

CHAPTER SIX

Inference Statistics

6.1 Hypothesis Testing

Hypothesis testing is a procedure that examines two alternative positions in which a test or an experiment is made to determine which of the positions may be true within certain level of confidence. The position that states that things are the same is called the **null hypothesis** (H_0); the position that states that things are not the same is called the **alternative hypothesis** (H_a). In accepting the alternative hypothesis one rejects the null hypothesis and in accepting the alternative hypothesis, one rejects the null hypothesis. More formal definitions of these concepts are shown below. Hypothesis testing always uses sample data to test or make inference about a population parameter. The researcher's hypothesis is often the alternative hypothesis; therefore, the null hypothesis is the opposite of the researcher's hypothesis.

The null hypothesis is a statement asserting no change or difference about a population parameter; it is denoted by the symbol, H_0 .

The alternative hypothesis is a statement that rejects the null hypothesis or a statement that might be true if the null hypothesis is not; it is denoted by the symbol, H_a . This is the hypothesis the research is designed to investigate.

The alternative hypothesis may contain the symbol, $>$, $<$, \neq

If one wishes to test that the sample mean equals a hypothesized or known population mean given normal errors of experiments, then one would, for example, compare the value of the sample mean with that of the population mean and if the difference is small, then one would accept that both are the same or that the null hypothesis is not rejected (almost saying that it **is true**) or $M_1 = M_2$. Knowing how much the difference in comparing both means will be in order to reject the null hypothesis (that both means are not the same) is a matter of probability or how confident or certain you want to be in making that judgment.

In hypothesis testing we present this view of deciding on a level of acceptance of the null hypothesis: If there is a low probability that we can reject the null hypothesis, then we say that the null hypothesis “is true.” This low probability of rejecting the null hypothesis is called the **significant level**, and is called alpha or denoted by the symbol, α or α .

Note statisticians never say that we accept the null hypothesis, but rather that there is no significant reason to reject it.

Acceptance Criterion

The goal of statistical hypothesis testing for this chapter and many others using a distribution for the test, say a normal distribution, is to accept the null hypothesis if the statistical estimator falls within the acceptable ***middle region*** of the confidence interval based on the level of significance or probability of acceptance or to reject if its falls in the ***tail***. Therefore, if the hypothesized population parameter falls within an estimated confidence interval based upon sample statistics, we do not reject the null hypothesis.

Figure 6.1.1 shows both the acceptance region and rejection regions for the population parameter for a sample that is normally distributed. The acceptance region of the null hypothesis is the region between the sample's lower and upper limits of the confidence interval. If the *middle* region is the acceptable region then the *tail* region is the rejected regions.

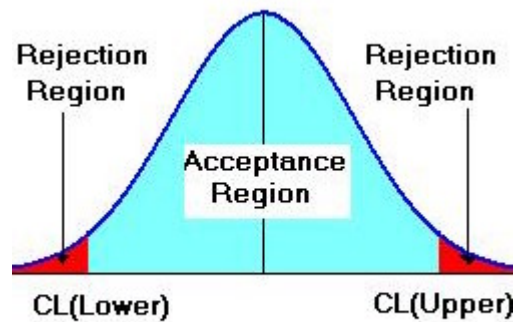


Figure 6.1.1 Acceptance and rejection regions of the normal curve

The **test statistics** is the value or quantity that is used to make a decision in testing the hypothesis.

For every possible types of statistical tests there are preferred test statistics used to make judgment about the null hypothesis, in the next several chapters we will study many of these. The **confidence interval (CI)** is determined from the test statistics, the critical value of the underlying test distribution, and the standard error of the sampling distribution. The critical value is based on the level of significance of the test in addition to the underlying distribution of the test.

$$\text{Confidence Interval} = \text{Test Statistics} \pm (\text{Critical Value})(\text{Standard Error})$$

Level of Significance

To find the lower and upper confidence interval limits that separate the acceptance region from the rejection regions of a sample distribution, we must define a few terms. Recall, we stated earlier that the low probability of rejecting the null hypothesis is called the **significant level**, and is called alpha or denoted by the symbol, α or a . This alpha value is a small probability that is used to identify the rare chance that sample data (often the mean or difference of means) would unlikely occur. Commonly used alpha levels are $a = 0.05$ (5%) and $a = 0.01$ (1%). For test that requires more critical assurance of evaluating the null hypothesis, we may even use a smaller alpha level of 0.001 (0.1%). By selecting an alpha level of 0.05, we identify the limits or extreme values that the most unlikely 5% of the sample data would fall outside of and the most likely 95% that the sample data will fall within (the 95% confidence interval).

The unlikely values, defined by the alpha level, make up the *critical region* of the test. The likely values that we consider typical or normal of a sample distribution make up the acceptance region of the test. Figure 6.1.2 shows a sample distribution with critical regions for rejecting the null hypothesis and the acceptance region if the null hypothesis is true. Figure 6.1.2 also shows the probability in percent for each region.

