LECTURE 08

HYPOTHESIS TESTING - II

Outline of today's lecture:

I. Comments about Hypothesis Testing

Attention 1 Since there is always a room for Type II error, β when we carry out hypothesis testing:

- You *should say* "Fail to Reject" the null hypothesis, H_0 → You *should not* say "accept" null hypothesis, H_0

Attention 2 Use the equality sign in null hypothesis when you do a one tailed-test.

- In fact it is a controversial issue in statistical science. Some statisticians prefer to use ≤ or ≥ signs in null hypothesis. However, when we use the null hypothesis such as $H_0 : \mu = 0.8$, calculating the level of α is simple: we just found $P(\text{value of test statistic is in RR when } H_0 : \mu = 0.8 \text{ is true})$. If we used $H_0 : \mu \geq 0.8$ as the null hypothesis, our previous definition of α [P(value of test statistic is in RR when $H_0 : \mu = 0.8$ is true)] would be inadequate because the value of P(value of test statistic is in RR) will change for different possible values of μ since in this case any value for μ is possible within the interval $\mu \geq 0.8$. In cases like this, α is defined to be the maximum (over all values of $\mu \geq 0.8$) value of P(value of test statistic is in RR). This maximum value happens when $\mu = 0.8$, the "boundary value" of μ in $H_0 : \mu \geq 0.8$. Thus, using $H_0 : \mu = 0.8$ instead of $H_0 : \mu \geq 0.8$ leads to the correct testing procedure and the correct calculation of α without needlessly raising additional considerations.

II. Small Sample Hypothesis Testing

In order to apply large-sample hypothesis-testing procedures, the sample size must be large enough that $Z = \left(\hat{\theta} - \theta_0\right) / \sigma_{\hat{\theta}}$ has approximately a standard normal distribution.

In this section, we will develop formal procedures for testing hypotheses about μ and μ_1 - μ_2, procedures that are appropriate for small samples from normal populations.

A. Small Sample Tests for Mean

- We assume that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a normal distribution with unknown mean μ and unknown variance σ^2.
- If $\overline{Y}$ and S denote the sample mean and sample standard deviation, respectively, and if $H_0 : \mu = \mu_0$ is true, then

$$T = \frac{\overline{Y} - \mu_0}{S / \sqrt{n}}$$

has a t distribution with n - 1 degrees of freedom

Small Sample Tests for Population Mean

Assumptions: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with $E(Y_i) = \mu$.

$H_0 : \mu = \mu_0$

$$H_a : \begin{cases} \mu > \mu_0 & \text{(upper tail alternative)} \\ \mu < \mu_0 & \text{(lower tail alternative)} \\ \mu \neq \mu_0 & \text{(two-tailed alternative)} \end{cases}$$

Test statistic: $T = \dfrac{\overline{Y} - \mu_0}{S / \sqrt{n}}$

Rejection Region: $RR : \begin{cases} \{t > t_\alpha\} & \text{(upper-tail RR)} \\ \{t < -t_\alpha\} & \text{(lower-tail RR)} \\ \{|t| > t_{\alpha/2}\} & \text{(two-tailed RR)} \end{cases}$.

See t-distribution table for values of t_α, with $v = n - 1$ degrees of freedom.

Example 10.12 A manufacturer of gunpowder has developed a new powder, which was tested in eight shells. The sample mean and sample standard deviation, $\bar{Y} = 2959$ and $s = 39.1$, respectively.

- The manufacturer *claims that the new gunpowder produces an average velocity of not less than 3000 feet per second*.
- Do the sample data provide sufficient evidence <u>to contradict</u> the manufacturer's claim at the 0.025 level of significance? (Assume that muzzle velocities are approximately normally distributed)

<u>*Solution*</u>

We want to test $H_0 : \mu = 3000$ versus the alternative, $H_A : \mu < 3000$.[1]

The rejection region is given by $t < -t_{0.025} = -2.365$, where t possesses $v = (n - 1) = 7$ degrees of freedom. Computing, we find that the observed value of the test statistic is

$$T = \frac{\bar{Y} - \mu_0}{S / \sqrt{n}} = \frac{2959 - 3000}{39.1 / \sqrt{8}} = -2.966$$

This value falls in the rejection region (that is, $t = -2.966$ is less than -2.365); hence the null hypothesis is rejected at the α=0.025 level of

[1] Do not forget: generally, the *desired conclusion* of the study is stated in the alternative hypothesis. In other words, often an alternative hypothesis is the *desired conclusion* of the investigator. In this question: the desired conclusion is if the sample data provides sufficient evidence to contradict the manufacturer's claim at the 0.025 level of significance, hence $H_A : \mu < 3000$

significance. So we conclude that sufficient evidence exists to contradict the manufacturer's claim, and we conclude that the true mean muzzle velocity is less than 3000 feet per second.

