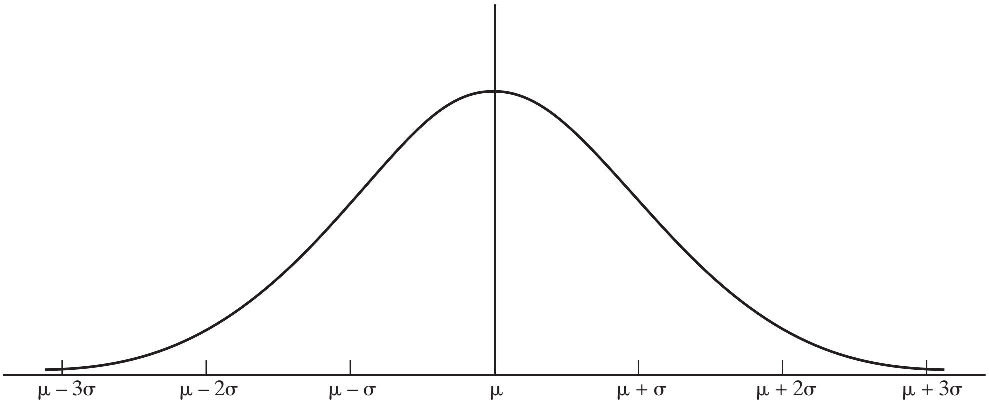
Chapter 7

INFERENCES BASED ON THE NORMAL DISTRIBUTION

CHAPTER OUTLINE



I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the "law of frequency of error" (the normal distribution). The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self effacement amidst the wildest confusion. The huger the mob, and the greater the anarchy, the more perfect is its sway. It is the supreme law of Unreason.

—Francis Galton

7.1 INTRODUCTION

Finding probability distributions to describe—and, ultimately, to predict—empirical data is one of the most important contributions a statistician can make to the research scientist. Already we have seen a number of functions playing that role. The binomial is an obvious model for the number of correct responses in the Pratt-Woodruff ESP experiment (Case Study 4.3.1); the probability of holding a winning ticket in the Florida Lottery is given by the hypergeometric (Example 3.2.6); and applications of the Poisson have run the gamut from radioactive decay (Case Study 4.2.2) to the number of wars starting in a given year (Case Study 4.2.3). Those examples notwithstanding, by far the most widely used probability model in statistics is the *normal* (or *Gaussian*) distribution,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(y-\mu)/\sigma]^2}, \quad -\infty < y < \infty \tag{7.1.1}$$

Some of the history surrounding the normal curve has already been discussed in Chapter 4—how it first appeared as a limiting form of the binomial, but then soon found itself used most often in nonbinomial situations. We also learned how to find areas under normal curves and did some problems involving sums and averages. Chapter 5 provided estimates of the parameters of the normal density and showed their role in fitting normal curves to data. In this chapter, we will take a second look at the properties and applications of this singularly important pdf, this time paying attention to the part it plays in estimation and hypothesis testing.

7.2 Comparing $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$

Suppose that a random sample of n measurements, $Y_1, Y_2, \ldots, Y_n$, is to be taken on a trait that is thought to be normally distributed, the objective being to draw an inference about the underlying pdf's true mean, μ. *If the variance* σ^2 *is known*, we already know how to proceed: A decision rule for testing $H_0: \mu = \mu_0$ is given in Theorem 6.2.1, and the construction of a confidence interval for μ is described in Section 5.3. As we learned, both of those procedures are based on the fact that the ratio $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ has a standard normal distribution, $f_Z(z)$.

In practice, though, the parameter σ^2 is seldom known, so the ratio $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ cannot be calculated, even if a value for the mean—say, μ_0—is substituted for μ. Typically, the only information experimenters have about σ^2 is what can be gleaned from the Y_i's themselves. The usual estimator for the population variance, of course, is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$, the unbiased version of the maximum likelihood estimator for σ^2. The question is, what effect does replacing σ with S have on the Z ratio? Are there probabilistic differences between $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$?

Historically, many early practitioners of statistics felt that replacing σ with S had, in fact, no effect on the distribution of the Z ratio. Sometimes they were right. If the sample size is very large (which was a common state of affairs in many of the early applications of statistics), the estimator S is essentially a constant and for all intents and purposes equal to the true σ. Under those conditions, the ratio $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ *will* behave much like a standard normal random variable, Z. When the sample size n is small, though, replacing σ with S *does* matter, and it changes the way we draw inferences about μ.

Credit for recognizing that $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ do not have the same distribution goes to William Sealy Gossett. After graduating in 1899 from Oxford with First Class degrees in Chemistry and Mathematics, Gossett took a position at Arthur Guinness, Son & Co., Ltd., a firm that brewed a thick, dark ale known as stout. Given the task of making the art of brewing more scientific, Gossett quickly realized that any experimental studies would necessarily face two obstacles. First, for a variety of economic and logistical reasons, sample sizes would invariably be small; and second, there would never be any way to know the exact value of the true variance, σ^2, associated with any set of measurements.

So, when the objective of a study was to draw an inference about μ, Gossett found himself working with the ratio $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$, where n was often on the order of four or five. The more he encountered that situation, the more he became convinced that ratios of that sort are *not* adequately described by the standard normal pdf. In particular, the distribution of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ seemed to have the same general bell-shaped configuration as $f_Z(z)$, but the tails were "thicker"—that is, ratios much smaller than zero or much greater than zero were not as rare as the standard normal pdf would predict.

Figure 7.2.1 illustrates the distinction between the distributions of $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ that caught Gossett's attention. In Figure 7.2.1a, five hundred samples of size $n = 4$ have been drawn from a normal distribution where the value of σ is known. For each sample, the ratio $\frac{\overline{Y}-\mu}{\sigma/\sqrt{4}}$ has been computed. Superimposed over the shaded histogram of those five hundred ratios is the standard normal curve, $f_Z(z)$. Clearly, the probabilistic behavior of the random variable $\frac{\overline{Y}-\mu}{\sigma/\sqrt{4}}$ is entirely consistent with $f_Z(z)$.

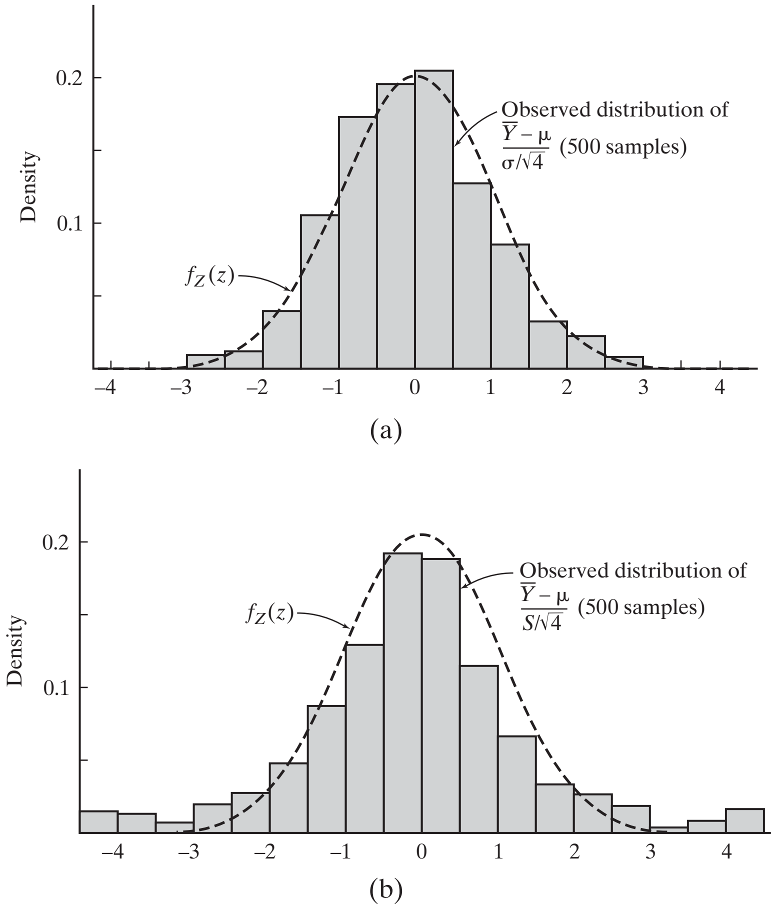


Figure 7.2.1

The histogram pictured in Figure 7.2.1b is also based on five hundred samples of size $n = 4$ drawn from a normal distribution. Here, though, S has been calculated for each sample, so the ratios comprising the histogram are $\frac{\overline{Y}-\mu}{S/\sqrt{4}}$ rather than $\frac{\overline{Y}-\mu}{\sigma/\sqrt{4}}$. In this case, the superimposed standard normal pdf does *not* adequately describe the histogram—specifically, it underestimates the number of ratios much less than zero as well as the number much larger than zero (which is exactly what Gossett had noted).

Gossett published a paper in 1908 entitled "The Probable Error of a Mean," in which he derived a formula for the pdf of the ratio $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$. To prevent disclosure of confidential company information, Guinness prohibited its employees from publishing any papers, regardless of content. So, Gossett's work, one of the major statistical breakthroughs of the twentieth century, was published under the name "Student."

Initially, Gossett's discovery attracted very little attention. Virtually none of his contemporaries had the slightest inkling of the impact that Gossett's paper would have on modern statistics. Indeed, fourteen years after its publication, Gossett sent R.A. Fisher a tabulation of his distribution, with a note saying, "I am sending you a copy of Student's Tables as you are the only man that's ever likely to use them."

Fisher very much understood the value of Gossett's work and believed that Gossett had effected a "logical revolution." Fisher presented a rigorous mathematical derivation of Gossett's pdf in 1924, the core of which appears in Appendix 7.A.1. Fisher somewhat arbitrarily chose the letter t for the $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ statistic. Consequently, its pdf is known as the *Student t distribution*.

7.3 Deriving the Distribution of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$

Broadly speaking, the set of probability functions that statisticians have occasion to use fall into two categories. There are a dozen or so that can effectively model the *individual* measurements taken on a variety of real-world phenomena. These are the distributions we studied in Chapters 3 and 4—most notably, the normal, binomial, Poisson, exponential, hypergeometric, and uniform. There is a smaller set of probability distributions that model the behavior of *functions* based on sets of n random variables. These are called *sampling distributions*, and they are typically used for inference purposes.

The normal distribution belongs to both categories. We have seen a number of scenarios (IQ scores, for example) where the Gaussian distribution is very effective at describing the distribution of repeated measurements. At the same time, the normal distribution is used to model the probabilistic behavior of $Z = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$. In the latter capacity, it serves as a sampling distribution.

Next to the normal distribution, the three most important sampling distributions are the *Student t distribution,* the *chi square distribution,* and the *F distribution*. All three will be introduced in this section, partly because we need the latter two to derive $f_T(t)$, the pdf for the t ratio, $T = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$. So, although our primary objective in this section is to study the Student t distribution, we will in the process introduce the two other sampling distributions that we will be encountering over and over again in the chapters ahead.

Deriving the pdf for a t ratio is not a simple matter. That may come as a surprise, given that deducing the pdf for $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is quite easy (using moment-generating functions). But going from $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ to $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ creates some major mathematical complications because T (unlike Z) is the ratio of *two* random variables, $\overline{Y}$ and S, both of which are functions of n random variables, $Y_1, Y_2, \ldots, Y_n$. In general—and this ratio is no exception—finding pdfs of quotients of random variables is difficult, especially when the numerator and denominator random variables have cumbersome pdfs to begin with.

As we will see in the next few pages, the derivation of $f_T(t)$ plays out in several steps. First, we show that $\sum_{j=1}^{m} Z_j^2$, where the Z_j's are independent standard normal random variables, has a gamma distribution (more specifically, a special case of the gamma distribution, called a *chi square distribution*). Then we show that $\overline{Y}$ and S^2, based on a random sample of size n from a normal distribution, are independent random variables and that $\frac{(n-1)S^2}{\sigma^2}$ has a chi square distribution. Next we derive the pdf of the ratio of two independent chi square random variables (which is called the *F distribution*). The final step in the proof is to show that $T^2 = \left(\frac{\overline{Y}-\mu}{S/\sqrt{n}}\right)^2$ can be written as the quotient of two independent chi square random variables, making it a special case of the *F* distribution. Knowing the latter allows us to deduce $f_T(t)$.

Theorem 7.3.1 *Let $U = \sum_{j=1}^{m} Z_j^2$, where $Z_1, Z_2, \ldots, Z_m$ are independent standard normal random variables. Then U has a gamma distribution with $r = \frac{m}{2}$ and $\lambda = \frac{1}{2}$. That is,*

$$f_U(u) = \frac{1}{2^{m/2}\Gamma\left(\frac{m}{2}\right)} u^{(m/2)-1} e^{-u/2}, \quad u \geq 0$$

Proof First take $m = 1$. For any $u \geq 0$,

$$F_{Z^2}(u) = P(Z^2 \leq u) = P\left(-\sqrt{u} \leq Z \leq \sqrt{u}\right) = 2P\left(0 \leq Z \leq \sqrt{u}\right)$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{u}} e^{-z^2/2}\, dz$$

Differentiating both sides of the equation for $F_{Z^2}(u)$ gives $f_{Z^2}(u)$:

$$f_{Z^2}(u) = \frac{d}{du} F_{Z^2}(u) = \frac{2}{\sqrt{2\pi}} \frac{1}{2\sqrt{u}} e^{-u/2} = \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)} u^{(1/2)-1} e^{-u/2}$$

Notice that $f_U(u) = f_{Z^2}(u)$ has the form of a gamma pdf with $r = \frac{1}{2}$ and $\lambda = \frac{1}{2}$. By Theorem 4.6.4, then, the sum of m such squares has the stated gamma distribution with $r = m\left(\frac{1}{2}\right) = \frac{m}{2}$ and $\lambda = \frac{1}{2}$.

The distribution of the sum of squares of independent standard normal random variables is sufficiently important that it gets its own name, despite the fact that it represents nothing more than a special case of the gamma distribution.

Definition 7.3.1
The pdf of $U = \sum_{j=1}^{m} Z_j^2$, where $Z_1, Z_2, \ldots, Z_m$ are independent standard normal random variables, is called the *chi square distribution with m degrees of freedom*.

The next theorem is especially critical in the derivation of $f_T(t)$. Using simple algebra, it can be shown that the square of a t ratio can be written as the quotient of two chi square random variables, one a function of $\overline{Y}$ and the other a function of S^2. By showing that $\overline{Y}$ and S^2 are independent (as Theorem 7.3.2 does), Theorem 3.8.4 can be used to find an expression for the pdf of the quotient.

Theorem 7.3.2 *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then*

a. *S^2 and $\overline{Y}$ are independent.*

b. $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ *has a chi square distribution with $n - 1$ degrees of freedom.*

Proof See Appendix 7.A.1

As we will see shortly, the square of a t ratio is a special case of an F random variable. The next definition and theorem summarize the properties of the F distribution that we will need to find the pdf associated with the Student t distribution.

