

---

This tutorial talks about the Analysis of Variance, also known as ANOVA. Now, the name ANOVA-- you can see where it comes from, the AN from Analysis, the O in Of, and the VA from the Variance. Now, an Analysis of Variance, or ANOVA, is a hypothesis test that's used to compare means from three or more samples. It does this by testing the quality of three or more population means by analyzing the sample variances.

Generally, you calculate a test F-statistic to find a p-value. In order to do this, we almost always need technology. If it's a really simple example, we can probably do it by hand, but ANOVA is usually supported by technology. And then the conditions for doing an ANOVA test is that the observations are taken independently and from a random sample, and then the population standard deviations are about the same, and the populations are normally distributed. So let's look at an example.

Here is an example of something we could do with an ANOVA test. We have the pulse rate, the beats per minute on the y-axis, and the participation in a sport on the x-axis. So we can compare whether or not the variation within each of these categories-- how that compares to the variation across the categories, and start to do those comparisons to draw some conclusions. For people who participate once a fortnight, there's a pretty big range here. There's a lot of variance.

And we can say whether or not this variance-- how that compares to the variance between once a fortnight and more than weekly. Depending on how those compare is to say whether or not the means from these are significantly different. If we had a different chart and these values were all very small, there was very little variance, something like that, then we could conclude more easily that the differences between the different treatment levels is what causes the differences between the beats per minute.

So when we're calculating an F-statistic, we're looking at the variance between the samples divided by the variance within the samples. If we end up with a large F-statistic, then the sample means differ more than the data within the samples. And that would be unlikely to occur if the null hypothesis were true. If we have a small F, that means that the sample means differ less than the data within the samples, which would be consistent with the null hypothesis. This has been your tutorial introducing the concept of an ANOVA test.