10.3: Difference between Two Variances - the F Distributions

Here we have to assume that the two *populations* (as opposed to sample mean distributions) have a distribution that is almost normal as shown in Figure 10.2.

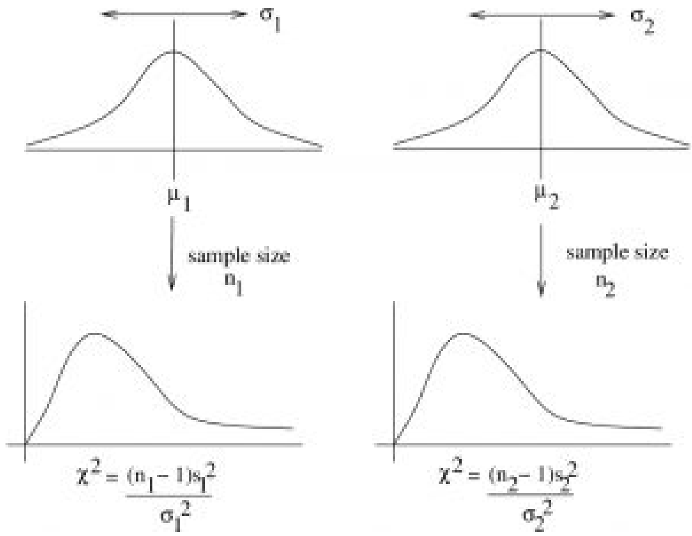


Figure 10.2: Two normal populations lead to two χ^2 distributions that represent distributions of sample variances. The F distribution results when you build up a distribution of the ratio of the two χ^2 sample values.

The ratio $\frac{s_1^2}{s_2^2}$ follows an F-distribution if $\sigma_1 = \sigma_2$. That F distribution has two degrees of freedom: one for the numerator (d.f.N. or ν_1) and one for the denominator (d.f.D. or ν_2). So we denote the distribution more specifically as F_{ν_1,ν_2}. For the case of Figure 10.2, $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. The F ratio, in general is the result of the following stochastic process. Let X_1 be random variable produced by a stochastic process with a $\chi^2_{\nu_1}$ distribution and let X_2 be random variable produced by a stochastic process with a $\chi^2_{\nu_2}$ distribution. Then the random variable $F = X_1/X_2$ will, by definition, have a F_{ν_1,ν_2} distribution.

The exact shape of the F_{ν_1,ν_2} distribution depends on the choice of ν_1 and ν_2, But it roughly looks like a χ^2 distribution as shown in Figure 10.3.

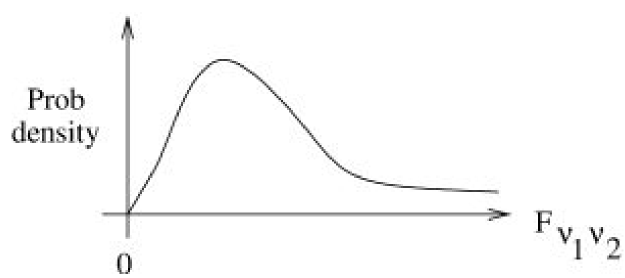


Figure 10.3: A generic F distribution.

F and t are related :

$$F_{1,\nu} = t_\nu^2 \tag{10.3.1}$$

so the t statistic can be viewed as a special case of the F statistic.

For comparing variances, we are interested in the follow hypotheses pairs :

Right-tailed	Left-tailed	Two-tailed
$H_0 : \sigma_1^2 \leq \sigma_2^2$	$H_0 : \sigma_1^2 \geq \sigma_2^2$	$H_0 : \sigma_1^2 = \sigma_2^2$
$H_1 : \sigma_1^2 > \sigma_2^2$	$H_1 : \sigma_1^2 < \sigma_2^2$	$H_1 : \sigma_1^2 \neq \sigma_2^2$

We'll always compare variances (σ^2) and not standard deviations (σ) to keep life simple.

The test statistic is

$$F_{\text{test}} = F_{\nu_1,\nu_2} = \frac{s_1^2}{s_2^2} \tag{10.3.2}$$

where (for finding the critical statistic), $\mu_1 = n_1 - 1$ and $\mu_2 = n_2 - 1$.

Note that $F_{\nu_1,\nu_2} = 1$ when $s_1^2 = s_2^2$, a fact you can use to get a feel for the meaning of this test statistic.

Values for the various F critical values are given in the ***F* Distribution Table** in the Appendix. We will denote a critical value of F with the notation :

$$F_{\text{crit}} = F_{\alpha,\quad \nu_1,\quad \nu_2} \tag{10.3.3}$$

Where:

α = Type I error rate
ν_1 = d.f.N.
ν_2 = d.f.D.

