

4. Hypothesis Tests about the Mean of a Normal Population When σ^2 is Known

Hypothesis testing procedures compare a conjecture we have about a population to the information contained in a sample of data. More specifically, the conjectures we test here concern the mean of a normal population. In the context of the problem faced by the airplane seat designer, suppose that airplanes since 1970 have been designed assuming the mean population hip width is 17 inches. Is that figure still valid in the year 2000? As with point and interval estimation procedures, hypothesis testing procedures are established based on our assumptions. In this section we assume that we have a sample of data from a normal population with unknown mean β but a known variance σ^2 . This is not the usual case in practice, but for pedagogic reasons it is a good place to start. The procedures we develop can be applied to any sample of data that is obtained. Because the procedures are general, they depend on random quantities. Thus a test's characteristics are stated in terms of probabilities.

Hypothesis testing uses the information about a parameter that is contained *in a sample of data*, namely its least squares point estimate and its standard error, to draw a conclusion about the conjecture, or hypothesis. In each and every hypothesis test four 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

The Null Hypothesis

The “null” hypothesis, which is denoted H_0 (*H-naught*), specifies a value c for a parameter. We specify the null hypothesis as $H_0: \beta = c$. 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.

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 the problem

at hand. For the null hypothesis $H_0: \beta = c$, where c is some constant, three possible alternative hypotheses are:

- $H_1: \beta \neq c$. Rejecting the null hypothesis that $\beta = c$ implies the conclusion that β takes some other value.
- $H_1: \beta > c$. Rejecting the null hypothesis that $\beta = c$ leads to the conclusion that β is greater than c . Using this alternative *completely* discounts the possibility that $\beta < c$, based on economic theory, and implies that these values are logically unacceptable alternatives to the null hypothesis. Inequality alternative hypotheses are widely used in economics.
- $H_1: \beta < c$. Following the discussion above, use this alternative when the rejecting the null hypothesis that $\beta = c$ leads to the conclusion that β is less than c . Use this alternative when there is no chance that $\beta > c$

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, which itself is a random variable, we decide either to reject the null hypothesis or not to reject it. A test statistic has a very special characteristic: its probability distribution must be (*completely*) *known when the null hypothesis is true*, and it must have some *other* distribution if the null hypothesis is not true.

Consider the null hypothesis $H_0: \beta = c$. We will introduce hypothesis testing concepts using the alternative $H_1: \beta \neq c$. If the sample data come from a normal population with unknown mean β but a known variance σ^2 , then

$$Z = \frac{b - \beta}{\sqrt{\sigma^2/T}} = \frac{b - \beta}{\sigma/\sqrt{T}} \sim N(0,1) \quad (4.1)$$

If the null hypothesis $H_0: \beta = c$ is ***true***, then

$$Z = \frac{b - c}{\sigma/\sqrt{T}} \sim N(0,1) \quad (4.2)$$

If the null hypothesis is ***not true***, then the Z-statistic in (4.2) does ***not*** have a standard normal distribution.

The Rejection Region

The *rejection region* 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. In practice the rejection region is a set of test statistic values that, when

the null hypothesis is true, are *unlikely* and have *low probability* of occurring. If, using a sample of data, 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.

In Figure 1, the standard normal $N(0,1)$ is depicted. The critical value z_c is such that $P[Z \geq z_c] = P[Z \leq -z_c] = \alpha/2$. The values of z_c that correspond to different values of α can be found in Table 1 at the back of *UE/2*. If $\alpha = .05$ then $z_c = 1.96$; if $\alpha = .01$ then $z_c = 2.58$. The random variable Z in (4.2) can be used as a *test statistic* since its distribution is known to be $N(0,1)$ when the null hypothesis is true. If b is the least squares estimate of β based on a sample of data, then use it to compute the *value* z of the test statistic Z .

If the null hypothesis $H_0: \beta = c$ is *true* then the test statistic (4.2) has the distribution $N(0,1)$. Thus, if the hypothesis is true, then the distribution of Z is that shown in Figure 1. If the alternative hypothesis $H_1: \beta \neq c$ is true, then values of the test statistic Z 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 frequently chosen to be .01, .05 or .10. The rejection region is determined by the critical values z_c . Thus the rejection region consists of the two “tails” of the $N(0,1)$ -distribution.

When the null hypothesis is true the probability of obtaining from a sample of data a value of the test statistic that falls in *either* tail area is “small,” and, combined, is equal to α . Sample values of the test statistic which are in the tail areas are *incompatible with the null hypothesis* and are evidence *against* the null hypothesis being true. When testing the null hypothesis $H_0: \beta = c$ against the alternative $H_1: \beta \neq c$ we are led to the following rule:

Rejection rule for a two-tailed test (σ^2 known): If the value of the test statistic Z falls in the rejection region, either tail of the $N(0,1)$ -distribution, then we reject the null hypothesis and accept the alternative.

