6 Statistical Inference for Regression

The previous chapter developed linear least-squares regression as a descriptive technique for fitting a linear surface to data. The subject of the present chapter, in contrast, is statistical inference. I will discuss point estimation of regression coefficients, along with elementary but powerful procedures for constructing confidence intervals and performing hypothesis tests in simple and multiple regression.[1] I will also develop two topics related to inference in regression: the distinction between empirical and structural relationships and the consequences of random measurement error in regression.

6.1 Simple Regression

6.1.1 The Simple-Regression Model

Standard statistical inference in simple regression is based on a *statistical model*, assumed to be descriptive of the population or process that is sampled:

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

The coefficients α and β are the *population regression parameters*; the central object of simple-regression analysis is to estimate these coefficients. The *error* ε_i represents the aggregated omitted causes of Y (i.e., the causes of Y beyond the explanatory variable X), other explanatory variables that could have been included in the regression model (at least in principle), measurement error in Y, and whatever component of Y is inherently random. A Greek letter, epsilon, is used for the errors because, without knowledge of the values of α and β, the errors are not directly observable. 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.* The expectation of the error—that is, the average value of ε given the value of X—is 0: $E(\varepsilon_i) \equiv E(\varepsilon|x_i) = 0$. Equivalently, the expected value of the response variable is a linear function of the explanatory variable:

$$\begin{aligned}\mu_i \equiv E(Y_i) \equiv E(Y|x_i) &= E(\alpha + \beta x_i + \varepsilon_i)\\ &= \alpha + \beta x_i + E(\varepsilon_i)\\ &= \alpha + \beta x_i + 0\\ &= \alpha + \beta x_i\end{aligned}$$

  We can remove $\alpha + \beta x_i$ from the expectation operator because α and β are fixed parameters, while the value of X is conditionally fixed to x_i.[2]

[1] The focus here is on the procedures themselves: The statistical theory underlying these methods and some extensions are developed in Chapters 9 and 10.

[2] I use a lowercase x here to stress that the value x_i is fixed—either literally, as in experimental research (see below), or by conditioning on the observed value x_i of X_i.

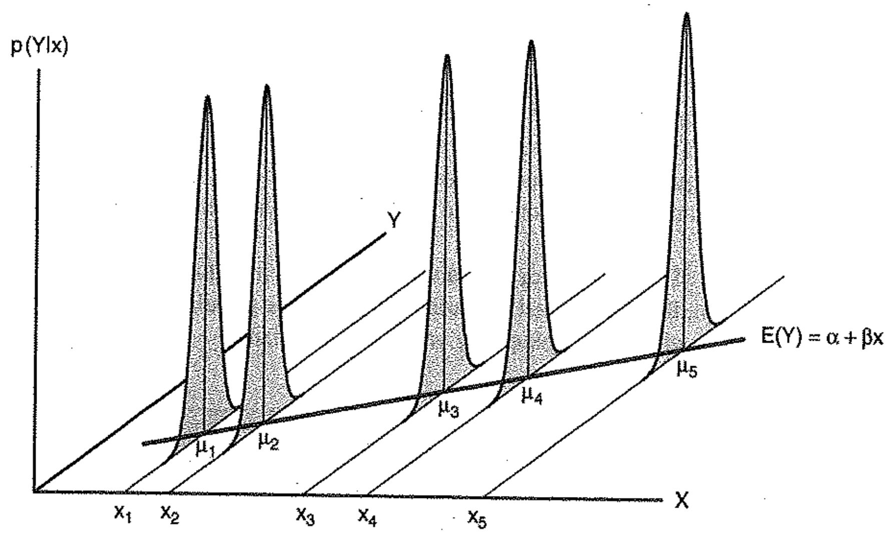

Figure 6.1 The assumptions of linearity, constant variance, and normality in simple regression. The graph shows the conditional population distributions $p(Y|x)$ of Y for several values of the explanatory variable X, labeled $x_1, \ldots, x_5$. The conditional means of Y given x are denoted $\mu_1, \ldots, \mu_5$. (Repeating Figure 2.4)

- *Constant variance.* The variance of the errors is the same regardless of the value of X: $V(\varepsilon|x_i) = \sigma_\varepsilon^2$. Because the distribution of the errors is the same as the distribution of the response variable around the population regression line, constant error variance implies constant conditional variance of Y given X:

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

  Note that because the mean of ε_i is 0, its variance is simply $E(\varepsilon_i^2)$.
- *Normality.* The errors are normally distributed: $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. Equivalently, the conditional distribution of the response variable 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 6.1. It should be abundantly clear from the graph that these assumptions place very strong constraints on the structure of the data.
- *Independence.* The observations are sampled independently: Any pair of errors ε_i and ε_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. For example, if the data constitute a simple random sample drawn from a large population, then the assumption of independence will be met to a close approximation. In contrast, if the data comprise a time series, then the assumption of independence may be very wrong.[3]
- *Fixed X, or X measured without error and independent of the error.* Depending on 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, in which the value of the explanatory variable is under the direct control of the researcher; if the experiment were replicated, then—at least in principle—the values of X would remain the same. Most social research, however,

[3] Chapter 16 discusses regression analysis with time-series data.

is observational, and therefore, X values are sampled, not fixed by design. Under these circumstances, we assume that the explanatory variable is measured without error and 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. This is in an important respect the most problematic of the assumptions underlying least-squares estimation because causal inference in nonexperimental research hinges on this assumption and because the assumption cannot be checked directly from the observed data.[4]

- *X is not invariant.* If the explanatory variable is fixed, then its values cannot all be the same; and if it is random, then there must be variation in X in the population. It is not possible to fit a line to data in which the explanatory variable is invariant.[5]

> Standard statistical inference for least-squares simple-regression analysis is based on the statistical model $Y_i = \alpha + \beta x_i + \varepsilon_i$. The key assumptions of the model concern the behavior of the errors ε_i: (1) Linearity, $E(\varepsilon_i) = 0$; (2) constant variance, $V(\varepsilon_i) = \sigma_\varepsilon^2$; (3) normality, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$; (4) independence, $\varepsilon_i, \varepsilon_j$ are independent for $i \neq j$; (5) the X values are fixed or, if random, are measured without error and are independent of the errors; and (6) X is not invariant.

