Introductory Applied Econometrics Analysis using Stata

November 14 – 18, 2016

Dushanbe, Tajikistan

Allen Park and Jarilkasin Ilyasov

Linear Regression with One Regressor

Outline

- 1. The population linear regression model
- 2. The ordinary least squares (OLS) estimator and the sample regression line
- 3. Measures of fit of the sample regression
- 4. The least squares assumptions
- 5. The sampling distribution of the OLS estimator

Based on Chapter 4. Stock and Watson. “Introduction to Econometrics” 3rd Edition.

Linear Regression with One Regressor

The ***population regression line***:

$$Test\ Score = \beta_0 + \beta_1 STR$$

$$\beta_1 = \text{slope of population regression line}$$
$$= \frac{\Delta \text{Test score}}{\Delta STR}$$
$$= \text{change in test score for a unit change in } STR$$

- *Why are* β_0 *and* β_1 *"population" parameters*?
- We would like to know the population value of β_1.
- We don't know β_1, so must estimate it using data.

Linear Regression with One Regressor

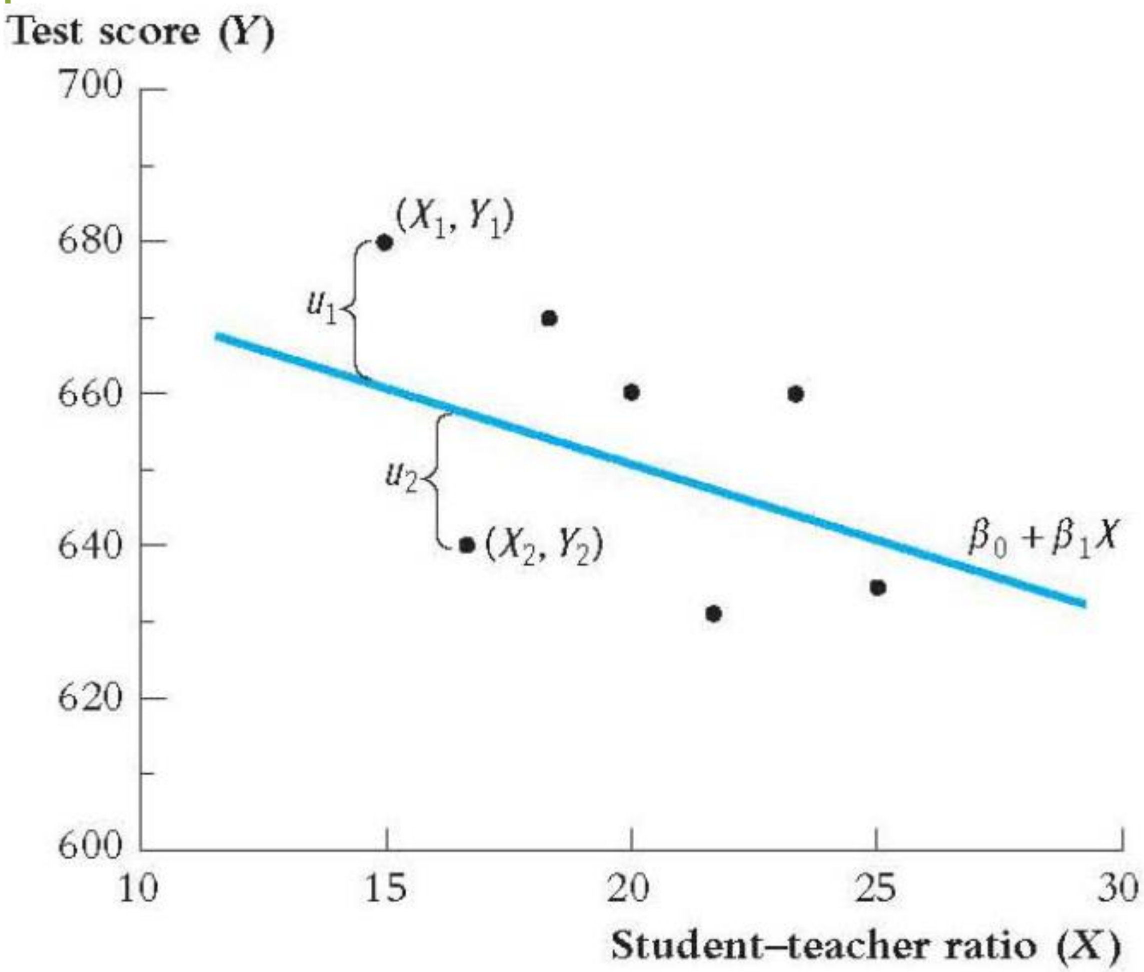
The Population Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, \ldots, n$$

- We have *n* observations, (*Xi*, *Yi*), *i* = 1,.., *n*.
- X_i is the ***independent variable*** or ***regressor***
- Y_i is the ***dependent variable***
- β_0 = ***intercept*** (the value of population when X=0)
- β_1 = ***slope*** (change in Y associated with a unit change in X)
- u_1 = the regression ***error*** (all of the factors)
- The regression error consists of omitted factors. In general, these omitted factors are other factors that influence *Y*, other than the variable *X*. The regression error also includes error in the measurement of *Y*.

Linear Regression with One Regressor

The population regression model in a picture: Observations on *Y* and *X* (n = 7); the population regression line; and the regression error (the "error term"):



Linear Regression with One Regressor

Estimating the coefficient

How can we estimate β_0 and β_1 from data?

Recall that $\bar{Y}$ was the least squares estimator of μ_Y: $\bar{Y}$ solves,

$$\min_m \sum_{i=1}^{n} (Y_i - m)^2$$

By analogy, **we will focus on the least squares ("*ordinary least squares*" or "*OLS*") estimator of the unknown parameters β_0 and β_1**. The OLS estimator solves,

