ACE 562
Fall 2005

Lecture 5: The Simple Linear Regression Model: Sampling Properties of the Least Squares Estimators

by
Professor Scott H. Irwin

Required Reading:

Griffiths, Hill and Judge. "Inference in the Simple Regression Model: Estimator Sampling Characteristics and Properties," Ch. 6 in *Learning and Practicing Econometrics*

Optional Reading:

Kennedy. "Appendix A: Sampling Distributions: The Foundation of Statistics," in *A Guide to Econometrics* (Ag Library reserve)

Overview

In the previous section, we specified a linear economic model that led to the following statistical model, known as the simple linear regression model

$$y_t = \beta_1 + \beta_2 x_t + e_t$$

We assumed that e_t and y_t are independent and identically distributed normal random variables

In compact form, the assumptions of the simple linear regression model are,

SR1. $y_t = \beta_1 + \beta_2 x_t + e_t, \quad t = 1,\ldots,T$

SR2. $E(e_t) = 0 \Leftrightarrow E(y_t) = \beta_1 + \beta_2 x_t$

SR3. $\text{var}(e_t) = \text{var}(y_t) = \sigma^2$

SR4. $\text{cov}[e_t, e_s] = \text{cov}[y_t, y_s] = 0 \qquad t \neq s$

SR5. The variable x_t is not random and must take on at least two different values

SR6. $e_t \sim N(0, \sigma^2) \Leftrightarrow y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$

One sample of data was obtained consisting of observations on food expenditure and income for forty households

⇒We assumed that the sample data was generated by the previous statistical model

Given this sample of data, we developed the following rules (estimators) for estimating the intercept and slope parameters of the (true, but unknown) linear statistical model

$$b_2 = \frac{T\sum_{t=1}^{T} y_t x_t - \sum_{t=1}^{T} x_t \sum_{t=1}^{T} y_t}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}$$

$$b_1 = \bar{y} - b_2 \bar{x}$$

Using the sample data and least squares estimators, we computed the following least squares estimates of the unknown intercept and slope for the statistical model

$$b_1 = 7.3832 \qquad b_2 = 0.2323$$

At this point, the estimates are simply computed numbers that have no statistical properties!

We can never know how close these particular numbers are to the true values we are trying to estimate

While we cannot know the accuracy of the least squares estimates, we can examine the properties of the estimator under repeated sampling

- As before, we imagine "hitting" the estimator with many hypothetical samples and examining its performance across the samples

Repeated sampling in the food expenditure example can be thought of as setting income levels to be the same across samples, so that we just randomly select new households for the given levels of income

- In effect, assume that we can perform a controlled experiment where the set of values for x_t are fixed across repeated samples, but the values for y_t vary randomly

Applying the least squares estimation rules to each new (hypothetical) sample leads to different b_1 and b_2 estimates

Consequently, the least squares estimation rules b_1 and b_2 are random variables

Viewing b_1 and b_2 as random variables leads to the following important questions

- What are the means, variances, covariances, and forms of the sampling distributions for the random variables b_1 and b_2?

- Since the least squares estimators are only one way of using the sample data to obtain estimates of the unknown parameters β_1 and β_2, how well do the least squares estimators compare to other estimators in repeated sampling?

If someone proposes an estimation rule for a particular statistical model, your next question should be: What are its sampling characteristics and how good is it?

---Griffiths, Hill and Judge, *LPE*, p.209

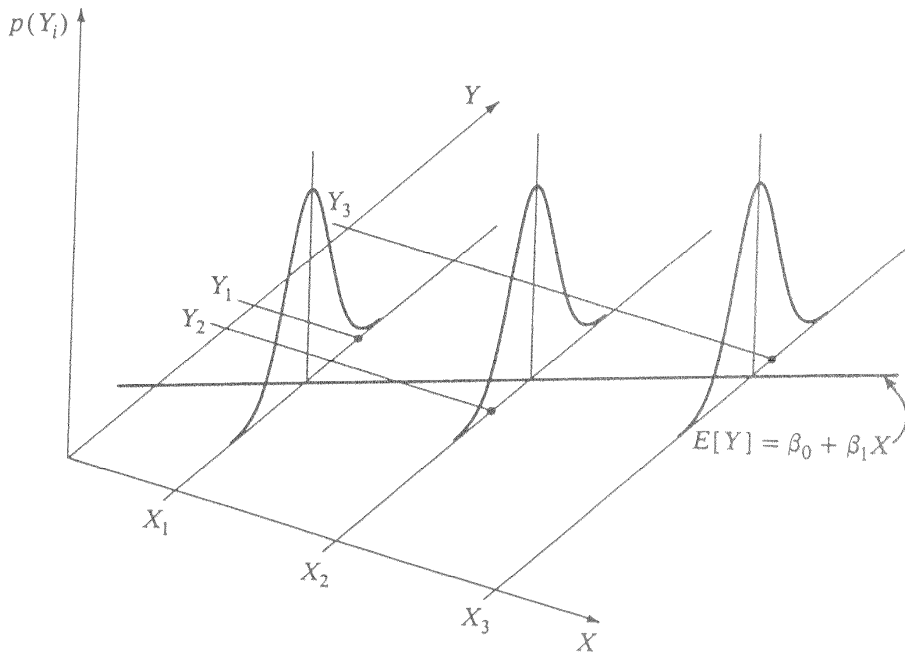


FIGURE 11.2 In the normal regression model, the value of Y for each observation is considered to be an outcome of a separate normal probability distribution. Each distribution has a mean of $\beta_0 + \beta_1 X_i$ and a standard deviation of σ_u as in Figure 11.1. In this three-dimensional diagram, probability density values are drawn vertically. The set of values labeled here as Y_1, Y_2, and Y_3 are thought of as just one possible set of values for Y corresponding to the three given values of X.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

