

## 4. Statistical Inference for Regression

### 1. Goals:

- ▶ To introduce the standard statistical models and assumptions for simple and multiple linear regression.
- ▶ To describe properties of the least-squares coefficients as estimators of the parameters of the regression model.
- ▶ To introduce flexible and general procedures for statistical inference based on least-squares estimators.
- ▶ To explore further the interpretation of regression equations.

## 2. Simple Regression

### 2.1 The Simple-Regression Model

Standard statistical inference in simple regression is based upon a statistical 'model':

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- ▶ The coefficients  $\alpha$  and  $\beta$  are the population regression parameters to be estimated.
- ▶ The *error*  $\varepsilon_i$  represents the aggregated, omitted causes of  $Y$ :
  - Other explanatory variables that could have been included.
  - Measurement error in  $Y$ .
  - Whatever component of  $Y$  is inherently random.

- ▶ The key assumptions of the simple-regression model concern the behavior of the errors — or, equivalently, of the distribution of  $Y$  conditional on  $X$ :
  - **Linearity.**  $E(\varepsilon_i) \equiv E(\varepsilon|x_i) = 0$ . Equivalently, the average value of  $Y$  is a linear function of  $X$ :

$$\begin{aligned} \mu_i &\equiv E(Y_i) \equiv E(Y|x_i) = E(\alpha + \beta x_i + \varepsilon_i) \\ &= \alpha + \beta x_i + E(\varepsilon_i) \\ &= \alpha + \beta x_i \end{aligned}$$

- **Constant Variance.**  $V(\varepsilon|x_i) = \sigma_\varepsilon^2$ . Equivalently, the variance of  $Y$  around the regression line is constant:

$$V(Y|x_i) = E[(Y_i - \mu_i)^2] = E[(Y_i - \alpha - \beta x_i)^2] = E(\varepsilon_i^2) = \sigma_\varepsilon^2$$

- **Normality.**  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ . Equivalently, the conditional distribution of  $Y$  given  $x$  is normal:  $Y_i \sim N(\alpha + \beta x_i, \sigma_\varepsilon^2)$ . The assumptions of linearity, constant variance, and normality are illustrated in Figure 1.

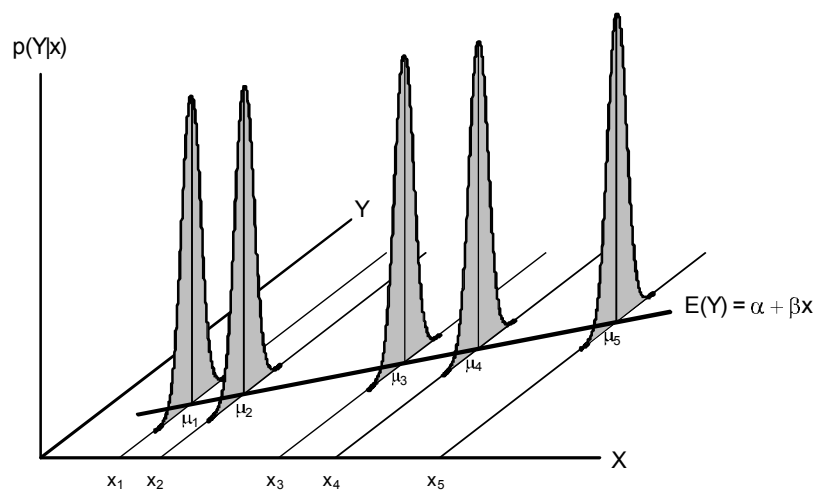


Figure 1. The assumptions of linearity, normality, and constant variance in the simple-regression model.

- **Independence.** The observations are sampled independently: Any pair of errors  $\varepsilon_i$  and  $\varepsilon_j$  (or, equivalently, of conditional response-variable values,  $Y_i$  and  $Y_j$ ) are independent for  $i \neq j$ . The assumption of independence needs to be justified by the procedures of data collection.
- **Fixed X or X independent of the error.** Depending upon the design of a study, the values of the explanatory variable may be fixed in advance of data collection or they may be sampled along with the response variable.
  - Fixed X corresponds almost exclusively to experimental research.
  - When, as is more common, X is sampled along with Y, we assume that the explanatory variable and the error are independent in the population from which the sample is drawn: That is, the error has the same distribution  $[N(0, \sigma_\varepsilon^2)]$  for every value of X in the population.