6.1.2 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 α and β:[6]

- The least-squares intercept and slope are *linear estimators*, in the sense that they are linear functions of the observations Y_i. For example, for fixed explanatory-variable values x_i,

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

where

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

While unimportant in itself, this property makes it simple to derive the sampling distributions of A and B.

- The sample least-squares coefficients are *unbiased estimators* of the population regression coefficients:

$$E(A) = \alpha$$
$$E(B) = \beta$$

Only the assumption of linearity is required to establish this result.[7]

[4] See Sections 1.2, 6.3, and 9.7 for further discussion of causal inference from observational data.

[5] See Figure 5.3 on page 81.

[6] I will simply state and briefly explain these properties here; derivations can be found in the exercises to this chapter and in Chapter 9.

[7] See Exercise 6.1.

- Both A and B have simple sampling variances:

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

The assumptions of linearity, constant variance, and independence are employed in the derivation of these formulas.[8]

It is instructive to examine the formula for $V(B)$ more closely to understand the conditions under which least-squares estimation is precise. Rewriting the formula,

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

Thus, the sampling variance of the slope estimate will be small when (1) the error variance σ_ε^2 is small, (2) the sample size n is large, and (3) the explanatory-variable values are spread out (i.e., have a large variance, S_X^2). The estimate of the intercept has small sampling variance under similar circumstances and, in addition, when the X values are centered near 0 and, hence, $\sum x_i^2$ is not much larger than $\sum (x_i - \overline{x})^2$.[9]

- Of all the linear unbiased estimators, the least-squares estimators are the most efficient—that is, they have the smallest sampling variance and hence the smallest mean-squared error. This result, called the Gauss-Markov theorem, requires the assumptions of linearity, constant variance, and independence, but not the assumption of normality.[10] Under normality, moreover, the least-squares estimators are the most efficient among *all* unbiased estimators, not just among linear estimators. This is a much more compelling result, because the restriction to linear estimators is merely a matter of convenience. When the error distribution is heavier tailed than normal, for example, the least-squares estimators may be much less efficient than certain robust-regression estimators, which are *not* linear functions of the data.[11]
- Under the full suite of assumptions, the least-squares coefficients A and B are the maximum-likelihood estimators of α and β.[12]
- Under the assumption of normality, the least-squares coefficients are themselves normally distributed. Summing up,

$$A \sim N\left[\alpha, \frac{\sigma_\varepsilon^2 \sum x_i^2}{n \sum (x_i - \overline{x})^2}\right]$$
$$B \sim N\left[\beta, \frac{\sigma_\varepsilon^2}{\sum (x_i - \overline{x})^2}\right] \tag{6.1}$$

[8] See Exercise 6.2.
[9] See Exercise 6.3.
[10] The theorem is named after the 19th-century German mathematical genius Carl Friedrich Gauss and the 20th-century Russian mathematician A. A. Markov. Although Gauss worked in the context of measurement error in the physical sciences, much of the general statistical theory of linear models is due to him.
[11] See Chapter 19.
[12] See Exercise 6.5. For an explanation of maximum-likelihood estimation, see Appendix D on probability and estimation.

Even if the errors are not normally distributed, the distributions of A and B are approximately normal, under very broad conditions, with the approximation improving as the sample size grows.[13]

> Under the assumptions of the regression model, the least-squares coefficients have certain desirable properties as estimators of the population regression coefficients. The least-squares coefficients are linear functions of the data, and therefore have simple sampling distributions; unbiased estimators of the population regression coefficients; the most efficient unbiased estimators of the population regression coefficients; maximum-likelihood estimators; and normally distributed.

6.1.3 Confidence Intervals and Hypothesis Tests

The distributions of A and B, given in Equation 6.1, cannot be directly employed for statistical inference because the error variance, σ_ε^2, is never known in practice. The variance of the residuals provides an unbiased estimator of σ_ε^2:[14]

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

With the estimated error variance in hand, we can estimate the sampling variances of A and B:

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

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

As in statistical inference for the mean, 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 $100(1-a)\%$ confidence interval for the slope, we take

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

where $t_{a/2}$ is the critical value of t with $n-2$ degrees of freedom and a probability of $a/2$ to the right, and $\text{SE}(B)$ is the standard error of B [i.e., the square root of $\widehat{V}(B)$]. For a 95% confidence interval, $t_{.025} \approx 2$, unless n is very small. Similarly, to test the hypothesis, H_0: $\beta = \beta_0$, that the population slope is equal to a specific value (most commonly, the null hypothesis 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 the hypothesis H_0. Confidence intervals and hypothesis tests for α are usually of less interest, but they follow the same pattern.

[13]The asymptotic normality of A and B follows from the central-limit theorem, because the least-squares coefficients are linear functions of the Y_is.

[14]See Section 10.3.

The standard error of the slope coefficient B in simple regression is $\mathrm{SE}(B) = S_E/\sqrt{\sum(x_i - \overline{x})^2}$, which can be used to construct t-tests and t-intervals for β.

