

10.1: Introduction to Dependent Samples

Another Kind of t-test

Remember way back when we learned about [one-sample t-tests](#)? We compared the mean of a sample to a population mean.

And then [last chapter](#), we learned about mean differences between two different groups. Independent t-tests compare TWO unrelated groups from ONE time point.

Now, we will find the average value of difference scores, that is: difference scores came from ONE group at TWO time points (or two perspectives).

It is important to understand the distinctions between them because they assess very different questions and require different approaches to the data. When in doubt, think about how the data were collected and where they came from. If they came from two time points with the same people (sometimes referred to as “longitudinal” data), you know you are working with repeated measures data (the measurement literally was repeated) and will use a dependent samples t-test because the repeated measures become pairs of data that are related. If it came from a single time point that used separate groups, you need to look at the nature of those groups and if they are related. Can individuals in one group being meaningfully matched up with one and only one individual from the other group? For example, are they a romantic couple? Twins? If so, we call those data matched and also use a dependent samples t-test.

However, if there’s no logical or meaningful way to link individuals across groups, or if there is no overlap between the groups, then we say the groups are independent and use the independent samples t-test, the subject of last chapter. And if we only have one group and population mean, then we use the one-sample t-test.

Dependent t-test

Researchers are often interested in change over time. Sometimes we want to see if change occurs naturally, and other times we are hoping for change in response to some manipulation. In each of these cases, we measure a single variable at different times, and what we are looking for is whether or not we get the same score at time 2 as we did at time 1. The absolute value of our measurements does not matter – all that matters is the change. Let’s look at an example:

Table 10.1.1: Raw and difference scores before and after training.

Employee	Before	After	Difference
A	8	9	1
B	7	7	0
C	4	1	-3
D	6	8	2
E	2	8	6

Table 10.1.1 shows scores on a quiz that five employees received before they took a training course and after they took the course. The difference between these scores (i.e. the score after minus the score before) represents improvement in the employees’ ability. This third column (Difference) is what we look at when assessing whether or not our training was effective. We want to see positive scores, which indicate improvement (the employees’ performance went up). What we are not interested in is how good they were before they took the training or after the training. Notice that the lowest scoring employee before the training (Employee E with a score of 2) improved so much that they ended up (After) as skilled as the highest scoring employee from Before (Employee A (with a score of 8), and ended up only 1 point lower at the end. There’s also one Difference score of 0 (Employee B), meaning that the training did not help this employee, and one negative Difference score, meaning that Employee C actually performed worse after the training! An important factor in this is that the participants received the same assessment at both time points. To calculate improvement or any other difference score, we must measure only a single variable.

When looking at change scores like the ones in Table 10.1.1, we calculate our difference scores by taking the time 2 score and subtracting the time 1 score. That is:

$$D = X_{T2} - X_{T1}$$

Where D is the difference score, X_{T1} is the score on the variable at Time 1 (Before), and X_{T2} is the score on the variable at Time 2 (After). Notice that we start with Time 2 so that positive scores show improvement, and negative scores show that skills decreased. Often, Time 1 is called a Pretest and Time 2 is called a Post-Test, making this a Pretest/Post-Test design. The difference score, D , will be the data we use to test for improvement.

We can also test to see if people who are matched or paired in some way agree on a specific topic. For example, we can see if a parent and a child agree on the quality of home life, or we can see if two romantic partners agree on how serious and committed their relationship is. In these situations, we also subtract one score from the other to get a difference score. This time, however, it doesn't matter which score we subtract from the other because what we are concerned with is the agreement.

In both of these types of data, what we have are multiple scores on a single variable. That is, a single observation or data point is comprised of two measurements that are put together into one difference score. This is what makes the analysis of change unique – our ability to link these measurements in a meaningful way. This type of analysis would not work if we had two separate samples of people who weren't related at the individual level, such as samples of people from different states that we gathered independently. Such datasets and analyses should be analyzed with an independent t-test, not the dependent t-test that we are discussing in this chapter.

A rose by any other name...

It is important to point out that this form of t-test has been called many different things by many different people over the years: “matched pairs”, “paired samples”, “repeated measures”, “dependent measures”, “dependent samples”, and many others. What all of these names have in common is that they describe the analysis of two scores that are related in a systematic way within people or within pairs, which is what each of the datasets usable in this analysis have in common. As such, all of these names are equally appropriate, and the choice of which one to use comes down to preference.

Now that we have an understanding of what difference scores are and know how to calculate them, we can use them to test hypotheses. As we will see, this works exactly the same way as testing hypotheses about one sample mean or two independent samples, but now with a different formula that focuses on the difference between each pair.

Interpreting Dependent t-tests

Null hypothesis significance testing and p-values are the same with dependent t-tests as the other types of t-tests:

Table 10.1.2- Small p-values Versus Large p-values

Small p-values ($p < .05$)	Large p-values ($p > .05$)
A small p-value means a small probability that the two means are similar (suggesting that the means are different...).	A large p-value means a large probability that the two means are similar.
We conclude that: <ul style="list-style-type: none"> • The means are different. • The samples are <i>not</i> from the same population 	We conclude that <ul style="list-style-type: none"> • The means are similar. • The samples <i>are</i> from the same population.
The <u>calculated t-score</u> is further from zero (more extreme) than the critical t-score . (Draw the standard normal curve and mark the <u>calculated t-score</u> and the critical t-score to help visualize this.)	The <u>calculated t-score</u> is closer to zero (less extreme) than the critical t-score . (Draw the standard normal curve and mark the <u>calculated t-score</u> and the critical t-score to help visualize this.)
Reject the null hypothesis (which says that the means are similar).	Retain (or fail to reject) the null hypothesis (which says that the means are similar).
Support the Research Hypothesis? MAYBE. Look at the actual means: <ul style="list-style-type: none"> • <i>Support the Research Hypothesis if the mean that was hypothesized to be bigger really is bigger.</i> • <i>Do not support the Research Hypothesis if the mean that was hypothesized to be bigger is actually smaller.</i> 	Do not support the Research Hypothesis (which said that one mean would be bigger, but the means are similar).

Small p-values ($p < .05$)	Large p-values ($p > .05$)
Write “The mean from Sample 1 (= ##) differs from the mean of Sample 2 (, suggesting that the samples are from different populations ($t(df) = \underline{\hspace{1cm}}$, $p < .05$). This supports (OR DOES NOT SUPPORT) the Research Hypothesis.”	Write “The mean from Sample 1 (= ##) is similar to the mean from Sample 2 (suggesting that the samples are from the same population ($t(df) = \underline{\hspace{1cm}}$, $p > .05$). This does not support the Research Hypothesis.”

Let's figure out what we do with dependent pairs of scores.

Contributors and Attributions

- [Foster et al.](#) (University of Missouri-St. Louis, Rice University, & University of Houston, Downtown Campus)

•

[Dr. MO](#) (Taft College)

This page titled [10.1: Introduction to Dependent Samples](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Michelle Oja](#).