For an alpha of 0.05, we get an acceptance or middle region of 95% and for a standard normal distribution the critical region of 5% is split into two halves of 2.5% each. Therefore, for this situation, the acceptable region for a test or the *middle* region will give a confidence interval of $-1.96 = z = 1.96$ (Figure 6.1.4). A value of $z = 2.0$ would fall outside this region so the null hypothesis is rejected; however a value of $z = 1.14$ falls within this region so there is no reason to reject the null hypothesis.

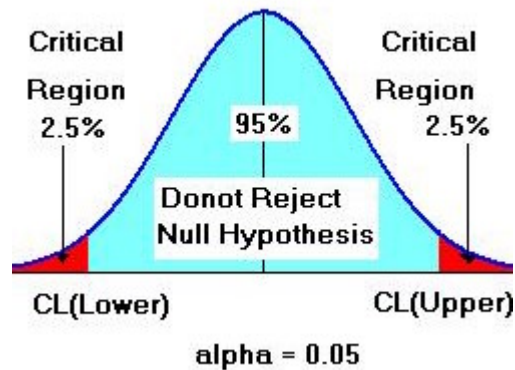


Figure 6.1.2 Critical and Acceptance Regions for Alpha 0.05

Throughout this text we will use the phrase “*there is no reason to reject the null hypothesis*” when “accepting” the null hypothesis. The **critical region** is the region outside the confidence interval for z that favors the alternative hypothesis - The *tail* region. The values of z at the endpoints of the critical region are called the **critical values**. In the last example, the **critical values** are $z = \pm 1.96$. A confidence interval is constructed to test many hypotheses at once; the width of the CI is dependent upon the critical value and standard error, which is primarily dependent upon the sample size.

The **critical region** is the values of the test statistics that provides evidence in favor of the alternative hypothesis. Therefore, a value in the critical region results in a decision to reject the null hypothesis.

Alpha or α is the **level of significance** of the hypothesis test and it is the probability that the test statistics will fall in the critical region or the red area if the null hypothesis is true.

If the underlying distribution for testing the null hypothesis is normal, the sample mean and standard deviation can be used to compute the z -score.

$$z = \frac{X - \mu}{\sigma}, \text{ where } \mu \text{ population mean and } \sigma \text{ the population standard deviation}$$

The **critical value** for the standard normal distribution would be the value of z at or beyond which we reject the null hypothesis, H_0 ; for $\alpha = 0.05$, the critical value in the example above is -1.96 and 1.96 and the critical region would be $z = -1.96$ (lower confidence limit) and $z = 1.96$ (upper confidence limit).

The **critical value** is the value of a test statistics at or beyond which we will reject the null hypothesis, H_0 . It is the boundary of the critical region.

One- and Two-Tailed Tests

The location of the critical region is dependent upon the directional statement of the alternative hypothesis. In comparing the sample mean, M to the population mean, μ the null hypothesis is

$$H_0: M = \mu \quad (\text{the sample mean equals the population mean});$$

however, the alternative hypothesis may be stated in one of three ways:

$$H_a: M \neq \mu \quad (\text{statement 1: non-directional})$$

or

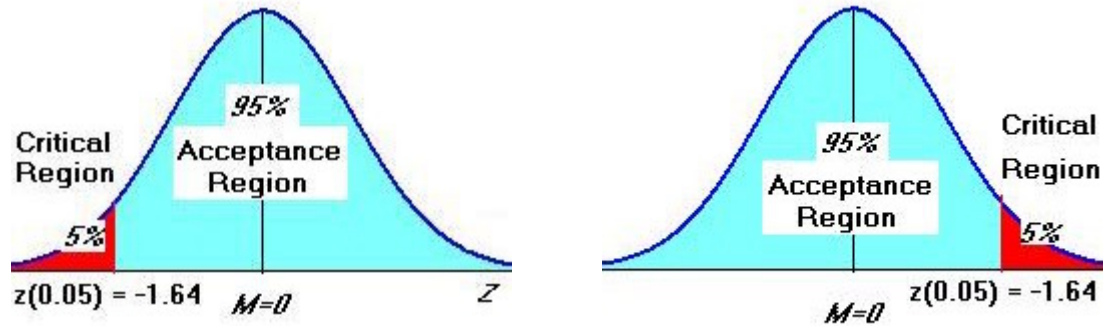
$$H_a: M < \mu \quad (\text{statement 2: directional})$$

or

$$H_a: M > \mu \quad (\text{statement 3: directional}).$$

Using an alpha, $\alpha = 0.05$, Figure 6.1.3 and Figure 6.1.4 show the critical regions that would result from the three possible alternative hypothesis statements. Figure 6.1.3a shows the critical region for statement 2 (the sample mean, M is less than the population mean). The lowest value for accepting the null hypothesis is $z(0.05)$ or $z_{0.05} = -1.64$ (the 5 percentile of the z distribution), so we call this test a **lower one-tailed test**. Figure 6.1.3b shows the critical region for statement 3 (the sample mean is greater than the population

mean). The highest value for accepting the null hypothesis is $z(0.95)$ or $z_{0.95} = 1.64$ (the 95 percentile of the z distribution), so we call this test an **upper one-tailed test**.



