{u})^2 - 2(\hat{\beta}_2 - \beta_2)^2 S_{xx} + (\hat{\beta}_2 - \beta_2)^2 S_{xx} = \sum (u_i - \bar{u})^2 - (\hat{\beta}_2 - \beta_2)^2 S_{xx}.$$

期待値をとります。$\mathrm{E}\sum (u_i - \bar{u})^2 = (N-1)\sigma^2$（標本分散の期待値は $\frac{N-1}{N}$ 倍の補正が必要）、$\mathrm{E}[(\hat{\beta}_2 - \beta_2)^2 S_{xx}] = \mathrm{Var}(\hat{\beta}_2 \mid x) S_{xx} = \sigma^2$ なので

$$\mathrm{E}\Big(\sum \hat{u}_i^2 \,\Big|\, x \Big) = (N-1)\sigma^2 - \sigma^2 = (N-2)\sigma^2.$$

よって $\hat{\sigma}^2 = \frac{1}{N-2} \sum \hat{u}_i^2$ は不偏です。引かれる「2」は、残差が2本の関係式 $\sum \hat{u}_i = 0$、$\sum x_i \hat{u}_i = 0$ を満たす（自由に動ける残差が $N-2$ 個）ことに対応します。

8.3 不均一分散と頑健な標準誤差

不均一分散とは、$\mathrm{Var}(u \mid x)$ が x によって異なることです。例えば所得が高い人ほど支出のばらつきが大きい、など、現実のデータではむしろ普通の状態です。

不均一分散のもとでは、上の分散の公式は誤りになります。$\mathrm{Var}(u_i \mid x_i) = \sigma_i^2$（人によって違う）とすると、正しくは

$$\mathrm{Var}(\hat{\beta}_2 \mid x) = \sum_i w_i^2 \sigma_i^2 = \frac{\sum (x_i - \bar{x})^2 \sigma_i^2}{\left[\sum (x_i - \bar{x})^2\right]^2}. \tag{8.3}$$

未知の σ_i^2 を $\hat{u}_i^2$ で置き換えて平方根をとったものが、**不均一分散に頑健な標準誤差**（White の標準誤差、Eicker–Huber–White の標準誤差）です。

計算例：5人のデータ（参考）

$\sum (x_i - \bar{x})^2 \hat{u}_i^2 = 4(0.64) + 1(0.36) + 0(1) + 1(0.36) + 4(0.04) = 3.44$。頑健な標準誤差 $= \sqrt{3.44}/10 \approx 0.185$。ただし $N = 5$ ではどちらの標準誤差も当てになりません。計算の練習としてのみ見てください。頑健な標準誤差は N が大きいときに正しく働きます。

コードで確かめる：Stata で確かめる

次の手順で、5人のデータの結果を確かめられます。1行ずつ「コマンドウィンドウ」に入力して実行します。

```
input x y
1 2
2 4
3 5
4 4
5 5
end
regress y x
regress y x, vce(robust)
```

- `input ... end`：データを手入力するコマンドです。`end` で入力を終えます。
- `regress y x`：y を x に回帰します。係数 0.6、定数項 2.2、標準誤差 0.283 が出れば本文と一致です。
- `vce(robust)`：頑健な標準誤差です。Stata は小標本の補正（$N/(N-K)$ 倍）を掛けるため約 0.239 になり、本文の 0.185 とは異なります。N が大きければほぼ一致します。

8.4 OLSは最良か：ガウス＝マルコフ定理（BLUE）

OLS が不偏であることは分かりましたが、不偏な推定量は他にもたくさんあります。その中で OLS はどういう意味で優れているのでしょうか。答えが**ガウス＝マルコフ定理**です。

BLUE の3つの言葉　BLUE は Best Linear Unbiased Estimator の頭文字です。

- **Linear（線形）**：推定量が $y_1, \dots, y_N$ の1次結合 $\tilde{\beta}_2 = \sum_i c_i y_i$ で書けること。ここで c_i は $x_1, \dots, x_N$ だけに依存する定数です。OLS は $\hat{\beta}_2 = \sum_i w_i y_i$ なので線形です。
- **Unbiased（不偏）**：$\mathrm{E}(\tilde{\beta}_2 \mid x) = \beta_2$ が、β_1, β_2 の値によらず成り立つこと。
- **Best（最良）**：その線形不偏推定量の中で、分散が最小であること。

要点：定理

MLR.1–5（正規性は不要）のもとで、$\hat{\beta}_2$ の条件付き分散は、線形不偏推定量の中で最小です。

証明（単回帰）　線形推定量 $\tilde{\beta}_2 = \sum c_i y_i$ が不偏になる条件から求めます。$y_i = \beta_1 + \beta_2 x_i + u_i$ を代入し、$\mathrm{E}(u_i \mid x) = 0$ を使うと

$$\mathrm{E}(\tilde{\beta}_2 \mid x) = \beta_1 \sum c_i + \beta_2 \sum c_i x_i .$$

