

- 9.1 Introduction
- 9.2 Testing $H_0: \mu_X = \mu_Y$
- 9.3 Testing $H_0: \sigma_X^2 = \sigma_Y^2$ —The F Test
- 9.4 Binomial Data: Testing $H_0: p_X = p_Y$
- 9.5 Confidence Intervals for the Two-Sample Problem
- 9.6 Taking a Second Look at Statistics (Choosing Samples)
- Appendix 9.A.1** A Derivation of the Two-Sample t Test (A Proof of Theorem 9.2.2)
- Appendix 9.A.2** Minitab Applications

After earning an Oxford degree in mathematics and chemistry, Gosset began working in 1899 for Messrs. Guinness, a Dublin brewery. Fluctuations in materials and temperature and the necessarily small-scale experiments inherent in brewing convinced him of the necessity for a new, small-sample theory of statistics. Writing under the pseudonym “Student,” he published work with the t ratio that was destined to become a cornerstone of modern statistical methodology.

—William Sealy Gosset (“Student”) (1876–1937)

9.1 Introduction

The simplicity of the one-sample model makes it the logical starting point for any discussion of statistical inference, but it also limits its applicability to the real world. Very few experiments involve just a single treatment or a single set of conditions. On the contrary, researchers almost invariably design experiments to compare responses to *several* treatment levels—or, at the very least, to compare a single treatment with a control.

In this chapter we examine the simplest of these multilevel designs, *two-sample inferences*. Structurally, two-sample inferences always fall into one of two different formats: Either two (presumably) different treatment levels are applied to two independent sets of similar subjects or the same treatment is applied to two (presumably) different kinds of subjects. Comparing the effectiveness of germicide A relative to that of germicide B by measuring the zones of inhibition each one produces in two sets of similarly cultured Petri dishes would be an example of the first type. On the other hand, examining the bones of sixty-year-old men and sixty-year-old women, all lifelong residents of the same city, to see whether both sexes absorb environmental strontium-90 at the same rate would be an example of the second type.

Inference in two-sample problems usually reduces to a comparison of *location* parameters. We might assume, for example, that the population of responses associated with, say, treatment X is normally distributed with mean μ_X and standard

deviation σ_X while the Y distribution is normal with mean μ_Y and standard deviation σ_Y . Comparing location parameters, then, reduces to testing $H_0: \mu_X = \mu_Y$. As always, the alternative may be either one-sided, $H_1: \mu_X < \mu_Y$ or $H_1: \mu_X > \mu_Y$, or two-sided, $H_1: \mu_X \neq \mu_Y$. (If the data are binomial, the location parameters are p_X and p_Y , the true “success” probabilities for treatments X and Y, and the null hypothesis takes the form $H_0: p_X = p_Y$.)

Sometimes, although much less frequently, it becomes more relevant to compare the *variabilities* of two treatments, rather than their locations. A food company, for example, trying to decide which of two types of machines to buy for filling cereal boxes would naturally be concerned about the *average* weights of the boxes filled by each type, but they would also want to know something about the *variabilities* of the weights. Obviously, a machine that produces high proportions of “underfills” and “overfills” would be a distinct liability. In a situation of this sort, the appropriate null hypothesis is $H_0: \sigma_X^2 = \sigma_Y^2$.

For comparing the means of two normal populations when $\sigma_X = \sigma_Y$, the standard procedure is the *two-sample t test*. As described in Section 9.2, this is a relatively straightforward extension of Chapter 7’s one-sample t test. If $\sigma_X \neq \sigma_Y$, an approximate t test is used. For comparing variances, though, it will be necessary to introduce a completely new test—this one based on the F distribution of Section 7.3. The binomial version of the two-sample problem, testing $H_0: p_X = p_Y$, is taken up in Section 9.4.

It was mentioned in connection with one-sample problems that certain inferences, for various reasons, are more aptly phrased in terms of confidence intervals rather than hypothesis tests. The same is true of two-sample problems. In Section 9.5, confidence intervals are constructed for the location *difference* of two populations, $\mu_X - \mu_Y$ (or $p_X - p_Y$), and the *variability quotient*, σ_X^2/σ_Y^2 .

9.2 Testing $H_0: \mu_X = \mu_Y$

We will suppose that the data for a given experiment consist of two independent random samples, X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m , representing either of the models referred to in Section 9.1. Furthermore, the two populations from which the X ’s and Y ’s are drawn will be presumed normal. Let μ_X and μ_Y denote their means. Our objective is to derive a procedure for testing $H_0: \mu_X = \mu_Y$.

As it turns out, the precise form of the test we are looking for depends on the variances of the X and Y populations. If it can be assumed that σ_X^2 and σ_Y^2 are equal, it is a relatively straightforward task to produce the GLRT for $H_0: \mu_X = \mu_Y$. (This is, in fact, what we will do in Theorem 9.2.2.) But if the variances of the two populations are *not* equal, the problem becomes much more complex. This second case, known as the Behrens-Fisher problem, is more than seventy-five years old and remains one of the more famous “unsolved” problems in statistics. What headway investigators *have* made has been confined to approximate solutions. These will be discussed later in this section. For what follows next, it can be assumed that $\sigma_X^2 = \sigma_Y^2$.

For the one-sample test $\mu = \mu_0$, the GLRT was shown to be a function of a special case of the t ratio introduced in Definition 7.3.3 (recall Theorem 7.3.5). We begin this section with a theorem that gives still another special case of Definition 7.3.3.

Theorem 9.2.1

Let X_1, X_2, \dots, X_n be a random sample of size n from a normal distribution with mean μ_X and standard deviation σ and let Y_1, Y_2, \dots, Y_m be an independent random sample of size m from a normal distribution with mean μ_Y and standard deviation σ .

Let S_X^2 and S_Y^2 be the two corresponding sample variances, and S_p^2 the pooled variance, where

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n+m-2}$$

Then

$$T_{n+m-2} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has a Student t distribution with $n+m-2$ degrees of freedom.