For Davis's regression of measured on reported weight (described in the preceding chapter), for example, we have the following results:

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

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

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

Because $t_{.025}$ for $101 - 2 = 99$ degrees of freedom is 1.9842, the 95% confidence intervals for α and β are

$$\alpha = 1.7775 \pm 1.9842 \times 1.7444 = 1.777 \pm 3.461$$

$$\beta = 0.97722 \pm 1.9842 \times 0.030529 = 0.9772 \pm 0.0606$$

The estimates of α and β are therefore quite precise. Furthermore, the confidence intervals include the values $\alpha = 0$ and $\beta = 1$, which, recall, imply unbiased prediction of measured weight from reported weight.[15]

6.2 Multiple Regression

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

6.2.1 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, $\varepsilon_i \equiv \varepsilon | x_{i1}, \ldots, x_{ik}$, 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:* $\varepsilon_i, \varepsilon_j$ are independent for $i \neq j$.
- *Fixed Xs or Xs measured without error and independent of ε.*

In addition, we assume that the Xs are not invariant and that no X is a perfect linear function of the others.[16]

---

[15] There is, however, a subtlety here: To construct separate confidence intervals for α and β is not quite the same as constructing a *joint confidence region* for both coefficients simultaneously. See Section 9.4.4 for a discussion of confidence regions in regression.

[16] We saw in Section 5.2.1 that when explanatory variables in regression are invariant or perfectly collinear, the least-squares coefficients are not uniquely defined.

Under these assumptions (or particular subsets of them), the least-squares estimators A, $B_1, \ldots, B_k$ of $\alpha, \beta_1, \ldots, \beta_k$ 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[17]

$$V(B_j) = \frac{1}{1-R_j^2} \times \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_{ij} - \overline{x}_j)^2}$$
$$= \frac{\sigma_\varepsilon^2}{\sum_{i=1}^n (x_{ij} - \widehat{x}_{ij})^2} \tag{6.2}$$

where R_j^2 is the squared multiple correlation from the regression of X_j on all the other Xs, and the $\widehat{x}_{ij}$ are the fitted values from this auxiliary regression. The second factor in the first line of Equation 6.2 is similar to the sampling variance of the slope in simple regression, although the error variance σ_ε^2 is smaller than before because some of the explanatory variables that were implicitly in the error in simple regression are now incorporated into the systematic part of the model. The first factor—called the *variance-inflation factor*—is new, however. The variance-inflation factor $1/(1-R_j^2)$ is large when the explanatory variable X_j is strongly correlated with other explanatory variables. The denominator in the second line of Equation 6.2 is the residual sum of squares from the regression of X_j on the other Xs, and it makes a similar point: When the *conditional* variation in X_j given the other Xs is small, the sampling variance of B_j is large.[18]

We saw in Chapter 5 that when one explanatory variable is perfectly collinear with others, the least-squares regression coefficients are not uniquely determined; in this case, the variance-inflation factor is infinite. The variance-inflation factor tells us that strong, though less-than-perfect, collinearity presents a problem for estimation, for although we can calculate least-squares estimates under these circumstances, their sampling variances may be very large. Equation 6.2 reveals that the other sources of imprecision of estimation in multiple regression are the same as in simple regression: large error variance, a small sample, and explanatory variables with little variation.[19]

6.2.2 Confidence Intervals and Hypothesis Tests

Individual Slope Coefficients

Confidence intervals and hypothesis tests for individual coefficients closely follow the pattern of simple-regression analysis: To find the standard error of a slope coefficient, we need to substitute an estimate of the error variance for the unknown σ_ε^2 in Equation 6.2. The variance of the residuals provides an unbiased estimator:

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

[17] Although we are usually less interested in inference about α, it is also possible to find the sampling variance of the intercept A. See Section 9.3.

[18] I am grateful to Georges Monette of York University for pointing this out to me.

[19] Collinearity is discussed further in Chapter 13.

Then, the standard error of B_j is

$$\text{SE}(B_j) = \frac{1}{\sqrt{1-R_j^2}} \times \frac{S_E}{\sqrt{\sum(x_{ij}-\overline{x}_j)^2}}$$

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

> The standard error of the slope coefficient B_j in multiple regression is $\text{SE}(B_j) = S_E/\sqrt{(1-R_j^2)\sum(x_{ij}-\overline{x}_j)^2}$. The coefficient standard error can be used in t-intervals and t-tests for β_j.

For example, for Duncan's regression of occupational prestige on education and income (from the previous chapter), we have

$$S_E^2 = \frac{7506.7}{45-2-1} = 178.73$$
$$r_{12} = .72451$$
$$\text{SE}(B_1) = \frac{1}{\sqrt{1-.72451^2}} \times \frac{\sqrt{178.73}}{\sqrt{38{,}971}} = 0.098252$$
$$\text{SE}(B_2) = \frac{1}{\sqrt{1-.72451^2}} \times \frac{\sqrt{178.73}}{\sqrt{26{,}271}} = 0.11967$$

With only two explanatory variables, $R_1^2 = R_2^2 = r_{12}^2$; this simplicity and symmetry are peculiar to the two-explanatory-variable case. To construct 95% confidence intervals for the slope coefficients, we use $t_{.025} = 2.0181$ from the t-distribution with $45-2-1=42$ degrees of freedom. Then,

$$\text{Education}: \quad \beta_1 = 0.54583 \pm 2.0181 \times 0.098252 = 0.5459 \pm 0.1983$$
$$\text{Income}: \quad \beta_2 = 0.59873 \pm 2.0181 \times 0.11967 = 0.5987 \pm 0.2415$$

Although they are far from 0, these confidence intervals are quite broad, indicating that the estimates of the education and income coefficients are imprecise—as is to be expected in a sample of only 45 occupations.

