

9.1: Change and Differences

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 9.1.1: Raw and difference scores before and after training.

Before	After	Improvement
6	9	3
7	7	0
4	10	6
1	3	2
8	10	2

Table 9.1.1 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 is what we look at when assessing whether or not our training was effective. We want to see positive scores, which indicate that 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 (with a score of 1) improved just as much as the highest scoring employee before the training (with a score of 8), regardless of how far apart they were to begin with. There’s also one improvement score of 0, meaning that the training did not help this employee. 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 9.1.1, we calculate our difference scores by taking the time 2 score and subtracting the time 1 score. That is:

$$X_d = X_{T2} - X_{T1} \quad (9.1.1)$$

Where X_d is the difference score, X_{T1} is the score on the variable at time 1, and X_{T2} is the score on the variable at time 2. The difference score, X_d , will be the data we use to test for improvement or change. We subtract time 2 minus time 1 for ease of interpretation; if scores get better, then the difference score will be positive. Similarly, if we’re measuring something like reaction time or depression symptoms that we are trying to reduce, then better outcomes (lower scores) will yield negative difference scores.

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 that weren’t related at the individual level, such as samples of people from different states that we gathered independently. Such datasets and analyses are the subject of the following 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. In this text, we will refer to paired samples, though the

appearance of any of the other names throughout this chapter should not be taken to refer to a different analysis: they are all the same thing.

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 with a t -statistic. The only difference is in the format of the null and alternative hypotheses.

This page titled [9.1: Change and Differences](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Foster et al.](#) ([University of Missouri's Affordable and Open Access Educational Resources Initiative](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.