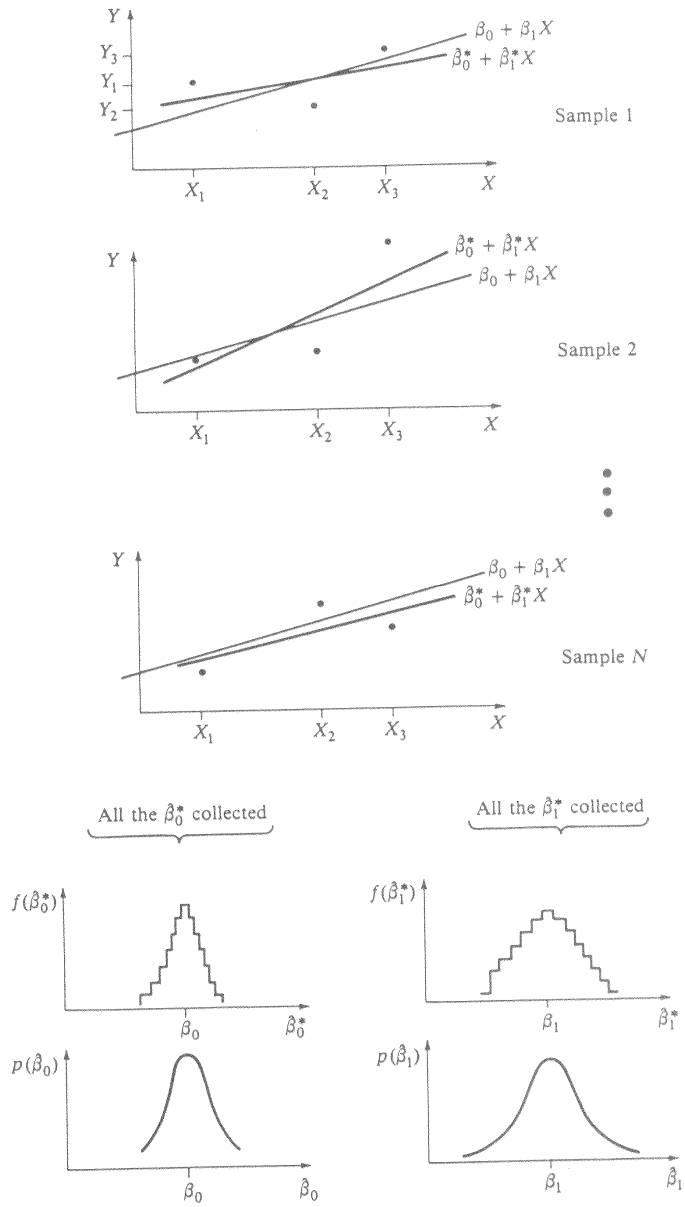


FIGURE 11.3 A thought experiment helps explain the nature of the sampling distributions of the OLS estimators of β_0 and β_1. The same three X values are used in each sample, but different sets of Y values are produced because different values for the disturbances occur in each sample. The $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ values computed in each sample are collected, and their frequency distributions are constructed. The sampling distribution of an estimator is the limiting form of the frequency distribution as N approaches infinity. The actual $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ that we calculate from a set of data are thought of as just one pair of outcomes from these sampling distributions.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

Means, Variances and Covariances of b_1 and b_2

The sampling distributions shown on the previous page can be developed theoretically

We will concentrate on the derivation of two important properties of the sampling distribution of b_2: mean and variance

- The derivations for the mean and variance of b_1 and the covariance of b_1 and b_2 are similar, so there is no need to repeat the process

To develop the desired formulas for the mean and variance of b_2, it is helpful to first derive a new version of the least squares formula for b_2

We start by re-stating the original version,

$$b_2 = \frac{T\sum_{t=1}^{T} y_t x_t - \sum_{t=1}^{T} x_t \sum_{t=1}^{T} y_t}{T\sum_{t=1}^{T} x_t^2 - \left(\sum_{t=1}^{T} x_t\right)^2}$$

The previous formula can be re-written as,

$$b_2 = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$$

Simply multiplying out the right hand term of the numerator, we get,

$$b_2 = \frac{\sum_{t=1}^{T}(x_t - \overline{x})(y_t - \overline{y})}{\sum_{t=1}^{T}(x_t - \overline{x})^2} = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t - \sum_{t=1}^{T}(x_t - \overline{x})\overline{y}}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$$

or,

$$b_2 = \frac{\sum_{t=1}^{T}(x_t - \overline{x})y_t - \overline{y}\sum_{t=1}^{T}(x_t - \overline{x})}{\sum_{t=1}^{T}(x_t - \overline{x})^2}$$

Since, $\sum_{t=1}^{T}(x_t - \overline{x}) = 0$, this can be simplified to

$$b_2 = \frac{\sum_{t=1}^{T}(x_t - \bar{x})y_t}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

Next,

$$b_2 = \sum_{t=1}^{T} w_t y_t$$

where,

$$w_t = \frac{(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

Finally, substitute the original formula for the statistical model, $y_t = \beta_1 + \beta_2 x_t + e_t$, for y_t,

$$b_2 = \sum_{t=1}^{T} w_t y_t = \sum_{t=1}^{T} w_t(\beta_1 + \beta_2 x_t + e_t)$$

or,

$$b_2 = \beta_1 \sum_{t=1}^{T} w_t + \beta_2 \sum_{t=1}^{T} w_t x_t + \sum_{t=1}^{T} w_t e_t$$

and,

$$b_2 = \beta_2 + \sum_{t=1}^{T} w_t e_t$$

since $\sum_{t=1}^{T} w_t = 0$ and $\sum_{t=1}^{T} w_t x_t = 1$

This new formula for the least squares estimator is quite valuable in deriving properties of b_2

Mean

The mean is derived by taking the expectation of b_2,

$$E(b_2) = E\left(\beta_2 + \sum_{t=1}^{T} w_t e_t\right)$$

$$E(b_2) = E(\beta_2) + \sum_{t=1}^{T} w_t E(e_t)$$

$$E(b_2) = \beta_2$$

This shows that the mean of the sampling distribution of b_2 is β_2, the population slope parameter

What happens if our statistical model is not correctly specified?

In this case, omitting an important variable will make

$$E(e_t) \neq 0$$

and

$$E(b_2) = \beta_2 + \sum_{t=1}^{T} w_t E(e_t)$$

If β_2 is positive and the omitted variable tends to increase the size of positive errors, then

$$E(b_2) > \beta_2$$

If β_2 is positive and the omitted variable tends to increase the size of negative errors, then

$$E(b_2) < \beta_2$$

Discussion highlights the importance of using economic theory to correctly specify the statistical model

- Specification questions dominate applied econometric work

- We will discuss this issue extensively next semester

Variance

The variance of b_2 in repeated samples is

$$\mathrm{var}(b_2) = E[b_2 - E(b_2)]^2 = E[b_2 - \beta_2]^2$$

Variance measures the precision of b_2 in the sense that it tells us how much the estimates produced by b_2 vary from sample-to-sample

- The lower the variance of an estimator, the greater the sampling precision

- The greater the variance of an estimator, the lower the sampling precision

Key point: An estimator is considered more precise than another estimator if its sampling variance is less than that of another estimator

Econometricians place a high priority on developing precise estimators

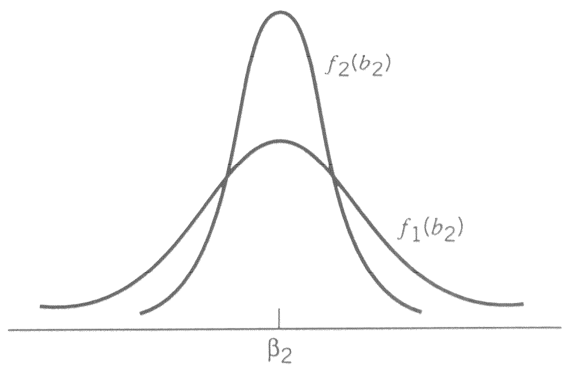


FIGURE **4.1** Two possible probability density functions for b_2

Hill, C., W. Griffiths, and G. Judge. *Undergraduate Econometrics.* John Wiley & Sons, Inc., New York, NY 1997.