## 2.2 Properties of the Least-Squares Estimator

Under the strong assumptions of the simple-regression model, the sample least-squares coefficients  $A$  and  $B$  have several desirable properties as estimators of the population regression coefficients  $\alpha$  and  $\beta$ :

- The least-squares intercept and slope are *linear estimators*, in the sense that they are linear functions of the observations  $Y_i$ . For example,

$$B = \sum_{i=1}^n m_i Y_i$$

where

$$m_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

- This result is not important in itself, but it makes the distributions of the least-squares coefficients simple.

- Under the assumption of linearity,  $A$  and  $B$  are unbiased estimators of  $\alpha$  and  $\beta$ :

$$E(A) = \alpha$$

$$E(B) = \beta$$

- Under the assumptions of linearity, constant variance, and independence,  $A$  and  $B$  have simple sampling variances:

$$V(A) = \frac{\sigma_\varepsilon^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

$$V(B) = \frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2}$$

- It is instructive to rewrite the formula for  $V(B)$ :

$$V(B) = \frac{\sigma_\varepsilon^2}{(n-1)S_X^2}$$

- ▶ **The Gauss-Markov theorem:** Of all linear unbiased estimators, the least-squares estimators are most efficient.
  - Under normality, the least-squares estimators are most efficient among *all* unbiased estimators, not just among linear estimators. This is a much more compelling result.
- ▶ Under the full suite of assumptions, the least-squares coefficients  $A$  and  $B$  are the maximum-likelihood estimators of  $\alpha$  and  $\beta$ .

- ▶ Under the assumption of normality, the least-squares coefficients are themselves normally distributed:

$$A \sim N \left[ \alpha, \frac{\sigma_\varepsilon^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} \right]$$

$$B \sim N \left[ \beta, \frac{\sigma_\varepsilon^2}{\sum (x_i - \bar{x})^2} \right]$$

- Even if the errors are not normally distributed, the distributions of  $A$  and  $B$  are approximately normal, with the approximation improving as the sample size grows.

## 2.3 Confidence Intervals and Hypothesis Tests

- ▶ The distributions of  $A$  and  $B$  cannot be directly employed for statistical inference since  $\sigma_\varepsilon^2$  is never known in practice.
- ▶ The variance of the residuals provides an unbiased estimator of  $\sigma_\varepsilon^2$ ,

$$S_E^2 = \frac{\sum E_i^2}{n - 2}$$

and a basis for estimating the variances of  $A$  and  $B$ :

$$\widehat{V}(A) = \frac{S_E^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

$$\widehat{V}(B) = \frac{S_E^2}{\sum (x_i - \bar{x})^2}$$

- ▶ The added uncertainty induced by estimating the error variance is reflected in the use of the  $t$ -distribution, in place of the normal distribution, for confidence intervals and hypothesis tests.

- To construct a  $100(1 - \alpha)\%$  confidence interval for the slope, we take

$$\beta = B \pm t_{\alpha/2} \text{SE}(B)$$

where  $t_{\alpha/2}$  is the critical value of  $t$  with  $n - 2$  degrees of freedom and a probability of  $\alpha/2$  to the right, and  $\text{SE}(B)$  is the square root of  $\widehat{V}(B)$ . (This is just like a confidence interval for a population mean.)

- Similarly, to test the hypothesis  $H_0: \beta = \beta_0$  (most commonly,  $H_0: \beta = 0$ ), calculate the test statistic

$$t_0 = \frac{B - \beta_0}{\text{SE}(B)}$$

which is distributed as  $t$  with  $n - 2$  degrees of freedom under  $H_0$ .

- Confidence intervals and hypothesis tests for  $\alpha$  follow the same pattern.

- For Davis's regression of measured on reported weight, for example:

$$S_E = \sqrt{\frac{418.87}{101 - 2}} = 2.0569$$

$$SE(A) = \frac{2.0569 \times \sqrt{329,731}}{\sqrt{101 \times 4539.3}} = 1.7444$$

$$SE(B) = \frac{2.0569}{\sqrt{4539.3}} = 0.030529$$

- Since  $t_{.025}$  for  $101 - 2 = 99$  degrees of freedom is 1.984, 95-percent confidence intervals for  $\alpha$  and  $\beta$  are

$$\alpha = 1.778 \pm 1.984 \times 1.744 = 1.778 \pm 3.460$$

$$\beta = 0.9772 \pm 1.984 \times 0.03053 = 0.9772 \pm 0.06057$$

©

### 3. Multiple Regression

Most of the results for multiple-regression analysis parallel those for simple regression.