On the other hand, if the null hypothesis $H_0: \beta = c$ is true, then the probability of obtaining a value of the test statistic Z in the central *non-rejection* region, $P(-z_c \leq Z \leq z_c) = 1 - \alpha$, is high. Sample values of the test statistic in the central non-rejection area are *compatible with the null hypothesis* and are not taken as evidence *against* the null hypothesis being true. Care must be taken here in interpreting the outcome of a statistical test, however, because one of the basic precepts of hypothesis testing is that *finding a sample value of the test statistic in the non-rejection region does not make the null hypothesis true!* Intuitively, if the true value of β is *near* c (but not equal to it), then the value of the test statistic will still fall in the non-

rejection region with high probability. In this case we would not reject the null hypothesis even though it is false. Consequently, when testing the null hypothesis $H_0: \beta = c$ against the alternative $H_1: \beta \neq c$, the decision rule is:

If the value of the test statistic Z falls between the critical values $-z_c$ and z_c , in the non-rejection region, then we **do not reject** the null hypothesis.

Avoid saying that “we accept the null hypothesis.” This statement implies that we are concluding that the null hypothesis is true, which is not the case at all based on the discussion above. The weaker statements “we do not reject the null hypothesis,” or “we fail to reject the null hypothesis,” do not send any misleading message.

A standard testing format that summarizes the four test ingredients and the test outcome is:

Format 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.

(4.1) Type I and Type II Errors

Whenever we reject, or do not reject, a null hypothesis there is a chance that we may be making a mistake. This is unavoidable. In any hypothesis testing situation there are two ways that we can make a correct decision and two ways that we can make an incorrect decision. We make a correct decision if:

- The null hypothesis is *false* and we decide to *reject* it.
- The null hypothesis is *true* and we decide *not* to reject it.

Our decision is incorrect if:

- The null hypothesis is *true* and we decide to *reject* it (a Type I error)

- The null hypothesis is *false* and we decide *not* to reject it (a Type II error)

When we reject the null hypothesis we risk what is called a Type I error. The probability of a Type I error is α , the *level of significance* of the test. A value of the test statistic in the rejection region, the range of unlikely values for the test statistic where we reject the null hypothesis, occurs with probability α when the null hypothesis is true. Thus the hypothesis testing procedure we use will *reject* a true hypothesis with probability α . The only good news here is that we can control the probability of a Type I error for the testing procedure by choosing the level of significance of the test. If this type of decision error is a *costly* one then we should choose the level of significance to be small, perhaps $\alpha = .01$ or $.05$.

We risk a Type II error when we do not reject the null hypothesis. Our testing procedure will lead us to *fail to reject* null hypotheses that are *false* with a certain probability. The magnitude of the probability of a Type II error is *not* under our control and *cannot* be computed as it depends on the true but unknown value of the parameter in question. The facts that we can tell you about the probability of a Type II error are these:

- The probability of a Type II error varies inversely with the level of significance of the test, α , which is the probability of a Type I error. If you choose to make α smaller, the probability of a Type II error increases.
- The closer the true value of the parameter is to the hypothesized parameter value the larger is the probability of a Type II error. If in the null hypothesis we hypothesize that $\beta = c$, and if the true (unknown) value of β is *close* to c , then the probability of a Type II error is high. Intuitively, the test loses the *power* to discriminate between the true parameter value and the (false) hypothesized value if they are similar in magnitude.
- The larger the sample size T , the lower the probability of a Type II error, given a level of Type I error α .
- For most types of hypotheses that economists test there is no one best test that can be used in all situations. By “best” test we mean one that has the minimum Type II error for any given level of Type I error, α . However, the test based on the t -distribution that we have described is a very good test and it is without question the test used most frequently in the situation that we have described.

(4.2) The p -Value of a Hypothesis Test

When reporting the outcome of statistical hypothesis tests it has become common practice to report the p -value of a test. In the context of the food expenditure example, when testing the null hypothesis $H_0:\beta = c$ against the alternative hypothesis $H_1:\beta \neq c$, we reject the null hypothesis when the absolute value of the Z -statistic is greater than or equal to the critical value z_c that corresponds to the level of significance we have chosen. That is, we reject the null hypothesis when $|z| \geq z_c$. The p -value of a test is calculated by finding the probability that the Z -distribution can take a value greater than or equal to the absolute value of the *sample value of the test statistic*.

To illustrate, suppose we were testing $H_0:\beta = c$ against the alternative hypothesis $H_1:\beta \neq c$. Using a sample of data we obtain the value

$$Z = \frac{b - c}{\sigma/\sqrt{T}} = 2.16$$

The p -value of this test is depicted in Figure 3.

