11. Explain why, when we sample with replacement from a finite population, the results of Theorem 1 apply rather than those of Theorem 6.

12. This question has been intentionally omitted for this edition.

13. Use MINITAB or some other statistical computer program to generate 20 samples of size 10 each from the uniform density function $f(x) = 1, 0 \leq x \leq 1$.
(a) Find the mean of each sample and construct a histogram of these sample means.
(b) Calculate the mean and the variance of the 20 sample means.

14. Referring to Exercise 13, now change the sample size to 30.
(a) Does this histogram more closely resemble that of a normal distribution than that of Exercise 13? Why?
(b) Which resembles it more closely?
(c) Calculate the mean and the variance of the 20 sample means.

15. If a random sample of size n is selected without replacement from the finite population that consists of the integers $1, 2, \ldots, N$, show that
(a) the mean of $\overline{X}$ is $\dfrac{N+1}{2}$;
(b) the variance of $\overline{X}$ is $\dfrac{(N+1)(N-n)}{12n}$;
(c) the mean and the variance of $Y = n \cdot \overline{X}$ are

$$E(Y) = \frac{n(N+1)}{2} \quad \text{and} \quad \text{var}(Y) = \frac{n(N+1)(N-n)}{12}$$

16. Find the mean and the variance of the finite population that consists of the 10 numbers 15, 13, 18, 10, 6, 21, 7, 11, 20, and 9.

17. Show that the variance of the finite population $\{c_1, c_2, \ldots, c_N\}$ can be written as

$$\sigma^2 = \frac{\sum_{i=1}^{N} c_i^2}{N} - \mu^2$$

Also, use this formula to recalculate the variance of the finite population of Exercise 16.

18. Show that, analogous to the formula of Exercise 17, the formula for the sample variance can be written as

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n-1} - \frac{n\overline{X}^2}{n-1}$$

Also, use this formula to calculate the variance of the following sample data on the number of service calls received by a tow truck operator on eight consecutive working days: 13, 14, 13, 11, 15, 14, 17, and 11.

19. Show that the formula for the sample variance can be written as

$$S^2 = \frac{n\left(\sum_{i=1}^{n} X_i^2\right) - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

Also, use this formula to recalculate the variance of the sample data of Exercise 18.

4 The Chi-Square Distribution

If X has the standard normal distribution, then X^2 has the special gamma distribution, which is referred to as the **chi-square distribution**, and this accounts for the important role that the chi-square distribution plays in problems of sampling from normal populations. Theorem 11 will show the importance of this distribution in making inferences about sample variances.

The chi-square distribution is often denoted by "χ^2 distribution," where χ is the lowercase Greek letter *chi*. We also use χ^2 for values of random variables having chi-square distributions, but we shall refrain from denoting the corresponding random variables by X^2, where X is the capital Greek letter *chi*. This avoids having to reiterate in each case whether X is a random variable with values x or a random variable with values χ.

If a random variable X has the chi-square distribution with ν degrees of freedom if its probability density is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-x/2} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The mean and the variance of the chi-square distribution with ν degrees of freedom are ν and 2ν, respectively, and its moment-generating function is given by

$$M_X(t) = (1-2t)^{-\nu/2}$$

The chi-square distribution has several important mathematical properties, which are given in Theorems 7 through 10.

THEOREM 7. If X has the standard normal distribution, then X^2 has the chi-square distribution with $\nu = 1$ degree of freedom.

More generally, let us prove the following theorem.

THEOREM 8. If $X_1, X_2, \ldots, X_n$ are independent random variables having standard normal distributions, then

$$Y = \sum_{i=1}^{n} X_i^2$$

has the chi-square distribution with $\nu = n$ degrees of freedom.

Proof Using the moment-generating function given previously with $\nu = 1$ and Theorem 7, we find that

$$M_{X_i^2}(t) = (1-2t)^{-\frac{1}{2}}$$

and it follows the theorem "$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$" that

$$M_Y(t) = \prod_{i=1}^{n} (1-2t)^{-\frac{1}{2}} = (1-2t)^{-\frac{n}{2}}$$

This moment-generating function is readily identified as that of the chi-square distribution with $\nu = n$ degrees of freedom.

