最小 2 乗法

別所俊一郎

慶應義塾大学経済学部

2012 年 4 月

回帰分析とは

- ある変数 X が他の変数 Y に与える効果を分析する手法の一つ：$Y = f(X; Z)$
- 線形回帰モデル（linear regression model）：変数 X と変数 Y の線形の関係に注目。傾き（slope）は X の 1 単位の変化が与える Y の変化のことで、(X, Y) の同時分布の特性値の 1 つ
- 「真の」関係が線形であることはほとんどない
  - 変形したものの線形関数
  - 近似として有効
- (X, Y) の無作為標本があるときの統計的推測
  - 傾きの推定
  - 仮説検定
  - 信頼区間の形成

線形回帰モデル：例

- 問題の設定：「少人数クラスにすると教育の効果は高まるか？」
  - 「教師 1 人当たりの児童数を減らすと標準テストの点はどう変化するか？」
- 児童数の変化がテストの点数をどれほど変化させるか？

$$\beta_{\text{児童数}} = \frac{\text{テストの点数の変化}}{\text{クラスの児童数の変化}} \equiv \frac{\Delta\text{点数}}{\Delta\text{児童数}}$$

この $\beta_{\text{児童数}}$ がわかれば、点数の変化分を計算できる

$$\Delta\text{点数} = \beta_{\text{児童数}} \Delta\text{児童数} \tag{4.2}$$

- この式から線形関数を考えることができて、

$$\text{点数} = \beta_0 + \beta_{\text{児童数}} \times \text{児童数} \tag{4.3}$$

線形回帰モデル：例

- しかし、点数（の変化）は児童数のみでは決まらない
  - 児童の家庭環境などのバックグラウンド
  - 教師や教材の質
  - テスト当日に関する不確定要因
- だから、線形関数はあくまで「平均的な」関係を示すに過ぎない
- すべての学校区に当てはまる関係式は、これら「その他」の要因を取り入れる必要がある。たとえば

$$\text{点数} = \beta_0 + \beta_{\text{児童数}} \times \text{児童数} + \text{その他の要因} \tag{4.4}$$

- 関係が「線形」であるとは限らないが…

より一般的に、説明変数が 1 つの線形回帰モデルは

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, 2, ..., n \tag{4.5}$$

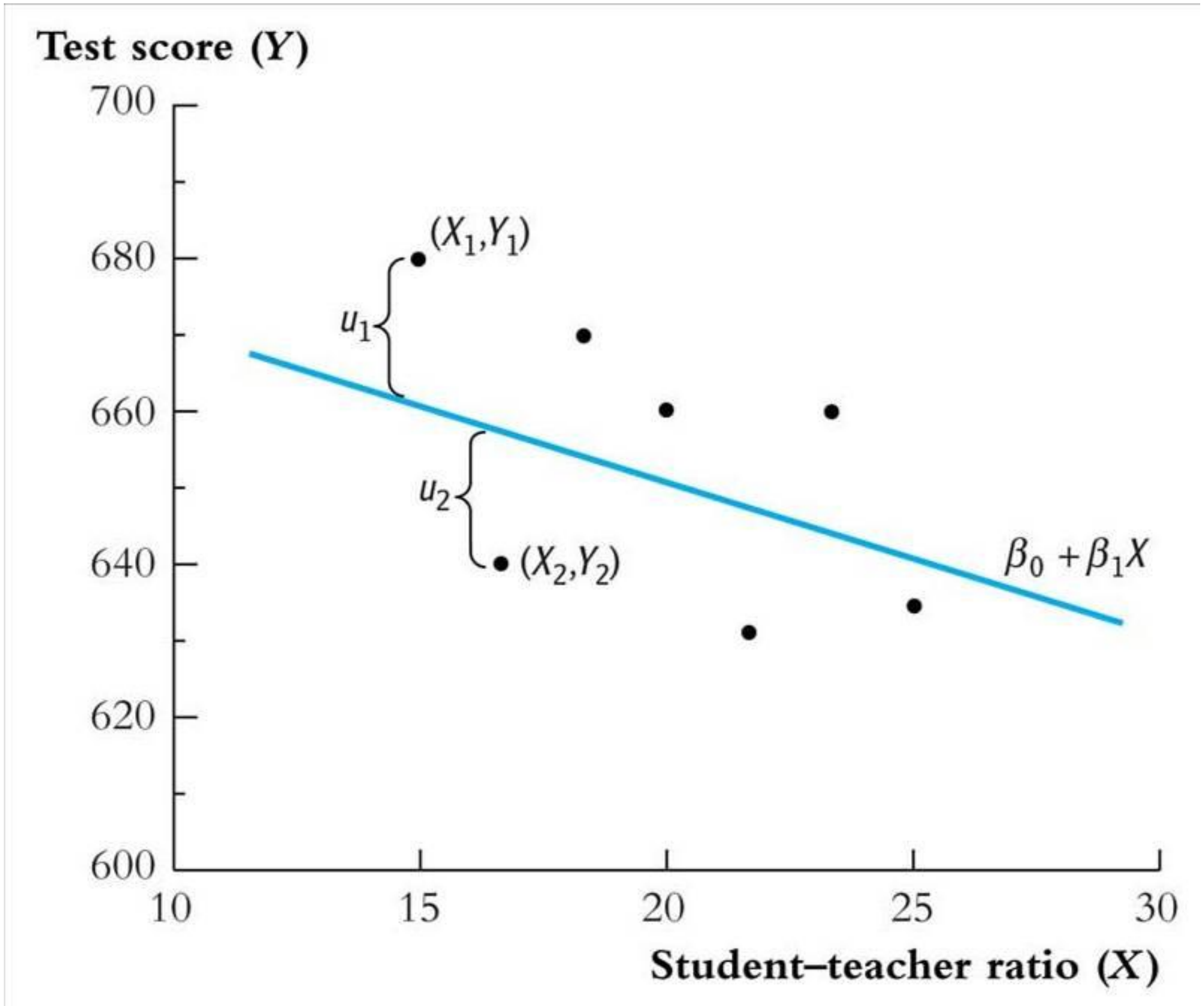
- Y_i ：被説明変数（dependent variable）
- X_i ：説明変数（explanatory variable, regressor）、独立変数（independent variable）。定数項も説明変数に含むこともある
- $\beta_0 + \beta_1 X$ ：回帰線（population regression line (function)）。所与の X に対応する平均的な Y の値を示す関数
- β_0 ：切片（intercept）。$X = 0$ のときの平均的な Y の値だが、経済的に意味のない数値の場合も。
- β_1 ：傾き（slope）。β_0 とあわせて、パラメタ（parameter）、係数（coefficient）とも呼ばれる。
- u_i ：誤差項（error term）。「その他の要因」を代表する確率変数。平均的な値（$\beta_0 + \beta_1 X_i$）と実現値（Y_i）の差を説明するもので、X_i 以外のすべての要因を含む。

線形回帰モデルにおける統計的推測

$$仮定：Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, 2, ..., n \tag{4.5}$$

- β_0 と β_1 の真の値がわかっていれば、X_i と u_i の実現値に応じて Y_i の値を計算できる（Figure 4.1）
- 手許にあるデータは (X_i, Y_i) の（無作為抽出）標本だけであり、ここから β_0 と β_1 を推測する
- もう 1 つの確率変数 u_i は実現値もわかっていない
- (X_i, Y_i, u_i) が線形の関係にあるかどうかもほんとうは定かではない
  - もし (X, Y, u) が上の式を満たす確率変数の組だとすれば，係数の値はいくつか？
- β_0 と β_1 の真の値を標本から統計的に推測するから、仮説検定や信頼区間の形成という手続きが可能
- β_0 と β_1 の真の値をどのように推測するのか？

Figure 4.1 Scatter plot of test score vs. student-teacher ratio (hypothetical data)



線形回帰モデルの係数の推定

係数 β_0 と β_1 の推定 未知の母平均を無作為抽出の標本平均から推定するのに似ている

標本統計量 Table 4.1.