#### 3.1 The Multiple-Regression Model

- The statistical model for multiple regression is

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \varepsilon_i$$

- The assumptions underlying the model concern the errors,  $\varepsilon_i$ , and are identical to the assumptions in simple regression:

- **Linearity.**  $E(\varepsilon_i) = 0$ .
- **Constant Variance.**  $V(\varepsilon_i) = \sigma_\varepsilon^2$ .
- **Normality.**  $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ .
- **Independence.**  $\varepsilon_i, \varepsilon_j$  independent for  $i \neq j$ .
- **Fixed X's or X's independent of  $\varepsilon$ .**

©

- Under these assumptions (or particular subsets of them), the least-squares estimators  $A, B_1, \dots, B_k$  of  $\alpha, \beta_1, \dots, \beta_k$  are

- linear functions of the data, and hence relatively simple;
- unbiased;
- maximally efficient among unbiased estimators;
- maximum-likelihood estimators;
- normally distributed.

- The slope coefficient  $B_j$  in multiple regression has sampling variance

$$V(B_j) = \frac{1}{1 - R_j^2} \times \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}$$

where  $R_j^2$  is the multiple correlation from the regression of  $X_j$  on all of the other  $X$ 's.

- The second factor is essentially the sampling variance of the slope in simple regression, although the error variance  $\sigma_\varepsilon^2$  is smaller than before.

©

- The first factor — called the *variance-inflation factor* — is large when the explanatory variable  $X_j$  is strongly correlated with other explanatory variables (the problem of collinearity).

©

## 3.2 Confidence Intervals and Hypothesis Tests

### 3.2.1 Individual Slope Coefficients

- Confidence intervals and hypothesis tests for individual coefficients closely follow the pattern of simple-regression analysis:

- The variance of the residuals provides an unbiased estimator of  $\sigma_\varepsilon^2$ :

$$S_E^2 = \frac{\sum E_i^2}{n - k - 1}$$

- Using  $S_E^2$ , we can calculate the standard error of  $B_j$ :

$$SE(B_j) = \frac{1}{\sqrt{1 - R_j^2}} \times \frac{S_E}{\sqrt{\sum (X_{ij} - \bar{X}_j)^2}}$$

- Confidence intervals and tests, based on the  $t$ -distribution with  $n - k - 1$  degrees of freedom, follow straightforwardly.

©

- For example, for Duncan's regression of occupational prestige on education and income:

$$S_E^2 = \frac{7506.7}{45 - 2 - 1} = 178.73$$

$$r_{12} = .72451$$

$$SE(B_1) = \frac{1}{\sqrt{1 - .72451^2}} \times \frac{\sqrt{178.73}}{\sqrt{38,971}} = 0.098252$$

$$SE(B_2) = \frac{1}{\sqrt{1 - .72451^2}} \times \frac{\sqrt{178.73}}{\sqrt{26,271}} = 0.11967$$

- With only two explanatory variables,  $R_1^2 = R_2^2 = r_{12}^2$ .

©

- To construct 95-percent confidence intervals for the slope coefficients, we use  $t_{.025} = 2.018$  from the  $t$ -distribution with  $45 - 2 - 1 = 42$  degrees of freedom:

$$\text{Education: } \beta_1 = 0.5459 \pm 2.018 \times 0.09825 = 0.5459 \pm 0.1983$$

$$\text{Income: } \beta_2 = 0.5987 \pm 2.018 \times 0.1197 = 0.5987 \pm 0.2415$$

©

### 3.2.2 All Slopes

- We can also test the global or 'omnibus' null hypothesis that all of the regression slopes are zero:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

which is not quite the same as testing the separate hypotheses

$$H_0^{(1)}: \beta_1 = 0; H_0^{(2)}: \beta_2 = 0; \dots; H_0^{(k)}: \beta_k = 0$$

- An  $F$ -test for the omnibus null hypothesis is given by

$$F_0 = \frac{\frac{\text{RegSS}}{k}}{\frac{\text{RSS}}{n - k - 1}} = \frac{n - k - 1}{k} \times \frac{R^2}{1 - R^2}$$

- Under the null hypothesis, this test statistic follows an  $F$ -distribution with  $k$  and  $n - k - 1$  degrees of freedom.

