CHAPTER 3

Interval Estimation and Hypothesis Testing

LEARNING OBJECTIVES

Based on the material in this chapter, you should be able to

1. Discuss how "sampling theory" relates to interval estimation and hypothesis testing.
2. Explain why it is important for statistical inference that given **x** the least squares estimators b_1 and b_2 are normally distributed random variables.
3. Explain the "level of confidence" of an interval estimator, and exactly what it means in a sampling context, and give an example.
4. Explain the difference between an interval estimator and an interval estimate. Explain how to interpret an interval estimate.
5. Explain the terms null hypothesis, alternative hypothesis, and rejection region, giving an example and a sketch of the rejection region.
6. Explain the logic of a statistical test, including why it is important that a test statistic has a known probability distribution if the null hypothesis is true.
7. Explain the term *p*-value and how to use a *p*-value to determine the outcome of a hypothesis test; provide a sketch showing a *p*-value.
8. Explain the difference between one-tail and two-tail tests. Explain, intuitively, how to choose the rejection region for a one-tail test.
9. Explain Type I error and illustrate it in a sketch. Define the level of significance of a test.
10. Explain the difference between economic and statistical significance.
11. Explain how to choose what goes in the null hypothesis and what goes in the alternative hypothesis.

KEYWORDS

alternative hypothesis
confidence intervals
critical value
degrees of freedom
hypotheses
hypothesis testing
inference
interval estimation
level of significance
linear combination of parameters
linear hypothesis
null hypothesis
one-tail tests
pivotal statistic
point estimates
probability value
p-value
rejection region
test of significance
test statistic
two-tail tests
Type I error
Type II error

In Chapter 2, we used the least squares estimators to develop **point estimates** for the parameters in the simple linear regression model. These estimates represent an **inference** about the regression function $E(y|x) = \beta_1 + \beta_2 x$ describing a relationship between economic variables. *Infer* means "to conclude by reasoning from something known or assumed." This dictionary definition describes statistical inference as well. We have assumed a relationship between economic variables and made various assumptions (SR1–SR5) about the regression model. Based on these assumptions, and given empirical estimates of regression parameters, we want to make inferences about the population from which the data were obtained.

In this chapter, we introduce additional tools of statistical inference: **interval estimation** and **hypothesis testing**. Interval estimation is a procedure for creating ranges of values, sometimes called **confidence intervals**, in which the unknown parameters are likely to be located. Hypothesis tests are procedures for comparing conjectures that we might have about the regression parameters to the parameter estimates we have obtained from a sample of data. Hypothesis tests allow us to say that the data are compatible, or are not compatible, with a particular conjecture or hypothesis.

The procedures for hypothesis testing and interval estimation depend very heavily on assumption SR6 of the simple linear regression model and the resulting conditional normality of the least squares estimators. If assumption SR6 does not hold, then the sample size must be sufficiently large so that the distributions of the least squares estimators are *approximately* normal. In this case, the procedures we develop in this chapter can be used but are also approximate. In developing the procedures in this chapter, we will be using the "Student's" t-distribution. You may want to refresh your memory about this distribution by reviewing Appendix B.3.7. In addition, it is sometimes helpful to see the concepts we are about to discuss in a simpler setting. In Appendix C, we examine statistical inference, interval estimation, and hypothesis testing in the context of estimating the mean of a normal population. You may want to review this material now or read it along with this chapter as we proceed.

3.1 Interval Estimation

In Chapter 2, in Example 2.4, we estimated that household food expenditure would rise by \$10.21 given a \$100 increase in weekly income. The estimate $b_2 = 10.21$ is a *point* estimate of the unknown population parameter β_2 in the regression model. Interval estimation proposes a range of values in which the true parameter β_2 is likely to fall. Providing a range of values gives a sense of what the parameter value might be, and the precision with which we have estimated it. Such intervals are often called **confidence intervals**. We prefer to call them **interval estimates** because the term "confidence" is widely misunderstood and misused. As we will see, our confidence is in the procedure we use to obtain the intervals, not in the intervals themselves. This is consistent with how we assessed the properties of the least squares estimators in Chapter 2.

3.1.1 The t-Distribution

Let us assume that assumptions SR1–SR6 hold for the simple linear regression model. In this case, we know that given $\mathbf{x}$ the least squares estimators b_1 and b_2 have normal distributions, as discussed in Section 2.6. For example, the normal distribution of b_2, the least squares estimator of β_2, is

$$b_2|\mathbf{x} \sim N\left(\beta_2, \frac{\sigma^2}{\sum(x_i - \bar{x})^2}\right)$$

A standardized normal random variable is obtained from b_2 by subtracting its mean and dividing by its standard deviation:

$$Z = \frac{b_2 - \beta_2}{\sqrt{\sigma^2 / \sum(x_i - \bar{x})^2}} \sim N(0, 1) \tag{3.1}$$

The standardized random variable Z is normally distributed with mean 0 and variance 1. By standardizing the conditional normal distribution of $b_2|\mathbf{x}$, we find a statistic Z whose $N(0, 1)$ sampling distribution does not depend on any unknown parameters or on $\mathbf{x}$! Such statistics are called **pivotal**, and this means that when making probability statements about Z we do not have to worry about whether $\mathbf{x}$ is fixed or random. Using a table of normal probabilities (Statistical Table 1) we know that

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

Substituting (3.1) into this expression, we obtain

$$P\left(-1.96 \leq \frac{b_2 - \beta_2}{\sqrt{\sigma^2 / \sum\left(x_i - \bar{x}\right)^2}} \leq 1.96\right) = 0.95$$

Rearranging gives us

$$P\left(b_2 - 1.96\sqrt{\sigma^2 / \sum\left(x_i - \bar{x}\right)^2} \leq \beta_2 \leq b_2 + 1.96\sqrt{\sigma^2 / \sum\left(x_i - \bar{x}\right)^2}\right) = 0.95$$

This defines an interval that has probability 0.95 of containing the parameter β_2. The two endpoints $\left(b_2 \pm 1.96\sqrt{\sigma^2 / \sum\left(x_i - \bar{x}\right)^2}\right)$ provide an interval estimator. If we construct intervals this way using all possible samples of size N from a population, then 95% of the intervals will contain the true parameter β_2. This easy derivation of an interval estimator is based on both assumption SR6 *and* our knowing the variance of the error term σ^2.

Although we do not know the value of σ^2, we can estimate it. The least squares residuals are $\hat{e}_i = y_i - b_1 - b_2 x_i$, and our estimator of σ^2 is $\hat{\sigma}^2 = \sum \hat{e}_i^2/(N-2)$. Replacing σ^2 by $\hat{\sigma}^2$ in (3.1) creates a random variable we can work with, but this substitution changes the probability distribution from standard normal to a t-distribution with $N - 2$ **degrees of freedom**,

$$t = \frac{b_2 - \beta_2}{\sqrt{\hat{\sigma}^2 / \sum\left(x_i - \bar{x}\right)^2}} = \frac{b_2 - \beta_2}{\sqrt{\widehat{\operatorname{var}}\left(b_2\right)}} = \frac{b_2 - \beta_2}{\operatorname{se}\left(b_2\right)} \sim t_{(N-2)} \tag{3.2}$$

The ratio $t = \left(b_2 - \beta_2\right)/\operatorname{se}\left(b_2\right)$ has a t-distribution with $N-2$ degrees of freedom, which we denote as $t \sim t_{(N-2)}$. By standardizing the conditional normal distribution of $b_2|\mathbf{x}$ and inserting the estimator $\hat{\sigma}^2$, we find a statistic t whose $t_{(N-2)}$ sampling distribution does not depend on any unknown parameters or on $\mathbf{x}$! It too is a **pivotal statistic**, and when making probability statements with a t-statistic, we do not have to worry about whether $\mathbf{x}$ is fixed or random. A similar result holds for b_1, so in general we can say, if assumptions SR1–SR6 hold in the simple linear regression model, then

$$t = \frac{b_k - \beta_k}{\operatorname{se}\left(b_k\right)} \sim t_{(N-2)} \quad \text{for} \quad k = 1, 2 \tag{3.3}$$

This equation will be the basis for interval estimation and hypothesis testing in the simple linear regression model. The statistical argument of how we go from (3.1) to (3.2) is in Appendix 3A.

When working with the t-distribution, remember that it is a bell-shaped curve centered at zero. It looks like the standard normal distribution, except that it is more spread out, with a larger variance and thicker tails. The shape of the t-distribution is controlled by a single parameter called the **degrees of freedom**, often abbreviated as df. We use the notation $t_{(m)}$ to specify a t-distribution with m degrees of freedom. In Statistical Table 2, there are percentile values of the t-distribution for various degrees of freedom. For m degrees of freedom, the 95th percentile of the t-distribution is denoted $t_{(0.95,\, m)}$. This value has the property that 0.95 of the probability falls to its left, so $P\left[t_{(m)} \leq t_{(0.95,\, m)}\right] = 0.95$. For example, if the degrees of freedom are $m = 20$, then, from Statistical Table 2, $t_{(0.95,\, 20)} = 1.725$. Should you encounter a problem requiring percentiles

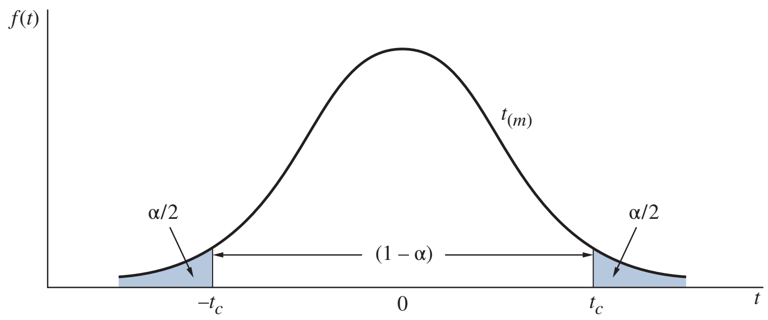

FIGURE 3.1 **Critical values from a *t*-distribution.**

that we do not give, you can interpolate for an approximate answer or use your computer software to obtain an exact value.

3.1.2 Obtaining Interval Estimates

From Statistical Table 2, we can find a "**critical value**" t_c from a t-distribution such that $P(t \geq t_c) = P(t \leq -t_c) = \alpha/2$, where α is a probability often taken to be $\alpha = 0.01$ or $\alpha = 0.05$. The critical value t_c for degrees of freedom m is the percentile value $t_{(1-\alpha/2,\, m)}$. The values t_c and $-t_c$ are depicted in Figure 3.1.

Each shaded "tail" area contains $\alpha/2$ of the probability, so that $1 - \alpha$ of the probability is contained in the center portion. Consequently, we can make the probability statement

$$P(-t_c \leq t \leq t_c) = 1 - \alpha \tag{3.4}$$

For a 95% confidence interval, the critical values define a central region of the t-distribution containing probability $1 - \alpha = 0.95$. This leaves probability $\alpha = 0.05$ divided equally between the two tails, so that $\alpha/2 = 0.025$. Then the critical value $t_c = t_{(1-0.025,\, m)} = t_{(0.975,\, m)}$. In the simple regression model, the degrees of freedom are $m = N - 2$, so expression (3.4) becomes

$$P\left[-t_{(0.975,\, N-2)} \leq t \leq t_{(0.975,\, N-2)}\right] = 0.95$$

We find the percentile values $t_{(0.975,\, N-2)}$ in Statistical Table 2.

Now, let us see how we can put all these bits together to create a procedure for interval estimation. Substitute t from (3.3) into (3.4) to obtain

$$P\left[-t_c \leq \frac{b_k - \beta_k}{\text{se}(b_k)} \leq t_c\right] = 1 - \alpha$$

Rearrange this expression to obtain

$$P\left[b_k - t_c\text{se}(b_k) \leq \beta_k \leq b_k + t_c\text{se}(b_k)\right] = 1 - \alpha \tag{3.5}$$

The interval endpoints $b_k - t_c\text{se}(b_k)$ and $b_k + t_c\text{se}(b_k)$ are random because they vary from sample to sample. These endpoints define an **interval estimator** of β_k. The probability statement in (3.5) says that the interval $b_k \pm t_c\text{se}(b_k)$ has probability $1 - \alpha$ of containing the true but unknown parameter β_k.

When b_k and $\text{se}(b_k)$ in (3.5) are estimated values (numbers), based on a given sample of data, then $b_k \pm t_c\text{se}(b_k)$ is called a $100(1 - \alpha)\%$ **interval estimate** of β_k. Equivalently, it is called a $100(1 - \alpha)\%$ **confidence interval**. Usually, $\alpha = 0.01$ or $\alpha = 0.05$, so that we obtain a 99% confidence interval or a 95% confidence interval.

The interpretation of confidence intervals requires a great deal of care. The properties of the interval estimation procedure are based on the notion of sampling. If we collect all possible samples of size N from a population, compute the least squares estimate b_k and its standard error $\text{se}(b_k)$ for each sample, and then construct the interval estimate $b_k \pm t_c\text{se}(b_k)$ for each sample, then $100(1-\alpha)\%$ of all the intervals constructed would contain the true parameter β_k. In Appendix 3C, we carry out a Monte Carlo simulation to demonstrate this sampling property.

Any *one* interval estimate, based on one sample of data, may or may not contain the true parameter β_k, and because β_k is unknown, we will never know whether it does or does not. When "confidence intervals" are discussed, remember that our confidence is in the *procedure* used to construct the interval estimate; it is *not* in any one interval estimate calculated from a sample of data.

EXAMPLE 3.1 | Interval Estimate for Food Expenditure Data

For the food expenditure data, $N = 40$ and the degrees of freedom are $N - 2 = 38$. For a 95% confidence interval, $\alpha = 0.05$. The critical value $t_c = t_{(1-\alpha/2,\, N-2)} = t_{(0.975,\, 38)} = 2.024$ is the 97.5 percentile from the t-distribution with 38 degrees of freedom. For β_2, the probability statement in (3.5) becomes

$$P\left[b_2 - 2.024\text{se}(b_2) \le \beta_2 \le b_2 + 2.024\text{se}(b_2)\right] = 0.95 \tag{3.6}$$

To construct an interval estimate for β_2, we use the least squares estimate $b_2 = 10.21$ and its standard error

$$\text{se}(b_2) = \sqrt{\widehat{\text{var}}(b_2)} = \sqrt{4.38} = 2.09$$

Substituting these values into (3.6), we obtain a "95% confidence interval estimate" for β_2:

$$b_2 \pm t_c\text{se}(b_2) = 10.21 \pm 2.024(2.09) = [5.97, 14.45]$$

That is, we estimate "with 95% confidence" that from an additional \$100 of weekly income households will spend between \$5.97 and \$14.45 on food.

Is β_2 actually in the interval [5.97, 14.45]? We do not know, and we will never know. What we *do* know is that when the procedure we used is applied to all possible samples of data from the same population, then 95% of all the interval estimates constructed using this procedure will contain the true parameter. The interval estimation procedure "works" 95% of the time. What we can say about the interval estimate based on our one sample is that, given the reliability of the procedure, we would be "surprised" if β_2 is not in the interval [5.97, 14.45].

What is the usefulness of an interval estimate of β_2? When reporting regression results, we always give a point estimate, such as $b_2 = 10.21$. However, the point estimate alone gives no sense of its reliability. Thus, we might also report an interval estimate. Interval estimates incorporate both the point estimate and the standard error of the estimate, which is a measure of the variability of the least squares estimator. The interval estimate includes an allowance for the sample size as well because for lower degrees of freedom the t-distribution critical value t_c is larger. If an interval estimate is wide (implying a large standard error), it suggests that there is not much information in the sample about β_2. If an interval estimate is narrow, it suggests that we have learned more about β_2.

What is "wide" and what is "narrow" depend on the problem at hand. For example, in our model, $b_2 = 10.21$ is an estimate of how much weekly household food expenditure will rise given a \$100 increase in weekly household income. A CEO of a supermarket chain can use this estimate to plan future store capacity requirements, given forecasts of income growth in an area. However, no decision will be based on this one number alone. The prudent CEO will carry out a sensitivity analysis by considering values of β_2 around 10.21. The question is "Which values?" One answer is provided by the interval estimate [5.97, 14.45]. Though β_2 may or may not be in this interval, the CEO knows that the procedure used to obtain the interval estimate "works" 95% of the time. If varying β_2 within the interval has drastic consequences on company sales and profits, then the CEO may conclude that there is insufficient evidence upon which to make a decision and order a new and larger data sample.

3.1.3 The Sampling Context

In Section 2.4.3, we illustrated the sampling properties of the least squares estimators using 10 data samples. Each sample of size $N = 40$ includes households with the same incomes as in Table 2.1 but with food expenditures that vary. These hypothetical data are in the data file *table2_2*. In Table 3.1, we present the OLS estimates, the estimates of σ^2, and the coefficient standard errors from each sample. Note the sampling variation illustrated by these estimates. The

TABLE 3.1 **Least Squares Estimates from 10 Hypothetical Random Samples**

Sample	b_1	$\text{se}(b_1)$	b_2	$\text{se}(b_2)$	$\hat{\sigma}^2$
1	93.64	31.73	8.24	1.53	4282.13
2	91.62	31.37	8.90	1.51	4184.79
3	126.76	48.08	6.59	2.32	9828.47
4	55.98	45.89	11.23	2.21	8953.17
5	87.26	42.57	9.14	2.05	7705.72
6	122.55	42.65	6.80	2.06	7735.38
7	91.95	42.14	9.84	2.03	7549.82
8	72.48	34.04	10.50	1.64	4928.44
9	90.34	36.69	8.75	1.77	5724.08
10	128.55	50.14	6.99	2.42	10691.61

TABLE 3.2 **Interval Estimates from 10 Hypothetical Random Samples**

Sample	$b_1 - t_c\text{se}(b_1)$	$b_1 + t_c\text{se}(b_1)$	$b_2 - t_c\text{se}(b_2)$	$b_2 + t_c\text{se}(b_2)$
1	29.40	157.89	5.14	11.34
2	28.12	155.13	5.84	11.96
3	29.44	224.09	1.90	11.29
4	−36.91	148.87	6.75	15.71
5	1.08	173.43	4.98	13.29
6	36.21	208.89	2.63	10.96
7	6.65	177.25	5.73	13.95
8	3.56	141.40	7.18	13.82
9	16.07	164.62	5.17	12.33
10	27.04	230.06	2.09	11.88

variation is due to the fact that in each sample household food expenditures are different. The 95% confidence intervals for the parameters β_1 and β_2 are given in Table 3.2 for the same samples.