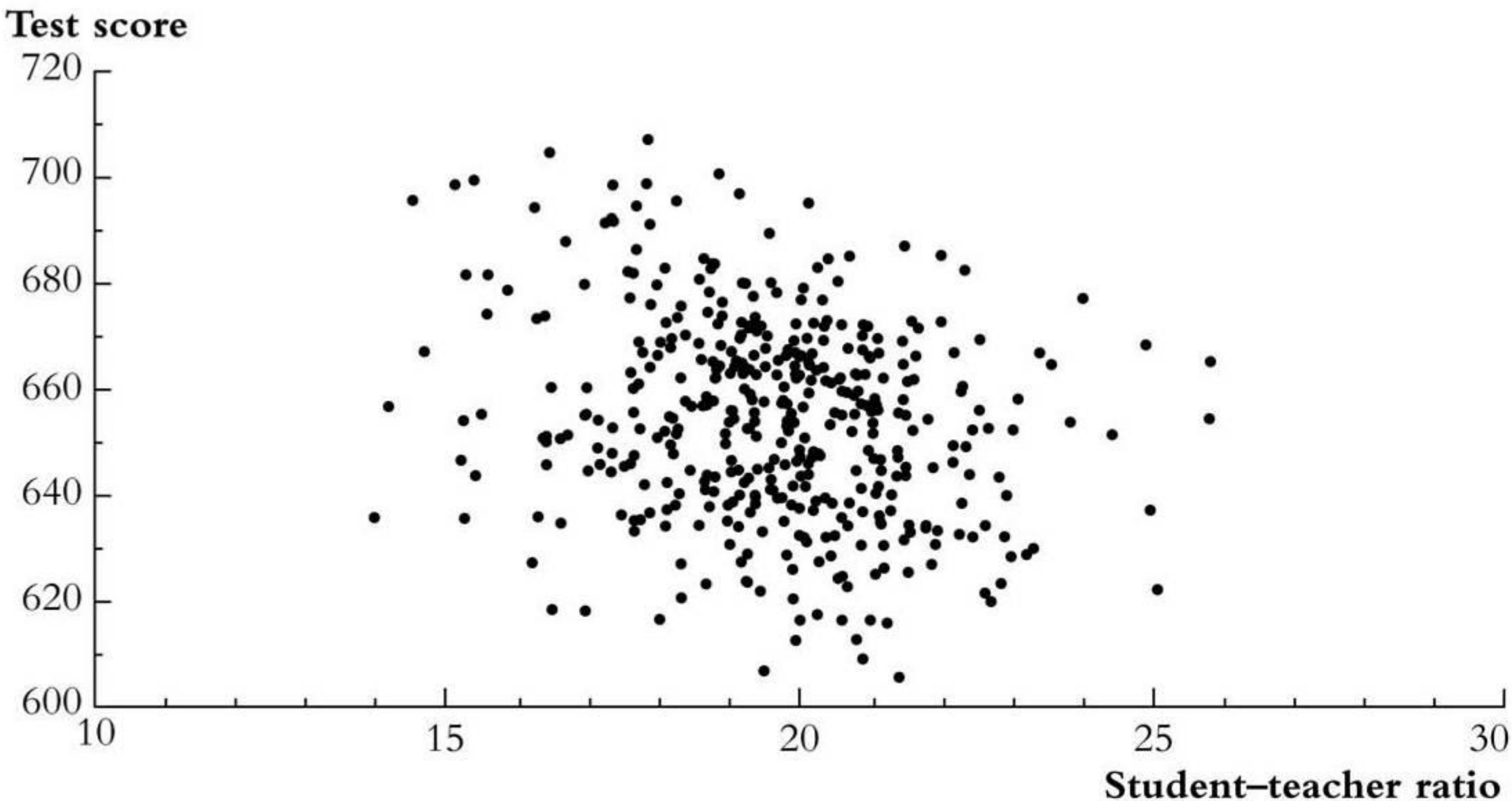
散布図 Figure 4.2. のばあい、

- 直線上に乗っているわけではない：他の要因 u_i の影響
- 標本平均は-0.23 だから、右下がりの回帰線を書くことができる？
- 回帰線の推定の方法は自由：母平均の推定と同じで、どのような直線を考えることもできるが…。
- もっともまっとうな直線の引き方は？

Table 4.1 Summary statistics of California school district data (Eviews output)

	TESTSCR	STR
Mean	654.1565	19.64043
Median	654.45	19.72321
Maximum	706.75	25.8
Minimum	605.55	14
Std. Dev.	19.05335	1.891812
Skewness	0.091615	-0.02537
Kurtosis	2.745712	3.609597
Jarque-Bera	1.719129	6.548185
Probability	0.423346	0.037851
Sum	274745.8	8248.979
Sum Sq. Dev.	152109.6	1499.581
Observations	420	420

Figure 4.2 Scatter plot of Test score vs. student-teacher ratio
(California school district data)



最小二乗推定量

もっとも有名な推定量の 1 つ

- データになるべく「近い」ような回帰線を描き、それを推定値とする
- 「近さ」は X が与えられたときの Y の予測値と実現値の差の 2 乗の和で測る
- $\overline{Y}$ が $E[Y]$ の最小二乗推定量であったことを思い出そう

$$\overline{Y} = \mathrm{argmin}_m \sum_{i=1}^{n} (Y_i - m)^2$$

これと同様に、二乗和で定義された「近さ」を最小化するような (β_0, β_1) の候補の組み合わせを探して、それを推定値とする

(β_0, β_1) の最小二乗推定量を (b_0, b_1) と書く。このとき、

- 回帰直線は $b_0 + b_1 X$ となる
- $X = X_i$ のときの Y の予測値は $b_0 + b_1 X_i$
- 実現値との差は $Y_i - (b_0 + b_1 X_i)$ だから、最小化すべき二乗和は、

$$\sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2 \tag{4.6}$$

  もし説明変数がなければ、母平均の最小二乗推定と同じ式になることに注意

- 母平均の最小二乗推定値が唯一に定まったのと同様、この最小化問題の解 (b_0, b_1) も一般に一意に定まる

最小二乗推定量は最小化問題の解として求まるから、微分してゼロとおけばよい。すなわち連立方程式

$$\frac{\partial}{\partial b_0}\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2 = 0$$

$$\frac{\partial}{\partial b_1}\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_i)^2 = 0$$

の解 (b_0, b_1) が求める値を与える。計算すると、

$$\hat{\beta}_1 = b_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} = \frac{\text{標本共分散}}{\text{説明変数の標本分散}}$$

$$\hat{\beta}_0 = b_0 = \overline{Y} - b_1\overline{X}$$

最小二乗推定の基礎用語

OLS 推定量　最小二乗推定量 Ordinary Least Squares estimators の略称。しばしば $(\hat{\beta}_0, \hat{\beta}_1)$ で表す

OLS regression line　OLS 推定量を用いて描かれる回帰直線のこと

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X$$

予測値、当てはめ値（fitted value）　所与の X_i に対する回帰直線上の点のことで、X_i に対する Y の平均的な値を与える.
一定の条件のもとでは，X_i で条件付けられた Y_i の期待値に対応.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \tag{4.9}$$