All Slopes

We can also test the null hypothesis that all the regression slopes are 0:

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_k = 0 \tag{6.3}$$

Testing this global or "omnibus" null hypothesis is not quite the same as testing the k separate hypotheses

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

If the explanatory variables are very highly correlated, for example, we might be able to reject the omnibus hypothesis (Equation 6.3) without being able to reject *any* of the individual hypotheses.

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

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

$$= \frac{n-k-1}{k} \times \frac{R^2}{1-R^2}$$

Under the omnibus null hypothesis, this test statistic has an F-distribution with k and $n-k-1$ degrees of freedom. The omnibus F-test follows from the analysis of variance for the regression, and the calculation of the test statistic can be organized in an *analysis-of-variance table*, which shows the partition of total variation into its components:

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	$n-1$		

Note that the degrees of freedom (*df*) add in the same manner as the sums of squares and that the *residual mean square*, *RMS*, is simply the estimated error variance, S_E^2.

It turns out that when the null hypothesis is true, the *regression mean square*, *RegMS*, provides an independent estimate of the error variance, so the ratio of the two mean squares should be close to 1. When, alternatively, the null hypothesis is false, the RegMS estimates the error variance plus a positive quantity that depends on the βs, tending to make the numerator of F_0 larger than the denominator:

$$E(F_0) \approx \frac{E(\text{RegMS})}{E(\text{RMS})} = \frac{\sigma_\varepsilon^2 + \text{positive quantity}}{\sigma_\varepsilon^2} > 1$$

We consequently reject the omnibus null hypothesis for values of F_0 that are sufficiently larger than 1.[20]

> An omnibus F-test for the null hypothesis that all the slopes are 0 can be calculated from the analysis of variance for the regression.

[20] The reasoning here is only approximate because the expectation of the ratio of two independent random variables is not the ratio of their expectations. Nevertheless, when the sample size is large, the null distribution of the F-statistic has an expectation very close to 1. See Appendix D on probability and estimation for information about the F-distribution.

For Duncan's regression, we have the following analysis-of-variance table:

Source	Sum of Squares	df	Mean Square	F	p
Regression	36181.	2	18090.	101.2	≪.0001
Residuals	7506.7	42	178.73		
Total	43688.	44			

The p-value for the omnibus null hypothesis—that is, $\Pr(F > 101.2)$ for an F-distribution with 2 and 42 degrees of freedom—is very close to 0.

A Subset of Slopes

It is, finally, possible to test a null hypothesis about a *subset* of the regression slopes

$$H_0\colon \beta_1 = \beta_2 = \cdots = \beta_q = 0 \tag{6.4}$$

where $1 \leq q \leq k$. Purely for notational convenience, I have specified a hypothesis on the *first* q coefficients; we can, of course, equally easily test a hypothesis for *any* q slopes. The "full" regression model, including all the explanatory variables, can be written as

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

If the null hypothesis is correct, then the first q of the βs are 0, yielding the "null" model

$$\begin{aligned} Y_i &= \alpha + 0x_{i1} + \cdots + 0x_{iq} + \beta_{q+1} x_{i,q+1} + \cdots + \beta_k x_{ik} + \varepsilon_i \\ &= \alpha + \beta_{q+1} x_{i,q+1} + \cdots + \beta_k x_{ik} + \varepsilon_i \end{aligned}$$

In effect, then, 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 in Equation 6.4 is based on a comparison of these two models. Let RSS_1 and RegSS_1 represent, respectively, the residual and regression sums of squares for the full model; similarly, RSS_0 and RegSS_0 are the residual and regression sums of squares for the null model. Because the null model is *nested within* (i.e., is a special case of) the full model, constraining the first q slopes to 0, $\text{RSS}_0 \geq \text{RSS}_1$. The residual and regression sums of squares in the two models add to the same total sum of squares; it follows that $\text{RegSS}_0 \leq \text{RegSS}_1$. If the null hypothesis is wrong and (some of) $\beta_1, \ldots, \beta_q$ 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 in Equation 6.4 is

$$\begin{aligned} F_0 &= \frac{(\text{RegSS}_1 - \text{RegSS}_0)/q}{\text{RSS}_1/(n-k-1)} \\ &= \frac{n-k-1}{q} \times \frac{R_1^2 - R_0^2}{1 - R_1^2} \end{aligned}$$

where R_1^2 and R_0^2 are the squared multiple correlations from the full and null models, respectively. Under the null hypothesis, this test statistic has an F-distribution with q and $n - k - 1$ degrees of freedom.

> An F-test for the null hypothesis that a subset of slope coefficients is 0 is based on a comparison of the regression sums of squares for two models: the full regression model and a null model that deletes the explanatory variables in the null hypothesis.

The motivation for testing a subset of coefficients will become clear in the next chapter, which takes up regression models that incorporate qualitative explanatory variables. I will, for the present, illustrate the incremental F-test by applying it to the trivial case in which $q = 1$ (i.e., a single coefficient).

In Duncan's data set, the regression of prestige on income alone produces $\text{RegSS}_0 = 30{,}665$, while 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 $36,181 - 30,665 = 5516$. The F-statistic for testing H_0: $\beta_{\text{Education}} = 0$ is, then,

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

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

When, as here, $q = 1$, the incremental F-test is equivalent to the t-test obtained by dividing the regression coefficient by its estimated standard error: $F_0 = t_0^2$. For the current example,

$$t_0 = \frac{0.54583}{0.098252} = 5.5554$$
$$t_0^2 = 5.5554^2 = 30.86$$

(which is the same as F_0).