Sampling variability causes the center of each of the interval estimates to change with the values of the least squares estimates, and it causes the widths of the intervals to change with the standard errors. If we ask the question "How many of these intervals contain the true parameters, and which ones are they?" we must answer that we do not know. But since 95% of all interval estimates constructed this way contain the true parameter values, we would expect perhaps 9 or 10 of these intervals to contain the true but unknown parameters.

Note the difference between point estimation and interval estimation. We have used the least squares estimators to obtain point estimates of unknown parameters. The estimated variance $\widehat{\text{var}}(b_k)$, for $k = 1$ or 2, and its square root $\sqrt{\widehat{\text{var}}(b_k)} = \text{se}(b_k)$ provide information about the sampling variability of the least squares estimator from one sample to another. Interval estimators are a convenient way to report regression results because they combine point estimation with a measure of sampling variability to provide a range of values in which the unknown parameters might fall. When the sampling variability of the least squares estimator is relatively small, then the interval estimates will be relatively narrow, implying that the least squares estimates are "reliable." If the least squares estimators suffer from large sampling variability, then the interval estimates will be wide, implying that the least squares estimates are "unreliable."

3.2 Hypothesis Tests

Many business and economic decision problems require a judgment as to whether or not a parameter is a specific value. In the food expenditure example, it may make a good deal of difference for decision purposes whether β_2 is greater than 10, indicating that a \$100 increase in income will increase expenditure on food by more than \$10. In addition, based on economic theory, we believe that β_2 should be positive. One check of our data and model is whether this theoretical proposition is supported by the data.

Hypothesis testing procedures compare a conjecture we have about a population to the information contained in a sample of data. Given an economic and statistical model, **hypotheses** are formed about economic behavior. These hypotheses are then represented as statements about model parameters. Hypothesis tests use the information about a parameter that is contained in a sample of data, its least squares point estimate, and its standard error to draw a conclusion about the hypothesis.

In each and every hypothesis test, five ingredients must be present:

Components of Hypothesis Tests

1. A null hypothesis H_0
2. An alternative hypothesis H_1
3. A test statistic
4. A rejection region
5. A conclusion

3.2.1 The Null Hypothesis

The **null hypothesis**, which is denoted by H_0 (*H-naught*), specifies a value for a regression parameter, which for generality we denote as β_k, for $k = 1$ or 2. The null hypothesis is stated as $H_0 : \beta_k = c$, where c is a constant, and is an important value in the context of a specific regression model. A null hypothesis is the belief we will maintain until we are convinced by the sample evidence that it is not true, in which case we *reject* the null hypothesis.

3.2.2 The Alternative Hypothesis

Paired with every null hypothesis is a logical **alternative hypothesis** H_1 that we will accept if the null hypothesis is rejected. The alternative hypothesis is flexible and depends, to some extent, on economic theory. For the null hypothesis $H_0 : \beta_k = c$, the three possible alternative hypotheses are as follows:

- $H_1 : \beta_k > c$. Rejecting the null hypothesis that $\beta_k = c$ leads us to accept the conclusion that $\beta_k > c$. Inequality alternative hypotheses are widely used in economics because economic theory frequently provides information about the *signs* of relationships between variables. For example, in the food expenditure example, we might well test the null hypothesis $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 > 0$ because economic theory strongly suggests that necessities such as food are normal goods and that food expenditure will rise if income increases.
- $H_1 : \beta_k < c$. Rejecting the null hypothesis that $\beta_k = c$ in this case leads us to accept the conclusion that $\beta_k < c$.
- $H_1 : \beta_k \neq c$. Rejecting the null hypothesis that $\beta_k = c$ in this case leads us to accept the conclusion that β_k takes a value either larger or smaller than c.

3.2.3 The Test Statistic

The sample information about the null hypothesis is embodied in the sample value of a **test statistic**. Based on the value of a test statistic, we decide either to reject the null hypothesis or not to reject it. A test statistic has a special characteristic: its probability distribution is completely *known* when the null hypothesis is true, and it has some *other* distribution if the null hypothesis is not true.

It all starts with the key result in (3.3), $t = \left(b_k - \beta_k\right)/\mathrm{se}\left(b_k\right) \sim t_{(N-2)}$. **If** the null hypothesis $H_0 : \beta_k = c$ is *true*, **then** we can substitute c for β_k and it follows that

$$t = \frac{b_k - c}{\mathrm{se}\left(b_k\right)} \sim t_{(N-2)} \tag{3.7}$$

If the null hypothesis is *not true*, then the t-statistic in (3.7) does *not* have a t-distribution with $N - 2$ degrees of freedom. This point is elaborated in Appendix 3B.

3.2.4 The Rejection Region

The **rejection region** depends on the form of the alternative. It is the range of values of the test statistic that leads to *rejection* of the null hypothesis. It is possible to construct a rejection region only if we have

- A test statistic whose distribution is known when the null hypothesis is true
- An alternative hypothesis
- A level of significance

The rejection region consists of values that are *unlikely* and that have low probability of occurring when the null hypothesis is true. The chain of logic is "If a value of the test statistic is obtained that falls in a region of low probability, then it is unlikely that the test statistic has the assumed distribution, and thus, it is unlikely that the null hypothesis is true." If the alternative hypothesis is true, then values of the test statistic will tend to be unusually large or unusually small. The terms "large" and "small" are determined by choosing a probability α, called the **level of significance** of the test, which provides a meaning for "an *unlikely* event." The level of significance of the test α is usually chosen to be 0.01, 0.05, or 0.10.

Remark

When no other specific choice is made, economists and statisticians often use a significance level of 0.05. That is, an occurrence "one time in twenty" is regarded as an unusual or improbable event by chance. This threshold for statistical significance is clung to as the Holy Grail but in reality is simply a historical precedent based on quotes by Sir Ronald Fisher who promoted the standard that t-values larger than two be regarded as significant.[1] A stronger threshold for significance, such as "one time in a hundred," or 0.01, might make more sense. The importance of the topic is quickly evident with a web search. The issues are discussed in *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, by Stephen T. Ziliak and Deirdre N. McCloskey, 2008, The University of Michigan Press.

If we reject the null hypothesis when it is true, then we commit what is called a **Type I error**. The level of significance of a test *is* the probability of committing a Type I error, so

[1]Mark Kelly (2013) "Emily Dickinson and monkeys on the stair. Or: What is the significance of the 5% significance level," *Significance*, Vol. 10(5), October, 21–22.

$P(\text{Type I error}) = \alpha$. Any time we reject a null hypothesis, it is possible that we have made such an error—there is no avoiding it. The good news is that we can specify the amount of Type I error we will tolerate by setting the level of significance α. If such an error is costly, then we make α small. If we do not reject a null hypothesis that is false, then we have committed a **Type II error**. In a real-world situation, we cannot control or calculate the probability of this type of error because it depends on the unknown true parameter β_k. For more about Type I and Type II errors, see Appendix C.6.9.

3.2.5 A Conclusion

When you have completed testing a hypothesis, you should state your conclusion. Do you reject the null hypothesis, or do you not reject the null hypothesis? As we will argue below, you should avoid saying that you "accept" the null hypothesis, which can be very misleading. Moreover, we urge you to make it standard practice to say what the conclusion means in the economic context of the problem you are working on and the economic significance of the finding. Statistical procedures are not ends in themselves. They are carried out for a reason and have meaning, which you should be able to explain.

3.3 Rejection Regions for Specific Alternatives

In this section, we hope to be very clear about the nature of the rejection rules for each of the three possible alternatives to the null hypothesis $H_0 : \beta_k = c$. As noted in the previous section, to have a rejection region for a null hypothesis, we need a test statistic, which we have; it is given in (3.7). Second, we need a specific alternative, $\beta_k > c$, $\beta_k < c$, or $\beta_k \neq c$. Third, we need to specify the level of significance of the test. The level of significance of a test, α, is the probability that we reject the null hypothesis when it is actually true, which is called a Type I error.

3.3.1 One-Tail Tests with Alternative "Greater Than" (>)

When testing the null hypothesis $H_0 : \beta_k = c$, if the *alternative* hypothesis $H_1 : \beta_k > c$ is true, then the value of the t-statistic (3.7) tends to become larger than usual for the t-distribution. We will reject the null hypothesis if the test statistic is larger than the critical value for the level of significance α. The critical value that leaves probability α in the right tail is the $(1 - \alpha)$-percentile $t_{(1-\alpha, N-2)}$, as shown in Figure 3.2. For example, if $\alpha = 0.05$ and $N - 2 = 30$, then from Statistical Table 2, the critical value is the 95th percentile value $t_{(0.95, 30)} = 1.697$.

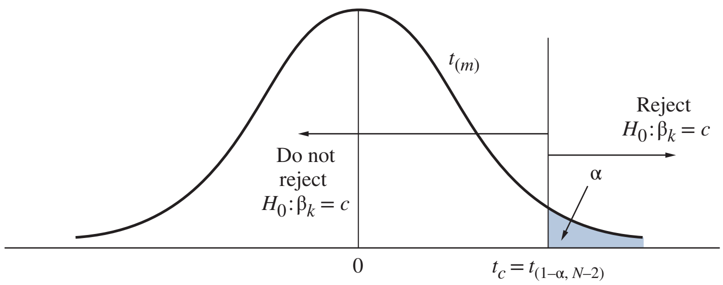

FIGURE 3.2 **Rejection region for a one-tail test of $H_0 : \beta_k = c$ against $H_1 : \beta_k > c$.**

The rejection rule is

> When testing the null hypothesis $H_0: \beta_k = c$ against the alternative hypothesis $H_1: \beta_k > c$, reject the null hypothesis and accept the alternative hypothesis if $t \geq t_{(1-\alpha, N-2)}$.

The test is called a "one-tail" test because unlikely values of the t-statistic fall only in one tail of the probability distribution. If the null hypothesis is true, then the test statistic (3.7) has a t-distribution, and its value would tend to fall in the center of the distribution, to the left of the critical value, where most of the probability is contained. The level of significance α is chosen so that if the null hypothesis is true, then the probability that the t-statistic value falls in the extreme right tail of the distribution is small; an event that is improbable and unlikely to occur by chance. If we obtain a test statistic value in the rejection region, we take it as evidence *against* the null hypothesis, leading us to conclude that the null hypothesis is unlikely to be true. Evidence against the null hypothesis is evidence in support of the alternative hypothesis. Thus, if we reject the null hypothesis then we conclude that the alternative is true.

If the null hypothesis $H_0: \beta_k = c$ is *true,* then the test statistic (3.7) has a t-distribution and its values fall in the nonrejection region with probability $1 - \alpha$. If $t < t_{(1-\alpha, N-2)}$, then there is no statistically significant evidence against the null hypothesis, and we do not reject it.

3.3.2 One-Tail Tests with Alternative "Less Than" (<)

If the alternative hypothesis $H_1: \beta_k < c$ is true, then the value of the t-statistic (3.7) tends to become smaller than usual for the t-distribution. We reject the null hypothesis if the test statistic is smaller than the critical value for the level of significance α. The critical value that leaves probability α in the left tail is the α-percentile $t_{(\alpha, N-2)}$, as shown in Figure 3.3.

When using Statistical Table 2 to locate critical values, recall that the t-distribution is symmetric about zero, so that the α-percentile $t_{(\alpha, N-2)}$ is the negative of the $(1 - \alpha)$-percentile $t_{(1-\alpha, N-2)}$. For example, if $\alpha = 0.05$ and $N - 2 = 20$, then from Statistical Table 2, the 95th percentile of the t-distribution is $t_{(0.95, 20)} = 1.725$ and the 5th percentile value is $t_{(0.05, 20)} = -1.725$.

The rejection rule is:

> When testing the null hypothesis $H_0: \beta_k = c$ against the alternative hypothesis $H_1: \beta_k < c$, reject the null hypothesis and accept the alternative hypothesis if $t \leq t_{(\alpha, N-2)}$.

The nonrejection region consists of t-statistic values greater than $t_{(\alpha, N-2)}$. When the null hypothesis is true, the probability of obtaining such a t-value is $1 - \alpha$, which is chosen to be large. Thus if $t > t_{(\alpha, N-2)}$ then do not reject $H_0: \beta_k = c$.

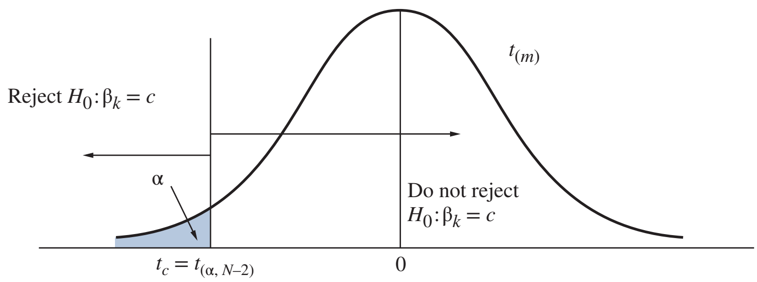

FIGURE 3.3 **The rejection region for a one-tail test of $H_0: \beta_k = c$ against $H_1: \beta_k < c$.**

Remembering where the rejection region is located may be facilitated by the following trick:

Memory Trick

The rejection region for a one-tail test is in the direction of the arrow in the alternative. If the alternative is >, then reject in the right tail. If the alternative is <, reject in the left tail.

3.3.3 Two-Tail Tests with Alternative "Not Equal To" (≠)

When testing the null hypothesis $H_0\!:\beta_k = c$, if the alternative hypothesis $H_1\!:\beta_k \neq c$ is true, then the value of the t-statistic (3.7) tends to become either larger *or* smaller than usual for the t-distribution. To have a test with the level of significance α, we define the critical values so that the probability of the t-statistic falling in either tail is $\alpha/2$. The left-tail critical value is the percentile $t_{(\alpha/2,\,N-2)}$ and the right-tail critical value is the percentile $t_{(1-\alpha/2,\,N-2)}$. We reject the null hypothesis that $H_0\!:\beta_k = c$ in favor of the alternative that $H_1\!:\beta_k \neq c$ if the test statistic $t \le t_{(\alpha/2,\,N-2)}$ or $t \ge t_{(1-\alpha/2,\,N-2)}$, as shown in Figure 3.4. For example, if $\alpha = 0.05$ and $N - 2 = 30$, then $\alpha/2 = 0.025$ and the left-tail critical value is the 2.5-percentile value $t_{(0.025,\,30)} = -2.042$; the right-tail critical value is the 97.5-percentile $t_{(0.975,\,30)} = 2.042$. The right-tail critical value is found in Statistical Table 2, and the left-tail critical value is found using the symmetry of the t-distribution.

Since the rejection region is composed of portions of the t-distribution in the left and right tails, this test is called a **two-tail test**. When the null hypothesis is true, the probability of obtaining a value of the test statistic that falls in *either* tail area is "small." The sum of the tail probabilities is α. Sample values of the test statistic that are in the tail areas are incompatible with the null hypothesis and are evidence against the null hypothesis being true. On the other hand, if the null hypothesis $H_0\!:\beta_k = c$ is true, then the probability of obtaining a value of the test statistic t in the central nonrejection region is high. Sample values of the test statistic in the central nonrejection area are compatible with the null hypothesis and are not taken as evidence against the null hypothesis being true. Thus, the rejection rule is

When testing the null hypothesis $H_0\!:\beta_k = c$ against the alternative hypothesis $H_1\!:\beta_k \neq c$, reject the null hypothesis and accept the alternative hypothesis if $t \le t_{(\alpha/2,\,N-2)}$ **or** if $t \ge t_{(1-\alpha/2,\,N-2)}$.

We do not reject the null hypothesis if $t_{(\alpha/2,\,N-2)} < t < t_{(1-\alpha/2,\,N-2)}$.

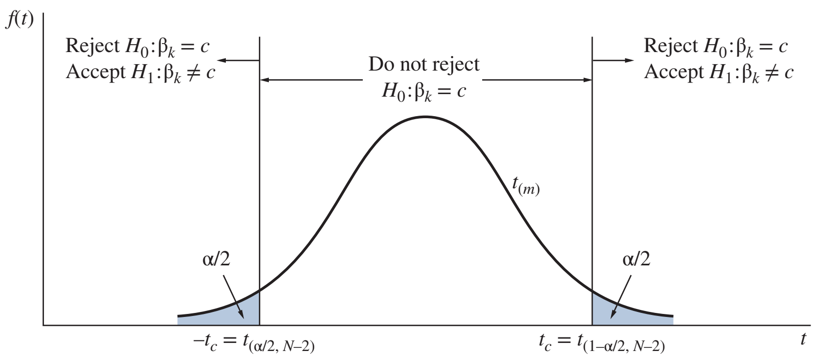

FIGURE 3.4 **Rejection region for a test of $H_0\!:\beta_k = c$ against $H_1\!:\beta_k \neq c$.**

3.4 Examples of Hypothesis Tests

We illustrate the mechanics of hypothesis testing using the food expenditure model. We give examples of right-tail, left-tail, and two-tail tests. In each case, we will follow a prescribed set of steps, closely following the list of required components for all hypothesis tests listed at the beginning of Section 3.2. A standard procedure for all hypothesis-testing problems and situations is

Step-by-Step Procedure for Testing Hypotheses

1. Determine the null and alternative hypotheses.
2. Specify the test statistic and its distribution if the null hypothesis is true.
3. Select α and determine the rejection region.
4. Calculate the sample value of the test statistic.
5. State your conclusion.

EXAMPLE 3.2 | Right-Tail Test of Significance

Usually, our first concern is whether there is a relationship between the variables, as we have specified in our model. If $\beta_2 = 0$, then there is no linear relationship between food expenditure and income. Economic theory suggests that food is a normal good and that as income increases food expenditure will also increase and thus that $\beta_2 > 0$. The least squares estimate of β_2 is $b_2 = 10.21$, which is certainly greater than zero. However, simply observing that the estimate has the correct sign does not constitute scientific proof. We want to determine whether there is convincing, or *significant*, statistical evidence that would lead us to conclude that $\beta_2 > 0$. When testing the null hypothesis that a parameter is zero, we are asking if the estimate b_2 is significantly different from zero, and the test is called a **test of significance**.

A statistical test procedure cannot prove the truth of a null hypothesis. When we fail to reject a null hypothesis, all the hypothesis test can establish is that the information in a sample of data is *compatible* with the null hypothesis. Conversely, a statistical test can lead us to *reject* the null hypothesis, with only a small probability α of rejecting the null hypothesis when it is actually true. Thus, rejecting a null hypothesis is a stronger conclusion than failing to reject it. For this reason, the null hypothesis is usually stated in such a way that if our theory is correct, then we will reject the null hypothesis. In our example, economic theory implies that there should be a positive relationship between income and food expenditure. We would like to establish that there is statistical evidence to support this theory using a hypothesis test. With this goal, we set up the null hypothesis that there is *no* relation between the variables, $H_0 : \beta_2 = 0$. In the alternative hypothesis, we put the conjecture that we would like to establish, $H_1 : \beta_2 > 0$. If we then reject the null hypothesis, we can make a direct statement, concluding that β_2 is positive, with only a small (α) probability that we are in error.