これが任意の β_1, β_2 で β_2 に等しいためには、次の2条件が必要十分です。

$$\sum c_i = 0, \qquad \sum c_i x_i = 1. \tag{8.4}$$

OLS の重み w_i は性質(a)(b)よりこの条件を満たします。そこで、任意の不偏な線形推定量の重みを $c_i = w_i + d_i$ と書きます。式(8.4)と(a)(b)から

$$\sum d_i = 0, \qquad \sum d_i x_i = 0. \tag{8.5}$$

分散を計算します。u_i は無相関で分散が σ^2 なので、第1節と同じく

$$\begin{aligned}
\mathrm{Var}(\tilde{\beta}_2 \mid x) &= \sigma^2 \sum c_i^2 && (\textstyle\sum c_i u_i \text{ の分散}) \\
&= \sigma^2 \sum (w_i + d_i)^2 && (c_i = w_i + d_i) \\
&= \sigma^2 \Big[\sum w_i^2 + 2 \sum w_i d_i + \sum d_i^2 \Big] && (\text{展開})
\end{aligned}$$

中央の項は、$w_i = (x_i - \bar{x})/S_{xx}$ と式(8.5)より

$$\sum w_i d_i = \frac{\sum x_i d_i - \bar{x} \sum d_i}{S_{xx}} = \frac{0 - \bar{x} \cdot 0}{S_{xx}} = 0.$$

したがって

$$\mathrm{Var}(\tilde{\beta}_2 \mid x) = \underbrace{\frac{\sigma^2}{S_{xx}}}_{=\mathrm{Var}(\hat{\beta}_2 \mid x)} + \sigma^2 \sum d_i^2 \;\geq\; \mathrm{Var}(\hat{\beta}_2 \mid x). \tag{8.6}$$

等号は全ての $d_i = 0$、つまり $\tilde{\beta}_2 = \hat{\beta}_2$ のときだけです。 ■

計算例：5人のデータで確かめる

$x = (1, 2, 3, 4, 5)$ で、「最初と最後の2人だけを使う」推定量 $\tilde{\beta}_2 = (y_5 - y_1)/(x_5 - x_1)$ を考えます。重みは $c = (-\frac{1}{4}, 0, 0, 0, \frac{1}{4})$。$\sum c_i = 0$、$\sum c_i x_i = \frac{5-1}{4} = 1$ なので不偏です。

- このデータでの値は $(5 - 2)/4 = 0.75$（OLS は 0.6）。
- $\sum c_i^2 = \frac{1}{16} + \frac{1}{16} = 0.125$。OLS は $\sum w_i^2 = 1/S_{xx} = 0.1$。

分散は $0.125\sigma^2$ 対 $0.1\sigma^2$ で OLS の方が小さく、定理のとおりです。差の $0.025 = \sum d_i^2$ は、使わなかった3人分の情報を捨てた損失です。

補足：BLUE を読むときの注意

- 「最良」は**線形かつ不偏**な推定量の中での話です。偏りを許せば、分散がもっと小さい推定量が存在しえます（リッジ回帰など）。また非線形な推定量は比べていません。
- 均一分散（MLR.5）が証明の中で使われています。不均一分散のもとでは $\mathrm{Var}(\sum c_i u_i \mid x) = \sum c_i^2 \sigma_i^2$ となり、上の展開は成り立ちません。このとき最良の線形不偏推定量は、分散の逆数で重み付けした**加重最小二乗法**（GLS の一種）です。
- 誤差が正規分布に従う（MLR.6）なら、「線形」という限定を外した全ての不偏推定量の中でも OLS が最小分散になることが知られています。

発展：行列で書いた頑健な分散（サンドイッチ型）

大学院では最初から不均一分散を許すので、均一分散の公式は使いません。頑健な分散推定量は行列で $\hat{V} = \hat{A}^{-1}\hat{B}\hat{A}^{-1}$（$\hat{A} = \frac{1}{N}\sum \mathbf{x}_i^\top \mathbf{x}_i$、$\hat{B} = \frac{1}{N}\sum \hat{u}_i^2 \mathbf{x}_i^\top \mathbf{x}_i$）と書かれ、**サンドイッチ型**と呼ばれます。上の式の分子 $\sum (x_i - \bar{x})^2 \hat{u}_i^2$ が $\hat{B}$ に、分母の2乗が $\hat{A}^{-1}$ で前後からはさむ形に対応します。

要点

- 均一分散のもとで $\mathrm{Var}(\hat{\beta}_2 \mid x) = \sigma^2 / \sum (x_i - \bar{x})^2$。$\sigma^2$ が小さく、N や x のばらつきが大きいほど精密です。
- 標準誤差は $\mathrm{se}(\hat{\beta}_2) = \sqrt{\hat{\sigma}^2 / \sum (x_i - \bar{x})^2}$、$\hat{\sigma}^2$ は $N - 2$ で割ります。
- 不均一分散のもとでは頑健な標準誤差を使います。