We can derive the variance of b_2 as follows,

$$\text{var}(b_2) = \text{var}\left(\beta_2 + \sum_{t=1}^{T} w_t e_t\right) = \text{var}\left(\sum_{t=1}^{T} w_t e_t\right)$$

$$\text{var}(b_2) = \sum_{t=1}^{T} w_t^2 \,\text{var}(e_t) + \sum_{t=1}^{T}\sum_{s=1}^{T} w_t w_s \,\text{cov}(e_t, e_s) \qquad t \neq s$$

$$\text{var}(b_2) = \sum_{t=1}^{T} w_t^2 \,\text{var}(e_t)$$

$$\text{var}(b_2) = \sum_{t=1}^{T} w_t^2 \sigma^2 = \sigma^2 \sum_{t=1}^{T} w_t^2$$

and,

$$\text{var}(b_2) = \sigma^2 \left[\frac{1}{\sum_{t=1}^{T}(x_t - \bar{x})^2}\right]$$

where $\sum_{t=1}^{T} w_t^2 = \dfrac{1}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$

The standard deviation of b_2 is found in the usual manner

$$se(b_2) = \sqrt{\text{var}(b_2)} = \sigma \left[\frac{1}{\sqrt{\sum_{t=1}^{T}(x_t - \bar{x})^2}} \right]$$

Notice that the term "standard error" (se) is normally used in place of standard deviation of the sampling distribution

To see why, define the error that arises in estimating the true slope parameter as $f = b_2 - \beta_2$

Applying our rules for the mean and variance of the transformation of a random variable we find that

$$E(f) = 0 \text{ and } \text{var}(f) = \text{var}(b_2)$$

Since the expected estimation error is zero, we can say that the typical error (without regard to sign) is given by the standard deviation of f, which equals the standard deviation of b_2

If we replace "typical" with "standard" we can say that $se(b_2)$ measures the standard estimation error for b_2, or in abbreviated form, standard error

Form

Now that we have derived the mean and variance of the sampling distribution of b_2, we can turn our attention to the form of the sampling distribution

Earlier, we noted that

$$b_2 = \sum_{t=1}^{T} w_t y_t$$

where,

$$w_t = \frac{(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

Writing the formula in the above format shows that the least squares rule b_2 is a linear function of the y_t

- The y_t are normally distributed (by assumption)
- Any linear function of normally distributed random variables is itself normally distributed

Thus, b_2 is normally distributed in repeated sampling

$$b_2 \sim N\left(\beta_2, \sigma^2\left[\frac{1}{\sum_{t=1}^{T}(x_t-\overline{x})^2}\right]\right)$$

We will simply state the mean and variance results for b_1,

$$E(b_1)=\beta_1$$

$$\mathrm{var}(b_1)=\sigma^2\left[\frac{\sum_{t=1}^{T}x_t^2}{T\sum_{t=1}^{T}(x_t-\overline{x})^2}\right]$$

Using a similar argument as we did for b_2, it can be shown that

$$b_1 \sim N\left(\beta_1, \sigma^2\left[\frac{\sum_{t=1}^{T}x_t^2}{T\sum_{t=1}^{T}(x_t-\overline{x})^2}\right]\right)$$

Covariance

Finally, the covariance between random variables b_1 and b_2 in repeated sampling is,

$$\operatorname{cov}(b_1, b_2) = \sigma^2 \left[\frac{-\bar{x}}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

Probability Calculations Using Sampling Distributions

While it is fun (!) to simply derive sampling distributions, their real usefulness is found in making probability statements about our estimates

Let's suppose that the true regression model is

$$y_t = 8 + 0.25x_t + e_t$$

and all of the standard assumptions hold for the error term and $\sigma^2 = 50$

Now assume that we are "given" a set of $T = 25$ values for the independent variable x_t and

$$\sum_{t=1}^{T}(x_t - \overline{x})^2 = 15{,}000$$

We now have all the information to derive the sampling distribution of b_2

The expected value of b_2 is given in this example as

$$E(b_2) = \beta_2 = 0.25$$

The sampling variance of b_2 is

$$\text{var}(b_2) = \sigma^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right] = 50 \cdot \frac{1}{15{,}000} = 0.0033$$

Putting the two together with our normality assumption

$$b_2 \sim N(0.25, 0.0033)$$

Suppose we want to determine the probability that our estimate from a single sample is between 0.20 and 0.30

$$\Pr(0.20 \leq b_2 \leq 0.30) = ?$$

We can transform this into a standard normal probability statement to obtain the answer

$$\Pr\left(\frac{0.20 - 0.25}{\sqrt{0.0033}} \leq \frac{b_2 - 0.25}{\sqrt{0.0033}} \leq \frac{0.30 - 0.25}{\sqrt{0.0033}} \right) = ?$$

$$\Pr(-0.8660 \leq Z \leq 0.8660) \approx 0.6158$$

Factors That Affect the Variances and Covariance of the Sampling Distributions of b_1 and b_2

$$\text{var}(b_1) = \sigma^2 \left[\frac{\sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2} \right] \qquad \text{var}(b_2) = \sigma^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

$$\text{cov}(b_1, b_2) = \sigma^2 \left[\frac{-\bar{x}}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

Important observations:

1. The variance of the error term (σ^2) appears in each formula

   - The larger the variance, σ^2, the greater the uncertainty about where the values of y_t will fall relative to the population mean, $E(y_t)$

     $\Rightarrow$ Sample information available to estimate β_1 and β_2 is less precise the larger is σ^2