Proof The method of proof here is very similar to what was used for Theorem 7.3.5. Note that an equivalent formulation of T_{n+m-2} is

$$\begin{aligned} T_{n+m-2} &= \frac{\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{S_p^2 / \sigma^2}} \\ &= \frac{\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{1}{n+m-2} \left[\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \sum_{i=1}^m \left(\frac{Y_i - \bar{Y}}{\sigma} \right)^2 \right]}} \end{aligned}$$

But $E(\bar{X} - \bar{Y}) = \mu_X - \mu_Y$ and $\text{Var}(\bar{X} - \bar{Y}) = \sigma^2/n + \sigma^2/m$, so the numerator of the ratio has a standard normal distribution, $f_Z(z)$.

In the denominator,

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{(n-1)S_X^2}{\sigma^2}$$

and

$$\sum_{i=1}^m \left(\frac{Y_i - \bar{Y}}{\sigma} \right)^2 = \frac{(m-1)S_Y^2}{\sigma^2}$$

are independent χ^2 random variables with $n-1$ and $m-1$ df, respectively, so

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 + \sum_{i=1}^m \left(\frac{Y_i - \bar{Y}}{\sigma} \right)^2$$

has a χ^2 distribution with $n+m-2$ df (recall Theorem 7.3.1 and Theorem 4.6.4). Also, by Appendix 7.A.2, the numerator and denominator are independent. It follows from Definition 7.3.3, then, that

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has a Student t distribution with $n+m-2$ df. □

Theorem 9.2.2 Let x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m be independent random samples from normal distributions with means μ_X and μ_Y , respectively, and with the same standard deviation σ . Let

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- a. To test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X > \mu_Y$ at the α level of significance, reject H_0 if $t \geq t_{\alpha, n+m-2}$.
- b. To test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X < \mu_Y$ at the α level of significance, reject H_0 if $t \leq -t_{\alpha, n+m-2}$.
- c. To test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X \neq \mu_Y$ at the α level of significance, reject H_0 if t is either (1) $\leq -t_{\alpha/2, n+m-2}$ or (2) $\geq t_{\alpha/2, n+m-2}$.

Proof See Appendix 9.A.1. □

Case Study 9.2.1

The mystery surrounding the nature of Mark Twain's participation in the Civil War was discussed (but not resolved) in Case Study 1.2.2. Recall that historians are still unclear as to whether the creator of *Huckleberry Finn* and *Tom Sawyer* was a civilian or a combatant in the early 1860s and whether his sympathies lay with the North or with the South.

A tantalizing clue that might shed some light on the matter is a set of ten war-related essays written by one Quintus Curtius Snodgrass, who claimed to be in the Louisiana militia, although no records documenting his service have ever been found. If Snodgrass was just a pen name Twain used, as some suspect, then these essays are basically a diary of Twain's activities during the war, and the mystery is solved. If Quintus Curtius Snodgrass was *not* a pen name, these essays are just a red herring, and all questions about Twain's military activities remain unanswered.

Assessing the likelihood that Twain and Snodgrass were one and the same would be the job of a "forensic statistician." Authors have characteristic word-length profiles that effectively serve as verbal fingerprints (much like incriminating evidence left at a crime scene). If Authors A and B tend to use, say, three-letter words with significantly different frequencies, a reasonable inference would be that A and B are different people.

Table 9.2.1 shows the proportions of three-letter words in each of the ten Snodgrass essays and in eight essays known to have been written by Mark Twain. If x_i denotes the i th Twain proportion, $i = 1, 2, \dots, 8$, and y_i denotes the i th Snodgrass proportion, $i = 1, 2, \dots, 10$, then

$$\sum_{i=1}^8 x_i = 1.855 \text{ so } \bar{x} = 1.855/8 = 0.2319$$

(Continued on next page)

Table 9.2.1 Proportion of Three-Letter Words

Twain	Proportion	QCS	Proportion
Sergeant Fathom letter	0.225	Letter I	0.209
Madame Caprell letter	0.262	Letter II	0.205
Mark Twain letters in <i>Territorial Enterprise</i>		Letter III	0.196
First letter	0.217	Letter IV	0.210
Second letter	0.240	Letter V	0.202
Third letter	0.230	Letter VI	0.207
Fourth letter	0.229	Letter VII	0.224
First <i>Innocents Abroad</i> letter		Letter VIII	0.223
First half	0.235	Letter IX	0.220
Second half	0.217	Letter X	0.201

and

$$\sum_{i=1}^{10} y_i = 2.097 \text{ so } \bar{y} = 2.097/10 = 0.2097$$

The question to be answered is whether the difference between 0.2319 and 0.2097 is statistically significant.

Let μ_X and μ_Y denote the true average proportions of three-letter words that Twain and Snodgrass, respectively, tended to use. Our objective is to test

$$H_0: \mu_X = \mu_Y$$

versus

$$H_1: \mu_X \neq \mu_Y$$

Since

$$\sum_{i=1}^8 x_i^2 = 0.4316 \quad \text{and} \quad \sum_{i=1}^{10} y_i^2 = 0.4406$$

the two sample variances are

$$\begin{aligned} s_X^2 &= \frac{8(0.4316) - (1.855)^2}{8(7)} \\ &= 0.0002103 \end{aligned}$$

and

$$\begin{aligned} s_Y^2 &= \frac{10(0.4406) - (2.097)^2}{10(9)} \\ &= 0.0000955 \end{aligned}$$

(Continued on next page)

(Case Study 9.2.1 continued)

Combined, they give a pooled standard deviation of 0.0121:

$$\begin{aligned}
 s_p &= \sqrt{\frac{\sum_{i=1}^8 (x_i - 0.2319)^2 + \sum_{i=1}^{10} (y_i - 0.2097)^2}{n + m - 2}} \\
 &= \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n + m - 2}} \\
 &= \sqrt{\frac{7(0.0002103) + 9(0.0000955)}{8 + 10 - 2}} \\
 &= \sqrt{0.0001457} \\
 &= 0.0121
 \end{aligned}$$

According to Theorem 9.2.1, if $H_0: \mu_X = \mu_Y$ is true, the sampling distribution of

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{8} + \frac{1}{10}}}$$

is described by a Student t curve with 16 ($= 8 + 10 - 2$) degrees of freedom.