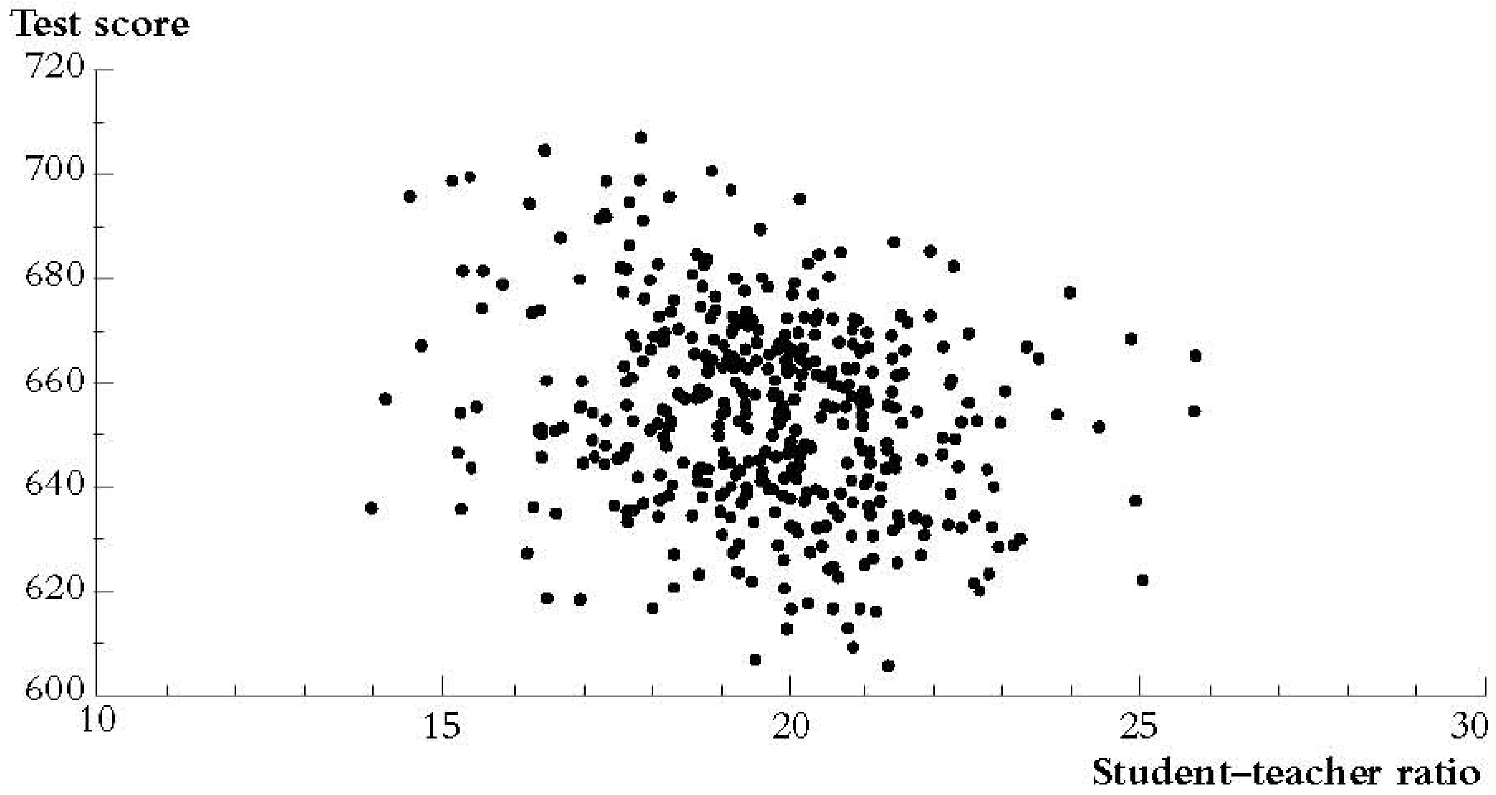
$$\min_{b_0, b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$$

Linear Regression with One Regressor

Mechanics of OLS

The population regression line: $Test\ Score = \beta_0 + \beta_1 STR$

$$\beta_1 = \frac{\Delta \text{Test score}}{\Delta STR} = ??$$



Linear Regression with One Regressor

The OLS estimator solves:

$$\min_{b_0, b_1} \sum_{i=1}^{n} [Y_i - (b_0 + b_1 X_i)]^2$$

- The OLS estimator minimizes the average squared difference between the actual values of *Yi* and the prediction ("predicted value") based on the estimated line.
- This minimization problem can be solved using calculus (App. 4.2).

The result is the OLS estimators of β_0 and β_1.

Linear Regression with One Regressor

The OLS Estimator, Predicted Values, and Residuals

The OLS estimators of the slope β_1 and the intercept β_0 are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \quad (4.7)$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \quad (4.8)$$

The OLS predicted values $\hat{Y}_i$ and residuals $\hat{u}_i$ are

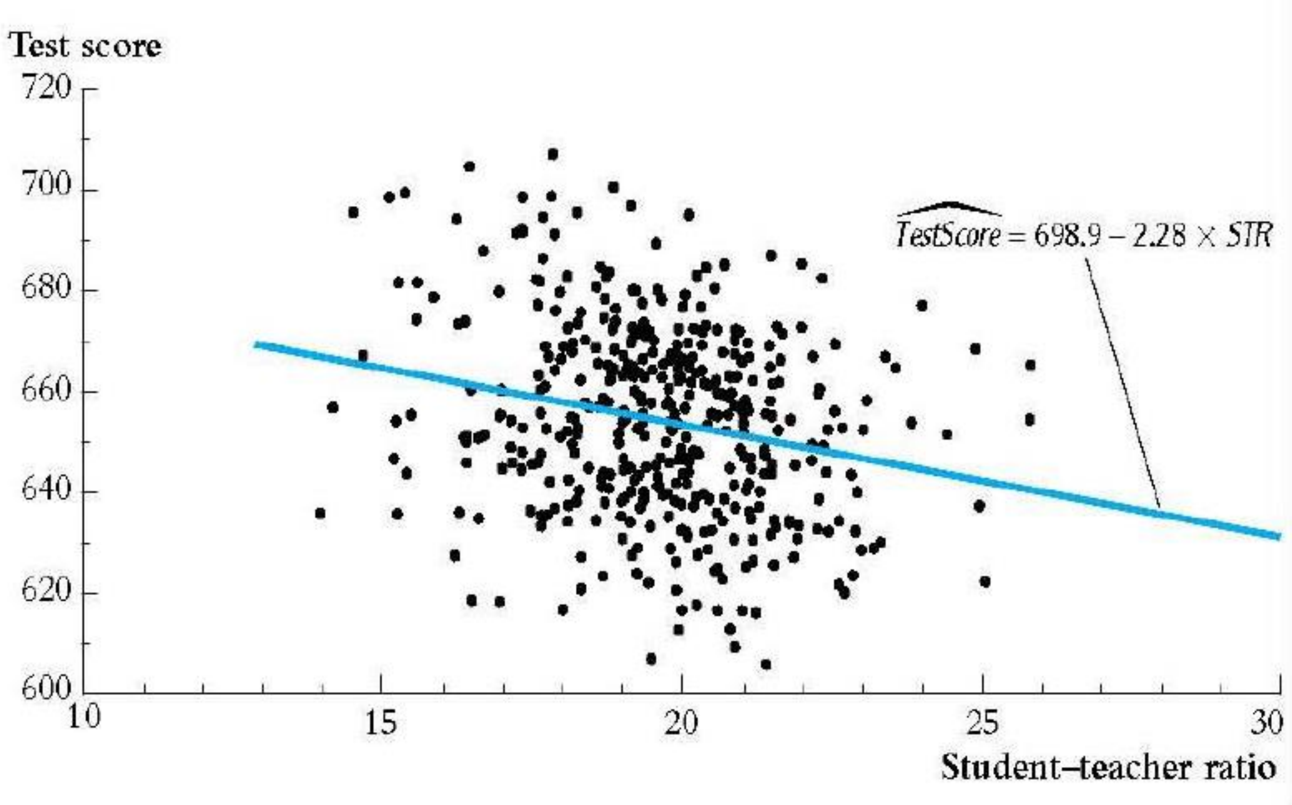
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,\ i = 1, \ldots, n \quad (4.9)$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \quad (4.10)$$

The estimated intercept $(\hat{\beta}_0)$, slope $(\hat{\beta}_1)$, and residual $(\hat{u}_i)$ are computed from a sample of n observations of X_i and Y_i, $i = 1, \ldots, n$. These are estimates of the unknown true population intercept (β_0), slope (β_1), and error term (u_i).

