

# CHAPTER 5

## STATISTICAL INFERENCE

---

- 1.0 Hypothesis Testing
- 2.0 Decision Errors
- 3.0 How a Hypothesis is Tested
- 4.0 Test for Goodness of Fit
- 5.0 Inferences about Two Means

**“It ain't easy being green.”**

**Kermit**

# STATISTICAL INFERENCE

## 1.0 HYPOTHESIS TESTING

A hypothesis is a statement, inference or tentative explanation about a population that can be tested by further investigation. A hypothesis test is a statistical test that assists in the decision to prove or disprove the statement. In most cases, it may be easier to disprove a hypothesis than to verify it. Sample measurements are taken and inferences made about the total population based on the sample statistics. The inferences and the resulting decisions could be correct or incorrect. In hypothesis testing as well as all other statistical analyses, the risks of making a wrong decision are predetermined.

### 1.1 Null and Alternate Hypotheses

The symbol  $\mu$  represents the population or universe mean and  $\bar{x}$  represents the sample mean. The symbol  $\sigma$  represents the population standard deviation and  $\sigma'$  or  $s$  represent the standard deviation calculated from the sample. The population values  $\mu$  and  $\sigma$  are parameters and the sample values  $\bar{x}$  and  $\sigma'$  or  $s$  are statistics.

The following hypothesis is to be tested: The number 5 wheel bearings made by the Canape Bearing Company have an average diameter of 0.65 inches.

Based on the sample statistics, is this statement true or false? The objective of hypothesis testing is to verify the validity of the statement.

The null hypothesis is the hypothesis to be proven or disproven. It is represented by  $H_0$ .

The alternate hypothesis is the hypothesis that is accepted if the null hypothesis is rejected by the test. It is represented by  $H_1$ .

For the test, the hypotheses are stated as follows:

$$\begin{array}{ll} \text{Null Hypothesis,} & H_0: \mu = 0.65 \\ \text{Alternate Hypothesis,} & H_1: \mu \neq 0.65 \end{array}$$

### 1.2 Types of Hypothesis Tests

Hypotheses concerning both measurements (variables or continuous data) or counts (attribute or discrete data) may be tested. The hypothesis  $H_0: \mu = 0.65$  is an example using continuous data.

A hypothesis concerning discrete data may be stated as follows:

The population from which a sample of one hundred parts was taken has a process average of 2% defective.

$$H_0 : p = .02 \quad \text{or} \quad H_0 : np = 100 \times .02 = 2$$

The symbol  $\mu$  represents the population mean and  $\bar{p}$  represents the sample mean. The symbol  $np$  represents the average number of defective parts in the sample for both the hypergeometric and the binomial distributions. For the Poisson distribution,  $np$  is the average number of defects in the sample.

In hypothesis testing, a general rule is to use the normal curve areas (Z - statistics) when the sample size  $n$  is thirty or more. When  $n$  is less than thirty, the t distribution areas (t - statistics) and corresponding tables are used. When  $n$  is very large ( $n \rightarrow \infty$ ), the Z table and t table are virtually identical. The method of testing a hypothesis is exactly the same; the only difference is the table that is used.

In the previous chapter  $n - 1$  was used instead of  $n$  in the denominator of the variance formula. Division by  $n - 1$  is called division by degrees of freedom. The term degrees of freedom is so named because only  $n - 1$  linear comparisons can be made among  $n$  choices. If there are three job candidates and three jobs to fill, there are three choices for the first job. Once it is filled, there are two choices for the second job. After the second job is filled, there is only one choice for the third job. The third job can only be filled by the remaining person, so there are  $n - 1$  or two degrees of freedom. To use the t table, the number of degrees of freedom is required.

## 2.0 DECISION ERRORS

Anytime sample data are used to make decisions or inferences about the entire population, decision errors can be made. In hypothesis testing and acceptance sampling, two types of errors can occur.