Suppose we let $\alpha = 0.01$. By part (c) of Theorem 9.2.2, H_0 should be rejected in favor of a two-sided H_1 if either (1) $t \leq -t_{\alpha/2, n+m-2} = -t_{0.005, 16} = -2.9208$ or (2) $t \geq t_{\alpha/2, n+m-2} = t_{0.005, 16} = 2.9208$ (see Figure 9.2.1). But

$$\begin{aligned}
 t &= \frac{0.2319 - 0.2097}{0.0121 \sqrt{\frac{1}{8} + \frac{1}{10}}} \\
 &= 3.88
 \end{aligned}$$

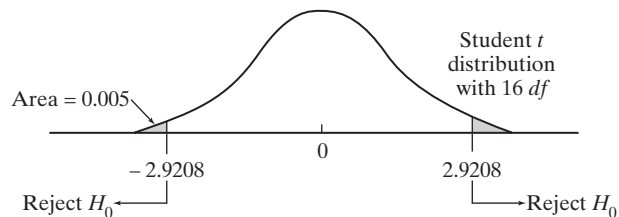


Figure 9.2.1

a value falling considerably to the right of $t_{0.005, 16}$. Therefore, we should *reject* H_0 —it appears that Twain and Snodgrass were not the same person. So, unfortunately, nothing that Twain did can be inferred from anything that Snodgrass wrote.

About the Data The X_i 's and Y_i 's in Table 9.2.1, being proportions, are necessarily *not* normally distributed random variables with the same variance, so the basic conditions of Theorem 9.2.2 are not met. Fortunately, the consequences of violated assumptions on the probabilistic behavior of T_{n+m-2} are frequently minimal. The

robustness property of the *one-sample t ratio* that we investigated in Chapter 7 also holds true for the *two-sample t ratio*.

Case Study 9.2.2

Dislike your statistics instructor? Retaliation time will come at the end of the semester, when you pepper the student course evaluation form with 1's. Were you pleased? Then send a signal with a load of 5's. Either way, students' evaluations of their instructors do matter. These instruments are commonly used for promotion, tenure, and merit raise decisions.

Studies of student course evaluations show that they do have value. They tend to show reliability and consistency. Yet questions remain as to the ability of these questionnaires to identify good teachers and courses.

A veteran instructor of developmental psychology decided to do a study (201) on how a single changed factor might affect his students' course evaluations. He had attended a workshop extolling the virtue of an enthusiastic style in the classroom—more hand gestures, increased voice pitch variability, and the like. The vehicle for the study was the large-lecture undergraduate developmental psychology course he had taught in the fall semester. He set about to teach the spring-semester offering in the same way, with the exception of a more enthusiastic style.

The professor fully understood the difficulty of controlling for the many variables. He selected the spring class to have the same demographics as the one in the fall. He used the same textbook, syllabus, and tests. He listened to audiotapes of the fall lectures and reproduced them as closely as possible, covering the same topics in the same order.

The first step in examining the effect of enthusiasm on course evaluations is to establish that students have, in fact, perceived an increase in enthusiasm. Table 9.2.2 summarizes the ratings the instructor received on the “enthusiasm” question for the two semesters. Unless the increase in sample means (2.14 to 4.21) is statistically significant, there is no point in trying to compare fall and spring responses to other questions.

Table 9.2.2

Fall, x_i	Spring, y_i
$n = 229$	$m = 243$
$\bar{x} = 2.14$	$\bar{y} = 4.21$
$s_X = 0.94$	$s_Y = 0.83$

Let μ_X and μ_Y denote the true means associated with the two different teaching styles. There is no reason to think that increased enthusiasm on the part of the instructor would *decrease* the students' perception of enthusiasm, so it can be argued here that H_1 should be one-sided. That is, we want to test

$$H_0: \mu_X = \mu_Y$$

versus

$$H_1: \mu_X < \mu_Y$$

(Continued on next page)

(Case Study 9.2.2 continued)

Let $\alpha = 0.05$.

Since $n = 229$ and $m = 243$, the t statistic has $229 + 243 - 2 = 470$ degrees of freedom. Thus, the decision rule calls for the rejection of H_0 if

$$t = \frac{\bar{x} - \bar{y}}{s_P \sqrt{\frac{1}{229} + \frac{1}{243}}} \leq -t_{\alpha, n+m-2} = -t_{0.05, 470}$$

A glance at Table A.2 in the Appendix shows that for any value $n > 100$, z_α is a good approximation of $t_{\alpha, n}$. That is, $-t_{0.05, 470} \doteq -z_{.05} = -1.64$.

The pooled standard deviation for these data is 0.885:

$$s_P = \sqrt{\frac{228(0.94)^2 + 242(0.83)^2}{229 + 243 - 2}} = 0.885$$

Therefore,

$$t = \frac{2.14 - 4.21}{0.885 \sqrt{\frac{1}{229} + \frac{1}{243}}} = -25.42$$

and our conclusion is a resounding rejection of H_0 —the increased enthusiasm was, indeed, noticed.

The real question of interest is whether the change in enthusiasm produced a *perceived* change in some other aspect of teaching that we know did *not* change. For example, the instructor did not become more knowledgeable about the material over the course of the two semesters. The student ratings, though, disagree.

Table 9.2.3 shows the instructor's fall and spring ratings on the "knowledgeable" question. Is the increase from $\bar{x} = 3.61$ to $\bar{y} = 4.05$ statistically significant? Yes. For these data, $s_P = 0.898$ and

$$t = \frac{3.61 - 4.05}{0.898 \sqrt{\frac{1}{229} + \frac{1}{243}}} = -5.33$$

which falls far to the left of the 0.05 critical value ($= -1.64$).

What we can glean from these data is both reassuring yet a bit disturbing. Table 9.2.2 appears to confirm the widely held belief that enthusiasm is an important factor in effective teaching. Table 9.2.3, on the other hand, strikes a more cautionary note. It speaks to another widely held belief—that student evaluations can sometimes be difficult to interpret. Questions that purport to be measuring one trait may, in fact, be reflecting something entirely different.

Table 9.2.3

Fall, x_i	Spring, y_i
$n = 229$	$m = 243$
$\bar{x} = 3.61$	$\bar{y} = 4.05$
$s_X = 0.84$	$s_Y = 0.95$

About the Data The five-choice responses in student evaluation forms are very common in survey questionnaires. Such questions are known as Likert items, named after the psychologist Rensis Likert. The item typically asks the respondent to choose his or her level of agreement with a statement, for example, “The instructor shows concern for students.” The choices start with “strongly disagree,” which is scored with a “1,” and go up to a “5” for “strongly agree.” The statistic for a given question in a survey is the *average* value taken over all responses.