The steps of this hypothesis test are as follows:

1. The null hypothesis is $H_0 : \beta_2 = 0$. The alternative hypothesis is $H_1 : \beta_2 > 0$.
2. The test statistic is (3.7). In this case, $c = 0$, so $t = b_2/\text{se}(b_2) \sim t_{(N-2)}$ if the null hypothesis is true.
3. Let us select $\alpha = 0.05$. The critical value for the right-tail rejection region is the 95th percentile of the t-distribution with $N - 2 = 38$ degrees of freedom, $t_{(0.95,\, 38)} = 1.686$. Thus, we will reject the null hypothesis if the calculated value of $t \geq 1.686$. If $t < 1.686$, we will not reject the null hypothesis.
4. Using the food expenditure data, we found that $b_2 = 10.21$ with standard error $\text{se}(b_2) = 2.09$. The value of the test statistic is
$$t = \frac{b_2}{\text{se}(b_2)} = \frac{10.21}{2.09} = 4.88$$
5. Since $t = 4.88 > 1.686$, we reject the null hypothesis that $\beta_2 = 0$ and accept the alternative that $\beta_2 > 0$. That is, we reject the hypothesis that there is no relationship between income and food expenditure and conclude that there is a *statistically significant* positive relationship between household income and food expenditure.

The last part of the conclusion is important. When you report your results to an audience, you will want to describe the outcome of the test in the context of the problem you are investigating, not just in terms of Greek letters and symbols.

What if we had not been able to reject the null hypothesis in this example? Would we have concluded that economic theory is wrong and that there is no relationship between income and food expenditure? No. Remember that failing to reject a null hypothesis **does not** mean that the null hypothesis is true.

EXAMPLE 3.3 | Right-Tail Test of an Economic Hypothesis

Suppose that the economic profitability of a new supermarket depends on households spending more than \$5.50 out of each additional \$100 weekly income on food and that construction will not proceed unless there is strong evidence to this effect. In this case, the conjecture we want to establish, the one that will go in the alternative hypothesis, is that $\beta_2 > 5.5$. If $\beta_2 \leq 5.5$, then the supermarket will be unprofitable and the owners would not want to build it. The least squares estimate of β_2 is $b_2 = 10.21$, which is greater than 5.5. What we want to determine is whether there is convincing statistical evidence that would lead us to conclude, based on the available data, that $\beta_2 > 5.5$. This judgment is based on not only the estimate b_2 but also its precision as measured by $\text{se}(b_2)$.

What will the null hypothesis be? We have been stating null hypotheses as equalities, such as $\beta_2 = 5.5$. This null hypothesis is too limited because it is theoretically possible that $\beta_2 < 5.5$. It turns out that the hypothesis testing procedure for testing the null hypothesis that $H_0 : \beta_2 \leq 5.5$ against the alternative hypothesis $H_1 : \beta_2 > 5.5$ is *exactly the same* as testing $H_0 : \beta_2 = 5.5$ against the alternative hypothesis $H_1 : \beta_2 > 5.5$. The test statistic and rejection region are exactly the same. For a right-tail test, you can form the null hypothesis in either of these ways depending on the problem at hand.

The steps of this hypothesis test are as follows:

1. The null hypothesis is $H_0 : \beta_2 \leq 5.5$. The alternative hypothesis is $H_1 : \beta_2 > 5.5$.
2. The test statistic $t = (b_2 - 5.5)/\text{se}(b_2) \sim t_{(N-2)}$ if the null hypothesis is true.
3. Let us select $\alpha = 0.01$. The critical value for the right-tail rejection region is the 99th percentile of the t-distribution with $N - 2 = 38$ degrees of freedom, $t_{(0.99, 38)} = 2.429$. We will reject the null hypothesis if the calculated value of $t \geq 2.429$. If $t < 2.429$, we will not reject the null hypothesis.
4. Using the food expenditure data, $b_2 = 10.21$ with standard error $\text{se}(b_2) = 2.09$. The value of the test statistic is
$$t = \frac{b_2 - 5.5}{\text{se}(b_2)} = \frac{10.21 - 5.5}{2.09} = 2.25$$
5. Since $t = 2.25 < 2.429$, we do not reject the null hypothesis that $\beta_2 \leq 5.5$. We are *not* able to conclude that the new supermarket will be profitable and will not begin construction.

In this example, we have posed a situation where the choice of the level of significance α becomes of great importance. A construction project worth millions of dollars depends on having *convincing* evidence that households will spend more than \$5.50 out of each additional \$100 income on food. Although the "usual" choice is $\alpha = 0.05$, we have chosen a conservative value of $\alpha = 0.01$ because we seek a test that has a low chance of rejecting the null hypothesis when it is actually true. Recall that the level of significance of a test defines what we mean by an unlikely value of the test statistic. In this example, if the null hypothesis is true, then building the supermarket will be unprofitable. We want the probability of building an unprofitable market to be very small, and therefore, we want the probability of rejecting the null hypothesis when it is true to be very small. In each real-world situation, the choice of α must be made on an assessment of *risk* and the *consequences* of making an incorrect decision.

A CEO unwilling to make a decision based on the available evidence may well order a new and larger sample of data to be analyzed. Recall that as the sample size increases, the least squares estimator becomes more precise (as measured by estimator variance), and consequently, hypothesis tests become more powerful tools for statistical inference.

EXAMPLE 3.4 | Left-Tail Test of an Economic Hypothesis

For completeness, we will illustrate a test with the rejection region in the left tail. Consider the null hypothesis that $\beta_2 \geq 15$ and the alternative hypothesis $\beta_2 < 15$. Recall our memory trick for determining the location of the rejection region for a t-test. The rejection region is in the direction of the arrow $<$ in the alternative hypothesis. This fact tells us that the rejection region is in the left tail of the t-distribution. The steps of this hypothesis test are as follows:

1. The null hypothesis is $H_0: \beta_2 \geq 15$. The alternative hypothesis is $H_1: \beta_2 < 15$.
2. The test statistic $t = (b_2 - 15)/\text{se}(b_2) \sim t_{(N-2)}$ if the null hypothesis is true.
3. Let us select $\alpha = 0.05$. The critical value for the left-tail rejection region is the 5th percentile of the t-distribution with $N - 2 = 38$ degrees of freedom, $t_{(0.05, 38)} = -1.686$. We will reject the null hypothesis if the calculated value of $t \leq -1.686$. If $t > -1.686$, we will not reject the null hypothesis. A left-tail rejection region is illustrated in Figure 3.3.
4. Using the food expenditure data, $b_2 = 10.21$ with standard error $\text{se}(b_2) = 2.09$. The value of the test statistic is
$$t = \frac{b_2 - 15}{\text{se}(b_2)} = \frac{10.21 - 15}{2.09} = -2.29$$
5. Since $t = -2.29 < -1.686$, we reject the null hypothesis that $\beta_2 \geq 15$ and accept the alternative that $\beta_2 < 15$. We conclude that households spend less than \$15 from each additional \$100 income on food.

EXAMPLE 3.5 | Two-Tail Test of an Economic Hypothesis

A consultant voices the opinion that based on other similar neighborhoods the households near the proposed market will spend an additional \$7.50 per additional \$100 income. In terms of our economic model, we can state this conjecture as the null hypothesis $\beta_2 = 7.5$. If we want to test whether this is true or not, then the alternative is that $\beta_2 \neq 7.5$. This alternative makes no claim about whether β_2 is greater than 7.5 or less than 7.5, simply that it is not 7.5. In such cases, we use a two-tail test, as follows:

1. The null hypothesis is $H_0: \beta_2 = 7.5$. The alternative hypothesis is $H_1: \beta_2 \neq 7.5$.
2. The test statistic $t = (b_2 - 7.5)/\text{se}(b_2) \sim t_{(N-2)}$ if the null hypothesis is true.
3. Let us select $\alpha = 0.05$. The critical values for this two-tail test are the 2.5-percentile $t_{(0.025, 38)} = -2.024$ and the 97.5-percentile $t_{(0.975, 38)} = 2.024$. Thus, we will reject the null hypothesis if the calculated value of $t \geq 2.024$ **or** if $t \leq -2.024$. If $-2.024 < t < 2.024$, then we will not reject the null hypothesis.
4. For the food expenditure data, $b_2 = 10.21$ with standard error $\text{se}(b_2) = 2.09$. The value of the test statistic is
$$t = \frac{b_2 - 7.5}{\text{se}(b_2)} = \frac{10.21 - 7.5}{2.09} = 1.29$$
5. Since $-2.204 < t = 1.29 < 2.204$, we do not reject the null hypothesis that $\beta_2 = 7.5$. The sample data are consistent with the conjecture households will spend an additional \$7.50 per additional \$100 income on food.

We must avoid reading into this conclusion more than it means. We **do not** conclude from this test that $\beta_2 = 7.5$, only that the data are not incompatible with this parameter value. The data are also compatible with the null hypotheses $H_0: \beta_2 = 8.5$ ($t = 0.82$), $H_0: \beta_2 = 6.5$ ($t = 1.77$), and $H_0: \beta_2 = 12.5$ ($t = -1.09$). A hypothesis test **cannot** be used to prove that a null hypothesis is true.

There is a trick relating two-tail tests and confidence intervals that is sometimes useful. Let q be a value within a $100(1 - \alpha)\%$ confidence interval, so that if $t_c = t_{(1-\alpha/2, N-2)}$, then
$$b_k - t_c\text{se}(b_k) \leq q \leq b_k + t_c\text{se}(b_k)$$
If we test the null hypothesis $H_0: \beta_k = q$ against $H_1: \beta_k \neq q$, when q is inside the confidence interval, then we will *not* reject the null hypothesis at the level of significance α. If q is outside the confidence interval, then the two-tail test will reject the null hypothesis. We do not advocate using confidence intervals to test hypotheses, they serve a different purpose, but if you are given a confidence interval, this trick is handy.

EXAMPLE 3.6 | Two-Tail Test of Significance

While we are confident that a relationship exists between food expenditure and income, models are often proposed that are more speculative, and the purpose of hypothesis testing is to ascertain whether a relationship between variables exists or not. In this case, the null hypothesis is $\beta_2 = 0$; that is, no linear relationship exists between x and y. The alternative is $\beta_2 \neq 0$, which would mean that a relationship exists but that there may be either a positive or negative association between the variables. This is the most common form of a **test of significance**. The test steps are as follows:

1. The null hypothesis is $H_0: \beta_2 = 0$. The alternative hypothesis is $H_1: \beta_2 \neq 0$.
2. The test statistic $t = b_2/\text{se}(b_2) \sim t_{(N-2)}$ *if the null hypothesis is true*.
3. Let us select $\alpha = 0.05$. The critical values for this two-tail test are the 2.5-percentile $t_{(0.025, 38)} = -2.024$ and the 97.5-percentile $t_{(0.975, 38)} = 2.024$. We will reject the null hypothesis if the calculated value of $t \geq 2.024$ *or* if $t \leq -2.024$. If $-2.024 < t < 2.024$, we will not reject the null hypothesis.
4. Using the food expenditure data, $b_2 = 10.21$ with standard error $\text{se}(b_2) = 2.09$. The value of the test statistic is $t = b_2/\text{se}(b_2) = 10.21/2.09 = 4.88$.
5. Since $t = 4.88 > 2.024$, we reject the null hypothesis that $\beta_2 = 0$ and conclude that there is a statistically significant relationship between income and food expenditure.

Two points should be made about this result. First, the value of the t-statistic we computed in this two-tail test is the same as the value computed in the one-tail test of significance in Example 3.2. The difference between the two tests is the rejection region and the critical values. Second, the two-tail test of significance is something that should be done each time a regression model is estimated, and consequently, computer software automatically calculates the t-values for null hypotheses that the regression parameters are zero. Refer back to Figure 2.9. Consider the portion that reports the estimates:

Variable	Coefficient	Standard Error	*t*-Statistic	Prob.
C	83.41600	43.41016	1.921578	0.0622
INCOME	10.20964	2.093264	4.877381	0.0000

Note that there is a column-labeled t-statistic. This is the t-statistic value for the null hypothesis that the corresponding parameter is zero. It is calculated as $t = b_k/\text{se}(b_k)$. Dividing the least squares estimates (Coefficient) by their standard errors (Std. error) gives the t-statistic values (t-statistic) for testing the hypothesis that the parameter is zero. The t-statistic value for the variable *INCOME* is 4.877381, which is relevant for testing the null hypothesis $H_0: \beta_2 = 0$. We have rounded this value to 4.88 in our discussions.

The t-value for testing the hypothesis that the intercept is zero equals 1.92. The $\alpha = 0.05$ critical values for these two-tail tests are $t_{(0.025, 38)} = -2.024$ and $t_{(0.975, 38)} = 2.024$ whether we are testing a hypothesis about the slope or intercept, so we fail to reject the null hypothesis that $H_0: \beta_1 = 0$ given the alternative $H_1: \beta_1 \neq 0$.

The final column, labeled "Prob.," is the subject of the following section.

Remark

"Statistically significant" does not necessarily imply "economically significant." For example, suppose that the CEO of a supermarket chain plans a certain course of action if $\beta_2 \neq 0$. Furthermore, suppose that a large sample is collected from which we obtain the estimate $b_2 = 0.0001$ with $\text{se}(b_2) = 0.00001$, yielding the t-statistic $t = 10.0$. We would reject the null hypothesis that $\beta_2 = 0$ and accept the alternative that $\beta_2 \neq 0$. Here, $b_2 = 0.0001$ is statistically different from zero. However, 0.0001 may not be "economically" different from zero, and the CEO may decide not to proceed with the plans. The message here is that one must think carefully about the importance of a statistical analysis before reporting or using the results.

3.5 The *p*-Value

When reporting the outcome of statistical hypothesis tests, it has become standard practice to report the ***p*-value** (an abbreviation for **probability value**) of the test. If we have the p-value of a test, p, we can determine the outcome of the test by comparing the p-value to the chosen level of significance, α, *without* looking up or calculating the critical values. The rule is

p-Value Rule

Reject the null hypothesis when the p-value is less than, or equal to, the level of significance α. That is, if $p \leq \alpha$, then reject H_0. If $p > \alpha$, then do not reject H_0.

If you have chosen the level of significance to be $\alpha = 0.01, 0.05, 0.10$, or any other value, you can compare it to the p-value of a test and then reject, or not reject, without checking the critical value. In written works, reporting the p-value of a test allows the reader to apply his or her own judgment about the appropriate level of significance.

How the p-value is computed depends on the alternative. If t is the calculated value of the t-statistic, then

- if $H_1 : \beta_k > c$, p = probability to the right of t
- if $H_1 : \beta_k < c$, p = probability to the left of t
- if $H_1 : \beta_k \neq c$, p = *sum* of probabilities to the right of $|t|$ *and* to the left of $-|t|$

Memory Trick

The direction of the alternative indicates the tail(s) of the distribution in which the p-value falls.

EXAMPLE 3.3 (continued) | p-Value for a Right-Tail Test

In Example 3.3, we tested the null hypothesis $H_0 : \beta_2 \leq 5.5$ against the one-sided alternative $H_1 : \beta_2 > 5.5$. The calculated value of the t-statistic was

$$t = \frac{b_2 - 5.5}{\operatorname{se}(b_2)} = \frac{10.21 - 5.5}{2.09} = 2.25$$

In this case, since the alternative is "greater than" (>), the p-value of this test is the probability that a t-random variable with $N - 2 = 38$ degrees of freedom is greater than 2.25, or $p = P\left[t_{(38)} \geq 2.25\right] = 0.0152$.

This probability value cannot be found in the usual t-table of critical values, but it is easily found using the computer. Statistical software packages, and spreadsheets such as Excel, have simple commands to evaluate the *cumulative distribution function* (*cdf*) (see Appendix B.1) for a variety of probability distributions. If $F_X(x)$ is the *cdf* for a random variable X, then for any value $x = c$, the cumulative probability is $P[X \leq c] = F_X(c)$. Given such a function for the t-distribution, we compute the desired p-value as

$$p = P\left[t_{(38)} \geq 2.25\right] = 1 - P\left[t_{(38)} \leq 2.25\right] = 1 - 0.9848$$
$$= 0.0152$$

Following the p-value rule, we conclude that at $\alpha = 0.01$ we do not reject the null hypothesis. If we had chosen $\alpha = 0.05$, we would reject the null hypothesis in favor of the alternative.

The logic of the p-value rule is shown in Figure 3.5. The probability of obtaining a t-value greater than 2.25 is 0.0152, $p = P\left[t_{(38)} \geq 2.25\right] = 0.0152$. The 99th percentile $t_{(0.99, 38)}$, which is the critical value for a right-tail test with the level of significance of $\alpha = 0.01$ must fall to the right of 2.25. This means that $t = 2.25$ does not fall in the rejection region if $\alpha = 0.01$ and we will not reject the null hypothesis at this level of significance. This is consistent with the *p-value rule*: When the p-value (0.0152) is greater than the chosen level of significance (0.01), we do not reject the null hypothesis.

On the other hand, the 95th percentile $t_{(0.95, 38)}$, which is the critical value for a right-tail test with $\alpha = 0.05$, must be to the left of 2.25. This means that $t = 2.25$ falls in the rejection region, and we reject the null hypothesis at the level of significance $\alpha = 0.05$. This is consistent with the *p-value rule*: When the p-value (0.0152) is less than or equal to the chosen level of significance (0.05), we will reject the null hypothesis.

