Chapter 5

Properties of our Estimators

Terminology: These two things are the same

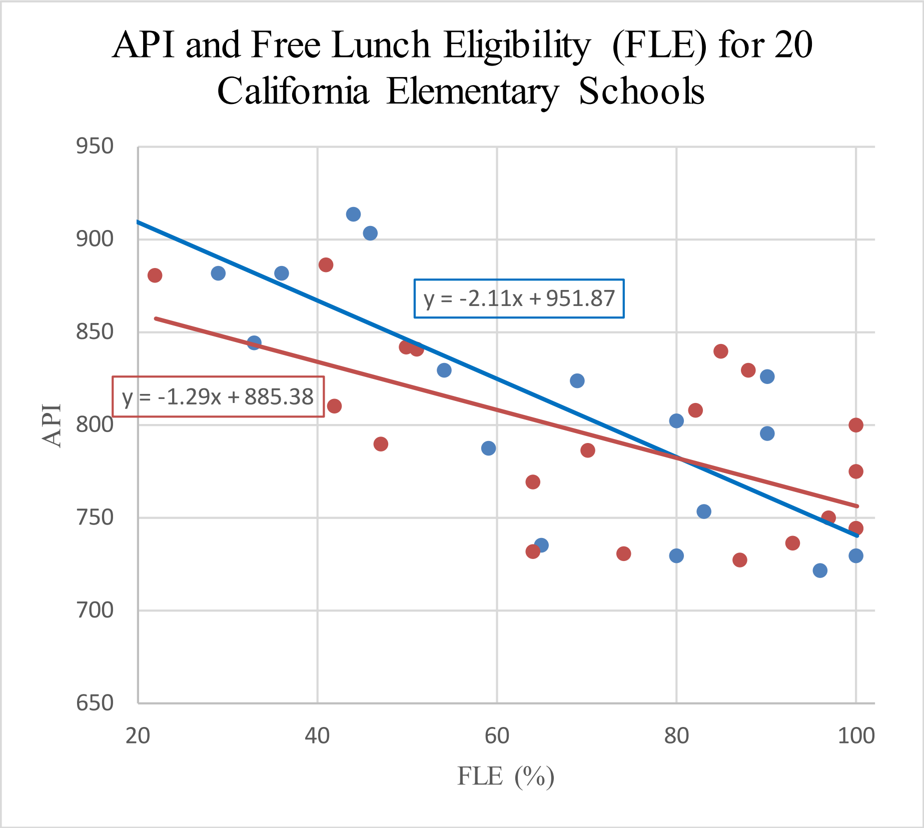
1. “the OLS estimator”

2. b

Learning Objectives

- Demonstrate the concept of sampling error
- Derive the Best, Linear and Unbiased properties of the ordinary least-squares (OLS) estimator
- Develop the formula for the standard error of the OLS coefficient
- Describe the consistency property of the OLS estimator