Is a t test an appropriate way to analyze data of this sort? Maybe, but the nature of the responses raises some serious concerns. First of all, the fact that students talk with each other about their instructors suggests that not all the sample values will be independent. More importantly, the five-point Likert scale hardly resembles the normality assumption implicit in a Student t analysis. For many practitioners—but not all—the robustness of the t test would be enough to justify the analysis described in Case Study 9.2.2.

The Behrens-Fisher Problem

Finding a statistic with known density for testing the equality of two means from normally distributed random samples when the standard deviations of the samples are not equal is known as the *Behrens-Fisher problem*. No exact solution is known, but a widely used approximation is based on the test statistic

$$W = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

where, as usual, \bar{X} and \bar{Y} are the sample means, and S_X^2 and S_Y^2 are the unbiased estimators of the variance. B. L. Welch, a faculty member at University College, London, in a 1938 *Biometrika* article showed that W is approximately distributed as a Student t random variable with degrees of freedom given by the nonintuitive expression

$$\frac{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^2}{\frac{\sigma_1^4}{n_1^2(n_1-1)} + \frac{\sigma_2^4}{n_2^2(n_2-1)}}$$

To understand Welch’s approximation, it helps to rewrite the random variable W as

$$W = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \cdot \frac{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

In this form, the numerator is a standard normal variable. Suppose there is a chi square random variable V with ν degrees of freedom such that the square of the denominator is equal to V/ν . Then the expression would indeed be a Student t variable with ν degrees of freedom. However, in general, the denominator will not have exactly that distribution. The strategy, then, is to find an *approximate* equality for

$$\frac{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} = \frac{V}{\nu}$$

or, equivalently,

$$\frac{S_X^2}{n} + \frac{S_Y^2}{m} = \left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right) \frac{V}{v}$$

At issue is the value of v . The method of moments (recall Section 5.2) suggests a solution. If the means and variances of both sides are equated, it can be shown that

$$v = \frac{\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m} \right)^2}{\frac{\sigma_X^4}{n^2(n-1)} + \frac{\sigma_Y^4}{m^2(m-1)}}$$

Moreover, the expression for v depends only on the *ratio* of the variances, $\theta = \frac{\sigma_X^2}{\sigma_Y^2}$. To see why, divide the numerator and denominator by σ_Y^4 . Then

$$\frac{\left(\frac{1}{n} \frac{\sigma_X^2}{\sigma_Y^2} + \frac{1}{m} \right)^2}{\frac{1}{n^2(n-1)} \left(\frac{\sigma_X^2}{\sigma_Y^2} \right)^2 + \frac{1}{m^2(m-1)}} = \frac{\left(\frac{1}{n} \theta + \frac{1}{m} \right)^2}{\frac{1}{n^2(n-1)} \theta^2 + \frac{1}{m^2(m-1)}}$$

and multiplying numerator and denominator by n^2 gives the somewhat more appealing form

$$v = \frac{\left(\theta + \frac{n}{m} \right)^2}{\frac{1}{(n-1)} \theta^2 + \frac{1}{(m-1)} \left(\frac{n}{m} \right)^2}$$

Of course, the main application of this theory occurs when σ_X^2 and σ_Y^2 are unknown and θ must thus be estimated, the obvious choice being $\theta = \frac{s_X^2}{s_Y^2}$.

This leads us to the following theorem for testing the equality of means when the variances cannot be assumed equal.

**Theorem
9.2.3**

Let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m be independent random samples from normal distributions with means μ_X and μ_Y , and standard deviations σ_X and σ_Y , respectively. Let

$$W = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

Using $\hat{\theta} = \frac{s_X^2}{s_Y^2}$, take v to be the expression $\frac{\left(\hat{\theta} + \frac{n}{m} \right)^2}{\frac{1}{(n-1)} \hat{\theta}^2 + \frac{1}{(m-1)} \left(\frac{n}{m} \right)^2}$, rounded to the nearest integer. Then W has approximately a Student t distribution with v degrees of freedom.

Case Study 9.2.3

Does size matter? While a successful company's large number of sales should mean bigger profits, does it yield greater *profitability*? *Forbes* magazine periodically rates the top two hundred small companies (52), and for each gives the profitability as measured by the five-year percentage return on equity. Using data from the *Forbes* article, Table 9.2.4 gives the return on equity for the twelve companies with the largest number of sales (ranging from \$679 million to \$738

(Continued on next page)

million) and for the twelve companies with the smallest number of sales (ranging from \$25 million to \$66 million). Based on these data, can we say that the return on equity differs between the two types of companies?

Table 9.2.4

Large-Sales Companies	Return on Equity (%)	Small-Sales Companies	Return on Equity (%)
Deckers Outdoor	21	NVE	21
Jos. A. Bank Clothiers	23	Hi-Shear Technology	21
National Instruments	13	Bovie Medical	14
Dolby Laboratories	22	Rocky Mountain Chocolate Factory	31
Quest Software	7	Rochester Medical	19
Green Mountain Coffee Roasters	17	Anika Therapeutics	19
Lufkin Industries	19	Nathan's Famous	11
Red Hat	11	Somanetics	29
Matrix Service	2	Bolt Technology	20
DXP Enterprises	30	Energy Recovery	27
Franklin Electric	15	Transcend Services	27
LSB Industries	43	IEC Electronics	24

Let μ_X and μ_Y be the respective average returns on equity. The indicated test of hypotheses is

$$H_0: \mu_X = \mu_Y$$

versus

$$H_1: \mu_X \neq \mu_Y$$

For the data in the table, $\bar{x} = 18.6$, $\bar{y} = 21.9$, $s_X^2 = 115.9929$, and $s_Y^2 = 35.7604$. The test statistic is

$$w = \frac{\bar{x} - \bar{y} - (\mu_X - \mu_Y)}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} = \frac{18.6 - 21.9}{\sqrt{\frac{115.9929}{12} + \frac{35.7604}{12}}} = -0.928$$

Also,

$$\hat{\theta} = \frac{s_X^2}{s_Y^2} = \frac{115.9929}{35.7604} = 3.244$$

so

$$\frac{(3.244 + \frac{12}{12})^2}{\frac{1}{11}(3.244)^2 + \frac{1}{11}(\frac{12}{12})^2} = 17.2$$

which implies that $\nu = 17$.