次の章へ　推定値と標準誤差がそろいました。これらを使って「$\beta_2 = 0$ か」を判断する仮説検定に進みます。

大学院とのつながり　大学院では「heteroscedasticity robust standard error」を標準として使います。

第9章

仮説検定と信頼区間

> **ここまでの流れ**　推定値 $\hat{\beta}_2$ と、そのぶれの大きさ $\mathrm{se}(\hat{\beta}_2)$ が計算できるようになりました。
> **この章の問い**　「教育は賃金に影響しない（$\beta_2 = 0$）」かどうかを、データでどう判断するのでしょうか。
> **この章で分かること**　t検定、p値、信頼区間の作り方と読み方です。

9.1　検定の考え方

推定値 $\hat{\beta}_2 = 0.6$ は0ではありませんが、推定値は標本によってぶれるので、真の β_2 が0でも偶然0.6が出たのかもしれません。「偶然とは言いにくいほど0から離れているか」を判断するのが仮説検定です。

1. **帰無仮説** $H_0 : \beta_2 = 0$ と**対立仮説** $H_1 : \beta_2 \neq 0$ を立てます。
2. H_0 が正しいと仮定したときに分布が分かる**検定統計量**を作ります。
3. その分布のもとで「めったに出ない値」の範囲（**棄却域**）を決めます。めったに出ない確率の基準を**有意水準**（5%など）といいます。
4. 実際の値が棄却域に入れば H_0 を棄却します。入らなければ棄却しません。

補足：つまずきポイント

「棄却しない」は「H_0 が正しいと証明された」という意味ではありません。「H_0 を否定するだけの証拠がなかった」というだけです。

9.2　t検定

$$t = \frac{\hat{\beta}_2 - 0}{\mathrm{se}(\hat{\beta}_2)}. \tag{9.1}$$

t 統計量は、「推定値が0から標準誤差何個分離れているか」です。$H_0 : \beta_2 = b$ の一般形なら分子は $\hat{\beta}_2 - b$ になります。

要点：t がどんな分布に従うか

- **学部（MLR.1–6、小標本でも正確）**：H_0 のもとで t は自由度 $N-K$ の t分布に正確に従います（K は定数項を含む係数の数）。
- **大学院（大標本・漸近理論）**：MLR.5, 6 なしでも、N が大きければ t は近似的に Normal$(0,1)$ に従います（中心極限定理）。有意水準5%なら $|t|>1.96$ で棄却します。

t分布は自由度が大きくなると標準正規分布に近づくので（自由度30で5%臨界値は約2.04、自由度120で約1.98）、N が大きい実証研究では両者の違いは小さくなります。

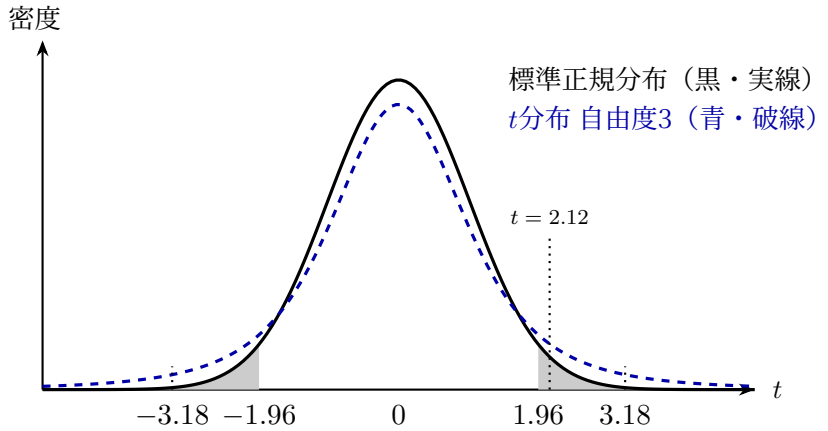
計算例：5人のデータ

$t=0.6/0.283\approx 2.12$。

- 正規近似を使うと $2.12>1.96$ で棄却となります。
- しかし $N=5$ は小さすぎて正規近似は信頼できません。正確な t分布（自由度 $5-2=3$）の5%臨界値は約3.18なので、H_0 は棄却されません。

漸近理論は「N が十分大きいとき」の近似である、という注意点を示す例です。

図 9.1　標準正規分布と自由度3の t分布



曲線は標準正規分布の密度 $\phi(t)=e^{-t^2/2}/\sqrt{2\pi}$ と、自由度3の t分布の密度 $f(t)=\frac{2}{\pi\sqrt{3}}(1+t^2/3)^{-2}$。灰色の部分は標準正規分布の両側2.5%ずつの棄却域（$|t|>1.96$）。