Two further properties of the chi-square distribution are given in the two theorems that follow; the reader will be asked to prove them in Exercises 20 and 21.

THEOREM 9. If $X_1, X_2, \ldots, X_n$ are independent random variables having chi-square distributions with $\nu_1, \nu_2, \ldots, \nu_n$ degrees of freedom, then

$$Y = \sum_{i=1}^{n} X_i$$

has the chi-square distribution with $\nu_1 + \nu_2 + \cdots + \nu_n$ degrees of freedom.

THEOREM 10. If X_1 and X_2 are independent random variables, X_1 has a chi-square distribution with ν_1 degrees of freedom, and $X_1 + X_2$ has a chi-square distribution with $\nu > \nu_1$ degrees of freedom, then X_2 has a chi-square distribution with $\nu - \nu_1$ degrees of freedom.

The chi-square distribution has many important applications. Foremost are those based, directly or indirectly, on the following theorem.

THEOREM 11. If $\overline{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the standard deviation σ, then

1. $\overline{X}$ and S^2 are independent;
2. the random variable $\dfrac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with $n-1$ degrees of freedom.

Proof Since a detailed proof of part 1 would go beyond the scope of this chapter we shall assume the independence of $\overline{X}$ and S^2 in our proof of part 2. In addition to the references to proofs of part 1 at the end of this chapter, Exercise 31 outlines the major steps of a somewhat simpler proof based on the idea of a conditional moment-generating function, and in Exercise 30 the reader will be asked to prove the independence of $\overline{X}$ and S^2 for the special case where $n = 2$.

To prove part 2, we begin with the identity

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2 + n(\overline{X}-\mu)^2$$

which the reader will be asked to verify in Exercise 22. Now, if we divide each term by σ^2 and substitute $(n-1)S^2$ for $\sum_{i=1}^{n}(X_i-\overline{X})^2$, it follows that

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

With regard to the three terms of this identity, we know from Theorem 8 that the one on the left-hand side of the equation is a random variable having a chi-square distribution with n degrees of freedom. Also, according to Theorems 4 and 7, the second term on the right-hand side of the equation is a random variable having a chi-square distribution with 1 degree of freedom. Now, since $\overline{X}$ and S^2 are assumed to be independent, it follows that the two terms on the right-hand side of the equation are independent, and we conclude that $\dfrac{(n-1)S^2}{\sigma^2}$ is a random variable having a chi-square distribution with $n-1$ degrees of freedom.

Since the chi-square distribution arises in many important applications, integrals of its density have been extensively tabulated. Table V of "Statistical Tables" contains values of $\chi^2_{\alpha,\nu}$ for $\alpha = 0.995, 0.99, 0.975, 0.95, 0.05, 0.025, 0.01, 0.005$, and $\nu = 1, 2, \ldots, 30$, where $\chi^2_{\alpha,\nu}$ is such that the area to its right under the chi-square curve with ν degrees of freedom (see Figure 1) is equal to α. That is, $\chi^2_{\alpha,\nu}$ is such that if X is a random variable having a chi-square distribution with ν degrees of freedom, then

$$P(X \geqq \chi^2_{\alpha,\nu}) = \alpha$$

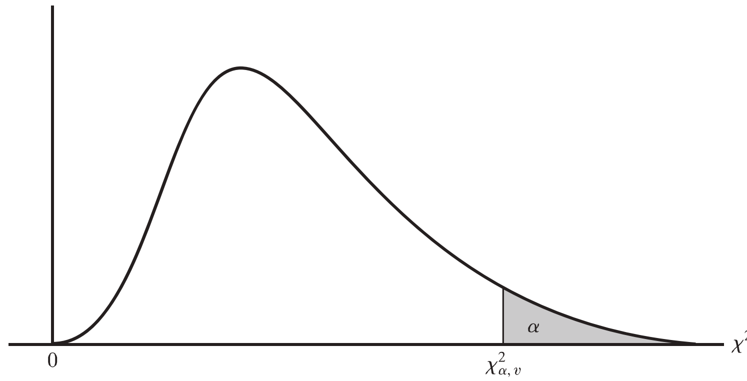

Figure 1. Chi-square distribution.

When ν is greater than 30, Table V of "Statistical Tables" cannot be used and probabilities related to chi-square distributions are usually approximated with normal distributions, as in Exercise 25 or 28.