(a) Lower One-Tailed Test

(b) Upper One-Tailed Test

Figure 6.1.3 Directional: One-Tailed Hypothesis Testing

Figure 6.1.4 shows the critical regions for statement 1 (the sample mean is not the same as the population mean). Notice, there is no directional implication of the sample mean relative to the population mean in this statement. We would accept the null hypothesis if the population mean (μ) is within $[M - 1.96(S^*)] = \mu = [M + 1.96(S^*)]$.

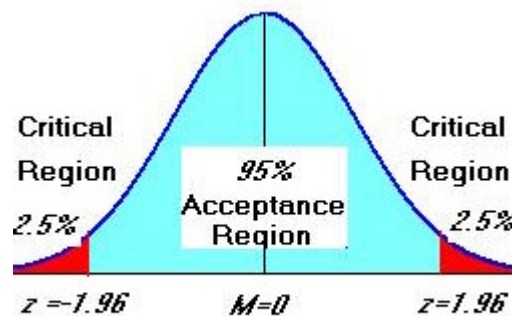


Figure 6.1.4 Non-directional: Two-Tailed Hypothesis Testing

Type I and Type II Errors

Type I Error

Even though it is unlikely that the test statistics will fall into the critical region (*tail*) when the null hypothesis is true, it is still possible. When this occurs, we reject H_0 , when indeed it is true, and therefore make an error in doing so. This type of error is also called the **consumer's risk** and is the most serious error for the producer because it may cause rejection when the hypothesis is true.

A **Type I error** is an error in rejecting the null hypothesis when it is true, and this happens if the test statistics falls inside the critical region (*tail*).

The probability (Pr) of rejecting the H_0 when it is true is called α , where

$$\alpha = \text{Pr}[\text{Type I error}]$$

Type II error

Another type of error is to not reject the null hypothesis when it is false. This is called a type II error. It is the probability that not rejecting the null hypothesis when it is indeed false. This happens when the test statistics does not fall in the critical region when H_0 is false. This type of error in hypothesis testing is called the **producer's risk** and is the most serious error for the consumer because it may cause acceptance when the hypothesis is false.

A **Type II error** is an error in accepting the null hypothesis when it is false, and this happens if the test statistics falls inside the acceptable region (*middle*) when it should have fallen in the *tail* region or critical region.

The probability of accepting the H_0 when it is false is called β , where

$$\beta = \text{Pr}[\text{Type II error}]$$

The **power** of a statistical test is the probability that it will correctly reject a false null hypothesis. That is, the power is the probability that the test will identify a treatment effect if one really exists. The concept of effect size is introduced here as one measure of the power of a statistical test.

The test statistics used to evaluate statistical tests is often inversely related to the sampling error which has one of its components the sample size. What this means is that larger sample size, n has a greater chance of making a statistical test significant than smaller sample size. Because as the sample size gets larger, the chance of rejecting the null hypothesis increases, we need to find tests to measure the significance of the hypothesis that is not dependent upon sample size.

Typical hypothesis testing involving test statistics based on sampling error, has a few limitations, First, it focus more on the sample information rather than the hypothesis itself. This limitation does not make any strong argument about the probability of the hypothesis itself being true or not; just that a specific sample statistics is very unlikely or not based upon a specified alpha level. Second, finding a significant treatment effect does not indicate a substantial treatment effect. It only addresses the supposition that the differences due to the treatment effect are very small relative to what is expected by chance if the null hypothesis is true; however, the measurement of the effect due to chance (standard error) is influenced by the sample size. Also it evaluates the relative size of treatment rather than absolute size. Cohen (1988) proposed a useful and practical means to measure the power or effect of the hypothesis testing by the following formula:

$$\text{Cohen's } d = \text{Effect Size} = \text{ES} = \frac{\text{mean difference}}{\text{standard deviation}}$$

For example a person after treatment for eating disorder may experience a weight change of 10 pounds, if the standard deviations for the population is 10 pounds, then the effect size would be $ES = 10/10 = 1$. This means that the treatment was able to change the person's weight by one standard deviation; from a practical prospective, this is a large change. Table 6.1.1 shows how the evaluate Cohen's d .