Example 10.13 What is the *p-value* associated with the statistical test in Example 10.12?

<u>*Solution*</u>

Because the null hypothesis should be rejected if t is "small," the smallest value of a for which the null hypothesis can be rejected is $p\text{-value} = P(T < -2.966)$, where T has a t distribution with n-1=7 degrees of freedom.

Unlike the table of areas under the normal curve, T tables does not give areas corresponding to many values of t. Rather, it gives the values o f t corresponding to upper-tail areas equal to 0.10, 0.05, 0.025, 0.010, and 0.005. Because the t distribution is symmetric about 0, we can use these upper-tail areas to provide corresponding lower-tail areas. In this instance, the t statistic is based on 7 degrees of freedom; hence, we consult the $df = 7$ row of T table and find that -2.966 falls between $-t_{0.025} = -2.365$ and $-t_{0.01} = -2.998$. These values are indicated in Figure 10.8.

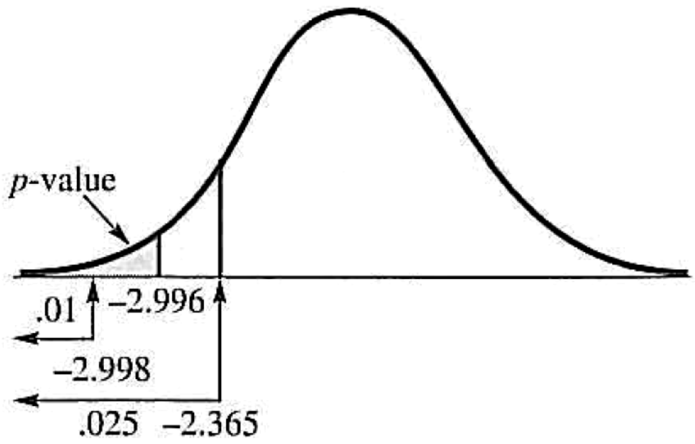

Figure 1 Determination of *p-value* for Example 10.13

Because the observed value of T (-2.966) is less than $-t_{0.025}$ = -2.365 but not less than $-t_{0.01}$ = = -2.998, we reject H_0 for $\alpha = 0.025$ but not for α= 0.01. Thus the *p-value* for the test satisfies:

$$0.01 \leq p\text{-}value \leq 0.025.$$

Indeed, if you want to obtain exact p-value, you can use *Excel*. The following formula will give you the *exact p-value*:

=TDIST(2,966;7;1)

The exact p-value is *0,010463*. The general formula is as follows:

TDIST(x; df; tails) [2]

[2] If the regional setting of your computer is not set to Turkish, then the formula will be of form: TDIST(x, df, tails).

where

x The numeric value at which to evaluate the distribution.

df The number indicating the number of degrees of freedom.

tails The number of distribution tails to return (If "tails" = 1, then the *one-tailed distribution* is used, if "tails" = 2, then the two-tailed distribution is used.)

B. Small Sample Tests for Comparing Two Population Means

A second application of the t distribution is in constructing a small-sample test to compare the means of two normal populations that possess *equal variances*.

- Suppose that independent random samples are selected from each of two normal populations: $Y_{11}, Y_{12}, ..., Y_{1n_1}$, from the first, and $Y_{21}, Y_{22}, ..., Y_{2n_2}$ from the second, where the mean and variance of the i^{th} population are μ_i and σ^2, for i=1,2.
- Further, assume that $\bar{Y}_i$, and S_i^2, for i=1,2, are the corresponding sample means and variances.
- When these assumptions are satisfied, then

$$T = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

has a Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Here, $S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ is the *pooled* estimator for σ^2.

