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 fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. Exact "F-tests" mainly arise when the models have been fitted to the data using least squares. 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.

Common examples of F-tests

Common examples of the use of F-tests include the study of the following cases:

1. The hypothesis that the means of a given set of normally distributed populations, all having the same standard deviation, are equal. This is perhaps the best-known F-test, and plays an important role in the analysis of variance (ANOVA).

2. The hypothesis that a proposed regression model fits the data well. See Lack-of-fit sum of squares.

3. The hypothesis that a data set in a regression analysis follows the simpler of two proposed linear models that are nested within each other.

In addition, some statistical procedures, such as Scheffé's method for multiple comparisons adjustment in linear models, also use F-tests.

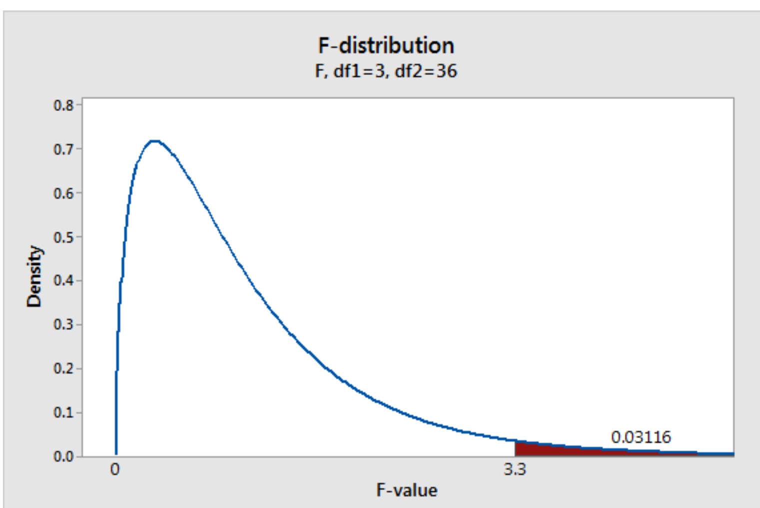
F-distribution

In probability theory and statistics, the F-distribution, also known as Snedecor's F distribution or the Fisher–Snedecor distribution (after Ronald Fisher and George W. Snedecor) is a continuous probability distribution that arises frequently as the null distribution of a test statistic, most notably in the analysis of variance (ANOVA), e.g., F-test

For one-way ANOVA, the ratio of the between-group variability to the within-group variability follows an F-distribution when the null hypothesis is true.

When you perform a one-way ANOVA for a single study, you obtain a single F-value. However, if we drew multiple random samples of the same size from the same population and performed the same one-way ANOVA, we would obtain many F-values and we could plot a distribution of all of them. This type of distribution is known as a sampling distribution. Because the F-distribution assumes that the null hypothesis is true, we can place the F-value from our study in the F-distribution to determine how consistent our results are with the null hypothesis and to calculate probabilities.

The probability that we want to calculate is the probability of observing an F-statistic that is at least as high as the value that our study obtained. That probability allows us to determine how common or rare our F-value is under the assumption that the null hypothesis is true. If the probability is low enough, we can conclude that our data is inconsistent with the null hypothesis. The evidence in the sample data is strong enough to reject the null hypothesis for the entire population.This probability that we're calculating is also known as the p-value!

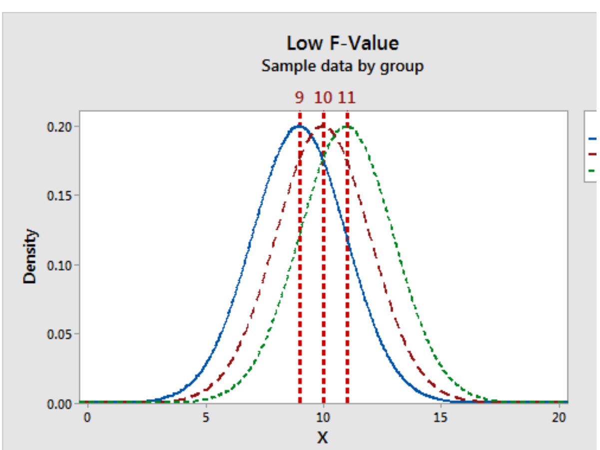


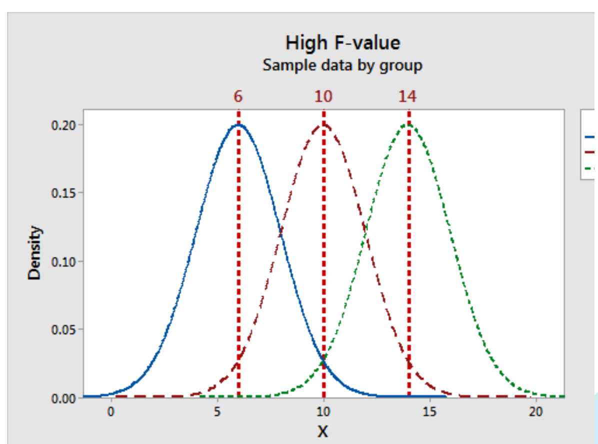
To plot the F-distribution for our plastic strength example, I'll use Minitab's probability distribution plots. In order to graph the F-distribution that is appropriate for our specific design and sample size, we'll need to specify the correct number of DF.

Looking at our one-way ANOVA output, we can see that we have 3 DF for the numerator and 36 DF for the denominator.

Multiple-comparison ANOVA problems

The F-test in one-way analysis of variance is used to assess whether the expected values of a quantitative variable within several pre-defined groups differ from each other. For example, suppose that a medical trial compares four treatments. The ANOVA F-test can be used to assess whether any of the treatments is on average superior, or inferior, to the others versus the null hypothesis that all four treatments yield the same mean response. This is an example of an "omnibus" test, meaning that a single test is performed to detect any of several possible differences. Alternatively, we could carry out pairwise tests among the treatments (for instance, in the medical trial example with four treatments we could carry out six tests among pairs of treatments). The advantage of the ANOVA F-test is that we do not need to pre-specify which treatments are to be compared, and we do not need to adjust for making multiple comparisons. The disadvantage of the ANOVA F-test is that if we reject the null hypothesis, we do not know which treatments can be said to be significantly different from the others, nor, if the F-test is performed at level α, can we state that the treatment pair with the greatest mean difference is significantly different at level α. The F-statistic is the test statistic for F-tests. In general, an F-statistic is a ratio of two quantities that are expected to be roughly equal under the null hypothesis, which produces an F-statistic of approximately 1.





The F-statistic incorporates both measures of variability discussed above. Let's take a look at how these measures can work together to produce low and high F-values. Look at the graphs below and compare the width of the spread of the group means to the width of the spread within each group.

The low F-value graph shows a case where the group means are close together (low variability) relative to the variability within each group. The high F-value graph shows a case where the variability of group means is large relative to the within group variability. In order to reject the null hypothesis that the group means are equal, we need a high F-value.

Assessing Means by Analyzing Variation

ANOVA uses the F-test to determine whether the variability between group means is larger than the variability of the observations within the groups. If that ratio is sufficiently large, you can conclude that not all the means are equal.

This brings us back to why we analyze variation to make judgments about means. Think about the question: "Are the group means different?" You are implicitly asking about the variability of the means. After all, if the group means *don't* vary, or don't vary by more than random chance allows, then you can't say the means are different. And that's why you use analysis of variance to test the means.