6.3 Empirical Versus Structural Relations

There are two fundamentally different interpretations of regression coefficients, and failure to distinguish clearly between them is the source of much confusion. 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 because the notion is simpler. Suppose that, in a population of interest, the relationship between two variables, Y and X_1, is well described by the simple-regression model:[21]

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

That is to say, the conditional mean of Y is a linear function of X. We do not assume that X_1 necessarily causes Y or, if it does, that the omitted causes of Y, incorporated in ε', are independent of X_1. There is, quite simply, a linear empirical relationship between Y and X_1 in the population. If we proceed to draw a random sample from this population, then the least-squares sample slope B_1' is an unbiased estimator of β_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 Xs is linear:

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

[21] Because this discussion applies to observational data, where the explanatory variables are random, I use uppercase Xs.

That is, the conditional mean of Y is a linear function of X_1 and X_2. The slope β_1 of the population regression plane can, and generally will, differ from β_1', the simple-regression slope (see below). The sample least-squares coefficients for the multiple regression, B_1 and B_2, are unbiased estimators of the corresponding population coefficients, β_1 and β_2.

That the simple-regression slope β_1' differs from the multiple-regression slope β_1 and that, therefore, the sample *simple*-regression coefficient B_1' is a *biased* estimator of the population *multiple*-regression slope β_1 is not problematic, for these are simply empirical relationships, and 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*—fitting a false model to the data—does not arise, as long as the linear regression model adequately describes the empirical relationship between the response variable and the explanatory variables in the population. This would not be the case, for example, if the relationship in the population were nonlinear.

The situation is different, however, if we view the regression equation as representing a structural relation—that is, a model of how response-variable scores are determined.[22] Imagine now that response-variable scores are *constructed* according to the multiple-regression model

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{6.5}$$

where the error ε satisfies the usual regression assumptions; in particular, $E(\varepsilon) = 0$ and ε is independent of X_1 and X_2.

If we use least squares to fit this model to sample data, we obtain unbiased estimators of β_1 and β_2. Suppose, however, that instead we fit the simple-regression model

$$Y = \alpha + \beta_1 X_1 + \varepsilon' \tag{6.6}$$

where, implicitly, the effect of X_2 on Y is absorbed by the error $\varepsilon' \equiv \varepsilon + \beta_2 X_2$ because X_2 is now among the omitted causes of Y. In the event that X_1 and X_2 are correlated, there is a correlation induced between X_1 and ε'. If we proceed to assume wrongly that X_1 and ε' are *uncorrelated*, as we do if we fit the model in Equation 6.6 by least squares, then we make an error of specification. The consequence of this error is that our simple-regression estimator of β_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 Equation 6.5, obtaining

$$\mu_Y = \alpha + \beta_1 \mu_1 + \beta_2 \mu_2 + 0 \tag{6.7}$$

where, for example, μ_Y is the population mean of Y; to obtain Equation 6.7, we use the fact that $E(\varepsilon)$ is 0. Subtracting this equation from Equation 6.5 has the effect of eliminating the constant α and expressing the variables as deviations from their population means:

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

Next, multiply this equation through by $X_1 - \mu_1$:

$$(X_1 - \mu_1)(Y - \mu_Y) = \beta_1(X_1 - \mu_1)^2 + \beta_2(X_1 - \mu_1)(X_2 - \mu_2) + (X_1 - \mu_1)\varepsilon$$

Taking the expectation of both sides of the equation produces

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

[22] In the interest of clarity, I am making this distinction more categorically than I believe is justified. I argued in Chapter 1 that it is unreasonable to treat statistical models as literal representations of social processes. Nevertheless, it is useful to distinguish between purely empirical descriptions and descriptions from which we intend to infer causation.

where σ_{1Y} is the covariance between X_1 and Y, σ_1^2 is the variance of X_1, and σ_{12} is the covariance of X_1 and X_2.[23] Solving for β_1, we get

$$\beta_1 = \frac{\sigma_{1Y}}{\sigma_1^2} - \beta_2 \frac{\sigma_{12}}{\sigma_1^2} \tag{6.8}$$

Recall that 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 β_1 but rather $\sigma_{1Y}/\sigma_1^2 \equiv \beta_1'$. Solving Equation 6.8 for β_1' produces $\beta_1' = \beta_1 +$ bias, where bias $= \beta_2\sigma_{12}/\sigma_1^2$.

It is instructive to take a closer look at the bias in the simple-regression estimator. For the bias to be nonzero, two conditions must be met: (1) X_2 must be a *relevant* explanatory variable—that is, $\beta_2 \neq 0$; and (2) X_1 and X_2 must be *correlated*—that is, $\sigma_{12} \neq 0$. Moreover, depending on the signs of β_2 and σ_{12}, the bias in the simple-regression estimator may be either positive or negative.

> 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 by incorporating the omitted explanatory variable in the error.

There is one final subtlety: The proper interpretation of the "bias" in the simple-regression estimator depends on the nature of the causal relationship between X_1 and X_2. Consider the situation depicted in Figure 6.2(a), where 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, because σ_{12}/σ_1^2 is the population slope for the regression of X_2 on X_1. If, however, as in Figure 6.2(b), X_2 is a *common prior cause* of both X_1 and Y, then the bias term represents a *spurious*—that is, noncausal—component of the empirical association between X_1 and Y. In the latter event, but not in the former, it is critical to control for X_2 in examining the relationship between Y and X_1.[24] An omitted common prior cause that accounts (or partially accounts) for the association between two variables is sometimes called a "lurking variable." It is the always-possible existence of lurking variables that makes it difficult to infer causation from observational data.

6.4 Measurement Error in Explanatory Variables*

Variables are rarely—if ever—measured without error.[25] Even relatively straightforward characteristics, such as education, income, height, and weight, are imperfectly measured, especially

[23] This result follows from the observation that the expectation of a mean-deviation product is a covariance, and the expectation of a mean-deviation square is a variance (see Appendix D on probability and estimation). $E[(X_1 - \mu_1)\varepsilon] = \sigma_{1\varepsilon}$ is 0 because of the independence of X_1 and the error.

