

## **F-Distribution** <sup>[1]</sup>

[Explorable.com](#) <sup>[2]</sup>lest 36.9K ganger

The F-distribution, also known as the Snedecor's F-distribution or the Fisher-Snedecor distribution (after R.A. Fisher and George W. Snedecor), is the distribution of ratios of two independent estimators of the population variances.

Suppose we have two samples <sup>[3]</sup> with  $n_1$  and  $n_2$  observations, the ratio  $F = s_1^2 / s_2^2$  where  $s_1^2$  and  $s_2^2$  are the sample variances, is distributed according to an F-distribution with  $v_1 = n_1 - 1$  numerator degrees of freedom, and  $v_2 = n_2 - 1$  denominator degrees of freedom.

For example, if  $F$  follows an F-distribution and the degrees of freedom for the numerator are 4 and the degrees of freedom for the denominator are 10, then  $F \sim F_{4,10}$ . For each combination of these degrees of freedom there is a different F-distribution. The F-distribution <sup>[4]</sup> is most spread out when the degrees of freedom are small. As the degrees of freedom increase, the F-distribution is less dispersed.

## **Properties**

The F-distribution has the following properties:

The mean of the distribution is equal to  $v_1 / (v_2 - 2)$ . The variance is equal to  $[v_2^2 * (v_1 + 2)] / [v_1 * (v_2 - 2) * (v_2 - 4)]$

The F-distribution is skewed to the right, and the F-values can be only positive. The curve reaches a peak not far to the right of 0, and then gradually approaches the horizontal axis. The F-distribution approaches, but never quite touches the horizontal axis.

## **Uses**

The main use of F-distribution is to test whether two independent samples <sup>[3]</sup> have been drawn for the normal populations <sup>[5]</sup> with the same variance, or if two independent estimates of the population variance are homogeneous or not, since it is often desirable to compare two variances <sup>[6]</sup> rather than two averages <sup>[7]</sup>. For instance, college administrators would prefer two college professors grading exams to have the same variation in their grading. For this, the F-test <sup>[8]</sup> can be used, and after examining the p-value <sup>[9]</sup>, inference can be drawn on the variation.

# Assumptions

In order to perform F-test of two variances, it is important that the following are true:

- The populations [10] from which the two samples are drawn are normally distributed [5].
- The two populations are independent of each other.

If the two populations have equal variances, then  $s_1^2$  and  $s_2^2$  are close in value and  $F$  is close to 1. But if the two population variances are very different,  $s_1^2$  and  $s_2^2$  tend to be very different, too.

Choosing  $s_1^2$  as the larger sample variance causes the ratio to be greater than 1. If  $s_1^2$  and  $s_2^2$  are far apart, then  $F$  is a large number. Therefore, if  $F$  is close to 1, the evidence favours the null hypothesis [11] (the two population variances are equal). But if  $F$  is much larger than 1, then the evidence [12] is against the null hypothesis, and we can infer that possibly the population variances differ to a large extent.

## Anova and F

In the technique known as Analysis of Variance [13] (ANOVA) which plays a very important role in Design of Experiments [14], the variance ratio test is applied to test the significance [15] of different components of variation against error variation.

For example, a new drug for treating Osteoporosis could need to be field tested. Since severity of this disease is generally a function of age, the new drug could be administered randomly to  $n$  patients in each age group. Put differently, this would be an experiment in  $m$  age groups and  $n$  different dosage levels of the drug allocated randomly to the patients. With figures provided from patients for each age group  $\times$  dose combination, we can use the variance ratio test (F-test [8]) to test for difference between dose levels and if this variation can be attributed to chance.

The other uses include testing the significance of the correlation [16] ratio between two random variables, and to test the linearity of regression [17].

---

**Kilde URL:** <https://staging.explorables.com/f-distribution>

### Lenker

- [1] <https://staging.explorables.com/f-distribution>
- [2] <https://staging.explorables.com/en>
- [3] <https://staging.explorables.com/sample-group>
- [4] <http://en.wikipedia.org/wiki/F-distribution>
- [5] <https://staging.explorables.com/normal-probability-distribution>
- [6] <https://staging.explorables.com/statistical-variance>
- [7] <https://staging.explorables.com/arithmetical-mean>
- [8] <https://staging.explorables.com/f-test>
- [9] <https://staging.explorables.com/p-value>
- [10] <https://staging.explorables.com/population-sampling>
- [11] <https://staging.explorables.com/null-hypothesis>
- [12] <https://staging.explorables.com/empirical-evidence>
- [13] <https://staging.explorables.com/anova>

- [14] <https://staging.explorable.com/design-of-experiment>
- [15] <https://staging.explorable.com/significance-test>
- [16] <https://staging.explorable.com/statistical-correlation>
- [17] <https://staging.explorable.com/linear-regression-analysis>