Type I error, alpha ( $\alpha$ ), rejection of a true null hypothesis

Type II error, beta ( $\beta$ ), acceptance of a false null hypothesis

	Null hypothesis ( $H_0$ ) is true	Null hypothesis ( $H_0$ ) is false
Do not reject $H_0$	no error ( $1 - \alpha$ )	Type II error ( $\beta$ )
Reject $H_0$	Type I error ( $\alpha$ )	no error ( $1 - \beta$ )

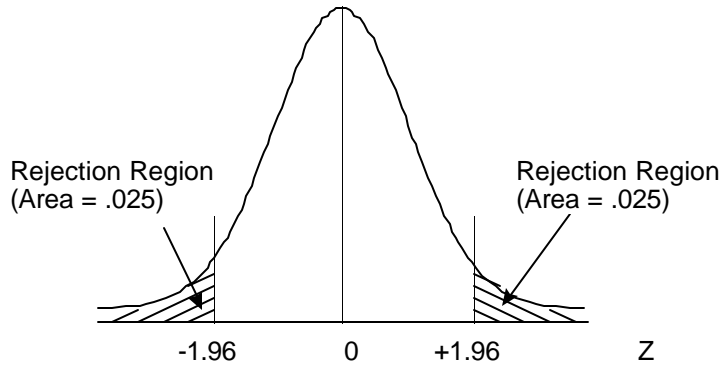
The level of significance is denoted by  $\alpha$ . This is the probability of rejecting a null hypothesis when it is true or the probability of making a wrong decision. The confidence interval is denoted by  $1 - \alpha$ . The confidence interval is the probability of not rejecting a null hypothesis when it is true or the probability of making a right decision. The probability of not rejecting a null hypothesis when it is false is  $\beta$ . This is the probability of making a wrong decision. The term  $1 - \beta$  is referred to as the power of test and is the probability of rejecting a null hypothesis when it is false. This is the probability of making a right decision when the null hypothesis is truly false. In acceptance sampling,  $\alpha$  and  $\beta$  are referred to as the producer's risk and consumer's risk respectively.

### 3.0 HOW A HYPOTHESIS IS TESTED

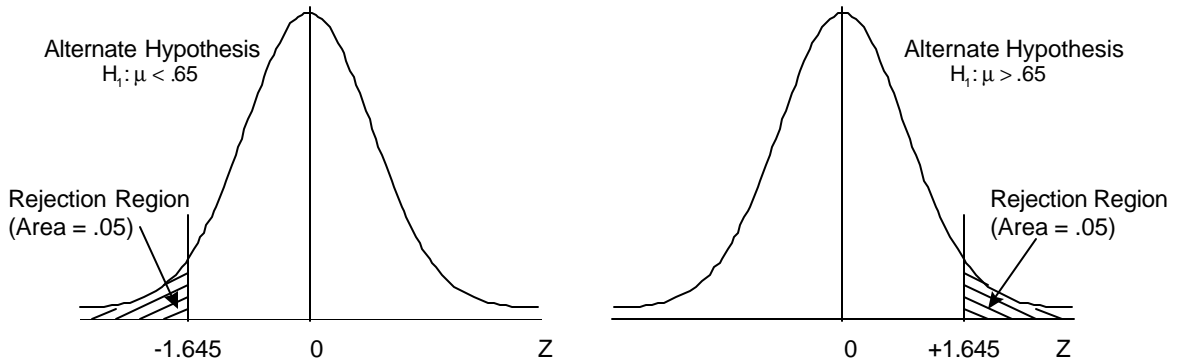
#### 3.1 One or two tailed test?

The alternate hypothesis determines whether the test will be a one tailed test or a two tailed test. If  $H_1$  is stated as  $H_1: \mu \neq .65$ , then the test is a two tailed test. If  $H_1$  is stated as  $H_1: \mu < .65$  or  $H_1: \mu > .65$ , the test is a one tailed test.