Small Sample Tests for Comparing Two Population Means

Assumptions: Independent samples from normal distributions with $\sigma_1^2 = \sigma_2^2$

$$H_0 : \mu_1 - \mu_2 = D_0$$

$$H_a : \begin{cases} \mu_1 - \mu_2 > D_0 & \text{(upper tail alternative)} \\ \mu_1 - \mu_2 < D_0 & \text{(lower tail alternative)} \\ \mu_1 - \mu_2 \neq D_0 & \text{(two-tailed alternative)} \end{cases}$$

Test statistic: $T = \dfrac{\left(\overline{Y}_1 - \overline{Y}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$, where

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Rejection Region: $RR : \begin{cases} \{t > t_\alpha\} & \text{(upper-tail RR)} \\ \{t < -t_\alpha\} & \text{(lower-tail RR)} \\ \{|t| > t_{\alpha/2}\} & \text{(two-tailed RR)} \end{cases}$.

See t-distribution table for values of t_α, with $v = n_1 + n_2 - 2$ degrees of freedom.

Examples 10.14 and 10.15 Solve them!

III. Testing Hypotheses Concerning Variances

A. Tests for a Population Variance

We again assume that we have a random sample $Y_1, Y_2, ..., Y_n$ from a normal distribution with unknown mean μ and unknown variance σ^2. Here, we consider the problem of testing $H_0 : \sigma^2 = \sigma_0^2$ for some fixed value σ_0^2 versus various alternative hypotheses.

If H_0 is true and $\sigma^2 = \sigma_0^2$, then we know that (Chapter 8),

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

has a χ^2 distribution with *n-1* degrees of freedom. If we desire to test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_A : \sigma^2 > \sigma_0^2$,we can use $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ as our test statistic.

- If H_A is true, then the actual value of σ^2 is larger than σ_0^2. In this case, we would expect S^2 (which estimates the true value of σ^2) to be larger than σ_0^2.

- The larger S^2 is relative to σ_0^2, the stronger is the evidence to support $H_A: \sigma^2 > \sigma_0^2$.
- Notice that S^2 is large relative to σ_0^2; then $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ would be large. is large.
- Thus we see that a rejection region of the form $RR = \{\chi^2 > k\}$ for some constant k is appropriate for testing $H_0: \sigma^2 = \sigma_0^2$ versus $H_A: \sigma^2 > \sigma_0^2$.

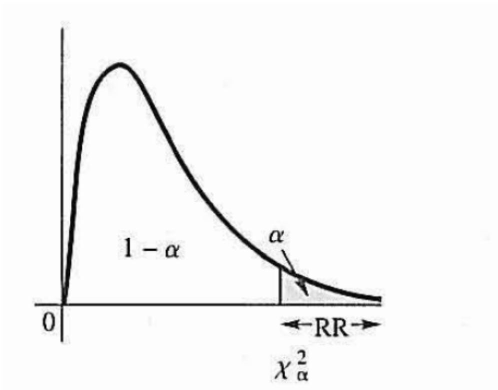

Figure 2 Rejection region RR for testing
$H_0: \sigma^2 = \sigma_0^2$ versus $H_A: \sigma^2 > \sigma_0^2$

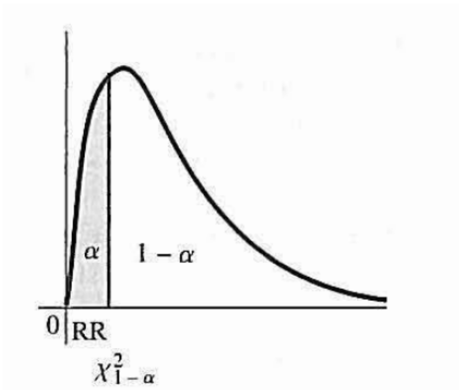

Figure 3 Rejection region RR for testing
$H_0: \sigma^2 = \sigma_0^2$ versus $H_A: \sigma^2 < \sigma_0^2$

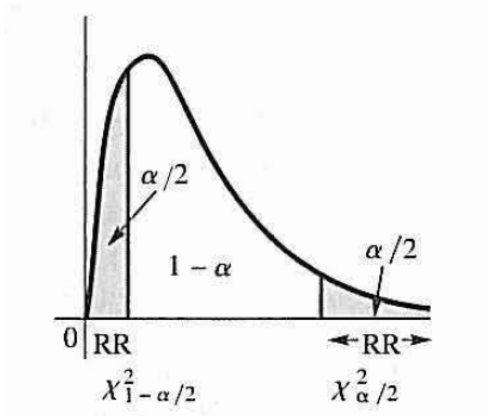

Figure 4 Rejection region RR for testing
$H_0: \sigma^2 = \sigma_0^2$ versus $H_A: \sigma^2 \neq \sigma_0^2$

Tests of Hypotheses Concerning a Population Variance

Assumptions: $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$.