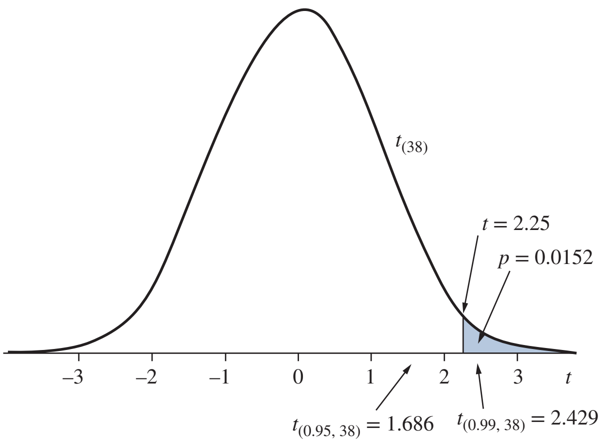

FIGURE 3.5 **The *p*-value for a right-tail test.**

EXAMPLE 3.4 (continued) | *p*-Value for a Left-Tail Test

In Example 3.4, we carried out a test with the rejection region in the left tail of the t-distribution. The null hypothesis was $H_0\!:\beta_2 \geq 15$, and the alternative hypothesis was $H_1\!:\beta_2 < 15$. The calculated value of the t-statistic was $t = -2.29$. To compute the p-value for this left-tail test, we calculate the probability of obtaining a t-statistic to the left of -2.29. Using your computer software, you will find this value to be $P\left[t_{(38)} \leq -2.29\right] = 0.0139$. Following the p-value rule, we conclude that at $\alpha = 0.01$, we do not reject the null hypothesis. If we choose $\alpha = 0.05$, we will reject the null hypothesis in favor of the alternative. See Figure 3.6 to see this graphically. Locate the 1st and 5th percentiles. These will be the critical values for left-tail tests with $\alpha = 0.01$ and $\alpha = 0.05$ levels of significance. When the p-value (0.0139) is greater than the level of significance ($\alpha = 0.01$), then the t-value -2.29 is not in the test rejection region. When the p-value (0.0139) is less than or equal to the level of significance ($\alpha = 0.05$), then the t-value -2.29 is in the test rejection region.

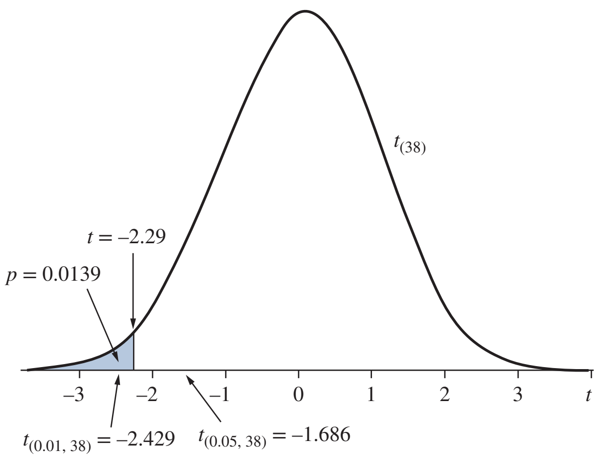

FIGURE 3.6 **The *p*-value for a left-tail test.**

EXAMPLE 3.5 (continued) | *p*-Value for a Two-Tail Test

For a two-tail test, the rejection region is in the two tails of the t-distribution, and the p-value is similarly calculated in the two tails of the distribution. In Example 3.5, we tested the null hypothesis that $\beta_2 = 7.5$ against the alternative hypothesis $\beta_2 \neq 7.5$. The calculated value of the t-statistic was $t = 1.29$. For this two-tail test, the p-value is the combined probability to the right of 1.29 and to the left of -1.29:

$$p = P\left[t_{(38)} \geq 1.29\right] + P\left[t_{(38)} \leq -1.29\right] = 0.2033$$

This calculation is depicted in Figure 3.7. Once the p-value is obtained, its use is unchanged. If we choose $\alpha = 0.05$, $\alpha = 0.10$, or even $\alpha = 0.20$, we will fail to reject the null hypothesis because $p > \alpha$.

At the beginning of this section, we stated the following rule for computing p-values for two-tail tests: if $H_1: \beta_k \neq c$, p = *sum* of probabilities to the right of $|t|$ *and* to the left of $-|t|$. The reason for the use of absolute values in this rule is that it will apply equally well if the value of the t-statistic turns out to be positive or negative.

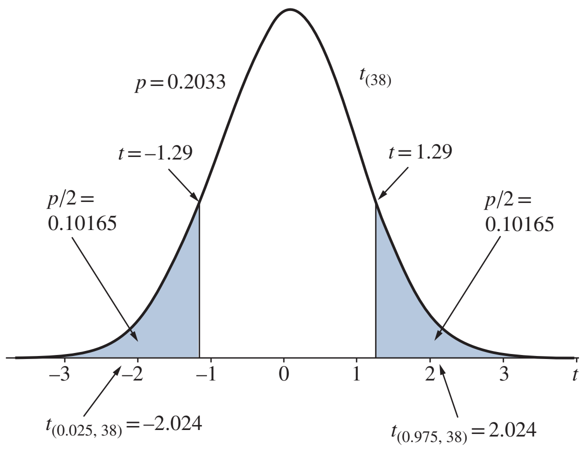

FIGURE 3.7 **The *p*-value for a two-tail test of significance.**

EXAMPLE 3.6 (continued) | *p*-Value for a Two-Tail Test of Significance

All statistical software computes the p-value for the two-tail test of significance for each coefficient when a regression analysis is performed. In Example 3.6, we discussed testing the null hypothesis $H_0: \beta_2 = 0$ against the alternative hypothesis $H_1: \beta_2 \neq 0$. For the calculated value of the t-statistic $t = 4.88$, the p-value is

$$p = P\left[t_{(38)} \geq 4.88\right] + P\left[t_{(38)} \leq -4.88\right] = 0.0000$$

Your software will automatically compute and report this p-value for a two-tail test of significance. Refer back to Figure 2.9 and consider just the portion reporting the estimates:

Variable	Coefficient	Standard Error	*t*-Statistic	Prob.
C	83.41600	43.41016	1.921578	0.0622
INCOME	10.20964	2.093264	4.877381	0.0000

Next to each t-statistic value is the two-tail p-value, which is labeled "Prob." by the EViews software. Other software packages will use similar names. When inspecting computer output, we can immediately decide if an estimate is statistically significant (statistically different from zero using a two-tail test) by comparing the p-value to whatever level of significance we care to use. The estimated intercept has p-value 0.0622, so it is not statistically different from zero at the level of significance $\alpha = 0.05$, but it is statistically significant if $\alpha = 0.10$.

The estimated coefficient for income has a p-value that is zero to four places. Thus, $p \leq \alpha = 0.01$ or even $\alpha = 0.0001$, and thus, we reject the null hypothesis that income has no effect on food expenditure at these levels of significance. The p-value for this two-tail test of significance is not actually zero. If more places are used, then $p = 0.00001946$. Regression software usually does not print out more than four places because in practice levels of significance less than $\alpha = 0.001$ are rare.

3.6 Linear Combinations of Parameters

So far, we have discussed statistical inference (point estimation, interval estimation, and hypothesis testing) for a single parameter, β_1 or β_2. More generally, we may wish to estimate and test hypotheses about a **linear combination of parameters** $\lambda = c_1\beta_1 + c_2\beta_2$, where c_1 and c_2 are constants that we specify. One example is if we wish to estimate the expected value of a dependent variable $E(y|x)$ when x takes some specific value, such as $x = x_0$. In this case, $c_1 = 1$ and $c_2 = x_0$, so that, $\lambda = c_1\beta_1 + c_2\beta_2 = \beta_1 + x_0\beta_2 = E(y|x = x_0)$.

Under assumptions SR1–SR5, the least squares estimators b_1 and b_2 are the best linear unbiased estimators of β_1 and β_2. It is also true that $\hat{\lambda} = c_1b_1 + c_2b_2$ is the best linear unbiased estimator of $\lambda = c_1\beta_1 + c_2\beta_2$. The estimator $\hat{\lambda}$ is unbiased because

$$E\left(\hat{\lambda}|\mathbf{x}\right) = E\left(c_1b_1 + c_2b_2|\mathbf{x}\right) = c_1E\left(b_1|\mathbf{x}\right) + c_2E\left(b_2|\mathbf{x}\right) = c_1\beta_1 + c_2\beta_2 = \lambda$$

Then, using the law of iterated expectations, $E\left(\hat{\lambda}\right) = E_{\mathbf{x}}\left[E\left(\hat{\lambda}|\mathbf{x}\right)\right] = E_{\mathbf{x}}[\lambda] = \lambda$. To find the variance of $\hat{\lambda}$, recall from the Probability Primer, Section P.5.6, that if X and Y are random variables, and if a and b are constants, then the variance $\text{var}(aX + bY)$ is given in equation (P.20) as

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\,\text{cov}(X, Y)$$

In the estimator $\left(c_1b_1 + c_2b_2\right)$, both b_1 and b_2 are random variables, as we do not know what their values will be until a sample is drawn and estimates calculated. Applying (P.20), we have

$$\text{var}\left(\hat{\lambda}|\mathbf{x}\right) = \text{var}\left(c_1b_1 + c_2b_2|\mathbf{x}\right) = c_1^2\text{var}\left(b_1|\mathbf{x}\right) + c_2^2\text{var}\left(b_2|\mathbf{x}\right) + 2c_1c_2\text{cov}\left(b_1, b_2|\mathbf{x}\right) \tag{3.8}$$

The variances and covariances of the least squares estimators are given in (2.14)–(2.16). We estimate $\text{var}\left(\hat{\lambda}|\mathbf{x}\right) = \text{var}\left(c_1b_1 + c_2b_2|\mathbf{x}\right)$ by replacing the unknown variances and covariances with their estimated variances and covariances in (2.20)–(2.22). Then

$$\widehat{\text{var}}\left(\hat{\lambda}|\mathbf{x}\right) = \widehat{\text{var}}\left(c_1b_1 + c_2b_2|\mathbf{x}\right) = c_1^2\widehat{\text{var}}\left(b_1|\mathbf{x}\right) + c_2^2\widehat{\text{var}}\left(b_2|\mathbf{x}\right) + 2c_1c_2\widehat{\text{cov}}\left(b_1, b_2|\mathbf{x}\right) \tag{3.9}$$

The standard error of $\hat{\lambda} = c_1b_1 + c_2b_2$ is the square root of the estimated variance,

$$\text{se}\left(\hat{\lambda}\right) = \text{se}\left(c_1b_1 + c_2b_2\right) = \sqrt{\widehat{\text{var}}\left(c_1b_1 + c_2b_2|\mathbf{x}\right)} \tag{3.10}$$

If in addition SR6 holds, or if the sample is large, the least squares estimators b_1 and b_2 have normal distributions. It is also true that linear combinations of normally distributed variables are normally distributed, so that

$$\hat{\lambda}|\mathbf{x} = c_1b_1 + c_2b_2 \sim N\left[\lambda, \text{var}\left(\hat{\lambda}|\mathbf{x}\right)\right]$$

where $\text{var}\left(\hat{\lambda}|\mathbf{x}\right)$ is given in (3.8). You may be thinking of how long such calculations will take using a calculator, but don't worry. Most computer software will do the calculations for you. Now it's time for an example.

EXAMPLE 3.7 | Estimating Expected Food Expenditure

An executive might ask of the research staff, "Give me an estimate of average weekly food expenditure by households with \$2,000 weekly income." Interpreting the executive's word "average" to mean "expected value," for the food expenditure model this means estimating

$$E(FOOD\_EXP|INCOME) = \beta_1 + \beta_2 INCOME$$

Recall that we measured income in \$100 units in this example, so a weekly income of \$2,000 corresponds to $INCOME = 20$. The executive is requesting an estimate of

$$E(FOOD\_EXP|INCOME = 20) = \beta_1 + \beta_2 20$$

which is a linear combination of the parameters.

Using the 40 observations in the data file *food*, in Section 2.3.2, we obtained the fitted regression,

$$\widehat{FOOD\_EXP} = 83.4160 + 10.2096 INCOME$$

The point estimate of average weekly food expenditure for a household with \$2,000 income is

$$\begin{aligned}\widehat{E(FOOD\_EXP|INCOME = 20)} &= b_1 + b_2 20\\ &= 83.4160 + 10.2096(20) = 287.6089\end{aligned}$$

We estimate that the expected food expenditure by a household with \$2,000 income is \$287.61 per week.

EXAMPLE 3.8 | An Interval Estimate of Expected Food Expenditure

If assumption SR6 holds, and given **x**, the estimator $\hat{\lambda}$ has a normal distribution. We can form a standard normal random variable as

$$Z = \frac{\hat{\lambda} - \lambda}{\sqrt{\operatorname{var}\left(\hat{\lambda}|\mathbf{x}\right)}} \sim N(0, 1)$$

Replacing the true variance in the denominator with the estimated variance, we form a pivotal t-statistic

$$t = \frac{\hat{\lambda} - \lambda}{\sqrt{\widehat{\operatorname{var}}\left(\hat{\lambda}\right)}} = \frac{\hat{\lambda} - \lambda}{\operatorname{se}\left(\hat{\lambda}\right)} = \frac{\left(c_1 b_1 + c_2 b_2\right) - \left(c_1 \beta_1 + c_2 \beta_2\right)}{\operatorname{se}\left(c_1 b_1 + c_2 b_2\right)} \sim t_{(N-2)} \tag{3.11}$$

If t_c is the $1 - \alpha/2$ percentile value from the $t_{(N-2)}$ distribution, then $P\left(-t_c \le t \le t_c\right) = 1 - \alpha$. Substitute (3.11) for t and rearrange to obtain

$$P\left[\left(c_1 b_1 + c_2 b_2\right) - t_c \operatorname{se}\left(c_1 b_1 + c_2 b_2\right) \le c_1 \beta_1 + c_2 \beta_2 \le \left(c_1 b_1 + c_2 b_2\right) + t_c \operatorname{se}\left(c_1 b_1 + c_2 b_2\right)\right] = 1 - \alpha$$

Thus, a $100(1 - \alpha)\%$ interval estimate for $c_1 \beta_1 + c_2 \beta_2$ is

$$\left(c_1 b_1 + c_2 b_2\right) \pm t_c \operatorname{se}\left(c_1 b_1 + c_2 b_2\right)$$

In Example 2.5, we obtained the estimated covariance matrix

$$\begin{bmatrix} \widehat{\operatorname{var}}\left(b_1\right) & \widehat{\operatorname{cov}}\left(b_1, b_2\right) \\ \widehat{\operatorname{cov}}\left(b_1, b_2\right) & \widehat{\operatorname{var}}\left(b_2\right) \end{bmatrix} = \begin{array}{c|cc} & C & INCOME \\ \hline C & 1884.442 & -85.9032 \\ INCOME & -85.9032 & 4.3818 \end{array}$$

To obtain the standard error for $b_1 + b_2 20$, we first calculate the estimated variance

$$\begin{aligned} \widehat{\operatorname{var}}\left(b_1 + 20 b_2\right) &= \widehat{\operatorname{var}}\left(b_1\right) + \left(20^2 \times \widehat{\operatorname{var}}\left(b_2\right)\right) \\ &\quad + \left(2 \times 20 \times \widehat{\operatorname{cov}}\left(b_1, b_2\right)\right) \\ &= 1884.442 + \left(20^2 \times 4.3818\right) \\ &\quad + (2 \times 20 \times (-85.9032)) \\ &= 201.0169 \end{aligned}$$

Given $\widehat{\operatorname{var}}\left(b_1 + 20 b_2\right) = 201.0169$, the corresponding standard error is[2]

$$\begin{aligned} \operatorname{se}\left(b_1 + 20 b_2\right) &= \sqrt{\widehat{\operatorname{var}}\left(b_1 + 20 b_2\right)} = \sqrt{201.0169} \\ &= 14.1780 \end{aligned}$$

A 95% interval estimate of $E(FOOD\_EXP|INCOME = 20) = \beta_1 + \beta_2(20)$ is $\left(b_1 + b_2 20\right) \pm t_{(0.975,38)} \operatorname{se}\left(b_1 + b_2 20\right)$ or

$$\begin{aligned} &[287.6089 - 2.024(14.1780),\ 287.6089 + 2.024(14.1780)] \\ &= [258.91,\ 316.31] \end{aligned}$$

We estimate with 95% confidence that the expected food expenditure by a household with \$2,000 income is between \$258.91 and \$316.31.

3.6.1 Testing a Linear Combination of Parameters

So far, we have tested hypotheses involving only one regression parameter at a time. That is, our hypotheses have been of the form $H_0 : \beta_k = c$. A more **general linear hypothesis** involves both parameters and may be stated as

$$H_0 : c_1 \beta_1 + c_2 \beta_2 = c_0 \tag{3.12a}$$

where c_0, c_1, and c_2 are specified constants, with c_0 being the hypothesized value. Despite the fact that the null hypothesis involves both coefficients, it still represents a single hypothesis to be tested using a t-statistic. Sometimes, it is written equivalently in implicit form as

$$H_0 : \left(c_1 \beta_1 + c_2 \beta_2\right) - c_0 = 0 \tag{3.12b}$$

[2]The value 201.0169 was obtained using computer software. If you do the calculation by hand using the provided numbers, you obtain 201.034. Do not be alarmed if you obtain small differences like this occasionally, as it most likely is the difference between a computer-generated solution and a hand calculation.

The alternative hypothesis for the null hypothesis in (3.12a) might be

i. $H_1: c_1\beta_1 + c_2\beta_2 \neq c_0$ leading to a two-tail t-test

ii. $H_1: c_1\beta_1 + c_2\beta_2 > c_0$ leading to a right-tail t-test [Null may be "≤"]

iii. $H_1: c_1\beta_1 + c_2\beta_2 < c_0$ leading to a left-tail t-test [Null may be "≥"]

If the implicit form is used, the alternative hypothesis is adjusted as well.

The test of the hypothesis (3.12) uses the pivotal t-statistic

$$t = \frac{\left(c_1 b_1 + c_2 b_2\right) - c_0}{\operatorname{se}\left(c_1 b_1 + c_2 b_2\right)} \sim t_{(N-2)} \text{ if the null hypothesis is true} \tag{3.13}$$

The rejection regions for the one- and two-tail alternatives (i)–(iii) are the same as those described in Section 3.3, and conclusions are interpreted the same way as well.

The form of the t-statistic is very similar to the original specification in (3.7). In the numerator, $\left(c_1 b_1 + c_2 b_2\right)$ is the best linear unbiased estimator of $\left(c_1\beta_1 + c_2\beta_2\right)$, and if the errors are normally distributed, or if we have a large sample, this estimator is normally distributed as well.

EXAMPLE 3.9 | Testing Expected Food Expenditure

The food expenditure model introduced in Section 2.1 and used as an illustration throughout provides an excellent example of how the **linear hypothesis** in (3.12) might be used in practice. For most medium and larger cities, there are forecasts of income growth for the coming year. A supermarket or food retail store of any type will consider this before a new facility is built. Their question is, if income in a locale is projected to grow at a certain rate, how much of that will be spent on food items? An executive might say, based on years of experience, "I expect that a household with \$2,000 weekly income will spend, on average, more than \$250 a week on food." How can we use econometrics to test this conjecture?