20 More Schools

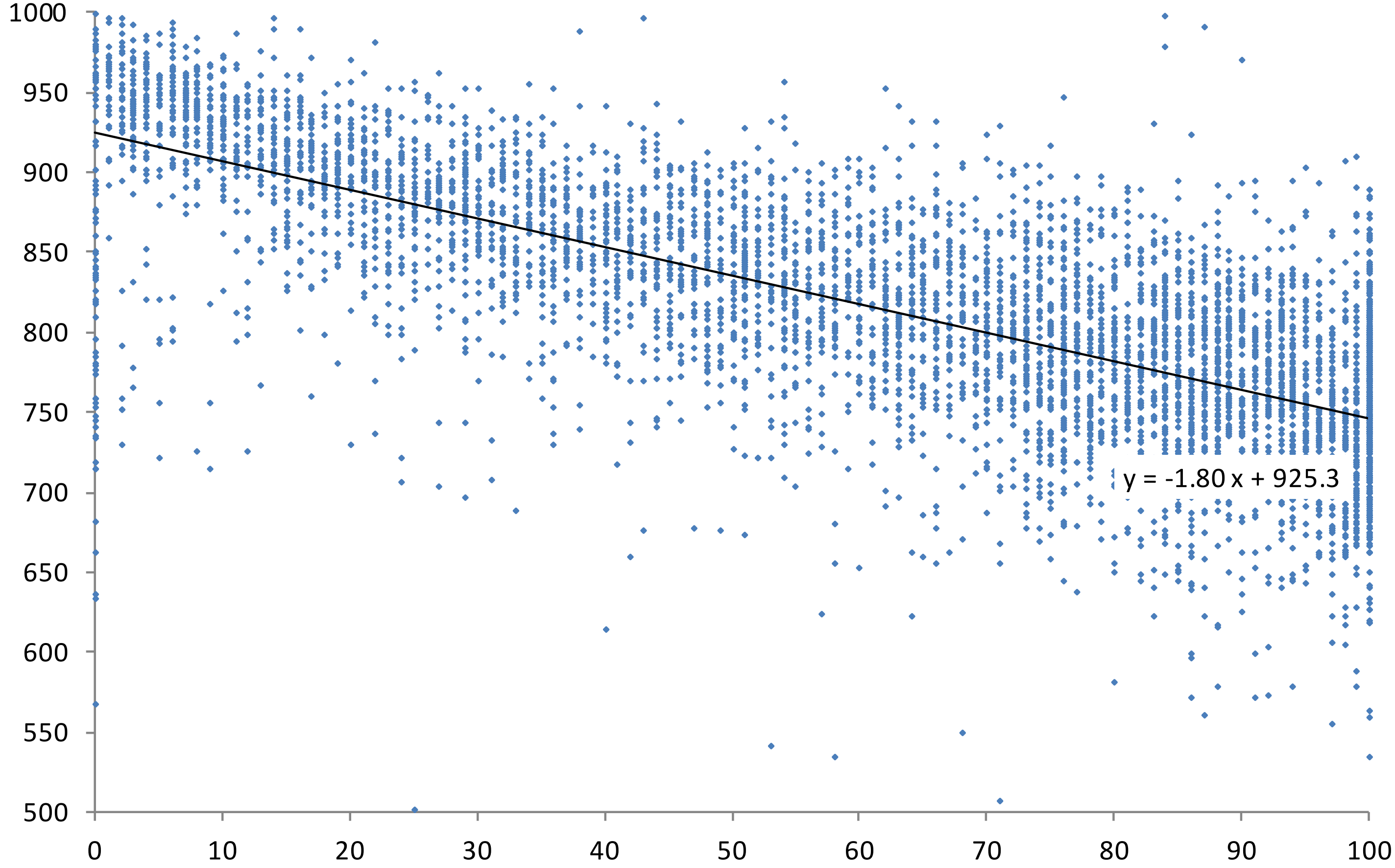


$$API_i = 951.87 - 2.11 FLE_i + e_i$$

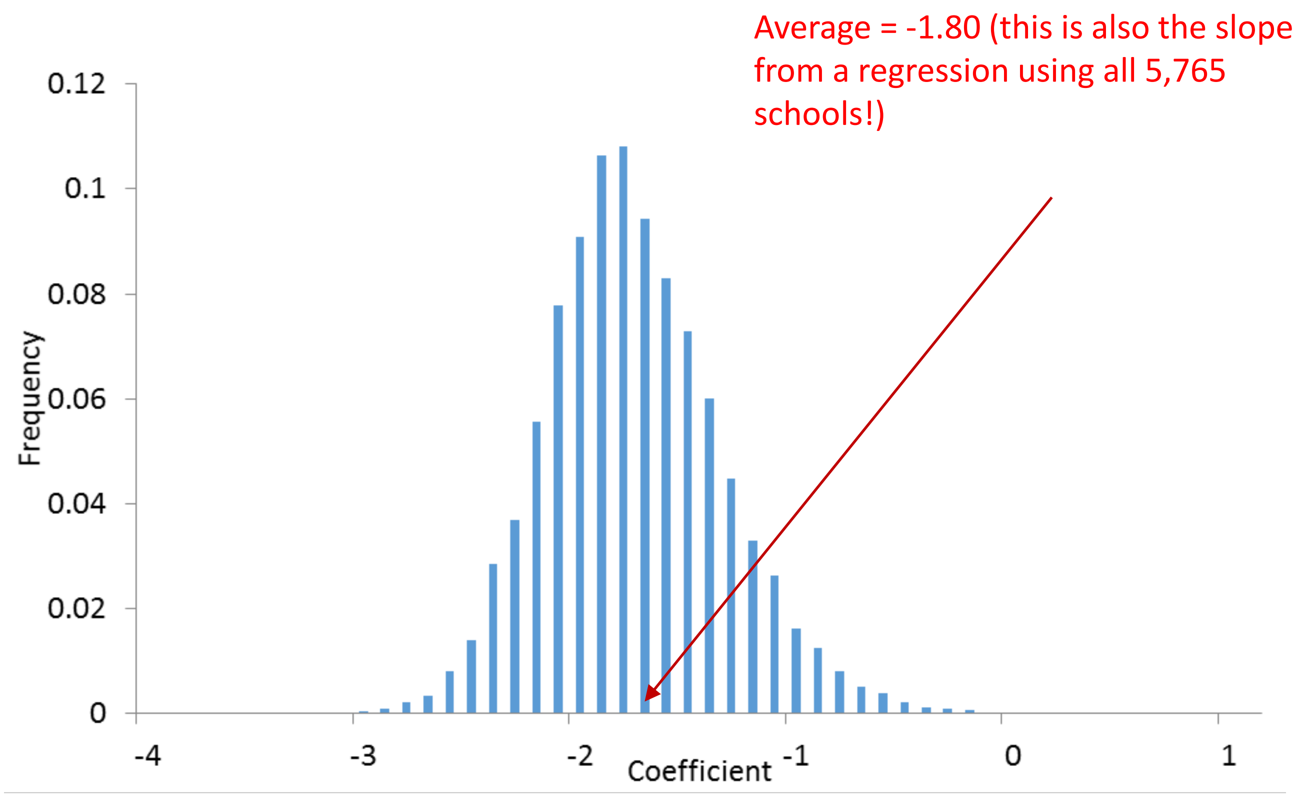
- **This relationship is different for a different 20 schools**
- **What if we draw random samples of 20 schools 10,000 times**
- **…and use them to estimate our model 10,000 times?**

The Population: All 5,765 Schools

$$API_i = 925.3 - 1.80FLE_i + \varepsilon_i$$



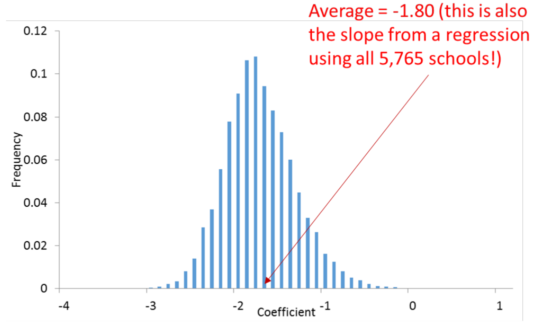
Slope Estimates from 10,000 Random Samples of 20 Schools Each



Most of the time we get close to the true population parameter…but not always! (*On average* we get the true population parameter.)

The Econometric Property of *Unbiasedness* in Action

- Choosing a sample of 20 schools and running a regression gives us an *unbiased* estimate of the population coefficient
  - The average estimate across samples of 20 equals the population value
- We are not systematically overestimating or underestimating the value of the parameter
- The *expected value of our estimate is the population parameter value*
- In this illustration, the key to unbiasedness is *random sampling* from the population



Accuracy: How Close Can We Expect Our Sample Estimate to Be to the Population Parameter?

- The smaller is the standard deviation across samples, the more accurate a given sample estimate is likely to be

