5.5 HYPOTHESIS TESTING: GENERAL COMMENTS

Having discussed the problem of point and interval estimation, we shall now consider the topic of hypothesis testing. In this section we discuss briefly some general aspects of this topic; **Appendix A** gives some additional details.

The problem of statistical hypothesis testing may be stated simply as follows: *Is a given observation or finding compatible with some stated hypothesis or not?* The word "compatible," as used here, means "sufficiently" close to the hypothesized value so that we do not reject the stated hypothesis. Thus, if some theory or prior experience leads us to believe that the true slope coefficient β_2 of the consumption–income example is unity, is the observed $\hat{\beta}_2 = 0.5091$ obtained from the sample of Table 3.2 consistent with the stated hypothesis? If it is, we do not reject the hypothesis; otherwise, we may reject it.

In the language of statistics, the stated hypothesis is known as the **null hypothesis** and is denoted by the symbol H_0. The null hypothesis is usually tested against an **alternative hypothesis** (also known as **maintained hypothesis**) denoted by H_1, which may state, for example, that true β_2 is different from unity. The alternative hypothesis may be **simple** or **composite.**[6] For example, $H_1: \beta_2 = 1.5$ is a simple hypothesis, but $H_1: \beta_2 \neq 1.5$ is a composite hypothesis.

The theory of hypothesis testing is concerned with developing rules or procedures for deciding whether to reject or not reject the null hypothesis. There are two *mutually complementary* approaches for devising such rules, namely, **confidence interval** and **test of significance.** Both these approaches predicate that the variable (statistic or estimator) under consideration has some probability distribution and that hypothesis testing involves making statements or assertions about the value(s) of the parameter(s) of such distribution. For example, we know that with the normality assumption $\hat{\beta}_2$ is normally distributed with mean equal to β_2 and variance given by (4.3.5). If we hypothesize that $\beta_2 = 1$, we are making an assertion about one

---

[6]A statistical hypothesis is called a **simple hypothesis** if it specifies the precise value(s) of the parameter(s) of a probability density function; otherwise, it is called a **composite hypothesis.** For example, in the normal pdf $(1/\sigma\sqrt{2\pi})\exp\{-\frac{1}{2}[(X-\mu)/\sigma]^2\}$, if we assert that $H_1: \mu = 15$ and $\sigma = 2$, it is a simple hypothesis; but if $H_1: \mu = 15$ and $\sigma > 15$, it is a composite hypothesis, because the standard deviation does not have a specific value.

of the parameters of the normal distribution, namely, the mean. Most of the statistical hypotheses encountered in this text will be of this type—making assertions about one or more values of the parameters of some assumed probability distribution such as the normal, F, t, or χ^2. How this is accomplished is discussed in the following two sections.

5.6 HYPOTHESIS TESTING: THE CONFIDENCE-INTERVAL APPROACH

Two-Sided or Two-Tail Test

To illustrate the confidence-interval approach, once again we revert to the consumption–income example. As we know, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$, is 0.5091. Suppose we postulate that

$$H_0: \beta_2 = 0.3$$
$$H_1: \beta_2 \neq 0.3$$

that is, the true MPC is 0.3 under the null hypothesis but it is less than or greater than 0.3 under the alternative hypothesis. The null hypothesis is a simple hypothesis, whereas the alternative hypothesis is composite; actually it is what is known as a **two-sided hypothesis.** Very often such a two-sided alternative hypothesis reflects the fact that we do not have a strong a priori or theoretical expectation about the direction in which the alternative hypothesis should move from the null hypothesis.

Is the observed $\hat{\beta}_2$ compatible with H_0? To answer this question, let us refer to the confidence interval (5.3.9). We know that in the long run intervals like (0.4268, 0.5914) will contain the true β_2 with 95 percent probability. Consequently, in the long run (i.e., repeated sampling) such intervals provide a range or limits within which the true β_2 may lie with a confidence coefficient of, say, 95%. Thus, the confidence interval provides a set of plausible null hypotheses. Therefore, if β_2 under H_0 falls within the $100(1-\alpha)\%$ confidence interval, we do not reject the null hypothesis; if it lies outside the interval, we may reject it.[7] This range is illustrated schematically in Figure 5.2.

Decision Rule: Construct a $100(1-\alpha)\%$ confidence interval for β_2. If the β_2 under H_0 falls within this confidence interval, do not reject H_0, but if it falls outside this interval, reject H_0.