Linear Regression with One Regressor

Application to the California *Test Score – Class Size* data



Estimated slope $= \hat{\beta}_1 = -2.28$

Estimated intercept $= \hat{\beta}_0 = 698.9$

Estimated regression line: $\widehat{TestScore} = 698.9 - 2.28 \times STR$

Linear Regression with One Regressor

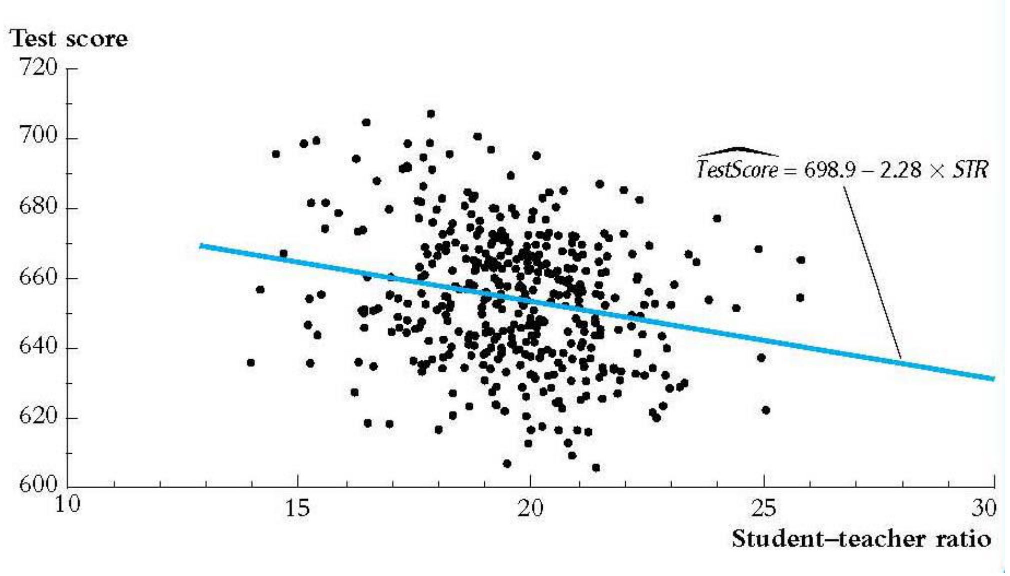
Interpretation of the estimated slope and intercept

$$\widehat{TestScore} = 698.9 - 2.28 * STR$$

- Districts with one more student per teacher on average have test scores that are 2.28 points lower.
- The intercept (taken literally) means that, according to this estimated line, districts with zero students per teacher would have a (predicted) test score of 698.9. But this interpretation of the intercept makes no sense – it extrapolates the line outside the range of the data – here, the intercept is not economically meaningful.

Linear Regression with One Regressor

Predicted values & residuals:



- One of the districts in the data set is Antelope, CA, for which *STR* = 19.33 and *Test Score* = 657.8

predicted value: $\hat{Y}_{Antelope} = 698.9 - 2.28 \times 19.33 = 654.8$

residual: $\hat{u}_{Antelope} = 657.8 - 654.8 = 3.0$

Linear Regression with One Regressor

OLS regression: STATA output

```
regress testscr str, robust

Regression with robust standard errors               Number of obs =     420
                                                     F(  1,   418) =   19.26
                                                     Prob > F      =  0.0000
                                                     R-squared     =  0.0512
                                                     Root MSE      =  18.581

-------------------------------------------------------------------------
        |               Robust
testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+----------------------------------------------------------------
    str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
  _cons |    698.933   10.36436    67.44   0.000     678.5602    719.3057
-------------------------------------------------------------------------
```

(We'll discuss the rest of this output later.)

Linear Regression with One Regressor

Measures of Fit

- Two regression statistics provide complementary measures of how well the regression line “fits” or explains the data:
- The ***regression*** R^2 measures the fraction of the variance of *Y* that is explained by *X*; it is unitless and ranges between zero (no fit) and one (perfect fit)
- The ***standard error of the regression* (*SER*)** measures the magnitude of a typical regression residual in the units of *Y*.

Linear Regression with One Regressor

The *regression* R^2 is the fraction of the sample variance of Y_i "explained" by the regression.

$$Y_i = \hat{Y}_i + \hat{u}_i \;\; = \text{OLS prediction} + \text{OLS residual}$$

- Total sum of squares = "explained" SS + "residual" SS

Definition of R^2:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- $R^2 = 0$ means $ESS = 0$
- $R^2 = 1$ means $ESS = TSS$
- $0 \leq R^2 \leq 1$
- For regression with a single X, R^2 = the square of the correlation coefficient between X and Y

Linear Regression with One Regressor

The Standard Error of the Regression (SER)

- The *SER* measures the spread of the distribution of *u*. The *SER* is (almost) the sample standard deviation of the OLS residuals:

$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(\hat{u}_i - \bar{\hat{u}})^2}$$

$$= \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

- The *SER*:
  - has the units of *u*, which are the units of *Y*
  - measures the average "size" of the OLS residual (the average "mistake" made by the OLS regression line)

Linear Regression with One Regressor

Technical note: why divide by *n*–2 instead of *n*–1?

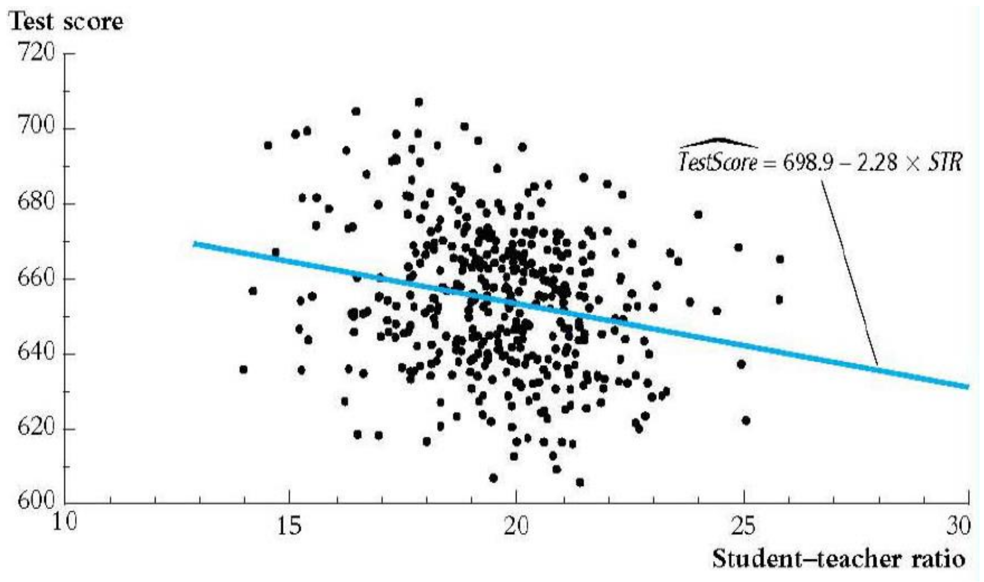
$$SER = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2}$$

- Division by *n*–2 is a "degrees of freedom" correction – just like division by *n*–1 in s_y^2, except that for the *SER*, two parameters have been estimated (β_0 and β_1, by $\hat{\beta}_0$ and $\hat{\beta}_1$), whereas in s_y^2 only one has been estimated (μ_Y, by $\bar{Y}$).
- When *n* is large, it doesn't matter whether *n*, *n*–1, or *n*–2 are used – although the conventional formula uses *n*–2 when there is a single regressor.

For details, see Section 17.4

Linear Regression with One Regressor

Example of the R^2 and the *SER*



TestScore = 698.9 – 2.28**STR*, **R2 = .05, *SER* = 18.6**

- *STR explains only a small fraction of the variation in test scores. Does this make sense? Does this mean the STR is unimportant in a policy sense?*

Linear Regression with One Regressor

The Least Squares Assumptions

- What, in a precise sense, are the properties of the sampling distribution of the OLS estimator? When will β_1 be unbiased? What is its variance?
- To answer these questions, we need to make some assumptions about how *Y* and *X* are related to each other, and about how they are collected (the sampling scheme)
- These three assumptions are known as the Least Squares Assumptions.