$H_0: \sigma^2 = \sigma_0^2$

$$H_a: \begin{cases} \sigma^2 > \sigma_0^2 & \text{(upper tail alternative)} \\ \sigma^2 < \sigma_0^2 & \text{(lower tail alternative)} \\ \sigma^2 \neq \sigma_0^2 & \text{(two-tailed alternative)} \end{cases}$$

Test statistic: $\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$

Rejection Region:

$$RR: \begin{cases} \{\chi^2 > \chi^2_\alpha\} & \text{(upper-tail RR)} \\ \{\chi^2 < \chi^2_{1-\alpha}\} & \text{(lower-tail RR)} \\ \{\chi^2 > \chi^2_{\alpha/2} \text{ or } \chi^2 < \chi^2_{1-(\alpha/2)}\} & \text{(two-tailed RR)} \end{cases}$$

Note that χ^2_α is chosen so that, for $v = n-1$ degrees of freedom, $P(\chi^2 > \chi^2_\alpha) = \alpha$.

Example 10.16 A company produces machined engine parts that are supposed to have a diameter variance no larger than 0.0002 (diameters measured in inches). A random sample of 10 parts gave a sample variance of 0.0003. Test, at the 5% level, $H_0 : \sigma^2 = 0.0002$ against $H_A : \sigma^2 > 0.0002$.

(Note: Assume that the measured diameters are normally distributed)

<u>*Solution*</u>

If it is reasonable to assume that the measured diameters are normally distributed, the appropriate test statistic is:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Since we have posed an upper-tail test, we reject H_0 for values of this statistic larger than $\chi^2_{0.05}$=16.919 (based on 9 *df*).

The observed value of the test statistic is:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{9.(0.0003)}{0.0002} = 13.5$$

Thus, H_0 is not rejected. There is not sufficient evidence to indicate that σ^2 exceeds 0.0002 at the 5% level of significance.

Example 10.17 Determine the *p-value* associated with the statistical test of Example 10.16

<u>*Solution*</u>

The *p-value* is the probability that a χ^2 random variable with 9 d.f. is larger than the observed value of 13.5. The area corresponding to this probability is shaded in Figure, below.

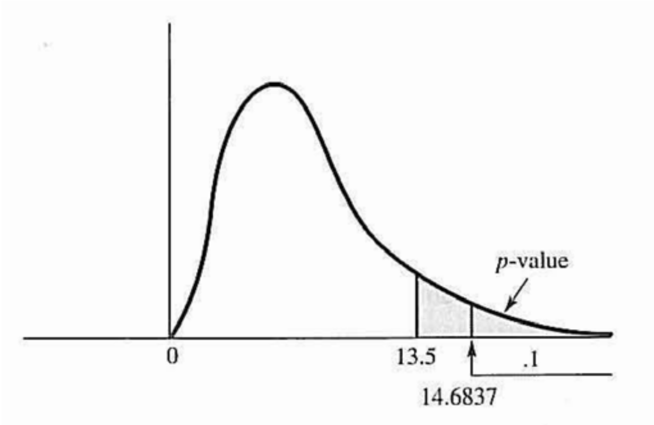

Figure 5 Illustration of the *p-value* for Example 10.17

We find that $\chi^2_{0.1}$= 14.6837. As the figure above indicates, the shaded area exceeds 0.1 and thus *p-value>0.1*. That is, for any value of α<0.l, the null hypothesis cannot be rejected.

B. Tests for Comparing Two Population Variances

Sometimes we wish to compare the variances of two normal distributions, particularly by testing to determine whether they are equal.

- These problems are encountered in comparing the precision of two measuring instruments,
- the variation in quality characteristics of a manufactured product, or
- the variation in scores for two testing procedures.