[24] Note that panels (a) and (b) in Figure 6.2 simply exchange the roles of X_1 and X_2.

[25] Indeed, one of the historical sources of statistical theory in the 18th and 19th centuries was the investigation of measurement errors in the physical sciences by great mathematicians like Gauss (mentioned previously) and Pierre Simon Laplace.

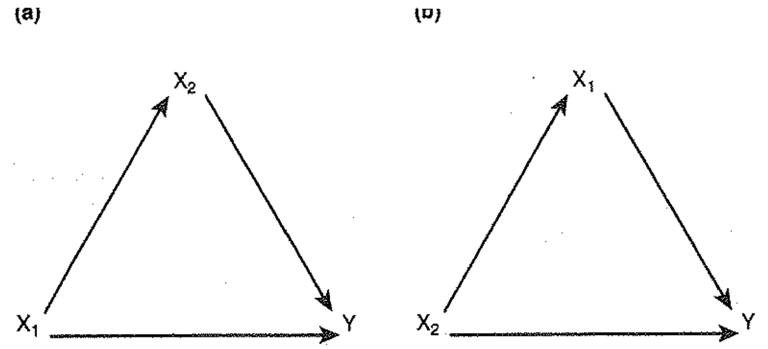

Figure 6.2 Two causal schemes relating a response variable to two explanatory variables: In (a) X_2 intervenes causally between X_1 and Y, while in (b) X_2 is a common prior cause of both X_1 and Y. In the second case, but not in the first, it is important to control for X_2 in examining the effect of X_1 on Y.

when we rely on individuals' verbal reports. Measures of "subjective" characteristics, such as racial prejudice and conservatism, almost surely have substantial components of error. Measurement error affects not only characteristics of individuals: As you are likely aware, official statistics relating to crime, the economy, and so on, are also subject to a variety of measurement errors.

The regression model accommodates measurement error in the *response* variable, because measurement error can be conceptualized as a component of the general error term ε, but the explanatory variables in regression analysis are assumed to be measured without error. In this section, I will explain the consequences of violating this assumption. To do so, we will examine the multiple-regression equation

$$Y = \beta_1 \tau + \beta_2 X_2 + \varepsilon \tag{6.9}$$

To keep the notation as simple as possible, all the variables in Equation 6.9 are expressed as deviations from their expectations, so the constant term disappears from the regression equation.[26] One of the explanatory variables, X_2, is measured without error, but the other, τ, is not directly observable. Instead, we have a fallible *indicator* X_1 of τ:

$$X_1 = \tau + \delta \tag{6.10}$$

where δ represents measurement error.

In addition to the usual assumptions about the regression errors ε, I will assume that the measurement errors δ are "random" and "well behaved"; in particular,

- $E(\delta) = 0$, so there is no systematic tendency for measurements to be too large or too small.
- The measurement errors δ are uncorrelated with the "true-score" variable τ. This assumption could easily be wrong. If, for example, individuals who are lighter than average tend to overreport their weights and individuals who are heavier than average tend to underreport their weights, then there will be a negative correlation between the measurement errors and true weight.
- The measurement errors δ are uncorrelated with the regression errors ε and with the other explanatory variable X_2.

[26]There is no loss of generality here, because we can always subtract the mean from each variable. See the previous section.

Because $\tau = X_1 - \delta$, we can rewrite Equation 6.9 as

$$\begin{aligned} Y &= \beta_1(X_1 - \delta) + \beta_2 X_2 + \varepsilon \\ &= \beta_1 X_1 + \beta_2 X_2 + (\varepsilon - \beta_1 \delta) \end{aligned} \tag{6.11}$$

As in the previous section, we can proceed by multiplying Equation 6.11 through by X_1 and X_2 and taking expectations; because all variables are in mean-deviation form, expected products are covariances and expected squares are variances:[27]

$$\begin{aligned} \sigma_{Y1} &= \beta_1 \sigma_1^2 + \beta_2 \sigma_{12} - \beta_1 \sigma_\delta^2 \\ \sigma_{Y2} &= \beta_1 \sigma_{12} + \beta_2 \sigma_2^2 \end{aligned} \tag{6.12}$$

Then, solving for the regression coefficients,

$$\begin{aligned} \beta_1 &= \frac{\sigma_{Y1}\sigma_2^2 - \sigma_{12}\sigma_{Y2}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2 - \sigma_\delta^2\sigma_2^2} \\ \beta_2 &= \frac{\sigma_{Y2}\sigma_1^2 - \sigma_{12}\sigma_{Y1}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} - \frac{\beta_1\sigma_{12}\sigma_\delta^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \end{aligned} \tag{6.13}$$

Suppose, now, that we (temporarily) ignore the measurement error in X_1, and proceed by least-squares regression of Y on X_1 and X_2. The population analogs of the least-squares regression coefficients are as follows:[28]

$$\begin{aligned} \beta_1' &= \frac{\sigma_{Y1}\sigma_2^2 - \sigma_{12}\sigma_{Y2}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \\ \beta_2' &= \frac{\sigma_{Y2}\sigma_1^2 - \sigma_{12}\sigma_{Y1}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} \end{aligned} \tag{6.14}$$

Comparing Equations 6.13 and 6.14 reveals the consequences of ignoring the measurement error in X_1. The denominator of β_1 in Equation 6.13 is necessarily positive, and its component $-\sigma_\delta^2\sigma_2^2$ is necessarily negative. Ignoring this component therefore inflates the denominator of β_1' in Equation 6.14, driving the coefficient β_1' toward 0. Put another way, ignoring measurement error in an explanatory variable tends to *attenuate* its coefficient, which makes intuitive sense.