Linear Regression with One Regressor

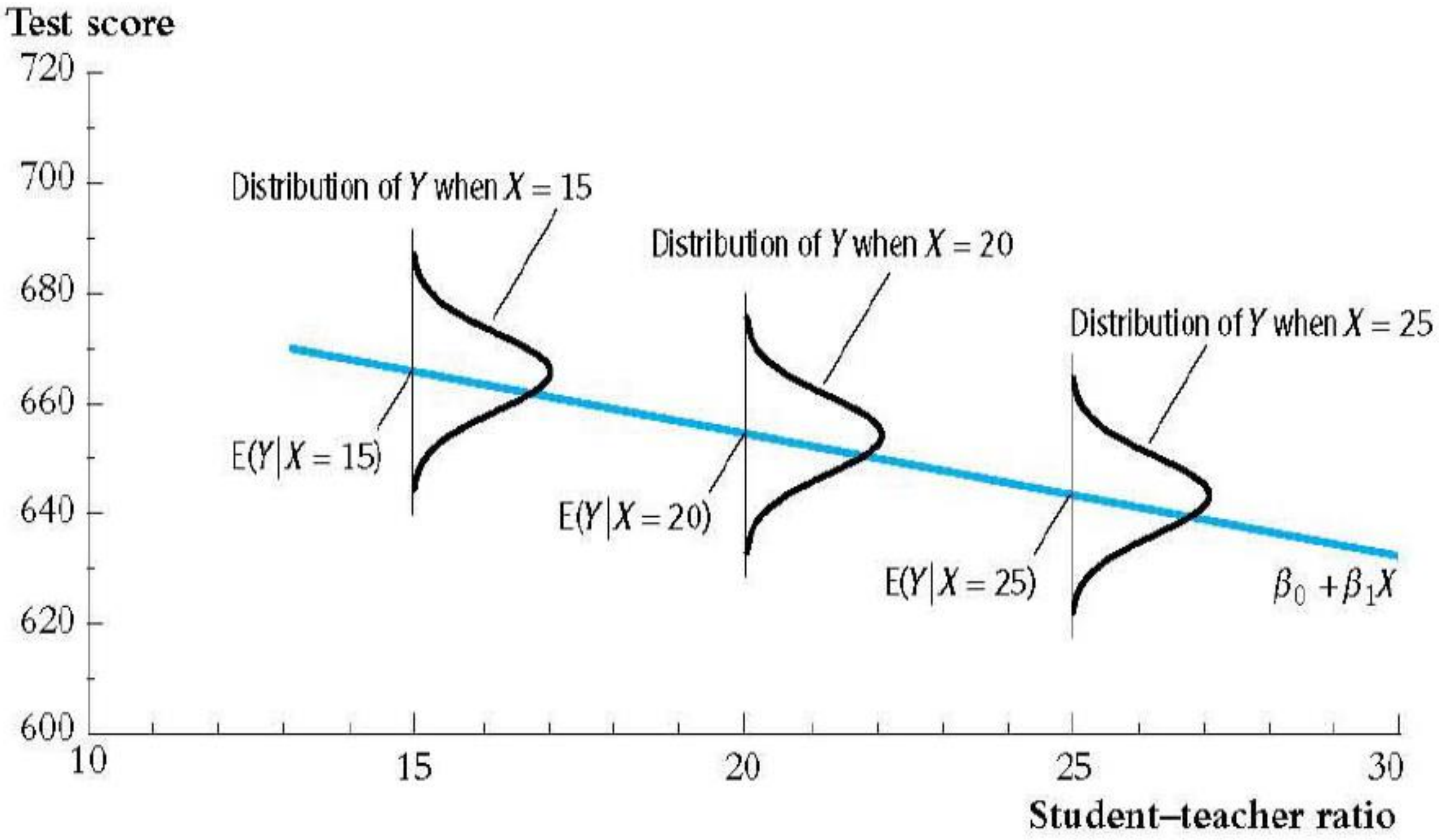
The Least Squares Assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1,\ldots, n$$

1. The conditional distribution of u given X has mean zero, that is, $E(u|X = x) = 0$.
    - *This implies that $\hat{\beta}_1$ is unbiased*
2. (X_i, Y_i), $i = 1,\ldots,n$, are i.i.d.
    - *This is true if (X, Y) are collected by simple random sampling*
    - *This delivers the sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$*
3. Large outliers in X and/or Y are rare.
    - *Technically, X and Y have finite fourth moments*
    - *Outliers can result in meaningless values of $\hat{\beta}_1$*

Linear Regression with One Regressor

Least squares assumption #1: $E(u|X = x) = 0$ (*for any given value of X, the mean of u is zero*):



Linear Regression with One Regressor

Least squares assumption #1 (cont.):

- A benchmark for thinking about this assumption is to consider an ideal randomized controlled experiment:
  - X is randomly assigned to people (students randomly assigned to different size classes; patients randomly assigned to medical treatments). Randomization is done by computer – using no information about the individual.
  - Because X is assigned randomly, all other individual characteristics – the things that make up u – are distributed independently of X, so u and X are independent
  - Thus, in an ideal randomized controlled experiment, $\boldsymbol{E(u|X = x) = 0}$ (that is, LSA #1 holds)
  - In actual experiments, or with observational data, we will need to think hard about whether $E(u|X = x) = 0$ holds.

Linear Regression with One Regressor

Least squares assumption #2: (X_i,Y_i), i = 1,…,n are i.i.d.

- This arises automatically if the entity (individual, district) is sampled by simple random sampling:
  - The entities are selected from the same population, so (X_i, Y_i) are *identically distributed* for all i = 1,…, n.
  - The entities are selected at random, so the values of (X,Y) for different entities are *independently distributed*.
- The main place we will encounter non-i.i.d. sampling is when data are recorded over time for the same entity (panel data and time series data) – you deal with that complication when we cover panel data.