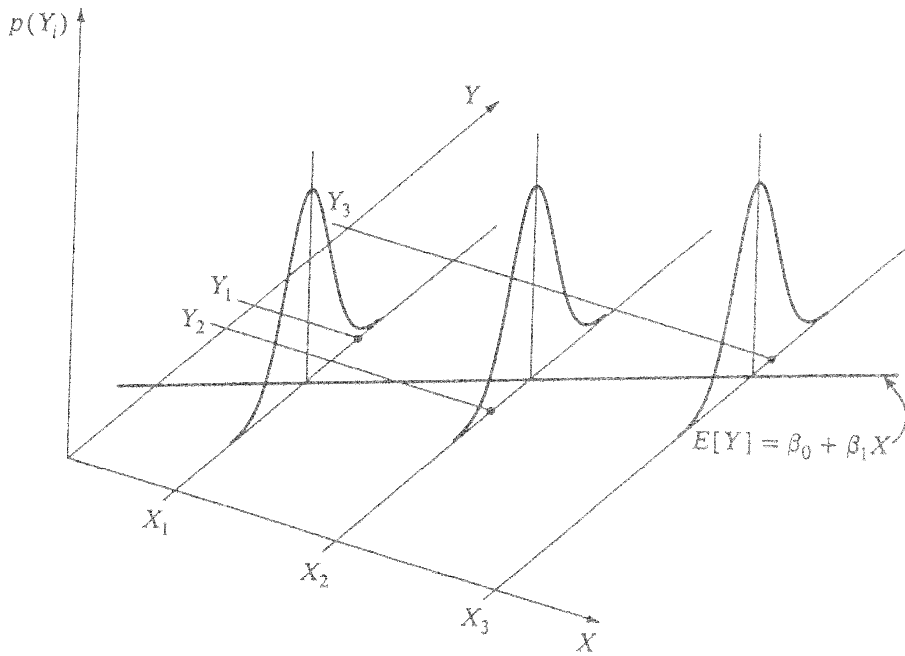


FIGURE 11.2 In the normal regression model, the value of Y for each observation is considered to be an outcome of a separate normal probability distribution. Each distribution has a mean of $\beta_0 + \beta_1 X_i$ and a standard deviation of σ_u as in Figure 11.1. In this three-dimensional diagram, probability density values are drawn vertically. The set of values labeled here as Y_1, Y_2, and Y_3 are thought of as just one possible set of values for Y corresponding to the three given values of X.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

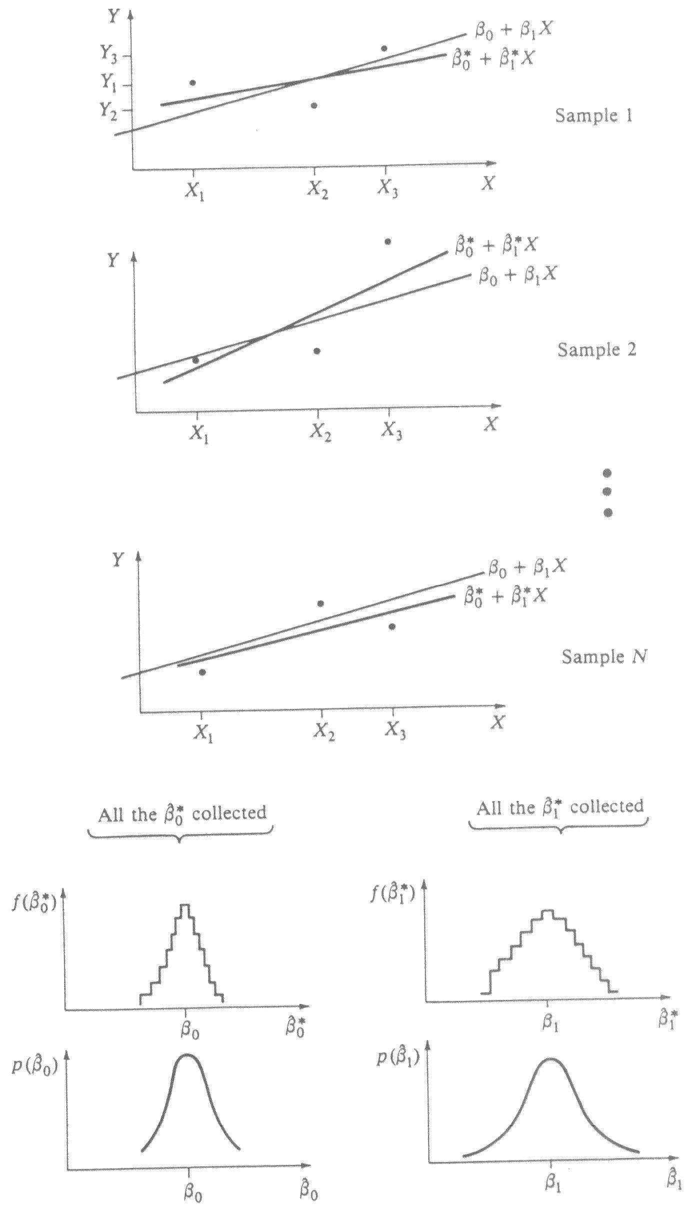


FIGURE 11.3 A thought experiment helps explain the nature of the sampling distributions of the OLS estimators of β_0 and β_1. The same three X values are used in each sample, but different sets of Y values are produced because different values for the disturbances occur in each sample. The $\hat\beta_0^*$ and $\hat\beta_1^*$ values computed in each sample are collected, and their frequency distributions are constructed. The sampling distribution of an estimator is the limiting form of the frequency distribution as N approaches infinity. The actual $\hat\beta_0^*$ and $\hat\beta_1^*$ that we calculate from a set of data are thought of as just one pair of outcomes from these sampling distributions.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

2. The sum of squares for x_t $\left(\sum_{t=1}^{T}(x_t - \bar{x})^2\right)$ appears in the denominator of each formula

- Sum of squares for x_t measures the spread, or variation, of x_t

- The larger the variation in x_t, the smaller are the variances and covariance of b_1 and b_2

  $\Rightarrow$ The more information we have about x_t, the more precisely can we estimate β_1 and β_2

3. The larger the sample size T, the smaller the variances and covariance of b_1 and b_2

- As T increases, the sum of squares for increases unambiguously

- Effect is clear for $\text{var}(b_2)$ and $\text{cov}(b_1, b_2)$

- Same impact on $\text{var}(b_1)$ because T and variation in x_t appear in denominator

  $\Rightarrow$ The more information we have about x_t, the more precisely can we estimate β_1 and β_2