For example, suppose that $Y_{11}, Y_{12}, ..., Y_{1n_1}$ and $Y_{21}, Y_{22}, ..., Y_{2n_2}$ are independent random samples from normal distributions with unknown means and that $V(Y_{1i}) = \sigma_1^2$ and $V(Y_{2i}) = \sigma_2^2$, where σ_1^2 and σ_2^2 are unknown. Suppose that we want to test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_A : \sigma_1^2 > \sigma_2^2$.

Because the sample variances S_1^2 and S_2^2 estimate the respective population variances σ_1^2 and σ_2^2, we will reject H_0 in favor of H_A if S_1^2 is much larger than S_2^2. That is, we will use a rejection region RR of the form

$$RR = \left\{ \frac{S_1^2}{S_2^2} > k \right\}$$

where k is chosen so that the probability of a type I error is α.

The appropriate value of *k* depends upon the probability distribution of the statistic $\frac{S_1^2}{S_2^2}$. Notice that $\frac{(n_1-1)S_1^2}{\sigma_1^2}$ and $\frac{(n_2-1)S_2^2}{\sigma_2^2}$ are independent chi-square random variables. From the definition of *F distribution* (Definition 7.3), it follows that

$$F=\frac{\dfrac{\left(\dfrac{\cancel{(n_1-1)}S_1^2}{\sigma_1^2}\right)}{\cancel{(n_1-1)}}}{\dfrac{\left(\dfrac{\cancel{(n_2-1)}S_2^2}{\sigma_2^2}\right)}{\cancel{(n_2-1)}}}=\frac{\dfrac{S_1^2}{\sigma_1^2}}{\dfrac{S_2^2}{\sigma_2^2}}=\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$$

has an F distribution with (n_1-1) numerator degrees of freedom and (n_2-1) denominator degrees of freedom.

Under the null hypothesis that $\sigma_1^2=\sigma_2^2$, it fol1ows that $F=\frac{S_1^2}{S_2^2}$ and the rejection region RR given earlier is equivalent to $RR=\{F>k\}=\{F>F_\alpha\}$ where $k=F_\alpha$. is the value of the F distribution with $v_1=(n_1-1)$ and $v_2=(n_2-1)$ such that $P(F>F_\alpha)=\alpha$.

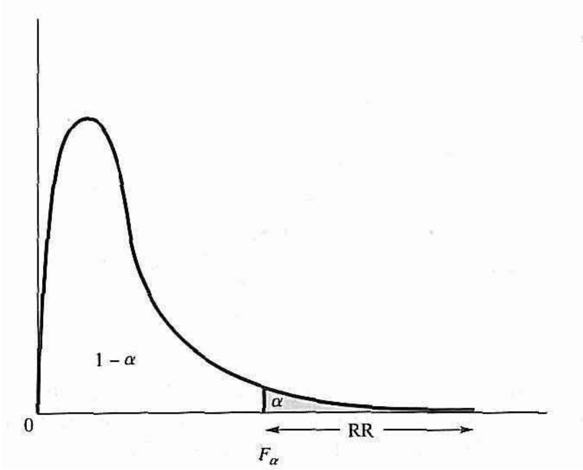

Figure 6 Rejection region RR for testing
$H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_A : \sigma_1^2 > \sigma_2^2$

Example 10.19 Suppose that we wish to compare the variation in diameters of parts produced by the company in Example 10.16 with the variation in diameters of parts produced by a competitor. Recall that the sample variance for our company, based on *n=10* diameters, was $s_1^2 = 0.0003$. In contrast, the sample variance of the diameter measurements for 20 of the competitor's parts was s_1^2=0.0001. Do the data provide sufficient information to indicate a smaller variation in diameters for the competitor? Test with $\alpha = 0.05$.

Solution

We are testing $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative $H_A : \sigma_1^2 > \sigma_2^2$. The test statistic, $F = \frac{S_1^2}{S_2^2}$ is based on v_1=9 numerator and v_2=19 denominator degrees of freedom, and we reject H_0 for values of F larger than $F_{0.05} = 2.42$.

Because the observed value of the test statistic is

$$F = \frac{S_1^2}{S_2^2} = \frac{0.0003}{0.0001} = 3$$

we see that $F > F_{0.05}$; therefore, at the $\alpha = 0.05$ level, we reject $H_0 : \sigma_1^2 = \sigma_2^2$ in favor of $H_A : \sigma_1^2 > \sigma_2^2$ and conclude that the competing company produces parts with smaller variation in their diameters.

