Section 3. Simple Regression – Confidence Intervals

1. **Review: Least Squares Estimators**
2. **OLS assumptions**
3. **Sampling distributions of OLS estimators**
4. **Standard Errors**
5. **Confidence Intervals for β_0, β_1**
6. **CLT demonstrations using Stata**

1. Review: OLS estimators

The OLS Estimator, Predicted Values, and Residuals

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

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{s_{XY}}{s_X^2} \tag{4.8}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X}. \tag{4.9}$$

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

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,\ i = 1, \ldots, n \tag{4.10}$$

$$\hat{u}_i = Y_i - \hat{Y}_i,\ i = 1, \ldots, n. \tag{4.11}$$

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).

2. OLS Assumptions

The Least Squares Assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, \ldots, n, \text{ where:}$$

1. The error term u_i has conditional mean zero given X_i, that is, $E(u_i \mid X_i) = 0$;
2. (X_i, Y_i), $i = 1, \ldots, n$ are independent and identically distributed (i.i.d.) draws from their joint distribution; and
3. (X_i, u_i) have nonzero finite fourth moments.

3. Sampling Properties of OLS estimators

- Unbiased: $E(b_0) = \boldsymbol{\beta}_0$, $E(b_1) = \boldsymbol{\beta}_1$
- Consistent: $\text{plim}(b_0) = \boldsymbol{\beta}_0$, $\text{plim}(b_1) = \boldsymbol{\beta}_1$
- Asymptotically Normal by the CLT

Derivation

Derivation

Sampling Properties of OLS estimators (cont.)

Large-Sample Distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$

If the least squares assumptions in Key Concept 4.3 hold, then in large samples $\hat{\beta}_0$ and $\hat{\beta}_1$ have a jointly normal sampling distribution. The large-sample normal distribution of $\hat{\beta}_1$ is $N(\beta_1, \sigma^2_{\hat{\beta}_1})$, where the variance of this distribution, $\sigma^2_{\hat{\beta}_1}$, is

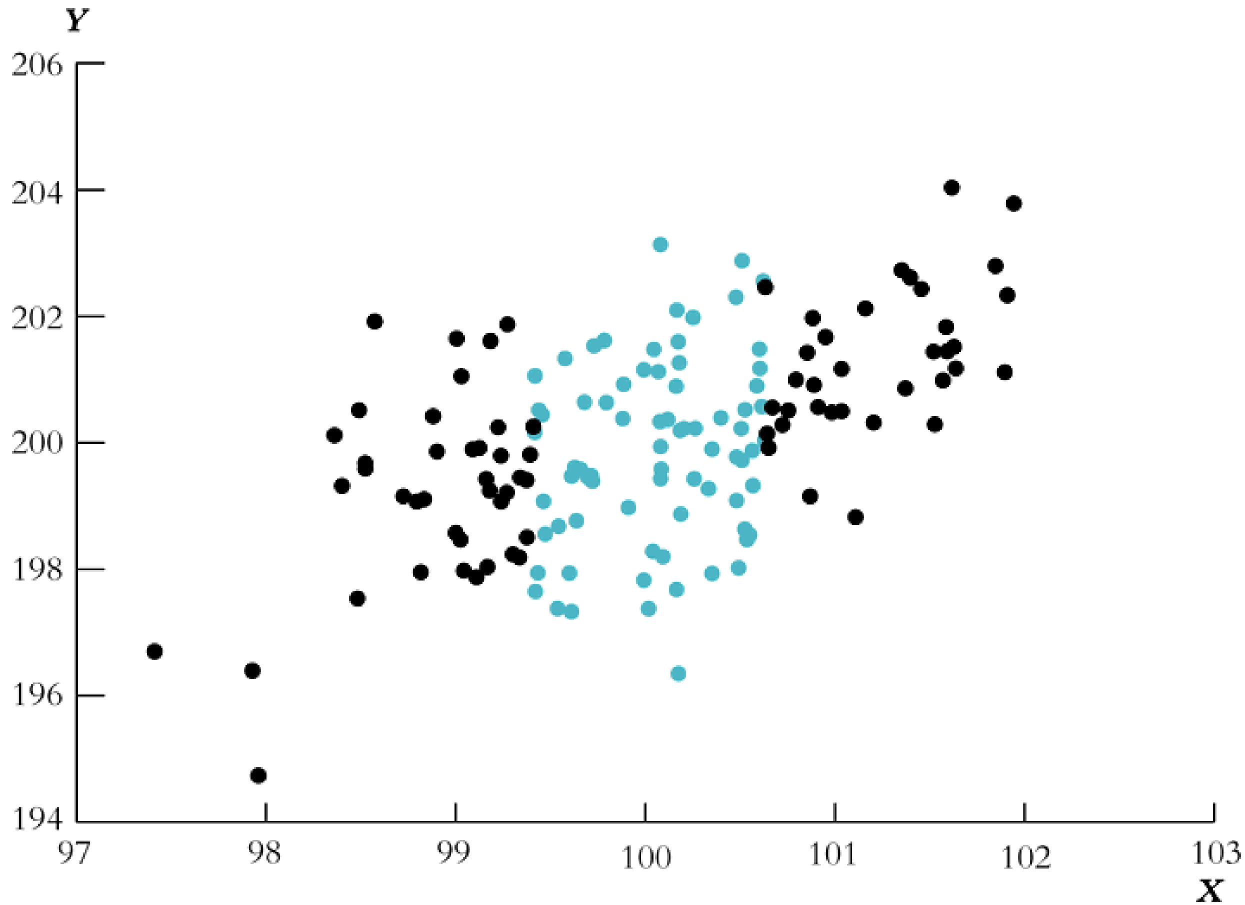
$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{\text{var}[(X_i - \mu_X)u_i]}{[\text{var}(X_i)]^2}. \quad (4.14)$$