We should reject H_0 at the $\alpha = 0.05$ level of significance if $w > t_{0.025, 17} = 2.1098$ or $w < -t_{0.025, 17} = -2.1098$. Here, $w = -0.928$ falls in between the two critical values, so the difference between \bar{x} and \bar{y} is not statistically significant.

Comment It occasionally happens that an experimenter wants to test $H_0: \mu_X = \mu_Y$ and *knows* the values of σ_X^2 and σ_Y^2 . For those situations, the t test of Theorem 9.2.2 is inappropriate. If the n X_i 's and m Y_i 's are normally distributed, it follows from the corollary to Theorem 4.3.3 that

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \tag{9.2.1}$$

has a standard normal distribution. Any such test of $H_0: \mu_X = \mu_Y$, then, should be based on an observed Z ratio rather than an observed t ratio.

If the degrees of freedom for a t test exceed 100, then the test statistic of Equation 9.2.1 is used, but it is treated as a Z ratio. In either the test of Theorem 9.2.2 or 9.2.3, if the degrees of freedom exceed 100, the statistic of Theorem 9.2.3 is used with the z tables.

Questions

9.2.1. Some states that operate a lottery believe that restricting the use of lottery profits to supporting education makes the lottery more profitable. Other states permit general use of the lottery income. The profitability of the lottery for a group of states in each category is given below.

State Lottery Profits			
For Education		For General Use	
State	% Profit	State	% Profit
New Mexico	24	Massachusetts	21
Idaho	25	Maine	22
Kentucky	28	Iowa	24
South Carolina	28	Colorado	27
Georgia	28	Indiana	27
Missouri	29	Dist. Columbia	28
Ohio	29	Connecticut	29
Tennessee	31	Pennsylvania	32
Florida	31	Maryland	32
California	35		
North Carolina	35		
New Jersey	35		

Source: *New York Times*, National Section, October 7, 2007, p. 14.

Test at the $\alpha = 0.01$ level whether the mean profit of states using the lottery for education is higher than that of states permitting general use. Assume that the variances of the two random variables are equal.

9.2.2. As the United States has struggled with the growing obesity of its citizens, diets have become big business. Among the many competing regimens for those seeking weight reduction are the Atkins and Zone diets. In a comparison of these two diets for one-year weight loss, a study (59) found that seventy-seven subjects on the Atkins diet had an average weight loss of $\bar{x} = -4.7$ kg and a sample standard deviation of $s_X = 7.05$ kg. Similar figures for the

seventy-nine people on the Zone diet were $\bar{y} = -1.6$ kg and $s_Y = 5.36$ kg. Is the greater reduction with the Atkins diet statistically significant? Test for $\alpha = 0.05$.

9.2.3. A medical researcher believes that women typically have lower serum cholesterol than men. To test this hypothesis, he took a sample of 476 men between the ages of nineteen and forty-four and found their mean serum cholesterol to be 189.0 mg/dl with a sample standard deviation of 34.2. A group of 592 women in the same age range averaged 177.2 mg/dl and had a sample standard deviation of 33.3. Is the lower average for the women statistically significant? Set $\alpha = 0.05$.

9.2.4. In the academic year 2004–05, 1126 high school freshmen took the SAT Reasoning Test. On the Critical Reasoning portion, this group had a mean score of 491 with a standard deviation of 119. The following year, 5042 sophomores (none of them in the 2004–05 freshmen group) scored an average of 498, with a standard deviation of 129. Is the higher average score for the sophomores a result of such factors as additional schooling and increased maturity or simply a random effect? Test at the $\alpha = 0.05$ level of significance.

Source: College Board SAT, Total Group Profile Report, 2008.

9.2.5. The University of Missouri–St. Louis gave a validation test to entering students who had taken calculus in high school. The group of ninety-three students receiving no college credit had a mean score of 4.17 on the validation test with a sample standard deviation of 3.70. For the twenty-eight students who received credit from a high school dual-enrollment class, the mean score was 4.61 with a sample standard deviation of 4.28. Is there a significant difference in these means at the $\alpha = 0.01$ level?

Source: MAA Focus, December 2008, p. 19.

9.2.6. Ring Lardner was one of this country's most popular writers during the 1920s and 1930s. He was also a

chronic alcoholic who died prematurely at the age of forty-eight. The following table lists the life spans of some of Lardner's contemporaries (36). Those in the sample on the left were all problem drinkers; they died, on the average, at age sixty-five. The twelve (sober) writers on the right tended to live a full ten years longer. Can it be argued that an increase of that magnitude is statistically significant? Test an appropriate null hypothesis against a one-sided H_1 . Use the 0.05 level of significance. (*Note:* The pooled sample standard deviation for these two samples is 13.9.)

<i>Authors Noted for Alcohol Abuse</i>		<i>Authors Not Noted for Alcohol Abuse</i>	
Name	Age at Death	Name	Age at Death
Ring Lardner	48	Carl Van Doren	65
Sinclair Lewis	66	Ezra Pound	87
Raymond Chandler	71	Randolph Bourne	32
Eugene O'Neill	65	Van Wyck Brooks	77
Robert Benchley	56	Samuel Eliot Morrison	89
J.P. Marquand	67	John Crowe Ransom	86
Dashiell Hammett	67	T.S. Eliot	77
e.e. cummings	70	Conrad Aiken	84
Edmund Wilson	77	Ben Ames Williams	64
Average:	65.2	Henry Miller	88
		Archibald MacLeish	90
		James Thurber	67
		Average:	75.5

9.2.7. Poverty Point is the name given to a number of widely scattered archaeological sites throughout Louisiana, Mississippi, and Arkansas. These are the remains of a society thought to have flourished during the period from 1700 to 500 B.C. Among their characteristic artifacts are ornaments that were fashioned out of clay and then baked. The following table shows the dates (in years B.C.) associated with four of these baked clay ornaments found in two different Poverty Point sites, Terral Lewis and Jacketown (86). The averages for the two samples are 1133.0 and 1013.5, respectively. Is it believable that these two settlements developed the technology to manufacture baked clay ornaments at the same time? Set up and test an appropriate H_0 against a two-sided H_1 at the $\alpha = 0.05$ level of significance. For these data $s_x = 266.9$ and $s_y = 224.3$.