The regression function for the food expenditure model is

$$E(FOOD\_EXP|INCOME) = \beta_1 + \beta_2 INCOME$$

The executive's conjecture is that

$$E(FOOD\_EXP|INCOME = 20) = \beta_1 + \beta_2 20 > 250$$

To test the validity of this statement, we use it as the alternative hypothesis

$$H_1: \beta_1 + \beta_2 20 > 250, \text{ or } H_1: \beta_1 + \beta_2 20 - 250 > 0$$

The corresponding null hypothesis is the logical alternative to the executive's statement

$$H_0: \beta_1 + \beta_2 20 \leq 250, \text{ or } H_0: \beta_1 + \beta_2 20 - 250 \leq 0$$

Notice that the null and alternative hypotheses are in the same form as the general linear hypothesis with $c_1 = 1$, $c_2 = 20$, and $c_0 = 250$.

The rejection region for a right-tail test is illustrated in Figure 3.2. For a right-tail test at the $\alpha = 0.05$ level of significance, the t-critical value is the 95th percentile of the $t_{(38)}$ distribution, which is $t_{(0.95,38)} = 1.686$. If the calculated t-statistic value is greater than 1.686, we will reject the null hypothesis and accept the alternative hypothesis, which in this case is the executive's conjecture.

Computing the t-statistic value

$$t = \frac{\left(b_1 + 20b_2\right) - 250}{\operatorname{se}\left(b_1 + 20b_2\right)}$$

$$= \frac{(83.4160 + 20 \times 10.2096) - 250}{14.1780}$$

$$= \frac{287.6089 - 250}{14.1780} = \frac{37.6089}{14.1780} = 2.65$$

Since $t = 2.65 > t_c = 1.686$, we reject the null hypothesis that a household with weekly income of \$2,000 will spend \$250 per week or less on food and conclude that the executive's conjecture that such households spend more than \$250 is correct, with the probability of Type I error 0.05.

In Example 3.8, we estimated that a household with \$2,000 weekly income will spend \$287.6089, which is greater than the executive's speculated value of \$250. However, simply observing that the estimated value is greater than \$250 is not a statistical test. It might be numerically greater, but is it **significantly** greater? The t-test takes into account the precision with which we have estimated this expenditure level and also controls the probability of Type I error.

3.7 Exercises

3.7.1 Problems

3.1 There were 64 countries in 1992 that competed in the Olympics and won at least one medal. Let *MEDALS* be the total number of medals won, and let *GDPB* be GDP (billions of 1995 dollars). A linear regression model explaining the number of medals won is $MEDALS = \beta_1 + \beta_2 GDPB + e$. The estimated relationship is

$$\widehat{MEDALS} = b_1 + b_2 GDPB = \underset{(2.38994)}{7.61733} + \underset{(0.00215)}{0.01309} GDPB \quad \text{(se)} \qquad \text{(XR3.1)}$$

a. We wish to test the hypothesis that there is no relationship between the number of medals won and *GDP* against the alternative there is a positive relationship. State the null and alternative hypotheses in terms of the model parameters.
b. What is the test statistic for part (a) and what is its distribution if the null hypothesis is true?
c. What happens to the distribution of the test statistic for part (a) if the alternative hypothesis is true? Is the distribution shifted to the left or right, relative to the usual t-distribution? [*Hint*: What is the expected value of b_2 if the null hypothesis is true, and what is it if the alternative is true?]
d. For a test at the 1% level of significance, for what values of the t-statistic will we reject the null hypothesis in part (a)? For what values will we fail to reject the null hypothesis?
e. Carry out the t-test for the null hypothesis in part (a) at the 1% level of significance. What is your economic conclusion? What does 1% level of significance mean in this example?

3.2 There were 64 countries in 1992 that competed in the Olympics and won at least one medal. Let *MEDALS* be the total number of medals won, and let *GDPB* be GDP (billions of 1995 dollars). A linear regression model explaining the number of medals won is $MEDALS = \beta_1 + \beta_2 GDPB + e$. The estimated relationship is given in equation (XR3.1) in Exercise 3.1.

a. We wish to test the null hypothesis that a one-billion dollar increase in *GDP* leads to an increase in average, or expected, number of medals won by 0.015, against the alternative that it does not. State the null and alternative hypotheses in terms of the model parameters.
b. What is the test statistic for part (a) and what is its distribution if the null hypothesis is true?
c. What happens to the distribution of the test statistic in part (a) if the alternative hypothesis is true? Is the distribution shifted to the left or right, relative to the usual t-distribution, or is the direction of the shift uncertain? [*Hint*: What is the expected value of b_2 if the null hypothesis is true, and what is it if the alternative is true?]
d. For a test at the 10% level of significance, for what values of the t-statistic, will we reject the null hypothesis in part (a)? For what values, will we fail to reject the null hypothesis?
e. Carry out the t-test for the null hypothesis in part (a). What is your economic conclusion?
f. If we carry out the test in part (a) at the 5% level of significance, what do we conclude? At the 1% level of significance, what do we conclude?
g. Carry out the same test at the 5% level of significance, but changing the null hypothesis value of interest to 0.016, then 0.017. What is the calculated t-statistic value in each case? Which hypotheses do you reject, and which do you fail to reject?

3.3 There were 64 countries in 1992 that competed in the Olympics and won at least one medal. Let *MEDALS* be the total number of medals won, and let *GDPB* be GDP (billions of 1995 dollars). A linear regression model explaining the number of medals won is $MEDALS = \beta_1 + \beta_2 GDPB + e$. The estimated relationship is given in equation (XR3.1) in Exercise 3.1.

The estimated covariance between the slope and intercept estimators is −0.00181 and the estimated error variance is $\hat{\sigma}^2 = 320.336$. The sample mean of *GDPB* is $\overline{GDPB} = 390.89$ and the sample variance of *GDPB* is $s^2_{GDPB} = 1099615$.

a. Estimate the expected number of medals won by a country with $GDPB = 25$.
b. Calculate the standard error of the estimate in (a) using for the variance $\widehat{\text{var}}(b_1) + (25)^2 \widehat{\text{var}}(b_2) + (2)(25)\widehat{\text{cov}}(b_1, b_2)$.

c. Calculate the standard error of the estimate in (a) using for the variance $\hat{\sigma}^2 \left\{ (1/N) + \left[\left(25 - \overline{GDPB}\right)^2 \Big/ \left((N-1)\, s^2_{GDPB}\right) \right] \right\}$.

d. Construct a 95% interval estimate for the expected number of medals won by a country with $GDPB = 25$.

e. Construct a 95% interval estimate for the expected number of medals won by a country with $GDPB = 300$. Compare and contrast this interval estimate to that in part (d). Explain the differences you observe.

3.4 Assume that assumptions SR1–SR6 hold for the simple linear regression model, $y_i = \beta_1 + \beta_2 x_i + e_i$, $i = 1, \ldots, N$. Generally, as the sample size N becomes larger, confidence intervals become narrower.

a. Is a narrower confidence interval for a parameter, such as β_2, desirable? Explain why or why not.

b. Give **two** specific reasons why, as the sample size gets larger, a confidence interval for β_2 tends to become narrower. The reasons should relate to the properties of the least squares estimator and/or interval estimation procedures.

3.5 If we have a large sample of data, then using critical values from the standard normal distribution for constructing a ***p*-value** is justified. But how large is "large"?

a. For a t-distribution with 30 degrees of freedom, the right-tail p-value for a t-statistic of 1.66 is 0.05366666. What is the approximate p-value using the cumulative distribution function of the standard normal distribution, $\Phi(z)$, in Statistical Table 1? Using a right-tail test with $\alpha = 0.05$, would you make the correct decision about the null hypothesis using the approximate p-value? Would the exact p-value be larger or smaller for a t-distribution with 90 degrees of freedom?

b. For a t-distribution with 200 degrees of freedom, the right-tail p-value for a t-statistic of 1.97 is 0.0251093. What is the approximate p-value using the standard normal distribution? Using a **two-tail test** with $\alpha = 0.05$, would you make the correct decision about the null hypothesis using the approximate p-value? Would the exact p-value be larger or smaller for a t-distribution with 90 degrees of freedom?

c. For a t-distribution with 1000 degrees of freedom, the right-tail p-value for a t-statistic of 2.58 is 0.00501087. What is the approximate p-value using the standard normal distribution? Using a two-tail test with $\alpha = 0.05$, would you make the correct decision about the null hypothesis using the approximate p-value? Would the exact p-value be larger or smaller for a t-distribution with 2000 degrees of freedom?

3.6 We have data on 2323 randomly selected households consisting of three persons in 2013. Let *ENTERT* denote the monthly entertainment expenditure ($) per person per month and let *INCOME* ($100) be monthly household income. Consider the simple linear regression model $ENTERT_i = \beta_1 + \beta_2 INCOME_i + e_i, i = 1, \ldots, 2323$. Assume that assumptions SR1–SR6 hold. The least squares estimated equation is $\widehat{ENTERT}_i = 9.820 + 0.503 INCOME_i$. The standard error of the slope coefficient estimator is $\text{se}(b_2) = 0.029$, the standard error of the intercept estimator is $\text{se}(b_1) = 2.419$, and the estimated covariance between the least squares estimators b_1 and b_2 is -0.062.

a. Construct a 90% confidence interval estimate for β_2 and interpret it for a group of CEOs from the entertainment industry.

b. The CEO of AMC Entertainment Mr. Lopez asks you to estimate the average monthly entertainment expenditure per person for a household with monthly income (for the three-person household) of $7500. What is your estimate?

c. AMC Entertainment's staff economist asks you for the estimated variance of the estimator $b_1 + 75b_2$. What is your estimate?

d. AMC Entertainment is planning to build a luxury theater in a neighborhood with average monthly income, for three-person households, of $7500. Their staff of economists has determined that in order for the theater to be profitable the average household will have to spend more than $45 per person per month on entertainment. Mr. Lopez asks you to provide conclusive statistical evidence, beyond reasonable doubt, that the proposed theater will be profitable. Carefully set up the null and alternative hypotheses, give the test statistic, and test rejection region using $\alpha = 0.01$. Using the information from the previous parts of the question, carry out the test and provide your result to the AMC Entertainment CEO.

e. The income elasticity of entertainment expenditures at the point of the means is $\varepsilon = \beta_2\left(\overline{INCOME}/\overline{ENTERT}\right)$. The sample means of these variables are $\overline{ENTERT} = 45.93$ and

$\overline{INCOME} = 71.84$. Test the null hypothesis that the elasticity is 0.85 against the alternative that it is not 0.85, using the $\alpha = 0.05$ level of significance.

f. Using Statistical Table 1, compute the approximate two-tail p-value for the t-statistic in part (e). Using the p-value rule, do you reject the null hypothesis $\varepsilon = \beta_2\left(\overline{INCOME}/\overline{ENTERT}\right) = 0.85$, versus the alternative $\varepsilon \neq 0.85$, at the 10% level of significance? Explain.

3.7 We have 2008 data on $INCOME =$ income per capita (in thousands of dollars) and $BACHELOR =$ percentage of the population with a bachelor's degree or more for the 50 U.S. States plus the District of Columbia, a total of $N = 51$ observations. The results from a simple linear regression of $INCOME$ on $BACHELOR$ are

$$\begin{array}{rccc} \widehat{INCOME} = & (a) & + & 1.029BACHELOR \\ \text{se} & (2.672) & & (c) \\ t & (4.31) & & (10.75) \end{array}$$

a. Using the information provided calculate the estimated intercept. Show your work.

b. Sketch the estimated relationship. Is it increasing or decreasing? Is it a positive or inverse relationship? Is it increasing or decreasing at a constant rate or is it increasing or decreasing at an increasing rate?

c. Using the information provided calculate the standard error of the slope coefficient. Show your work.

d. What is the value of the t-statistic for the null hypothesis that the intercept parameter equals 10?

e. The p-value for a two-tail test that the intercept parameter equals 10, from part (d), is 0.572. Show the p-value in a sketch. On the sketch, show the rejection region if $\alpha = 0.05$.

f. Construct a 99% interval estimate of the slope. Interpret the interval estimate.

g. Test the null hypothesis that the slope coefficient is one against the alternative that it is not one at the 5% level of significance. State the economic result of the test, in the context of this problem.

3.8 Using 2011 data on 141 U.S. public research universities, we examine the relationship between cost per student and full-time university enrollment. Let $ACA =$ real academic cost per student (thousands of dollars), and let $FTESTU =$ full-time student enrollment (thousands of students). The least squares fitted relation is $\widehat{ACA} = 14.656 + 0.266FTESTU$.

a. For the regression, the 95% interval estimate for the intercept is [10.602, 18.710]. Calculate the standard error of the estimated intercept.

b. From the regression output, the standard error for the slope coefficient is 0.081. Test the null hypothesis that the true slope, β_2, is 0.25 (or less) against the alternative that the true slope is greater than 0.25 using the 10% level of significance. Show all steps of this hypothesis test, including the null and alternative hypotheses, and state your conclusion.

c. On the regression output, the automatically provided p-value for the estimated slope is 0.001. What is the meaning of this value? Use a sketch to illustrate your answer.

d. A member of the board of supervisors states that ACA should fall if we admit more students. Using the estimated equation and the information in parts (a)–(c), test the null hypothesis that the slope parameter β_2 is zero, or positive, against the alternative hypothesis that it is negative. Use the 5% level of significance. Show all steps of this hypothesis test, including the null and alternative hypotheses, and state your conclusion. Is there any statistical support for the board member's conjecture?

e. In 2011, Louisiana State University (LSU) had a full-time student enrollment of 27,950. Based on the estimated equation, the least squares estimate of $E(ACA|FTESTU = 27,950)$ is 22.079, with standard error 0.964. The actual value of ACA for LSU that year was 21.403. Would you say that this value is surprising or not surprising? Explain.

3.9 Using data from 2013 on 64 black females, the estimated linear regression between $WAGE$ (earnings per hour, in \$) and years of education, $EDUC$ is $\widehat{WAGE} = -8.45 + 1.99EDUC$.

a. The standard error of the estimated slope coefficient is 0.52. Construct and interpret a 95% interval estimate for the effect of an additional year of education on a black female's expected hourly wage rate.

b. The standard error of the estimated intercept is 7.39. Test the null hypothesis that the intercept $\beta_1 = 0$ against the alternative that the true intercept is not zero, using the $\alpha = 0.10$ level of significance. In your answer, show (i) the formal null and alternative hypotheses, (ii) the test statistic and its distribution under the null hypothesis, (iii) the rejection region (in a figure), (iv) the calculated value of the test statistic, and (v) state your conclusion, with its economic interpretation.

c. Estimate the expected wage for a black female with 16 years of education, $E(WAGE|EDUC = 16)$.

d. The estimated covariance between the intercept and slope is −3.75. Construct a 95% interval estimate for the expected wage for a black female with 16 years of education.

e. It is conjectured that a black female with 16 years of education will have an expected wage of more than $23 per hour. Use this as the "alternative hypothesis" in a test of the conjecture at the 10% level of significance. Does the evidence support the conjecture or not?

3.10 Using data from 2013 on 64 black females, the estimated log-linear regression between *WAGE* (earnings per hour, in $) and years of education, *EDUC* is $\widehat{\ln(WAGE)} = 1.58 + 0.09EDUC$. The reported t-statistic for the slope coefficient is 3.95.

a. Test at the 5% level of significance, the null hypothesis that the return to an additional year of education is less than or equal to 8% against the alternative that the rate of return to education is more than 8%. In your answer, show (i) the formal null and alternative hypotheses, (ii) the test statistic and its distribution under the null hypothesis, (iii) the rejection region (in a figure), (iv) the calculated value of the test statistic, and (v) state your conclusion, with its economic interpretation.

b. Testing the null hypothesis that the return to education is 8%, against the alternative that it is not 8%, we obtain the p-value 0.684. What is the p-value for the test in part (a)? In a sketch, show for the test in part (a) the p-value and the 5% critical value from the t-distribution.

c. Construct a 90% interval estimate for the return to an additional year of education and state its interpretation.

3.11 The theory of labor supply indicates that more labor services will be offered at higher wages. Suppose that *HRSWK* is the usual number of hours worked per week by a randomly selected person and *WAGE* is their hourly wage. Our regression model is specified as $HRSWK = \beta_1 + \beta_2 WAGE + e$. Using a sample of 9799 individuals from 2013, we obtain the estimated regression $\widehat{HRSWK} = 41.58 + 0.011WAGE$. The estimated variances and covariance of the least squares estimators are as follows:

	INTERCEPT	*WAGE*
INTERCEPT	0.02324	−0.00067
WAGE	−0.00067	0.00003

a. Test the null hypothesis that the relationship has slope that is less than, or equal to, zero at the 5% level of significance. State the null and alternative hypotheses in terms of the model parameters. Using the results, do we confirm or refute the theory of labor supply?

b. Use Statistical Table 1 of normal probabilities to calculate an approximate p-value for the test in (a). Draw a sketch representing the p-value.

c. Under assumptions SR1–SR6 of the simple regression model, the expected number of hours worked per week is $E(HRSWK|WAGE) = \beta_1 + \beta_2 WAGE$. Construct a 95% interval estimate for the expected number of hours worked per week for a person earning $20/h.

d. In the sample, there are 203 individuals with hourly wage $20. The average number of hours worked for these people is 41.68. Is this result compatible with the interval estimate in (c)? Explain your reasoning.

e. Test the null hypothesis that the expected hours worked for a person earning $20 per hour is 41.68, against the alternative that it is not, at the 1% level of significance.

3.12 Consider a log-linear regression for the weekly sales (number of cans) of a national brand of canned tuna ($SAL1$ = target brand sales) as a function of the ratio of its price to the price of a competitor, $RPRICE3 = 100$(price of target brand ÷ price competitive brand #3), $\ln(SAL1) = \gamma_1 + \gamma_2 RPRICE3 + e$. Using $N = 52$ weekly observations the least squares estimated equation is

$$\begin{array}{rl} \ln(SAL1) = & 11.481 - 0.031RPRICE3 \\ (\text{se}) & (0.535) \quad (0.00529) \end{array}$$

a. The variable *RPRICE3* is the price of the target brand as a percentage of the price of competitive brand #3 or more simply "the relative price." The sample mean of *RPRICE3* is 99.66, its median is 100, its minimum value is 70.11, and its maximum value is 154.24. What do these summary statistics tell us about the prices of the target brand relative to the prices of its competitor?

b. Interpret the coefficient of *RPRICE*3. Does its sign make economic sense?

c. Construct and interpret a 95% interval estimate for the effect on the weekly sales, *SAL*1, of a 1% increase in the price of the target brand as a percentage of the price of competitive brand #3, which is relative price *RPRICE*3.

d. Carry out a test of the null hypothesis $H_0 : \gamma_2 \geq -0.02$ against the alternative $H_1 : \gamma_2 < -0.02$ using the $\alpha = 0.01$ level of significance. Include in your answer (i) the test statistic and its distribution if the null hypothesis is true, (ii) a sketch of the rejection region, (iii) show the location of the test statistic value, (iv) state your conclusion, and (v) show on the sketch the region that would represent the *p*-value.

e. "Hypothesis tests and interval estimators for the regression model are valid as long as the regression error terms are normally distributed." Is this true or false? Explain.