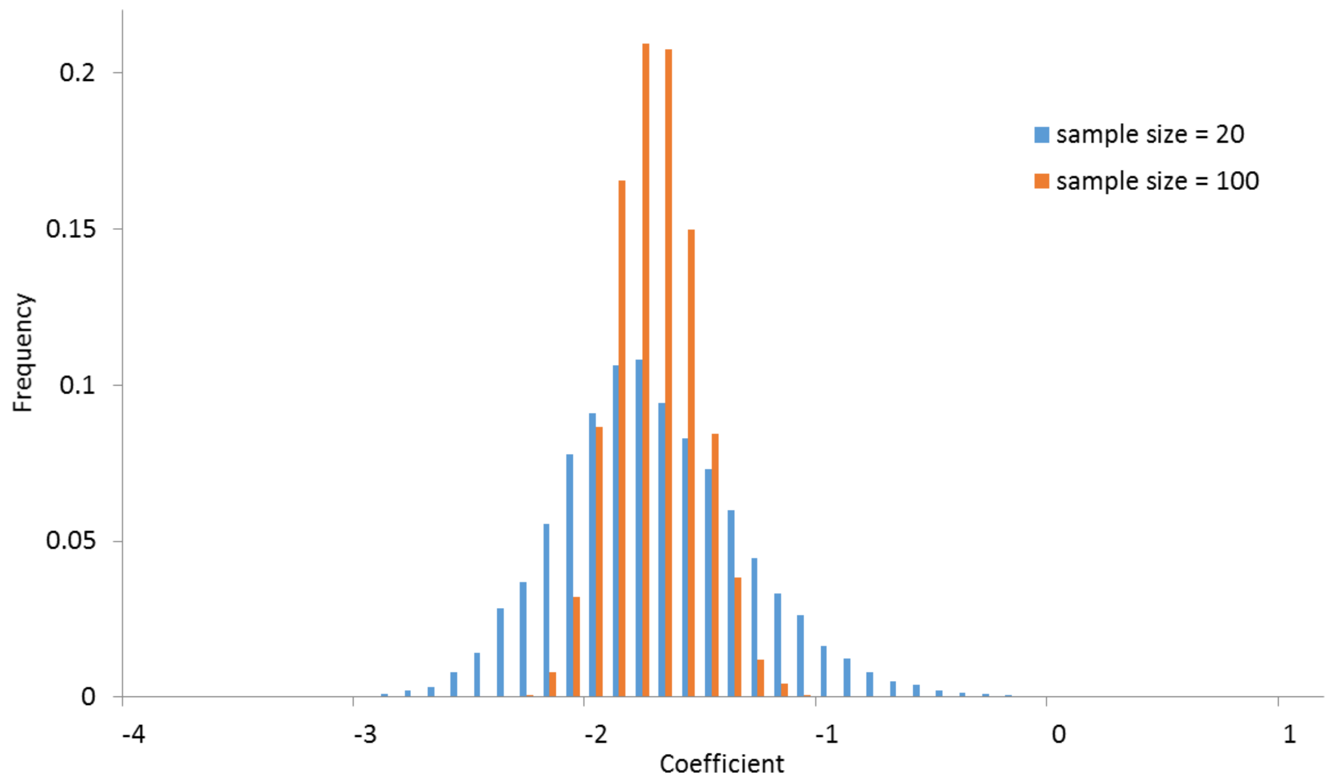
- The standard deviation across samples is known as a standard error

- We want to use statistical procedures that give us the smallest possible standard error

- One way to get a small standard error is to use a large sample.

Larger Samples Generally Give Smaller Standard Errors

If we had used samples of 100 instead of 20 in our experiment, it turns out that we would have obtained a standard error of 0.18 instead of 0.41—and a more accurate estimate.



Another Way to Get a Small Standard Error Is to Use an Efficient Estimation Procedure

- If you want to estimate the population mean, the most efficient (best) linear unbiased estimator (BLUE) is:

$$\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

- If you want to estimate the slope of a regression of *Y* on *X*, the BLUE turns out to be:

$$b_1 = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}$$

Next:

(1) Show these are BLUE

(2) Along the way, estimate Var(b_1), which we use in Ch 6 for hypothesis tests.

Sample Mean Estimator is Linear and Unbiased (Given CR1)

$$E[\bar{Y}] = E\left[\frac{1}{N}\sum_{i=1}^{N} Y_i\right] \quad \text{(by definition)}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E[Y_i] \quad \text{(expectation of sum = sum of expectation)}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \mu \quad \text{(assume CR1)}$$

$$= \frac{N\mu}{N}$$

$$= \mu$$

Deriving the Variance Requires CR1, CR2 & CR3

$$V[\bar{Y}] = V\left[\frac{1}{N}\sum_{i=1}^{N} Y_i\right] \qquad \text{(by definition)}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} V[Y_i] \qquad \text{(assume CR3)}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} \sigma^2 \qquad \text{(assume CR1, CR2)}$$

$$= \frac{1}{N^2} N\sigma^2$$

$$= \frac{\sigma^2}{N}$$

Estimate this with:

$$s_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

$$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{N}\right) \qquad s.e.(\bar{Y}) = \frac{\sigma}{\sqrt{N}}$$

Classical Regression Assumptions

$$\text{Sample Model} \quad Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_K X_{Ki} + e_i$$

$$\text{Population Model} \quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_K X_{Ki} + \varepsilon_i$$

CR1: Representative sample. Our sample is representative of the population we want to say something about.