Linear Regression with One Regressor

Least squares assumption #3: *Large outliers are rare*

$$\textit{Technical statement: } E(X^4) < \infty \text{ and } E(Y^4) < \infty$$

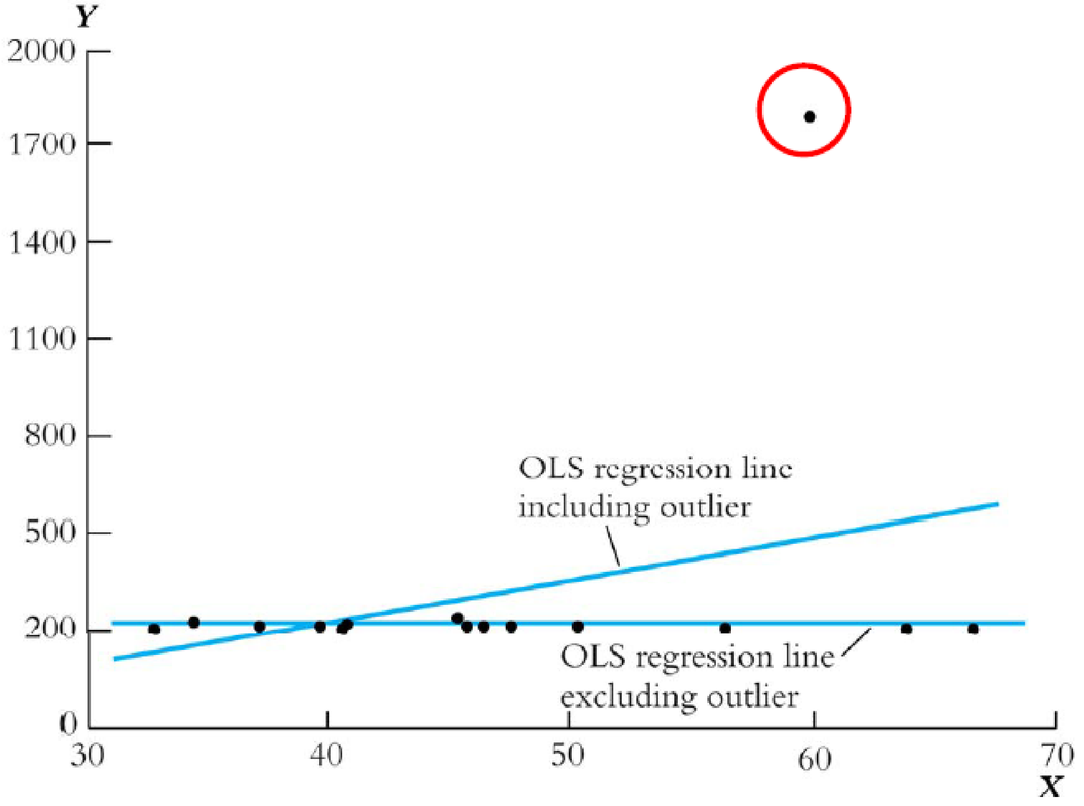
- A large outlier is an extreme value of *X* or *Y*
- On a technical level, if *X* and *Y* are bounded, then they have finite fourth moments. (Standardized test scores automatically satisfy this; *STR*, family income, etc. satisfy this too.)
- The substance of this assumption is that a large outlier can strongly influence the results – so we need to rule out large outliers.
- Look at your data! If you have a large outlier, is it a typo?
- Does it belong in your data set? Why is it an outlier?

Linear Regression with One Regressor

OLS can be sensitive to an outlier:

Is the lone point an outlier in X or Y?

In practice, outliers are often data glitches (coding or recording problems). Sometimes they are observations that really shouldn't be in your data set. Plot your data!



Linear Regression with One Regressor

The Sampling Distribution of the OLS Estimator

- The OLS estimator is computed from a sample of data. A different sample yields a different value of $\hat{\beta}_1$. This is the source of the "sampling uncertainty" of $\hat{\beta}_1$. We want to:
  - quantify the sampling uncertainty associated with $\hat{\beta}_1$
  - use $\hat{\beta}_1$ to test hypotheses such as $\hat{\beta}_1 = 0$
  - construct a confidence interval for $\hat{\beta}_1$
  - All these require figuring out the sampling distribution of the OLS estimator. Two steps to get there…
    - Probability framework for linear regression
    - Distribution of the OLS estimator

Linear Regression with One Regressor

The probability framework for linear regression is summarized by the three least squares assumptions.

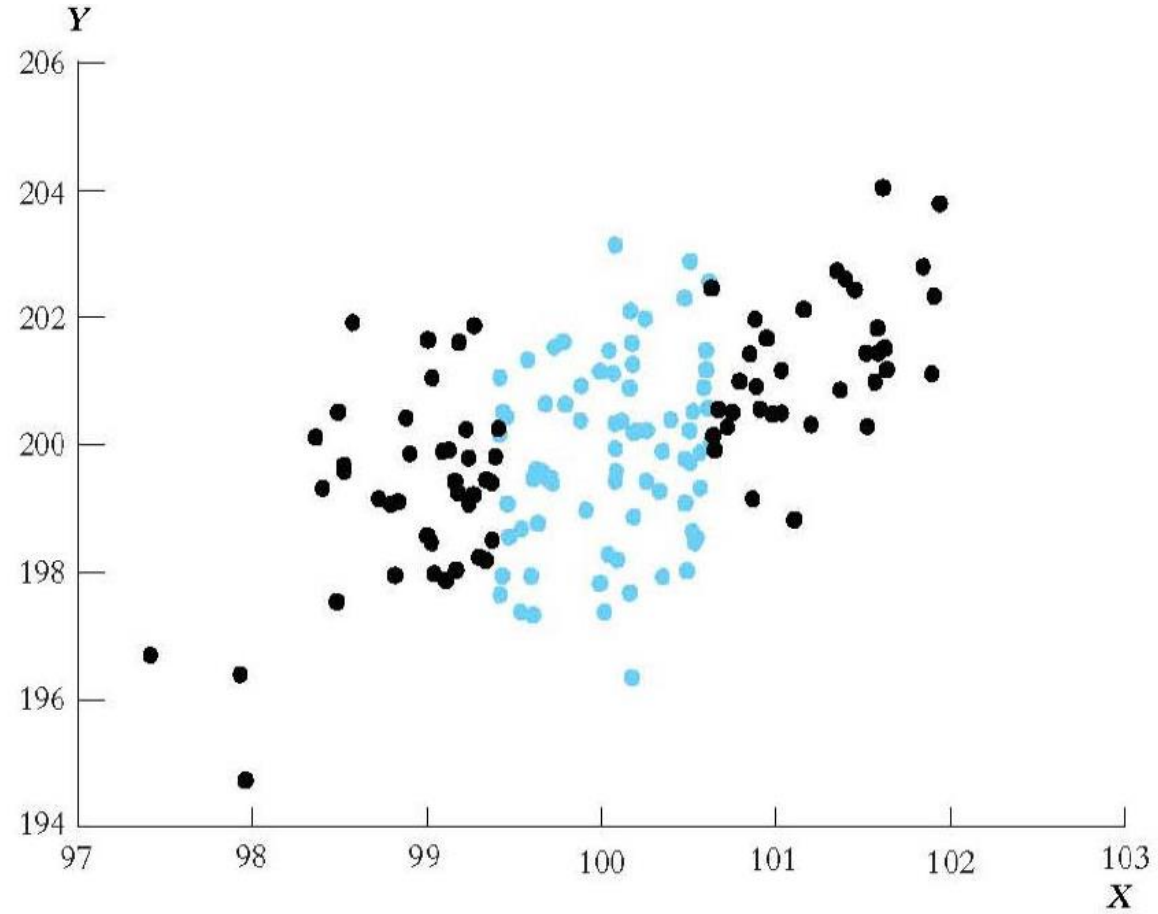
- ***Population***
  - The group of interest (ex: all possible school districts)
- ***Random variables*: *Y*, *X***
  - Ex: (*Test Score, STR*)
- ***Joint distribution of* (*Y*, *X*). We assume:**
  - The population regression function is linear
  - $E(u|X) = 0$ (1st Least Squares Assumption)
  - *X*, *Y* have nonzero finite fourth moments (3rd L.S.A.)
- ***Data Collection by simple random sampling implies***:
  - $\{(X_i, Y_i)\}$, $i = 1,\ldots, n$, are i.i.d. (2nd L.S.A.)

Linear Regression with One Regressor

Summary of Sampling Distribution

- $\hat{\beta}_1$ is unbiased: $E(\hat{\beta}_1) = \beta_1$ – just like $\bar{Y}$!
- var($\hat{\beta}_1$) is inversely proportional to *n* – just like $\bar{Y}$!
  - The exact sampling distribution is complicated – it depends on the population distribution of (*Y*, *X*) – but when *n* is large we get some simple (and good) approximations.
  - The larger the variance of *X*, the smaller the variance of $\hat{\beta}_1$

Linear Regression with One Regressor



- The number of black and blue dots is the same. Using which would you get a more accurate regression line?