The effect of measurement error in X_1 on the coefficient of X_2 is even more pernicious. Here, we can write $\beta_2' = \beta_2 + \text{bias}$, where

$$\text{bias} = \frac{\beta_1\sigma_{12}\sigma_\delta^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}$$

The bias term can be positive or negative, toward 0 or away from it. To get a better grasp on the bias in the least-squares *estimand* β_2', imagine that the measurement error variance σ_δ^2 grows larger and larger. Because σ_δ^2 is a component of σ_1^2, this latter quantity also grows larger, but because the measurement errors δ are uncorrelated with variables other than X_1, other variances and covariances are unaffected.[29]

[27]See Exercise 6.10.
[28]See Exercise 6.11.
[29]See Exercise 6.12.

Using Equation 6.14,

$$\lim_{\sigma_\delta^2 \to \infty} \beta_2' = \frac{\sigma_{Y2}\sigma_1^2}{\sigma_1^2\sigma_2^2} = \frac{\sigma_{Y2}}{\sigma_2^2}$$

which is the population analog of the least-squares slope for the *simple* regression of Y on X_2 alone. Once more, the result is simple and intuitively plausible: Substantial measurement error in X_1 renders it an ineffective statistical control, driving β_2' toward the marginal relationship between X_2 and Y, and away from the partial relationship between these two variables.[30]

> Measurement error in an explanatory variable tends to attenuate its regression coefficient and to make the variable an imperfect statistical control.

Although there are statistical methods that attempt to estimate regression equations taking account of measurement errors, these methods are beyond the scope of the presentation in this book and, in any event, involve assumptions that are difficult to justify in practice.[31] Perhaps the most important lessons to be drawn from the results of this section are (1) that large measurement errors in the Xs can invalidate a regression analysis; (2) that, therefore, where measurement errors are likely to be substantial, we should not view the results of a regression as definitive; and (3) that it is worthwhile to expend effort to improve the quality of social measurements.

Exercises

Exercise 6.1. *Demonstrate the unbias of the least-squares estimators A and B of α and β in simple regression:

(a) Expressing the least-squares slope B as a linear function of the observations, $B = \sum m_i Y_i$ (as in the text), and using the assumption of linearity, $E(Y_i) = \alpha + \beta x_i$, show that $E(B) = \beta$. [*Hint*: $E(B) = \sum m_i E(Y_i)$.]
(b) Show that A can also be written as a linear function of the Y_is. Then, show that $E(A) = \alpha$.

Exercise 6.2. *Using the assumptions of linearity, constant variance, and independence, along with the fact that A and B can each be expressed as a linear function of the Y_is, derive the sampling variances of A and B in simple regression. [*Hint*: $V(B) = \sum m_i^2 V(Y_i)$.]

Exercise 6.3. Examining the formula for the sampling variance of A in simple regression,

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

why is it intuitively sensible that the variance of A is large when the mean of the xs is far from 0? Illustrate your explanation with a graph.

---

[30] I am grateful to Georges Monette, of York University, for this insight. See Exercise 6.13 for an illustration.

[31] Measurement errors in explanatory variables are often discussed in the context of *structural-equation models*, which are multiple-equation regression models in which the response variable in one equation can appear as an explanatory variable in others. Duncan (1975, Chapters 9 and 10) presents a fine elementary treatment of the topic, part of which I have adapted for the presentation in this section. A more advanced development may be found in Bollen (1989).

Exercise 6.4. The formula for the sampling variance of B in simple regression,

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

shows that, to estimate β precisely, it helps to have spread out xs. Explain why this result is intuitively sensible, illustrating your explanation with a graph. What happens to $V(B)$ when there is *no* variation in X?

Exercise 6.5. *Maximum-likelihood estimation of the simple-regression model: Deriving the maximum-likelihood estimators of α and β in simple regression is straightforward. Under the assumptions of the model, the Y_is are independently and normally distributed random variables with expectations $\alpha + \beta x_i$ and common variance σ_ε^2. Show that if these assumptions hold, then the least-squares coefficients A and B are the maximum-likelihood estimators of α and β and that $\widehat{\sigma}_\varepsilon^2 = \sum E_i^2/n$ is the maximum-likelihood estimator of σ_ε^2. Note that the MLE of the error variance is biased. (*Hints*: Because of the assumption of independence, the joint probability density for the Y_is is the product of their marginal probability densities

$$p(y_i) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left[-\frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_\varepsilon^2}\right]$$

Find the log-likelihood function; take the partial derivatives of the log likelihood with respect to the parameters α, β, and σ_ε^2; set these partial derivatives to 0; and solve for the maximum-likelihood estimators.) A more general result is proved in Section 9.3.3

Exercise 6.6. Linear transformation of X and Y in simple regression (continuation of Exercise 5.4):

(a) Suppose that the X values in Davis's regression of measured on reported weight are transformed according to the equation $X' = 10(X - 1)$ and that Y is regressed on X'. Without redoing the regression calculations in detail, find $\mathrm{SE}(B')$ and $t_0' = B'/\mathrm{SE}(B')$.
(b) Now, suppose that the Y values are transformed according to the equation $Y'' = 5(Y + 2)$ and that Y'' is regressed on X. Find $\mathrm{SE}(B'')$ and $t_0'' = B''/\mathrm{SE}(B'')$.
(c) In general, how are hypothesis tests and confidence intervals for β affected by linear transformations of X and Y?