Following this rule, for our hypothetical example, $H_0: \beta_2 = 0.3$ clearly lies outside the 95% confidence interval given in (5.3.9). Therefore, we can reject

[7]Always bear in mind that there is a 100α percent chance that the confidence interval does not contain β_2 under H_0 even though the hypothesis is correct. In short, there is a 100α percent chance of committing a **Type 1 error.** Thus, if $\alpha = 0.05$, there is a 5 percent chance that we could reject the null hypothesis even though it is true.

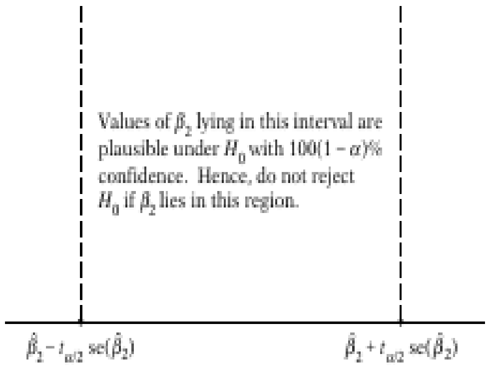

FIGURE 5.2 A 100(1 − α)% confidence interval for β_2.

the hypothesis that the true MPC is 0.3, with 95% confidence. If the null hypothesis were true, the probability of our obtaining a value of MPC of as much as 0.5091 by sheer chance or fluke is at the most about 5 percent, a small probability.

In statistics, when we reject the null hypothesis, we say that our finding is **statistically significant.** On the other hand, when we do not reject the null hypothesis, we say that our finding is **not statistically significant.**

Some authors use a phrase such as "highly statistically significant." By this they usually mean that when they reject the null hypothesis, the probability of committing a Type I error (i.e., α) is a small number, usually 1 percent. But as our discussion of the **p value** in Section 5.8 will show, it is better to leave it to the researcher to decide whether a statistical finding is "significant," "moderately significant," or "highly significant."

One-Sided or One-Tail Test

Sometimes we have a strong a priori or theoretical expectation (or expectations based on some previous empirical work) that the alternative hypothesis is one-sided or unidirectional rather than two-sided, as just discussed. Thus, for our consumption–income example, one could postulate that

$$H_0\colon \beta_2 \le 0.3 \qquad \text{and} \qquad H_1\colon \beta_2 > 0.3$$

Perhaps economic theory or prior empirical work suggests that the marginal propensity to consume is greater than 0.3. Although the procedure to test this hypothesis can be easily derived from (5.3.5), the actual mechanics are better explained in terms of the test-of-significance approach discussed next.[8]

[8]If you want to use the confidence interval approach, construct a (100 − α)% *one-sided* or *one-tail* confidence interval for β_2. Why?

5.7 HYPOTHESIS TESTING: THE TEST-OF-SIGNIFICANCE APPROACH

Testing the Significance of Regression Coefficients: The *t* Test

An *alternative but complementary approach* to the confidence-interval method of testing statistical hypotheses is the **test-of-significance approach** developed along independent lines by R. A. Fisher and jointly by Neyman and Pearson.[9] **Broadly speaking, a test of significance is a procedure by which sample results are used to verify the truth or falsity of a null hypothesis.** The key idea behind tests of significance is that of a **test statistic** (estimator) and the sampling distribution of such a statistic under the null hypothesis. The decision to accept or reject H_0 is made on the basis of the value of the test statistic obtained from the data at hand.

As an illustration, recall that under the normality assumption the variable

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \tag{5.3.2}$$

follows the t distribution with $n - 2$ df. If the value of true β_2 is specified under the null hypothesis, the t value of (5.3.2) can readily be computed from the available sample, and therefore it can serve as a test statistic. And since this test statistic follows the t distribution, confidence-interval statements such as the following can be made:

$$\Pr\left[-t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2^*}{\text{se}(\hat{\beta}_2)} \le t_{\alpha/2}\right] = 1 - \alpha \tag{5.7.1}$$

where β_2^* is the value of β_2 under H_0 and where $-t_{\alpha/2}$ and $t_{\alpha/2}$ are the values of t (the **critical** t values) obtained from the t table for $(\alpha/2)$ level of significance and $n - 2$ df [cf. (5.3.4)]. The t table is given in **Appendix D**.

