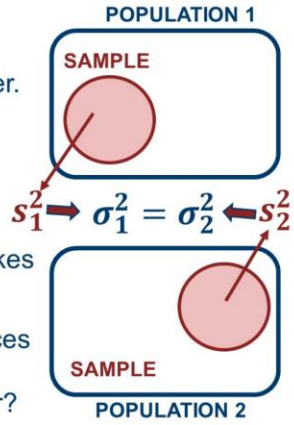


Comparing population variances

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?

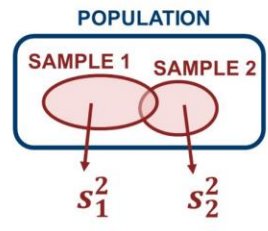


The F distribution

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.

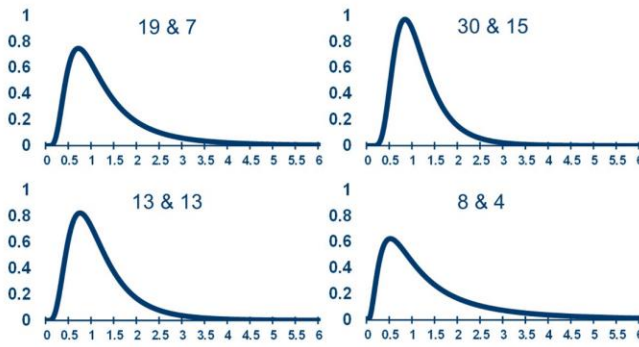
The probability distribution of these ratios is the F distribution.



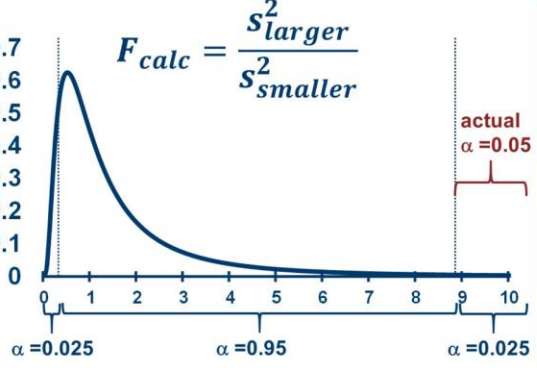
Ratio 1: $F = \frac{s_1^2}{s_2^2}$ Ratio 2: $F = \frac{s_2^2}{s_1^2}$

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)



$\alpha = 0.025$

Table of critical F values for $\alpha = 0.025$. The degrees of freedom of the numerator are listed horizontally. The degrees of freedom of the denominator are listed vertically. This table is used for one-tailed F tests and ANOVA F tests.

df numerator \ df denominator	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	227.990	236.767	243.906	249.590	254.418	258.585	262.191
2	18.5128	16.0000	15.0000	14.1801	13.5803	13.1008	12.7188	12.3951	12.1114	11.8594
3	10.1286	9.00000	8.45138	8.01080	7.65039	7.34142	7.06329	6.81006	6.57770	6.36188
4	7.70853	7.00000	6.59089	6.21070	5.87035	5.57064	5.30329	5.06106	4.83970	4.63488
5	6.59126	6.00000	5.63082	5.28030	4.96000	4.67035	4.41329	4.18106	3.96970	3.77488
6	5.96446	5.50000	5.17082	4.84030	4.54000	4.27035	4.02329	3.79106	3.57970	3.39488
7	5.59126	5.15000	4.84082	4.52030	4.24000	3.98035	3.74329	3.52106	3.31970	3.14488
8	5.33126	4.95000	4.65082	4.34030	4.06000	3.81035	3.58329	3.37106	3.17970	3.01488
9	5.13126	4.80000	4.51082	4.21030	3.94000	3.69035	3.47329	3.27106	3.08970	2.93488
10	4.97126	4.70000	4.42082	4.13030	3.86000	3.62035	3.41329	3.22106	3.04970	2.90488