Terral Lewis Estimates, x_i	Jaketown Estimates, y_i
1492	1346
1169	942
883	908
988	858

9.2.8. A major source of "mercury poisoning" comes from the ingestion of methylmercury (CH_3^{203}), which is

found in contaminated fish (recall Question 5.3.3). Among the questions pursued by medical investigators trying to understand the nature of this particular health problem is whether methylmercury is equally hazardous to men and women. The following (114) are the half-lives of methylmercury in the systems of six women and nine men who volunteered for a study where each subject was given an oral administration of CH_3^{203} . Is there evidence here that women metabolize methylmercury at a different rate than men do? Do an appropriate two-sample t test at the $\alpha = 0.01$ level of significance. The two sample standard deviations for these data are $s_x = 15.1$ and $s_y = 8.1$.

Methylmercury (CH_3^{203}) Half-Lives (in Days)	
Females, x_i	Males, y_i
52	72
69	88
73	87
88	74
87	78
56	70
	78
	93
	74

9.2.9. Lipton, a company primarily known for tea, considered using coupons to stimulate sales of its packaged dinner entrees. The company was particularly interested whether there was a difference in the effect of coupons on singles versus married couples. A poll of consumers asked them to respond to the question "Do you use coupons regularly?" by a numerical scale, where 1 stands for agree strongly, 2 for agree, 3 for neutral, 4 for disagree, and 5 for disagree strongly. The results of the poll are given in the following table (19).

<i>Use Coupons Regularly</i>	
Single (X)	Married (Y)
$n = 31$	$n = 57$
$\bar{x} = 3.10$	$\bar{y} = 2.43$
$s_x = 1.469$	$s_y = 1.350$

Is the observed difference significant at the $\alpha = 0.05$ level?

9.2.10. A company markets two brands of latex paint—regular and a more expensive brand that claims to dry an hour faster. A consumer magazine decides to test this claim by painting ten panels with each product. The average drying time of the regular brand is 2.1 hours with a sample standard deviation of 12 minutes. The fast-drying version has an average of 1.6 hours with a sample standard deviation of 16 minutes. Test the null hypothesis that the more expensive brand dries an hour quicker. Use a one-sided H_1 . Let $\alpha = 0.05$.

9.2.11. (a) Suppose $H_0: \mu_X = \mu_Y$ is to be tested against $H_1: \mu_X \neq \mu_Y$. The two sample sizes are 6 and 11. If $s_p = 15.3$, what is the smallest value for $|\bar{x} - \bar{y}|$ that will result in H_0 being rejected at the $\alpha = 0.01$ level of significance?

(b) What is the smallest value for $\bar{x} - \bar{y}$ that will lead to the rejection of $H_0: \mu_X = \mu_Y$ in favor of $H_1: \mu_X > \mu_Y$ if $\alpha = 0.05$, $s_p = 214.9$, $n = 13$, and $m = 8$?

9.2.12. Suppose that $H_0: \mu_X = \mu_Y$ is being tested against $H_1: \mu_X \neq \mu_Y$, where σ_X^2 and σ_Y^2 are known to be 17.6 and 22.9, respectively. If $n = 10$, $m = 20$, $\bar{x} = 81.6$, and $\bar{y} = 79.9$, what P -value would be associated with the observed Z ratio?

9.2.13. An executive has two routes that she can take to and from work each day. The first is by interstate; the second requires driving through town. On the average it takes her 33 minutes to get to work by the interstate and 35 minutes by going through town. The standard deviations for the two routes are 6 and 5 minutes, respectively. Assume the distributions of the times for the two routes are approximately normally distributed.

- (a)** What is the probability that on a given day, driving through town would be the quicker of her choices?
- (b)** What is the probability that driving through town for an entire week (ten trips) would yield a lower average time than taking the interstate for the entire week?

9.2.14. Prove that the Z ratio given in Equation 9.2.1 has a standard normal distribution.

9.2.15. If X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m are independent random samples from normal distributions with the same σ^2 , prove that their pooled sample variance, s_p^2 , is an unbiased estimator for σ^2 .

9.2.16. Let X_1, X_2, \dots, X_n and Y_1, Y_2, \dots, Y_m be independent random samples drawn from normal distributions with means μ_X and μ_Y , respectively, and with the same known variance σ^2 . Use the generalized likelihood ratio criterion to derive a test procedure for choosing between $H_0: \mu_X = \mu_Y$ and $H_1: \mu_X \neq \mu_Y$.

9.2.17. A person exposed to an infectious agent, either by contact or by vaccination, normally develops antibodies to that agent. Presumably, the severity of an infection is related to the number of antibodies produced. The degree of antibody response is indicated by saying that the person's blood serum has a certain *titer*, with higher titers indicating greater concentrations of antibodies. The following table gives the titers of twenty-two persons involved in a tularemia epidemic in Vermont (18). Eleven were quite ill; the other eleven were asymptomatic. Use an approximate t ratio to test $H_0: \mu_X = \mu_Y$ against a one-sided H_1 at the 0.05 level of significance.

The sample standard deviations for the "Severely Ill" and "Asymptomatic" groups are 428 and 183, respectively.

<i>Severely Ill</i>		<i>Asymptomatic</i>	
Subject	Titer	Subject	Titer
1	640	12	10
2	80	13	320
3	1280	14	320
4	160	15	320
5	640	16	320
6	640	17	80
7	1280	18	160
8	640	19	10
9	160	20	640
10	320	21	160
11	160	22	320

9.2.18. For the approximate two-sample t test described in Question 9.2.17, it will be true that

$$v < n + m - 2$$

Why is that a disadvantage for the approximate test? That is, why is it better to use the Theorem 9.2.1 version of the t test if, in fact, $\sigma_X^2 = \sigma_Y^2$?

9.2.19. The two-sample data described in Question 8.2.2 would be analyzed by testing $H_0: \mu_X = \mu_Y$, where μ_X and μ_Y denote the true average motorcycle-related fatality rates for states having "limited" and "comprehensive" helmet laws, respectively.

- (a)** Should the t test for $H_0: \mu_X = \mu_Y$ follow the format of Theorem 9.2.2 or the approximation given in Theorem 9.2.3? Explain.
- (b)** Is there anything unusual about these data? Explain.