Rearranging (5.7.1), we obtain

$$\Pr\left[\beta_2^* - t_{\alpha/2}\,\text{se}(\hat{\beta}_2) \le \hat{\beta}_2 \le \beta_2^* + t_{\alpha/2}\,\text{se}(\hat{\beta}_2)\right] = 1 - \alpha \tag{5.7.2}$$

which gives the interval in which $\hat{\beta}_2$ will fall with $1 - \alpha$ probability, given $\beta_2 = \beta_2^*$. In the language of hypothesis testing, the $100(1 - \alpha)\%$ confidence interval established in (5.7.2) is known as the **region of acceptance** (of

[9] Details may be found in E. L. Lehman, *Testing Statistical Hypotheses,* John Wiley & Sons, New York, 1959.

the null hypothesis) and the *region(s)* outside the confidence interval is (are) called the **region(s) of rejection** (of H_0) or the **critical region(s)**. As noted previously, the confidence limits, the endpoints of the confidence interval, are also called **critical values.**

The intimate connection between the confidence-interval and test-of-significance approaches to hypothesis testing can now be seen by comparing (5.3.5) with (5.7.2). In the confidence-interval procedure we try to establish a range or an interval that has a certain probability of including the true but unknown β_2, whereas in the test-of-significance approach we hypothesize some value for β_2 and try to see whether the computed $\hat{\beta}_2$ lies within reasonable (confidence) limits around the hypothesized value.

Once again let us revert to our consumption–income example. We know that $\hat{\beta}_2 = 0.5091$, se $(\hat{\beta}_2) = 0.0357$, and df = 8. If we assume $\alpha = 5$ percent, $t_{\alpha/2} = 2.306$. If we let H_0: $\beta_2 = \beta_2^* = 0.3$ and H_1: $\beta_2 \neq 0.3$, (5.7.2) becomes

$$\Pr\,(0.2177 \le \hat{\beta}_2 \le 0.3823) = 0.95 \qquad \textbf{(5.7.3)}^{10}$$

as shown diagrammatically in Figure 5.3. Since the observed $\hat{\beta}_2$ lies in the critical region, we reject the null hypothesis that true $\beta_2 = 0.3$.

In practice, there is no need to estimate (5.7.2) explicitly. One can compute the t value in the middle of the double inequality given by (5.7.1) and see whether it lies between the critical t values or outside them. For our example,

$$t = \frac{0.5091 - 0.3}{0.0357} = 5.86 \qquad \textbf{(5.7.4)}$$

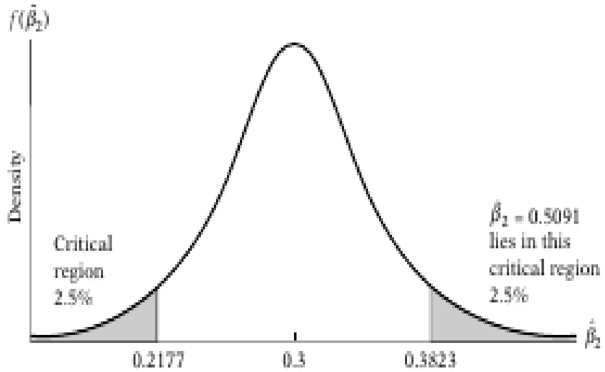

FIGURE 5.3 The 95% confidence interval for $\hat{\beta}_2$ under the hypothesis that $\beta_2 = 0.3$.

[10] In Sec. 5.2, point 4, it was stated that we *cannot* say that the probability is 95 percent that the fixed interval (0.4268, 0.5914) includes the true β_2. But we can make the probabilistic statement given in (5.7.3) because $\hat{\beta}_2$, being an estimator, is a random variable.

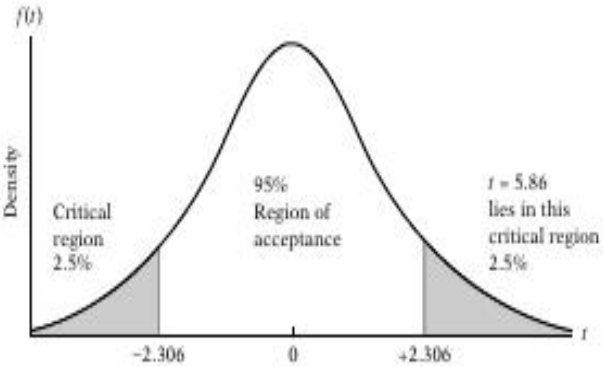

FIGURE 5.4 The 95% confidence interval for t(8 df).

which clearly lies in the critical region of Figure 5.4. The conclusion remains the same; namely, we reject H_0.

