F-Test (ANOVA)
&
Two-Way ANOVA

What is F-test?

An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fit to a data set, in order to identify the model that best fits the population from which the data were sampled.

The name was coined by George W. Snedecor, in honour of Sir Ronald A. Fisher. Fisher initially developed the statistic as the variance ratio in the 1920s.

The F-test is designed to test if two population variances are equal. It does this by comparing the ratio of two variances. So, if the variances are equal, the ratio of the variances will be 1.

There are several different F-tables. Each one has a different level of significance. So, find the correct level of significance first, and then look up the numerator degrees of freedom and the denominator degrees of freedom to find the critical value.

The numerator degrees of freedom will be the degrees of freedom for whichever sample has the larger variance (since it is in the numerator) and the denominator degrees of freedom will be the degrees of freedom for whichever sample has the smaller variance (since it is in the denominator).

The F-test statistic is found by dividing the between group variance by the within group variance. The degrees of freedom for the numerator are the **degrees of freedom for the between group (k-1)** and the degrees of freedom for the denominator are the **degrees of freedom for the within group (N-k).**

Characteristics of the *F* Distribution

- The values of *F* cannot be negative.

- The distribution is positively skewed.

- The mean value of *F* is approximately equal to 1.

- The *F* distribution is a family of curves based on the degrees of freedom of the variance of the numerator and denominator.
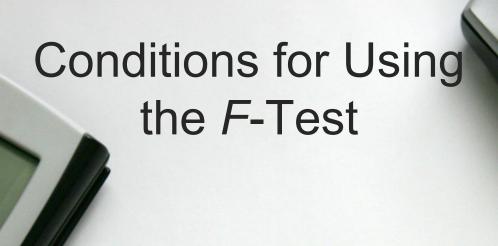
How to use F-test ?

1) Invoke the null hypothesis that states that the two variances we are comparing are from the same population. (*i.e.*, they are not statistically different)

2) Calculate the *F* value (the ratio of the two variances)

3) Look up the table value of *F* for the degrees of freedom used to calculate both variances and for a given confidence level.

4) If the calculated *F* is greater than the table value, then the null hypothesis is not correct. Else, the two could have come from the same population of measurements.

Assumptions for the F-test

- Samples are **independent random samples**
- Distribution of response variable is a **normal** curve within each population
- Different populations **may have different means.**
- All populations have **same standard dev.** , *s.*

The populations under H_0	The populations under H_a

$\mu_1 = \mu_2 = \mu_3$

μ_1 μ_2 μ_3

X

Conditions for Using the *F*-Test

- *F*-statistic can be used if data are **not extremely skewed**, there are **no extreme outliers**, and group standard deviations are not markedly different.

- Tests based on *F*-statistic are valid for data with skewness or outliers **if sample sizes are large**.

- A rough criterion for standard deviations is that the largest of the sample standard deviations should not be more than twice as large as the smallest of the sample standard deviations.