9.2.20. Some financial analysts believe that the election of a Republican president is good for the stock market. To test this claim, one study (155) recorded the ten-year growth in Standard & Poor's index following each election of a new president. The results are given in the table below.

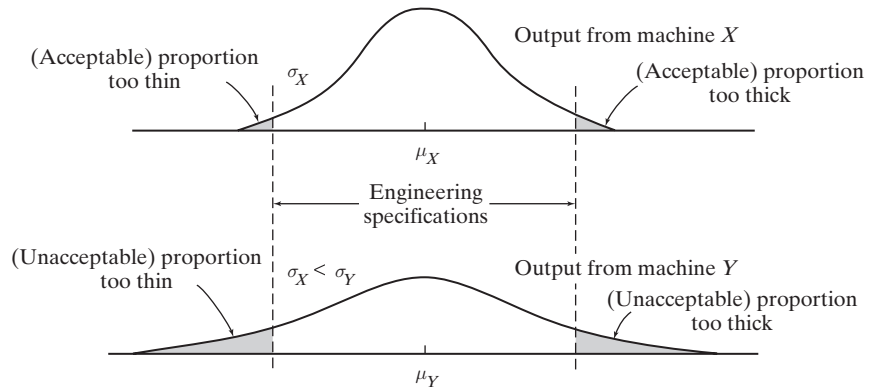
Democrats		Republicans	
Winner	S&P Growth	Winner	S&P Growth
Roosevelt '36	22.4	Eisenhower '52	45.7
Roosevelt '40	24.0	Eisenhower '56	28.6
Roosevelt '44	38.0	Nixon '68	14.2
Truman '48	45.7	Nixon '72	18.8
Kennedy '60	21.2	Reagan '80	50.3
Johnson '64	17.9	Reagan '84	40.1
Carter '76	38.2	Bush '88	52.4
Clinton '92	33.7		
Clinton '96	23.8		

Is the higher average for the Republicans statistically significant? Test at the 0.01 level. Do not assume the variances are equal.

9.3 Testing $H_0: \sigma_X^2 = \sigma_Y^2$ —The F Test

Although by far the majority of two-sample problems are set up to detect possible shifts in location parameters, situations sometimes arise where it is equally important—perhaps even more important—to compare variability parameters. Two machines on an assembly line, for example, may be producing items whose *average* dimensions (μ_X and μ_Y) of some sort—say, thickness—are not significantly different but whose variabilities (as measured by σ_X^2 and σ_Y^2) are. This becomes a critical piece of information if the increased variability results in an unacceptable proportion of items from one of the machines falling outside the engineering specifications (see Figure 9.3.1).

Figure 9.3.1 Variability of machine outputs.



In this section we will examine the generalized likelihood ratio test of $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_1: \sigma_X^2 \neq \sigma_Y^2$. The data will consist of two independent random samples of sizes n and m : The first— x_1, x_2, \dots, x_n —is assumed to have come from a normal distribution having mean μ_X and variance σ_X^2 ; the second— y_1, y_2, \dots, y_m —from a normal distribution having mean μ_Y and variance σ_Y^2 . (All four parameters are assumed to be unknown.) Theorem 9.3.1 gives the test procedure that will be used. The proof will not be given, but it follows the same basic pattern we have seen in other GLRTs; the important step is showing that the likelihood ratio is a monotonic function of the F random variable described in Definition 7.3.2.

Comment Tests of $H_0: \sigma_X^2 = \sigma_Y^2$ arise in another, more routine context. Recall that the procedure for testing the equality of μ_X and μ_Y depends on whether or not the two population variances are equal. This implies that a test of $H_0: \sigma_X^2 = \sigma_Y^2$ should precede every test of $H_0: \mu_X = \mu_Y$. If the former is accepted, the t test on μ_X and μ_Y is done according to Theorem 9.2.2; but if $H_0: \sigma_X^2 = \sigma_Y^2$ is rejected, Theorem 9.2.2 is not entirely appropriate. A frequently used alternative in that case is the approximate t test described in Theorem 9.2.3.

Theorem 9.3.1

Let x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m be independent random samples from normal distributions with means μ_X and μ_Y and standard deviations σ_X and σ_Y , respectively.

- a. To test $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_1: \sigma_X^2 > \sigma_Y^2$ at the α level of significance, reject H_0 if $s_Y^2/s_X^2 \leq F_{\alpha, m-1, n-1}$.

- b. To test $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_1: \sigma_X^2 < \sigma_Y^2$ at the α level of significance, reject H_0 if $s_Y^2/s_X^2 \geq F_{1-\alpha, m-1, n-1}$.
- c. To test $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_1: \sigma_X^2 \neq \sigma_Y^2$ at the α level of significance, reject H_0 if s_Y^2/s_X^2 is either (1) $\leq F_{\alpha/2, m-1, n-1}$ or (2) $\geq F_{1-\alpha/2, m-1, n-1}$.

Comment The GLRT described in Theorem 9.3.1 is *approximate* for the same sort of reason the GLRT for $H_0: \sigma^2 = \sigma_0^2$ is approximate (see Theorem 7.5.2). The distribution of the test statistic, S_Y^2/S_X^2 , is not symmetric, and the two ranges of variance ratios yielding λ 's less than or equal to λ^* (i.e., the left tail and right tail of the critical region) have slightly different areas. For the sake of convenience, though, it is customary to choose the two critical values so that each cuts off the same area, $\alpha/2$.

Case Study 9.3.1

Electroencephalograms are records showing fluctuations of electrical activity in the brain. Among the several different kinds of brain waves produced, the dominant ones are usually *alpha* waves. These have a characteristic frequency of anywhere from eight to thirteen cycles per second.

The objective of the experiment described in this example was to see whether sensory deprivation over an extended period of time has any effect on the alpha-wave pattern. The subjects were twenty inmates in a Canadian prison who were randomly split into two equal-sized groups. Members of one group were placed in solitary confinement; those in the other group were allowed to remain in their own cells. Seven days later, alpha-wave frequencies were measured for all twenty subjects (60), as shown in Table 9.3.1.