The large-sample normal distribution of $\hat{\beta}_0$ is $N(\beta_0, \sigma^2_{\hat{\beta}_0})$, where

$$\sigma^2_{\hat{\beta}_0} = \frac{1}{n}\frac{\text{var}(H_i u_i)}{[E(H_i^2)]^2}, \quad \text{where } H_i = 1 - \left(\frac{\mu_X}{E(X_i^2)}\right)X_i. \quad (4.15)$$

FIGURE 4.5 The Variance of $\hat{\beta}_1$ and the Variance of X

The colored dots represent a set of X_i's with a small variance. The black dots represent a set of X_i's with a large variance. The regression line can be estimated more accurately with the black dots than with the colored dots.



4. Standard Errors

- CLT: b_1 approx ~ $N(\beta_1, V(b_1))$
  b_0 approx ~ $N(\beta_0, V(b_0))$

- **"Standard Errors" are consistent estimators of standard deviations of b_o and b_1 . (S&W p. 133, 151,180).**
- **So "t-statistics" have a standard normal distribution.**

5. Confidence Intervals

Confidence Intervals for β_1

A 95% two-sided confidence interval for β_1 is an interval that contains the true value of β_1 with a 95% probability, that is, it contains the true value of β_1 in 95% of all possible randomly drawn samples. Equivalently, it is also the set of values of β_1 that cannot be rejected by a 5% two-sided hypothesis test. When the sample size is large, it is constructed as

$$95\% \text{ confidence interval for } \beta_1 = (\hat{\beta}_1 - 1.96SE(\hat{\beta}_1), \hat{\beta}_1 + 1.96SE(\hat{\beta}_1)). \quad (4.27)$$

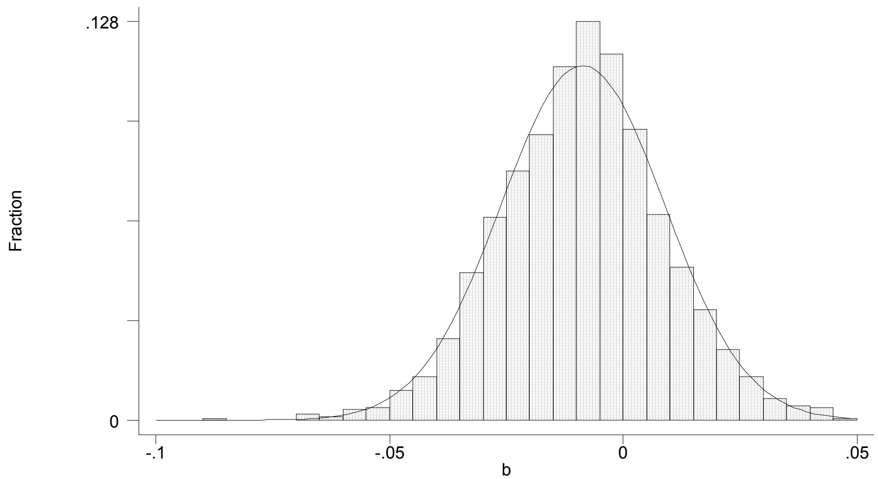
..and the same for $\boldsymbol{\beta_0}$