Notice that if the estimated β_2 ($= \hat{\beta}_2$) is equal to the hypothesized β_2, the t value in (5.7.4) will be zero. However, as the estimated β_2 value departs from the hypothesized β_2 value, $|t|$ (that is, the absolute t value; *note:* t can be positive as well as negative) will be increasingly large. *Therefore, a "large" $|t|$ value will be evidence against the null hypothesis.* Of course, we can always use the t table to determine whether a particular t value is large or small; the answer, as we know, depends on the degrees of freedom as well as on the probability of Type I error that we are willing to accept. If you take a look at the t table given in **Appendix D**, you will observe that for any given value of df the probability of obtaining an increasingly large $|t|$ value becomes progressively smaller. Thus, for 20 df the probability of obtaining a $|t|$ value of 1.725 or greater is 0.10 or 10 percent, but for the same df the probability of obtaining a $|t|$ value of 3.552 or greater is only 0.002 or 0.2 percent.

Since we use the t distribution, the preceding testing procedure is called appropriately the ***t* test. In the language of significance tests, a statistic is said to be statistically significant if the value of the test statistic lies in the critical region. In this case the null hypothesis is rejected. By the same token, a test is said to be statistically insignificant if the value of the test statistic lies in the acceptance region.** In this situation, the null hypothesis is not rejected. In our example, the t test is significant and hence we reject the null hypothesis.

Before concluding our discussion of hypothesis testing, note that the testing procedure just outlined is known as a **two-sided, or two-tail,** test-of-significance procedure in that we consider the two extreme tails of the relevant probability distribution, the rejection regions, and reject the null hypothesis if it lies in either tail. But this happens because our H_1 was a two-sided composite hypothesis; $\beta_2 \neq 0.3$ means β_2 is either greater than or less than 0.3. But suppose prior experience suggests to us that the MPC is expected to be greater than 0.3. In this case we have: $H_0\colon \beta_2 \leq 0.3$ and $H_1\colon \beta_2 > 0.3$. Although H_1 is still a composite hypothesis, it is now one-sided. To test this hypothesis, we use the **one-tail test** (the right tail), as shown in Figure 5.5. (See also the discussion in Section 5.6.)

The test procedure is the same as before except that the upper confidence limit or critical value now corresponds to $t_\alpha = t_{.05}$, that is, the 5 percent level. As Figure 5.5 shows, we need not consider the lower tail of the t distribution in this case. Whether one uses a two- or one-tail test of significance will depend upon how the alternative hypothesis is formulated, which, in turn, may depend upon some a priori considerations or prior empirical experience. (But more on this in Section 5.8.)

We can summarize the t test of significance approach to hypothesis testing as shown in Table 5.1.

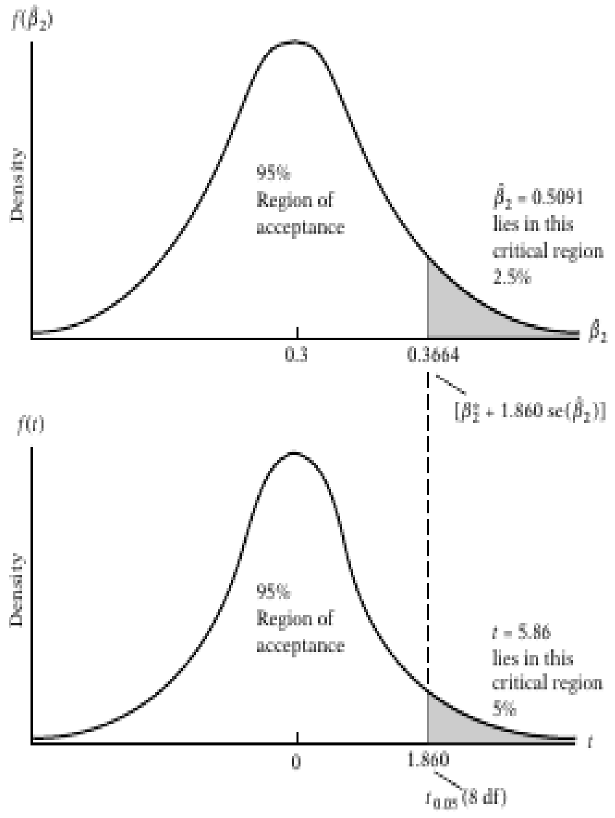

FIGURE 5.5 One-tail test of significance.