最小二乗推定の基礎用語

残差（residual） 各観測値の実現値 Y_i と当てはめ値 $\hat{Y}_i$ の差

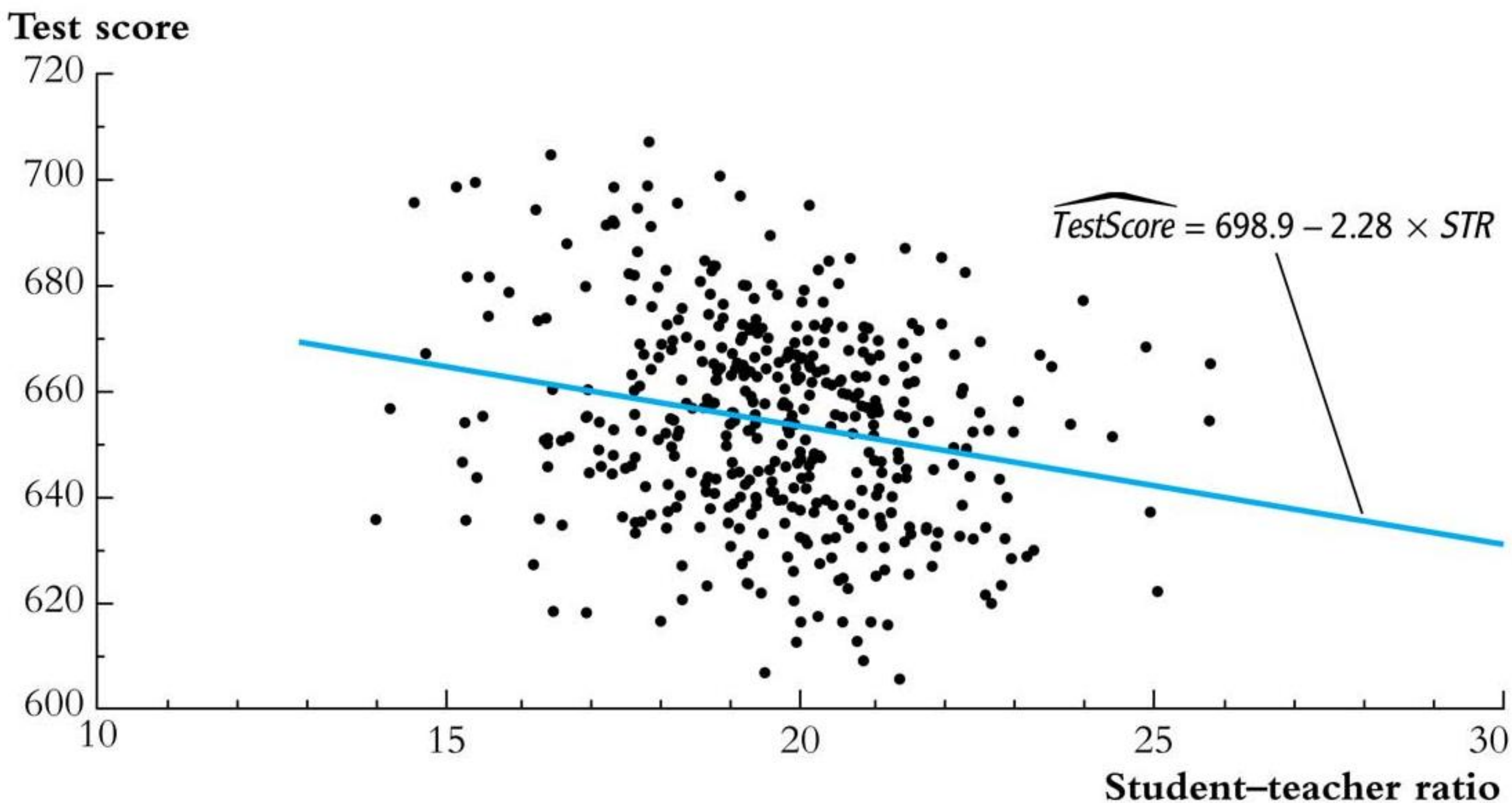
$$\hat{u}_i = Y_i - \hat{Y}_i \tag{4.10}$$

誤差（**error**）とは異なる概念であることに注意

OLS 推定量も確率変数 ：標本平均が確率変数であったのと同様.

- 同じ母集団であってもサンプルが異なれば OLS 推定 値 は異なる

Figure 4.3 The estimated regression line for the California data



なぜ最小二乗推定量なのか？

- 真の値 (β_0, β_1) の推測の方法はいくらでも考えられる
  - 母平均の推定量が標本平均に限らなかったのと同じ
  - 「予測値と実現値の差の絶対値の和」の最小化や、なんらかの加重平均でもよいはず
- OLS は望ましい性質を持っている
  - 理論的に、ある仮定のもとで、OLS 推定量は不偏性と一致性をもち、その分布は漸近的に正規分布に従う
  - さらにある仮定のもとで、OLS 推定量は線形推定量のなかでもっとも効率的（分散が小さい）
  - これらの仮定が満たされないケースでも OLS 推定量は計算可能だが…。

なぜ最小二乗推定量なのか？

- OLS は社会科学全般でよく使われている
  - 実証分析を進めるうえでの共通言語のひとつ
  - パッケージソフトも多い。MS-Excel にも組み込み。
  - 「手計算」が比較的容易だった…

