Print

Assumptions for the Two-Sample t Test

Before using a two-sample t test, you need to verify three statistical assumptions. Otherwise, the test might be invalid.

The first assumption is that the observations are independent, meaning that when you sampled the data, you collected each unit of information independently from one another.

For example, you can't have repeated measurements on the same observations weighing the results toward a specific result. As you can probably tell, you verify this assumption during the design stage of your analysis.

The second assumption is that you have normally distributed populations. If the populations from which you obtained your samples are normally distributed, then your sample data will most likely be normal too. For large samples, the two-sample t test is fairly robust to deviations from the assumption of normality. However, for small samples, it's important to verify this assumption by visually examining plots of the data. You should confirm that the histogram appears normal and QQ-plots follow a straight line, and we do this in the upcoming demonstration.

Finally, for the last assumption, you need to verify that you have equal population variances. To do this, you can use the folded F Test to test for equality of variances.

The null hypothesis is that the population variances are equal. The formula is sigma 1 squared equals sigma 2 squared. Remember that sigma squared represents variance.

$$H_0 : \sigma_1^2 = \sigma_2^2$$

The alternative is that the variances are not equal.

Your test statistic, or F value, is simply the ratio of the maximum sample variance of the two groups to the minimum sample variance of the two groups.

$$F = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}$$

By construction, the Fstatistic is always greater than or equal to 1.

If the null hypothesis is true and the variances in the two populations are equal, then the F value is close to 1 and the p-value for F is statistically nonsignificant, a p-value most likely greater than 0.05.

Consequently, a large value for the F statistic is evidence against the assumption of equality.

Close