Definition 7.3.2
Suppose that U and V are independent chi square random variables with n and m degrees of freedom, respectively. A random variable of the form $\frac{V/m}{U/n}$ is said to have an F *distribution with m and n degrees of freedom*.

Comment The F in the name of this distribution commemorates the renowned statistician Sir Ronald Fisher.

Theorem 7.3.3 *Suppose $F_{m,n} = \frac{V/m}{U/n}$ denotes an F random variable with m and n degrees of freedom. The pdf of $F_{m,n}$ has the form*

$$f_{F_{m,n}}(w) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2} w^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(n+mw)^{(m+n)/2}}, \quad w \geq 0$$

Proof We begin by finding the pdf for V/U. From Theorem 7.3.1 we know that $f_V(v) = \frac{1}{2^{m/2}\Gamma(m/2)} v^{(m/2)-1} e^{-v/2}$ and $f_U(u) = \frac{1}{2^{n/2}\Gamma(n/2)} u^{(n/2)-1} e^{-u/2}$.

From Theorem 3.8.4, we have that the pdf of $W = V/U$ is

$$\begin{aligned} f_{V/U}(w) &= \int_0^\infty |u| f_U(u) f_V(uw)\, du \\ &= \int_0^\infty u \frac{1}{2^{n/2}\Gamma(n/2)} u^{(n/2)-1} e^{-u/2} \frac{1}{2^{m/2}\Gamma(m/2)} (uw)^{(m/2)-1} e^{-uw/2}\, du \\ &= \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \int_0^\infty u^{\frac{n+m}{2}-1} e^{-[(1+w)/2]u}\, du \end{aligned}$$

The integrand is the variable part of a gamma density with $r = (n+m)/2$ and $\lambda = (1+w)/2$. Thus, the integral equals the inverse of the density's constant. This gives

$$f_{V/U} = \frac{1}{2^{(n+m)/2}\Gamma(n/2)\Gamma(m/2)} w^{(m/2)-1} \frac{\Gamma\left(\frac{n+m}{2}\right)}{[(1+w)/2]^{\frac{n+m}{2}}} = \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma(n/2)\Gamma(m/2)} \frac{w^{(m/2)-1}}{(1+w)^{\frac{n+m}{2}}}$$

The statement of the theorem, then, follows from Theorem 3.8.2:

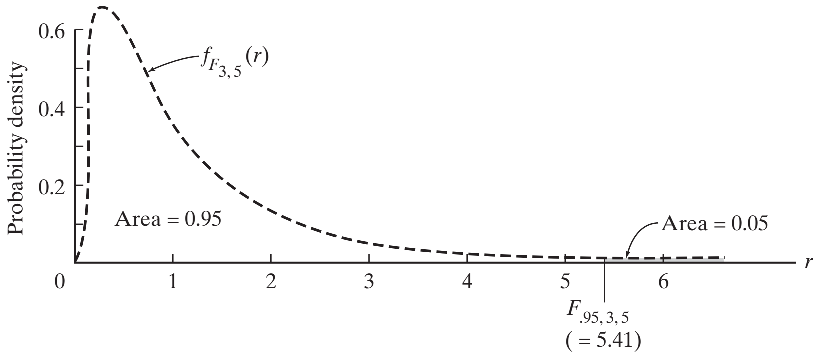
$$f_{\frac{V/m}{U/n}}(w) = f_{\frac{n}{m}V/U}(w) = \frac{1}{n/m} f_{V/U}\left(\frac{w}{n/m}\right) = \frac{m}{n} f_{V/U}\left(\frac{m}{n}w\right)$$

F TABLES

When graphed, an F distribution looks very much like a typical chi square distribution—values of $\frac{V/m}{U/n}$ can never be negative and the F pdf is skewed sharply to the right. Clearly, the complexity of $f_{F_{m,n}}(r)$ makes the function difficult to work with directly. Tables, though, are widely available that give various percentiles of F distributions for different values of m and n.

Figure 7.3.1 shows $f_{F_{3,5}}(r)$. In general, the symbol $F_{p,m,n}$ will be used to denote the 100 pth percentile of the F distribution with m and n degrees of freedom. Here, the 95th percentile of $f_{F_{3,5}}(r)$—that is, $F_{.95,3,5}$—is 5.41 (see Appendix Table A.4).

Figure 7.3.1



USING THE F DISTRIBUTION TO DERIVE THE pdf FOR t RATIOS

Now we have all the background results necessary to find the pdf of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$. Actually, though, we can do better than that because what we have been calling the "t ratio" is just one special case of an entire family of quotients known as t ratios. Finding the pdf for that entire family will give us the probability distribution for $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ as well.

Definition 7.3.3
Let Z be a standard normal random variable and let U be a chi square random variable independent of Z with n degrees of freedom. The *Student t ratio with n degrees of freedom* is denoted T_n, where

$$T_n = \frac{Z}{\sqrt{\frac{U}{n}}}$$

Comment The term "degrees of freedom" is often abbreviated by df.

Lemma *The pdf for T_n is symmetric: $f_{T_n}(t) = f_{T_n}(-t)$, for all t.*

Proof For convenience of notation, let $V = \sqrt{\frac{U}{n}}$. Then by Theorem 3.8.4 and the symmetry of the pdf of Z,

$$f_{T_n}(t) = \int_0^{\infty} v f_V(v) f_Z(tv)dv = \int_0^{\infty} v f_V(v) f_Z(-tv)dv = f_{T_n}(-t)$$

Theorem 7.3.4 *The pdf for a Student t random variable with n degrees of freedom is given by*

$$f_{T_n}(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}, \quad -\infty < t < \infty$$

Proof Note that $T_n^2 = \frac{Z^2}{U/n}$ has an F distribution with 1 and n df. Therefore,

$$f_{T_n^2}(t) = \frac{n^{n/2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)} t^{-1/2} \frac{1}{(n+t)^{(n+1)/2}}, \quad t > 0$$

Suppose that $t > 0$. By the symmetry of $f_{T_n}(t)$,

$$\begin{aligned} F_{T_n}(t) = P(T_n \le t) &= \frac{1}{2} + P(0 \le T_n \le t) \\ &= \frac{1}{2} + \frac{1}{2}P(-t \le T_n \le t) \\ &= \frac{1}{2} + \frac{1}{2}P\left(0 \le T_n^2 \le t^2\right) \\ &= \frac{1}{2} + \frac{1}{2}F_{T_n^2}(t^2) \end{aligned}$$

Differentiating $F_{T_n}(t)$ gives the stated result:

$$\begin{aligned} f_{T_n}(t) = F'_{T_n}(t) &= t \cdot f_{T_n^2}(t^2) \\ &= t\frac{n^{n/2}\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}(t^2)^{-(1/2)}\frac{1}{(n+t^2)^{(n+1)/2}} \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\left[1+\left(\frac{t^2}{n}\right)\right]^{(n+1)/2}} \end{aligned}$$

Comment Over the years, the lowercase t has come to be the accepted symbol for the random variable of Definition 7.3.3. We will follow that convention when the context allows some flexibility. In mathematical statements about distributions, though, we will be consistent with random variable notation and denote the Student t ratio as T_n.

All that remains to be verified, then, to accomplish our original goal of finding the pdf for $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ is to show that the latter is a special case of the Student t random variable described in Definition 7.3.3. Theorem 7.3.5 provides the details. Notice that a sample of size n yields a t ratio in this case having $n-1$ degrees of freedom.

Theorem 7.3.5 *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and standard deviation σ. Then*

$$T_{n-1} = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$$

has a Student t distribution with $n-1$ degrees of freedom.

Proof We can rewrite $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ in the form

$$\frac{\overline{Y}-\mu}{S/\sqrt{n}} = \frac{\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}$$

(Continued on next page)

(Theorem 7.3.5 continued)

But $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is a standard normal random variable and $\frac{(n-1)S^2}{\sigma^2}$ has a chi square distribution with $n-1$ df. Moreover, Theorem 7.3.2 shows that

$$\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2}$$

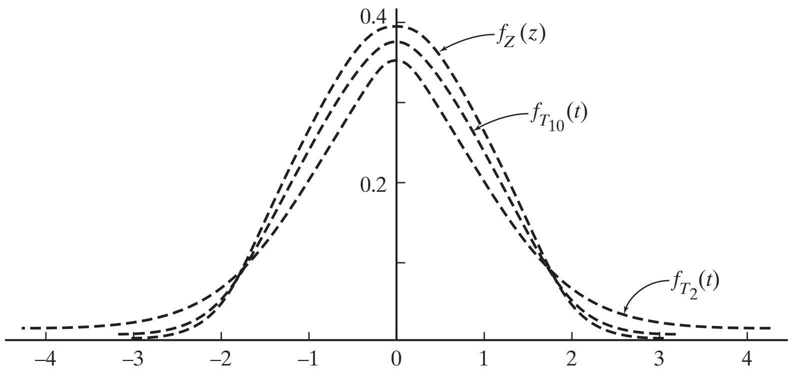
are independent. The statement of the theorem follows immediately, then, from Definition 7.3.3.

$f_{T_n}(t)$ AND $f_Z(z)$: HOW THE TWO PDFS ARE RELATED

Despite the considerable disparity in the appearance of the formulas for $f_{T_n}(t)$ and $f_Z(z)$, Student t distributions and the standard normal distribution have much in common. Both are bell shaped, symmetric, and centered around zero. Student t curves, though, are flatter.

Figure 7.3.2 is a graph of two Student t distributions—one with 2 df and the other with 10 df. Also pictured is the standard normal pdf, $f_Z(z)$. Notice that as n increases, $f_{T_n}(t)$ becomes more and more like $f_Z(z)$.

Figure 7.3.2



The convergence of $f_{T_n}(t)$ to $f_Z(z)$ is a consequence of two estimation properties:

1. The sample standard deviation is asymptotically unbiased for σ.
2. The standard deviation of S goes to 0 as n approaches ∞. (See Question 7.3.4.)

Therefore as n gets large, the probabilistic behavior of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ will become increasingly similar to the distribution of $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$—that is, to $f_Z(z)$.

Questions

7.3.1. Show directly—without appealing to the fact that χ^2_n is a gamma random variable—that $f_U(u)$ as stated in Definition 7.3.1 is a true probability density function.

7.3.2. Find the moment-generating function for a chi square random variable and use it to show that $E(\chi^2_n) = n$ and $\text{Var}(\chi^2_n) = 2n$.

7.3.3. Is it believable that the numbers 65, 30, and 55 are a random sample of size 3 from a normal distribution with $\mu = 50$ and $\sigma = 10$? Answer the question by using a chi square distribution. (*Hint:* Let $Z_i = (Y_i - 50)/10$ and use Theorem 7.3.1.)

7.3.4. Use the fact that $(n-1)S^2/\sigma^2$ is a chi square random variable with $n-1$ df to prove that

$$\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

(*Hint:* Use the fact that the variance of a chi square random variable with k df is $2k$.)

7.3.5. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution. Use the statement of Question 7.3.4 to prove that S^2 is consistent for σ^2.

7.3.6. If Y is a chi square random variable with n degrees of freedom, the pdf of $(Y - n)/\sqrt{2n}$ converges to $f_Z(z)$ as n goes to infinity (recall Question 7.3.2). Use the asymptotic normality of $(Y - n)/\sqrt{2n}$ to approximate the 40th percentile of a chi square random variable with 200 degrees of freedom.

7.3.7. Use Appendix Table A.4 to find

(a) $F_{.50,6,7}$

(b) $F_{.001,15,5}$

(c) $F_{.90,2,2}$

7.3.8. Let V and U be independent chi square random variables with 7 and 9 degrees of freedom, respectively. Is it more likely that $\frac{V/7}{U/9}$ will be between (1) 2.51 and 3.29 or (2) 3.29 and 4.20?

7.3.9. Use Appendix Table A.4 to find the values of x that satisfy the following equations:

(a) $P(0.109 < F_{4,6} < x) = 0.95$

(b) $P(0.427 < F_{11,7} < 1.69) = x$

(c) $P(F_{x,x} > 5.35) = 0.01$

(d) $P(0.115 < F_{3,x} < 3.29) = 0.90$

(e) $P\left(x < \frac{V/2}{U/3}\right) = 0.25$, where V is a chi square random variable with 2 df and U is an independent chi square random variable with 3 df.

7.3.10. Suppose that two independent samples of size n are drawn from a normal distribution with variance σ^2. Let S_1^2 and S_2^2 denote the two sample variances. Use the fact that $\frac{(n-1)S^2}{\sigma^2}$ has a chi square distribution with $n - 1$ df to explain why

$$\lim_{\substack{n\to\infty \\ m\to\infty}} F_{m,n} = 1$$

7.3.11. If the random variable F has an F distribution with m and n degrees of freedom, show that $1/F$ has an F distribution with n and m degrees of freedom.

7.3.12. Use the result claimed in Question 7.3.11 to express percentiles of $f_{F_{n,m}}(r)$ in terms of percentiles from $f_{F_{m,n}}(r)$. That is, if we know the values a and b for which $P(a \le F_{m,n} \le b) = q$, what values of c and d will satisfy the equation $P(c \le F_{n,m} \le d) = q$? "Check" your answer with Appendix Table A.4 by comparing the values of $F_{.05,2,8}$, $F_{.95,2,8}$, $F_{.05,8,2}$, and $F_{.95,8,2}$.

7.3.13. Show that as $n \to \infty$, the pdf of a Student t random variable with n df converges to $f_Z(z)$. (*Hint:* To show that the constant term in the pdf for T_n converges to $1/\sqrt{2\pi}$, use Stirling's formula,

$$n! \doteq \sqrt{2\pi n}\, n^n e^{-n})$$

Also, recall that $\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a$.

7.3.14. Evaluate the integral

$$\int_0^\infty \frac{1}{1+x^2}\, dx$$

using the Student t distribution.

7.3.15. For a Student t random variable T with n degrees of freedom and any positive integer k, show that $E(T^{2k})$ exists if $2k < n$. (*Hint:* Integrals of the form

$$\int_0^\infty \frac{1}{(1+y^\alpha)^\beta}\, dy$$

are finite if $\alpha > 0$, $\beta > 0$, and $\alpha\beta > 1$.)

7.4 Drawing Inferences About μ

One of the most common of all statistical objectives is to draw inferences about the *mean* of the population being represented by a set of data. Indeed, we already took a first look at that problem in Section 6.2. If the Y_i's come from a normal distibution where σ is known, the null hypothesis $H_0 : \mu = \mu_0$ can be tested by calculating a Z ratio, $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ (recall Theorem 6.2.1).