Example 10.20 Give bounds for the *p-value* associated with the data of Example 10.19.

<u>*Solution*</u>

The calculated *F* value for this upper-tail test is *F*=3. We know that his value is based on v_1=9 and v_2=19 numerator and denominator degrees of freedom, respectively. From F Table we see that $F_{0.025} = 2.88$, whereas $F_{0.01} = 3.52$. Thus, the observed value, F=3, would lead to rejection of the null hypothesis for $\alpha = 0.025$ but not for $\alpha = 0.01$.Hence, *0.01< p-value < 0.025.*

Attention If the research hypothesis is that the variance of one population is *larger* than the variance of another population, we identify the population with the hypothesized larger variance as population 1 and proceed as indicated in the solution to Example 10.19. (So we put the larger variance on numerator in statistic F)

Tests of Hypothesis $\sigma_1^2 = \sigma_2^2$

Assumptions: Independent samples from normal populations.

$H_0: \sigma_1^2 = \sigma_2^2$

$H_A: \sigma_1^2 > \sigma_2^2$

Test statistic: $F = \dfrac{S_1^2}{S_2^2}$

Rejection region: $F > F_\alpha$, where F_α is chosen so that $P(F>F_\alpha)=\alpha$ when F has $v_1 = (n_1 - 1)$ numerator degrees of freedom and $v_2 = (n_2 - 1)$ denominator degrees of freedom.

If we use $F = \dfrac{S_1^2}{S_2^2}$ as a test statistic for testing $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_A: \sigma_1^2 \neq \sigma_2^2$, the appropriate rejection region is

$$RR = \left\{ F > F^{n_1-1}_{n_2-1,\alpha/2} \quad or \quad F < \frac{1}{F^{n_2-1}_{n_2-1,\alpha/2}} \right\}$$

An equivalent test is obtained as follows. Let n_L and n_S denote the sample sizes associated with the larger and smaller sample variances, respectively. Place the larger sample variance in the numerator and the smaller sample variance in the denominator of the

F statistic, and reject $H_0: \sigma_1^2 = \sigma_2^2$ in favor of $H_A: \sigma_1^2 \neq \sigma_2^2$ if $F > F_{\alpha/2}$. where $F_{\alpha/2}$ is determined for $v_1 = n_L - 1$ and $v_1 = n_S - 1$ numerator and denominator degrees of freedom, respectively.

Example 10.21 An experiment to explore the pain thresholds to electrical shock for males and females resulted in the data summary given in Table below. Do the data provide sufficient evidence to indicate a significant difference in the variability of pain thresholds for men and women? Use $\alpha = 0.1$. What can be said about the *p-value*? (Assume that the pain thresholds for men and women are approximately normally distributed)

	Males	**Females**
n	14	10
$\bar{y}$	16.2	14.9
s^2	12.7	26.4

Solution

We desire to test $H_0: \sigma_M^2 = \sigma_F^2$ versus $H_A: \sigma_M^2 \neq \sigma_F^2$, where σ_M^2 and σ_F^2 are the variances of pain thresholds for men and women, respectively. The larger S^2 is 26.4 (the S^2 for women), and the sample size associated with the larger S^2 is $n_L=10$. The smaller S^2 is 2.7 (the S^2 for men), and $n_S=14$ (the number of men in the sample). Therefore, we compute

$$F = \frac{S_1^2}{S_2^2} = \frac{26.4}{12.7} = 2.079$$

and we compare this value to $F_{\alpha/2} = F_{0.05}$ with $v_1 = 10 - 1 = 9$ and $v_2 = 14 - 1 = 13$ numerator and denominator degrees of freedom, respectively. Because $F_{0.05} = 2.71$ and because 2.079 is not larger than the critical value (2.71), insufficient evidence exists to support a claim that the variability of pain thresholds for men and women differs.

The *p-value* associated with the observed value of F for this two-tailed test can be found as follows. Referring to F-Table, with $v_1 = 10 - 1 = 9$ numerator and $v_2 = 14 - 1 = 13$ denominator degrees of freedom, respectively, we find $F_{0.10}$=2.16. Thus, *p-value*>2(0.10)=0.20. Unless we were willing to work with a very large value of α (same value greater than .2), these results would not allow us to conclude that the variances of pain thresholds for men and women differ.