Linear Regression with One Regressor

Hypothesis Tests and Confidence Intervals

Outline

- 1. The standard error of $\hat{\beta}_1$
- 2. Hypothesis tests concerning β_1
- 3. Confidence intervals for β_1
- 4. Heteroskedasticity and homoskedasticity
- 5. Efficiency of OLS and the Student *t* distribution

Based on Chapter 5. Stock and Watson. “Introduction to Econometrics” 3rd Edition.

Linear Regression with One Regressor

A big picture review of where we are going…

- We want to learn about the slope of the population regression line. We have data from a sample, so there is sampling uncertainty. There are five steps towards this goal:

1. State the population object of interest
2. Provide an estimator of this population object
3. Derive the sampling distribution of the estimator (this requires certain assumptions). In large samples this sampling distribution will be normal by the CLT.
4. The square root of the estimated variance of the sampling distribution is the standard error (SE) of the estimator
5. Use the *SE* to construct *t*-statistics (for hypothesis tests) and confidence intervals.

Linear Regression with One Regressor

- **Object of interest:** β_1 in $Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, \ldots, n$
- **Estimator**: the OLS estimator β_1
- **The Sampling Distribution of** β_1 (three assumption of distribution)

Linear Regression with One Regressor

Hypothesis Testing and the Standard Error of β_1

- The objective is to test a hypothesis, like β_1 = 0, using data –to reach a tentative conclusion whether the (null) hypothesis is correct or incorrect.
  - Null hypothesis can be two-sided: $H_0: \ \beta_1 = \beta_{1,0}$ vs. $H_1: \ \beta_1 \neq \beta_{1,0}$
  - Can be one-sided: $H_0: \ \beta_1 = \beta_{1,0}$ vs. $H_1: \ \beta_1 < \beta_{1,0}$

Linear Regression with One Regressor

- ***General approach***: construct *t*-statistic, and compute *p*-value (or compare to the *N*(0,1) critical value)

$$t = \frac{\text{estimator - hypothesized value}}{\text{standard error of the estimator}}$$

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}$$

where $\beta_{1,0}$ is the hypothesized value under the null.

and $SE(\hat{\beta}_1)$= the square root of an estimator of the variance of the sampling distribution of $\hat{\beta}_1$

$$SE(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2_{\hat{\beta}_1}}$$

Linear Regression with One Regressor

Example: *Test Scores* and *STR*, California data

```
regress testscr str, robust

Regression with robust standard errors               Number of obs =     420
                                                     F(  1,   418) =   19.26
                                                     Prob > F      =  0.0000
                                                     R-squared     =  0.0512
                                                     Root MSE      =  18.581

-------------------------------------------------------------------------
        |               Robust
testscr |      Coef.   Std. Err.       t    P>|t|     [95% Conf. Interval]
--------+----------------------------------------------------------------
    str |  -2.279808   .5194892     -4.39   0.000    -3.300945   -1.258671
  _cons |    698.933   10.36436     67.44   0.000     678.5602    719.3057
-------------------------------------------------------------------------
```

Linear Regression with One Regressor

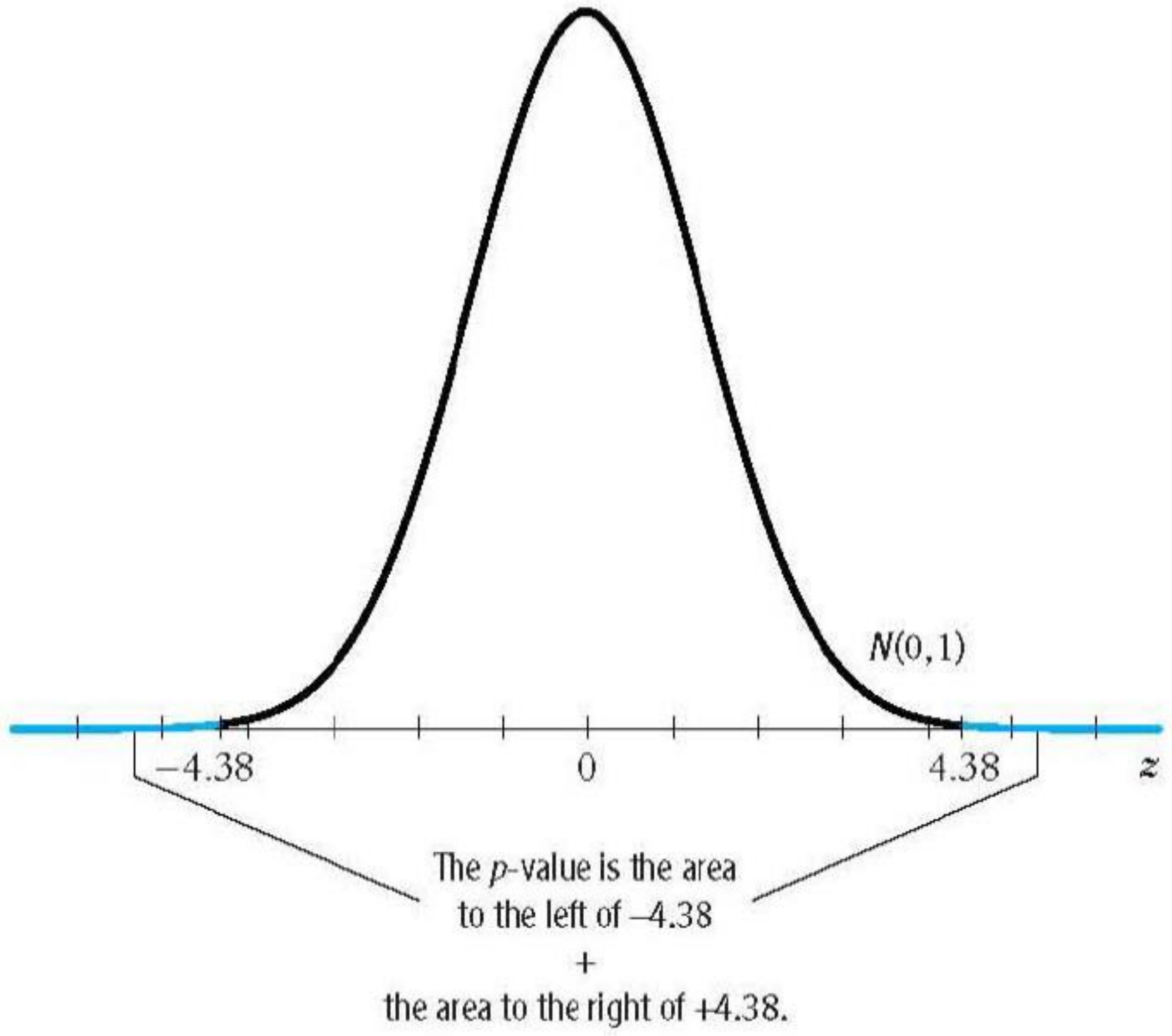
Estimated regression line: *TestScore* = 698.9 – 2.28**STR*

The standard errors: $SE(\hat{\beta}_0) = 10.4$ $SE(\hat{\beta}_1) = 0.52$

t-statistic testing: $\beta_{1,0} = 0 = \dfrac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)} = \dfrac{-2.28 - 0}{0.52} = -4.38$

- What is the 2-sided significance level for a degree of freedom >120 ??? Do we reject the null or not?
- Alternatively, we can compute the *p*-value…