Table 9.3.1 Alpha-Wave Frequencies (CPS)

Nonconfined, x_i	Solitary Confinement, y_i
10.7	9.6
10.7	10.4
10.4	9.7
10.9	10.3
10.5	9.2
10.3	9.3
9.6	9.9
11.1	9.5
11.2	9.0
10.4	10.9

Judging from Figure 9.3.2, there was an apparent *decrease* in alpha-wave frequency for persons in solitary confinement. There also appears to have been an *increase* in the variability for that group. We will use the F test to determine whether the observed difference in variability ($s_X^2 = 0.21$ versus $s_Y^2 = 0.36$) is statistically significant.

(Continued on next page)

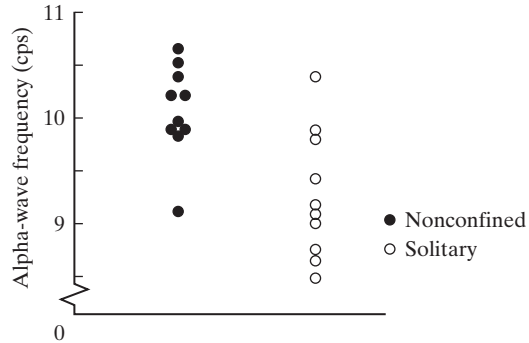


Figure 9.3.2 Alpha-wave frequencies (cps).

Let σ_X^2 and σ_Y^2 denote the true variances of alpha-wave frequencies for nonconfined and solitary-confined prisoners, respectively. The hypotheses to be tested are

$$H_0: \sigma_X^2 = \sigma_Y^2$$

versus

$$H_1: \sigma_X^2 \neq \sigma_Y^2$$

Let $\alpha = 0.05$ be the level of significance. Given that

$$\begin{aligned} \sum_{i=1}^{10} x_i &= 105.8 & \sum_{i=1}^{10} x_i^2 &= 1121.26 \\ \sum_{i=1}^{10} y_i &= 97.8 & \sum_{i=1}^{10} y_i^2 &= 959.70 \end{aligned}$$

the sample variances become

$$s_X^2 = \frac{10(1121.26) - (105.8)^2}{10(9)} = 0.21$$

and

$$s_Y^2 = \frac{10(959.70) - (97.8)^2}{10(9)} = 0.36$$

Dividing the sample variances gives an observed F ratio of 1.71:

$$F = \frac{s_Y^2}{s_X^2} = \frac{0.36}{0.21} = 1.71$$

Both n and m are ten, so we would expect S_Y^2/S_X^2 to behave like an F random variable with nine and nine degrees of freedom (assuming $H_0: \sigma_X^2 = \sigma_Y^2$ is true). From Table A.4 in the Appendix, we see that the values cutting off areas of 0.025 in either tail of that distribution are 0.248 and 4.03 (see Figure 9.3.3).

Since the observed F ratio falls between the two critical values, our decision is to *fail to reject* H_0 —a ratio of sample variances equal to 1.71 does not rule out

(Continued on next page)

(Case Study 9.3.1 continued)

the possibility that the two true variances are equal. (In light of the Comment preceding Theorem 9.3.1, it would now be appropriate to test $H_0: \mu_X = \mu_Y$ using the two-sample t test described in Section 9.2.)

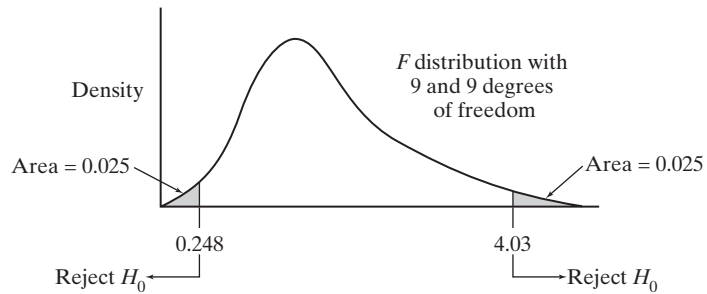


Figure 9.3.3 Distribution of S_Y^2/S_X^2 when H_0 is true.

Questions

9.3.1. Case Study 9.2.3 was offered as an example of testing means when the variances are not assumed equal. Was this a correct assumption about the variances? Test at the 0.05 level of significance.

9.3.2. Two popular forms of mortgage are the thirty-year fixed-rate mortgage, where the borrower has thirty years to repay the loan at a constant rate, and the adjustable-rate mortgage (ARM), one version of which is for five years with the possibility of yearly changes in the interest rate. Since the ARM offers less certainty, its rates are usually lower than those of fixed-rate mortgages. However, such vehicles should show more variability in rates. Test this hypothesis at the 0.10 level of significance using the following samples of mortgage offerings for a loan of \$160,000 (the borrower needs \$200,000, but must pay \$40,000 up front).

<i>\$160,000 Mortgage Rates</i>	
30-Year Fixed	ARM
5.500	3.875
5.500	5.125
5.250	5.000
5.125	4.750
5.875	4.375
5.625	
5.250	
4.875	

in which a subject is shown a series of pictures and is asked to make up a story about each one. Interpreted properly, the content of the stories can provide valuable insights into the subject's mental well-being. The following data show the TAT results for 40 women, 20 of whom were the mothers of normal children and 20 the mothers of schizophrenic children. In each case the subject was shown the same set of 10 pictures. The figures recorded were the numbers of stories (out of 10) that revealed a *positive* parent-child relationship, one where the mother was clearly capable of interacting with her child in a flexible, open-minded way (199).

	<i>TAT Scores</i>									
	Mothers of Normal Children					Mothers of Schizophrenic Children				
8	4	6	3	1	2	1	1	3	2	
4	4	6	4	2	7	2	1	3	1	
2	1	1	4	3	0	2	4	2	3	
3	2	6	3	4	3	0	1	2	2	

- (a) Test $H_0: \sigma_X^2 = \sigma_Y^2$ versus $H_1: \sigma_X^2 \neq \sigma_Y^2$, where σ_X^2 and σ_Y^2 are the variances of the scores of mothers of normal children and scores of mothers of schizophrenic children, respectively. Let $\alpha = 0.05$.
- (b) If $H_0: \sigma_X^2 = \sigma_Y^2$ is accepted in part (a), test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X \neq \mu_Y$. Set α equal to 0.05.

9.3.3. Among the standard personality inventories used by psychologists is the thematic apperception test (TAT)

9.3.4. In a study designed to investigate the effects of a strong magnetic field on the early development of mice