3.13 Consider the following estimated area response model for sugar cane (area of sugar cane planted in thousands of hectares in a region of Bangladesh), as a function of relative price (100 times the price of sugar cane divided by the price of jute, which is an alternative crop to sugar cane, planted by Bangladesh farmers), $\widehat{AREA}_t = -0.24 + 0.50RPRICE_t$ using 34 annual observations.

a. The sample average of *RPRICE* is 114.03, with a minimum of 74.9 and a maximum of 182.2. *RPRICE* is the price of sugar cane taken as a percentage of the price of jute. What do these sample statistics tell us about the relative price of sugar cane?

b. Interpret the intercept and slope of the estimated relation.

c. The *t*-statistic is −0.01 for the hypothesis that the intercept parameter is zero. What do you conclude? Is this an economically surprising result? Explain.

d. The sample mean area planted is 56.83 thousand hectares, and the sample mean for relative price is 114.03. Taking these values as given, test at the 5% level of significance the hypothesis that the elasticity of area response to price at the means is 1.0. The estimated variance of the coefficient of *RPRICE* is 0.020346.

e. The model is re-estimated in log-linear form, obtaining $\widehat{\ln(AREA_t)} = 3.21 + 0.0068RPRICE_t$. Interpret the coefficient of *RPRICE*. The standard error of the slope estimate is 0.00229. What does that tell us about the estimated relationship?

f. Using the model in (e), test the null hypothesis that a 1% increase in the price of sugar cane relative to the price of jute increases the area planted in sugar cane by 1%. Use the 5% level of significance and a two-tail test. Include (i) the test statistic and its distribution if the null hypothesis is true, (ii) a sketch of the rejection region, (iii) show the location of the test statistic value, (iv) state your conclusion, and (v) show on the sketch, the region that would represent the *p*-value.

3.14 What is the meaning of statistical significance and how valuable is this concept? A *t*-statistic is $t = (b - c)/\text{se}(b)$, where b is an estimate of a parameter β, c is the hypothesized value, and se(b) is the standard error. If the sample size N is large, then the statistic is approximately a standard normal distribution if the null hypothesis $\beta = c$ is true.

a. With a 5% level of significance, we assert that an event happening with less than a one in 20 chance is "statistically significant," while an event happening with more than a one in 20 chance is not statistically significant. True or False?

b. Would you say something happening one time in 10 by chance (10%) is very improbable or not very improbable? Would you say something happening one time in 100 by chance (1%) is very improbable or not?

c. If we adopt a rule that in large samples, a *t*-value greater than 2.0 (in absolute value) indicates statistical significance, and we use Statistical Table 1 of standard normal cumulative probabilities, what is the implied significance level? If we adopt a rule that in large samples, a *t*-value greater than 3.0 (in absolute value) indicates statistical significance, what is the implied significance level?

d. Suppose that we clinically test two diet pills, one called "Reliable" and another called "More." Using the Reliable pill, the estimated weight loss is 5 lbs with a standard error of 0.5 lbs. With the More pill, the estimated weight loss is 20 lbs with standard error 10 lbs. When testing whether the true weight loss is zero (the null, or none, hypothesis), what are the *t*-statistic values? What is the ratio of the *t*-values?

e. If the drugs Reliable and More were equivalent in safety, cost and every other comparison, and if your goal was weight loss, which drug would you take? Why?

3.15 In a capital murder trial, with a potential penalty of life in prison, would you as judge tell the jury to make sure that we accidently convict an innocent person only one time in a hundred, or use some other threshold? What would it be?

a. What is the economic cost of a Type I error in this example? List some of the factors that would have to be considered in such a calculation.

b. What is the economic cost of a Type II error in this example? List some of the factors that would have to be considered in such a calculation.

3.16 A big question in the United States, a question of "cause and effect," is whether mandatory health care will really make Americans healthier. What is the role of hypothesis testing in such an investigation?

a. Formulate null and alternative hypotheses based on the question.

b. What is a Type I error in the context of this question? What factors would you consider if you were assigned the task of calculating the economic cost of a Type I error in this example?

c. What is a Type II error in the context of this question? What factors would you consider if you were assigned the task of calculating the economic cost of a Type II error in this example?

d. If we observe that individuals who have health insurance are in fact healthier, does this prove that we should have mandatory health care?

e. There is a saying, "Correlation does not imply causation." How might this saying relate to part (d)?

f. Post hoc ergo propter hoc (Latin: "after this, therefore because of this") is a logical fallacy discussed widely in Principles of Economics textbooks. An example might be "A rooster crows and then the sun appears, thus the crowing rooster causes the sun to rise." How might this fallacy relate to the observation in part (d)?

3.17 Consider the regression model $WAGE = \beta_1 + \beta_2 EDUC + e$. Where $WAGE$ is hourly wage rate in US 2013 dollars. $EDUC$ is years of schooling. The model is estimated twice, once using individuals from an urban area, and again for individuals in a rural area.

$$\text{Urban} \qquad \begin{array}{l} \widehat{WAGE} = -10.76 + 2.46EDUC, \quad N = 986 \\ \quad\text{(se)} \qquad\quad (2.27) \quad (0.16) \end{array}$$

$$\text{Rural} \qquad \begin{array}{l} \widehat{WAGE} = -4.88 + 1.80EDUC, \quad N = 214 \\ \quad\text{(se)} \qquad\quad (3.29) \quad (0.24) \end{array}$$

a. Using the urban regression, test the null hypothesis that the regression slope equals 1.80 against the alternative that it is greater than 1.80. Use the $\alpha = 0.05$ level of significance. Show all steps, including a graph of the critical region and state your conclusion.

b. Using the rural regression, compute a 95% interval estimate for expected $WAGE$ if $EDUC = 16$. The required standard error is 0.833. Show how it is calculated using the fact that the estimated covariance between the intercept and slope coefficients is -0.761.

c. Using the urban regression, compute a 95% interval estimate for expected $WAGE$ if $EDUC = 16$. The estimated covariance between the intercept and slope coefficients is -0.345. Is the interval estimate for the urban regression wider or narrower than that for the rural regression in (b). Do you find this plausible? Explain.

d. Using the rural regression, test the hypothesis that the intercept parameter β_1 equals four, or more, against the alternative that it is less than four, at the 1% level of significance.

3.18 A life insurance company examines the relationship between the amount of life insurance held by a household and household income. Let $INCOME$ be household income (thousands of dollars) and $INSURANCE$ the amount of life insurance held (thousands of dollars). Using a random sample of $N = 20$ households, the least squares estimated relationship is

$$\begin{array}{l} \widehat{INSURANCE} = 6.855 + 3.880INCOME \\ \qquad\text{(se)} \qquad\qquad (7.383) \quad (0.112) \end{array}$$

a. Draw a sketch of the fitted relationship identifying the estimated slope and intercept. The sample mean of $INCOME = 59.3$. What is the sample mean of the amount of insurance held? Locate the point of the means in your sketch.

b. How much do we estimate that the average amount of insurance held changes with each additional \$1000 of household income? Provide both a point estimate and a 95% interval estimate. Explain the interval estimate to a group of stockholders in the insurance company.

c. Construct a 99% interval estimate of the expected amount of insurance held by a household with \$100,000 income. The estimated covariance between the intercept and slope coefficient is −0.746.

d. One member of the management board claims that for every \$1000 increase in income the average amount of life insurance held will increase by \$5000. Let the algebraic model be $INSURANCE = \beta_1 + \beta_2 INCOME + e$. Test the hypothesis that the statement is true against the alternative that it is not true. State the conjecture in terms of a null and alternative hypothesis about the model parameters. Use the 5% level of significance. Do the data support the claim or not? Clearly, indicate the test statistic used and the rejection region.

e. Test the hypothesis that as income increases the amount of life insurance held increases by the same amount. That is, test the null hypothesis that the slope is one. Use as the alternative that the slope is larger than one. State the null and alternative hypotheses in terms of the model parameters. Carry out the test at the 1% level of significance. Clearly indicate the test statistic used, and the rejection region. What is your conclusion?

3.7.2 Computer Exercises

3.19 The owners of a motel discovered that a defective product was used during construction. It took 7 months to correct the defects during which approximately 14 rooms in the 100-unit motel were taken out of service for 1 month at a time. The data are in the file *motel*.

a. Plot *MOTEL_PCT* and *COMP_PCT* versus *TIME* on the same graph. What can you say about the occupancy rates over time? Do they tend to move together? Which seems to have the higher occupancy rates? Estimate the regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$. Construct a 95% interval estimate for the parameter β_2. Have we estimated the association between *MOTEL_PCT* and *COMP_PCT* relatively precisely, or not? Explain your reasoning.

b. Construct a 90% interval estimate of the expected occupancy rate of the motel in question, *MOTEL_PCT*, given that $COMP\_PCT = 70$.

c. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0 : \beta_2 \leq 0$ against the alternative hypothesis $H_0 : \beta_2 > 0$ at the $\alpha = 0.01$ level of significance. Discuss your conclusion. Clearly define the test statistic used and the rejection region.

d. In the linear regression model $MOTEL\_PCT = \beta_1 + \beta_2 COMP\_PCT + e$, test the null hypothesis $H_0 : \beta_2 = 1$ against the alternative hypothesis $H_0 : \beta_2 \neq 1$ at the $\alpha = 0.01$ level of significance. If the null hypothesis were true, what would that imply about the motel's occupancy rate versus their competitor's occupancy rate? Discuss your conclusion. Clearly define the test statistic used and the rejection region.

e. Calculate the least squares residuals from the regression of *MOTEL_PCT* on *COMP_PCT* and plot them against *TIME*. Are there any unusual features to the plot? What is the predominant sign of the residuals during time periods 17–23 (July, 2004 to January, 2005)?

3.20 The owners of a motel discovered that a defective product was used during construction. It took seven months to correct the defects during which approximately 14 rooms in the 100-unit motel were taken out of service for one month at a time. The data are in the file *motel*.

a. Calculate the sample average occupancy rate for the motel during the time when there were no repairs being made. What is the sample average occupancy rate for the motel during the time when there were repairs being made? How big a difference is there?

b. Consider the linear regression $MOTEL\_PCT = \delta_1 + \delta_2 REPAIR + e$, where *REPAIR* is an indicator variable taking the value 1 during the repair period and 0 otherwise. What are the estimated coefficients? How do these estimated coefficients relate to the calculations in part (a)?

c. Construct a 95% interval estimate for the parameter δ_2 and give its interpretation. Have we estimated the effect of the repairs on motel occupancy relatively precisely, or not? Explain.

d. The motel wishes to claim economic damages because the faulty materials led to repairs which cost them customers. To do so, their economic consultant tests the null hypothesis $H_0 : \delta_2 \geq 0$ against the alternative hypothesis $H_1 : \delta_2 < 0$. Explain the logic behind stating the null and alternative hypotheses in this way. Carry out the test at the $\alpha = 0.05$ level of significance. Discuss your conclusions. Clearly state the test statistic, the rejection region, and the p-value.

e. To further the motel's claim, the consulting economist estimates a regression model $(MOTEL\_PCT - COMP\_PCT) = \gamma_1 + \gamma_2 REPAIR + e$, so that the dependent variable is the difference in the occupancy rates. Construct and discuss the economic meaning of the 95% interval estimate of γ_2.

f. Test the null hypothesis that $\gamma_2 = 0$ against the alternative that $\gamma_2 < 0$ at the $\alpha = 0.01$ level of significance. Discuss the meaning of the test outcome. Clearly state the test statistic, the rejection region, and the *p*-value.

3.21 The capital asset pricing model (CAPM) is described in Exercise 2.16. Use all available observations in the data file *capm5* for this exercise.

a. Construct 95% interval estimates of Exxon-Mobil's and Microsoft's "beta." Assume that you are a stockbroker. Explain these results to an investor who has come to you for advice.

b. Test at the 5% level of significance the hypothesis that Ford's "beta" value is one against the alternative that it is not equal to one. What is the economic interpretation of a beta equal to one? Repeat the test and state your conclusions for General Electric's stock and Exxon-Mobil's stock. Clearly state the test statistic used and the rejection region for each test, and compute the *p*-value.

c. Test at the 5% level of significance the null hypothesis that Exxon-Mobil's "beta" value is greater than or equal to one against the alternative that it is less than one. Clearly state the test statistic used and the rejection region for each test, and compute the *p*-value. What is the economic interpretation of a beta less than one?

d. Test at the 5% level of significance the null hypothesis that Microsoft's "beta" value is less than or equal to one against the alternative that it is greater than one. Clearly state the test statistic used and the rejection region for each test, and compute the *p*-value. What is the economic interpretation of a beta more than one?

e. Test at the 5% significance level, the null hypothesis that the intercept term in the CAPM model for Ford's stock is zero, against the alternative that it is not. What do you conclude? Repeat the test and state your conclusions for General Electric's stock and Exxon-Mobil's stock. Clearly state the test statistic used and the rejection region for each test, and compute the *p*-value.

3.22 The data file *collegetown* contains data on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price (in \$1000 units), *PRICE*, and total interior area in hundreds of square feet, *SQFT*.

a. Using the linear regression $PRICE = \beta_1 + \beta_2 SQFT + e$, estimate the elasticity of expected house *PRICE* with respect to *SQFT*, evaluated at the sample means. Construct a 95% interval estimate for the elasticity, treating the sample means as if they are given (not random) numbers. What is the interpretation of the interval?

b. Test the null hypothesis that the elasticity, calculated in part (a), is one against the alternative that the elasticity is not one. Use the 1% level of significance. Clearly state the test statistic used, the rejection region, and the test *p*-value. What do you conclude?

c. Using the linear regression model $PRICE = \beta_1 + \beta_2 SQFT + e$, test the hypothesis that the marginal effect on expected house price of increasing house size by 100 square feet is less than or equal to \$13000 against the alternative that the marginal effect will be greater than \$13000. Use the 5% level of significance. Clearly state the test statistic used, the rejection region, and the test *p*-value. What do you conclude?

d. Using the linear regression $PRICE = \beta_1 + \beta_2 SQFT + e$, estimate the expected price, $E(PRICE|SQFT) = \beta_1 + \beta_2 SOFT$, for a house of 2000 square feet. Construct a 95% interval estimate of the expected price. Describe your interval estimate to a general audience.

e. Locate houses in the sample with 2000 square feet of living area. Calculate the sample mean (average) of their selling prices. Is the sample average of the selling price for houses with $SQFT = 20$ compatible with the result in part (d)? Explain.

3.23 The data file *collegetown* contains data on 500 single-family houses sold in Baton Rouge, Louisiana, during 2009–2013. The data include sale price in \$1000 units, *PRICE*, and total interior area in hundreds of square feet, *SQFT*.

a. Using the quadratic regression model, $PRICE = \alpha_1 + \alpha_2 SOFT^2 + e$, test the hypothesis that the marginal effect on expected house price of increasing the size of a 2000 square foot house by 100 square feet is less than or equal to \$13000 against the alternative that the marginal effect will be greater than \$13000. Use the 5% level of significance. Clearly state the test statistic used, the rejection region, and the test *p*-value. What do you conclude?

b. Using the quadratic regression model in part (a), test the hypothesis that the marginal effect on expected house price of increasing the size of a 4000 square foot house by 100 square feet is less than or equal to \$13000 against the alternative that the marginal effect will be greater than \$13000.

Use the 5% level of significance. Clearly state the test statistic used, the rejection region, and the test p-value. What do you conclude?

c. Using the quadratic regression model in part (a), estimate the expected price $E(PRICE|SQFT) = \alpha_1 + \alpha_2 SQFT^2$ for a house of 2000 square feet. Construct a 95% interval estimate of the expected price. Describe your interval estimate to a general audience.

d. Locate houses in the sample with 2000 square feet of living area. Calculate the sample mean (average) of their selling prices. Is the sample average of the selling price for houses with $SQFT = 20$ compatible with the result in part (c)? Explain.

3.24 We introduced Professor Ray C. Fair's model for explaining and predicting U.S. presidential elections in Exercise 2.23. Fair's data, 26 observations for the election years from 1916 to 2016, are in the data file *fair5*. The dependent variable is $VOTE =$ percentage share of the popular vote won by the Democratic party. Define $GROWTH = INCUMB \times growth\ rate$, where growth rate is the annual rate of change in real per capita GDP in the first three quarters of the election year. If Democrats are the incumbent party, then $INCUMB = 1$; if the Republicans are the incumbent party then $INCUMB = -1$.

a. Estimate the linear regression, $VOTE = \beta_1 + \beta_2 GROWTH + e$, using data from 1916 to 2016. Construct a 95% interval estimate of the effect of economic growth on expected $VOTE$. How would you describe your finding to a general audience?

b. The expected $VOTE$ in favor of the Democratic candidate is $E(VOTE|GROWTH) = \beta_1 + \beta_2 GROWTH$. Estimate $E(VOTE|GROWTH = 4)$ and construct a 95% interval estimate and a 99% interval estimate. Assume a Democratic incumbent is a candidate for a second presidential term. Is achieving a 4% growth rate enough to ensure a victory? Explain.

c. Test the hypothesis that when $INCUMB = 1$ economic growth has either a zero or negative effect on expected $VOTE$ against the alternative that economic growth has a positive effect on expected $VOTE$. Use the 1% level of significance. Clearly state the test statistic used, the rejection region, and the test p-value. What do you conclude?

d. Define $INFLAT = INCUMB \times$ inflation rate, where the inflation rate is the growth in prices over the first 15 quarters of an administration. Using the data from 1916 to 2016, and the model $VOTE = \alpha_1 + \alpha_2 INFLAT + e$, test the hypothesis that inflation has no effect against the alternative that it does have an effect. Use the 1% level of significance. State the test statistic used, the rejection region, and the test p-value and state your conclusion.

