

## Paired Observations

A two-sample t test is used to compare population means based on samples drawn independently from two populations. For example, in the diameter scenario, random samples were collected, independently, from two different machines. But suppose your data form a pair of measurements for an item?

For example, your data might consist of before-and-after measurements on the same subject in a diet study, or measurements on the same parts with two different gauges in a measurement system study. These data form pairs of observations.

One of the most famous statistical examples is from the work of Sir Francis Galton, whose study of the heights of parents and children led to the development of regression analysis. When Galton graphed the heights of offspring against the average heights of parents, pairing adult children with their biological parents, he noticed a pattern that he called "regression to the mean." Galton observed that there is a tendency for parents with extreme heights to have children with heights that are less extreme.

The data table Galton.jmp contains 952 pairs of parent-child heights (in inches).

This file is in the JMP Sample Data Library. Notice how the observations are arranged. In earlier examples (for example, the Diameter Test data), we had one column containing all of the measurements, and a second with a label identifying the sample. Here, the two samples are side by side. Each row is a family pair of parent and child. In this situation, instead of performing a two-sample t test to compare the means, you'd use a paired t test.

Here's why. In a two-sample t test, you compare the group means, where each mean is simply the average of all of the values in the group. This ignores the fact that the data might be paired. However, as we see in this hypothetical scenario, paired data are often correlated. And, values within a pair might be more similar than the different pairs are to one another.

The paired t test accounts for this by using the fact that the observations are matched. You're still studying the difference between the means for the two groups, but the differences are calculated based on the difference for each individual pair. As a result, the paired t test is usually more sensitive in detecting true differences than the two-sample t test.

To conduct a paired t test, you create a new variable that is the difference between the two columns of data.

Here, you see the differences for the first ten rows of the Galton data. Then you conduct a one-sample t test on this new variable, with the null hypothesis that the mean difference is zero. The mean difference for the Galton data is -0.064 inches, which is only 0.8 standard errors below the hypothesized mean. Regardless of the alternative hypothesis, the observed difference is not statistically significant at any reasonable alpha level. How do we interpret this?

Based on these data, we don't have evidence that the heights of parents and their children are different, on average. For most software packages, JMP included, you don't have to create the column of differences yourself. There are automated routines that do this for you. This analysis in JMP is called matched pairs. As you will learn in a JMP demo, additional information is provided to help visualize and interpret the paired differences. Now, what if we mistakenly analyze paired data using a pooled two-sample t test instead? Let's consider a more applied example. You're interested in comparing a measurement, gas leak, for two different gauges.

The gas leak, in  $10^{-5}$  cc/min (ten to the minus fifth cubic centimeters per minute), was measured on 10 parts using two different gauges. You run a paired t test and learn that there is a significant difference between the gauges. On average, the measurements for Gauge 2 are 0.0015 units higher than the measurements for Gauge 1. However, what if the data are in this stacked form instead? When you ignore the pairing, all of the variability between the pairs is included in the estimate of the standard deviation used to calculate the test statistic.

If you conduct a two-sample t test, you can see that the standard error is larger, the t ratio is smaller, and the p-value is much larger than the corresponding values for the paired t test.

In this example, if you ignore the matched pairing, you would not detect the true difference between the gauges. For this reason, when the data are matched pairs, the two-sample t test will generally be less sensitive in detecting a true mean difference than the paired t test. As an analyst, it's always important to know how your data were collected, because this will determine the appropriate analysis.

---

*Statistical Thinking for Industrial Problem Solving*

Copyright © 2020 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close