Implicit in that solution, though, is an assumption not likely to be satisfied: rarely does the experimenter actually know the value of σ. Section 7.3 dealt with precisely that scenario and derived the pdf of the ratio $T_{n-1} = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$, where σ has been replaced by an estimator, S. Given T_{n-1} (which we learned has a Student t distribution with $n - 1$ degrees of freedom), we now have the tools necessary to draw inferences about μ in the all-important case where σ is not known. Section 7.4 illustrates these various techniques and also examines the key assumption underlying the "t test" and looks at what happens when that assumption is not satisfied.

t TABLES

We have already seen that doing hypothesis tests and constructing confidence intervals using $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ or some other Z ratio requires that we know certain upper and/or lower percentiles from the standard normal distribution. There will be a similar need to identify appropriate "cutoffs" from Student t distributions when the inference procedure is based on $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$, or some other t ratio.

Figure 7.4.1 shows a portion of the t table that typically appears in the back of statistics books. Each row corresponds to a different Student t pdf. The column headings give the area *to the right* of the number appearing in the body of the table.

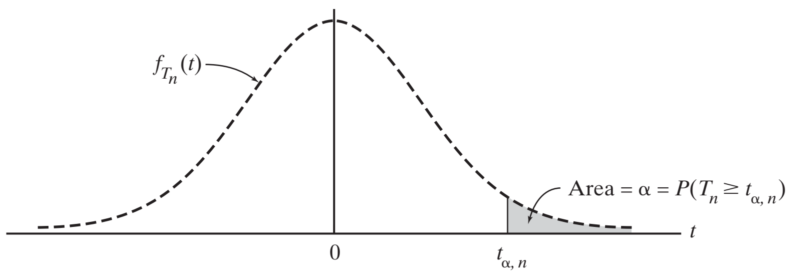
Figure 7.4.1

	α						
df	.20	.15	.10	.05	.025	.01	.005
1	1.376	1.963	3.078	6.3138	12.706	31.821	63.657
2	1.061	1.386	1.886	2.9200	4.3027	6.965	9.9248
3	0.978	1.250	1.638	2.3534	3.1825	4.541	5.8409
4	0.941	1.190	1.533	2.1318	2.7764	3.747	4.6041
5	0.920	1.156	1.476	2.0150	2.5706	3.365	4.0321
6	0.906	1.134	1.440	1.9432	2.4469	3.143	3.7074
⋮			⋮				
30	0.854	1.055	1.310	1.6973	2.0423	2.457	2.7500
∞	0.84	1.04	1.28	1.64	1.96	2.33	2.58

For example, the entry 4.541 listed in the $\alpha = .01$ column and the df $= 3$ row has the property that $P(T_3 \geq 4.541) = 0.01$.

More generally, we will use the symbol $t_{\alpha,n}$ to denote the $100(1-\alpha)$th percentile of $f_{T_n}(t)$. That is, $P(T_n \geq t_{\alpha,n}) = \alpha$ (see Figure 7.4.2). No lower percentiles of Student t curves need to be tabulated because the symmetry of $f_{T_n}(t)$ implies that $P(T_n \leq -t_{\alpha,n}) = \alpha$.

Figure 7.4.2



The number of different Student t pdfs summarized in a t table varies considerably. Many tables will provide cutoffs for degrees of freedom ranging only from 1 to 30; others will include df values from 1 to 50, or even from 1 to 100. The last row in most t tables, though, is labeled "∞": Those entries, of course, correspond to z_α.

CONSTRUCTING A CONFIDENCE INTERVAL FOR μ

The fact that $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ has a Student t distribution with $n-1$ degrees of freedom justifies the statement that

$$P\left(-t_{\alpha/2,n-1} \leq \frac{\overline{Y}-\mu}{S/\sqrt{n}} \leq t_{\alpha/2,n-1}\right) = 1-\alpha$$

or, equivalently, that

$$P\left(\overline{Y} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \overline{Y} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha \qquad (7.4.1)$$

(provided the Y_i's are a random sample from a normal distribution).

When the actual data values are then used to evaluate $\overline{Y}$ and S, the lower and upper endpoints identified in Equation 7.4.1 define a $100(1-\alpha)\%$ confidence interval for μ.

Theorem 7.4.1 *Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from a normal distribution with (unknown) mean μ. A $100(1-\alpha)\%$ confidence interval for μ is the set of values*

$$\left(\overline{y} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \overline{y} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right)$$

CASE STUDY 7.4.1

To hunt flying insects, bats emit high-frequency sounds and then listen for their echoes. Until an insect is located, these pulses are emitted at intervals of from fifty to one hundred milliseconds. When an insect *is* detected, the pulse-to-pulse interval suddenly decreases—sometimes to as low as ten milliseconds—thus enabling the bat to pinpoint its prey's position. This raises an interesting question: How far apart are the bat and the insect when the bat first senses that the insect is there? Or, put another way, what is the effective range of a bat's echolocation system?

The technical problems that had to be overcome in measuring the bat-to-insect detection distance were far more complex than the statistical problems involved in analyzing the actual data. The procedure that finally evolved was to put a bat into an eleven-by-sixteen-foot room, along with an ample supply of fruit flies, and record the action with two synchronized sixteen-millimeter sound-on-film cameras. By examining the two sets of pictures frame by frame, scientists could follow the bat's flight pattern and, at the same time, monitor its pulse frequency. For each insect that was caught (70), it was therefore possible to estimate the distance between the bat and the insect at the precise moment the bat's pulse-to-pulse interval decreased (see Table 7.4.1).

Table 7.4.1

Catch Number	Detection Distance (cm)
1	62
2	52
3	68
4	23
5	34
6	45
7	27
8	42
9	83
10	56
11	40

(Continued on next page)

(Case Study 7.4.1 continued)

Define μ to be a bat's true average detection distance. Use the eleven observations in Table 7.4.1 to construct a 95% confidence interval for μ.

Letting $y_1 = 62, y_2 = 52, \ldots, y_{11} = 40$, we have that

$$\sum_{i=1}^{11} y_i = 532 \quad \text{and} \quad \sum_{i=1}^{11} y_i^2 = 29{,}000$$

Therefore,

$$\bar{y} = \frac{532}{11} = 48.4 \text{ cm}$$

and

$$s = \sqrt{\frac{11(29{,}000) - (532)^2}{11(10)}} = 18.1 \text{ cm}$$

If the population from which the y_i's are being drawn is normal, the behavior of

$$\frac{\overline{Y} - \mu}{S/\sqrt{n}}$$

will be described by a Student t curve with 10 degrees of freedom. From Table A.2 in the Appendix,

$$P(-2.2281 < T_{10} < 2.2281) = 0.95$$

Accordingly, the 95% confidence interval for μ is

$$\left[\bar{y} - 2.2281\left(\frac{s}{\sqrt{11}}\right), \bar{y} + 2.2281\left(\frac{s}{\sqrt{11}}\right)\right]$$
$$= \left[48.4 - 2.2281\left(\frac{18.1}{\sqrt{11}}\right), 48.4 + 2.2281\left(\frac{18.1}{\sqrt{11}}\right)\right]$$
$$= (36.2 \text{ cm},\ 60.6 \text{ cm})$$

Example 7.4.1 The sample mean and sample standard deviation for the random sample of size $n = 20$ given in the following list are 2.6 and 3.6, respectively. Let μ denote the true mean of the distribution being represented by these y_i's.

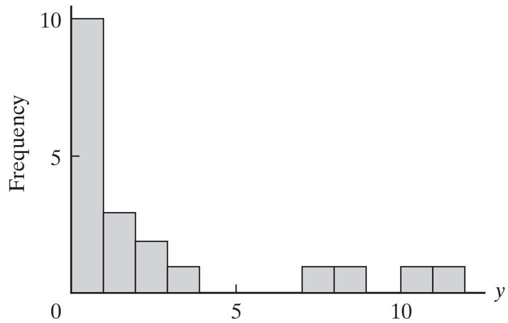
2.5	0.1	0.2	1.3
3.2	0.1	0.1	1.4
0.5	0.2	0.4	11.2
0.4	7.4	1.8	2.1
0.3	8.6	0.3	10.1

Is it correct to say that a 95% confidence interval for μ is the set of following values?

$$\left(\bar{y} - t_{.025,n-1} \cdot \frac{s}{\sqrt{n}}, \bar{y} + t_{.025,n-1} \cdot \frac{s}{\sqrt{n}}\right)$$
$$= \left(2.6 - 2.0930 \cdot \frac{3.6}{\sqrt{20}}, 2.6 + 2.0930 \cdot \frac{3.6}{\sqrt{20}}\right)$$
$$= (0.9, 4.3)$$

No. It *is* true that all the correct factors have been used in calculating (0.9, 4.3), but Theorem 7.4.1 does not apply in this case because the normality assumption it makes is clearly being violated. Figure 7.4.3 is a histogram of the twenty y_i's. The extreme skewness that is so evident there is not consistent with the presumption that the data's underlying pdf is a normal distribution. As a result, the pdf describing the probabilistic behavior of $\frac{\overline{Y}-\mu}{S/\sqrt{20}}$ would *not* be $f_{T_{19}}(t)$.

Figure 7.4.3



■

Comment To say that $\frac{\overline{Y}-\mu}{S/\sqrt{20}}$ in this situation is not *exactly* a T_{19} random variable leaves unanswered a critical question: Is the ratio *approximately* a T_{19} random variable? We will revisit the normality assumption—and what happens when that assumption is not satisfied—later in this section when we discuss a critically important property known as *robustness*.

Questions

7.4.1. Use Appendix Table A.2 to find the following probabilities:

(a) $P(T_6 \geq 1.134)$

(b) $P(T_{15} \leq 0.866)$

(c) $P(T_3 \geq -1.250)$

(d) $P(-1.055 < T_{29} < 2.462)$

7.4.2. What values of x satisfy the following equations?

(a) $P(-x \leq T_{22} \leq x) = 0.98$

(b) $P(T_{13} \geq x) = 0.85$

(c) $P(T_{26} < x) = 0.95$

(d) $P(T_2 \geq x) = 0.025$

7.4.3. Which of the following differences is larger? Explain.

$$t_{.05,n} - t_{.10,n} \quad \text{or} \quad t_{.10,n} - t_{.15,n}$$

7.4.4. A random sample of size $n = 9$ is drawn from a normal distribution with $\mu = 27.6$. Within what interval $(-a, +a)$ can we expect to find $\frac{\overline{Y}-27.6}{S/\sqrt{9}}$ 80% of the time? 90% of the time?

7.4.5. Suppose a random sample of size $n = 11$ is drawn from a normal distribution with $\mu = 15.0$. For what value of k is the following true?

$$P\left(\left|\frac{\overline{Y}-15.0}{S/\sqrt{11}}\right| \geq k\right) = 0.05$$

7.4.6. Let $\overline{Y}$ and S denote the sample mean and sample standard deviation, respectively, based on a set of $n = 20$ measurements taken from a normal distribution with $\mu = 90.6$. Find the function $k(S)$ for which

$$P[90.6 - k(S) \leq \overline{Y} \leq 90.6 + k(S)] = 0.99$$

7.4.7. Cell phones emit radio frequency energy that is absorbed by the body when the phone is next to the ear and may be harmful. The table in the next column gives the absorption rate for a sample of twenty high-radiation cell phones. (The Federal Communication Commission sets a maximum of 1.6 watts per kilogram for the absorption rate of such energy.) Construct a 90% confidence interval for the true average cell phone absorption rate.

1.54	1.41
1.54	1.40
1.49	1.40
1.49	1.39
1.48	1.39
1.45	1.39
1.44	1.38
1.42	1.38
1.41	1.37
1.41	1.33

Data from: http://cellphones.procon.org/view.resource.php?resourceID=003054

7.4.8. The following table lists the typical cost of repairing the bumper of a moderately priced midsize car damaged by a corner collision at 3 mph. Use these observations to construct a 95% confidence interval for μ, the true average repair cost for all such automobiles with similar damage. The sample standard deviation for these data is $s = \$369.02$.

Make/Model	Repair Cost	Make/Model	Repair Cost
Hyundai Sonata	\$1019	Honda Accord	\$1461
Nissan Altima	\$1090	Volkswagen Jetta	\$1525
Mitsubishi Galant	\$1109	Toyota Camry	\$1670
Saturn AURA	\$1235	Chevrolet Malibu	\$1685
Subaru Legacy	\$1275	Volkswagen Passat	\$1783
Pontiac G6	\$1361	Nissan Maxima	\$1787
Mazda 6	\$1437	Ford Fusion	\$1889
Volvo S40	\$1446	Chrysler Sebring	\$2484

Data from: www.iihs.org/ratings/bumpersbycategory.aspx?

7.4.9. Creativity, as any number of studies have shown, is very much a province of the young. Whether the focus is music, literature, science, or mathematics, an individual's best work seldom occurs late in life. Einstein, for example, made his most profound discoveries at the age of twenty-six; Newton, at the age of twenty-three. The following are twelve scientific breakthroughs dating from the middle of the sixteenth century to the early years of the twentieth century (217). All represented high-water marks in the careers of the scientists involved.

Discovery	Discoverer	Year	Age, y
Earth goes around sun	Copernicus	1543	40
Telescope, basic laws of astronomy	Galileo	1600	34
Principles of motion, gravitation, calculus	Newton	1665	23
Nature of electricity	Franklin	1746	40
Burning is uniting with oxygen	Lavoisier	1774	31
Earth evolved by gradual processes	Lyell	1830	33
Evidence for natural selection controlling evolution	Darwin	1858	49
Field equations for light	Maxwell	1864	33
Radioactivity	Curie	1896	34
Quantum theory	Planck	1901	43
Special theory of relativity, $E = mc^2$	Einstein	1905	26
Mathematical foundations for quantum theory	Schrödinger	1926	39

(a) What can be inferred from these data about the *true* average age at which scientists do their best work? Answer the question by constructing a 95% confidence interval.

(b) Before constructing a confidence interval for a set of observations extending over a long period of time, we should be convinced that the y_i's exhibit no biases or trends. If, for example, the age at which scientists made major discoveries decreased from century to century, then the parameter μ would no longer be a constant, and the confidence interval would be meaningless. Plot "date" versus "age" for these twelve discoveries. Put "date" on the abscissa. Does the variability in the y_i's appear to be random with respect to time?

7.4.10. How long does it take to fly from Atlanta to New York's LaGuardia airport? There are many components of the time elapsed, but one of the more stable measurements is the actual in-air time. For a sample of eighty-three flights between these destinations on Fridays in October, the time in minutes (y) gave the following results:

$$\sum_{i=1}^{83} y_i = 8622 \quad \text{and} \quad \sum_{i=1}^{83} y_i^2 = 899{,}750$$

Find a 99% confidence interval for the average flight time.

Data from: apps.bts.gov

7.4.11. In a nongeriatric population, platelet counts ranging from 140 to 440 (thousands per mm^3 of blood) are considered "normal." The following are the platelet counts recorded for twenty-four female nursing home residents (180).