OLS 推定の前提

OLS 推定量が「望ましい」性質を持つためにはいくつかの前提が必要。

- どのようなときに用いればよいか、がわかる
- どのようなときに用いにくいか、がわかる
- 以下の条件を満たしていないときでも OLS 推定量は計算は可能

ここで，母集団と標本が満たすべき条件は以下の 3 つ

1. 説明変数を所与としたときの誤差項の条件付分布の平均がゼロ
2. 各観測値の分布は i.i.d.
3. 説明変数と誤差項の 4 次モーメントは有限

もし，手許にあるデータ（標本）がこれらの仮定を満たしていれば，OLS は望ましい性質を持つ推定量となる

仮定 1：$E[u_i|X_i] = 0$

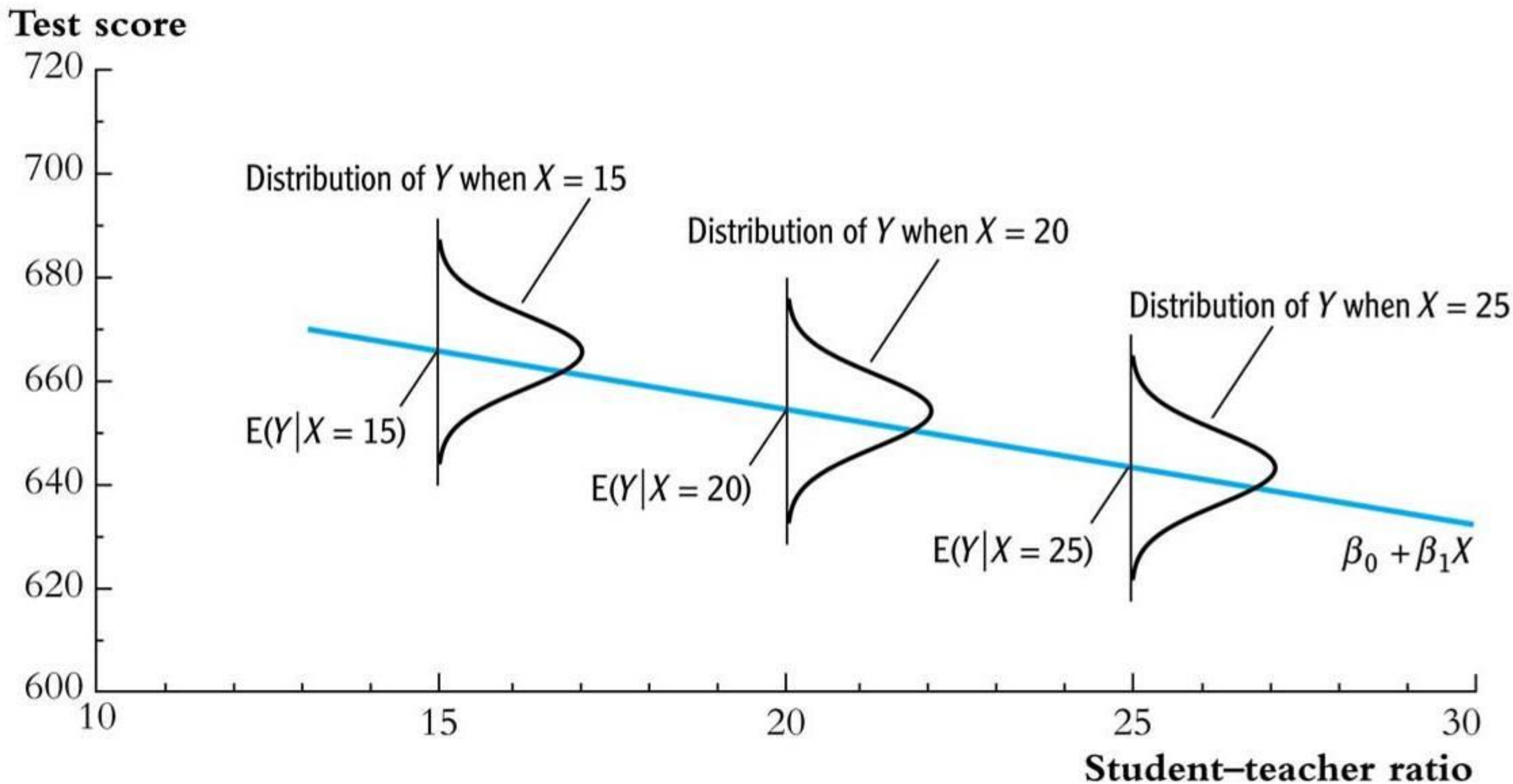
X_i を所与としたときの u_i の条件付分布の平均がゼロ。

- u_i が表している「その他の要素」についての仮定
- 説明変数を所与としたときに誤差項の平均がゼロ、という意味において、誤差項と説明変数のあいだに関係はない
- 観測値は（真の）回帰線の周囲に均等に分布
- 説明変数 X_i を所与としたとき、Y_i の条件付期待値は真の回帰直線上に並ぶ
- Figure 4.4

実際に成り立っているかどうかには慎重な検討と判断が必要

- Randomized experiment では $E[u|X = 0] = 0$ かつ $E[u|X = 1] = 0$

Figure 4.4 The conditional probability distributions and the population regression line



仮定 1：$E[u_i|X_i] = 0$

説明変数と誤差項の相関との関係

- 条件付き期待値がゼロなら共分散もゼロだから、

$$E[u_i|X_i] = 0 \Longrightarrow \mathrm{corr}(X_i, u_i) = 0$$

- ただし、逆は必ずしも成り立たないから、説明変数と誤差項の相関がゼロだからといって誤差項の条件付き期待値がゼロになるとは限らない
- 対偶は成り立つから、説明変数と誤差項に相関があれば誤差項の条件付き期待値はゼロにならない
- それゆえ、説明変数と誤差項の相関で考えてもよい。
- このような条件を「直交条件 orthogonality condition」と呼び，説明変数が「外生 exogenous」という．条件が満たされていないとき，説明変数が「内生 endogenous」という．

仮定 2：(X_i, Y_i) は **i.i.d.**

標本抽出の方法についての仮定

- 個人の無作為抽出のケースには成り立つ

成り立たないケース

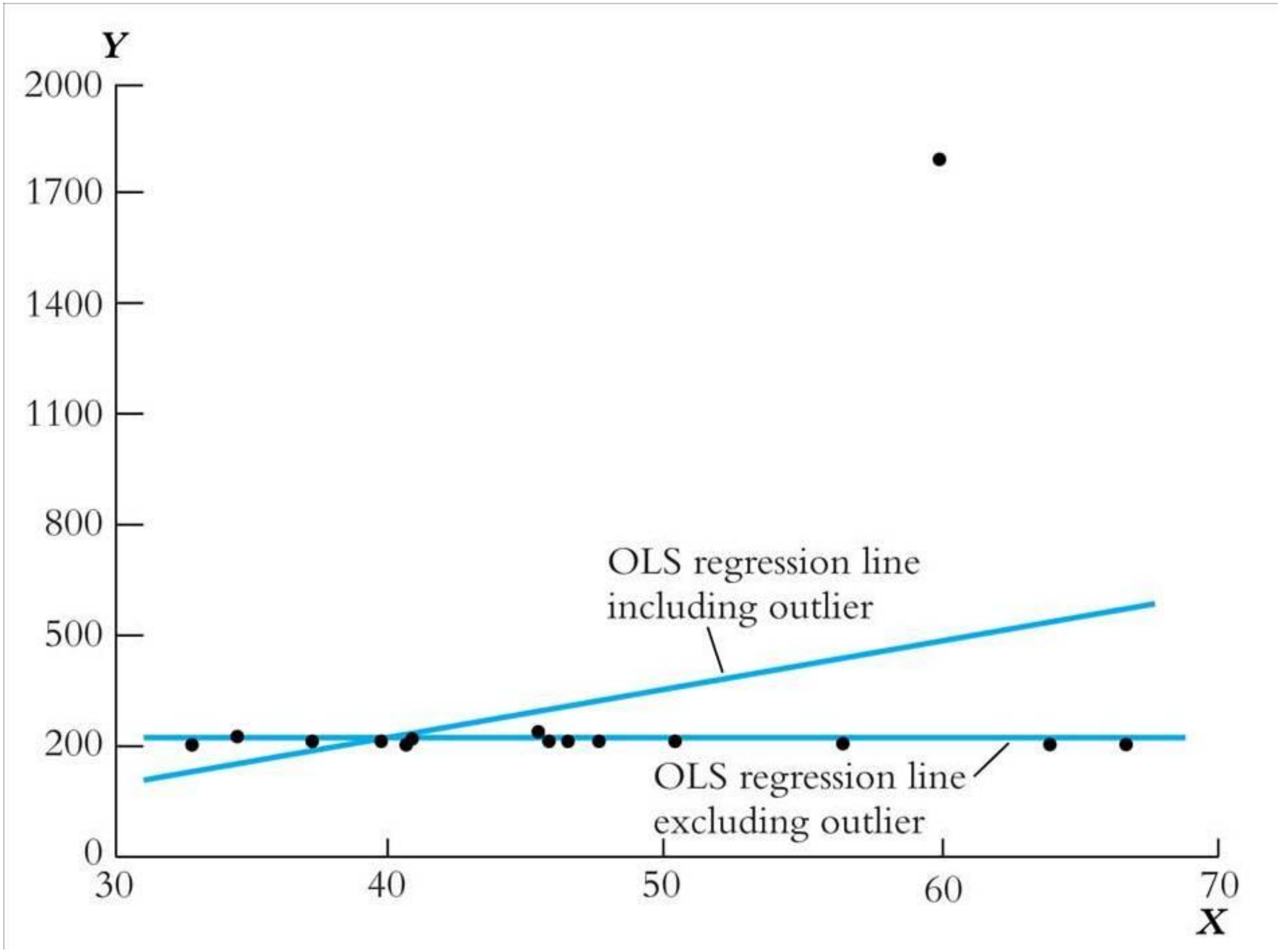
- X_i が実験の一部として設定されているケース（稀）
- 時系列データ：同じ主体の通時的変化を追っているケース
  - 時点が近ければ相関を持つ可能性が高い（独立でない）
  - 特殊な扱いが必要
- （Oversampling のケース）

仮定 3：$0 < E[X_i^4], E[u_i^4] < \infty$

X_i と u_i が有限の 4 次モーメントを持つ

- 極端な外れ値を持たない
- OLS の検定統計量の大標本近似を正当化する仮定：中心極限定理の応用のため（標本分散の一致性の証明にも用いたことを想起せよ）
- 確認するのは困難だが、成立しているものとして扱うことがほとんど（観測される値は有限個）。
- 正規分布の 4 次モーメントは有限
- Figure 4.5 は極端な例