Sketch for a two tailed test ( $\alpha = .05$ )  
 Alternate Hypothesis,  $H_1: \mu \neq .65$



Sketches for a one tailed test ( $\alpha = .05$ )



#### 3.2 Examples

##### Example 1

Example 1 is a step by step procedure on how to perform a hypothesis test. A sample of 100 bearings is taken. The diameter is measured and recorded for each part in the sample. The mean and standard deviation are calculated. The following results are obtained:

$$\bar{x} = 0.67 \text{ inches, } s = 0.08 \text{ inches, } n = 100$$

The distribution of individual data values has a mean of 0.67" and a standard deviation of 0.08". Since a sample mean is to be tested against a population mean, the test distribution is a distribution of averages. The distribution of averages has a mean of 0.67" and a standard deviation or standard error of  $0.08/\sqrt{n}$ .

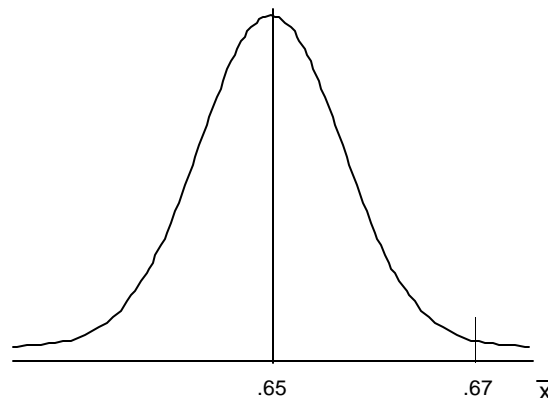
Could the population from which this sample was selected have an average diameter of 0.65 inches? Test this hypothesis against a level of significance of .05 ( $\alpha = .05$ ).

**Step 1)** State the null and alternate hypotheses.

$$H_0: \mu = 0.65$$

$$H_1: \mu \neq 0.65$$

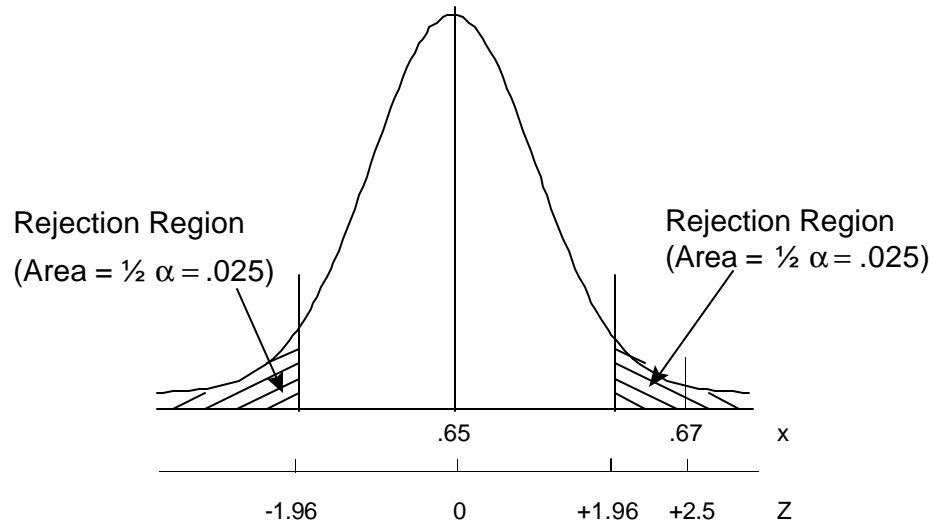
**Step 2)** Draw a sketch showing  $\mu = .65$  and  $\bar{x} = .67$  on the  $\bar{x}$  scale.



**Step 3)** Determine the critical values of Z and the region of rejection.

Since  $H_1: \mu \neq .65$ , this will be a two tailed test. Half of  $\alpha$  will define the rejection region on the left tail of the curve and half of  $\alpha$  will define the rejection region on the right tail of the curve. Half of  $\alpha$  is  $.05/2 = .025$ . The sample size is greater than thirty, therefore the correct table to use is the standard normal curve or Z table.