©

- The calculation of the test statistic can be organized in an *analysis-of-variance table*:

Source	Sum of Squares	df	Mean Square	F
Regression	RegSS	$k$	$\frac{\text{RegSS}}{k}$	$\frac{\text{RegMS}}{\text{RMS}}$
Residuals	RSS	$n - k - 1$	$\frac{\text{RSS}}{n - k - 1}$	
Total	TSS	$n - 1$		

- When the null hypothesis is true, RegMS and RMS provide independent estimates of the error variance, so the ratio of the two mean squares should be close to one.

- When the null hypothesis is false, RegMS estimates the error variance plus a positive quantity that depends upon the  $\beta$ 's:

$$E(F_0) \approx \frac{E(\text{RegMS})}{E(\text{RMS})} = \frac{\sigma_\varepsilon^2 + \text{positive quantity}}{\sigma_\varepsilon^2}$$

- We consequently reject the omnibus null hypothesis for values of  $F_0$  that are sufficiently larger than 1.

► For Duncan's regression:

Source	SS	df	MS	F	p
Regression	36,181.	2	18,090.	101.2	$\ll .0001$
Residuals	7506.7	42	178.73		
Total	43,688.	44			

### 3.2.3 A Subset of Slopes

► Consider the hypothesis

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_q = 0$$

where  $1 \leq q \leq k$ .

- The 'full' regression model, including all of the explanatory variables, may be written:

$$Y_i = \alpha + \beta_1 X_{i1} + \dots + \beta_q X_{iq} + \beta_{q+1} X_{i,q+1} + \dots + \beta_k X_{ik} + \varepsilon_i$$

- If the null hypothesis is correct, then the first  $q$  of the  $\beta$ 's are zero, yielding the 'null' model

$$Y_i = \alpha + \beta_{q+1} X_{i,q+1} + \dots + \beta_k X_{ik} + \varepsilon_i$$

- The null model omits the first  $q$  explanatory variables, regressing  $Y$  on the remaining  $k - q$  explanatory variables.

- An  $F$ -test of the null hypothesis is based upon a comparison of these two models:

- $\text{RSS}_1$  and  $\text{RegSS}_1$  are the residual and regression sums of squares for the full model.
- $\text{RSS}_0$  and  $\text{RegSS}_0$  are the residual and regression sums of squares for the null model.
- Because the null model is a special case of the full model,  $\text{RSS}_0 \geq \text{RSS}_1$ . Equivalently,  $\text{RegSS}_0 \leq \text{RegSS}_1$ .
- If the null hypothesis is wrong and (some of)  $\beta_1, \dots, \beta_q$  are nonzero, then the *incremental* (or 'extra') *sum of squares* due to fitting the additional explanatory variables

$$\text{RSS}_0 - \text{RSS}_1 = \text{RegSS}_1 - \text{RegSS}_0$$

should be large.

– The  $F$ -statistic for testing the null hypothesis is

$$F_0 = \frac{\frac{\text{RegSS}_1 - \text{RegSS}_0}{q}}{\frac{\text{RSS}_1}{n - k - 1}}$$

$$= \frac{n - k - 1}{q} \times \frac{R_1^2 - R_0^2}{1 - R_1^2}$$

– Under the null hypothesis, this test statistic has an  $F$ -distribution with  $q$  and  $n - k - 1$  degrees of freedom.

► I will, for the present, illustrate the incremental  $F$ -test by applying it to the trivial case in which  $q = 1$ :

- In Duncan's dataset, the regression of prestige on income alone produces  $\text{RegSS}_0 = 30,655$
- The regression of prestige on both income and education produces  $\text{RegSS}_1 = 36,181$  and  $\text{RSS}_1 = 7506.7$ .

©

• Consequently, the incremental sum of squares due to education is

$$\text{RegSS}_1 - \text{RegSS}_0 = 36,181 - 30,655 = 5516$$

• The  $F$ -statistic for testing  $H_0: \beta_{\text{Education}} = 0$  is

$$F_0 = \frac{\frac{5516}{1}}{\frac{7506.7}{45 - 2 - 1}} = 30.86$$

with 1 and 42 degrees of freedom, for which  $p < .0001$ .

• When  $q = 1$ , the incremental  $F$ -test is equivalent to the  $t$ -test,  $F_0 = t_0^2$ :

$$t_0 = \frac{0.5459}{0.09825} = 5.556$$

$$t_0^2 = 5.556^2 = 30.87$$

©

## 4. Empirical vs. Structural Relations

There are two fundamentally different interpretations of regression coefficients.