Subject	Count	Subject	Count
1	125	13	180
2	170	14	180
3	250	15	280
4	270	16	240
5	144	17	270
6	184	18	220
7	176	19	110
8	100	20	176
9	220	21	280
10	200	22	176
11	170	23	188
12	160	24	176

Use the following sums:

$$\sum_{i=1}^{24} y_i = 4645 \quad \text{and} \quad \sum_{i=1}^{24} y_i^2 = 959{,}265$$

How does the definition of "normal" above compare with the 90% confidence interval?

7.4.12. If a normally distributed sample of size $n = 16$ produces a 95% confidence interval for μ that ranges from 44.7 to 49.9, what are the values of $\bar{y}$ and s?

7.4.13. Two samples, each of size n, are taken from a normal distribution with unknown mean μ and unknown standard deviation σ. A 90% confidence interval for μ is constructed with the first sample, and a 95% confidence interval for μ is constructed with the second. Will the 95% confidence interval necessarily be longer than the 90% confidence interval? Explain.

7.4.14. Revenues reported last week from nine boutiques franchised by an international clothier averaged \$59,540 with a standard deviation of \$6860. Based on those figures, in what range might the company expect to find the average revenue of all of its boutiques?

7.4.15. What "confidence" is associated with each of the following random intervals? Assume that the Y_i's are normally distributed.

(a) $\left[\overline{Y} - 2.0930\left(\frac{S}{\sqrt{20}}\right), \overline{Y} + 2.0930\left(\frac{S}{\sqrt{20}}\right)\right]$

(b) $\left[\overline{Y} - 1.345\left(\frac{S}{\sqrt{15}}\right), \overline{Y} + 1.345\left(\frac{S}{\sqrt{15}}\right)\right]$

(c) $\left[\overline{Y} - 1.7056\left(\frac{S}{\sqrt{27}}\right), \overline{Y} + 2.7787\left(\frac{S}{\sqrt{27}}\right)\right]$

(d) $\left[-\infty, \overline{Y} + 1.7247\left(\frac{S}{\sqrt{21}}\right)\right]$

7.4.16. The weather station at Dismal Swamp, California, recorded monthly precipitation (y) for twenty-eight years. For these data, $\sum_{i=1}^{336} y_i = 1392.6$ and $\sum_{i=1}^{336} y_i^2 = 10,518.84$.

(a) Find the 95% confidence interval for the mean monthly precipitation.

(b) The table on the right gives a frequency distribution for the Dismal Swamp precipitation data. Does this distribution raise questions about using Theorem 7.4.1?

Rainfall in Inches	Frequency
0–1	85
1–2	38
2–3	35
3–4	41
4–5	28
5–6	24
6–7	18
7–8	16
8–9	16
9–10	5
10–11	9
11–12	21

Data from: www.wcc.nrcs.usda.gov.

TESTING $H_0 : \mu = \mu_0$ (THE ONE-SAMPLE t TEST)

Suppose a normally distributed random sample of size n is observed for the purpose of testing the null hypothesis that $\mu = \mu_0$. If σ is unknown—which is usually the case—the procedure to use is called a *one-sample t test*. Conceptually, the latter is much like the Z test of Theorem 6.2.1, except that the decision rule is defined in terms of $t = \frac{\overline{y}-\mu_0}{s/\sqrt{n}}$ rather than $z = \frac{\overline{y}-\mu_0}{\sigma/\sqrt{n}}$ [which requires that the critical values come from $f_{T_{n-1}}(t)$ rather than $f_Z(z)$].

Theorem 7.4.2 *Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from a normal distribution where σ is unknown. Let $t = \frac{\overline{y}-\mu_0}{s/\sqrt{n}}$.*

a. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ at the α level of significance, reject H_0 if $t \geq t_{\alpha,n-1}$.*

b. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ at the α level of significance, reject H_0 if $t \leq -t_{\alpha,n-1}$.*

c. *To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at the α level of significance, reject H_0 if t is either (1) $\leq -t_{\alpha/2,n-1}$ or (2) $\geq t_{\alpha/2,n-1}$.*

Proof Appendix 7.A.2 gives the complete derivation that justifies using the procedure described in Theorem 7.4.2. In short, the test statistic $t = \frac{\overline{y}-\mu_0}{s/\sqrt{n}}$ is a monotonic function of the λ that appears in Definition 6.5.2, which makes the one-sample t test a GLRT.

CASE STUDY 7.4.2

Not all rectangles are created equal. Since antiquity, societies have expressed aesthetic preferences for rectangles having certain width (w) to length (l) ratios.

One "standard" calls for the width-to-length ratio to be equal to the ratio of the length to the sum of the width and the length. That is,

$$\frac{w}{l} = \frac{l}{w+l} \tag{7.4.2}$$

Equation 7.4.2 implies that the width is $\frac{1}{2}(\sqrt{5}-1)$, or approximately 0.618, times as long as the length. The Greeks called this the golden rectangle and used it often in their architecture (see Figure 7.4.4). Many other cultures were similarly inclined. The Egyptians, for example, built their pyramids out of stones whose faces were golden rectangles. Today in our society, the golden rectangle remains an architectural and artistic standard, and even items such as driver's licenses, business cards, and picture frames often have w/l ratios close to 0.618.

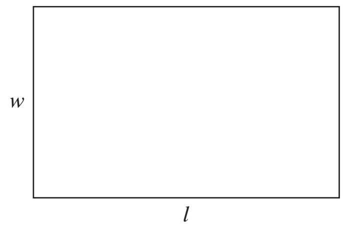


Figure 7.4.4

The fact that many societies have embraced the golden rectangle as an aesthetic standard has two possible explanations. One, they "learned" to like it because of the profound influence that Greek writers, philosophers, and artists have had on cultures all over the world. Or two, there is something unique about human perception that predisposes a preference for the golden rectangle.

Researchers in the field of experimental aesthetics have tried to test the plausibility of those two hypotheses by seeing whether the golden rectangle is accorded any special status by societies that had no contact whatsoever with the Greeks or with their legacy. One such study (41) examined the w/l ratios of beaded rectangles sewn by the Shoshoni Indians as decorations on their blankets and clothes. Table 7.4.2 lists the ratios found for twenty such rectangles.

If, indeed, the Shoshonis also had a preference for golden rectangles, we would expect their ratios to be "close" to 0.618. The average value of the entries in Table 7.4.2, though, is 0.661. What does that imply? Is 0.661 close enough to 0.618 to support the position that liking the golden rectangle is a human characteristic, or is 0.661 so far from 0.618 that the only prudent conclusion is that the Shoshonis did *not* agree with the aesthetics espoused by the Greeks?

Table 7.4.2 Width-to-Length Ratios of Shoshoni Rectangles

0.693	0.749	0.654	0.670
0.662	0.672	0.615	0.606
0.690	0.628	0.668	0.611
0.606	0.609	0.601	0.553
0.570	0.844	0.576	0.933

Let μ denote the true average width-to-length ratio of Shoshoni rectangles. The hypotheses to be tested are

$$H_0 : \mu = 0.618$$

versus

$$H_1 : \mu \neq 0.618$$

For tests of this nature, the value of $\alpha = 0.05$ is often used. For that value of α and a two-sided test, the critical values, using part (c) of Theorem 7.4.2 and Appendix Table A.2, are $t_{.025,19} = 2.0930$ and $-t_{.025,19} = -2.0930$.

The data in Table 7.4.2 have $\overline{y} = 0.661$ and $s = 0.093$. Substituting these values into the t ratio gives a test statistic that lies just inside of the interval between -2.0930 and 2.0930:

$$t = \frac{\overline{y} - \mu_0}{s/\sqrt{n}} = \frac{0.661 - 0.618}{0.093/\sqrt{20}} = 2.068$$

Thus, these data do not rule out the possibility that the Shoshoni Indians also embraced the golden rectangle as an aesthetic standard.

About the Data Like π and e, the ratio w/l for golden rectangles (more commonly referred to as either *phi* or the *golden ratio*), is an irrational number with all sorts of fascinating properties and connections.

Algebraically, the solution of the equation

$$\frac{w}{l} = \frac{l}{w + l}$$

is the continued fraction

$$\frac{w}{l} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}}$$

Among the curiosities associated with phi is its relationship with the *Fibonacci series*. The latter, of course, is the famous sequence in which each term is the sum of its two predecessors—that is,

$$1 \quad 1 \quad 2 \quad 3 \quad 5 \quad 8 \quad 13 \quad 21 \quad 34 \quad 55 \quad 89 \quad \ldots$$

Quotients of consecutive terms in the Fibonacci sequence alternate above and below the golden ratio and converge to its value ($= 0.61803\ldots$) : $1/1 = 1.0000$, $1/2 = 0.5000$, $2/3 = 0.6667$, $3/5 = 0.6000$, $5/8 = 0.6250$, $8/13 = 0.6154$, $13/21 = 0.6190$, and so on.

Example 7.4.2 Three banks serve a metropolitan area's inner-city neighborhoods: Federal Trust, American United, and Third Union. The state banking commission is concerned that loan applications from inner-city residents are not being accorded the same consideration that comparable requests have received from individuals in rural areas. Both constituencies claim to have anecdotal evidence suggesting that the other group is being given preferential treatment.

Records show that last year these three banks approved 62% of all the home mortgage applications filed by rural residents. Listed in Table 7.4.3 are the approval rates posted over that same period by the twelve branch offices of Federal Trust (FT), American United (AU), and Third Union (TU) that work primarily with the inner-city community. Do these figures lend any credence to the contention that the banks are treating inner-city residents and rural residents differently? Analyze the data using an $\alpha = 0.05$ level of significance.

Table 7.4.3

Bank	Location	Affiliation	Percent Approved
1	3rd & Morgan	AU	59
2	Jefferson Pike	TU	65
3	East 150th & Clark	TU	69
4	Midway Mall	FT	53
5	N. Charter Highway	FT	60
6	Lewis & Abbot	AU	53
7	West 10th & Lorain	FT	58
8	Highway 70	FT	64
9	Parkway Northwest	AU	46
10	Lanier & Tower	TU	67
11	King & Tara Court	AU	51
12	Bluedot Corners	FT	59

As a starting point, we might want to test

$$H_0 : \mu = 62$$

versus

$$H_1 : \mu \neq 62$$

where μ is the true average approval rate for all inner-city banks. Table 7.4.4 summarizes the analysis. The two critical values are $\pm t_{.025,11} = \pm 2.2010$, and the observed t ratio is $-1.66 \left(= \frac{58.667-62}{6.946/\sqrt{12}}\right)$, so our decision is "Fail to reject H_0."

Table 7.4.4

Banks	n	$\bar{y}$	s	t Ratio	Critical Value	Reject H_0?
All	12	58.667	6.946	-1.66	± 2.2010	No

■

About the Data The "overall" analysis of Table 7.4.4, though, may be too simplistic. Common sense would tell us to look also at the three banks separately. What emerges, then, is an entirely different picture (see Table 7.4.5). Now we can see why both groups felt discriminated against: American United ($t = -3.63$) and Third

Table 7.4.5

Banks	n	$\bar{y}$	s	t Ratio	Critical Value	Reject H_0?
American United	4	52.25	5.38	-3.63	± 3.1825	Yes
Federal Trust	5	58.80	3.96	-1.81	± 2.7764	No
Third Union	3	67.00	2.00	$+4.33$	± 4.3027	Yes

Union ($t = +4.33$) each had rates that differed significantly from 62%—*but in opposite directions!* Only Federal Trust seems to be dealing with inner-city residents and rural residents in an even-handed way. ■

Questions

7.4.17. Recall the *Bacillus subtilis* data in Question 5.3.2. Test the null hypothesis that exposure to the enzyme does not affect a worker's respiratory capacity (as measured by the FEV_1/VC ratio). Use a one-sided H_1 and let $\alpha = 0.05$. Assume that σ is not known.

7.4.18. Recall Case Study 5.3.1. Assess the credibility of the theory that Etruscans were native Italians by testing an appropriate H_0 against a two-sided H_1. Set α equal to 0.05. Use 143.8 mm and 6.0 mm for $\bar{y}$ and s, respectively, and let $\mu_0 = 132.4$. Do these data appear to satisfy the distribution assumption made by the t test? Explain.

7.4.19. MBAs R Us advertises that its program increases a person's score on the GMAT by an average of forty points. As a way of checking the validity of that claim, a consumer watchdog group hired fifteen students to take both the review course and the GMAT. Prior to starting the course, the fifteen students were given a diagnostic test that predicted how well they would do on the GMAT in the absence of any special training. The following table gives each student's actual GMAT score minus his or her predicted score. Set up and carry out an appropriate hypothesis test. Use the 0.05 level of significance.

Subject	y_i = act. GMAT − pre. GMAT	y_i^2
SA	35	1225
LG	37	1369
SH	33	1089
KN	34	1156
DF	38	1444
SH	40	1600
ML	35	1225
JG	36	1296
KH	38	1444
HS	33	1089
LL	28	784
CE	34	1156
KK	47	2209
CW	42	1764
DP	46	2116

7.4.20. In addition to the Shoshoni data of Case Study 7.4.2, a set of rectangles that might tend to the golden ratio are national flags. The table below gives the width-to-length ratios for a random sample of the flags of thirty-four countries. Let μ be the width-to-length ratio for national flags. At the $\alpha = 0.01$ level, test $H_0 : \mu = 0.618$ versus $H_1 : \mu \neq 0.618$.

Country	Ratio Width to Height	Country	Ratio Width to Height
Afghanistan	0.500	Iceland	0.720
Albania	0.714	Iran	0.571
Algeria	0.667	Israel	0.727
Angola	0.667	Laos	0.667
Argentina	0.667	Lebanon	0.667
Bahamas	0.500	Liberia	0.526
Denmark	0.757	Macedonia	0.500
Djibouti	0.553	Mexico	0.571
Ecuador	0.500	Monaco	0.800
Egypt	0.667	Namibia	0.667
El Salvador	0.600	Nepal	1.250
Estonia	0.667	Romania	0.667
Ethiopia	0.500	Rwanda	0.667
Gabon	0.750	South Africa	0.667
Fiji	0.500	St. Helena	0.500
France	0.667	Sweden	0.625
Honduras	0.500	United Kingdom	0.500

Data from: http://www.anyflag.com/country/costaric.php.

7.4.21. A manufacturer of pipe for laying underground electrical cables is concerned about the pipe's rate of corrosion and whether a special coating may retard that rate. As a way of measuring corrosion, the manufacturer examines a short length of pipe and records the depth of the maximum pit. The manufacturer's tests have shown that in a year's time in the particular kind of soil the manufacturer must deal with, the average depth of the maximum pit in a foot of pipe is 0.0042 inch. To see whether that average can be reduced, ten pipes are coated with a new plastic and buried in the same soil. After one year, the following maximum pit depths are recorded (in inches): 0.0039, 0.0041, 0.0038, 0.0044, 0.0040, 0.0036, 0.0034, 0.0036, 0.0046, and 0.0036. Given that the sample standard deviation for these ten measurements is 0.000383 inch, can it be concluded at the $\alpha = 0.05$ level of significance that the plastic coating is beneficial?