Exercise 6.7. Consider the regression model $Y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. How can the incremental-sum-of-squares approach be used to test the hypothesis that the two population slopes are equal to each other, H_0: $\beta_1 = \beta_2$? [*Hint*: Under H_0, the model becomes $Y = \alpha + \beta x_1 + \beta x_2 + \varepsilon = Y = \alpha + \beta(x_1 + x_2) + \varepsilon$, where β is the common value of β_1 and β_2.] Under what circumstances would a hypothesis of this form be meaningful? (*Hint*: Consider the units of measurement of x_1 and x_2.) Now, test the hypothesis that the "population" regression coefficients for education and income in Duncan's occupational prestige regression are equal to each other. Is this test sensible?

Exercise 6.8. Examples of specification error (also see the discussion in Section 9.7):

(a) Describe a nonexperimental research situation—real or contrived—in which failure to control statistically for an omitted variable induces a correlation between the error and an explanatory variable, producing erroneous conclusions. (For example: An educational researcher discovers that university students who study more get lower grades on average; the researcher concludes that studying has an adverse effect on students' grades.)

(b) Describe an experiment—real or contrived—in which faulty experimental practice induces an explanatory variable to become correlated with the error, compromising the validity of the results produced by the experiment. (For example: In an experimental study of a promising new therapy for depression, doctors administering the treatments tend to use the new therapy with patients for whom more traditional approaches have failed; it is discovered that subjects receiving the new treatment tend to do worse, on average, than those receiving older treatments or a placebo; the researcher concludes that the new treatment is not effective.)

(c) Is it fair to conclude that a researcher is *never* able absolutely to rule out the possibility that an explanatory variable of interest is correlated with the error? Is experimental research no better than observational research in this respect? Explain your answer.

Exercise 6.9. Suppose that the "true" model generating a set of data is $Y = \alpha + \beta_1 X_1 + \varepsilon$, where the error ε conforms to the usual linear-regression assumptions. A researcher fits the model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, which includes the irrelevant explanatory variable X_2—that is, the true value of β_2 is 0. Had the researcher fit the (correct) simple-regression model, the variance of B_1 would have been $V(B_1) = \sigma_\varepsilon^2 / \sum(X_{i1} - \overline{X}_1)^2$.

(a) Is the model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ wrong? Is B_1 for this model a biased estimator of β_1?

(b) The variance of B_1 in the multiple-regression model is

$$V(B_1) = \frac{1}{1 - r_{12}^2} \times \frac{\sigma_\varepsilon^2}{\sum(X_{i1} - \overline{X}_1)^2}$$

What, then, is the cost of including the irrelevant explanatory variable X_2? How does this cost compare to that of *failing* to include a relevant explanatory variable?

Exercise 6.10. *Derive Equation 6.12 by multiplying Equation 6.11 through by each of X_1 and X_2. (*Hints*: Both X_1 and X_2 are uncorrelated with the regression error ε. Likewise, X_2 is uncorrelated with the measurement error δ. Show that the covariance of X_1 and δ is simply the measurement error variance σ_δ^2 by multiplying $X_1 = \tau + \delta$ through by δ and taking expectations.)

Exercise 6.11. *Show that the population analogs of the regression coefficients can be written as in Equation 6.14. (*Hint*: Ignore the measurement errors, and derive the population analogs of the normal equations by multiplying the "model" $Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ through by each of X_1 and X_2 and taking expectations.)

Exercise 6.12. *Show that the variance of $X_1 = \tau + \delta$ can be written as the sum of "true-score variance," σ_τ^2, and measurement error variance, σ_δ^2. (*Hint*: Square both sides of Equation 6.10 and take expectations.)

Exercise 6.13. Recall Duncan's regression of occupational prestige on the educational and income levels of occupations. Following Duncan, regress prestige on education and income. Also, perform a simple regression of prestige on income alone. Then add random measurement errors to education. Sample these measurement errors from a normal distribution with mean 0, repeating the exercise for each of the following measurement error variances: $\sigma_\delta^2 = 10^2, 25^2, 50^2, 100^2$. In each case, recompute the regression of prestige on income and education. Then, treating the initial multiple regression as corresponding to $\sigma_\delta^2 = 0$, plot the coefficients of education and income as a function of σ_δ^2. What happens to the education coefficient as measurement error in education grows? What happens to the income coefficient?

Summary

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

$$Y_i = \alpha + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

The key assumptions of the model concern the behavior of the errors ε_i:

1. *Linearity:* $E(\varepsilon_i) = 0$.
2. *Constant variance:* $V(\varepsilon_i) = \sigma_\varepsilon^2$.
3. *Normality:* $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.
4. *Independence:* $\varepsilon_i, \varepsilon_j$ are independent for $i \neq j$.
5. The X values are fixed or, if random, are measured without error and are independent of the errors.

In addition, we assume that the Xs are not invariant, and that no X is a perfect linear function of the others.

- Under these assumptions, or particular subsets of them, the least-squares coefficients have certain desirable properties as estimators of the population regression coefficients. The least-squares coefficients are

1. linear functions of the data and therefore have simple sampling distributions,
2. unbiased estimators of the population regression coefficients,
3. the most efficient unbiased estimators of the population regression coefficients,
4. maximum-likelihood estimators, and
5. normally distributed.

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

$$\text{SE}(B) = \frac{S_E}{\sqrt{\sum(x_i - \overline{x})^2}}$$

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

$$\text{SE}(B_j) = \frac{1}{\sqrt{1 - R_j^2}} \times \frac{S_E}{\sqrt{\sum(x_{ij} - \overline{x}_j)^2}}$$

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

- An F-test for the omnibus null hypothesis that all the slopes are 0 can be calculated from the analysis of variance for the regression

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

- There is an incremental F-test for the null hypothesis that a subset of q slope coefficients is 0. This test is based on 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:

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

  This 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 by incorporating the omitted explanatory variable in the error.
- Measurement error in an explanatory variable tends to attenuate its regression coefficient and to make the variable an imperfect statistical control.