CR2: Homoscedastic Errors. $Var[\varepsilon_i] = \sigma^2$

The variance of the error is constant over all of our data, i.e., it is homoskedastic.

CR3: Uncorrelated Errors. $Cov[\varepsilon_i, \varepsilon_j] = 0$

The errors are uncorrelated across observations.

CR4: Normally distributed errors. This assumption is only required of small samples.

CR5: Exogenous *X*. The values of the explanatory variable are exogenous, or given (there are no errors in the X-direction). Only required for causality.

OLS Estimator is Linear and Unbiased (Given CR1)

$$b_1 = \sum_{i=1}^{N} w_i Y_i \qquad w_i = \frac{x_i}{\sum_{i=1}^{N} x_i^2}$$

$$E[b_1] = E\left[\sum_{i=1}^{N} w_i Y_i\right]$$

$$= E\left[\sum_{i=1}^{N} w_i(\beta_0 + \beta_1 X_i + \varepsilon_i)\right] \qquad \text{(by definition of } Y_i\text{)}$$

$$= E\left[\beta_0 \sum_{i=1}^{N} w_i + \beta_1 \sum_{i=1}^{N} w_i X_i + \sum_{i=1}^{N} w_i \varepsilon_i\right]$$

$$\sum_{i=1}^{N} w_i = \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} X_i - N\bar{X}}{\sum_{i=1}^{N} x_i^2} = \frac{0}{\sum_{i=1}^{N} x_i^2}$$

$$\sum_{i=1}^{N} w_i X_i = \frac{\sum_{i=1}^{N} x_i X_i}{\sum_{i=1}^{N} x_i^2} = \frac{\sum_{i=1}^{N} x_i x_i}{\sum_{i=1}^{N} x_i^2} = 1$$

So…

$$E[b_1] = E\left[\beta_0 * 0 + \beta_1 * 1 + \sum_{i=1}^{N} w_i \varepsilon_i\right]$$

$$= E[\beta_1] + E\left[\sum_{i=1}^{N} w_i \varepsilon_i\right]$$

$$= \beta_1 + \sum_{i=1}^{N} E[w_i \varepsilon_i] \qquad \text{(expectation of sum = sum of expectation)}$$

$$= \beta_1$$

Remember: X is uncorrelated with ε in the population model!

Deriving Its Variance Requires CR2 & CR3

$$V[b_1] = V\left[\sum_{i=1}^{N} w_i Y_i\right] \qquad \text{(by definition of } b_1\text{)}$$

$$= \sum_{i=1}^{N} V\left[w_i Y_i\right] \qquad \text{(assume CR3)}$$

$$= \sum_{i=1}^{N} w_i^2 V[Y_i] \qquad \text{(we condition on } X\text{, i.e., take it as given)}$$

$$= \sum_{i=1}^{N} w_i^2 V[\varepsilon_i] \qquad \text{(we condition on } X\text{, i.e., take it as given)}$$

$$= \sigma^2 \sum_{i=1}^{N} w_i^2 \qquad \text{(assume CR2)}$$

↓

$$V[b_1] = \sigma^2 \sum_{i=1}^{N} w_i^2 = \sigma^2 \frac{\sum_{i=1}^{N} x_i^2}{\left(\sum_{i=1}^{N} x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}$$

How to estimate the standard error

Standard error is: $$s.e[b_1] = \sqrt{\frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}}$$

Estimate σ² using: $$s^2 = \frac{1}{N-K-1}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2 = \frac{Sum\ of\ \ squared\ errors}{N-K-1}$$

Estimated std error: $$est\ s.e.[b_1] = \frac{s}{\sqrt{\sum_{i=1}^{N} x_i^2}}$$

How to Get a Smaller Standard Error

$$s.e.[b_1] = \frac{\sigma}{\sqrt{\sum_{i=1}^{N} x_i^2}}$$

- Sample drawn from a population with a smaller σ
- Larger sample size (so denominator sums over more observations)
- Lots of variation in *X* (so big x_is—Why?)
- Use an efficient estimator (formula) to compute *b*