3.25 Using data on the "Ashcan School," we have an opportunity to study the market for art. What factors determine the value of a work of art? Use the data in the file *ashcan_small*. [Note: the file *ashcan* contains more variables.]

a. Define $YEARS\_OLD = DATE\_AUCTN - CREATION$, which is the age of the painting at the time of its sale. Use data on works that sold ($SOLD = 1$) to estimate the regression $\ln(RHAMMER) = \beta_1 + \beta_2 YEARS\_OLD + e$. Construct a 95% interval estimate for the percentage change in real hammer price given that a work of art is another year old at the time of sale. [*Hint*: Review the discussion of equation (2.28).] Explain the result to a potential art buyer.

b. Test the null hypothesis that each additional year of age increases the "hammer price" by 2%, against the two-sided alternative. Use the 5% level of significance.

c. The variable $DREC$ is an indicator variable taking the value one if a sale occurred during a recession and is zero otherwise. Use data on works that sold ($SOLD = 1$) to estimate the regression model $\ln(RHAMMER) = \alpha_1 + \alpha_2 DREC + e$. Construct a 95% interval estimate of the percentage reduction in hammer price when selling in a recession. Explain your finding to a client who is considering selling during a recessionary period.

d. Test the conjecture that selling a work of art during a recession reduces the hammer price by 2% or less, against the alternative that the reduction in hammer price is greater than 2%. Use the 5% level of significance. Clearly state the test statistic used, the rejection region, and the test p-value. What is your conclusion?

3.26 How much does experience affect wage rates? The data file *cps5_small* contains 1200 observations on hourly wage rates, experience, and other variables from the March 2013 Current Population Survey (CPS). [Note: The data file *cps5* contains more observations and variables.]

a. Estimate the linear regression $WAGE = \beta_1 + \beta_2 EXPER + e$ and discuss the results.

b. Test the statistical significance of the estimated relationship at the 5% level. Use a one-tail test. What is your alternative hypothesis? What do you conclude?

c. Estimate the linear regression $WAGE = \beta_1 + \beta_2 EXPER + e$ for individuals living in a metropolitan area, where $METRO = 1$. Is there a statistically significant positive relationship between expected wages and experience at the 1% level? How much of an effect is there?

d. Estimate the linear regression $WAGE = \beta_1 + \beta_2 EXPER + e$ for individuals not living in a metropolitan area, where $METRO = 0$. Is there a statistically significant positive relationship between expected wages and experience at the 1% level? Can we safely say that experience has no effect on wages for individuals living in nonmetropolitan areas? Explain.

3.27 Is the relationship between experience and wages constant over one's lifetime? We will investigate this question using a quadratic model. The data file *cps5_small* contains 1200 observations on hourly wage rates, experience, and other variables from the March 2013 Current Population Survey (CPS). [Note: the data file *cps5* contains more observations and variables.]

a. Create the variable $EXPER30 = EXPER - 30$. Describe this variable. When is it positive, negative or zero?

b. Estimate by least squares the quadratic model $WAGE = \gamma_1 + \gamma_2 (EXPER30)^2 + e$. Test the null hypothesis that $\gamma_2 = 0$ against the alternative $\gamma_2 \neq 0$ at the 1% level of significance. Is there a statistically significant quadratic relationship between expected $WAGE$ and $EXPER30$?

c. Create a plot of the fitted value $\widehat{WAGE} = \hat{\gamma}_1 + \hat{\gamma}_2 (EXPER30)^2$, on the y-axis, versus $EXPER30$, on the x-axis. Up to the value $EXPER30 = 0$ is the slope of the plot constant, or is it increasing, or decreasing? Up to the value $EXPER30 = 0$ is the function increasing at an increasing rate or increasing at a decreasing rate?

d. If $y = a + bx^2$ then $dy/dx = 2bx$. Using this result, calculate the estimated slope of the fitted function $\widehat{WAGE} = \hat{\gamma}_1 + \hat{\gamma}_2 (EXPER30)^2$, when $EXPER = 0$, when $EXPER = 10$, and when $EXPER = 20$.

e. Calculate the t-statistic for the null hypothesis that the slope of the function is zero, $H_0{:}2\gamma_2 EXPER30 = 0$, when $EXPER = 0$, when $EXPER = 10$, and when $EXPER = 20$.

3.28 The owners of a motel discovered that a defective product was used during construction. It took 7 months to correct the defects during which approximately 14 rooms in the 100-unit motel were taken out of service for 1 month at a time. The data are in the file *motel*.

a. Create a new variable, $RELPRICE2 = 100RELPRICE$, which equals the percentage of the competitor's price charged by the motel in question. Plot $RELPRICE2$ against $TIME$. Compute the summary statistics for this variable. What are the sample mean and median? What are the minimum and maximum values? Does the motel in question charge more than its competitors for a room, or less, or about the same? Explain.

b. Consider a linear regression with $y = MOTEL\_PCT$ and $x = RELPRICE2$. Interpret the estimated slope coefficient. Construct a 95% interval estimate for the slope. Have we estimated the slope of the relationship very well? Explain your answer.

c. Construct a 90% interval estimate of the expected motel occupancy rate if the motel's price is 80% of its competitor's price. Do you consider the interval relatively narrow or relatively wide? Explain your reasoning.

d. Test the null hypothesis that there is no relationship between the variables against the alternative that there is an inverse relationship between them, at the $\alpha = 0.05$ level of significance. Discuss your conclusion. Be sure to include in your answer the test statistic used, the rejection region, and the p-value.

e. Test the hypothesis that for each percent higher for the relative price that the motel in question charges, it loses 1% of its occupancy rate. Formulate the null and alternative hypotheses in terms of the model parameters, carry out the relevant test at the 5% level of significance, and state your conclusion. Be sure to state the test statistic used, the rejection region, and the p-value.

3.29 We introduced Tennessee's Project STAR (Student/Teacher Achievement Ratio) in Exercise 2.22. The data file is *star5_small*. [The data file *star5* contains more observations and more variables.] Three types of classes were considered: small classes [$SMALL = 1$], regular-sized classes with a teacher aide [$AIDE = 1$], and regular-sized classes [$REGULAR = 1$].

a. Compute the sample mean and standard deviation for student math scores, $MATHSCORE$, in small classes. Compute the sample mean and standard deviation for student math scores, $MATHSCORE$, in regular classes, with no teacher aide. Which type of class had the higher average score? What is the difference in sample average scores for small classes versus regular-sized classes? Which type of class had the higher score standard deviation?

b. Consider students only in small classes or regular-sized classes without a teacher aide. Estimate the regression model $MATHSCORE = \beta_1 + \beta_2 SMALL + e$. How do the estimates of the regression parameters relate to the sample average scores calculated in part (a)?

c. Using the model from part (b), construct a 95% interval estimate of the expected *MATHSCORE* for a student in a regular-sized class and a student in a small class. Are the intervals fairly narrow or not? Do the intervals overlap?

d. Test the null hypothesis that the expected mathscore is no different in the two types of classes versus the alternative that expected *MATHSCORE* is higher for students in small classes using the 5% level of significance. State these hypotheses in terms of the model parameters, clearly state the test statistic you use, and the test rejection region. Calculate the p-value for the test. What is your conclusion?

e. Test the null hypothesis that the expected *MATHSCORE* is 15 points higher for students in small classes versus the alternative that it is not 15 points higher using the 10% level of significance. State these hypotheses in terms of the model parameters, clearly state the test statistic you use, and the test rejection region. Calculate the p-value for the test. What is your conclusion?

3.30 We introduced Tennessee's Project STAR (Student/Teacher Achievement Ratio) in Exercise 2.22. The data file is *star5_small*. [The data file *star5* contains more observations and more variables.] Three types of classes were considered: small classes [$SMALL = 1$], regular-sized classes with a teacher aide [$AIDE = 1$], and regular-sized classes [$REGULAR = 1$].

a. Compute the sample mean and standard deviation for student math scores, *MATHSCORE*, in regular classes with no teacher aide. Compute the sample mean and standard deviation for student math scores, *MATHSCORE*, in regular classes with a teacher aide. Which type of class had the higher average score? What is the difference in sample average scores for regular-sized classes versus regular sized classes with a teacher aide? Which type of class had the higher score standard deviation?

b. Consider students only in regular sized classes without a teacher aide and regular sized classes with a teacher aide. Estimate the regression model $MATHSCORE = \beta_1 + \beta_2 AIDE + e$. How do the estimates of the regression parameters relate to the sample average scores calculated in part (a)?

c. Using the model from part (b), construct a 95% interval estimate of the expected *MATHSCORE* for a student in a regular-sized class without a teacher aide and a regular-sized class with a teacher aide. Are the intervals fairly narrow or not? Do the intervals overlap?

d. Test the null hypothesis that the expected *MATHSCORE* is no different in the two types of classes versus the alternative that expected *MATHSCORE* is higher for students in regular-sized classes with a teacher aide, using the 5% level of significance. State these hypotheses in terms of the model parameters, clearly state the test statistic you use, and the test rejection region. Calculate the p-value for the test. What is your conclusion?

e. Test the null hypothesis that the expected *MATHSCORE* is three points, or more, higher for students in regular-sized classes with a teacher aide versus the alternative that the difference is less than three points, using the 10% level of significance. State these hypotheses in terms of the model parameters, clearly state the test statistic you use and the test rejection region. Calculate the p-value for the test. What is your conclusion?

3.31 Data on weekly sales of a major brand of canned tuna by a supermarket chain in a large midwestern U.S. city during a mid-1990s calendar year are contained in the data file *tuna*. There are 52 observations for each of the variables. The variable $SAL1$ = unit sales of brand no. 1 canned tuna, and $APR1$ = price per can of brand no. 1 tuna (in dollars).

a. Calculate the summary statistics for $SAL1$ and $APR1$. What are the sample means, minimum and maximum values, and their standard deviations. Plot each of these variables versus *WEEK*. How much variation in sales and price is there from week to week?

b. Plot the variable $SAL1$ (y-axis) against $APR1$ (x-axis). Is there a positive or inverse relationship? Is that what you expected, or not? Why?

c. Create the variable $PRICE1 = 100APR1$. Estimate the linear regression $SAL1 = \beta_1 + \beta_2 PRICE1 + e$. What is the point estimate for the effect of a one cent increase in the price of brand no. 1 on the sales of brand no. 1? What is a 95% interval estimate for the effect of a one cent increase in the price of brand no. 1 on the sales of brand no. 1?

d. Construct a 90% interval estimate for the expected number of cans sold in a week when the price per can is 70 cents.

e. Construct a 95% interval estimate of the elasticity of sales of brand no. 1 with respect to the price of brand no. 1 "at the means." Treat the sample means as constants and not random variables. Do you find the sales are fairly elastic, or fairly inelastic, with respect to price? Does this make economic sense? Why?

f. Test the hypothesis that elasticity of sales of brand no. 1 with respect to the price of brand no. 1 from part (e) is minus three against the alternative that the elasticity is not minus three. Use the 10% level of significance. Clearly, state the null and alternative hypotheses in terms of the model parameters, give the rejection region, and the p-value for the test. What is your conclusion?

3.32 What is the relationship between crime and punishment? We use data from 90 North Carolina counties to examine the question. County crime rates and other characteristics are observed over the period 1981–1987. The data are in the file *crime*. Use the 1985 data for this exercise.

a. Calculate the summary statistics for *CRMRTE* (crimes committed per person) and *PRBARR* (the probability of arrest = the ratio of arrests to offenses), including the maximums and minimums. Does there appear to be much variation from county to county in these variables?

b. Plot *CRMRTE* versus *PRBARR*. Do you observe a relationship between these variables?

c. Estimate the linear regression model $CRMRTE = \beta_1 + \beta_2 PRBARR + e$. If we increase the probability of arrest by 10% what will be the effect on the crime rate? What is a 95% interval estimate of this quantity?

d. Test the null hypothesis that there is no relationship between the county crime rate and the probability of arrest versus the alternative that there is an inverse relationship. State the null and alternative hypotheses in terms of the model parameters. Clearly, state the test statistic and its distribution if the null hypothesis is true and the test rejection region. Use the 1% level of significance. What is your conclusion?

Appendix 3A Derivation of the *t*-Distribution

Interval estimation and hypothesis testing procedures in this chapter involve the t-distribution. Here we develop the key result.

The first result that is needed is the normal distribution of the least squares estimator. Consider, for example, the normal distribution of b_2 the least squares estimator of β_2, which we denote as

$$b_2|\mathbf{x} \sim N\left(\beta_2, \frac{\sigma^2}{\sum\left(x_i - \bar{x}\right)^2}\right)$$

A standardized normal random variable is obtained from b_2 by subtracting its mean and dividing by its standard deviation:

$$Z = \frac{b_2 - \beta_2}{\sqrt{\operatorname{var}\left(b_2|\mathbf{x}\right)}} \sim N(0, 1) \tag{3A.1}$$

That is, the standardized random variable Z is normally distributed with mean 0 and variance 1. Despite the fact that the distribution of the least squares estimator b_2 depends on $\mathbf{x}$, the standardization leaves us with a pivotal statistic whose distribution depends on neither unknown parameters nor $\mathbf{x}$.

The second piece of the puzzle involves a chi-square random variable. If assumption SR6 holds, then the random error term e_i has a conditional normal distribution, $e_i|\mathbf{x} \sim N\left(0, \sigma^2\right)$. Standardize the random variable by dividing by its standard deviation so that $e_i/\sigma \sim N(0, 1)$. The square of a standard normal random variable is a chi-square random variable (see Appendix B.5.2) with one degree of freedom, so $\left(e_i/\sigma\right)^2 \sim \chi^2_{(1)}$. If all the random errors are independent, then

$$\sum\left(\frac{e_i}{\sigma}\right)^2 = \left(\frac{e_1}{\sigma}\right)^2 + \left(\frac{e_2}{\sigma}\right)^2 + \cdots + \left(\frac{e_N}{\sigma}\right)^2 \sim \chi^2_{(N)} \tag{3A.2}$$

Since the true random errors are unobservable, we replace them by their sample counterparts, the least squares residuals $\hat{e}_i = y_i - b_1 - b_2 x_i$, to obtain

$$V = \frac{\sum \hat{e}_i^2}{\sigma^2} = \frac{(N-2)\,\hat{\sigma}^2}{\sigma^2} \tag{3A.3}$$

The random variable V in (3A.3) does not have a $\chi^2_{(N)}$ distribution because the least squares residuals are *not* independent random variables. All N residuals $\hat{e}_i = y_i - b_1 - b_2 x_i$ depend on the least squares estimators b_1 and b_2. It can be shown that only $N-2$ of the least squares residuals are independent in the simple linear regression model. Consequently, the random variable in (3A.3) has a chi-square distribution with $N-2$ degrees of freedom. That is, when multiplied by the constant $(N-2)/\sigma^2$, the random variable $\hat{\sigma}^2$ has a *chi-square distribution with* $N-2$ *degrees of freedom*,

$$V = \frac{(N-2)\,\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(N-2)} \tag{3A.4}$$

The random variable V has a distribution that depends only on the degrees of freedom, $N-2$. Like Z in (3A.1), V is a pivotal statistic. We have *not* established the fact that the chi-square random variable V is statistically independent of the least squares estimators b_1 and b_2, but it is. The proof is beyond the scope of this book. Consequently, V and the standard normal random variable Z in (3A.1) are independent.

From the two random variables V and Z, we can form a t-random variable. A t-random variable is formed by dividing a standard normal random variable, $Z \sim N(0,1)$, by the square root of an *independent* chi-square random variable, $V \sim \chi^2_{(m)}$, that has been divided by its degrees of freedom, m. That is,

$$t = \frac{Z}{\sqrt{V/m}} \sim t_{(m)} \tag{3A.5}$$

The t-distribution's shape is completely determined by the degrees of freedom parameter, m, and the distribution is symbolized by $t_{(m)}$. See Appendix B.5.3. Using Z and V from (3A.1) and (3A.4), respectively, we have

$$\begin{aligned} t &= \frac{Z}{\sqrt{V/(N-2)}} = \frac{(b_2-\beta_2)\Big/\sqrt{\sigma^2/\sum(x_i-\bar{x})^2}}{\sqrt{\dfrac{(N-2)\,\hat{\sigma}^2/\sigma^2}{N-2}}} \\ &= \frac{b_2-\beta_2}{\sqrt{\dfrac{\hat{\sigma}^2}{\sum(x_i-\bar{x})^2}}} = \frac{b_2-\beta_2}{\sqrt{\widehat{\operatorname{var}}(b_2)}} = \frac{b_2-\beta_2}{\operatorname{se}(b_2)} \sim t_{(N-2)} \end{aligned} \tag{3A.6}$$

The second line is the key result that we state in (3.2), with its generalization in (3.3).

Appendix 3B Distribution of the t-Statistic under H_1

To better understand how t-tests work, let us examine the t-statistic in (3.7) when the null hypothesis is not true. We can do that by writing it out in some additional detail. What happens to Z in (3A.1) if we test a hypothesis $H_0\!:\beta_2 = c$ that might not be true? Instead of subtracting β_2, we subtract c, to obtain

$$\frac{b_2-c}{\sqrt{\operatorname{var}(b_2)}} = \frac{b_2-\beta_2+\beta_2-c}{\sqrt{\operatorname{var}(b_2)}} = \frac{b_2-\beta_2}{\sqrt{\operatorname{var}(b_2)}} + \frac{\beta_2-c}{\sqrt{\operatorname{var}(b_2)}} = Z+\delta \sim N(\delta, 1)$$

The statistic we obtain is the standard normal Z plus another factor, $\delta = (\beta_2 - c)\Big/\sqrt{\text{var}(b_2)}$, that is zero only if the null hypothesis is true. A **noncentral *t*-random variable** is formed from the ratio

$$t|\mathbf{x} = \frac{Z + \delta}{\sqrt{V/m}} \sim t_{(m,\delta)} \tag{3B.1}$$

This is a more general t-statistic, with m degrees of freedom and noncentrality parameter δ, denoted $t_{(m,\,\delta)}$. It has a distribution that is not centered at zero unless $\delta = 0$. The non-central t-distribution is introduced in Appendix B.7.3. It is the factor δ that leads the t-test to reject a false null hypothesis with probability greater than α, which is the probability of a Type I error. Because δ depends on the sample data, we have indicated that the non-central t-distribution is conditional on $\mathbf{x}$. If the null hypothesis is true then $\delta = 0$ and the t-statistic does not depend on any unknown parameters or $\mathbf{x}$; it is a pivotal statistic.