Table 6.1.1 *Evaluating Effect Size with Cohen's d .*

Cohen's d	Magnitude of Effect Size
$0.2 <$	Small
$0.2 \text{ to } 0.8$	Medium
> 0.8	Large

The Null Hypothesis and Other Hypotheses:

The null hypothesis tests that $H_0: mean_1 = mean_2$ or $(mean_1 - mean_2) = 0$. An important point in hypothesis testing is that *the rejection of the null hypothesis does not necessarily make an experimental hypothesis true*. The test will tell you whether or not there is a significant difference between the two groups being evaluated, but it does not prove why a difference exists.

We don't always assume that a null hypothesis $H_0: mean_1 - mean_2 = 0$.

Occasionally, we may wish to test our obtained difference, D , against some other null hypothesis. We may wish to test that $H_0: mean_1 - mean_2 = E$, or some value, so our test statistics would be (where S_D is the standard error of the difference):

$$\text{test statistics} = \frac{(M_2 - M_1) - E}{S_D}$$

Basic Steps in Hypothesis Testing

Even though the formula for a test statistics will depend upon many factors, the basic steps involve are simple and few.

Step 1. Identify the null and alternative hypothesis.

The null hypothesis often contains the = sign and
The alternative hypothesis contains >, < and \neq sign.

Example: $H_0: m_1 = m_2$ and $H_a: m_1 > m_2$,

Step 2. Choose the level of significance of the test, α (0.05 or 0.01).

Since α is the **Type I error**, the probability of rejecting the null hypothesis when it is true, the smaller the size of the α value the more critical the test. The probability of acceptance of confidence interval is $1 - \alpha$

Step 3. Select the test statistics.

This is often based on sampling data or estimates about the population parameters with some **standard error of the estimate**. This is compared against an expected or reference value often looked up from some probability distribution table.

Step 4. Determine the critical region.

The critical region is those values of the test statistics that strongly favor the alternative hypothesis.

Often it is a good practice to sketch the critical region (red). (Figures 6.1.3 and 6.1.4)

Step 5. Make your decision.

If the test statistics falls into the critical region, reject H_0 .

When this occurs we say that the results are statistically significant.

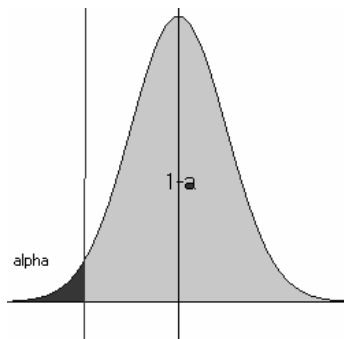
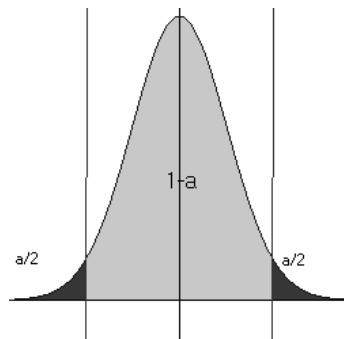
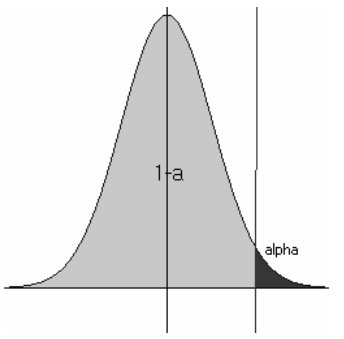
Summary

In inference statistics we will use sample estimates of population parameters to test hypotheses about population parameters or conditions or make comparisons about

sample statistics or distributions. We will define both null and alternative hypotheses given some level of confidence or significance about our sample. Then we will calculate a sample test statistics that will be compared against a critical value or region of the sampling distribution and then make a decision whether our null hypothesis is true (likely to occur). If the null hypothesis is unlikely, we will reject it and favor the alternative hypothesis (the hypothesis the research is trying to prove). Finally, we will compute some measure of effect size to test the power of our hypothesis testing.

Hypothesis Testing Worksheet

Type of Test (e.g. Two-sample t -test.): _____

Enter sample(s) and/or population parameters: Others: Sample size, $N_1 =$ _____ if appropriate, $N_2 =$ _____ $s_1 =$ _____ or s or $s_2 =$ _____ \bar{x} or $M_1 =$ _____ or μ or $M_2 =$ _____		
(1) State Null Hypothesis: H_0	State Alternative Hypothesis: H_a	
(2) Choose level of Significance, $\alpha =$ _____ Write $1 - \alpha =$ _____	Enter degrees of freedom, $d.f.$	
(3) Write and Compute the Test Statistics (e.g. t-test): 		
(4) Select Critical Region of Test Criterion: table-value		
Lower-Tailed Test 	Two-Tailed Test 	Upper-Tailed Test 
(5) Make Decision:		