Either draw another sketch or superimpose a Z scale on the first sketch. Look up  $.50 - .025 = .475$  in the table. The corresponding Z values are  $\pm 1.96$ . These are the critical values. The critical values are the Z values that separate the rejection region from the acceptance region. On the sketch, shade in the areas to the left of -1.96 and to the right of +1.96. The shaded areas indicate the regions of rejection. The area of the acceptance region is .95, therefore this test will have a confidence interval of .95.



**Step 4)** Calculate the value of  $Z$  that corresponds with  $\bar{x}$  using the  $\bar{x}$  to  $Z$  conversion formula.

$$Z = \frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}}$$

$$Z = \frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}} = \frac{(.67 - .65)}{\frac{.08}{\sqrt{100}}} = \frac{.02}{\frac{.08}{10}} = \frac{.02}{.008} = +2.5$$

If the calculated value of  $Z$  is between the critical values, it falls in the acceptance region and the hypothesis is not rejected. If the calculated value of  $Z$  is to the left of  $-1.96$  or to the right of  $+1.96$ , the hypothesis that  $\mu = .65$  is rejected in favor of the alternate hypothesis that  $\mu \neq .65$ .

**Step 5)** Make a decision to reject or not reject the null hypothesis.

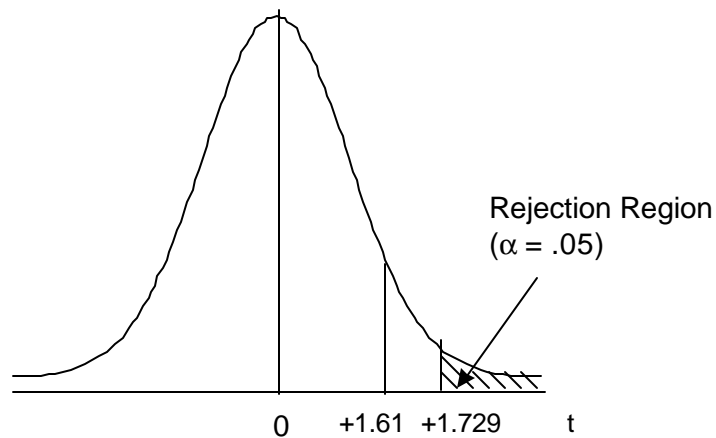
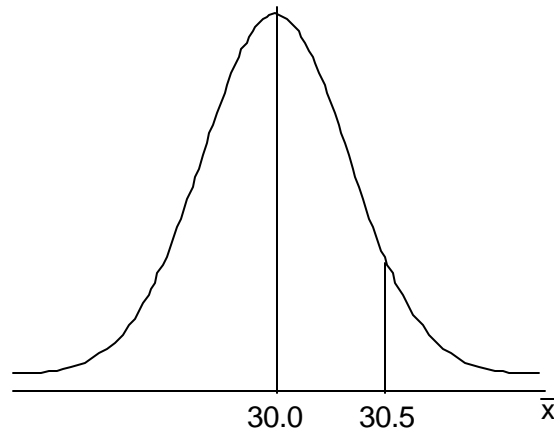
The calculated  $Z$  value of  $+2.5$  lies in the rejection region, therefore  $H_0$  is rejected in favor of  $H_1$ . It is concluded that  $\mu \neq .65$ .

### Example 2

Given  $\bar{x} = 30.5$ ,  $s = 1.4$ , and  $n = 20$ . Test the hypothesis  $\mu = 30$  against the alternative  $\mu > 30$  at a level of significance of  $.05$ . This will be a one tailed test because the alternate hypothesis is  $\mu > 30$ . Since  $n = 20$ , the  $t$  table is used to determine the critical value and rejection region.