- Borrowing Goldberger's (1973) terminology, we may interpret a regression descriptively, as an *empirical association* among variables, or causally, as a *structural relation* among variables.
- I will deal first with empirical associations.
  - Suppose that in a population, the relationship between  $Y$  and  $X_1$  is well described by a straight line:

$$Y = \alpha' + \beta_1' X_1 + \varepsilon'$$

- We do not assume that  $X_1$  necessarily causes  $Y$  or, if it does, that the omitted causes of  $Y$ , incorporated in  $\varepsilon'$ , are independent of  $X_1$ .
- If we draw a random sample from this population, then the least-squares sample slope  $B_1'$  is an unbiased estimator of  $\beta_1'$ .

©

• Suppose, now, that we introduce a second explanatory variable,  $X_2$ , and that, in the same sense as before, the population relationship between  $Y$  and the two  $X$ 's is linear:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

- The slope  $\beta_1$  generally will differ from  $\beta_1'$ .
- The sample least-squares coefficients for the multiple regression,  $B_1$  and  $B_2$ , are unbiased estimators of the corresponding population coefficients,  $\beta_1$  and  $\beta_2$ .

©

- That the simple-regression slope  $\beta'_1$  differs from the multiple-regression slope  $\beta_1$ , and that therefore the sample *simple*-regression coefficient  $B'_1$  is a *biased* estimator of the population *multiple*-regression slope  $\beta_1$ , is not problematic, for we do not in this context interpret a regression coefficient as the *effect* of an explanatory variable on the response variable.
  - The issue of *specification error* does not arise, as long as the linear-regression model adequately describes the empirical relationship between the variables in the population.

► Imagine now that response-variable scores are *constructed* according to the multiple-regression model

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

where  $E(\varepsilon) = 0$  and  $\varepsilon$  is independent of  $X_1$  and  $X_2$ .

- If we use least-squares to fit this model to sample data, then we will obtain unbiased estimators of  $\beta_1$  and  $\beta_2$ .

- Instead we mistakenly fit the simple-regression model

$$Y = \alpha + \beta_1 X_1 + \varepsilon'$$

where, implicitly, the effect of  $X_2$  on  $Y$  is absorbed by the error  $\varepsilon' = \varepsilon + \beta_2 X_2$ .

- If we assume wrongly that  $X_1$  and  $\varepsilon'$  are *uncorrelated* then we make an error of specification.
- The consequence is that our least-squares simple-regression estimator of  $\beta_1$  is biased: Because  $X_1$  and  $X_2$  are correlated, and because  $X_2$  is omitted from the model, part of the effect of  $X_2$  is mistakenly attributed to  $X_1$ .
- To make the nature of this specification error more precise, let us take the expectation of both sides of the true (multiple-regression) model:

$$\mu_Y = \alpha + \beta_1 \mu_1 + \beta_2 \mu_2 + 0$$

- Subtracting this equation from the model produces

$$Y - \mu_Y = \beta_1 (X_1 - \mu_1) + \beta_2 (X_2 - \mu_2) + \varepsilon$$

- Multiply this equation through by  $X_1 - \mu_1$

$$\begin{aligned} (X_1 - \mu_1)(Y - \mu_Y) &= \beta_1 (X_1 - \mu_1)^2 \\ &+ \beta_2 (X_1 - \mu_1)(X_2 - \mu_2) \\ &+ (X_1 - \mu_1)\varepsilon \end{aligned}$$

and take the expectations of both sides:

$$\sigma_{1Y} = \beta_1 \sigma_1^2 + \beta_2 \sigma_{12}$$

- Solving for  $\beta_1$ :

$$\beta_1 = \frac{\sigma_{1Y}}{\sigma_1^2} - \beta_2 \frac{\sigma_{12}}{\sigma_1^2}$$

- The least-squares coefficient for the simple regression of  $Y$  on  $X_1$  is  $B = S_{1Y}/S_1^2$ . The simple regression therefore estimates not  $\beta_1$  but rather  $\sigma_{1Y}/\sigma_1^2 \equiv \beta'_1$ .
- Put another way,  $\beta'_1 = \beta_1 + \text{bias}$ , where the *bias* =  $\beta_2 \sigma_{12}/\sigma_1^2$ .
- For the bias to be nonzero, two conditions must be met:
  - (A)  $X_2$  must be a *relevant* explanatory variable — that is,  $\beta_2 \neq 0$ .
  - (B)  $X_1$  and  $X_2$  must be *correlated* — that is,  $\sigma_{12} \neq 0$ .