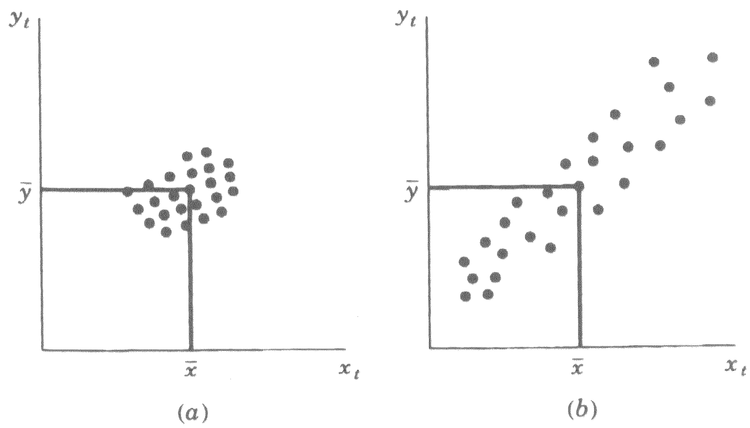
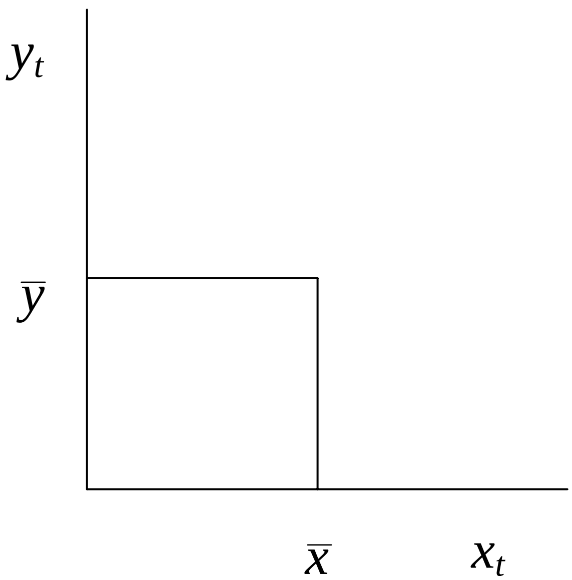
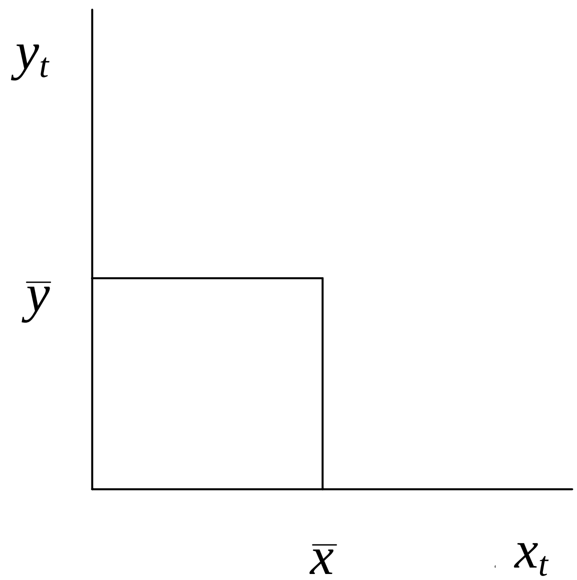


Figure 6.2 The influence of variation in the explanatory variable on precision of estimation. (*a*) Low variation, low precision. (*b*) High variation, high precision.

Griffiths, W.E., R.C. Hill and G.C. Judge. *Learning and Practicing Econometrics*. John Wiley & Sons, Inc., New York, NY, 1993.





Which data sample would you rather use to estimate the linear relationship between x and y?

4. The term $\sum_{t=1}^{T} x_t^2$ appears in the numerator of the formula for $\text{var}(b_1)$

- $\sum_{t=1}^{T} x_t^2$ measures the distance of the data on x_t from the origin

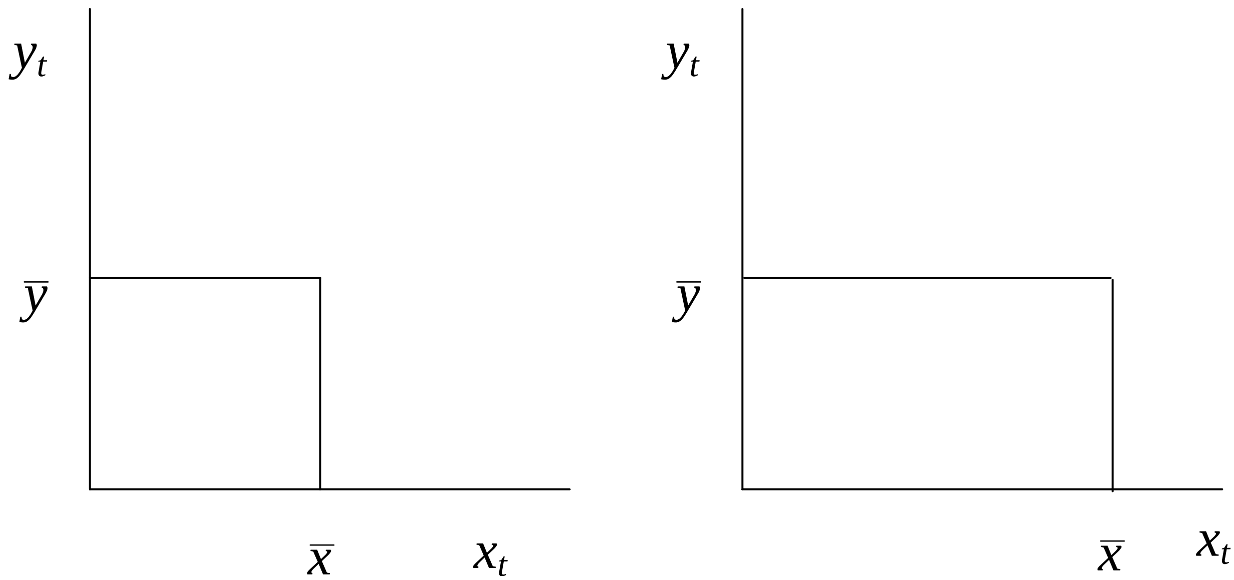
  $\Rightarrow$ The more distant from zero are the data on x_t, the independent variable, the more difficult it is to accurately estimate the intercept β_1

5. $\bar{x}$ appears in the numerator of $\text{cov}(b_1, b_2)$

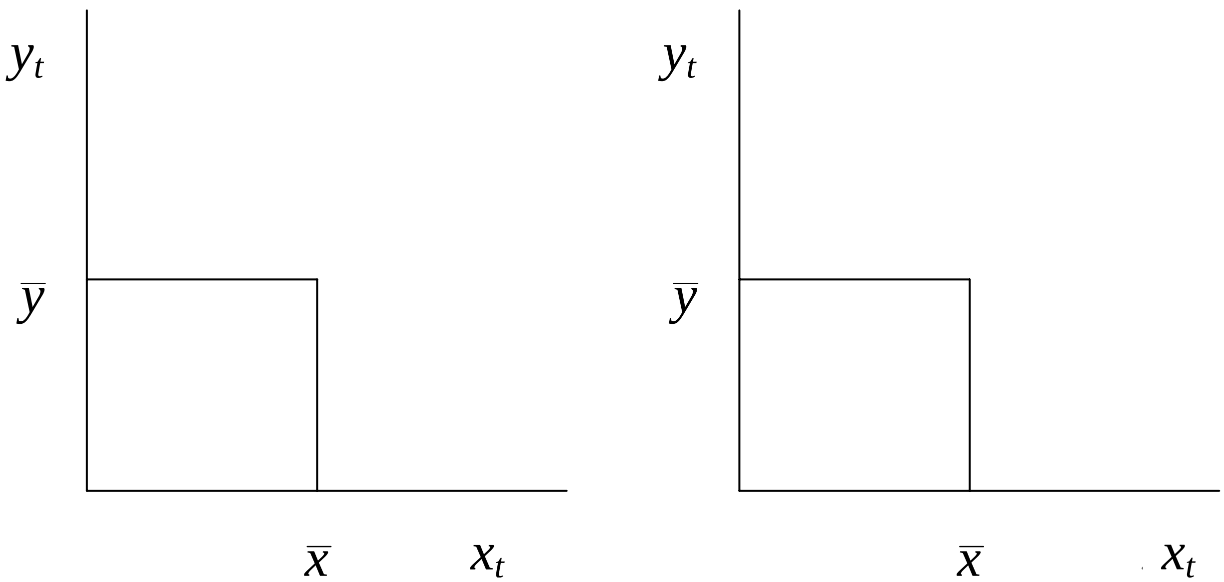
- Covariance has the opposite sign as $\bar{x}$

- Covariance increases in magnitude the larger is $\bar{x}$

An understanding of the previous five points is of great value when interpreting regression results in applied research



Which data sample would you rather use to estimate the intercept in the linear relationship between *x* and *y*?