7.4.22. In athletic contests, a wide-spread conviction exists that the home team has an advantage. However, one explanation for this is that a team schedules some much weaker opponents to play at home. To avoid this bias, a study of college football games considered only games between teams ranked in the top twenty-five. For three hundred seventeen such games, the margin of victory (y = home team score–visitors score) was recorded. For these data, $\bar{y} = 4.57$ and $s = 18.29$. Does this study confirm the existence of a home field advantage? Test $H_0 : \mu = 0$ versus $H_1 : \mu > 0$ at the 0.05 level of significance.

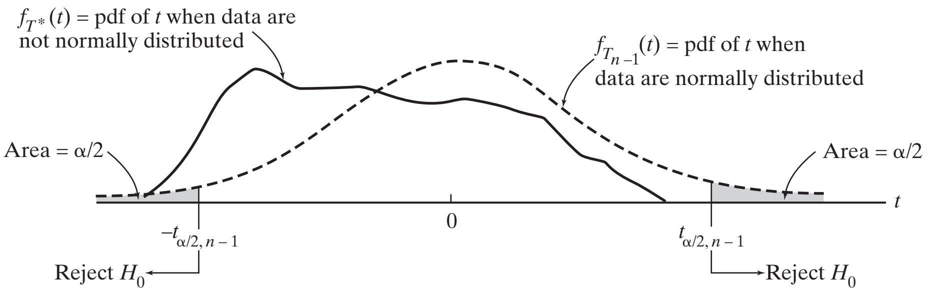
7.4.23. The first analysis done in Example 7.4.2 (using all $n = 12$ banks with $\bar{y} = 58.667$) failed to reject $H_0: \mu = 62$ at the $\alpha = 0.05$ level. Had μ_0 been, say, 61.7 or 58.6, the same conclusion would have been reached. What do we call the entire set of μ_0's for which $H_0: \mu = \mu_0$ would *not* be rejected at the $\alpha = 0.05$ level?

TESTING H_0: $\mu = \mu_0$ WHEN THE NORMALITY ASSUMPTION IS NOT MET

Every t test makes the same explicit assumption—namely, that the set of n y_i's is normally distributed. But suppose the normality assumption is *not* true. What are the consequences? Is the validity of the t test compromised?

Figure 7.4.5 addresses the first question. We know that if the normality assumption *is* true, the pdf describing the variation of the t ratio, $\frac{\bar{Y}-\mu_0}{S/\sqrt{n}}$, is $f_{T_{n-1}}(t)$. The latter, of course, provides the decision rule's critical values. If $H_0 : \mu = \mu_0$ is to be tested against $H_1 : \mu \neq \mu_0$, for example, the null hypothesis is rejected if t is either (1) $\leq -t_{\alpha/2,n-1}$ or (2) $\geq t_{\alpha/2,n-1}$ (which makes the Type I error probability equal to α).

Figure 7.4.5



If the normality assumption is *not* true, the pdf of $\frac{\bar{Y}-\mu_0}{S/\sqrt{n}}$ will not be $f_{T_{n-1}}(t)$ and in general

$$P\left(\frac{\bar{Y}-\mu_0}{S/\sqrt{n}} \leq -t_{\alpha/2,n-1}\right) + P\left(\frac{\bar{Y}-\mu_0}{S/\sqrt{n}} \geq t_{\alpha/2,n-1}\right) \neq \alpha$$

In effect, violating the normality assumption creates *two* α's: The "nominal" α is the Type I error probability we specify at the outset—typically, 0.05 or 0.01. The "true" α is the actual probability that $\frac{\bar{Y}-\mu_0}{S/\sqrt{n}}$ falls in the rejection region (when H_0 is true). For the two-sided decision rule pictured in Figure 7.4.5,

$$\text{true}\,\alpha = \int_{-\infty}^{-t_{\alpha/2,n-1}} f_{T^*}(t)\,dt + \int_{t_{\alpha/2,n-1}}^{\infty} f_{T^*}(t)\,dt$$

Whether or not the validity of the t test is "compromised" by the normality assumption being violated depends on the numerical difference between the two α's. If $f_{T^*}(t)$ is, in fact, quite similar in shape and location to $f_{T_{n-1}}(t)$, then the true α will be approximately equal to the nominal α. In that case, the fact that the y_i's are not normally distributed would be essentially irrelevant. On the other hand, if $f_{T^*}(t)$ and $f_{T_{n-1}}(t)$ are dramatically different (as they appear to be in Figure 7.4.5), it would follow that the normality assumption *is* critical, and establishing the "significance" of a t ratio becomes problematic.

Unfortunately, getting an exact expression for $f_{T^*}(t)$ is essentially impossible because the distribution depends on the pdf being sampled, and there is seldom any way of knowing precisely what that pdf might be. However, we can still meaningfully explore the sensitivity of the t ratio to violations of the normality assumption by simulating samples of size n from selected distributions and comparing the resulting histogram of t ratios to $f_{T_{n-1}}(t)$.

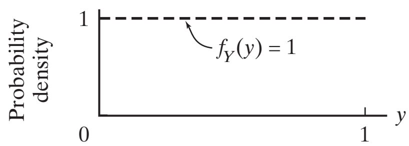
Figure 7.4.6 shows four such simulations, using Minitab; the first three consist of one hundred random samples of size $n = 6$. In Figure 7.4.6(a), the samples come from a uniform pdf defined over the interval [0, 1]; in Figure 7.4.6(b), the underlying pdf is the exponential with $\lambda = 1$; and in Figure 7.4.6(c), the data are coming from a Poisson pdf with $\lambda = 5$.

If the normality assumption were true, t ratios based on samples of size 6 would vary in accordance with the Student t distribution with 5 df. On pp. 401–402, $f_{T_5}(t)$ has been superimposed over the histograms of the t ratios coming from the three different pdfs. What we see there is really quite remarkable. The t ratios based on y_i's coming from a uniform pdf, for example, are behaving much the same way as t ratios would vary if the y_i's were normally distributed—that is, $f_{T^*}(t)$ in this case appears to be very similar to $f_{T_5}(t)$. The same is true for samples coming from a Poisson distribution (see Theorem 4.2.2). For both of those underlying pdfs, in other words, the true α would not be much different from the nominal α.

Figure 7.4.6(b) tells a slightly different story. When samples of size 6 are drawn from an exponential pdf, the t ratios are *not* in particularly close agreement with

Figure 7.4.6

(a)




```
MTB > random¨100¨c1-c6;
SUBC> uniform¨0¨1.
MTB > rmean¨c1-c6¨c7
MTB > rstdev¨c1-c6¨c8
MTB > let¨c9¨=¨sqrt(6)*(((c7)-0.5)/(c8))
MTB > histogram¨c9
```

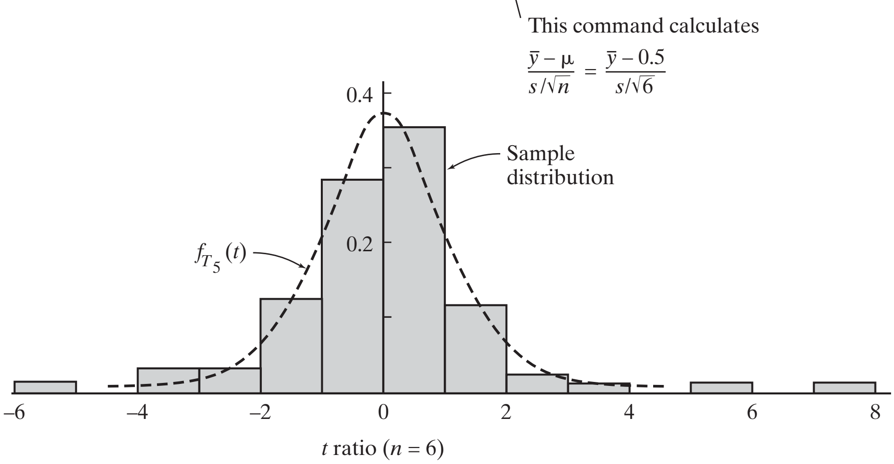
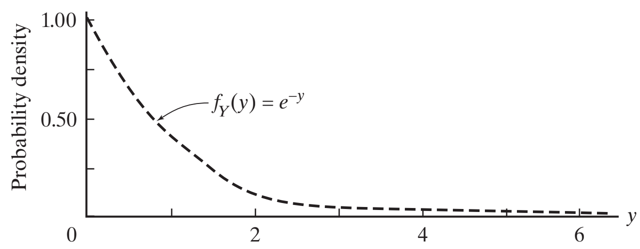


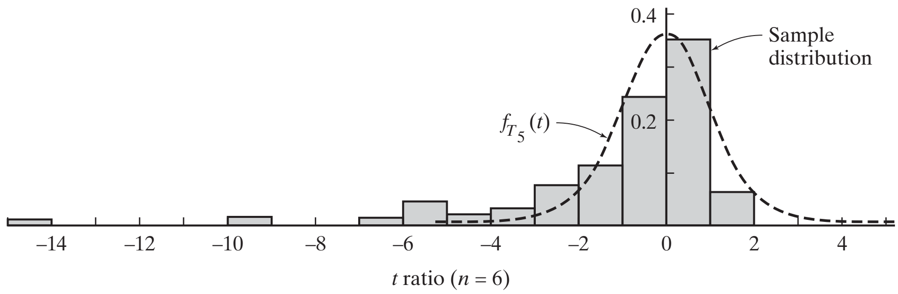
Figure 7.4.6 (*Continued*)

(**b**)

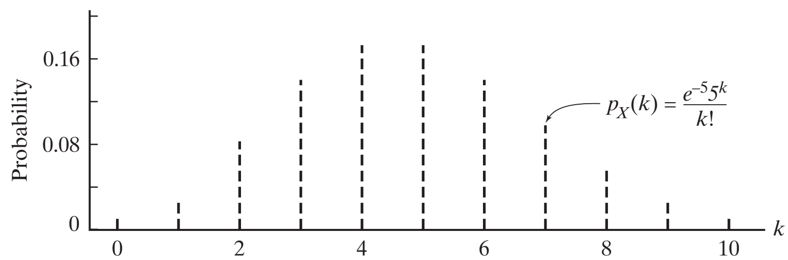



```
MTB > random¨100¨cl-c6;
SUBC> exponential¨1.
MTB > rmean¨cl-c6¨c7
MTB > rstdev¨cl-c6¨c8
MTB > let¨c9¨=¨sqrt(6)*(((c7)¨-¨1.0)/(c8))
MTB¨>¨histogram¨c9
```

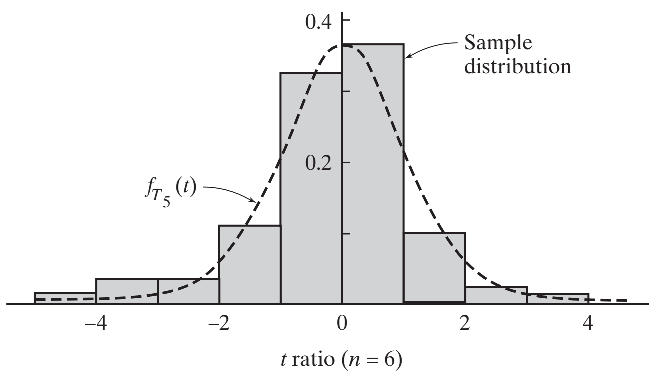
$$\left[= \frac{\bar{y} - \mu}{s/\sqrt{6}}\right]$$



(**c**)



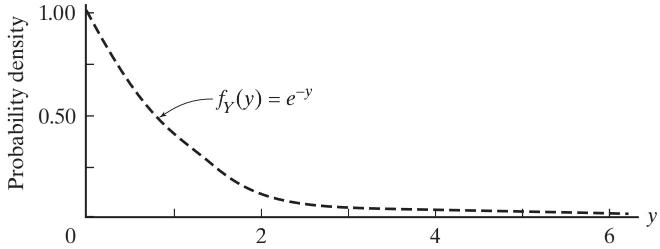

```
MTB¨>¨random¨100¨cl-c6;
SUBC¨>¨poisson¨5.
MTB¨>¨rmean¨cl-c6¨c7
MTB¨>¨rstdev¨cl-c6¨c8
MTB¨>¨let¨c9¨=¨sqrt(6)*(((c7)¨-¨5.0)/(c8))
MTB¨>¨histogram¨c9
```



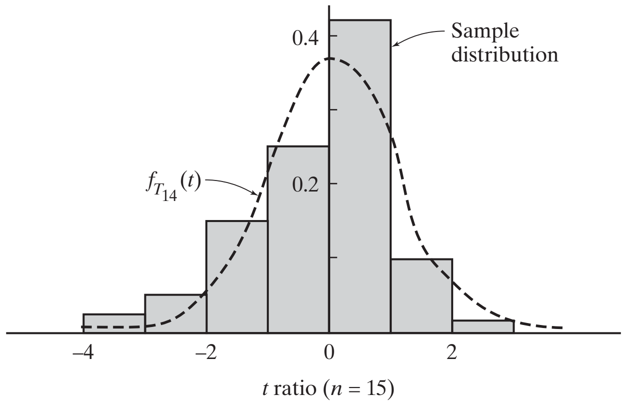
$f_{T_5}(t)$. Specifically, very negative t ratios are occurring much more often than the Student t curve would predict, while large positive t ratios are occurring less often (see Question 7.4.24). But look at Figure 7.4.6(d). When the sample size is increased to $n = 15$, the skewness so prominent in Figure 7.4.6(b) is mostly gone.

Figure 7.4.6 *(Continued)*

(d)




```
MTB > random 100 cl-c15;
SUBC> exponential 1.
MTB > rmean cl-c15 c16
MTB > rstdev cl-c15 c17
MTB > let c18 = sqrt(15)*(((c16 - 1.0)/(c17))
MTB > histogram c18
```



Reflected in these specific simulations are some general properties of the t ratio:

1. The distribution of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ is relatively unaffected by the pdf of the y_i's (provided $f_Y(y)$ is not too skewed and n is not too small).
2. As n increases, the pdf of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ becomes increasingly similar to $f_{T_{n-1}}(t)$.

In mathematical statistics, the term *robust* is used to describe a procedure that is not heavily dependent on whatever assumptions it makes. Figure 7.4.6 shows that the *t test is robust with respect to departures from normality*.

From a practical standpoint, it would be difficult to overstate the importance of the t test being robust. If the pdf of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$ varied dramatically depending on the origin of the y_i's, we would never know if the true α associated with, say, a 0.05 decision rule was anywhere near 0.05. That degree of uncertainty would make the t test virtually worthless.

Questions

7.4.24. Explain why the distribution of t ratios calculated from small samples drawn from the exponential pdf, $f_Y(y) = e^{-y}$, $y \geq 0$, will be skewed to the left [recall Figure 7.4.6(b)]. (*Hint:* What does the shape of $f_Y(y)$ imply about the possibility of each y_i being close to 0? If the entire sample did consist of y_i's close to 0, what value would the t ratio have?)