EXAMPLE 2

Suppose that the thickness of a part used in a semiconductor is its critical dimension and that the process of manufacturing these parts is considered to be under control if the true variation among the thicknesses of the parts is given by a standard deviation not greater than $\sigma = 0.60$ thousandth of an inch. To keep a check on the process, random samples of size $n = 20$ are taken periodically, and it is regarded to be "out of control" if the probability that S^2 will take on a value greater than or equal to the observed sample value is 0.01 or less (even though $\sigma = 0.60$). What can one conclude about the process if the standard deviation of such a periodic random sample is $s = 0.84$ thousandth of an inch?

Solution

The process will be declared "out of control" if $\dfrac{(n-1)s^2}{\sigma^2}$ with $n = 20$ and $\sigma = 0.60$ exceeds $\chi^2_{0.01,19} = 36.191$. Since

$$\frac{(n-1)s^2}{\sigma^2} = \frac{19(0.84)^2}{(0.60)^2} = 37.24$$

exceeds 36.191, the process is declared out of control. Of course, it is assumed here that the sample may be regarded as a random sample from a normal population.

---

5 The t Distribution

In Theorem 4 we showed that for random samples from a normal population with the mean μ and the variance σ^2, the random variable $\overline{X}$ has a normal distribution with the mean μ and the variance $\dfrac{\sigma^2}{n}$; in other words,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

has the standard normal distribution. This is an important result, but the major difficulty in applying it is that in most realistic applications the population standard deviation σ is unknown. This makes it necessary to replace σ with an estimate, usually with the value of the sample standard deviation S. Thus, the theory that follows leads to the exact distribution of $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ for random samples from normal populations.

To derive this sampling distribution, let us first study the more general situation treated in the following theorem.

THEOREM 12. If Y and Z are independent random variables, Y has a chi-square distribution with ν degrees of freedom, and Z has the standard normal distribution, then the distribution of

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

is given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{for } -\infty < t < \infty$$

and it is called the ***t* distribution with *ν* degrees of freedom**.

Proof Since Y and Z are independent, their joint probability density is given by

$$f(y,z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2} \cdot \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}}y^{\frac{\nu}{2}-1}e^{-\frac{y}{2}}$$

for $y > 0$ and $-\infty < z < \infty$, and $f(y,z) = 0$ elsewhere. Then, to use the change-of-variable technique, we solve $t = \dfrac{z}{\sqrt{y/\nu}}$ for z, getting $z = t\sqrt{y/\nu}$ and hence $\dfrac{\partial z}{\partial t} = \sqrt{y/\nu}$. Thus, the joint density of Y and T is given by

$$g(y,t) = \begin{cases} \dfrac{1}{\sqrt{2\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}}y^{\frac{\nu-1}{2}}e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)} & \text{for } y > 0 \text{ and } -\infty < t < \infty \\ 0 & \text{elsewhere} \end{cases}$$

and, integrating out y with the aid of the substitution $w = \dfrac{y}{2}\left(1+\dfrac{t^2}{\nu}\right)$, we finally get

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{for } -\infty < t < \infty$$

The t distribution was introduced originally by W. S. Gosset, who published his scientific writings under the pen name "Student," since the company for which he worked, a brewery, did not permit publication by employees. Thus, the t distribution is also known as the **Student *t* distribution**, or **Student's *t* distribution**. As shown in Figure 2, graphs of t distributions having different numbers of degrees of freedom resemble that of the standard normal distribution, but have larger variances. In fact, for large values of υ, the t distribution approaches the standard normal distribution.

In view of its importance, the t distribution has been tabulated extensively. Table IV of "Statistical Tables", for example, contains values of $t_{\alpha,\nu}$ for $\alpha = 0.10, 0.05, 0.025, 0.01, 0.005$ and $\nu = 1, 2, \ldots, 29$, where $t_{\alpha,\nu}$ is such that the area to its right under the curve of the t distribution with ν degrees of freedom (see Figure 3) is equal to α. That is, $t_{\alpha,\nu}$ is such that if T is a random variable having a t distribution with ν degrees of freedom, then

$$P(T \geqq t_{\alpha,\nu}) = \alpha$$

The table does not contain values of $t_{\alpha,\nu}$ for $\alpha > 0.50$, since the density is symmetrical about $t = 0$ and hence $t_{1-\alpha,\nu} = -t_{\alpha,\nu}$. When ν is 30 or more, probabilities related to the t distribution are usually approximated with the use of normal distributions (see Exercise 35).