Figure 4.5 The sensitivity of OLS to large outliers



OLS 推定の仮定とは…

数学的なもの

- これらの仮定が成り立てば、OLS 推定量の標本分布は漸近的に正規分布に従う
- 仮説検定や信頼区間の形成が可能になる

OLS が使いやすい／使いにくい状況の特定化

- 実際にはさまざまな事情でこれらの仮定は厳密には成り立たない
- とくに時系列データのばあい
- それらへの対処法はまた後ほど

OLS 推定量の標本分布

OLS 推定量は確率変数で標本分布を持つ

- OLS 推定量は標本によって決まるから、母集団が同じであっても標本が変われば値は変わる
- 小標本の分布は複雑だが、大標本ならば中心極限定理によって漸近的に正規分布に従う
- 仮説検定などを行うためには標本分布の性質を知っておくことが必要
- 前述の仮定のもとで、OLS 推定量は一致性と不偏性を持ち、漸近的に正規分布に従う

推定量の分布

標本平均の分布

- 小標本のときに分布の形状を特定化するのは困難だが、大標本（$n \to \infty$）のとき、無作為標本であれば

$$
\text{（不偏性）} \quad E[\overline{Y}] = \mu_y
$$

$$
\text{（一致性）} \quad \overline{Y} \xrightarrow{d} N\left(\mu_y, \sigma_{\overline{Y}}^2\right)
$$

OLS 推定量の分布

- 小標本のときに分布の形状を特定化するのは困難だが、サンプルの大きさによらず

$$
\text{（不偏性）} \quad E[\hat{\beta}_0] = \beta_0, \qquad E[\hat{\beta}_1] = \beta_1 \tag{4.20}
$$

OLS 推定量の不偏性の証明

OLS 推定量は

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2}, \quad \hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

いま、$Y_i = \beta_0 + \beta_1 X_i + u_i$ だから

$$Y_i - \overline{Y} = \beta_1 (X_i - \overline{X}) + u_i - \overline{u}$$

$\hat{\beta}_1$ の式の分子に代入して整理すると

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})(\beta_1 (X_i - \overline{X}) + u_i - \overline{u})}{\sum_{i=1}^n (X_i - \overline{X})^2} = \beta_1 + \frac{\sum_{i=1}^n (X_i - \overline{X}) u_i}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

両辺期待値をとると、

$$E[\hat{\beta}_1] = \beta_1 + E\left[E\left[\left. \frac{\sum_{i=1}^n (X_i - \overline{X}) u_i}{\sum_{i=1}^n (X_i - \overline{X})^2} \right| X_i \right] \right]$$

$$= \beta_1 + E\left[\frac{\sum_{i=1}^n (X_i - \overline{X}) E\left[u_i | X_i \right]}{\sum_{i=1}^n (X_i - \overline{X})^2} \right] = \beta_1 \qquad Q.E.D.$$

OLS 推定量の一致性

$n \to \infty$ のとき、OLS 推定量は中心極限定理によって 2 変量正規分布に漸近的に従う

（証明） OLS 推定量は

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(X_i - \overline{X})u_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

だから、まず分子に着目すると、$\overline{X}$ は μ_X の一致推定量だから分子は $v_i \equiv (X_i - \overline{X})u_i$ の標本平均で近似できる。ここで、$E[u_i|X_i] = 0$ だから $E[v_i] = 0$ であり、また標本は i.i.d.、$\mathrm{var}(v_i) = \mathrm{var}[(X_i - \overline{X})u_i] < \infty$ だから中心極限定理が成り立ち、

$$v \xrightarrow{d} N(0, \sigma_v^2/n)$$

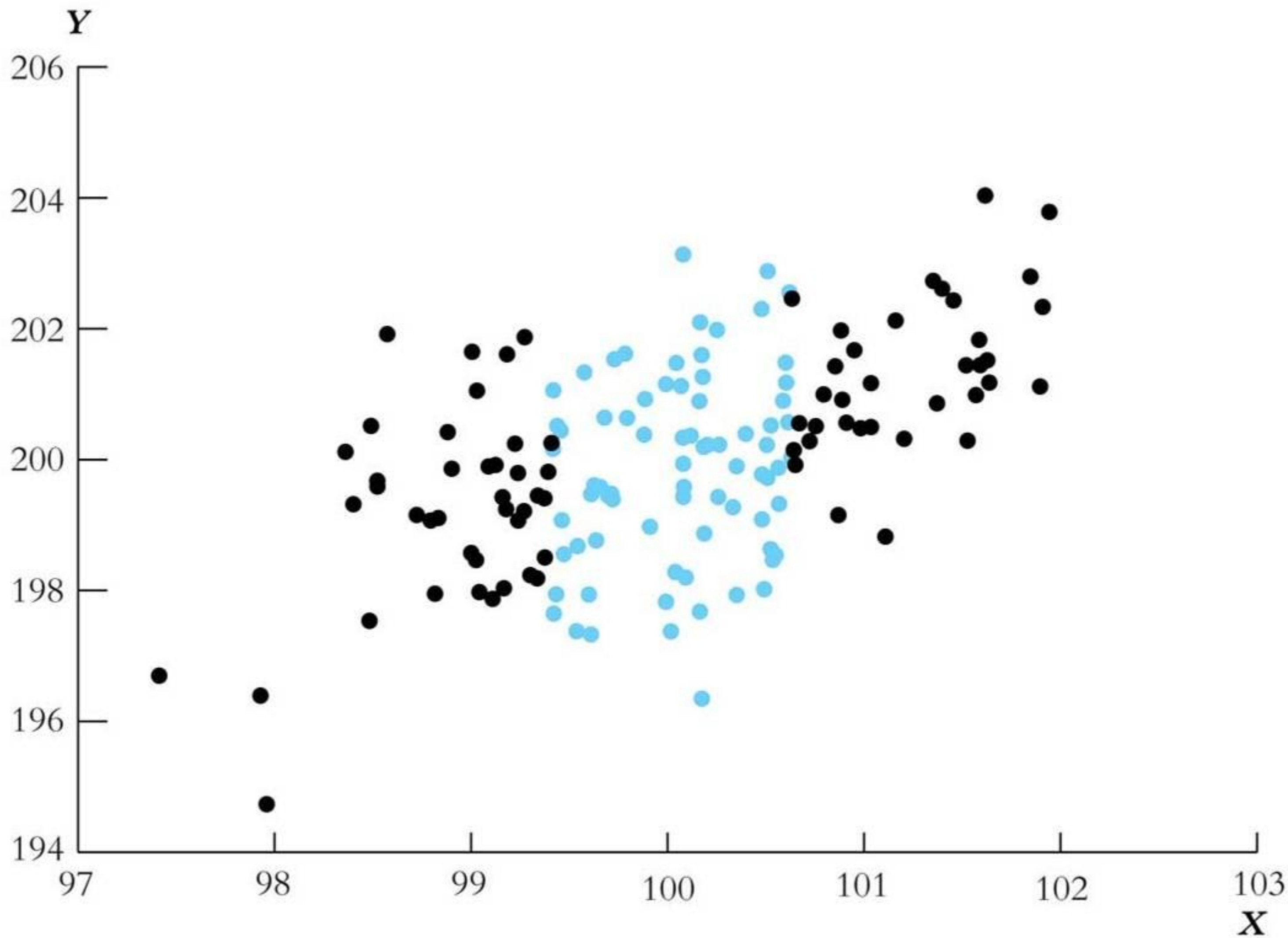
分母は $\mathrm{var}(X)$ の一致推定量だから、$\hat{\beta}_1 - \beta_1 \cong \overline{v}/\mathrm{var}(X)$。それゆえ

$$\hat{\beta}_1 \xrightarrow{d} N\left(\beta_1, \frac{\mathrm{var}((X - \mu_X)u)}{n(\mathrm{var}(X))^2}\right) \qquad Q.E.D.$$

大標本理論の適用可能性

- $n > 100$ もあれば十分。今後の他の推定量についても同様。
- OLS 推定量は一致性を持ち、その標準誤差はサンプルサイズが大きいほど小さくなる
- X_i の（標本）分散が大きいほど OLS 推定量（$\hat{\beta}_1$）の分散は小さい：散らばっているほうが正確な線を引きやすい（Fig 4.6）
- 正規分布に漸近的に従うという性質を使うと仮説検定や信頼区間の設定が容易。