7.4.25. Suppose one hundred samples of size $n = 3$ are taken from each of the pdfs

$$\textbf{(1)}\ f_Y(y) = 2y, \qquad 0 \leq y \leq 1$$

and

$$\textbf{(2)}\ f_Y(y) = 4y^3, \qquad 0 \leq y \leq 1$$

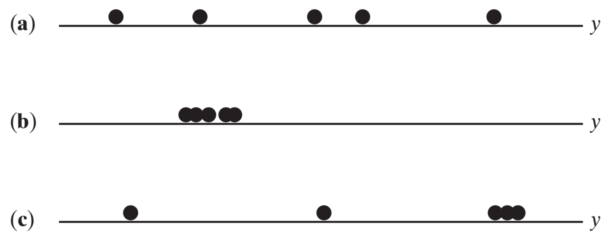
and for each set of three observations, the ratio

$$\frac{\overline{y} - \mu}{s/\sqrt{3}}$$

is calculated, where μ is the expected value of the particular pdf being sampled. How would you expect the distributions of the two sets of ratios to be different? How would they be similar? Be as specific as possible.

7.4.26. Suppose that random samples of size n are drawn from the uniform pdf, $f_Y(y) = 1, 0 \leq y \leq 1$. For each sample, the ratio $t = \frac{\overline{y}-0.5}{s/\sqrt{n}}$ is calculated. Parts (b) and (d) of Figure 7.4.6 suggest that the pdf of t will become increasingly similar to $f_{T_{n-1}}(t)$ as n increases. To which pdf is $f_{T_{n-1}}(t)$, itself, converging as n increases?

7.4.27. On which of the following sets of data would you be reluctant to do a t test? Explain.



7.5 Drawing Inferences About σ^2

When random samples are drawn from a normal distribution, it is usually the case that the parameter μ is the target of the investigation. More often than not, the mean mirrors the "effect" of a treatment or condition, in which case it makes sense to apply what we learned in Section 7.4—that is, either construct a confidence interval for μ or test the hypothesis that $\mu = \mu_0$.

But exceptions are not that uncommon. Situations occur where the "precision" associated with a measurement is, itself, important—perhaps even more important than the measurement's "location." If so, we need to shift our focus to the often used *measure of variability parameter,* σ^2. Two key facts that we learned earlier about the population variance will now come into play. First, an unbiased estimator for σ^2 based on its maximum likelihood estimator is the sample variance, S^2, where

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

And, second, the ratio

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

has a chi square distribution with $n-1$ degrees of freedom. Putting these two pieces of information together allows us to draw inferences about σ^2—in particular, we can construct confidence intervals for σ^2 and test the hypothesis that $\sigma^2 = \sigma_o^2$.

CHI SQUARE TABLES

Just as we need a t table to carry out inferences about μ (when σ^2 is unknown), we need a chi square table to provide the cutoffs for making inferences involving σ^2. The layout of chi square tables is dictated by the fact that all chi square pdfs (unlike Z and t distributions) are skewed (see, for example, Figure 7.5.1, showing a chi square curve having 5 degrees of freedom). Because of that asymmetry, chi square tables need to provide cutoffs for both the left-hand tail and the right-hand tail of each chi square distribution.

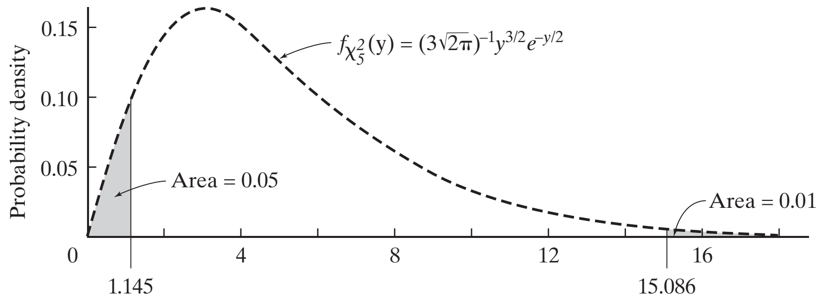


Figure 7.5.1

Figure 7.5.2 shows the top portion of the chi square table that appears in Appendix A.3. Successive rows refer to different chi square distributions (each having a different number of degrees of freedom). The column headings denote the areas *to the left* of the numbers listed in the body of the table.

				p				
df	.01	.025	.05	.10	.90	.95	.975	.99
1	0.000157	0.000982	0.00393	0.0158	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217

Figure 7.5.2

We will use the symbol $\chi^2_{p,n}$ to denote the number along the horizontal axis that cuts off, to its left, an area of p under the chi square distribution with n degrees of freedom. For example, from the fifth row of the chi square table, we see the numbers 1.145 and 15.086 under the column headings .05 and .99, respectively. It follows that

$$P\left(\chi^2_5 \le 1.145\right) = 0.05$$

and

$$P\left(\chi^2_5 \le 15.086\right) = 0.99$$

(see Figure 7.5.1). In terms of the $\chi^2_{p,n}$ notation, $1.145 = \chi^2_{.05,5}$ and $15.086 = \chi^2_{.99,5}$. (The area *to the right* of 15.086, of course, must be 0.01.)

CONSTRUCTING CONFIDENCE INTERVALS FOR σ^2

Since $\frac{(n-1)S^2}{\sigma^2}$ has a chi square distribution with $n-1$ degrees of freedom, we can write

$$P\left[\chi^2_{\alpha/2,n-1} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{1-\alpha/2,n-1}\right] = 1-\alpha \tag{7.5.1}$$

If Equation 7.5.1 is then inverted to isolate σ^2 in the center of the inequalities, the two endpoints will necessarily define a $100(1-\alpha)\%$ confidence interval for the population variance. The algebraic details are left as an exercise.

Theorem 7.5.1 *Let s^2 denote the sample variance calculated from a random sample of n observations drawn from a normal distribution with mean μ and variance σ^2. Then*

a. *a $100(1-\alpha)\%$ confidence interval for σ^2 is the set of values*

$$\left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\right)$$

b. *a $100(1-\alpha)\%$ confidence interval for σ is the set of values*

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}}\right)$$

CASE STUDY 7.5.1

The chain of events that define the geological evolution of the Earth began hundreds of millions of years ago. Fossils play a key role in documenting the *relative* times those events occurred, but to establish an *absolute* chronology, scientists rely primarily on radioactive decay.

One of the dating techniques uses a rock's potassium-argon ratio. Almost all minerals contain potassium (K) as well as certain of its isotopes, including ^{40}K. The latter, though, is unstable and decays into isotopes of argon and calcium, ^{40}Ar and ^{40}Ca. By knowing the rates at which the various daughter products are formed and by measuring the amounts of ^{40}Ar and ^{40}K present in a specimen, geologists can estimate the object's age.

Critical to the interpretation of any such dates, of course, is the precision of the underlying procedure. One obvious way to estimate that precision is to use the technique on a sample of rocks known to have the same age. Whatever variation occurs, then, from rock to rock is reflecting the inherent precision (or lack of precision) of the procedure.

Table 7.5.1 lists the potassium-argon estimated ages of nineteen mineral samples, all taken from the Black Forest in southeastern Germany (118). Assume that the procedure's estimated ages are normally distributed with (unknown) mean μ and (unknown) variance σ^2. Construct a 95% confidence interval for σ.

Table 7.5.1

Specimen	Estimated Age (millions of years)
1	249
2	254
3	243
4	268
5	253
6	269
7	287
8	241
9	273
10	306
11	303
12	280
13	260
14	256
15	278
16	344
17	304
18	283
19	310

Here

$$\sum_{i=1}^{19} y_i = 5261$$

$$\sum_{i=1}^{19} y_i^2 = 1{,}469{,}945$$

so the sample variance is 733.4:

$$s^2 = \frac{19(1{,}469{,}945) - (5261)^2}{19(18)} = 733.4$$

Since $n = 19$, the critical values appearing in the left-hand and right-hand limits of the σ confidence interval come from the chi square pdf with 18 df. According to Appendix Table A.3,

$$P\left(8.23 < \chi^2_{18} < 31.53\right) = 0.95$$

so the 95% confidence interval for the potassium-argon method's precision is the set of values

$$\left(\sqrt{\frac{(19-1)(733.4)}{31.53}}, \sqrt{\frac{(19-1)(733.4)}{8.23}}\right) = (20.5 \text{ million years, } 40.0 \text{ million years})$$

Example 7.5.1 The width of a confidence interval for σ^2 is a function of both n and S^2:

$$\begin{aligned}\text{Width} &= \text{upper limit} - \text{lower limit}\\ &= \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} - \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\\ &= (n-1)S^2\left(\frac{1}{\chi^2_{\alpha/2,n-1}} - \frac{1}{\chi^2_{1-\alpha/2,n-1}}\right) \qquad (7.5.2)\end{aligned}$$

As n gets larger, the interval will tend to get narrower because the unknown σ^2 is being estimated more precisely. What is the smallest number of observations that will guarantee that the average width of a 95% confidence interval for σ^2 is no greater than σ^2?

Since S^2 is an unbiased estimator for σ^2, Equation 7.5.2 implies that the expected width of a 95% confidence interval for the variance is the expression

$$E(\text{width}) = (n-1)\sigma^2\left(\frac{1}{\chi^2_{.025,n-1}} - \frac{1}{\chi^2_{.975,n-1}}\right)$$

Clearly, then, for the expected width to be less than or equal to σ^2, n must be chosen so that

$$(n-1)\left(\frac{1}{\chi^2_{.025,n-1}} - \frac{1}{\chi^2_{.975,n-1}}\right) \le 1$$

Trial and error can be used to identify the desired n. The first three columns in Table 7.5.2 come from the chi square distribution in Appendix Table A.3. As the computation in the last column indicates, $n = 39$ is the smallest sample size that will yield 95% confidence intervals for σ^2 whose average width is less than σ^2.

Table 7.5.2

n	$\chi^2_{.025,n-1}$	$\chi^2_{.975,n-1}$	$(n-1)\left(\frac{1}{\chi^2_{.025,n-1}} - \frac{1}{\chi^2_{.975,n-1}}\right)$
15	5.629	26.119	1.95
20	8.907	32.852	1.55
30	16.047	45.722	1.17
38	22.106	55.668	1.01
39	22.878	56.895	0.99

■

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

The generalized likelihood ratio criterion introduced in Section 6.5 can be used to set up hypothesis tests for σ^2. The complete derivation appears in Appendix 7.A.3. Theorem 7.5.2 states the resulting decision rule. Playing a key role—just as it did in the construction of confidence intervals for σ^2—is the chi square ratio from Theorem 7.3.2.

Theorem 7.5.2 *Let s^2 denote the sample variance calculated from a random sample of n observations drawn from a normal distribution with mean μ and variance σ^2. Let $\chi^2 = (n-1)s^2/\sigma_0^2$.*

a. *To test $H_0: \sigma^2 = \sigma_0^2$ versus $H_1: \sigma^2 > \sigma_0^2$ at the α level of significance, reject H_0 if $\chi^2 \geq \chi^2_{1-\alpha,n-1}$.*

b. *To test $H_0: \sigma^2 = \sigma_0^2$ versus $H_1: \sigma^2 < \sigma_0^2$ at the α level of significance, reject H_0 if $\chi^2 \leq \chi^2_{\alpha,n-1}$.*

c. *To test $H_0: \sigma^2 = \sigma_0^2$ versus $H_1: \sigma^2 \neq \sigma_0^2$ at the α level of significance, reject H_0 if χ^2 is either (1) $\leq \chi^2_{\alpha/2,n-1}$ or (2) $\geq \chi^2_{1-\alpha/2,n-1}$.*

CASE STUDY 7.5.2

Mutual funds are investment vehicles consisting of a portfolio of various types of investments. If such an investment is to meet annual spending needs, the owner of shares in the fund is interested in the average of the annual returns of the fund. Investors are also concerned with the volatility of the annual returns, measured by the variance or standard deviation. One common method of evaluating a mutual fund is to compare it to a benchmark, the Lipper Average being one of these. This index number is the average of returns from a universe of mutual funds.

The Global Rock Fund is a typical mutual fund, with heavy investments in international funds. It claimed to best the Lipper Average in terms of volatility over the period from 1989 through 2007. Its returns are given in the table below.

Year	Investment Return %	Year	Investment Return %
1989	15.32	1999	27.43
1990	1.62	2000	8.57
1991	28.43	2001	1.88
1992	11.91	2002	−7.96
1993	20.71	2003	35.98
1994	−2.15	2004	14.27
1995	23.29	2005	10.33
1996	15.96	2006	15.94
1997	11.12	2007	16.71
1998	0.37		

The sample standard deviation for these nineteen returns is 11.28%; the (true) corresponding volatility measure for Lipper Averages is known to be 11.67%. The question to be answered is whether the decrease from 11.67% to 11.28% is statistically significant.

Let σ^2 denote the true volatility characteristic of the Global Rock Fund. The hypothesis to be tested, then, would be written

(Continued on next page)

(Case Study 7.5.2 continued)

$$H_0 : \sigma^2 = (11.67)^2$$

versus

$$H_1 : \sigma^2 < (11.67)^2$$

Let $\alpha = 0.05$. With $n = 19$, the critical value for the chi square ratio (from part (b) of Theorem 7.5.2) is $\chi^2_{\alpha,n-1} = \chi^2_{.05,18} = 9.390$ (see Figure 7.5.3). But

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(19-1)(11.28)^2}{(11.67)^2} = 16.82$$

so our decision is clear: Do not reject H_0.

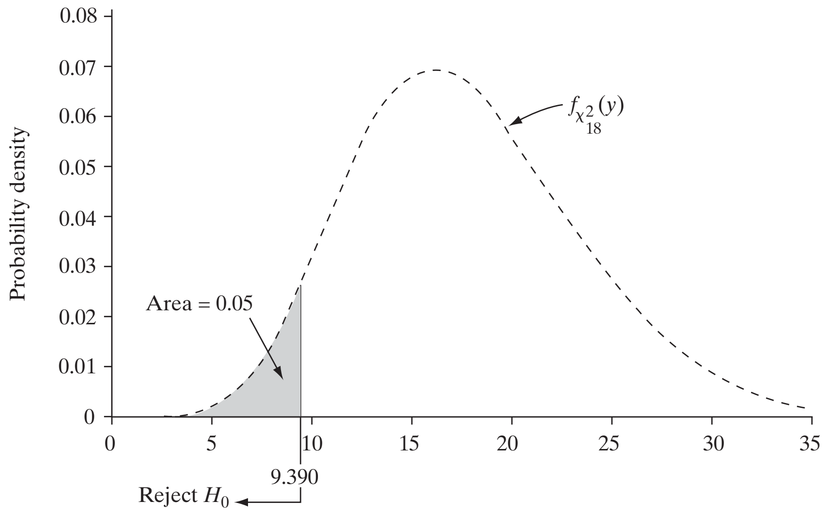


Figure 7.5.3

Questions

7.5.1. Use Appendix Table A.3 to find the following cutoffs and indicate their location on the graph of the appropriate chi square distribution.