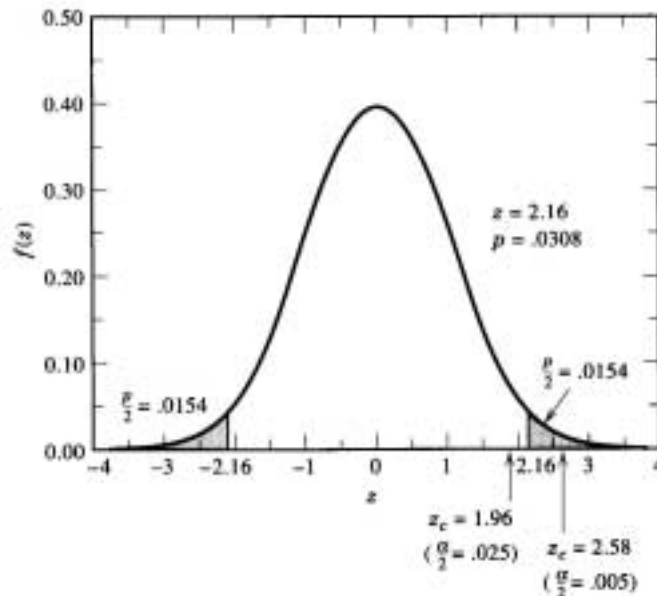


Figure 3. The p -value of a Z test statistic value $z = 2.16$

To calculate the p -value of $z = 2.16$ we proceed as follows:

$$\begin{aligned} p\text{-value}(z = 2.16) &= \Pr[Z \leq -2.16 \text{ or } Z \geq 2.16] \\ &= 2\Pr[Z \geq 2.16] \\ &= 2[\Pr(Z < 2.16)] \\ &= 2[1 - .9846] \\ &= 2[.0154] = .0308 \end{aligned}$$

The calculation of $\Pr[Z < 2.16]$ can be carried out in two ways. By using Table 1 at the back of *UE/2* we find that

$$\Pr[0 \leq Z \leq 2.16] = .4846$$

So

$$\Pr[Z < 2.16] = .5 + .4846 = .9846$$

Alternatively, your computer software will have a function that can compute this probability.

Using a p -value we can determine whether to reject a null hypothesis by comparing it to the level of significance α . The rule is:

Rejection rule for a two-tailed test: When the p -value of a hypothesis test is *smaller* than the chosen value of α , then the test procedure leads to rejection of the null hypothesis.

This rule is very handy. If we have the p -value of a test, 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 ourselves.

The reasoning of the rule is made clear by examining Figure 3. If the p -value is less than or equal to α , then $|z| \geq z_c$ and the null hypothesis is rejected. Thus by knowing that the p -value of $z = 2.16$ is $p = .0308$ we can determine immediately that the null hypothesis is rejected at the $\alpha = .01$ level, but it will not be rejected at the $\alpha = .01$ level of significance.

(4.3) A Monte Carlo Experiment

To illustrate hypothesis testing outcomes, we use the Monte Carlo simulation technique introduced in Section 2.1. There we assume data come from a normal population with mean $\beta = 10$ and variance $\sigma^2 = 10$. In this section we test the null hypothesis $H_0: \beta = 10$ against the alternative hypothesis $H_1: \beta \neq 10$

using the test statistic Z in (4.2). Note that in this artificial data example the hypothesis we are testing in this simulation is *true*. In Table 4 are estimates of β and the test statistics from 10 samples of data.

Table 4 Estimates and Test Statistics from 10 Samples

Sample	b	z
1	10.206	0.357
2	9.828	-0.298
3	11.194	2.068
4	8.822	-2.041
5	10.434	0.752
6	8.855	-1.984
7	10.511	0.886
8	9.212	-1.365
9	10.464	0.804
10	10.142	0.245

Notice the variation in the values of b , the least squares estimate of β , and in the values of the test statistic z . The variation we observe is *sampling variation*, caused by drawing different samples from the population. If we test at the $\alpha = .05$ level of significance the critical value from the $N(0,1)$ distribution is $z_c = 1.96$. Despite the fact that the hypothesis we are testing is true, in samples 3, 4 and 6 we reject the null hypothesis $H_0:\beta = 10$ and accept the alternative that $H_1:\beta \neq 10$. Recall that α is the probability of a Type I error, in which we reject the null hypothesis when it is true. In this case, we would make a Type I error in 3 of 10 samples, which is higher than we might expect. In 10,000 samples the percentage of Type I errors is .0514. These results remind you that the probability of a Type I error is a *repeated sampling property* of the test procedure, which holds if a large number of samples are drawn.