Properties of OLS

1. If CR1 holds, then *b* is unbiased (i.e., E[*b*] = β)

2. If CR1-CR3 hold, then *b* is BLUE (i.e., smaller standard error than any other linear unbiased estimator)

3. If CR1-CR3 hold, then the standard error formula is

$$s.e.[b_1] = \frac{\sigma}{\sqrt{\sum_{i=1}^{N} x_i^2}}$$

4. If CR1-CR4, then *b* is BUE (i.e., has a smaller standard error than any other unbiased estimator)

Showing the "Best" in "BLUE"

(N = 2; X_1 is person 1's income and X_2 is person 2's income)

Which of the following is an unbiased estimator of mean income? (Circle the best answer.)

$\frac{1}{2}X_1 + \frac{1}{2}X_2$ $\frac{1}{4}X_1 + \frac{3}{4}X_2$ Both Neither

Which of the following is the "best" (that is, least-variance) estimator of mean income?

$\frac{1}{2}X_1 + \frac{1}{2}X_2$ $\frac{1}{4}X_1 + \frac{3}{4}X_2$ Both Neither

Generally, for the class of linear unbiased estimator (LUE) of the mean:

$$\widetilde{X} = \sum_{i=1}^{N} c_i X_i$$

…only the weights $c_i = 1/N$ make it BLUE. That's why $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Same Logic Applies to OLS

$$b_1 = \sum_{i=1}^{N} w_i Y_i$$

- The OLS estimator has the lowest standard error of all linear unbiased estimators. Given assumptions CR1-CR3, it is BLUE.
- This is the **Gauss-Markov Theorem**.
- **If CR2 or CR3 breaks down, we lost the “B” in “BLUE”**
- **If CR1 is violated, we lose the “U” too!**

Multiple Regression

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_K X_{Ki} + e_i$$

$$E[b_k] = \beta_k \qquad \text{and} \qquad s.e.[b_k] = \frac{\sigma}{\sqrt{\sum_{i=1}^{N} v_{ki}^2}}$$

v_k is the part of variable X_k that is *not* correlated with any of the other right-hand variables, i.e., the residual from a regression of X_k on all the other explanatory variables

For two RHS variables:

$$X_{1i} = c_0 + c_1 X_{2i} + v_{1i}$$

Consistent Estimator

- Consistency means that the estimator is more likely to be close to the population parameter as the sample size increases

- Only need to assume CR1.

- When not consistent?
  - Silly estimator (estimate average height and count all the tall people twice)

  - Weird distribution (estimate average wages but some people have zero hours worked, so divide by zero)

  - CR1 fails (increasing the sample size will not improve your estimator if you are not using representative data)

Distribution of OLS Estimator

Assuming CR1, CR2, CR3:

$$b_1 \sim \left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \right) \quad \text{or} \quad \frac{b_1 - \beta_1}{s.e.[b_1]} \sim (0,1)$$

Assuming CR1, CR2, CR3, and CR4:

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \right) \quad \text{or} \quad \frac{b_1 - \beta_1}{s.e.[b_1]} \sim N(0,1)$$

In Ch. 6, we will see that distribution of b_1 is approximately normal if sample size is large, even when we don't assume CR4.

Features of the Sample OLS Regression

(you can check these on a spreadsheet or in Stata)

(1) The least-squares regression residuals sum to zero

(2) The actual and predicted values of Y_i have the same mean, since residuals sum to zero and $Y_i = \hat{Y}_i + e_i$

(3) The sample residuals are uncorrelated with X_i (nothing more in Y_i can be explained by X_i or by a linear function of X_i)

(4) The covariance between predicted $\hat{Y}_i$ and the residuals is zero; this follows from (3)

What We Learned

- An unbiased estimator gets the right answer in an average sample.
- Larger samples produce more accurate estimates (smaller standard error) than smaller samples.
- Under assumptions CR1-CR3, OLS is the best, linear unbiased estimator —it is BLUE.
- We can use our sample data to estimate the accuracy of our sample coefficient as an estimate of the population coefficient.
- Consistency means that the estimator will get the right answer if applied to the whole population