(a) $\chi^2_{.95,14}$

(b) $\chi^2_{.90,2}$

(c) $\chi^2_{.025,9}$

7.5.2. Evaluate the following probabilities:

(a) $P(\chi^2_{17} \geq 8.672)$

(b) $P(\chi^2_6 < 10.645)$

(c) $P(9.591 \leq \chi^2_{20} \leq 34.170)$

(d) $P(\chi^2_2 < 9.210)$

7.5.3. Find the value y that satisfies each of the following equations:

(a) $P(\chi^2_9 \geq y) = 0.99$

(b) $P(\chi^2_{15} \leq y) = 0.05$

(c) $P(9.542 \leq \chi^2_{22} \leq y) = 0.09$

(d) $P(y \leq \chi^2_{31} \leq 48.232) = 0.95$

7.5.4. For what value of n is each of the following statements true?

(a) $P(\chi^2_n \geq 5.009) = 0.975$

(b) $P(27.204 \leq \chi^2_n \leq 30.144) = 0.05$

(c) $P(\chi^2_n \leq 19.281) = 0.05$

(d) $P(10.085 \leq \chi^2_n \leq 24.769) = 0.80$

7.5.5. For df values beyond the range of Appendix Table A.3, chi square cutoffs can be approximated by using a formula based on cutoffs from the standard normal pdf, $f_Z(z)$. Define $\chi^2_{p,n}$ and z^*_p so that $P(\chi^2_n \le \chi^2_{p,n}) = p$ and $P(Z \le z^*_p) = p$, respectively. Then

$$\chi^2_{p,n} \doteq n\left(1 - \frac{2}{9n} + z^*_p\sqrt{\frac{2}{9n}}\right)^3$$

Approximate the 95th percentile of the chi square distribution with 200 df. That is, find the value of y for which

$$P(\chi^2_{200} \le y) \doteq 0.95$$

7.5.6. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution having mean μ and variance σ^2. What is the smallest value of n for which the following is true?

$$P\left(\frac{S^2}{\sigma^2} < 2\right) \ge 0.95$$

(*Hint:* Use a trial-and-error method.)

7.5.7. Start with the fact that $(n-1)S^2/\sigma^2$ has a chi square distribution with $n-1$ df (if the Y_i's are normally distributed) and derive the confidence interval formulas given in Theorem 7.5.1.

7.5.8. A random sample of size $n = 19$ is drawn from a normal distribution for which $\sigma^2 = 12.0$. In what range are we likely to find the sample variance, s^2? Answer the question by finding two numbers a and b such that

$$P(a \le S^2 \le b) = 0.95$$

7.5.9. How long sporting events last is quite variable. This variability can cause problems for TV broadcasters, since the amount of commercials and commentator blather varies with the length of the event. As an example of this variability, the table below gives the lengths for a random sample of middle-round contests at the 2008 Wimbledon Championships in women's tennis.

Match	Length (minutes)
Cirstea-Kuznetsova	73
Srebotnik-Meusburger	76
De Los Rios-V. Williams	59
Kanepi-Mauresmo	104
Garbin-Szavay	114
Bondarenko-Lisicki	106
Vaidisova-Bremond	79
Groenefeld-Moore	74
Govortsova-Sugiyama	142
Zheng-Jankovic	129
Perebiynis-Bammer	95
Bondarenko-V. Williams	56
Coin-Mauresmo	84
Petrova-Pennetta	142
Wozniacki-Jankovic	106
Groenefeld-Safina	75

Data from: 2008.usopen.org/en_US/scores/cmatch/index.html?promo=t.

(a) Assume that match lengths are normally distributed. Use Theorem 7.5.1 to construct a 95% confidence interval for the standard deviation of match lengths.

(b) Use these same data to construct two *one-sided* 95% confidence intervals for σ.

7.5.10. How much interest certificates of deposit (CDs) pay varies by financial institution and also by length of the investment. A large sample of national one-year CD offerings in 2009 showed an average interest rate of 1.84 and a standard deviation $\sigma = 0.262$. A five-year CD ties up an investor's money, so it usually pays a higher rate of interest. However, higher rates might cause more variability. The table lists the five-year CD rate offerings from $n = 10$ banks in the northeast United States. Find a 95% confidence interval for the standard deviation of five-year CD rates. Do these data suggest that interest rates for five-year CDs are more variable than those for one-year certificates?

Bank	Interest Rate (%)
Domestic Bank	2.21
Stonebridge Bank	2.47
Waterfield Bank	2.81
NOVA Bank	2.81
American Bank	2.96
Metropolitan National Bank	3.00
AIG Bank	3.35
iGObanking.com	3.44
Discover Bank	3.44
Intervest National Bank	3.49

Data from: Company reports.

7.5.11. In Case Study 7.5.1, the 95% confidence interval was constructed for σ rather than for σ^2. In practice, is an experimenter more likely to focus on the standard deviation or on the variance, or do you think that both formulas in Theorem 7.5.1 are likely to be used equally often? Explain.

7.5.12. (a) Use the asymptotic normality of chi square random variables (see Question 7.3.6) to derive large-sample confidence interval formulas for σ and σ^2.

(b) Use your answer to part (a) to construct an approximate 95% confidence interval for the standard deviation of estimated potassium-argon ages based on the nineteen y_i's in Table 7.5.1. How does this confidence interval compare with the one in Case Study 7.5.1?

7.5.13. If a 90% confidence interval for σ^2 is reported to be (51.47, 261.90), what is the value of the sample standard deviation?

7.5.14. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from the pdf

$$f_Y(y) = \left(\frac{1}{\theta}\right)e^{-y/\theta}, \quad y > 0; \quad \theta > 0$$

(a) Use moment-generating functions to show that the ratio $2n\overline{Y}/\theta$ has a chi square distribution with $2n$ df.

(b) Use the result in part (a) to derive a $100(1-\alpha)\%$ confidence interval for θ.

7.5.15. Another method for dating rocks was used before the advent of the potassium-argon method described in Case Study 7.5.1. Because of a mineral's lead content, it was capable of yielding estimates for this same time period with a standard deviation of 30.4 million years. The potassium-argon method in Case Study 7.5.1 had a smaller sample standard deviation of $\sqrt{733.4} = 27.1$ million years. Is this "proof" that the potassium-argon method is more precise? Using the data in Table 7.5.1, test at the 0.05 level whether the potassium-argon method has a smaller standard deviation than the older procedure using lead.

7.5.16. When working properly, the amounts of cement that a filling machine puts into 25-kg bags have a standard deviation (σ) of 1.0 kg. In the next column are the weights recorded for thirty bags selected at random from a day's production. Test H_0: $\sigma^2 = 1$ versus H_1: $\sigma^2 > 1$ using the $\alpha = 0.05$ level of significance. Assume that the weights are normally distributed.

26.18	24.22	24.22
25.30	26.48	24.49
25.18	23.97	25.68
24.54	25.83	26.01
25.14	25.05	25.50
25.44	26.24	25.84
24.49	25.46	26.09
25.01	25.01	25.21
25.12	24.71	26.04
25.67	25.27	25.23

Use the following sums:

$$\sum_{i=1}^{30} y_i = 758.62 \text{ and } \sum_{i=1}^{30} y_i^2 = 19{,}195.7938$$

7.5.17. A stock analyst claims to have devised a mathematical technique for selecting high-quality mutual funds and promises that a client's portfolio will have higher average ten-year annualized returns and lower volatility; that is, a smaller standard deviation. After ten years, one of the analyst's twenty-four-stock portfolios showed an average ten-year annualized return of 11.50% and a standard deviation of 10.17%. The benchmarks for the type of funds considered are a mean of 10.10% and a standard deviation of 15.67%.

(a) Let μ be the mean for a twenty-four-stock portfolio selected by the analyst's method. Test at the 0.05 level that the portfolio beat the benchmark; that is, test H_0: $\mu = 10.1$ versus H_1: $\mu > 10.1$.

(b) Let σ be the standard deviation for a twenty-four-stock portfolio selected by the analyst's method. Test at the 0.05 level that the portfolio beat the benchmark; that is, test H_0: $\sigma = 15.67$ versus H_1: $\sigma < 15.67$.

7.6 Taking a Second Look at Statistics (Type II Error)

For data that are normal, *and when the variance σ^2 is known*, both Type I errors and Type II errors can be determined, staying within the family of normal distributions. (See Example 6.4.1, for instance.) As the material in this chapter shows, the situation changes radically when σ^2 is not known. With the development of the Student t distribution, tests of a given level of significance α can be constructed. But what is the Type II error of such a test?

To answer this question, let us first recall the form of the test statistic and critical region testing, for example,

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

The null hypothesis is rejected if

$$\frac{\overline{Y} - \mu_0}{S/\sqrt{n}} \geq t_{\alpha,n-1}$$

The probability of the Type II error, β, of the test at some value $\mu_1 > \mu_0$ is

$$P\left(\frac{\overline{Y} - \mu_0}{S/\sqrt{n}} < t_{\alpha,n-1}\right)$$

However, since μ_0 is not the mean of $\overline{Y}$ under H_1, the distribution of $\frac{\overline{Y}-\mu_0}{S/\sqrt{n}}$ is *not* Student t. Indeed, a new distribution is called for.

The following algebraic manipulations help to place the needed density into a recognizable form.

$$\frac{\overline{Y}-\mu_0}{S/\sqrt{n}} = \frac{\overline{Y}-\mu_1+(\mu_1-\mu_0)}{S/\sqrt{n}} = \frac{\frac{\overline{Y}-\mu_1}{\sigma}+\frac{(\mu_1-\mu_0)}{\sigma}}{\frac{S/\sqrt{n}}{\sigma}} = \frac{\frac{\overline{Y}-\mu_1}{\sigma/\sqrt{n}}+\frac{(\mu_1-\mu_0)}{\sigma/\sqrt{n}}}{S/\sigma}$$

$$= \frac{\frac{\overline{Y}-\mu_1}{\sigma/\sqrt{n}}+\frac{(\mu_1-\mu_0)}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{\frac{\overline{Y}-\mu_1}{\sigma/\sqrt{n}}+\delta}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \frac{Z+\delta}{\sqrt{\frac{U}{n-1}}}$$

where $Z = \frac{\overline{Y}-\mu_1}{\sigma/\sqrt{n}}$ is normal, $U = \frac{(n-1)S^2}{\sigma^2}$ is a chi square variable with $n-1$ degrees of freedom, and $\delta = \frac{(\mu_1-\mu_0)}{\sigma/\sqrt{n}}$ is an (unknown) constant. Note that the random variable $\frac{Z+\delta}{\sqrt{\frac{U}{n-1}}}$ differs from the Student t with $n-1$ degrees of freedom $\frac{Z}{\sqrt{\frac{U}{n-1}}}$ only because of the additive term δ in the numerator. But adding δ changes the nature of the pdf significantly.

An expression of the form $\frac{Z+\delta}{\sqrt{\frac{U}{n-1}}}$ is said to have a *noncentral t distribution with $n-1$ degrees of freedom and noncentrality parameter* δ.

The probability density function for a noncentral t variable is now well known (105). Even though there are computer approximations to the distribution, not knowing σ^2 means that δ is also unknown. One approach often taken is to specify the difference between the true mean and the hypothesized mean *as a given proportion of* σ. That is, the Type II error is given as a function of $\frac{\mu_1-\mu_0}{\sigma}$ rather than μ_1. In some cases, this quantity can be approximated by $\frac{\mu_1-\mu_0}{s}$.

The following numerical example will help to clarify these ideas.

Example 7.6.1

Suppose we wish to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ at the $\alpha = 0.05$ level of significance. Let $n = 20$. In this case the test is to reject H_0 if the test statistic $\frac{\overline{y}-\mu_0}{s/\sqrt{n}}$ is greater than $t_{.05,19} = 1.7291$. What will be the Type II error if the mean has shifted by 0.5 standard deviations to the right of μ_0?

Saying that the mean has shifted by 0.5 standard deviations to the right of μ_0 is equivalent to setting $\frac{\mu_1-\mu_0}{\sigma} = 0.5$. In that case, the noncentrality parameter is $\delta = \frac{\mu_1-\mu_0}{\sigma/\sqrt{n}} = (0.5)\cdot\sqrt{20} = 2.236$.

The probability of a Type II error is

$$P(T_{19,2.236} \leq 1.7291)$$

where $T_{19,2.236}$ is a noncentral t variable with 19 degrees of freedom and noncentrality parameter 2.236.

To calculate this quantity, we need the cdf of $T_{19,2.236}$. Fortunately, many statistical software programs have this function.

Using such software for the Student t distribution with 19 DF and noncentrality parameter 2.236 gives the cdf $F_X(1.729) = 0.304828$.

Thus, the sought-after Type II error to three decimal places is 0.305. ■

SIMULATIONS

As we have seen, with enough distribution theory, the tools for finding Type II errors for the Student t test exist. Also, there are noncentral chi square and F distributions.

However, the assumption that the underlying data are normally distributed is necessary for such results. In the case of Type I errors, we have seen that the t test is somewhat robust with regard to the data deviating from normality. (See Section 7.4.)

In the case of the noncentral t, dealing with departures from normality presents significant analytical challenges. But the empirical approach of using simulations can bypass such difficulties and still give meaningful results.

To start, consider a simulation of the problem presented in Example 7.6.1. Suppose the data have a normal distribution with $\mu_0 = 5$ and $\sigma = 3$. The sample size is $n = 20$. Suppose we want to find the Type II error when the true $\delta = 2.236$. For the given $\sigma = 3$, this is equivalent to

$$2.236 = \frac{\mu_1 - \mu_0}{\sigma/\sqrt{n}} = \frac{\mu_1 - 5}{3/\sqrt{20}}$$

or $\mu_1 = 6.5$.

A Type II error occurs if the test statistic is less than 1.7291. In this case, H_0 would be accepted when rejection is the proper decision.

Using Minitab, two hundred samples of size 20 from the normal distribution with $\mu = 6.5$ and $\sigma^2 = 9$ are generated: Minitab produces a 200×20 array. For each row of the array, the test statistic $\frac{\bar{y}-5}{s/\sqrt{20}}$ is calculated and placed in column 21. If this value is less than 1.7291, a 1 is placed in that row of column 22; otherwise a 0 goes there. The sum of the entries in column 22 gives the observed number of Type II errors. Based on the computed value of the Type II error, 0.305, for the assumed value of δ, this observed number should be approximately $200(0.305) = 61$.

The Minitab simulation gave sixty-four observed Type II errors—a very close figure to what was expected.