Figure 4.6 The variance of beta and the variance of X



係数についての仮説検定

「クラスあたり児童数はテストの点数に影響しない」を調べたい

- 仮説を数字で表現する：$\beta_{児童数} = 0$
- 仮説検定を行う：どうやって？

復習：母平均についての仮説検定

1. 帰無仮説・対立仮説の設定：
   $H_0 : E[Y] = \mu_{Y,0} \quad \text{v.s.} \quad H_1 : E[Y] \neq \mu_{Y,0}$
2. 標本平均 $\overline{Y}$ の標準誤差（$\mathrm{SE}(\overline{Y})$）の推定
3. t 値の算出：$t = (\overline{Y} - \mu_{Y,0})/\mathrm{SE}(\overline{Y})$
4. p 値の算出：H_0 を棄却できる有意水準の最小値
   - H_0 が正しいとしたときに、得られた値よりも「離れた」値が得られる確率
   - $n \to \infty$ では $t \xrightarrow{d} N(0,1)$ だから $p = 2\Phi(-|t^{\mathrm{act}}|)$

$\hat{\beta} \xrightarrow{d} N$ だから、基本的な手続きは母平均の仮説検定と同じ

[1] 帰無仮説・対立仮説の設定

$$H_0 : \beta_1 = \beta_{1,0} \quad \text{v.s.} \quad H_1 : \beta_1 \neq \beta_{1,0} \tag{5.2}$$

[2] OLS 推定量 $\hat{\beta}_1$ の標準誤差（SE($\hat{\beta}_1$)）の推定

$$\mathrm{SE}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}} = \sqrt{\frac{1}{n}\frac{\frac{1}{n-2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \hat{u}_i^2}{[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2]^2}} \tag{5.3}$$

推定量の分散を対応する標本統計量で置き換えたもの

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{\mathrm{var}[(X_i - \mu_X)u_i]}{(\mathrm{var}[X_i])^2} \tag{5.4}$$

OLS 推定量の仮説検定

[3] t 値の算出：

$$t = \frac{\text{推定量} - \text{仮説の値}}{\text{推定量の標準誤差}} = \frac{\hat{\beta}_1 - \beta_{1,0}}{\mathrm{SE}(\hat{\beta}_1)} \tag{5.5}$$

[4] p 値の算出：H_0 を棄却できる有意水準の最小値

$$p = \Pr_{H_0}\left[|\hat{\beta}_1 - \beta_{1,0}| > |\hat{\beta}_1^{\mathrm{act}} - \beta_{1,0}|\right] = \Pr_{H_0}(|t| > |t|^{\mathrm{act}}) \tag{5.6}$$

$\hat{\beta} \xrightarrow{d} N$ だから、

$$p = \Pr_{H_0}(|Z| > |t|^{\mathrm{act}}) = 2\Phi(-|t^{\mathrm{act}}|) \tag{5.7}$$

- H_0 が正しいとしたときに、得られた値よりも「離れた」値が得られる確率

OLS 推定量の片側検定

- 帰無仮説の設定

$$H_0 : \beta_1 = \beta_{1,0} \quad \text{v.s.} \quad H_1 : \beta_1 < \beta_{1,0} \tag{5.9}$$

- t 値の解釈、p 値の算出

$$p = \Pr_{H_0}(Z < t^{\text{act}}) = \Phi(t^{\text{act}}) \tag{5.10}$$

片側検定を使うとき

- 仮説の値より大きな（小さな）値を取ることが理論的・実証的に自明なとき
- そのようなケースは多くないので、両側検定を使うケースが多い
  - 価格効果の符号条件？
  - 他の要因の混在の可能性

係数の信頼区間の設定

標本（データ）は確率的要素を含んでいるから、真の値を同定することはできない。(β_0, β_1) の信頼区間を求めることはできる

β_1 の信頼区間

- 95%信頼区間とは
  - 有意水準 5%の両側検定で棄却されない値の集合
  - 真の値 β を含む確率が 95%であるような区間（標本の 95%は真の値を含むような区間）
- 母平均の信頼区間の形成と同様
  - 5%の有意水準による検定で棄却されない値を集めてくればよい
  - $H_0 : \beta_1 = \beta_{1,0}$ が棄却されるのは、$\beta_{1,0}$ が $(\hat{\beta}_1 \pm 1.96\mathrm{SE}(\hat{\beta}_1))$ の外にあるとき
  - 信頼区間は、$(\hat{\beta}_1 - 1.96\mathrm{SE}(\hat{\beta}_1), \hat{\beta}_1 + 1.96\mathrm{SE}(\hat{\beta}_1))$

説明変数 X の値が Δx だけ変化したとする

- 対応する Y の変分は $\Delta y = \beta_1 \Delta x$
- 真の Δx は分からないから、推定値 $\hat{\beta}_1$ を用いる
- 点推定のほかに、信頼区間の形成もできる
- $\hat{\beta}_1$ の信頼区間は分かっているから、

$$\Big((\hat{\beta}_1 - 1.96\mathrm{SE}(\hat{\beta}_1))\Delta x, (\hat{\beta}_1 + 1.96\mathrm{SE}(\hat{\beta}_1))\Delta x\Big) \tag{5.13}$$

説明変数が 2 値変数のときの回帰

説明変数が 2 つの値しか取らない（$D_i = 0, 1$）ときのケース

- ダミー変数（indicator variable, dummy variable）
- 質的変数を代理（性別、都市／田舎、などなど）
- 連続変数を区切る（大きい／小さい）

β_1 を「傾き」と解釈するのは適切でない

- OLS 推定量の計算方法は同じ
- 「係数」は平均の差を意味する

$$D_i = 0 \quad \text{のとき} \quad Y_i = \beta_0 + u_i$$
$$D_i = 1 \quad \text{のとき} \quad Y_i = \beta_0 + \beta_1 + u_i$$

だから、

$$E[Y_i|D_i = 0] = \beta_0, \quad E[Y_i|D_i = 1] = \beta_0 + \beta_1 \qquad (5.16, 17)$$

説明変数が 2 値変数のときの回帰

仮説検定・信頼区間の形成

- 手続きは連続のケースと同じ
- β_1 は 2 つの条件付き期待値の差だから、母平均が同じという帰無仮説は $H_0 : \beta_1 = 0$
- β_1 の OLS 推定量は 2 つのグループの標本平均の差になる

OLS 推定のあてはまりのよさ

OLS 推定がどれくらいデータと合致しているかを示す指標

R^2（決定係数）　Y_i の変動のうち X_i の変動で説明される比率。0 と 1 の間の値をとり、1 に近いほど Y_i の予測がよくできている

回帰の標準誤差（Standard Error of the Regression）　Y_i が当てはめ値からどれくらい離れているかを示す

R^2

- Y_i の変動のうち X_i の変動で説明される比率
- 実現値を $Y_i = \hat{Y}_i + \hat{u}_i$ と分解したとき、

$$R^2 = \frac{\hat{Y}_i\text{の標本分散}}{Y_i\text{の標本分散}} = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{4.16}$$

  - ESS (explained sum of squares)
  - TSS (total sum of squares)
- 残差平方和（SSR: sum of squared residuals）でも定義できて、

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\text{TSS} - \text{SSR}}{\text{TSS}} = 1 - \frac{\text{SSR}}{\text{TSS}} = 1 - \frac{\sum_{i=1}^{n}\hat{u}_i^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{4.18}$$

全変動のうち、残差の変動で説明される部分を引いた比率。

$[0,1]$ のあいだの値を取る

- $\hat{\beta}_1 = 0$ であれば、X_i は Y_i の変動をまったく説明できず、$\hat{Y}_i = \overline{Y}, \forall i$。このとき ESS はゼロに等しい。
- $\hat{Y}_i = Y_i, \forall i$ のとき、$\hat{u}_i = 0, \forall i$ だから、ESS と TSS は等しく、$R^2 = 1$
- R^2 が 1 に近いほど Y_i の予測がよくできていることになる