Linear Regression with One Regressor



The *p*-value based on the large-*n* standard normal approximation to the *t*-statistic is 0.00001 (10^{-5})

Linear Regression with One Regressor

Confidence Intervals

- Recall that a 95% confidence is, equivalently:
  - The set of points that cannot be rejected at the 5% significance level;
  - A set-valued function of the data (an interval that is a function of the data) that contains the true parameter value 95% of the time in repeated samples.

$$\text{95\% confidence interval for } \beta_1 = \{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\}$$

Thus:

$$\{\hat{\beta}_1 \pm 1.96 \times SE(\hat{\beta}_1)\} = \{-2.28 \pm 1.96 \times 0.52\}$$
$$= (-3.30, -1.26)$$

Linear Regression with One Regressor

The following two statements are equivalent *(why?)*

- The 95% confidence interval does not include zero;
- The hypothesis $\beta_1 = 0$ is rejected at the 5% level

Linear Regression with One Regressor

A concise (and conventional) way to report regressions:

- Put standard errors in parentheses below the estimated coefficients to which they apply.

$$\widehat{TestScore} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28}\times STR,\ R^2 = .05,\ SER = 18.6$$

Standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$

Linear Regression with One Regressor

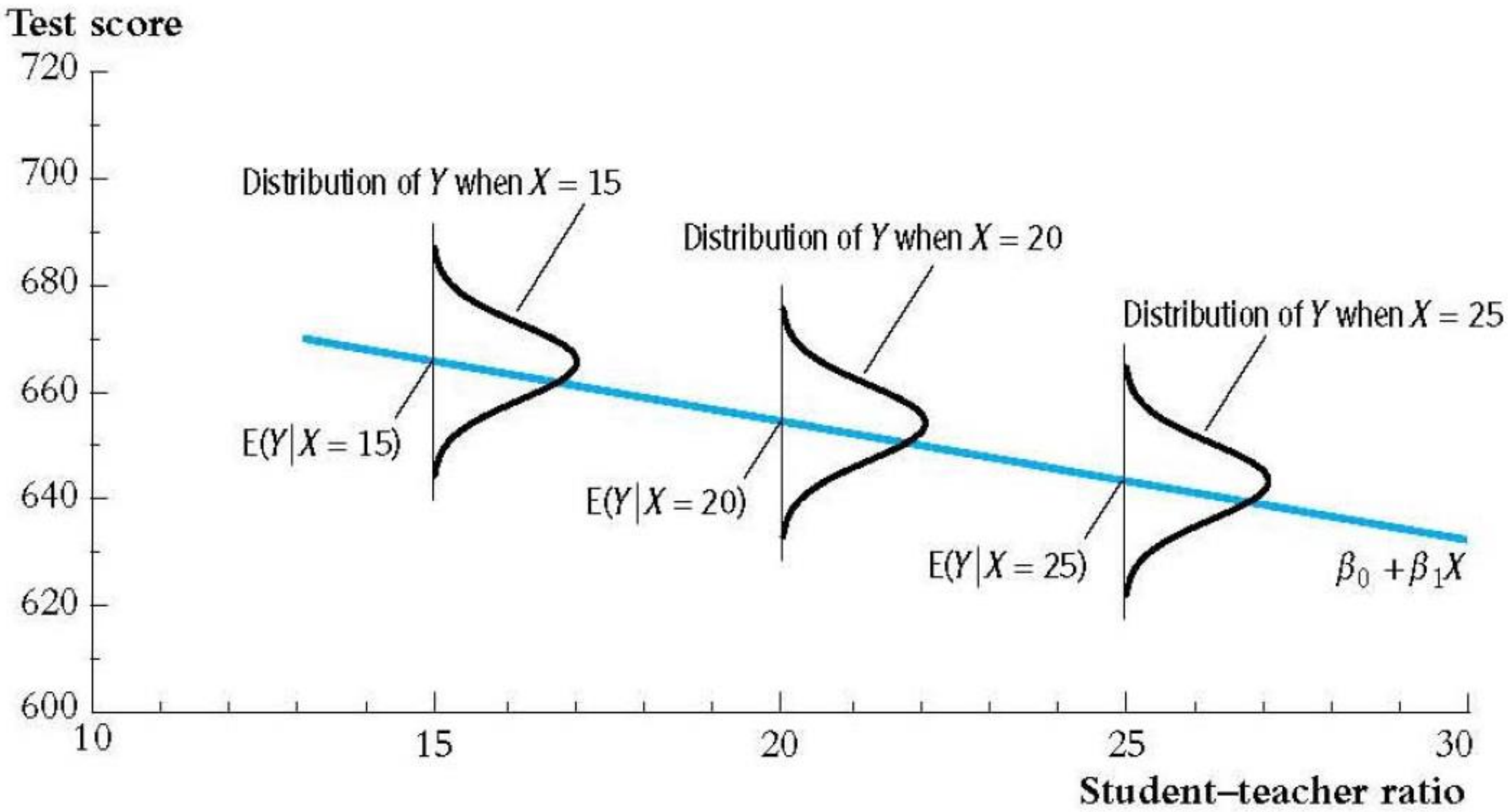
Heteroskedasticity and Homoskedasticity, and Homoskedasticity-Only Standard Errors

What do these two terms mean?

- If var(u|X=x) is constant – that is, if the variance of the conditional distribution of u given X does not depend on X – then u is said to be ***homoskedastic***. Otherwise, u is ***heteroskedastic***.