The robustness for Type II errors can lead to analytical thickets. However, simulation can again shed some light on Type II errors in some cases. As an example, suppose the data are not normal, but gamma with $r = 4.694$ and $\lambda = 0.722$. Even though the distribution is skewed, these values make the mean $\mu = 6.5$ and the variance $\sigma^2 = 9$, as in the normal case above. Again relying on Minitab to give two hundred random samples of size 20, the observed number of Type II errors is sixty, so the test has some robustness for Type II errors in that case. Even though the data are not normal, the key statistic in the analysis, $\bar{y}$, will be approximately normal by the Central Limit Theorem.

If the distribution of the underlying data is unknown or extremely skewed, nonparametric tests, like the ones covered in Chapter 14 and in (31) are advised.

Appendix 7.A.1 Some Distribution Results for $\overline{Y}$ and S^2

Theorem 7.A.1.1 *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Define*

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Then

a. *$\overline{Y}$ and S^2 are independent.*

b. *$\frac{(n-1)S^2}{\sigma^2}$ has a chi square distribution with $n-1$ degrees of freedom.*

Proof The proof of this theorem relies on certain linear algebra techniques as well as a change-of-variables formula for multiple integrals. Definition 7.A.1.1 and the Lemma that follows review the necessary background results. For further details, see (49) or (226).

Definition 7.A.1.1

a. A matrix A is said to be *orthogonal* if $AA^T = I$.

b. Let β be any n-dimensional vector over the real numbers. That is, $\beta = (c_1, c_2, \ldots, c_n)$, where each c_j is a real number. The *length* of β will be defined as

$$\| \beta \| = \left(c_1^2 + \cdots + c_n^2\right)^{1/2}$$

(Note that $\| \beta \|^2 = \beta\beta^T$.)

Lemma

a. *A matrix A is orthogonal if and only if*

$$\| A\beta \| = \| \beta \| \quad \textit{for each } \beta$$

b. *If a matrix A is orthogonal, then det $A = 1$.*

c. *Let g be a one-to-one mapping with a continuous derivative on a subset, D, of n-space. Then*

$$\int_{g(D)} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = \int_D f[g(y_1, \ldots, y_n)] \ \det J(g)\, dy_1 \cdots dy_n$$

where $J(g)$ is the Jacobian of the transformation.

Set $X_i = (Y_i - \mu)/\sigma$ for $i = 1, 2, \ldots, n$. Then all the X_i's are $N(0, 1)$. Let A be an $n \times n$ orthogonal matrix whose last row is $\left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right)$. Let $\vec{X} = (X_1, \ldots, X_n)^T$ and define $\vec{Z} = (Z_1, Z_2, \ldots, Z_n)^T$ by the transformation $\vec{Z} = A\vec{X}$. (Note that $Z_n = \left(\frac{1}{\sqrt{n}}\right)X_1 + \cdots + \left(\frac{1}{\sqrt{n}}\right)X_n = \sqrt{n}\,\overline{X}$.)

For any set D,

$$\begin{aligned} P(\vec{Z} \in D) &= P(A\vec{X} \in D) = P(\vec{X} \in A^{-1}D) \\ &= \int_{A^{-1}D} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \cdots dx_n \\ &= \int_D f_{X_1,\ldots,X_n}[g(\vec{z})] \ \det J(g)\, dz_1 \cdots dz_n \\ &= \int_D f_{X_1,\ldots,X_n}(A^{-1}\vec{z}) \cdot 1 \cdot dz_1 \cdots dz_n \end{aligned}$$

where $g(\vec{z}) = A^{-1}\vec{z}$. But A^{-1} is orthogonal, so setting $(x_1, \ldots, x_n)^T = A^{-1}z$, we have that

$$x_1^2 + \cdots + x_n^2 = z_1^2 + \cdots + z_n^2$$

Thus

$$\begin{aligned} f_{X_1,\ldots,X_n}(\vec{x}) &= (2\pi)^{-n/2} e^{-(1/2)\left(x_1^2 + \cdots + x_n^2\right)} \\ &= (2\pi)^{-n/2} e^{-(1/2)\left(z_1^2 + \cdots + z_n^2\right)} \end{aligned}$$

From this we conclude that

$$P(\vec{Z} \in D) = \int_D (2\pi)^{-n/2} e^{-(n/2)\left(z_1^2 + \cdots + z_n^2\right)} dz_1 \cdots dz_n$$

implying that the Z_j's are independent standard normals.

Finally,

$$\sum_{j=1}^{n} Z_j^2 = \sum_{j=1}^{n-1} Z_j^2 + n\overline{X}^2 = \sum_{j=1}^{n} X_j^2 = \sum_{j=1}^{n} (X_j - \overline{X})^2 + n\overline{X}^2$$

Therefore,

$$\sum_{j=1}^{n-1} Z_j^2 = \sum_{j=1}^{n} (X_j - \overline{X})^2$$

and $\overline{X}^2$ (and thus $\overline{X}$) is independent of $\sum_{j=1}^{n} (X_i - \overline{X})^2$, so the conclusion follows for standard normal variables. Also, since $\overline{Y} = \sigma\overline{X} + \mu$ and $\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sigma^2 \sum_{i=1}^{n} (X_i - \overline{X})^2$, the conclusion follows for $N(\mu, \sigma^2)$ variables.

Comment As part of the proof just presented, we established a version of *Fisher's lemma:*

> Let $X_1, X_2, \ldots, X_n$ be independent standard normal random variables and let A be an orthogonal matrix. Define $(Z_1, \ldots, Z_n)^T = A(X_1, \ldots, X_n)^T$. Then the Z_i's are independent standard normal random variables.

Appendix 7.A.2 A Proof That the One-Sample *t* Test Is a GLRT

Theorem 7.A.2.1 *The one-sample t test, as outlined in Theorem 7.4.2, is a GLRT.*

Proof Consider the test of $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$. The two parameter spaces restricted to H_0 and $H_0 \cup H_1$—that is, ω and Ω, respectively—are given by

$$\omega = \{(\mu, \sigma^2): \mu = \mu_0;\quad 0 \leq \sigma^2 < \infty\}$$

and

$$\Omega = \{(\mu, \sigma^2): -\infty < \mu < \infty;\quad 0 \leq \sigma^2 < \infty\}$$

Without elaborating the details (see Example 5.2.4 for a very similar problem), it can be readily shown that, under ω,

$$\mu_e = \mu_0 \quad \text{and} \quad \sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \mu_0)^2$$

Under Ω,

$$\mu_e = \overline{y} \quad \text{and} \quad \sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \overline{y})^2$$

Therefore, since

$$L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2\right]$$

direct substitution gives

$$L(\omega_e) = \left[\frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{\sum_{i=1}^{n}(y_i-\mu_0)^2}}\right]^n e^{-n/2}$$

$$= \left[\frac{ne^{-1}}{2\pi\sum_{i=1}^{n}(y_i-\mu_0)^2}\right]^{n/2}$$

and

$$L(\Omega_e) = \left[\frac{ne^{-1}}{2\pi\sum_{i=1}^{n}(y_i-\overline{y})^2}\right]^{n/2}$$

From $L(\omega_e)$ and $L(\Omega_e)$ we get the likelihood ratio:

$$\lambda = \frac{L(\omega_e)}{L(\Omega_e)} = \left[\frac{\sum_{i=1}^{n}(y_i-\overline{y})^2}{\sum_{i=1}^{n}(y_i-\mu_0)^2}\right]^{n/2}, \quad 0 < \lambda \leq 1$$

As is often the case, it will prove to be more convenient to base a test on a monotonic function of λ, rather than on λ itself. We begin by rewriting the ratio's denominator:

$$\sum_{i=1}^{n}(y_i-\mu_0)^2 = \sum_{i=1}^{n}[(y_i-\overline{y})+(\overline{y}-\mu_0)]^2$$

$$= \sum_{i=1}^{n}(y_i-\overline{y})^2 + n(\overline{y}-\mu_0)^2$$

Therefore,

$$\lambda = \left[1+\frac{n(\overline{y}-\mu_0)^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2}\right]^{-n/2}$$

$$= \left(1+\frac{t^2}{n-1}\right)^{-n/2}$$

where

$$t = \frac{\overline{y}-\mu_0}{s/\sqrt{n}}$$

Observe that as t^2 increases, λ decreases. This implies that the original GLRT—which, by definition, would have rejected H_0 for any λ that was too small, say, less than λ^*—is equivalent to a test that rejects H_0 whenever t^2 is too large. But t is an observation of the random variable

$$T = \frac{\overline{Y}-\mu_0}{S/\sqrt{n}} \quad (= T_{n-1} \text{ by Theorem 7.3.5})$$

(Continued on next page)

(Theorem 7.A.2.1 continued)

Thus "too large" translates numerically into $t_{\alpha/2,n-1}$:

$$0 < \lambda \le \lambda^* \Leftrightarrow t^2 \ge (t_{\alpha/2,n-1})^2$$

But

$$t^2 \ge (t_{\alpha/2,n-1})^2 \Leftrightarrow t \le -t_{\alpha/2,n-1} \quad \text{or} \quad t \ge t_{\alpha/2,n-1}$$

and the theorem is proved.

Appendix 7.A.3 A Proof of Theorem 7.5.2

We begin by considering the test of H_0: $\sigma^2 = \sigma_0^2$ against a two-sided H_1. The relevant parameter spaces are

$$\omega = \left\{(\mu, \sigma^2): -\infty < \mu < \infty, \quad \sigma^2 = \sigma_0^2\right\}$$

and

$$\Omega = \{(\mu, \sigma^2): -\infty < \mu < \infty, \quad 0 \le \sigma^2\}$$

In both, the maximum likelihood estimate for μ is $\overline{y}$. In ω, the maximum likelihood estimate for σ^2 is simply σ_0^2; in Ω, $\sigma_e^2 = (1/n)\sum_{i=1}^{n}(y_i - \overline{y})^2$ (see Example 5.4.4). Therefore, the two likelihood functions, maximized over ω and over Ω, are

$$L(\omega_e) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \overline{y}}{\sigma_0}\right)^2\right]$$

and

$$L(\Omega_e) = \left[\frac{n}{2\pi\sum_{i=1}^{n}(y_i - \overline{y})^2}\right]^{n/2} \exp\left\{-\frac{n}{2}\sum_{i=1}^{n}\left[\frac{y_i - \overline{y}}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}\right]^2\right\}$$

$$= \left[\frac{n}{2\pi\sum_{i=1}^{n}(y_i - \overline{y})^2}\right]^{n/2} e^{-n/2}$$

It follows that the generalized likelihood ratio is given by

$$\lambda = \frac{L(\omega_e)}{L(\Omega_e)}$$

$$= \left[\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{n\sigma_0^2}\right]^{n/2} \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \overline{y}}{\sigma_0}\right)^2 + \frac{n}{2}\right]$$

$$= \left(\frac{\sigma_e^2}{\sigma_0^2}\right)^{n/2} \cdot e^{-(n/2)\left(\sigma_e^2/\sigma_0^2\right) + n/2}$$

We need to know the behavior of λ, considered as a function of (σ_e^2/σ_0^2). For simplicity, let $x = (\sigma_e^2/\sigma_0^2)$. Then $\lambda = x^{n/2}e^{-(n/2)x+n/2}$ and the inequality $\lambda \leq \lambda^*$ is equivalent to $xe^{-x} \leq e^{-1}(\lambda^*)^{2/n}$. The right-hand side is again an arbitrary constant, say, k^*. Figure 7.A.3.1 is a graph of $y = xe^{-x}$. Notice that the values of $x = (\sigma_e^2/\sigma_0^2)$ for which $xe^{-x} \leq k^*$, and equivalently $\lambda \leq \lambda^*$, fall into two regions, one for values of σ_e^2/σ_0^2 close to 0 and the other for values of σ_e^2/σ_0^2 much larger than 1. According to the likelihood ratio principle, we should reject H_0 for any $\lambda \leq \lambda^*$, where $P(\Lambda \leq \lambda^*|H_0) = \alpha$. But λ^* determines (via k^*) numbers a and b, so the critical region is $C = \{(\sigma_e^2/\sigma_0^2) : (\sigma_e^2/\sigma_0^2) \leq a \text{ or } (\sigma_e^2/\sigma_0^2) \geq b\}$.

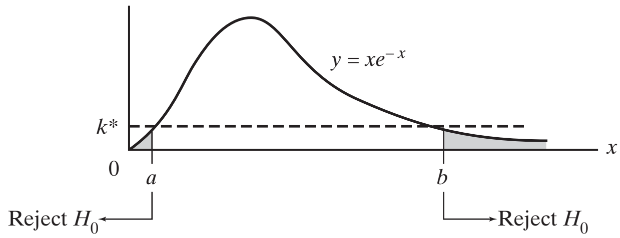


Figure 7.A.3.1

Comment At this point it is necessary to make a slight approximation. Just because $P(\Lambda \leq \lambda^*|H_0) = \alpha$, it does not follow that

$$P\left[\frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma_0^2} \leq a\right] = \frac{\alpha}{2} = P\left[\frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma_0^2} \geq b\right]$$

and, in fact, the two tails of the critical regions will *not* have exactly the same probability. Nevertheless, the two are numerically close enough that we will not substantially compromise the likelihood ratio criterion by setting each one equal to $\alpha/2$.

Note that

$$P\left[\frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma_0^2} \leq a\right] = P\left[\frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma_0^2} \leq na\right]$$

$$= P\left[\frac{(n-1)S^2}{\sigma_0^2} \leq na\right]$$

$$= P\left(\chi_{n-1}^2 \leq na\right)$$

and, similarly,

$$P\left[\frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{\sigma_0^2} \geq b\right] = P\left(\chi_{n-1}^2 \geq nb\right)$$

Thus we will choose as critical values $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ and reject H_0 if either

$$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{\alpha/2,n-1}$$

or

$$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{1-\alpha/2,n-1}$$

See Figure 7.A.3.2.

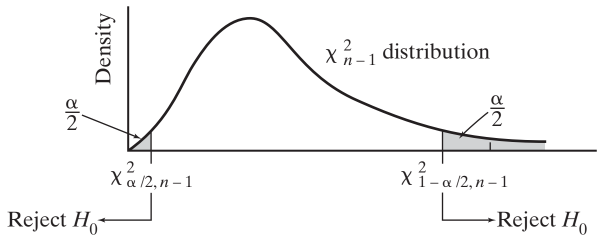


Figure 7.A.3.2

Comment One-sided tests for dispersion are set up in a similar fashion. In the case of

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

versus

$$H_1: \sigma^2 < \sigma_0^2$$

H_0 is rejected if

$$\frac{(n-1)s^2}{\sigma_0^2} \leq \chi^2_{\alpha,n-1}$$

For

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

versus

$$H_1: \sigma^2 > \sigma_0^2$$

H_0 is rejected if

$$\frac{(n-1)s^2}{\sigma_0^2} \geq \chi^2_{1-\alpha,n-1}$$