Demonstration of negative covariance between slope and intercept estimates when mean of *x* is positive

Summary of Key Factors Affecting Precision of Least Squares Estimators

In the following table, characteristics of the sample data are categorized in terms of impact on precision

Sample Characteristic	More Precision	Less Precision
High Variance of y_t		✓
Low Variance of y_t	✓	
High Variation in x_t	✓	
Low Variation in x_t		✓
Large T	✓	
Small T		✓
Small Distance from Origin and x_t	✓	
Large Distance from Origin and x_t		✓

Sampling Properties of the Least Squares Estimators

Since there are a number of different rules for obtaining estimates of β_1 and β_2, how can we be assured that b_1 and b_2 are the "best" rules?

Previously, we developed four main criteria for "good" estimators

- Computational cost: Estimator is a linear function of sample data
- Unbiasedness: In repeated sampling, the estimator generates estimates that on average equal the population parameter
- Efficiency: Of all possible unbiased estimators, there is no other estimator that has a smaller variance
- Consistency: As the sample size increases the probability mass of the of estimator "collapses" on the population parameter

We will examine the least squares estimator b_2 to see if it meets these four criteria

Similar results hold for b_1

Computational Cost

Earlier, we noted that,

$$b_2 = \sum_{t=1}^{T} w_t y_t$$

where,

$$w_t = \frac{(x_t - \overline{x})}{\sum_{t=1}^{T} (x_t - \overline{x})^2}$$

Writing the formula in the above format shows that the least squares rule b_2 is a linear function of the y_t

Unbiasedness

We want to know whether the expected value of b_2 is in fact equal to β_2

Earlier, we showed that the mean is derived by taking the expectation of the following version of the formula for b_2,

$$E(b_2) = E\left(\beta_2 + \sum_{t=1}^{T} w_t e_t\right)$$

$$E(b_2) = E(\beta_2) + \sum_{t=1}^{T} w_t E(e_t)$$

$$E(b_2) = \beta_2$$

Shows that the least squares estimator b_2 is unbiased

Efficiency

For a given sample size T, we want to know whether the sampling variance of b_2 is smaller than any other unbiased, linear estimator

- Desire an estimator that gives us the highest probability of obtaining an estimate close to the true parameter value

In other words, is there a different estimator that produces a sampling variance smaller than the following formula,

$$\text{var}(b_2) = \sigma^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

For all unbiased linear estimators, var(b_2) is the smallest sampling variance possible

- Proof is found on pp. 78-79 of Hill et. al, *Undergraduate Econometrics* and many other texts

Consistency

We want to show that as the sample size increases, the probability mass of the of estimator "collapses" on the population parameter

This can be demonstrated informally by noting the formula for the sampling variance of b_2,

$$\text{var}(b_2) = \sigma^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

and that

$$\lim_{T \to \infty} \text{var}(b_2) = \sigma^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right] = 0$$

Summary of Sampling Properties Discussion

The least squares estimators b_1 and b_2 of the population parameters β_1 and β_2 are,

- Linear
- Unbiased
- Efficient
- Consistent

The first three properties are sufficient to prove that b_1 and b_2 are the best linear unbiased estimators (BLUE) of β_1 and β_2

- In this context "best" implies minimum variance sampling distribution
- Known as the Gauss-Markov Theorem

Key Points:

- The estimators b_1 and b_2 are “best” when compared to similar estimators, those that are linear and unbiased. The Gauss-Markov Theorem does not say that b_1 and b_2 are the best of all possible estimators.

- The estimators b_1 and b_2 are best within their class because they have the minimum variance.

- In order for the Gauss-Markov Theorem to hold, the assumptions (SR1-SR5) must be true. If any of the assumptions 1-5 are not true, then b_1 and b_2 are not the best linear unbiased estimators of β_1 and β_2.

- The Gauss-Markov Theorem does not depend on the assumption of normality.

- In the simple linear regression model, if we want to use a linear and unbiased estimator, then we have to do no more searching.

- The Gauss-Markov theorem applies to the least squares estimators. It does not apply to the least squares estimates from a single sample.

Estimating the Variance of the Error Term

Recall that e_t and y_t were assumed to be *iid* with the following distributions,

$$e_t \sim N(0,\sigma^2) \qquad \text{and} \qquad y_t \sim N(\beta_1 + \beta_2 x_t, \sigma^2)$$

Unless σ^2 is known, which is highly unlikely, it will have to be estimated as well

Again, we cannot use the least squares principle, as σ^2 does not appear in the sum of squares function,

$$S(\beta_1,\beta_2) = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (y_t - \beta_1 - \beta_2 x_t)^2$$

Instead, we apply a "heuristic" procedure based on the definition of σ^2

The original definition of σ^2 in the statistical model is,

$$\text{var}(y_t) = \text{var}(e_t) = \sigma^2 = E[e_t^2]$$

In other words, the variance is the expected value of the squared errors

Given this definition, it would be natural to estimate σ^2 as the average of the squared errors

In order to do this, we must first obtain estimates of the population errors using our sample data as

$$\hat{e}_t = y_t - b_1 - b_2 x_t$$

We can then develop our sample estimator of σ^2 as,

$$\hat{\sigma}^2 = \frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots + \hat{e}_T^2}{T-2} = \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-2}$$

Notice the squared sample errors are averaged by dividing by *T-2* **not** *T* (or *T-1*)

- Accounts for the fact that 2 regression parameters (β_1, β_2) have to be estimated

Many regression packages report something called the "standard error of the regression"

This is simply the square root of the estimated variance of the error term,

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots + \hat{e}_T^2}{T-2}} = \sqrt{\frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-2}}$$

Warning: Do not confuse the standard error of the regression with the standard error of the sampling distribution of the least squares estimators b_1 and b_2

Sampling Properties of Variance Estimator

The rule derived for estimating the population variance of the error term (σ^2) is,