図の読み方

1. 黒い実線が標準正規分布、青い破線が自由度3の t分布です。t分布は裾が厚く、極端な値が出やすいことが分かります。
2. 灰色の部分は標準正規分布の棄却域（両側2.5%ずつ）で、境目が ± 1.96 です。t分布の5%臨界値は ± 3.18 と、ずっと外側になります。
3. $t = 2.12$ は正規分布なら棄却域の中、t分布なら棄却域の外です。サンプルサイズが小さいとき、正規近似で判断すると棄却しすぎてしまいます。

9.3 p値

H_0 が正しいとき、実際に得た値以上に極端な $|t|$ が出る確率を**p値**といいます。式で書くと、観測した統計量を t_{obs} として $p = \Pr(|T| \geq |t_{\text{obs}}| \mid H_0)$（両側）です。$p$値が有意水準（例：0.05）より小さければ棄却、と判断しても、$|t|$ と臨界値を比べるのと同じ結論になります。論文の回帰表で「***」などの記号がついているのは、p値の大きさを示しています。

計算例：5人のデータ

$t = 2.12$ に対する自由度3の t分布の両側 p値は約0.124で、0.05より大きいので棄却されません。正規近似なら約0.034で棄却されます。

9.4 信頼区間

95%信頼区間（大標本）は次のとおりです。

$$\left[\hat{\beta}_2 - 1.96\,\mathrm{se}(\hat{\beta}_2),\ \hat{\beta}_2 + 1.96\,\mathrm{se}(\hat{\beta}_2)\right]. \tag{9.2}$$

意味：この方法で区間を作ることを何度もくり返すと、そのうち95%の区間が真の β_2 を含みます。

補足：つまずきポイント

「1回作った区間 $[a, b]$ に、真の β_2 が95%の確率で入っている」という読み方は、厳密には正しくありません。真の β_2 は定数なので、入っているか入っていないかのどちらかです。95%は「区間の作り方」の性質です。

検定との関係：5%水準の t検定で $H_0 : \beta_2 = b$ が棄却されない $\iff$ b が95%信頼区間の中にあります。信頼区間は「棄却されない値の集まり」です。

9.5 複数の制約の検定：F検定

「$\beta_2 = \beta_3 = 0$（教育も経験も影響しない）」のように、複数の制約を同時に検定したいことがあります。

t検定を繰り返してはいけない理由

- 各 t検定を5%水準で行って「どちらか一方でも棄却されたら棄却」とすると、H_0 が正しくても誤って棄却する確率は5%を超えます（2回なら、独立な場合でも $1 - 0.95^2 = 9.75\%$）。
- 2つの推定量 $\hat{\beta}_2, \hat{\beta}_3$ は一般に相関しているので、1つずつの検定ではその相関を無視してしまいます。

考え方：制約を課すと当てはまりがどれだけ悪化するか　モデルを2通りに推定します。

- **制約なしモデル**（UR）：全ての変数を入れて OLS。残差2乗和を SSR_{UR} とします。
- **制約つきモデル**（R）：H_0 の制約を課して OLS。例えば $\beta_2 = \beta_3 = 0$ なら、x_2, x_3 を除いたモデル。残差2乗和を SSR_R とします。

制約を課すと選べる直線の範囲が狭まるので、必ず $\mathrm{SSR}_R \geq \mathrm{SSR}_{UR}$ です。H_0 が正しければ悪化は小さく、誤っていれば大きいはずです。その悪化の大きさを、σ^2 の推定値で割って測ります。

$$F = \frac{(\mathrm{SSR}_R - \mathrm{SSR}_{UR})/q}{\mathrm{SSR}_{UR}/(N-K)}. \tag{9.3}$$

数式の読み方：式(9.3)の読み方

- q：制約の本数（例：$\beta_2 = \beta_3 = 0$ なら $q = 2$）。K：制約なしモデルの係数の数（定数項を含む）。
- 分子：制約1本あたりの SSR の増加。分母：制約なしモデルの $\hat{\sigma}^2$（第8章。$N - K$ で割る不偏推定量）。
- F は「制約によるSSRの悪化」を「誤差のばらつき」と比べた比です。F が大きいほど H_0 は疑わしくなります。
- 検定は片側です。$F > F_\alpha(q, N-K)$（臨界値）なら棄却します。

なぜ F分布になるのか　MLR.1–6（誤差が正規分布）のもとで、H_0 が正しければ

$$\frac{\mathrm{SSR}_R - \mathrm{SSR}_{UR}}{\sigma^2} \sim \chi^2(q), \qquad \frac{\mathrm{SSR}_{UR}}{\sigma^2} \sim \chi^2(N-K),$$

で、この2つは互いに独立です。式(9.3)は、これらをそれぞれの自由度で割った比なので、σ^2 が約分されて自由度 $(q,\, N-K)$ の F**分布**に従います。F分布とは、「独立な2つの χ^2 変数を、それぞれの自由度で割った比」の分布です。