TABLE 5.1 THE t TEST OF SIGNIFICANCE: DECISION RULES

Type of hypothesis	H_0: the null hypothesis	H_1: the alternative hypothesis	Decision rule: reject H_0 if
Two-tail	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\|t\| > t_{\alpha/2,\text{df}}$
Right-tail	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{\alpha,\text{df}}$
Left-tail	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t < -t_{\alpha,\text{df}}$

Notes: β_2^* is the hypothesized numerical value of β_2.
$|t|$ means the absolute value of t.
t_α or $t_{\alpha/2}$ means the critical t value at the α or $\alpha/2$ level of significance.
df: degrees of freedom, $(n-2)$ for the two-variable model, $(n-3)$ for the three-variable model, and so on.
The same procedure holds to test hypotheses about β_1.

Testing the Significance of σ^2: The χ^2 Test

As another illustration of the test-of-significance methodology, consider the following variable:

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{5.4.1}$$

which, as noted previously, follows the χ^2 distribution with $n-2$ df. For the hypothetical example, $\hat{\sigma}^2 = 42.1591$ and df = 8. If we postulate that H_0: $\sigma^2 = 85$ vs. H_1: $\sigma^2 \neq 85$, Eq. (5.4.1) provides the test statistic for H_0. Substituting the appropriate values in (5.4.1), it can be found that under H_0, $\chi^2 = 3.97$. If we assume $\alpha = 5\%$, the critical χ^2 values are 2.1797 and 17.5346. Since the computed χ^2 lies between these limits, the data support the null hypothesis and we do not reject it. (See Figure 5.1.) This test procedure is called the **chi-square test of significance.** The χ^2 test of significance approach to hypothesis testing is summarized in Table 5.2.

TABLE 5.2 A SUMMARY OF THE χ^2 TEST

H_0: the null hypothesis	H_1: the alternative hypothesis	Critical region: reject H_0 if
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),\text{df}}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{\text{df}(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,\text{df}}$ or $< \chi^2_{(1-\alpha/2),\text{df}}$

Note: σ_0^2 is the value of σ^2 under the null hypothesis. The first subscript on χ^2 in the last column is the level of significance, and the second subscript is the degrees of freedom. These are critical chi-square values. Note that df is $(n-2)$ for the two-variable regression model, $(n-3)$ for the three-variable regression model, and so on.

5.8 HYPOTHESIS TESTING: SOME PRACTICAL ASPECTS

The Meaning of "Accepting" or "Rejecting" a Hypothesis

If on the basis of a test of significance, say, the t test, we decide to "accept" the null hypothesis, all we are saying is that on the basis of the sample evidence we have no reason to reject it; we are not saying that the null hypothesis is true beyond any doubt. Why? To answer this, let us revert to our consumption–income example and assume that H_0: β_2 (MPC) = 0.50. Now the estimated value of the MPC is $\hat{\beta}_2 = 0.5091$ with a $\text{se}(\hat{\beta}_2) = 0.0357$. Then on the basis of the t test we find that $t = (0.5091 - 0.50)/0.0357 = 0.25$, which is insignificant, say, at $\alpha = 5\%$. Therefore, we say "accept" H_0. But now let us assume H_0: $\beta_2 = 0.48$. Applying the t test, we obtain $t = (0.5091 - 0.48)/0.0357 = 0.82$, which too is statistically insignificant. So now we say "accept" this H_0. Which of these two null hypotheses is the "truth"? We do not know. Therefore, in "accepting" a null hypothesis we should always be aware that another null hypothesis may be equally compatible with the data. It is therefore preferable to say that we *may* accept the null hypothesis rather than we (do) accept it. Better still,

> . . . just as a court pronounces a verdict as "not guilty" rather than "innocent," so the conclusion of a statistical test is "do not reject" rather than "accept."[11]

The "Zero" Null Hypothesis and the "2-*t*" Rule of Thumb

A null hypothesis that is commonly tested in empirical work is H_0: $\beta_2 = 0$, that is, the slope coefficient is zero. This "zero" null hypothesis is a kind of straw man, the objective being to find out whether Y is related at all to X, the explanatory variable. If there is no relationship between Y and X to begin with, then testing a hypothesis such as $\beta_2 = 0.3$ or any other value is meaningless.

This null hypothesis can be easily tested by the confidence interval or the t-test approach discussed in the preceding sections. But very often such formal testing can be shortcut by adopting the "2-t" rule of significance, which may be stated as

> **"2-*t*" Rule of Thumb.** If the number of degrees of freedom is 20 or more and if α, the level of significance, is set at 0.05, then the null hypothesis $\beta_2 = 0$ can be rejected if the t value $[= \hat{\beta}_2/\text{se}(\hat{\beta}_2)]$ computed from (5.3.2) exceeds 2 in absolute value.