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-2}$$

Just as was the case with b_1 and b_2, we are interested in the sampling properties of $\hat{\sigma}^2$

The same four criteria are applied when asking whether $\hat{\sigma}^2$ is a "good" estimation rule

- Computational cost, unbiasedness, efficiency, consistency

It is obvious that $\hat{\sigma}^2$ is not a linear estimator

It can be shown that $\hat{\sigma}^2$ is unbiased, efficient, and consistent

- Best unbiased estimator (BUE)

- Proof can be found in advanced econometrics books

The next issue is the form of the sampling distribution of the variance estimator $\hat{\sigma}^2$

To derive the sampling distribution, first note that the population random errors are distributed as,

$$e_t \sim N(0,\sigma^2) \qquad t = 1,\ldots,T$$

Consequently,

$$\frac{e_t}{\sigma} \sim N(0,1) \qquad t = 1,\ldots,T$$

and,

$$\left(\frac{e_t}{\sigma}\right)^2 \sim \chi_1 \qquad t = 1,\ldots,T$$

If we sum over the T transformed random errors,

$$\sum_{i=1}^{T}\left(\frac{e_t}{\sigma}\right)^2 \sim \chi_T$$

Based on the previous result, we can generate the sampling distribution of $\hat{\sigma}^2$,

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{T-2}\chi_{T-2}$$

Note that the sampling distribution of $\hat{\sigma}^2$ is proportional to a chi-square with *T-2* degrees of freedom

Estimators of the Variances and Covariance of b_1 and b_2

Recall that the variances and covariance of the sampling distributions of b_1 and b_2 were functions of the unknown parameter σ^2

We can generate estimators of the variances and covariances of b_1 and b_2 by simply replacing σ^2 with $\hat{\sigma}^2$ in the earlier formulas,

$$\hat{\text{var}}(b_1) = \hat{\sigma}^2 \left[\frac{\sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

$$\hat{\text{var}}(b_2) = \hat{\sigma}^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

$$\hat{\text{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\bar{x}}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]$$

Likewise, the estimators for the standard errors of b_1 and b_2 are,

$$\hat{se}(b_1) = \sqrt{\hat{var}(b_1)} = \sqrt{\hat{\sigma}^2 \left[\frac{\sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2} \right]}$$

$$\hat{se}(b_2) = \sqrt{\hat{var}(b_2)} = \sqrt{\hat{\sigma}^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \bar{x})^2} \right]}$$

This material is tricky:

- We developed estimators of the variances, covariance and standard errors of the least squares estimators of b_1 and b_2 !!

- While we will not consider the extra complexity, the estimators of the variances, covariance and standard errors are themselves random variables whose values vary in repeated sampling

Sample Estimates of the Variances and Covariance of b_1 and b_2

We can now use our sample data on food expenditure and income to generate estimates of the variances and covariance of b_1 and b_2

First, estimate the variance of the error term,

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} \hat{e}_t^2}{T-2} = \frac{1780.4}{38} = 46.853$$

and the standard error of the regression,

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{46.853} = 6.845$$

Next, estimate the variances, standard errors, and covariance of b_1 and b_2,

$$\hat{\text{var}}(b_1) = \hat{\sigma}^2 \left[\frac{\sum_{t=1}^{T} x_t^2}{T\sum_{t=1}^{T}(x_t - \bar{x})^2} \right] = 46.853(0.3429) = 16.0669$$

$$\hat{se}(b_1) = \sqrt{\hat{\text{var}}(b_1)} = \sqrt{16.0669} = 4.0083$$

$$\hat{\text{var}}(b_2) = \hat{\sigma}^2 \left[\frac{1}{\sum_{t=1}^{T} (x_t - \overline{x})^2} \right] = 46.853(0.0000653) = 0.0031$$

$$\hat{se}(b_2) = \sqrt{\hat{\text{var}}(b_2)} = \sqrt{0.0031} = 0.0557$$

$$\hat{\text{cov}}(b_1, b_2) = \hat{\sigma}^2 \left[\frac{-\overline{x}}{\sum_{t=1}^{T} (x_t - \overline{x})^2} \right] = 46.853(-0.00455) = -0.2134$$

Sample Regression Output from Excel

SUMMARY OUTPUT

Regression Statistics	
Multiple R	**0.563096017**
R Square	**0.317077125**
Adjusted R Square	**0.29910547**
Standard Error	**6.844922384**
Observations	**40**

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	**1**	**826.6352172**	**826.6352**	**17.64318**	**0.000155136**
Residual	**38**	**1780.412573**	**46.85296**		
Total	**39**	**2607.04779**			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	**7.383217543**	**4.008356335**	**1.841956**	**0.073296**	**-0.731275911**	**15.497711**
X Variable 1	**0.23225333**	**0.055293429**	**4.200378**	**0.000155**	**0.120317631**	**0.34418903**

Summary of Estimates for Food Expenditure Data

$$b_1 = 7.3832 \quad b_2 = 0.2323 \quad \hat{\sigma}^2 = 46.853$$

$$\hat{\text{var}}(b_1) = 16.0669 \qquad \hat{\text{var}}(b_2) = 0.0031$$

$$\hat{\text{cov}}(b_1, b_2) = -0.2134$$

Based on this information and the assumption that the statistical model is correctly specified, we can estimate the distributions of e_t and y_t as,

$$e_t \sim N(0, 46.853) \qquad y_t \sim N(7.382 + 0.2323x_t, 46.853)$$

We also can estimate the sampling distributions of b_1 and b_2 as,

$$b_1 \sim N(7.382, 16.0669) \qquad b_2 \sim N(0.2323, 0.0031)$$

Interpretation Guidelines

Regression standard error

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{46.853} = 6.845$$

We say, "The typical error of the regression model, without regard to sign, is estimated to be $6.845/week."

Standard error of parameter estimates

$$\hat{se}(b_1) = \sqrt{\hat{\text{var}}(b_1)} = \sqrt{16.0669} = 4.0083$$

We say, "The typical error in estimating β_1, without regard to sign, is estimated to be 4.0083."

$$\hat{se}(b_2) = \sqrt{\hat{\text{var}}(b_2)} = \sqrt{0.0031} = 0.0557$$

We say, "The typical error in estimating β_2, without regard to sign, is estimated to be 0.0557."