決定係数による書き方　$\mathrm{SSR} = (1-R^2)\,\mathrm{SST}$ で、同じ被説明変数なら SST は共通なので

$$F = \frac{(R^2_{UR} - R^2_R)/q}{(1-R^2_{UR})/(N-K)}. \tag{9.4}$$

特に「定数項以外の全ての係数が0」という**回帰全体の有意性**の検定では、制約つきモデルは定数項だけなので $R^2_R = 0$、$q = K-1$ で

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}.$$

統計ソフトの回帰結果の表に出る「F値」は、この検定です。

制約が1本のとき：$F = t^2$　$q=1$ のとき、F は t 統計量の2乗に等しくなります。第3章の5人のデータで $H_0: \beta_2 = 0$ を確かめます。

計算例：5人のデータ

- 制約つきモデル（$\beta_2 = 0$、つまり $y = \beta_1 + u$）：$\hat\beta_1 = \bar y = 4$ で $\mathrm{SSR}_R = \mathrm{SST} = 6$。
- 制約なしモデル：$\mathrm{SSR}_{UR} = 2.4$、$K=2$、$N-K=3$。
- $F = \dfrac{(6-2.4)/1}{2.4/3} = \dfrac{3.6}{0.8} = 4.5$。
- 一方 $t = 2.12$ の2乗は $2.1213^2 = 4.5$ で一致します。
- 臨界値：$F_{0.05}(1,3) \approx 10.13$。これは自由度3の t分布の臨界値 3.18 の2乗 $3.182^2 \approx 10.13$ と一致するので、t検定と同じく H_0 は棄却されません。

補足：F検定と t検定の関係

$q = 1$ の両側 t検定と F検定は常に同じ結論になります。$q \geq 2$ のときだけ、F検定でないと「同時に」検定できません。F の分子 $\mathrm{SSR}_R - \mathrm{SSR}_{UR}$ は、H_0 からのずれを推定値の分散・共分散で調整した2次形式になっており、これが推定値どうしの相関を正しく反映しています。

発展：大学院では Wald検定

大学院では誤差の正規性を仮定せず、大標本の漸近理論を使います。F検定に対応するのが **Wald検定**で、その統計量 W は「制約からのずれを、その分散で割って2乗したもの」の多次元版です。H_0 のもとで W は近似的に $\chi^2(q)$ に従い、不均一分散のもとでも頑健な分散推定量を使って計算できます。大標本では $qF \approx W$ で、制約が1本なら $W = t^2$ です。

9.6 回帰表の読み方

実際の分析では、統計ソフトが次のような表を出力します。第3章の5人のデータ（$N = 5$）の結果を、ここまで学んだ言葉で読んでみましょう。

	係数	標準誤差	t 値	p 値	95%信頼区間
x（傾き）	0.600	0.283	2.12	0.124	$[-0.300,\ 1.500]$
定数項	2.200	0.938	2.35	0.101	$[-0.785,\ 5.185]$

$N = 5$、$R^2 = 0.6$。信頼区間と p 値は自由度3の t分布で計算しています。定数項の標準誤差は $\sqrt{\hat{\sigma}^2(\frac{1}{N} + \frac{\bar{x}^2}{S_{xx}})} = \sqrt{0.8(0.2 + 0.9)} \approx 0.938$ です（導出は省略）。

数式の読み方：各列の意味

- **係数**（第3章）：$\hat{\beta}_2 = 0.6$。「x が1増えると、y の平均が約0.6増える」。
- **標準誤差**（第8章）：0.283。標本が変わると係数が ± 0.28 程度ぶれる。
- **t 値**（本章）：$0.6/0.283 = 2.12$。係数が0から標準誤差何個分離れているか。
- **p 値**：0.124。$H_0 : \beta_2 = 0$ が正しいとして、これほど極端な t が出る確率。5%より大きいので棄却できない。
- **信頼区間**：$0.6 \pm 3.18 \times 0.283 = [-0.300,\ 1.500]$。0を含むので、$H_0 : \beta_2 = 0$ は5%水準で棄却されない（p 値と同じ結論）。
- R^2（第4章）：y のばらつきの60%を x で説明。

補足：結果の言い方

この表から言えるのは、「傾きの推定値は0.6だが、サンプルが5人と少ないため標準誤差が大きく、真の傾きが0である可能性を排除できない」という慎重な結論です。「教育は賃金に影響しない」と結論づけるのではなく、「影響がないとは言い切れないが、あるとも言い切れない」とするのが正確です。

要点：回帰結果を読む手順

1. 係数の符号と大きさが、経済学的に妥当か（単位に注意）。
2. 標準誤差と比べて、係数が十分大きいか（$|t|$ が1.96前後を超えるか）。
3. 信頼区間が、意味のある大きさの効果を含んでいるか。
4. R^2 は参考程度に見る。
5. 識別の仮定（$\mathrm{E}(u \mid x) = 0$）が現実的か、欠落変数がないかを考える。