The calculated value of t is

$$t = \frac{(\bar{x} - \mu)}{\frac{s}{\sqrt{n}}} = \frac{(30.5 - 30)}{\frac{1.4}{\sqrt{20}}} = \frac{.5}{.31} = +1.61$$



The critical value is  $t = +1.79$ . The area to the right of the critical value is the region of rejection. The calculated value  $t = 1.61$  falls in the acceptance region, therefore  $H_0$  is not rejected. It can be concluded that the sample could have reasonably come from a population whose mean is thirty ( $\mu = 30$ ).

### Example 3

A sample of fifty parts is taken from a large shipment of a purchased product. The parts are inspected and categorized as either defective or non-defective. The inspector finds two defective parts. ( $\bar{p} = 2/50 = .04$  or 4% defective). The supplier states that his process yields 3% defective product. Could the supplier be telling the truth? Test at a level of significance of .01. This problem can be worked out by stating the null hypothesis in terms of  $p$  or in terms of  $np$ .

The standard deviation is modified depending on which null hypothesis is used. The Z value and the resulting decision will be the same. The formulas below are from chapter 4, page 60.

The standard deviation in terms of np for the binomial distribution =  $\sqrt{npq}$

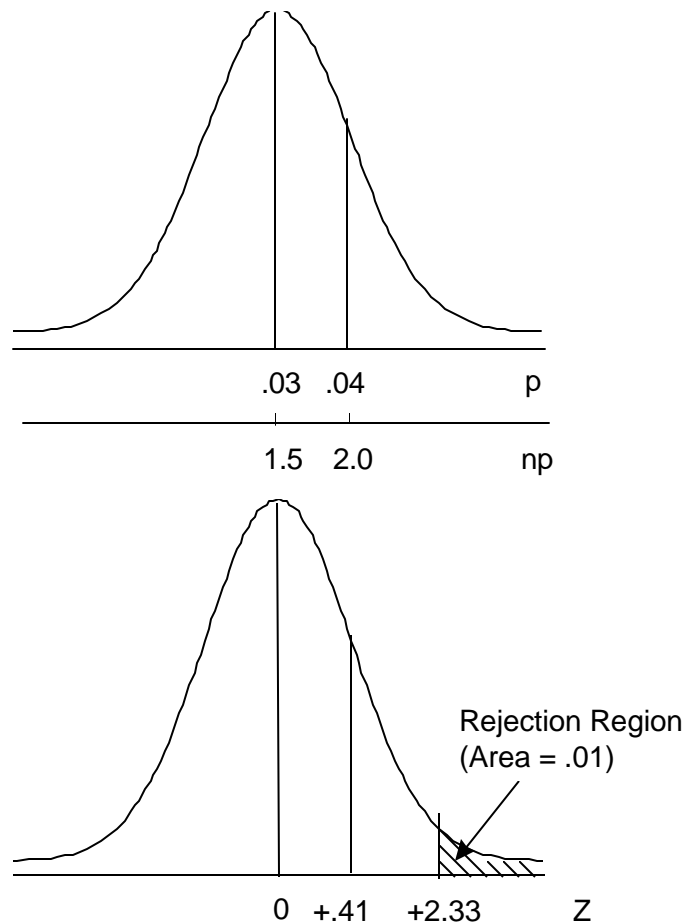
The standard deviation in terms of p for the binomial distribution =  $\sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$

The hypothesis test in terms of p is  $H_0: p = .03$ ,  $H_1: p > .03$ ,  $\bar{p} = .04$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.03(.97)}{50}} = .0241 \qquad Z = \frac{(\bar{p} - p)}{\sigma_p} = \frac{(.04 - .03)}{.0241} = .41$$

The hypothesis test in terms of np is  $H_0: np = 1.5$ ,  $H_1: np > 1.5$ ,  $n\bar{p} = 2$

$$\sigma_{np} = \sqrt{np(1-p)} = \sqrt{(50)(.03)(.97)} = 1.206 \qquad Z = \frac{(n\bar{p} - np)}{\sigma_{np}} = \frac{(2 - 1.5)}{1.206} = .41$$



The critical value of Z is +2.33. The calculated value of Z is +.41. This value lies in the acceptance region, therefore  $H_0$  is not rejected. The test supports the supplier's statement that the process yields 3% defective product.