- Depending upon the signs of  $\beta_2$  and  $\sigma_{12}$ , the bias in the simple-regression estimator may be either positive or negative.
- Final subtlety: The proper interpretation of the ‘bias’ in the simple-regression estimator depends upon the nature of the causal relationship between  $X_1$  and  $X_2$  (see Figure 2) :
  - In part (a) of the following figure,  $X_2$  *intervenes* causally between  $X_1$  and  $Y$ .
    - Here, the ‘bias’ term  $\beta_2 \sigma_{12}/\sigma_1^2$  is simply the *indirect effect* of  $X_1$  on  $Y$  transmitted through  $X_2$ , since  $\sigma_{12}/\sigma_1^2$  is the population slope for the regression of  $X_2$  on  $X_1$ .
  - In part (b),  $X_2$  is a *common prior cause* of both  $X_1$  and  $Y$ , and the bias term represents a *spurious* — that is, noncausal — component of the empirical association between  $X_1$  and  $Y$ .
  - In (b), but not in (a), it is critical to control for  $X_2$  in examining the relationship between  $Y$  and  $X_1$ .



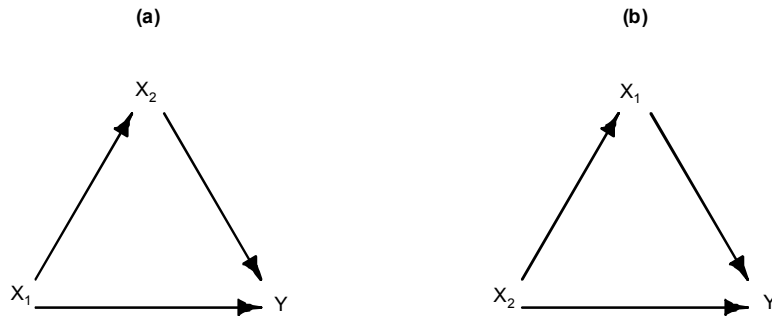


Figure 2. Two ‘causal models’ for an omitted explanatory variable  $X_2$ : (a)  $X_2$  intervenes between  $X_1$  and  $Y$ ; (b)  $X_2$  is a common prior cause of both  $X_1$  and  $Y$ .

## 5. Summary

- Standard statistical inference for least-squares regression analysis is based upon the statistical model

$$Y_i = \alpha + \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i$$

- The key assumptions of the model include linearity, constant variance, normality, and independence.
- The  $X$ -values are either fixed or, if random, are assumed to be independent of the errors.
- Under these assumptions, or particular subsets of them, the least-squares coefficients have certain desirable properties as estimators of the population regression coefficients.

- The estimated error of the slope coefficient  $B$  in simple regression is

$$SE(B) = \frac{S_E}{\sqrt{\sum(X_i - \bar{X})^2}}$$

- The standard error of the slope coefficient  $B_j$  in multiple regression is

$$SE(B_j) = \frac{1}{\sqrt{1 - R_j^2}} \times \frac{S_E}{\sqrt{\sum(X_{ij} - \bar{X}_j)^2}}$$

- In both cases, these standard errors can be used in  $t$ -intervals and hypothesis tests for the corresponding population slope coefficients.

- An  $F$ -test for the omnibus null hypothesis that all of the slopes are zero can be calculated from the analysis of variance for the regression:

$$F_0 = \frac{\frac{\text{RegSS}}{k}}{\frac{\text{RSS}}{n - k - 1}}$$

- The omnibus  $F$ -statistic has  $k$  and  $n - k - 1$  degrees of freedom.
- There is also an  $F$ -test for the hypothesis that a subset of  $q$  slope coefficients is zero, based upon a comparison of the regression sums of squares for the full regression model (model 1) and for a null model (model 0) that deletes the explanatory variables in the null hypothesis:

$$F_0 = \frac{\frac{\text{RegSS}_1 - \text{RegSS}_0}{q}}{\frac{\text{RSS}_1}{n - k - 1}}$$

- This incremental  $F$ -statistic has  $q$  and  $n - k - 1$  degrees of freedom.
- ▶ It is important to distinguish between interpreting a regression descriptively as an empirical association among variables and structurally as specifying causal relations among variables.
  - In the latter event, but not in the former, it is sensible to speak of bias produced by omitting an explanatory variable that (1) is a cause of  $Y$ , and (2) is correlated with an explanatory variable in the regression equation.
  - Bias in least-squares estimation results from the correlation that is induced between the included explanatory variable and the error.