回帰の標準誤差 SER

誤差項 u_i の標準誤差の推定値

- 誤差項 $\{u_1, u_2, ..., u_n\}$ は観測されないから、対応するものを用いる
- 残差 $\{\hat{u}_1, \hat{u}_2, ..., \hat{u}_n\}$ を用いると、残差の平均はゼロだから、

$$\mathrm{SER} = s_{\hat{u}}, \quad s_{\hat{u}}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2 = \frac{\mathrm{SSR}}{n-2} \tag{4.19}$$

$n-2$ で割っているのは、2 つの係数を推定したことによる自由度修正。n が大きくなれば無視できる。

分散均一と分散不均一

誤差項についての唯一の仮定は $E[u_i|X_i] = 0$
分散についての仮定は置いてこなかった

- 説明変数の実現値 X_i を所与としたときの誤差項の条件付き分散 $E[u_i^2|X_i]$ がすべての i について一定で、X_i に依存しないとき、分散均一（homoskedasticity）という
- 説明変数の実現値 X_i を所与としたときの誤差項の条件付き分散 $E[u_i^2|X_i]$ が一定でないとき、分散不均一（heteroskedasticity）という
- (X_i, Y_i) は i.i.d. なので、無条件分散は一定
- Fig4.4. と Fig 5.2. の比較

Figure 5.2 An example of homoskedasticity

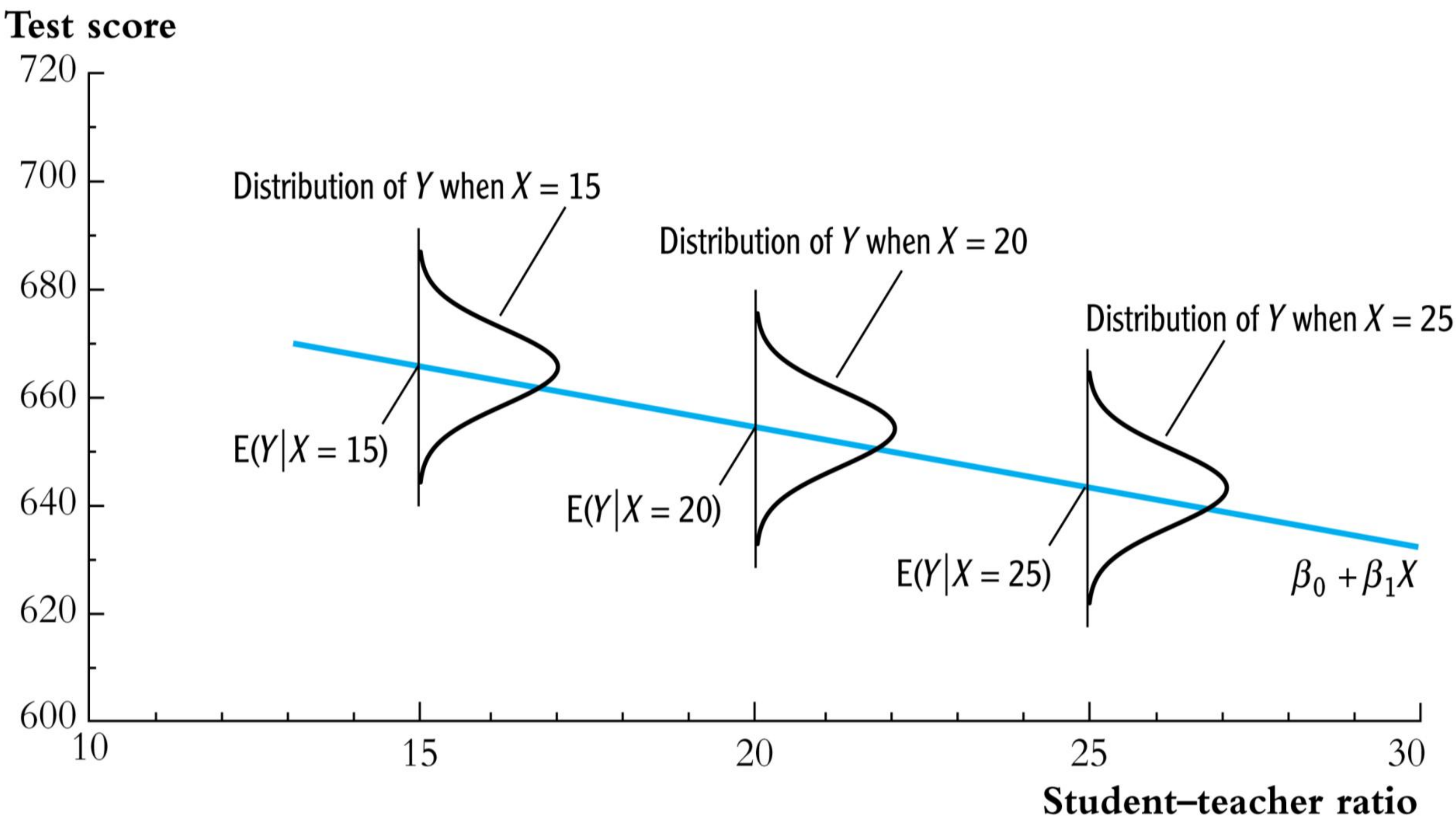
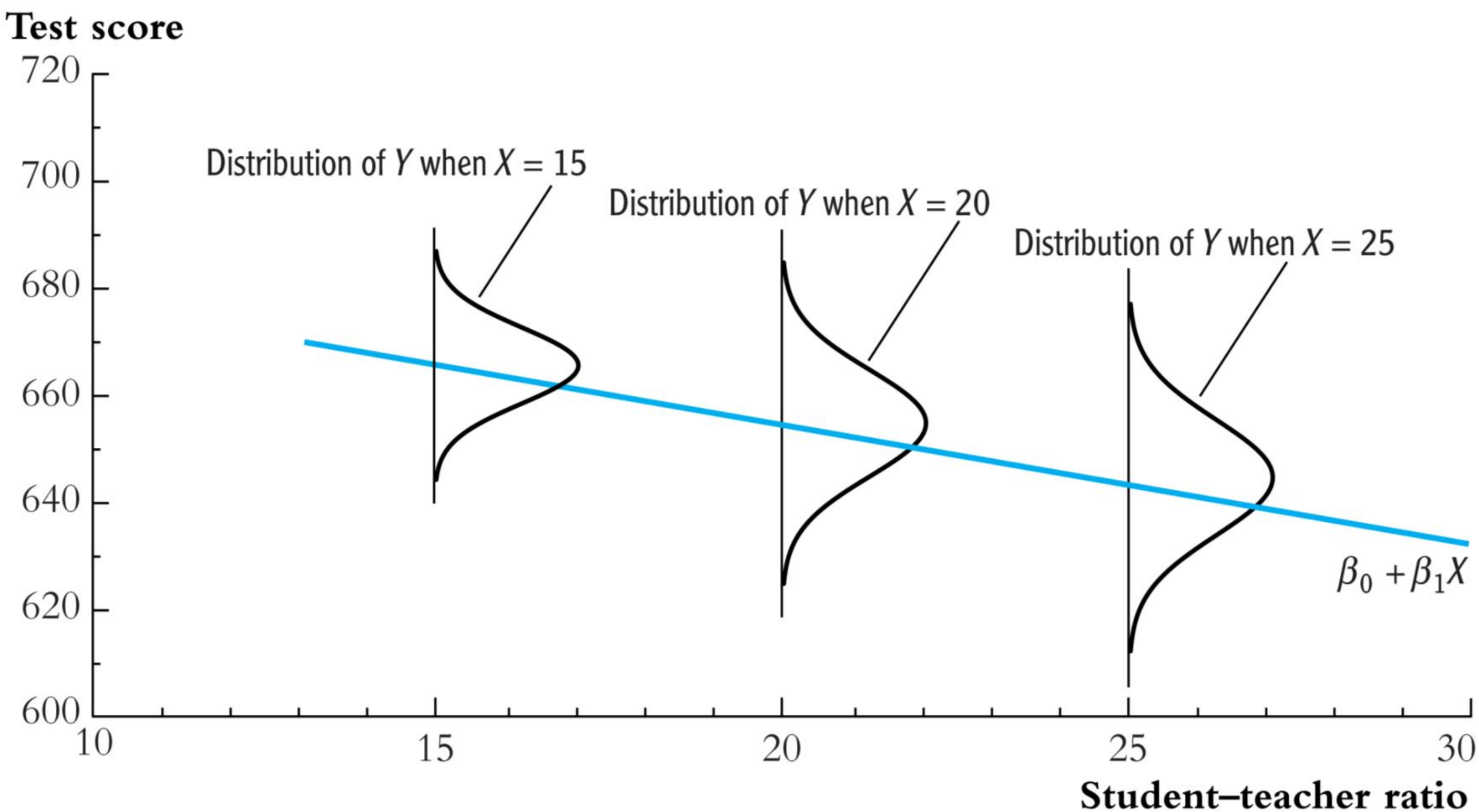