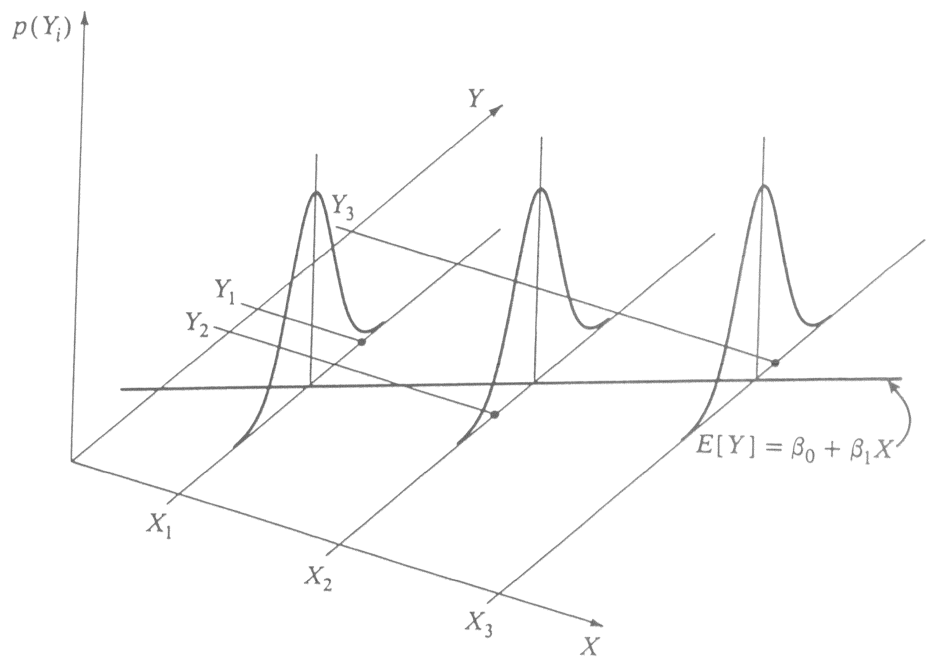


FIGURE 11.2 In the normal regression model, the value of Y for each observation is considered to be an outcome of a separate normal probability distribution. Each distribution has a mean of $\beta_0 + \beta_1 X_i$ and a standard deviation of σ_u as in Figure 11.1. In this three-dimensional diagram, probability density values are drawn vertically. The set of values labeled here as Y_1, Y_2, and Y_3 are thought of as just one possible set of values for Y corresponding to the three given values of X.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

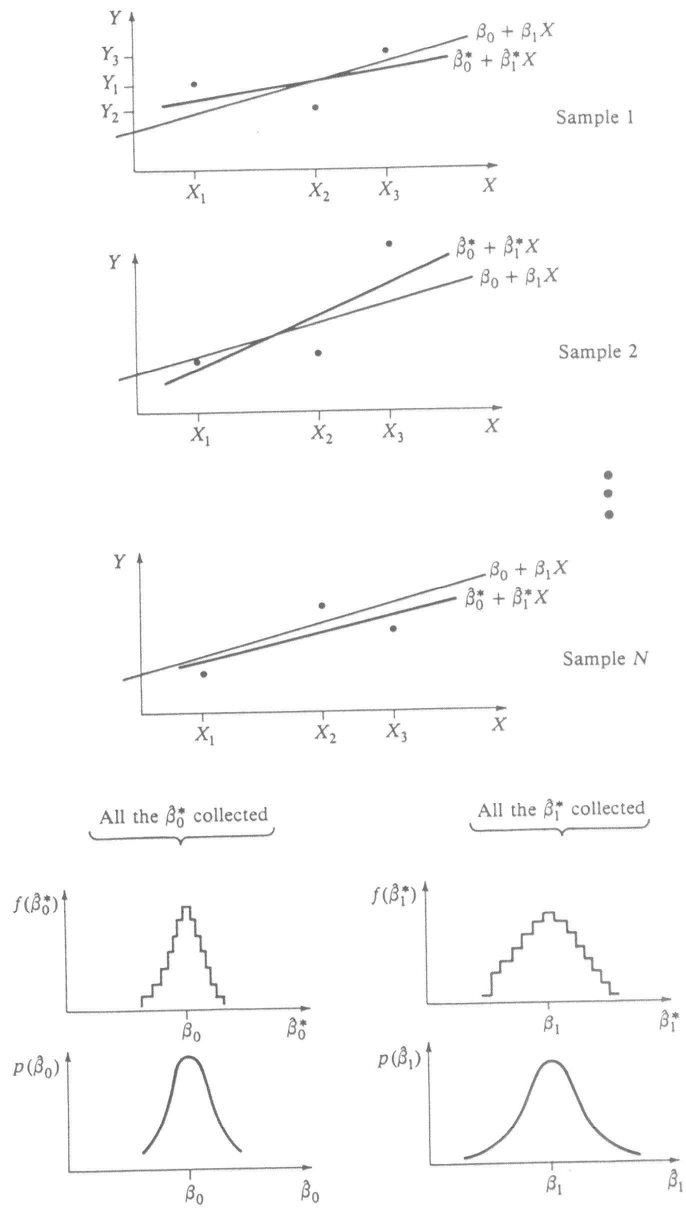


FIGURE 11.3 A thought experiment helps explain the nature of the sampling distributions of the OLS estimators of β_0 and β_1. The same three X values are used in each sample, but different sets of Y values are produced because different values for the disturbances occur in each sample. The $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ values computed in each sample are collected, and their frequency distributions are constructed. The sampling distribution of an estimator is the limiting form of the frequency distribution as N approaches infinity. The actual $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ that we calculate from a set of data are thought of as just one pair of outcomes from these sampling distributions.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ. 1995.

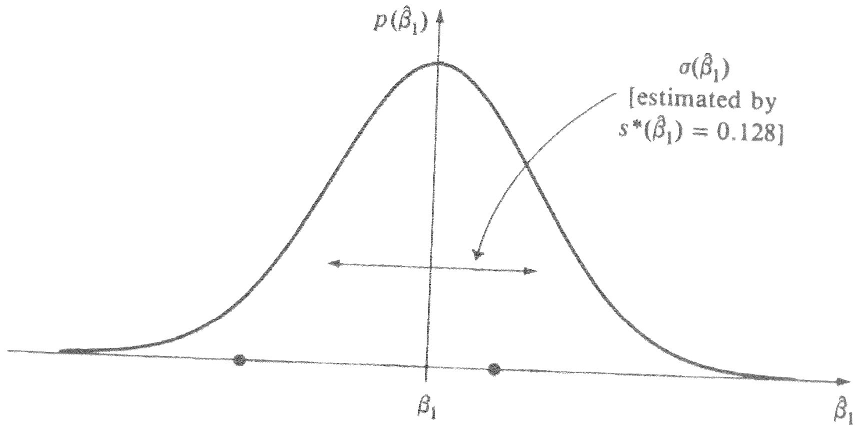


FIGURE 11.4 The sampling distribution of $\hat{\beta}_1$ in the earnings function is normal, with a mean equal to β_1 (unknown) and a standard deviation (i.e., standard error) estimated to be 0.128. The estimate of β_1 is $\hat{\beta}_1^* = 0.797$; this might correspond to either of the unlabeled points on the horizontal axis.

Mirer, Thad W. Economic Statistics and Econometrics, Third Edition. Prentice Hall, Englewood Cliffs, NJ 1995.