Linear Regression with One Regressor

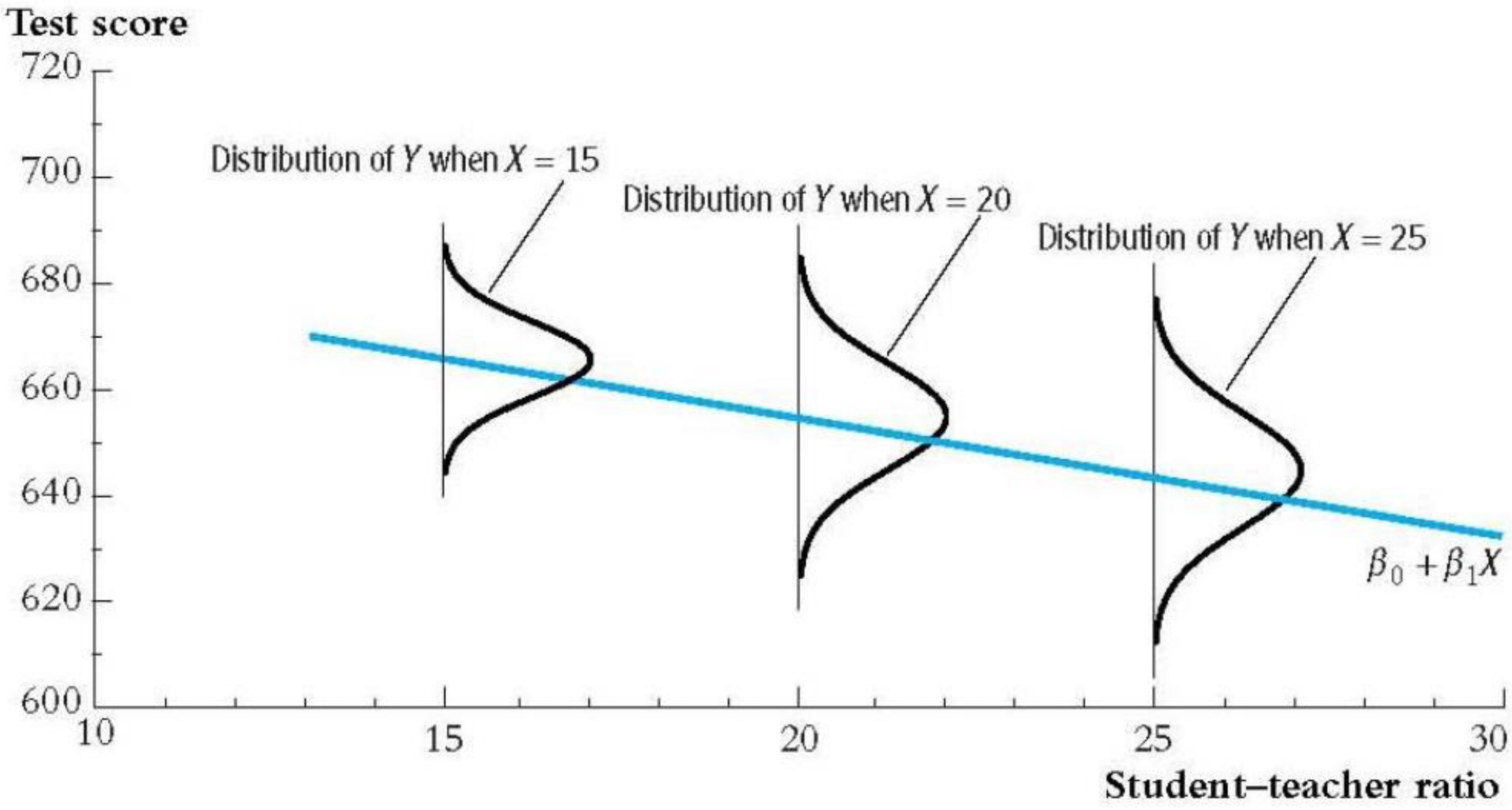
Homoskedasticity *in a picture:*



- $E(u|X=x) = 0$ (u satisfies Least Squares Assumption #1)
- The variance of u ***does not*** depend on x

Linear Regression with One Regressor

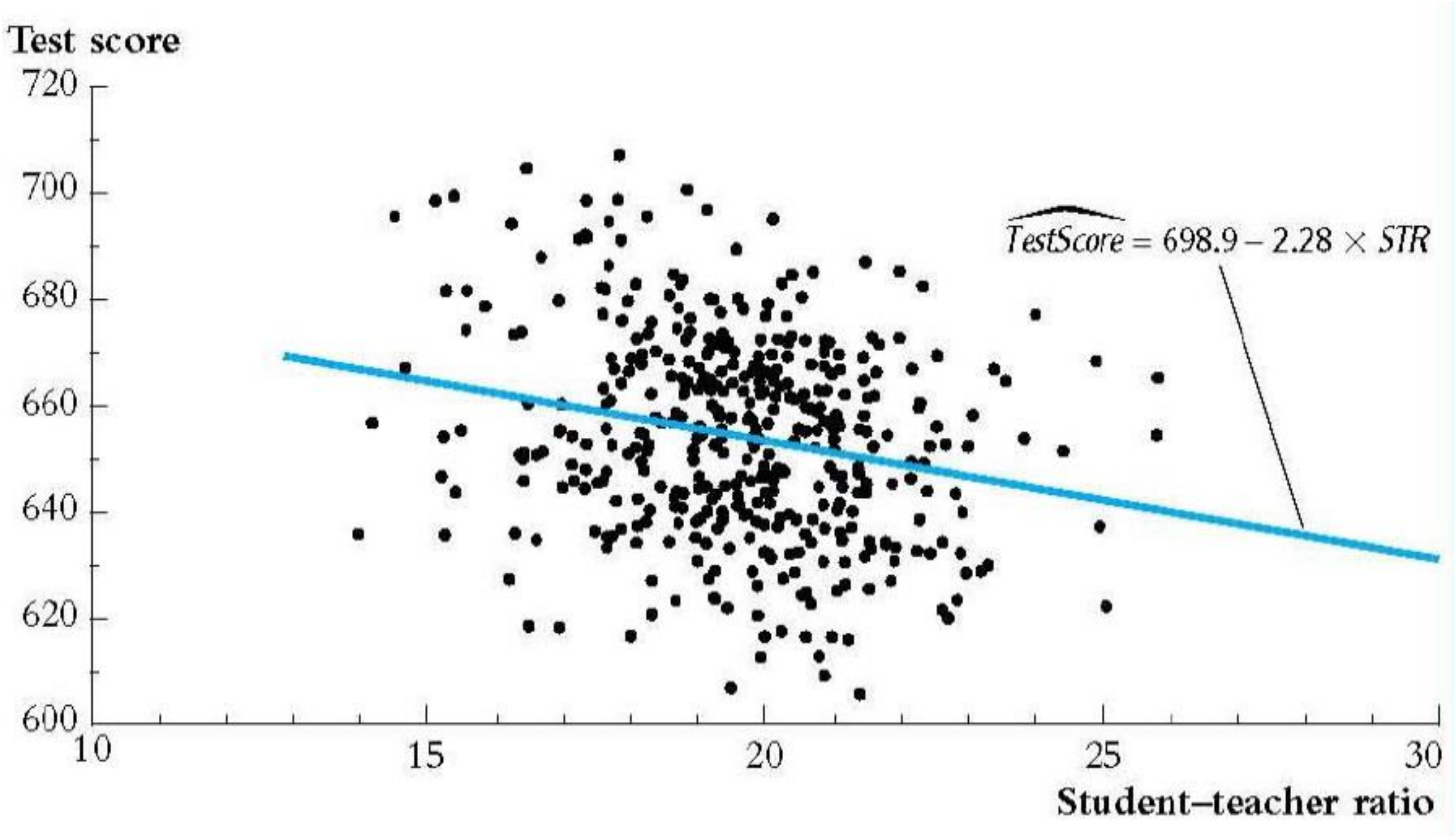
Heteroskedasticity *in a picture:*



- $E(u|X=x) = 0$ (u satisfies Least Squares Assumption #1)
- The variance of u ***does*** depend on x: u is heteroskedastic.

Linear Regression with One Regressor

The class size data:



Heteroskedastic or homoskedastic?

Linear Regression with One Regressor

***What if the errors are in fact homoskedastic*?**

- You can prove that OLS has the lowest variance among estimators that are linear in *Y*… a result called the Gauss-Markov theorem
- *We have formulas for standard errors for* $\hat{\beta}_1$
  - *Homoskedasticity-only standard errors*
  - *Heteroskedasticity – robust standard errors*
- The main advantage of the homoskedasticity-only standard errors is that the formula is simpler. But the disadvantage is that the formula is only correct if the errors are homoskedastic.

Linear Regression with One Regressor

Heteroskedasticity-robust standard errors in STATA

```
regress testscr str, robust

Regression with robust standard errors              Number of obs =     420
                                                    F(  1,   418) =   19.26
                                                    Prob > F      =  0.0000
                                                    R-squared     =  0.0512
                                                    Root MSE      =  18.581
-------------------------------------------------------------------------
        |               Robust
testscr |      Coef.   Std. Err.       t     P>|t|     [95% Conf. Interval]
--------+----------------------------------------------------------------
    str |  -2.279808   .5194892     -4.39   0.000    -3.300945   -1.258671
  _cons |    698.933   10.36436     67.44   0.000     678.5602    719.3057
-------------------------------------------------------------------------
```

- If you use the **", robust"** option, STATA computes heteroskedasticity-robust standard errors
- Otherwise, STATA computes homoskedasticity-only standard errors