Suppose that we have a sample of size $N = 40$ so that the degrees of freedom are $N - 2 = 38$ and we test a hypothesis concerning β_2 such that $\beta_2 - c = 1$. Using a right-tail test, the probability of rejecting the null hypothesis is $P(t > 1.686)$, where $t_{(0.95,\,38)} = 1.686$ is from Statistical Table 2, the percentiles of the usual t-distribution. If $\delta = 0$, this rejection probability is 0.05. With $\beta_2 - c = 1$, we must compute the right-tail probability using the non-central t-distribution with noncentrality parameter

$$\delta = \frac{\beta_2 - c}{\sqrt{\text{var}(b_2)}} = \frac{\beta_2 - c}{\sqrt{\sigma^2 / \sum (x_i - \bar{x})^2}} = \frac{\sqrt{\sum (x_i - \bar{x})^2}\,(\beta_2 - c)}{\sigma} \tag{3B.2}$$

For a numerical example, we use values arising from the simulation experiment used in Appendix 2H. The sample of x-values consists of $x_i = 10, i = 1, \ldots, 20$ and $x_i = 20$, $i = 21, \ldots, 40$. The sample mean is $\bar{x} = 15$ so that $\sum (x_i - \bar{x})^2 = 40 \times 5^2 = 1000$. Also, $\sigma^2 = 2500$. The noncentrality parameter is

$$\delta = \frac{\sqrt{\sum (x_i - \bar{x})^2}\,(\beta_2 - c)}{\sigma} = \frac{\sqrt{1000}\,(\beta_2 - c)}{\sqrt{2500}} = 0.63246(\beta_2 - c)$$

Thus, the probability of rejecting the null hypothesis $H_0 : \beta_2 = 9$ versus $H_1 : \beta_2 > 9$ when the true value of $\beta_2 = 10$ is

$$P\left(t_{(38,\,0.63246)} > 1.686\right) = 1 - P\left(t_{(38,\,0.63246)} \le 1.686\right) = 0.15301$$

The probability calculation uses the cumulative distribution function for the non-central t-distribution, which is available in econometric software and at some websites. Similarly, the probability of rejecting the null hypothesis $H_0 : \beta_2 = 8$ versus $H_1 : \beta_2 > 8$ when the true value of $\beta_2 = 10$ is

$$P\left(t_{(38,\,1.26491)} > 1.686\right) = 1 - P\left(t_{(38,\,1.26491)} \le 1.686\right) = 0.34367$$

Why does the probability of rejection increase? The effect of the noncentrality parameter is to shift the t-distribution rightward, as shown in Appendix B.7.3. For example, the probability of rejecting the null hypothesis $H_0 : \beta_2 = 9$ versus $H_1 : \beta_2 > 9$ is shown in Figure 3B.1.

The solid curve is the usual central t-distribution with 38 degrees of freedom. The area under the curve to the right of 1.686 is 0.05. The dashed curve is the non-central t-distribution with $\delta = 0.63246$. The area under the curve to the right of 1.686 is larger, approximately 0.153.

The probability of rejecting a false null hypothesis is called a test's **power**. In an ideal world, we would reject false null hypotheses always, and if we had an infinite amount of data we could. The keys to a t-test's power are the three ingredients making the noncentrality parameter larger.

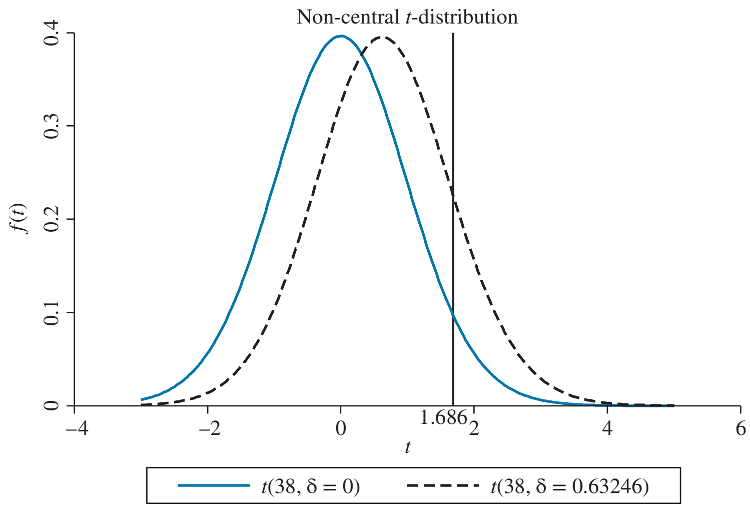

FIGURE 3B.1 **Probability of rejecting $H_0 : \beta_2 = 9$.**

A larger noncentrality parameter shifts the t-distribution further rightward and increases the probability of rejection. Thus, the probability of rejecting a false null hypothesis increases when

1. The magnitude of the hypothesis error $\beta_2 - c$ increases.
2. The smaller the true error variance, σ^2, that measures the overall model uncertainty.
3. The larger the total variation in the explanatory variable, which might be the result of a larger sample size.

In a real situation, the actual power of a test is unknown because we do not know β_2 or σ^2, and the power calculation depends on being given the x-values. Nevertheless, it is good to know the factors that will increase the probability of rejecting a false null hypothesis. In the following section, we carry out a Monte Carlo simulation experiment to illustrate the power calculations above.

Recall that a Type II error is failing to reject a hypothesis that is false. Consequently, the probability of a Type II error is the complement of the test's power. For example, the probability of a Type II error when testing $H_0 : \beta_2 = 9$ versus $H_1 : \beta_2 > 9$ when the true value of $\beta_2 = 10$ is

$$P\left(t_{(38,0.63246)} \leq 1.686\right) = 1 - 0.15301 = 0.84699$$

For testing $H_0 : \beta_2 = 8$ versus $H_1 : \beta_2 > 8$, when the true value is $\beta_2 = 10$, the probability of a Type II error is $P\left(t_{(38,\ 1.26491)} \leq 1.686\right) = 1 - 0.34367 = 0.65633$. As test power increases, the probability of a Type II error falls, and vice versa.

Appendix 3C Monte Carlo Simulation

In Appendix 2H, we introduced a Monte Carlo simulation to illustrate the repeated sampling properties of the least squares estimators. In this appendix, we use the same framework to illustrate the repeated sampling performances of interval estimators and hypothesis tests.

Recall that the **data generation process** for the simple linear regression model is given by

$$y_i = E\left(y_i|x_i\right) + e_i = \beta_1 + \beta_2 x_i + e_i, \quad i = 1, \ldots, N$$

The Monte Carlo parameter values are $\beta_1 = 100$ and $\beta_2 = 10$. The value of x_i is 10 for the first 20 observations and 20 for the remaining 20 observations, so that the regression functions are

$$E(y_i|x_i = 10) = 100 + 10x_i = 100 + 10 \times 10 = 200, \quad i = 1, \ldots, 20$$

$$E(y_i|x_i = 20) = 100 + 10x_i = 100 + 10 \times 20 = 300, \quad i = 21, \ldots, 40$$

The random errors are independently and normally distributed with mean 0 and variance $\text{var}(e_i|x_i) = \sigma^2 = 2{,}500$, or $e_i|x \sim N(0, 2500)$.

When studying the performance of hypothesis tests and interval estimators, it is necessary to use enough Monte Carlo samples so that the percentages involved are estimated precisely enough to be useful. For tests with probability of Type I error $\alpha = 0.05$, we should observe true null hypotheses being rejected 5% of the time. For 95% interval estimators, we should observe that 95% of the interval estimates contain the true parameter values. We use $M = 10{,}000$ Monte Carlo samples so that the experimental error is very small. See Appendix 3C.3 for an explanation.

3C.1 Sampling Properties of Interval Estimators

In Appendix 2H.4, we created one sample of data that is in the data file *mcl_fixed_x*. The least squares estimates using these data values are

$$\hat{y} = 127.2055 + 8.7325x$$
$$(23.3262) \quad (1.4753)$$

A 95% interval estimate of the slope is $b_2 \pm t_{(0.975, 38)}\text{se}(b_2) = [5.7460, 11.7191]$. We see that for this sample, the 95% interval estimate contains the true slope parameter value $\beta_2 = 10$.

We repeat the process of estimation and interval estimation 10,000 times. In these repeated samples, 95.03% of the interval estimates contain the true parameter. Table 3C.1 contains results for the Monte Carlo samples 321–330 for illustration purposes. The estimates are *B2*, the standard error is *SE*, the lower bound of the 95% interval estimate is *LB*, and the upper bound is *UB*. The variable *COVER* = 1 if the interval estimate contains the true parameter value. Two of the intervals do not contain the true parameter value $\beta_2 = 10$. The 10 sample results we are reporting were chosen to illustrate that interval estimates do not cover the true parameter in all cases.

The lesson is, that in many samples from the data generation process, and if assumptions SR1–SR6 hold, the procedure for constructing 95% interval estimates "works" 95% of the time.

TABLE 3C.1 **Results of 10000 Monte Carlo Simulations**

SAMPLE	*B2*	*SE*	*TSTAT*	*REJECT*	*LB*	*UB*	*COVER*
321	7.9600	1.8263	−1.1170	0	4.2628	11.6573	1
322	11.3093	1.6709	0.7836	0	7.9267	14.6918	1
323	9.8364	1.4167	−0.1155	0	6.9683	12.7044	1
324	11.4692	1.3909	1.0563	0	8.6535	14.2849	1
325	9.3579	1.5127	−0.4245	0	6.2956	12.4202	1
326	9.6332	1.5574	−0.2355	0	6.4804	12.7861	1
327	9.0747	1.2934	−0.7154	0	6.4563	11.6932	1
328	7.0373	1.3220	−2.2411	0	4.3611	9.7136	0
329	13.1959	1.7545	1.8215	1	9.6441	16.7478	1
330	14.4851	2.1312	2.1046	1	10.1708	18.7994	0

3C.2 Sampling Properties of Hypothesis Tests

The null hypothesis $H_0 : \beta_2 = 10$ is true. If we use the one-tail alternative $H_1 : \beta_2 > 10$ and level of significance $\alpha = 0.05$, the null hypothesis is rejected if the test statistic $t = (b_2 - 10)/\text{se}(b_2) > 1.68595$, which is the 95th percentile of the t-distribution with 38 degrees of freedom.[3] For the sample *mc1_fixed_x*, the calculated value of the t-statistic is -0.86, so we fail to reject the null hypothesis, which in this case is the correct decision.

We repeat the process of estimation and hypothesis testing 10,000 times. In these samples, 4.98% of the tests reject the null hypothesis that the parameter value is 10. In Table 3C.1, the t-statistic value is *TSTAT* and *REJECT* = 1 if the null hypothesis is rejected. We see that samples 329 and 330 incorrectly reject the null hypothesis.

The lesson is that in many samples from the data generation process, and if assumptions SR1–SR6 hold, the procedure for testing a true null hypothesis at significance level $\alpha = 0.05$ rejects the true null hypothesis 5% of the time. Or, stated positively, the test procedure does not reject the true null hypothesis 95% of the time.

To investigate the power of the t-test, the probability that it rejects a false hypothesis, we tested $H_0 : \beta_2 = 9$ versus $H_1 : \beta_2 > 9$ and $H_0 : \beta_2 = 8$ versus $H_1 : \beta_2 > 8$. The theoretical rejection rates we calculated in Appendix 3B are 0.15301 in the first case and 0.34367 in the second. In 10,000 Monte Carlo samples, the first hypothesis was rejected in 1515 samples for a rejection rate of 0.1515. The second hypothesis was rejected in 3500 of the samples, a rejection rate of 0.35. The Monte Carlo values are very close to the true rejection rates.

3C.3 Choosing the Number of Monte Carlo Samples

A 95% confidence interval estimator should contain the true parameter value 95% of the time in many samples. The M samples in a Monte Carlo experiment are independent experimental trials in which the probability of a "success," an interval containing the true parameter value, is $P = 0.95$. The number of successes follows a **binomial** distribution. The **proportion** of successes $\hat{P}$ in M trials is a random variable with expectation P and variance $P(1 - P)/M$. If the number of Monte Carlo samples M is large, a 95% interval estimate of the proportion of Monte Carlo successes is $P \pm 1.96\sqrt{P(1 - P)/M}$. If $M = 10{,}000$, this interval is [0.9457, 0.9543]. We chose $M = 10{,}000$ so that this interval would be narrow, giving us confidence that *if* the true probability of success is 0.95 we will obtain a Monte Carlo average close to 0.95 with a "high" degree of confidence. Our result that 95.03% of our interval estimates contain the true parameter β_2 is "within" the margin of error for such Monte Carlo experiments. On the other hand, if we had used $M = 1000$ Monte Carlo samples, the interval estimate of the proportion of Monte Carlo successes would be, [0.9365, 0.9635]. With this wider interval, the proportion of Monte Carlo successes could be quite different from 0.95, casting a shadow of doubt on whether our method was working as advertised or not.

Similarly, for a test with probability of rejection $\alpha = 0.05$, the 95% interval estimate of the proportion of Monte Carlo samples leading to rejection is $\alpha \pm 1.96\sqrt{\alpha(1 - \alpha)/M}$. If $M = 10{,}000$, this interval is [0.0457, 0.0543]. That our Monte Carlo experiments rejected the null hypothesis 4.98% of the time is within this margin of error. If we had chosen $M = 1000$, then the proportion of Monte Carlo rejections is estimated to be in the interval [0.0365, 0.0635], which again leaves just a little too much wiggle room for comfort.

The point is that if fewer Monte Carlo samples are chosen the "noise" in the Monte Carlo experiment can lead to a percent of successes or rejections that has too wide a margin of error for

[3]We use a t-critical value with more decimals, instead of the table value 1.686, to ensure accuracy in the Monte Carlo experiment.

us to tell whether the statistical procedure, interval estimation, or hypothesis testing is "working" properly or not.[4]

3C.4 Random-*x* Monte Carlo Results

We used the "fixed-x" framework in Monte Carlo results reported in Sections 3C.1 and 3C.2. In each Monte Carlo sample, the x-values were $x_i = 10$ for the first 20 observations and $x_i = 20$ for the next 20 observations. Now we modify the experiment to the random-x case, as in Appendix 2H.7. The data-generating equation remains $y_i = 100 + 10x_i + e_i$ with the random errors having a normal distribution with mean zero and standard deviation 50, $e_i \sim N(0, 50^2 = 2500)$. We randomly choose x-values from a normal distribution with mean $\mu_x = 15$ and standard deviation $\sigma_x = 1.6$, so $x \sim N(15, 1.6^2 = 2.56)$.

One sample of data is in the data file *mc1_random_x*. Using these values, we obtain the least squares estimates

$$\hat{y} = \underset{(84.7107)}{116.7410} + \underset{(5.5248)}{9.7628}x$$

A 95% interval estimate of the slope is $b_2 \pm t_{(0.975, 38)}\text{se}(b_2) = [-1.4216, 20.9472]$. For this sample, the 95% interval estimate contains the true slope parameter value $\beta_2 = 10$.

We generate 10,000 Monte Carlo samples using this design and compute the least squares estimates and 95% interval estimates. In these samples, with $\mathbf{x}$ varying from sample to sample, the 95% interval estimates for β_2 contain the true value in 94.87% of the samples. Table 3C.2 contains results for the Monte Carlo samples 321–330 for illustration purposes. The estimates are *B2*, the standard error *SE*, the lower bound of the 95% interval estimate is *LB*, and the upper bound is *UB*. The variable *COVER* = 1 if the interval contains the true parameter value. In the selected samples, one interval estimate, 323, does not contain the true parameter value.

In the Monte Carlo experiment, we test the null hypothesis $H_0: \beta_2 = 10$ against the alternative $H_1: \beta_2 > 10$ using the t-statistic $t = (b_2 - 10)/\text{se}(b_2)$. We reject the null hypothesis if $t \geq 1.685954$, which is the 95th percentile of the $t_{(38)}$ distribution. In Table 3C.2, the t-statistic values are *TSTAT* and *REJECT* = 1 if the test rejects the null hypothesis. In 5.36% of the 10,000 Monte Carlo samples, we reject the null hypothesis, which is within the margin of error discussed in Section 3C.2. In Table 3C.2, for sample 323, the true null hypothesis was rejected.

TABLE 3C.2 Results of 10,000 Monte Carlo Simulations with Random-*x*

SAMPLE	*B2*	*SE*	*TSTAT*	*REJECT*	*LB*	*UB*	*COVER*
321	9.6500	5.1341	−0.0682	0	−0.7434	20.0434	1
322	7.4651	4.3912	−0.5773	0	−1.4244	16.3547	1
323	22.9198	5.6616	2.2820	1	11.4584	34.3811	0
324	8.6675	4.8234	−0.2763	0	−1.0970	18.4320	1
325	18.7736	5.2936	1.6574	0	8.0573	29.4899	1
326	16.4197	3.8797	1.6547	0	8.5657	24.2738	1
327	3.7841	5.1541	−1.2060	0	−6.6500	14.2181	1
328	3.6013	4.9619	−1.2896	0	−6.4436	13.6462	1
329	10.5061	5.6849	0.0890	0	−1.0024	22.0145	1
330	9.6342	4.8478	−0.0755	0	−0.1796	19.4481	1

[4]Other details concerning Monte Carlo simulations can be found in *Microeconometrics: Methods and Applications*, by A. Colin Cameron and Pravin K. Trivedi (Cambridge University Press, 2005). The material is advanced.

We conclude from these simulations that in the random-x cases there is no evidence that inferences do not perform as expected, with 95% of intervals covering the true parameter value and 5% of tests rejecting a true null hypothesis.

To investigate the power of the t-test, the probability that it rejects a false hypothesis, we tested $H_0\colon\beta_2 = 9$ versus $H_1\colon\beta_2 > 9$ and $H_0\colon\beta_2 = 8$ versus $H_1\colon\beta_2 > 8$. In 10,000 Monte Carlo samples, the first hypothesis was rejected in 7.8% of the time and the second hypothesis was rejected 11.15% of the time. These rejection rates are far less than in the fixed-x results studied in Appendix 3B and less than the empirical rejection rates in the simulation results in Appendix 3C.2. We noted that the ability of the t-test to reject a false hypothesis was related to the magnitude of the noncentrality parameter in (3A.8), $\delta = \sqrt{\sum\left(x_i - \bar{x}\right)^2}\left(\beta_2 - c\right)/\sigma$. In these experiments, the factors $\left(\beta_2 - c\right) = 1$ and 2 and $\sigma = 50$ are the same as in the fixed-x example. What must have changed? The only remaining factor is the variation in the x-values, $\sum\left(x_i - \bar{x}\right)^2$. In the earlier example, $\sum\left(x_i - \bar{x}\right)^2 = 1000$ and the x-values were fixed in repeated samples. In this experiment, the x-values were not fixed but random, and for each sample of x-values, the amount of variation changes. We specified the variance of x to be 2.56, and in 10,000 Monte Carlo experiments, the average of the sample variance $s_x^2 = 2.544254$ and the average of the variation in x about its mean, $\sum\left(x_i - \bar{x}\right)^2$, was 99.22591, or about one-tenth the variation in the fixed-x case. It is perfectly clear why the power of the test in the random-x case was lower, it is because on average $\sum\left(x_i - \bar{x}\right)^2$ was smaller.