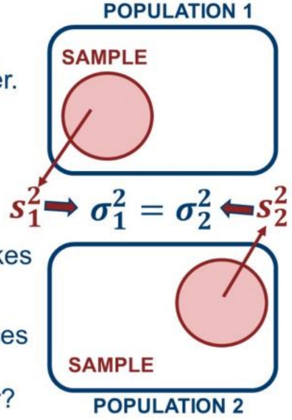
Comparing population variances

We want to know if population variances differ.
We can't measure the populations.
We take random samples.
We calculate sample variances.

The sample variances are estimates of the population variances, but sampling error makes them inexact.

What are the chances the population variances are the same (i.e., H_0), based on how much the sample variances differ from one another?



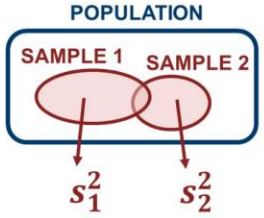
The F distribution

On average, two samples from the same population should be equal.

But sampling error causes them to differ.

We can measure different with a ratio.

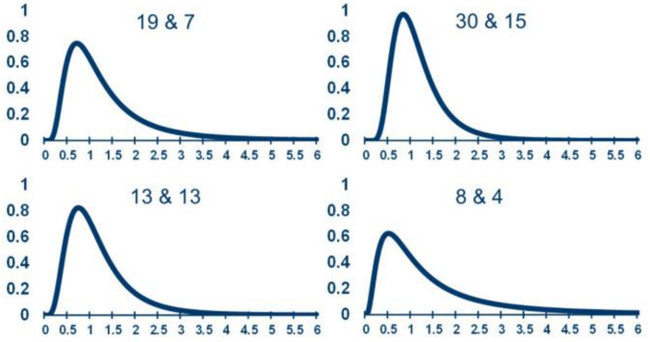
The probability distribution of these ratios is the F distribution.



Ratio 1: $F = \frac{s_1^2}{s_2^2}$ Ratio 2: $F = \frac{s_2^2}{s_1^2}$

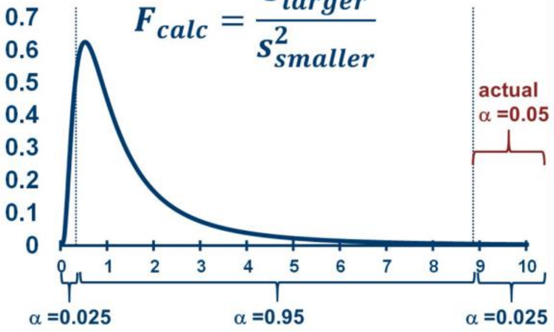
The F distribution

There is a different F distribution for every combination of degrees of freedom.

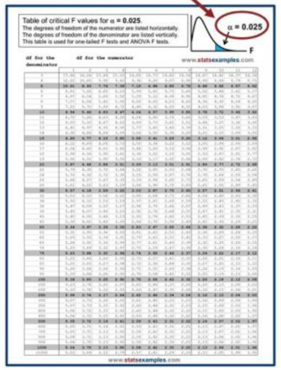


The variance ratio F test is two-tailed (in disguise)

$$F_{calc} = \frac{s_{larger}^2}{s_{smaller}^2}$$



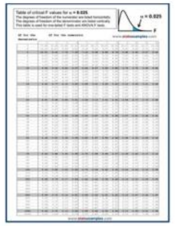
α =0.025



The variance ratio F test formal procedure

- Create null and alternative hypotheses:
  H_0: $\sigma_1^2 = \sigma_2^2$
  H_A: $\sigma_1^2 \neq \sigma_2^2$
- Calculate F ratio $$F_{calc} = \frac{s_{larger}^2}{s_{smaller}^2}$$
- Compare F_{calc} to various F_{crit} values (α=0.025 or α= 0.05)
- Determine probability, **p value**, of seeing F_{calc} as large as we do.
- Decide to "reject H_0" or "fail to reject H_0" based on the p value.
  H_0: $\sigma_1^2 = \sigma_2^2$ consistent with non-small p values.
  H_A: $\sigma_1^2 \neq \sigma_2^2$ would give us small p values.

Columns $df_{numerator}$
Rows $df_{denominator}$

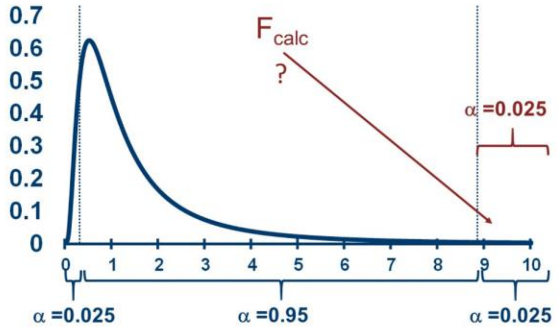


The F distribution

e.g., for df_1=8, df_2=4 this is the F distribution.

95% of the time F is in the center 95%

p<0.025 that F<0.198
p<0.025 that F>8.980

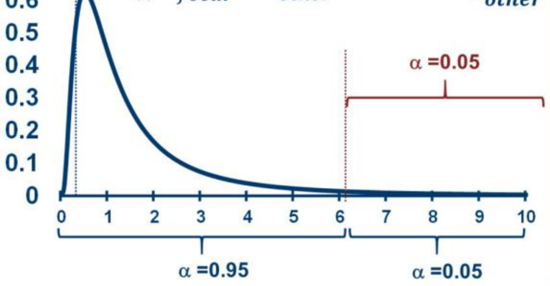


We could make tables with values at both ends ... or use top **α=0.025** if we always divide larger variance by smaller variance (which flips those small values over to the right side).

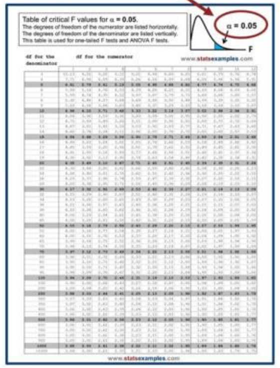
Sometimes the F test is one-tailed (e.g., ANOVA)

H_0: $\sigma_{focal}^2 \leq \sigma_{other}^2$
H_A: $\sigma_{focal}^2 > \sigma_{other}^2$

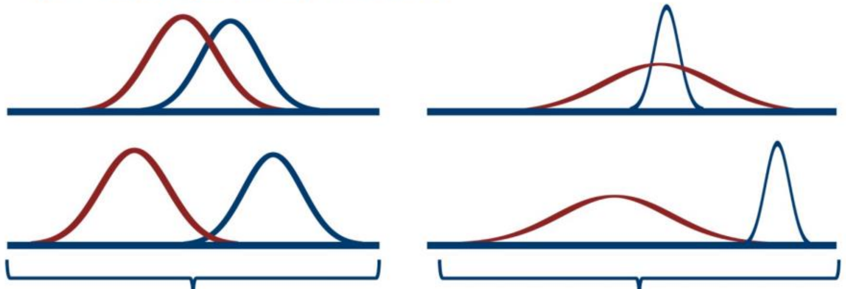
$$F_{calc} = \frac{s_{focal}^2}{s_{other}^2}$$



α =0.05



The F test only tests the variances



These will give same large p value (p>0.05)

These will give same small p value (p<0.05)

Caution about the F test

- F test requires normal population distributions. If the populations are not normally distributed, the test can easily give type I or II errors.

Why use the F test?

- Two-tailed F test is a pre-test for homoscedastic t-tests or other two-sample tests that require equality of variances.
- One-tailed F test is the basis of the ANOVA technique (which itself is the basis of correlation and regression).

StatsExamples.com