The rationale for this rule is not too difficult to grasp. From (5.7.1) we know that we will reject H_0: $\beta_2 = 0$ if

$$t = \hat{\beta}_2/\text{se}(\hat{\beta}_2) > t_{\alpha/2} \qquad \text{when } \hat{\beta}_2 > 0$$

[11]Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, p. 114.

or

$$t = \hat{\beta}_2/\text{se}(\hat{\beta}_2) < -t_{\alpha/2} \qquad \text{when } \hat{\beta}_2 < 0$$

or when

$$|t| = \left|\frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)}\right| > t_{\alpha/2} \tag{5.8.1}$$

for the appropriate degrees of freedom.

Now if we examine the t table given in **Appendix D**, we see that for df of about 20 or more a computed t value in excess of 2 (in absolute terms), say, 2.1, is statistically significant at the 5 percent level, implying rejection of the null hypothesis. Therefore, if we find that for 20 or more df the computed t value is, say, 2.5 or 3, we do not even have to refer to the t table to assess the significance of the estimated slope coefficient. Of course, one can always refer to the t table to obtain the precise level of significance, and one should always do so when the df are fewer than, say, 20.

In passing, note that if we are testing the one-sided hypothesis $\beta_2 = 0$ versus $\beta_2 > 0$ or $\beta_2 < 0$, then we should reject the null hypothesis if

$$|t| = \left|\frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)}\right| > t_{\alpha} \tag{5.8.2}$$

If we fix α at 0.05, then from the t table we observe that for 20 or more df a t value in excess of 1.73 is statistically significant at the 5 percent level of significance (one-tail). Hence, whenever a t value exceeds, say, 1.8 (in absolute terms) and the df are 20 or more, one need not consult the t table for the statistical significance of the observed coefficient. Of course, if we choose α at 0.01 or any other level, we will have to decide on the appropriate t value as the benchmark value. But by now the reader should be able to do that.

Forming the Null and Alternative Hypotheses[12]

Given the null and the alternative hypotheses, testing them for statistical significance should no longer be a mystery. But how does one formulate these hypotheses? There are no hard-and-fast rules. Very often the phenomenon under study will suggest the nature of the null and alternative hypotheses. For example, consider the capital market line (CML) of portfolio theory, which postulates that $E_i = \beta_1 + \beta_2\sigma_i$, where E = expected return on portfolio and σ = the standard deviation of return, a measure of risk. Since return and risk are expected to be positively related—the higher the risk, the

[12]For an interesting discussion about formulating hypotheses, see J. Bradford De Long and Kevin Lang, "Are All Economic Hypotheses False?" *Journal of Political Economy*, vol. 100, no. 6, 1992, pp. 1257–1272.

higher the return—the natural alternative hypothesis to the null hypothesis that $\beta_2 = 0$ would be $\beta_2 > 0$. That is, one would not choose to consider values of β_2 less than zero.

But consider the case of the demand for money. As we shall show later, one of the important determinants of the demand for money is income. Prior studies of the money demand functions have shown that the income elasticity of demand for money (the percent change in the demand for money for a 1 percent change in income) has typically ranged between 0.7 and 1.3. Therefore, in a new study of demand for money, if one postulates that the income-elasticity coefficient β_2 is 1, the alternative hypothesis could be that $\beta_2 \neq 1$, a two-sided alternative hypothesis.

Thus, theoretical expectations or prior empirical work or both can be relied upon to formulate hypotheses. But no matter how the hypotheses are formed, *it is extremely important that the researcher establish these hypotheses before carrying out the empirical investigation*. Otherwise, he or she will be guilty of circular reasoning or self-fulfilling prophesies. That is, if one were to formulate hypotheses after examining the empirical results, there may be the temptation to form hypotheses that justify one's results. Such a practice should be avoided at all costs, at least for the sake of scientific objectivity. Keep in mind the Stigler quotation given at the beginning of this chapter!

Choosing α, the Level of Significance

It should be clear from the discussion so far that whether we reject or do not reject the null hypothesis depends critically on α, the level of significance or the *probability of committing a **Type I** error*—the probability of rejecting the true hypothesis. In **Appendix A** we discuss fully the nature of a Type I error, its relationship to a ***Type II** error* (the probability of accepting the false hypothesis) and why classical statistics generally concentrates on a Type I error. But even then, why is α commonly fixed at the 1, 5, or at the most 10 percent levels? As a matter of fact, there is nothing sacrosanct about these values; any other values will do just as well.