Figure 5.2 An example of heteroskedasticity



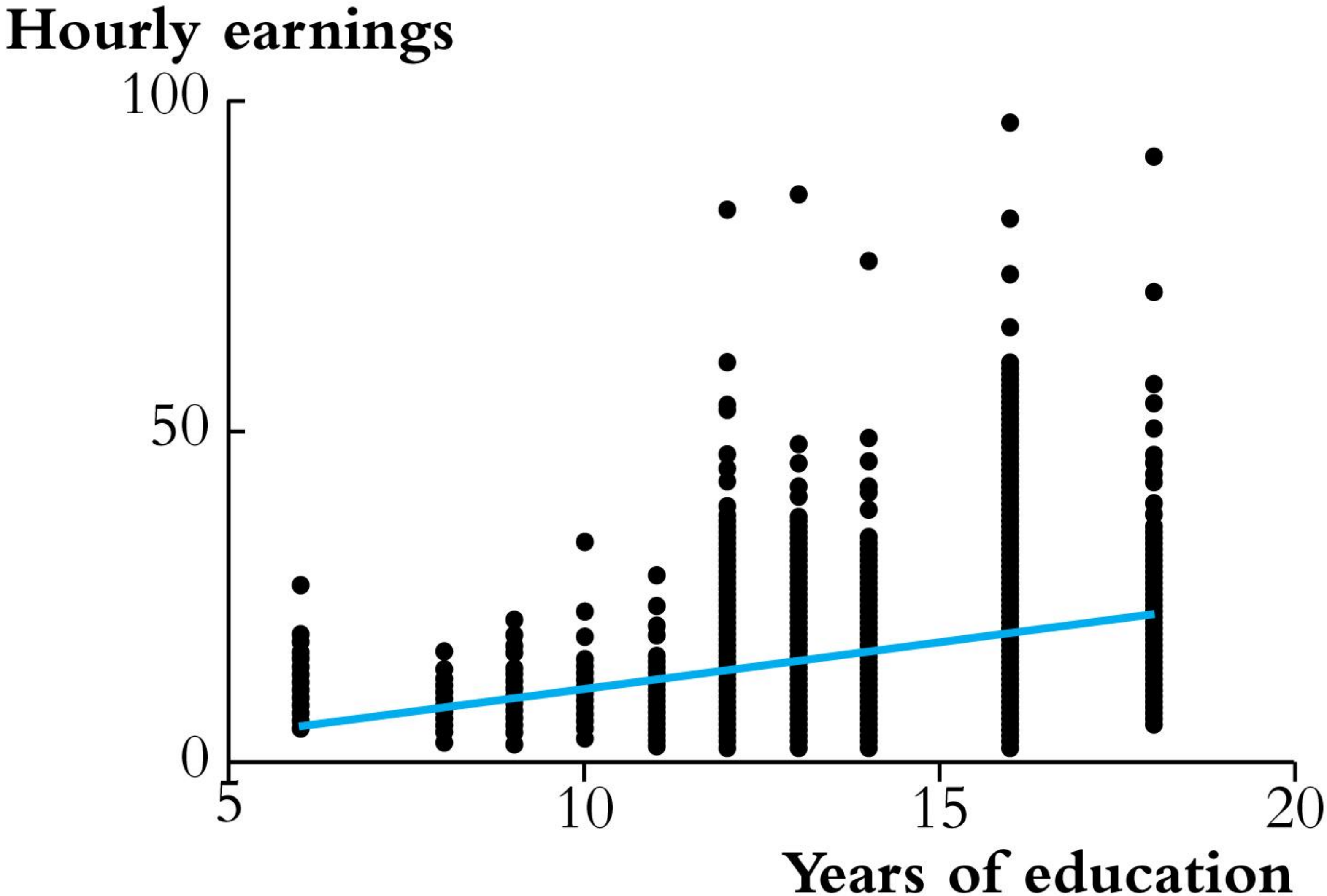
分散不均一：例

男女の賃金格差

$$\text{Earnings}_i = \beta_0 + \beta_1 \text{Male}_i + u_i \tag{5.19}$$

- 男性を表すダミー変数 Male_i の係数 β_1 は男女間の平均的な賃金格差を示す
- ここでの問題は、$\text{var}(u_i|\text{Male}_i)$ がダミー変数 Male_i に依存するかどうか
- 誤差 u_i は実際に観察できないが、この場合は、$D_i = 0, 1$ で場合わけして標本分散を計算すればよい
- 男女それぞれの賃金の分散が等しいかどうかという問題に帰着

Figure 5.3 Scatter plot of hourly earnings and years of education
Heteroskedastic or homoskedastic?



分散均一性の数学的含意

分散均一であれば、

- OLS 推定量は不偏性・一致性を持ち、漸近的に正規分布に従う
  - これらの性質は分散均一性の仮定がなくても成り立つ
  - 分散不均一のほうがより一般的な仮定
- Gauss-Markov の定理が成り立つ
  - 分散均一であれば、OLS 推定量は、$\{Y_1, Y_2, ..., Y_n\}$ について線形な不偏推定量のなかで最も efficient な（効率的、分散の小さい）推定量である
  - OLS は BLUE（Best Linear Unbiased Estimator）である
  - 逆に、分散不均一であれば、OLS 推定量よりも分散の小さい線形不偏推定量が存在する

分散均一のときの OLS 推定量

係数の推定量そのものは変わらないが、その標準誤差が簡単に

- Homoskedasticity-only な var$(\hat{\beta}_1)$

$$\mathrm{var}(\hat{\beta}_1) = \frac{\mathrm{var}[(X-\mu_X)u]}{n(\mathrm{var}(X))^2} = \frac{\mathrm{var}(u_i)}{n\mathrm{var}(X_i)} \tag{5.22}$$

- Homoskedasticity-only な var$(\hat{\beta}_1)$ は分散不均一のデータでは適切ではない。この var$(\hat{\beta}_1)$ を用いて計算された t 値は標準正規分布に従わない
- Heteroskedasticity-robust（分散不均一に頑健な）な標準誤差は分散均一のときにも適用可能（Eicker-Huber-White の標準誤差）
- 経済理論が分散均一性を含意することはあまりないので、常に robust な標準誤差を用いるほうがよい
- 計量ソフトではしばしばオプション指定が必要

WLS：Weighted Least Squares

分散不均一であるとき、OLS 推定量は BLUE ではない

- OLS より分散の小さい線形不偏推定量を作ることができる
- WLS：各観測値を $\sqrt{\mathrm{var}(u_i/X_i)}$ の逆数でウェイト付けしたデータを OLS 推定したもの
- このような変換によって分散均一の仮定が満たされることになるから。
- ただし実際には $\mathrm{var}(u_i/X_i)$ を知っている必要があるのであまり使われない
- 計量経済理論的には興味深い推定量ではある
- WLS のようなことをやろうとする推定量はまたべつに存在