#### 4.0 TEST FOR GOODNESS OF FIT

To test a sample distribution to determine if it fits the pattern of a specific distribution such as the normal, a Goodness of Fit test is performed. Goodness of Fit refers to the comparison of an observed sample distribution with a known theoretical distribution. In the following example the chi square distribution is used to compare a sample distribution with the normal distribution. This is a test to determine if the universe from where the sample data were obtained could be normally distributed.

$$\text{Chi Square} = \chi^2 = \frac{\sum(f_i - F_i)^2}{F_i}$$

$f_i$  = observed frequency  
 $F_i$  = theoretical frequency

Based on the sample data, a hypothesis that the universe may be normally distributed against an alternative that it is not normally distributed is to be tested.

$H_0$ : universe may be normally distributed  
 $H_1$ : universe is not normally distributed

For a chi square test involving intervals or cells, the degrees of freedom are  $k - 3$ , where  $k$  is the number of intervals. For tests using row and column matrices called contingency tables, the degrees of freedom are  $(\# \text{ rows} - 1) \times (\# \text{ columns} - 1)$  or  $(r - 1)(c - 1)$ .

##### Example 4

The measurements from a sample of one hundred number 10 bearings from the Canape Bearing Company are grouped into seven intervals or cells. The data table is on the next page. The number of intervals is arbitrary. More cells will give a higher degree of accuracy. Six to Twelve cells will yield satisfactory results. The number of occurrences in each cell is compared to the number of occurrences that should be in the cell based on the known theoretical distribution. In this example, the data are compared to the normal distribution. The chi square test sums the differences in each cell and the total value of chi square is used to test the hypothesis that the data may be normally distributed. The mean and standard deviation are calculated using the midpoint ( $m_i$ ) of the interval and the observed frequencies ( $f_i$ ).

$$\bar{x} = \frac{\sum(f_i m_i)}{n} = \frac{11310}{100} = 113.1$$

$$s = \sqrt{\frac{\sum[f_i (m_i - \bar{x})^2]}{n - 1}} = \sqrt{\frac{29339}{100}} = 17.21$$

Could the universe from where these data were obtained be normally distributed? Test at a level of significance of .05. The null and alternate hypotheses are

$H_0$ : universe may be normally distributed  
 $H_1$ : universe is not normally distributed

Interval	Observed Frequency (f <sub>i</sub> )	Theoretical Frequency (F <sub>i</sub> )	$\frac{(f_i - F_i)^2}{F_i}$
140 or above	8	5.9	.75
130 - < 140	11	10.4	.03
120 - < 130	15	18.1	.53
110 - < 120	20	22.7	.32
100 - < 110	21	20.5	.01
90 - < 100	16	13.4	.50
below 90	9	9.0	0
			$\chi^2 = 2.14$

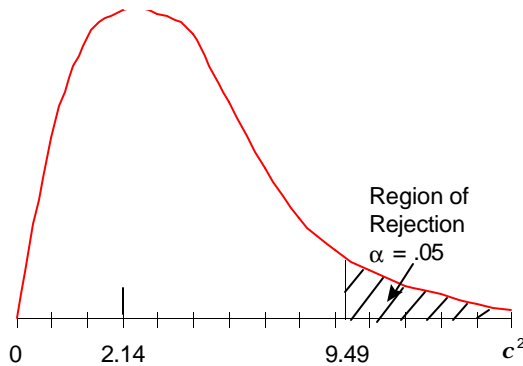
The degrees of freedom are k - 3. There are seven intervals, therefore d.f. are 7 - 3 = 4

Chi square =  $\chi^2 = 2.14$ . From the chi square table, the critical value for  $\alpha = .05$  and 4 degrees of freedom is 9.49. Therefore, the null hypothesis is not rejected. It is concluded that the universe could be normally distributed.