Among the many applications of the t distribution, its major application (for which it was originally developed) is based on the following theorem.

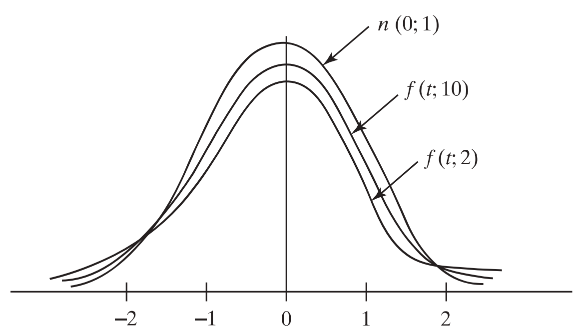

Figure 2. Comparison of t distributions and standard normal distribution.

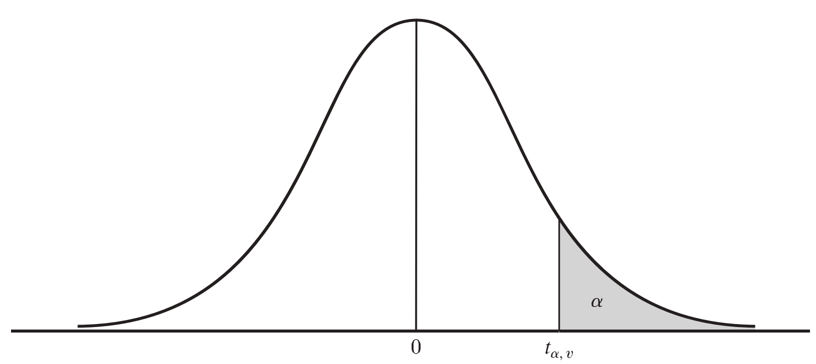

Figure 3. t distribution.

THEOREM 13. If $\overline{X}$ and S^2 are the mean and the variance of a random sample of size n from a normal population with the mean μ and the variance σ^2, then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

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

Proof By Theorems 11 and 4, the random variables

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

have, respectively, a chi-square distribution with $n-1$ degrees of freedom and the standard normal distribution. Since they are also independent by part 1 of Theorem 11, substitution into the formula for T of Theorem 12 yields

$$T = \frac{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

and this completes the proof.

EXAMPLE 3

In 16 one-hour test runs, the gasoline consumption of an engine averaged 16.4 gallons with a standard deviation of 2.1 gallons. Test the claim that the average gasoline consumption of this engine is 12.0 gallons per hour.

Solution

Substituting $n = 16, \mu = 12.0$, $\overline{x} = 16.4$, and $s = 2.1$ into the formula for t in Theorem 13, we get

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{16.4 - 12.0}{2.1/\sqrt{16}} = 8.38$$

Since Table IV of "Statistical Tables" shows that for $\nu = 15$ the probability of getting a value of T greater than 2.947 is 0.005, the probability of getting a value greater than 8 must be negligible. Thus, it would seem reasonable to conclude that the true average hourly gasoline consumption of the engine exceeds 12.0 gallons.

6 The *F* Distribution

Another distribution that plays an important role in connection with sampling from normal populations is the F distribution, named after Sir Ronald A. Fisher, one of the most prominent statisticians of the last century. Originally, it was studied as the sampling distribution of the ratio of two independent random variables with chi-square distributions, each divided by its respective degrees of freedom, and this is how we shall present it here.

Fisher's F distribution is used to draw statistical inferences about the ratio of two sample variances. As such, it plays a key role in the analysis of variance, used in conjunction with experimental designs.

THEOREM 14. If U and V are independent random variables having chi-square distributions with ν_1 and ν_2 degrees of freedom, then

$$F = \frac{U/\nu_1}{V/\nu_2}$$

is a random variable having an ***F* distribution**, that is, a random variable whose probability density is given by

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}$$

for $f > 0$ and $g(f) = 0$ elsewhere.