The ***F* Distribution Table** gives critical values for small right tail areas only. This means that they are useless for a left-tailed test. But that does not mean we cannot do a left-tail test. A left-tail test is easily converted into a right tail test by switching the assignments of populations 1 and 2. To get the assignments correct in the first place then, always define populations 1 and 2 so that $\sigma_1^2 > \sigma_2^2$. Assign population 1 so that it has the largest sample variance. Do this even for a two-tail test because we will have no idea what F_{crit} on the left side of the distribution is.

Example 10.3 : Given the following data for smokers and non-smokers (maybe its about some sort of disease occurrence, who cares, let's focus on dealing with the numbers), test if the population variances are equal or not at $\alpha = 0.05$.

Smokers	Nonsmokers
$n_1 = 26$	$n_2 = 18$
$s_1^2 = 36$	$s_2^2 = 10$

Note that $s_1^2 > s_2^2$ so we're good to go.

Solution :

1. Hypothesis.

$$\$H_0 : \sigma_1^2 = \sigma_2^2\$$$
$$\$H_1 : \sigma_1^2 \neq \sigma_2^2\$$$

2. Critical statistic.

Use the ***F* Distribution Table**; it is a bunch of tables labeled by "α" that we will designate at α_T, the table values that signify right tail areas. Since this is a two-tail test, we need $alpha_T = \alpha/2$. Next we need the degrees of freedom:

$$\text{d.f.N.} = \nu_1 = n_1 - 1 = 26 - 1 = 25 \tag{10.3.4}$$

$$\text{d.f.D.} = \nu_2 = n_2 - 1 = 18 - 1 = 17 \tag{10.3.5}$$

So the critical statistic is

Rendered by QuickLaTeX.com

3. Test statistic.

$$F_{\nu_1,\nu_2} = \frac{s_1^2}{s_2^2} \tag{10.3.6}$$

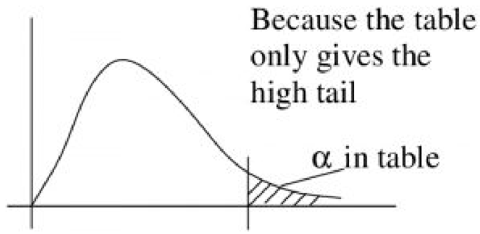
$$F_{\text{test}} = F_{25,17} = \frac{36}{10} = 3.6 \tag{10.3.7}$$

With this test statistic, we can estimate the p-value using the ***F* Distribution Table**. To find p, look up all the numbers with d.f.N = 25 and d.f.N = 17 (24 & 17 are the closest in the tables so use those) in all the the ***F* Distribution Table** and form your own table. For each column in your table record α_T and the F value corresponding to the degrees of freedom of interest. Again, α_T corresponds to $p/2$ for a two-tailed test. So make a row above the α_T row with $p = 2\alpha_T$. (For a one-tailed test, we would put $p = \alpha_T$.)

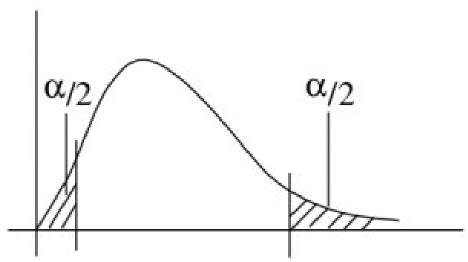
p	0.20 0.10 0.05 0.02 0.01
α_T	0.10 0.05 0.025 0.01 0.005
F	1.84 2.19 2.56 3.08 3.51 3.6 is over here somewhere so $p < 0.01$

Notice how we put an upper limit on p because F_{test} was larger than all the F values in our little table.

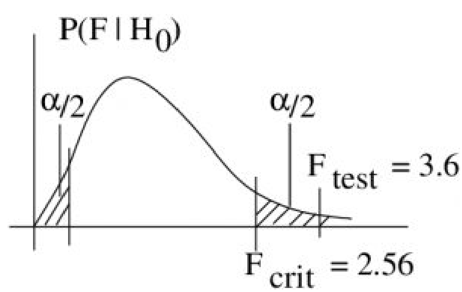
Let's take a graphical look at why we use $p = 2\alpha$ in the little table and $\alpha_T = \alpha/2$ for finding F_{crit} for two tailed tests :



But in a two-tailed test we want α split on both sides:



4. Decision.



Reject H_0. The p-value estimate supports this :

$$(p < 0.01) < (\alpha = 0.05) \tag{10.3.8}$$

5. Interpretation.

There is enough evidence to conclude, at $\alpha = 0.05$ with an F-test, that the variance of the smoker population is different from the non-smoker population.

□