要点

- $t = (\hat{\beta}_2 - b)/\mathrm{se}(\hat{\beta}_2)$ で、大標本では $|t| > 1.96$ なら5%水準で棄却します。
- 小標本の学部の理論では自由度 $N - K$ の t分布を使い、臨界値は1.96より大きくなります。
- 95%信頼区間は $\hat{\beta}_2 \pm 1.96\,\mathrm{se}(\hat{\beta}_2)$ で、「棄却されない値の集まり」です。
- 複数の制約は $F = \dfrac{(\mathrm{SSR}_R - \mathrm{SSR}_{UR})/q}{\mathrm{SSR}_{UR}/(N-K)}$ で同時に検定します。制約が1本なら $F = t^2$ です。
- 「棄却しない」は「H_0 が正しい」ではありません。

次の章へ　推定・検定の道具がそろいました。最後に、実証分析でよく使うダミー変数と対数の解釈を整理します。

大学院とのつながり　大学院の漸近理論では t の近似分布が正規分布、Wald統計量の近似分布が χ^2 分布になります。

第10章

ダミー変数と対数

ここまでの流れ　推定から検定まで、回帰分析の基本の流れが一通りそろいました。
この章の問い　「大卒か否か」のような質的な変数や、対数をとった変数の係数は、どう読むのでしょうか。
この章で分かること　ダミー変数の係数は「2グループの平均の差」、対数モデルの係数は「変化率」を表すことです。

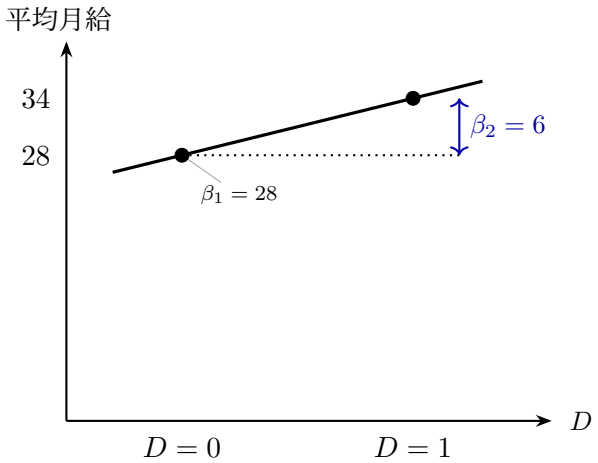
10.1　ダミー変数

$y = \beta_1 + \beta_2 D + u$（$D \in \{0, 1\}$）なら

$$\mathrm{E}(y \mid D = 0) = \beta_1, \qquad \mathrm{E}(y \mid D = 1) = \beta_1 + \beta_2 \implies \beta_2 = \mathrm{E}(y \mid D = 1) - \mathrm{E}(y \mid D = 0). \tag{10.1}$$

ダミー変数の係数は、2つのグループの平均の差です。第2章の仮想の母集団では $6 = 34 - 28$（大卒と非大卒の平均月給の差）でした。

図 10.1　ダミー変数の回帰：係数は2グループの平均の差



第2章の仮想の母集団の値（非大卒の平均月給28、大卒の平均月給34、単位は万円）を使い、直線 $28 + 6D$ を描いた。

図の読み方

1. 2つの黒い点は各グループの平均です。回帰直線は、この2点を結ぶ直線になります。
2. 切片は $D = 0$ の平均（28）、傾きは2点の縦の差（6）です。
3. D は0か1の値しかとらないので、直線の「傾き」は実質的に2グループの差を表しています。

10.2 対数を使ったモデル

モデル	β_2 の解釈	例
$y = \beta_1 + \beta_2 x$	x が1単位増→ y が β_2 単位変化	
$\log y = \beta_1 + \beta_2 x$	x が1単位増→ y がおよそ $100\beta_2\%$ 変化	教育1年で賃金が8%上昇
$\log y = \beta_1 + \beta_2 \log x$	x が1%増→ y がおよそ $\beta_2\%$ 変化（弾力性）	価格弾力性

理由：$\log y$ の変化は、変化が小さいとき y の変化率にほぼ等しいからです（$\log(1.08) \approx 0.077$）。

数式の読み方：弾力性との対応

$\dfrac{\partial \mathrm{E}[y \mid x]}{\partial x_j} \cdot \dfrac{x_j}{\mathrm{E}[y \mid x]} = \dfrac{\partial \log \mathrm{E}[y \mid x]}{\partial \log x_j}$ は「x_j が1%増えたとき $\mathrm{E}(y \mid x)$ が何%変わるか」を表します。両対数モデルの係数がおよそこれに当たります。

補足：注意

変化が大きいときは近似が粗くなります。例えば $\log y = \beta_1 + \beta_2 x$ で $\beta_2 = 0.08$ なら、正確な変化率は $e^{0.08} - 1 \approx 8.3\%$ で、「およそ8%」とほぼ同じです。また、厳密には $\mathrm{E}(\log y \mid x)$ と $\log \mathrm{E}(y \mid x)$ は異なりますが、学部の範囲では同一視されます。