In an introductory book like this it is not possible to discuss in depth why one chooses the 1, 5, or 10 percent levels of significance, for that will take us into the field of statistical decision making, a discipline unto itself. A brief summary, however, can be offered. As we discuss in **Appendix A**, for a given sample size, if we try to reduce a *Type I error*, a *Type II error* increases, and vice versa. That is, given the sample size, if we try to reduce the probability of rejecting the true hypothesis, we at the same time increase the probability of accepting the false hypothesis. So there is a tradeoff involved between these two types of errors, given the sample size. Now the only way we can decide about the tradeoff is to find out the relative costs of the two types of errors. Then,

> If the error of rejecting the null hypothesis which is in fact true (Error Type I) is costly relative to the error of not rejecting the null hypothesis which is in fact

false (Error Type II), it will be rational to set the probability of the first kind of error low. If, on the other hand, the cost of making Error Type I is low relative to the cost of making Error Type II, it will pay to make the probability of the first kind of error high (thus making the probability of the second type of error low).[13]

Of course, the rub is that we rarely know the costs of making the two types of errors. Thus, applied econometricians generally follow the practice of setting the value of α at a 1 or a 5 or at most a 10 percent level and choose a test statistic that would make the probability of committing a Type II error as small as possible. Since one minus the probability of committing a Type II error is known as the **power of the test,** this procedure amounts to maximizing the power of the test. (See **Appendix A** for a discussion of the power of a test.)

But all this problem with choosing the appropriate value of α can be avoided if we use what is known as the ***p* value** of the test statistic, which is discussed next.

The Exact Level of Significance: The *p* Value

As just noted, the Achilles heel of the classical approach to hypothesis testing is its arbitrariness in selecting α. Once a test statistic (e.g., the t statistic) is obtained in a given example, why not simply go to the appropriate statistical table and find out the actual probability of obtaining a value of the test statistic as much as or greater than that obtained in the example? This probability is called the ***p* value** (i.e., **probability value**), also known as the **observed or exact level of significance** or the **exact probability of committing a Type I error.** More technically, the p value is defined as **the lowest significance level at which a null hypothesis can be rejected.**

To illustrate, let us return to our consumption–income example. Given the null hypothesis that the true MPC is 0.3, we obtained a t value of 5.86 in (5.7.4). What is the p value of obtaining a t value of as much as or greater than 5.86? Looking up the t table given in **Appendix D**, we observe that for 8 df the probability of obtaining such a t value must be much smaller than 0.001 (one-tail) or 0.002 (two-tail). By using the computer, it can be shown that the probability of obtaining a t value of 5.86 or greater (for 8 df) is about 0.000189.[14] This is the p value of the observed t statistic. This observed, or exact, level of significance of the t statistic is much smaller than the conventionally, and arbitrarily, fixed level of significance, such as 1, 5, or 10 percent. As a matter of fact, if we were to use the p value just computed,

[13] Jan Kmenta, *Elements of Econometrics*, Macmillan, New York, 1971, pp. 126–127.

[14] One can obtain the p value using electronic statistical tables to several decimal places. Unfortunately, the conventional statistical tables, for lack of space, cannot be that refined. Most statistical packages now routinely print out the p values.

and reject the null hypothesis that the true MPC is 0.3, the probability of our committing a Type I error is only about 0.02 percent, that is, only about 2 in 10,000!

As we noted earlier, if the data do not support the null hypothesis, $|t|$ obtained under the null hypothesis will be "large" and therefore the p value of obtaining such a $|t|$ value will be "small." In other words, for a given sample size, as $|t|$ increases, the p value decreases, and one can therefore reject the null hypothesis with increasing confidence.

What is the relationship of the p value to the level of significance α? If we make the habit of fixing α equal to the p value of a test statistic (e.g., the t statistic), then there is no conflict between the two values. To put it differently, **it is better to give up fixing α arbitrarily at some level and simply choose the p value of the test statistic.** It is preferable to leave it to the reader to decide whether to reject the null hypothesis at the given p value. If in an application the p value of a test statistic happens to be, say, 0.145, or 14.5 percent, and if the reader wants to reject the null hypothesis at this (exact) level of significance, so be it. Nothing is wrong with taking a chance of being wrong 14.5 percent of the time if you reject the true null hypothesis. Similarly, as in our consumption–income example, there is nothing wrong if the researcher wants to choose a p value of about 0.02 percent and not take a chance of being wrong more than 2 out of 10,000 times. After all, some investigators may be risk-lovers and some risk-averters!