#### 4.1 Sketch of the Chi Square Distribution

The chi square distribution is not a symmetrical distribution. Its shape depends on the degrees of freedom. If it is known that a universe is normal, the distribution of sample variances has the form of the chi square distribution. The chi square values that are computed in the example are sample variances. This is why the chi square distribution is used to perform the test.

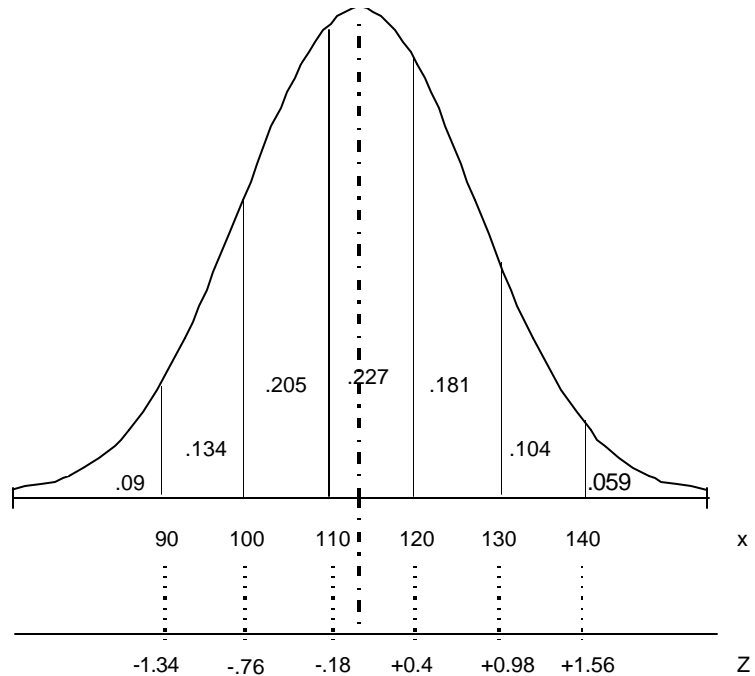
The critical value of chi square is defined as the value on the abscissa that indicates where the region of rejection begins. The numbers on the abscissa are the various values of chi square. The region of rejection is the area to the right of the critical value and is indicated by the shading.



If goodness of fit tests are made to compare sample data to other distributions, such as the uniform or binomial, the theoretical frequencies would be obtained using the specific distribution.

## 4.2 Theoretical Frequencies for the Example

The theoretical frequencies are obtained from the standard normal curve table.



$$\bar{x} = 113.1 \text{ and } s = 17.21$$

## 5.0 INFERENCES ABOUT TWO MEANS

When a control group is compared with an experimental group, or when two experimental groups are compared, the concern is primarily with differences in their means. This test assumes that the means are independent. Two samples are independent when the outcome of one does not affect the other. The pooled variance is the combined variance of two or more samples and is denoted by  $s_p^2$ .

The pooled variance for two samples is

$$s_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

and  $n_1$  and  $s_1^2$  are the sample size and variance for the first sample and  $n_2$  and  $s_2^2$  are the sample size and variance for the second. The pooled standard deviation is  $\sqrt{s_p^2} = s_p$ . The denominator  $(n_1 + n_2 - 2)$  indicates the total degrees of freedom.

Two processes are being evaluated. A study shows that 10 employees working on process A have an average assembly time of 250 seconds with variance of 100. In addition, 20 employees on process B have an average assembly time of 225 seconds with variance of 80. Are the average assembly times for the two processes significantly different?