要点

- ダミー変数の係数は、2グループの平均の差です。
- $\log y$ を被説明変数にすると、係数は「x が1単位増えたときの y のおよその変化率」です。
- 両対数モデルの係数は弾力性（x が1%増えたときの y の変化率）に当たります。

次の章へ　全体のまとめとして、学部と大学院のちがいを整理し、練習問題で理解を確かめます。

終章

まとめと練習問題

ここまでの流れ　モデル・識別・OLS・不偏性・標準誤差・検定と、回帰分析の基本を一通り学びました。
この章の問い　学部の内容と大学院の内容はどう違い、理解はできているでしょうか。
この章で分かること　学部と大学院のちがいの整理と、9問の練習問題（解答つき）です。

終.1　用語・記号の整理

用語・記号	意味
条件付き期待値 $\mathrm{E}(y \mid x)$	x が決まった人たちの y の平均（第1章）
誤差項 u	y と $\mathrm{E}(y \mid x)$ のずれ。観測できない（第1章）
識別	母集団全体が分かれば β がただ1つに決まること（第2章）
OLS（最小二乗法）	残差の2乗和を最小にする推定法（第3章）
残差 $\hat{u}_i$	$y_i - \hat{y}_i$。誤差の推定値で観測できる（第4章）
R^2	y のばらつきのうち x で説明できる割合（第4章）
部分効果	他を一定にしたときの x_j の効果（第5章）
欠落変数バイアス	入れ忘れた変数のせいで生じる偏り（第6章）
内生変数	誤差項と相関している説明変数（第6章）
不偏性	推定量の平均が真の値に一致すること（第7章）
標準誤差	推定量の標準偏差の推定値（第8章）
BLUE	線形不偏推定量の中で分散が最小（第8章）
t 統計量・p 値	係数が0から離れている度合い、極端さの確率（第9章）
F 検定	複数の制約を同時に検定する（第9章）
信頼区間	「棄却されない値の集まり」（第9章）

終.2 学部と大学院のちがい

	学部	大学院
表記	和の記号 $\sum$ 中心、単回帰が多い	行列 $X^\top X$、$\mathrm{E}(\mathbf{x}_i^\top \mathbf{x}_i)$
誤差の条件	$\mathrm{E}(u \mid x) = 0$	直交 $\mathrm{E}(\mathbf{x}^\top u) = 0$ と外生性 $\mathrm{E}(u \mid \mathbf{x}) = 0$ を区別
推定対象	条件付き期待値	線形射影 $L(y \mid \mathbf{x})$ も扱う
分散の仮定	均一分散	不均一分散を許す（頑健な標準誤差）
分布	正規誤差→正確な t, F 分布	$N \to \infty$ の漸近理論→正規、χ^2
主な性質	不偏性、BLUE	一致性、漸近正規性

右の列の「一致性」「漸近正規性」を理解するための道具が、大数の法則と中心極限定理などの漸近理論です。

終.3 練習問題

1. $x = (0, 1, 2)$、$y = (1, 3, 2)$ のとき、$\hat{\beta}_1, \hat{\beta}_2$、残差、$R^2$ を求めなさい。また、残差の2つの性質が成り立つことを確かめなさい。

2. 単回帰で $\sum \hat{u}_i = 0$ を、$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}$ から示しなさい。

3. 欠落変数バイアスの式で、$\gamma < 0$ かつ $\mathrm{Cov}(x, q) > 0$ のとき、OLS は β_2 を過大評価するか過小評価するか。具体的な経済の例を1つ考えなさい。

4. ある回帰で $\hat{\beta}_2 = 0.08$、$\mathrm{se}(\hat{\beta}_2) = 0.02$（$N = 2000$）。$H_0 : \beta_2 = 0$ の5%水準の t検定の結論と、95%信頼区間を答えなさい。$H_0 : \beta_2 = 0.1$ は棄却されるか。

5. 「x が定数項を含むとき、OLS 残差の平均は必ず0」を、正規方程式 $X^\top \hat{U} = 0$ の1行目から説明しなさい。

6. サンプルサイズを100から400に増やすと、$\mathrm{se}(\hat{\beta}_2)$ はおよそ何倍になるか。

7. 単回帰の不偏性の証明で、MLR.4（$\mathrm{E}(u \mid x) = 0$）はどのステップで使われたか。

8. $x = (0, 1, 2)$ のとき、「最初と最後の2点だけを使う」推定量 $\tilde{\beta}_2 = (y_3 - y_1)/(x_3 - x_1)$ が線形不偏であることを示し、その条件付き分散を OLS の分散と比べなさい（均一分散 σ^2 とする）。