The variance ratio F test formal procedure

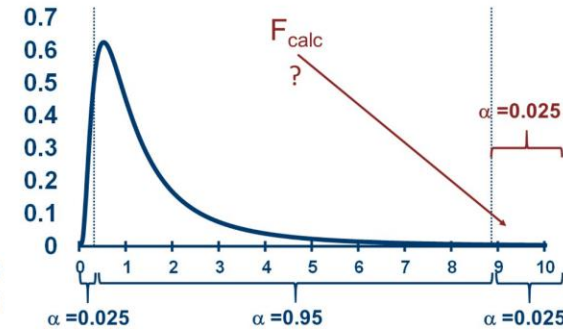
- ▶ Create null and alternative hypotheses: $H_0: \sigma_1^2 = \sigma_2^2$, $H_A: \sigma_1^2 \neq \sigma_2^2$
 - ▶ Calculate F ratio: $F_{calc} = \frac{s_{larger}^2}{s_{smaller}^2}$
 - ▶ Compare F_{calc} to various F_{crit} values ($\alpha=0.025$ or $\alpha=0.05$)
 - ▶ Determine probability, **p value**, 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

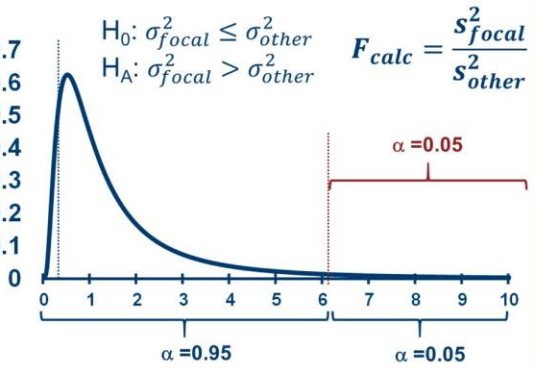
e.g., for $df_1=8$, $df_2=4$ this is the F distribution.

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 $\alpha=0.025$ 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)

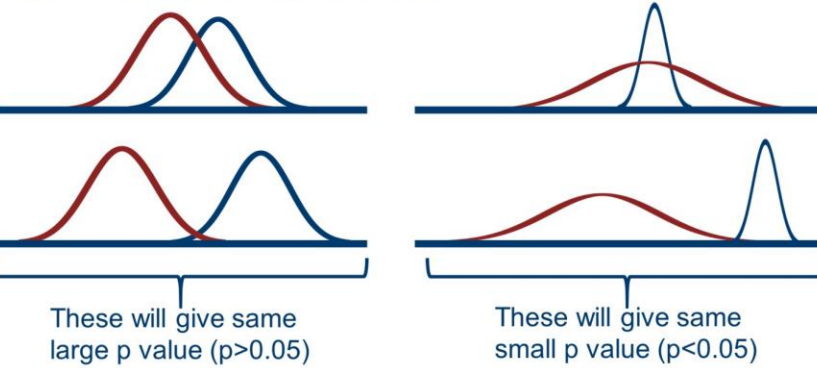


$\alpha = 0.05$

Table of critical F values for $\alpha = 0.05$. The degrees of freedom of the numerator are listed horizontally. The degrees of freedom of the denominator are listed vertically. This table is used for one-tailed F tests and ANOVA F tests.

df numerator \ df denominator	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	227.990	236.767	243.906	249.590	254.418	258.585	262.191
2	18.5128	16.0000	15.0000	14.1801	13.5803	13.1008	12.7188	12.3951	12.1114	11.8594
3	10.1286	9.00000	8.45138	8.01080	7.65039	7.34142	7.06329	6.81006	6.57770	6.36188
4	7.70853	7.00000	6.59089	6.21070	5.87035	5.57064	5.30329	5.06106	4.83970	4.63488
5	6.59126	6.00000	5.63082	5.28030	4.96000	4.67035	4.41329	4.18106	3.96970	3.77488
6	5.96446	5.50000	5.17082	4.84030	4.54000	4.27035	4.02329	3.79106	3.57970	3.39488
7	5.59126	5.15000	4.84082	4.52030	4.24000	3.98035	3.74329	3.52106	3.31970	3.14488
8	5.33126	4.95000	4.65082	4.34030	4.06000	3.81035	3.58329	3.37106	3.17970	3.01488
9	5.13126	4.80000	4.51082	4.21030	3.94000	3.69035	3.47329	3.27106	3.08970	2.93488
10	4.97126	4.70000	4.42082	4.13030	3.86000	3.62035	3.41329	3.22106	3.04970	2.90488

The F test only tests the variances



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).