Proof By virtue of independence, the joint density of U and V is given by

$$f(u,v) = \frac{1}{2^{\nu_1/2}\Gamma\left(\frac{\nu_1}{2}\right)} \cdot u^{\frac{\nu_1}{2}-1}e^{-\frac{u}{2}} \cdot \frac{1}{2^{\nu_2/2}\Gamma\left(\frac{\nu_2}{2}\right)} \cdot v^{\frac{\nu_2}{2}-1}e^{-\frac{v}{2}}$$

$$= \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot u^{\frac{\nu_1}{2}-1}v^{\frac{\nu_2}{2}-1}e^{-\frac{u+v}{2}}$$

for $u > 0$ and $v > 0$, and $f(u,v) = 0$ elsewhere. Then, to use the change-of-variable technique, we solve

$$f = \frac{u/\nu_1}{v/\nu_2}$$

for u, getting $u = \frac{\nu_1}{\nu_2} \cdot vf$ and hence $\frac{\partial u}{\partial f} = \frac{\nu_1}{\nu_2} \cdot v$. Thus, the joint density of F and V is given by

$$g(f,v) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot f^{\frac{\nu_1}{2}-1}v^{\frac{\nu_1+\nu_2}{2}-1}e^{-\frac{v}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right)}$$

for $f > 0$ and $v > 0$, and $g(f,v) = 0$ elsewhere. Now, integrating out v by making the substitution $w = \frac{v}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right)$, we finally get

$$g(f) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1+\frac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}$$

for $f > 0$, and $g(f) = 0$ elsewhere.

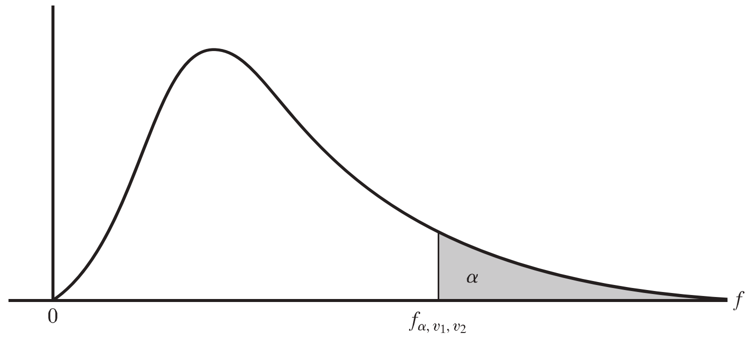

Figure 4. *F* distribution.

In view of its importance, the F distribution has been tabulated extensively. Table VI of "Statistical Tables", for example, contains values of f_{α,ν_1,ν_2} for $\alpha = 0.05$ and 0.01 and for various values of ν_1 and ν_2, where f_{α,ν_1,ν_2} is such that the area to its right under the curve of the F distribution with ν_1 and ν_2 degrees of freedom (see Figure 4) is equal to α. That is, f_{α,ν_1,ν_2} is such that

$$P(F \geqq f_{\alpha,\nu_1,\nu_2}) = \alpha$$

Applications of Theorem 14 arise in problems in which we are interested in comparing the variances σ_1^2 and σ_2^2 of two normal populations; for instance, in problems in which we want to estimate the ratio $\frac{\sigma_1^2}{\sigma_2^2}$ or perhaps to test whether $\sigma_1^2 = \sigma_2^2$. We base such inferences on **independent random samples** of sizes n_1 and n_2 from the two populations and Theorem 11, according to which

$$\chi_1^2 = \frac{(n_1 - 1)s_1^2}{\sigma_1^2} \quad \text{and} \quad \chi_2^2 = \frac{(n_2 - 1)s_2^2}{\sigma_2^2}$$

are values of random variables having chi-square distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom. By "independent random samples," we mean that the $n_1 + n_2$ random variables constituting the two random samples are all independent, so that the two chi-square random variables are independent and the substitution of their values for U and V in Theorem 14 yields the following result.

THEOREM 15. If S_1^2 and S_2^2 are the variances of independent random samples of sizes n_1 and n_2 from normal populations with the variances σ_1^2 and σ_2^2, then

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

is a random variable having an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom.

The F distribution is also known as the **variance-ratio distribution**.