9. 問1のデータ（$N = 3$）で $H_0 : \beta_2 = 0$ を F検定の式で計算し、t^2 と一致することを確かめなさい。

終.4 解答

問1　$\bar{x}=1$、$\bar{y}=2$。$x-\bar{x}=(-1,0,1)$、$y-\bar{y}=(-1,1,0)$。$\sum(x-\bar{x})(y-\bar{y})=1+0+0=1$、$\sum(x-\bar{x})^2=2$。よって $\hat{\beta}_2=0.5$、$\hat{\beta}_1=2-0.5\times 1=1.5$。$\hat{y}=(1.5,2,2.5)$、$\hat{u}=(-0.5,1,-0.5)$。$\mathrm{SSE}=0.25+0+0.25=0.5$、$\mathrm{SST}=1+1+0=2$ より $R^2=0.25$。残差の性質：$\sum\hat{u}_i=-0.5+1-0.5=0$、$\sum x_i\hat{u}_i=0(-0.5)+1(1)+2(-0.5)=0$。成り立っています。

問2

$$\sum\hat{u}_i=\sum(y_i-\hat{\beta}_1-\hat{\beta}_2x_i)=N\bar{y}-N\hat{\beta}_1-N\hat{\beta}_2\bar{x}=N(\bar{y}-\hat{\beta}_1-\hat{\beta}_2\bar{x}).$$

$\hat{\beta}_1=\bar{y}-\hat{\beta}_2\bar{x}$ を代入すると括弧の中は0です。

問3　バイアス $\gamma\mathrm{Cov}(x,q)/\mathrm{Var}(x)<0$ なので過小評価です。例は第6章の「肥料と土地のやせ具合」を参照してください。

問4　$t=0.08/0.02=4>1.96$ なので棄却します（β_2 は5%水準で有意に0と異なる）。95%信頼区間は $0.08\pm 1.96\times 0.02=[0.0408,\,0.1192]$。0.1 は信頼区間の中にあるので、$H_0:\beta_2=0.1$ は棄却されません（実際 $t=(0.08-0.1)/0.02=-1$、$|t|<1.96$）。

問5　$X^{\top}\hat{U}$ の1行目は「X の1列目（すべて1）と $\hat{U}$ の内積」$=\sum_i 1\cdot\hat{u}_i$ です。これが0なので残差の和、したがって平均は0です。定数項を含まない回帰では、この保証はありません。

問6　分散はおよそ $\sigma^2/(N\mathrm{Var}(x))$ なので、N が4倍になると分散は1/4、標準誤差は $\sqrt{1/4}=1/2$ 倍です。

問7　ステップ4で $\mathrm{E}(u_i\mid x_1,\ldots,x_N)=\mathrm{E}(u_i\mid x_i)=0$ とした最後の等号です（その前の等号では MLR.2 の無作為抽出を使っています）。

問8　重みは $c=(-\frac{1}{2},0,\frac{1}{2})$。$\sum c_i=0$、$\sum c_ix_i=\frac{1}{2}\cdot 2=1$ なので不偏です（$\mathrm{E}(\tilde{\beta}_2\mid x)=\beta_1\cdot 0+\beta_2\cdot 1$）。分散は $\sigma^2\sum c_i^2=\sigma^2(\frac{1}{4}+\frac{1}{4})=0.5\sigma^2$。OLS は $\sigma^2/S_{xx}=\sigma^2/2=0.5\sigma^2$ で、この場合は両者が等しくなります（$x=(0,1,2)$ では中央の点が $\bar{x}$ と一致し、$d_i=0$ になるため。実際、問1の OLS の重みは $w=(-\frac{1}{2},0,\frac{1}{2})$）。

問9　問1より $\mathrm{SSR}_{UR} = \sum \hat{u}_i^2 = 0.25 + 1 + 0.25 = 1.5$、$\mathrm{SST} = 2$ なので $\mathrm{SSR}_R = 2$（$\beta_2 = 0$ のとき $\hat{\beta}_1 = \bar{y}$）。$q = 1$、$N - K = 1$ より $F = \dfrac{(2 - 1.5)/1}{1.5/1} = \dfrac{1}{3}$。一方 $\hat{\sigma}^2 = 1.5/1 = 1.5$、$\mathrm{se}(\hat{\beta}_2) = \sqrt{1.5/2} = \sqrt{0.75}$、$t = 0.5/\sqrt{0.75}$ で $t^2 = 0.25/0.75 = \frac{1}{3}$ となり一致します。

要点

- 回帰分析は「モデル→識別→推定→精度・検定」の順に進み、OLS は識別結果の標本対応物であり残差2乗和の最小化でもあります。
- 重回帰の係数は「他を一定にした」部分効果で、欠落変数があると $\gamma \mathrm{Cov}(x, q)/\mathrm{Var}(x)$ だけ偏ります。
- $\mathrm{E}(u \mid x) = 0$ のもとで OLS は不偏で、精度は標準誤差、判断は t検定・信頼区間で行います。
- 大学院では、不均一分散を許す漸近理論（一致性・漸近正規性・頑健な標準誤差）に進みます。

次のステップ　$\hat{\beta} = \beta + \dfrac{\sum (x_i - \bar{x}) u_i}{\sum (x_i - \bar{x})^2}$ の分解に、大数の法則と中心極限定理を当てはめると、一致性と漸近正規性が得られます。これが漸近理論の出発点です。