In the rest of this text, we will generally quote the p value of a given test statistic. Some readers may want to fix α at some level and reject the null hypothesis if the p value is less than α. That is their choice.

Statistical Significance versus Practical Significance

Let us revert to our consumption–income example and now hypothesize that the true MPC is 0.61 (H_0: $\beta_2 = 0.61$). On the basis of our sample result of $\hat{\beta}_2 = 0.5091$, we obtained the interval (0.4268, 0.5914) with 95 percent confidence. Since this interval does not include 0.61, we can, with 95 percent confidence, say that our estimate is statistically significant, that is, significantly different from 0.61.

But what is the practical or substantive significance of our finding? That is, what difference does it make if we take the MPC to be 0.61 rather than 0.5091? Is the 0.1009 difference between the two MPCs that important practically?

The answer to this question depends on what we really do with these estimates. For example, from macroeconomics we know that the income multiplier is 1/(1 − MPC). Thus, if MPC is 0.5091, the multiplier is 2.04, but it is 2.56 if MPC is equal to 0.61. That is, if the government were to increase its expenditure by $1 to lift the economy out of a recession, income will eventually increase by $2.04 if the MPC is 0.5091 but by $2.56 if the MPC is 0.61. And that difference could very well be crucial to resuscitating the economy.

The point of all this discussion is that one should not confuse statistical significance with practical, or economic, significance. As Goldberger notes:

> When a null, say, $\beta_j = 1$, is specified, the likely intent is that β_j is *close* to 1, so close that for all practical purposes it may be treated *as if it were* 1. But whether 1.1 is "practically the same as" 1.0 is a matter of economics, not of statistics. One cannot resolve the matter by relying on a hypothesis test, because the test statistic $[t =](b_j - 1)/\hat{\sigma}_{bj}$ measures the estimated coefficient in standard error units, which are not meaningful units in which to measure the economic parameter $\beta_j - 1$. It may be a good idea to reserve the term "significance" for the statistical concept, adopting "substantial" for the economic concept.[15]

The point made by Goldberger is important. As sample size becomes very large, issues of statistical significance become much less important but issues of economic significance become critical. Indeed, since with very large samples almost any null hypothesis will be rejected, there may be studies in which the magnitude of the point estimates may be the only issue.

The Choice between Confidence-Interval and Test-of-Significance Approaches to Hypothesis Testing

In most applied economic analyses, the null hypothesis is set up as a straw man and the objective of the empirical work is to knock it down, that is, reject the null hypothesis. Thus, in our consumption–income example, the null hypothesis that the MPC $\beta_2 = 0$ is patently absurd, but we often use it to dramatize the empirical results. Apparently editors of reputed journals do not find it exciting to publish an empirical piece that does not reject the null hypothesis. Somehow the finding that the MPC is statistically different from zero is more newsworthy than the finding that it is equal to, say, 0.7!

Thus, J. Bradford De Long and Kevin Lang argue that it is better for economists

> . . . to concentrate on the magnitudes of coefficients and to report confidence levels and not significance tests. If all or almost all null hypotheses are false, there is little point in concentrating on whether or not an estimate is indistinguishable from its predicted value under the null. Instead, we wish to cast light on what models are good approximations, which requires that we know ranges of parameter values that are excluded by empirical estimates.[16]

In short, these authors prefer the confidence-interval approach to the test-of-significance approach. The reader may want to keep this advice in mind.[17]

---

[15]Arthur S. Goldberger, *A Course in Econometrics,* Harvard University Press, Cambridge, Massachusetts, 1991, p. 240. Note b_j is the OLS estimator of β_j and $\hat{\sigma}_{bj}$ is its standard error. For a corroborating view, see D. N. McCloskey, "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests," *American Economic Review,* vol. 75, 1985, pp. 201–205. See also D. N. McCloskey and S. T. Ziliak, "The Standard Error of Regression," *Journal of Economic Literature*, vol. 37, 1996, pp. 97–114.

[16]See their article cited in footnote 12, p. 1271.

[17]For a somewhat different perspective, see Carter Hill, William Griffiths, and George Judge, *Undergraduate Econometrics,* Wiley & Sons, New York, 2001, p. 108.