For this type of test, the variances are always tested first for significant differences. The F ratio and corresponding F table values are used to test the variances. If the variances are significantly different, the test for differences in means is not valid. Before testing for differences in means, the assumption is that the variances are homogeneous.

The t table is used for the test of differences between means since both sample sizes are less than thirty.

The formula for conversion from the  $\bar{x}$  scale to the t scale:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

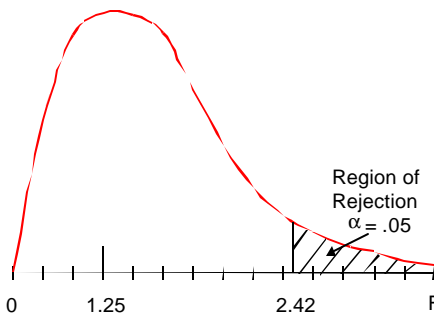
The variances are tested for significant differences ( $\alpha = .05$ ).

$$1) \quad H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2$$

$$2) \quad F = \frac{s_L^2}{s_s^2} = \frac{100}{80} = 1.25$$

The F ratio is a value on the abscissa of the F distribution curve. The symbol  $s_L^2$  represents the larger of the two variances and the symbol  $s_s^2$  represents the smaller of the two variances. The sample size for the larger variance is denoted by  $n_L$  and the sample size for the smaller variance is denoted by  $n_s$ . There are  $n_L - 1$  degrees of freedom for the numerator and  $n_s - 1$  degrees of freedom for the denominator.

- 3) Region of rejection:  $F \geq 2.42$ . This is the critical value of F from the  $F_{.05}$  table with 9 degrees of freedom for the numerator and 19 degrees of freedom for the denominator.



- 4) The calculated value of F is less than the critical value of 2.42. The null hypothesis that  $\sigma_1^2 = \sigma_2^2$  is not rejected. It is concluded that there is no significant difference between the two variances.

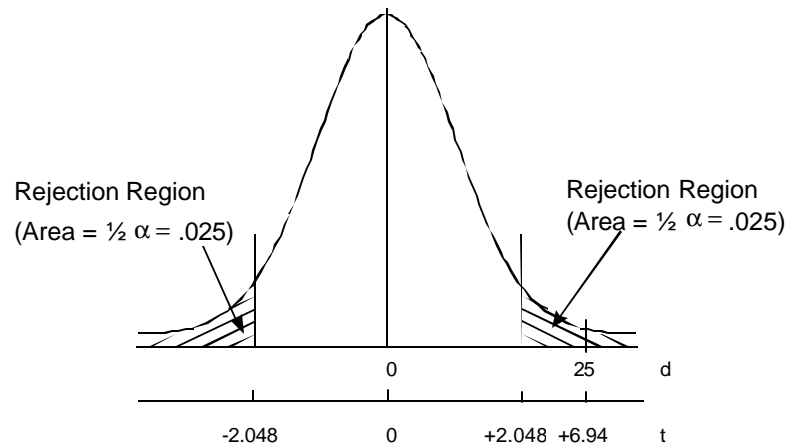
The means are now tested for significant differences ( $\alpha = .05$ ).

- 1)  $H_0: \mu_1 = \mu_2$   
 $H_1: \mu_1 \neq \mu_2$

- 2) Region of rejection:  $-2.048 > t > +2.048$  (These are the critical values of t).

$$3) s_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}, \quad s_p = \sqrt{s_p^2} = \sqrt{\frac{9(100) + 19(80)}{28}} = 9.2967$$

$$3) t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{250 - 225}{9.2967 \sqrt{\frac{1}{10} + \frac{1}{20}}} = \frac{25}{9.2967(.3873)} = +6.94$$



- 4) The calculated value of t,  $t = +6.94$ , falls beyond the critical value of  $t = +2.048$ , therefore the null hypothesis that  $\mu_1 = \mu_2$  is rejected. The conclusion is that there is a statistically significant difference between the two means.