6. CLT demonstrations using Stata

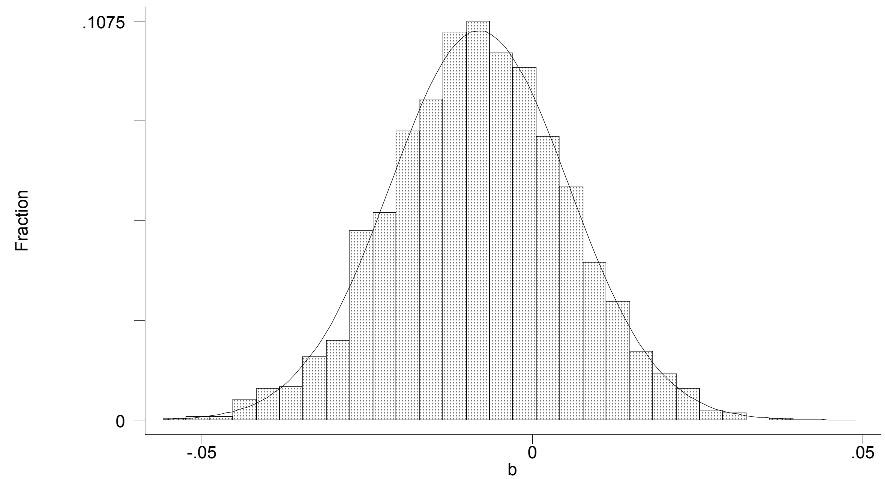
- Stata doesn't mind running a regression
  a few thousand times,
  - which allows us to observe a
  sampling distribution for $\mathbf{b_1}$
  e.g., bootregh00.do

..and the same for $\mathbf{b_0}$

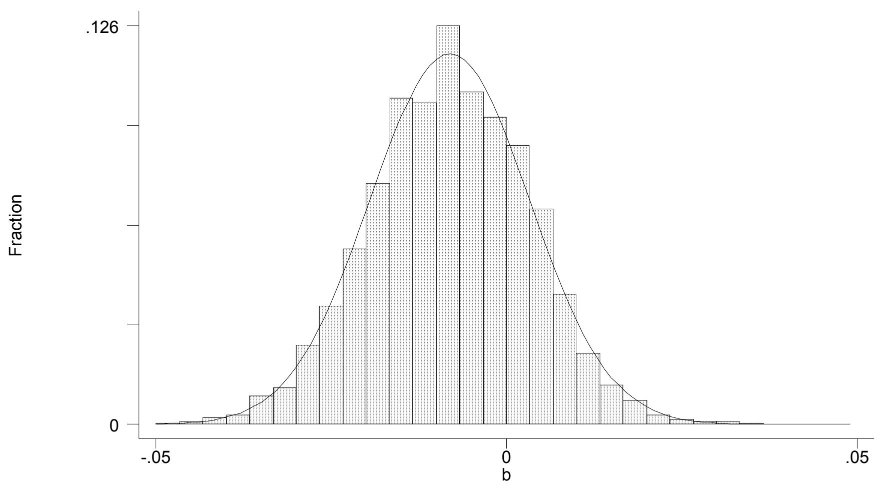
CLT in action: sampling distributions for b_1



n=20, 2000 reps



n=30, 2000 reps



n=40, 2000 reps

Note: In developing countries this slope is about -0.2 children per year of education. Vg

Next time..

- Omitted variable bias