

Wilderness Survival Guide to the T-Test

Laura Mills, Ph.D.

Pine River Institute

Abstract

Many programs that offer intensive experiential therapy for youths and young adults are part of a broad collaborative effort that explores the impact of treatment. This type of exploration has been made possible by the guidance of NATSAP leaders, but also by program-level dedication to collecting data on the health of our clients. These efforts have resulted in benefits that span evaluation-informed clinical dialogue, treatment planning, and agency decisions. Now is a great time to think about how else the data can expand knowledge in our field and how agencies can work with their own data to do so. This paper was designed as a fun, adventure-oriented explanation of one particular type of analysis, the *t*-test, which looks at differences between two groups (e.g., differences on depression scores between a treatment group and a non-treatment group). This paper also explains how this analysis can be easily accomplished in excel.

Keywords: youth treatment evaluation, outdoor behavioral health, research methods

Twenty-something years ago, I helped guide a group of twelfth grade students on a five day hike while my colleague helped lead a five-day canoe trip, both through Killarney Provincial Park in Ontario. The trip was a wonderful opportunity to gain leadership skills and enjoy Killarney's beautiful coniferous-deciduous blend that is striking against the backdrop of its many lakes and pink-white Canadian Shield. The students spent months planning the necessary details of their last adventure together before heading off to their myriad futures.

We had incredible warm, dry, sunny days coupled with chilly nights that begged for campfire. The students, who started out with diverse abilities for tripping, became increasingly cohesive and competent in the skills needed to reach our daily goals. Conversations became more thoughtful and reflective, and it seemed there was a general increase in self-confidence. My colleague agreed that her canoe trippers met each day with renewed excitement and she also sensed that they shifted toward higher self-confidence.

We wondered whether self-confidence was different between our two groups, but we didn't then have knowledge to resolve this curiosity. I would have appreciated being able to examine whether our curiosity played out in some measurable way, if self-confidence was *different* between the canoe and hike groups and whether confidence really did *increase* over the course of the trip. Of course, there is a way to answer these questions and I am excited to share this knowledge. In this paper, I will explain a simple approach to understanding differences between two groups and will provide simple steps for using excel to measure these differences. I first will explain what I mean by '*differences between groups*'. This actually refers theoretically to the question 'do two groups of scores come from the same population or are the averages of these groups different enough that they are probably derived from different populations?' I know it's not something we would say in conversation, but let me explain...

Imagine all the clients across North America who come to intensive experiential outdoor programs (shown on the right). This group is called a population of treatment-seeking clients. Suppose they all completed a tool that measured orientation skills. Some people would score extremely high and some extremely low, but the majority would score somewhere within these extremes and most would hover around the average score. As I'm sure you've guessed, I'm talking about a typical distribution of scores, or a bell-shaped curve. There are lots of scores in the middle of the bell-shape, and fewer and fewer scores as you get toward the tails.



Now, imagine how different everyone's score is from the overall average. If you estimated the average of *those* differences, you would have a rough estimate of the *standard deviation*, which is a measure of

WILDERNESS SURVIVAL GUIDE

how much variability exists in a group of scores. The average and variability of a group of scores give us an idea about the population. We know approximately what most people in the group will score, give or take a standard deviation.



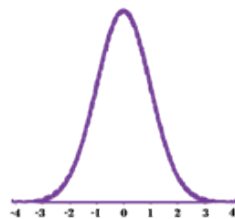
Think back to all the clients you have ever had in your program. These clients represent a sample of the population mentioned above. If they all completed a measure of orientation skills, you would have a sample with slightly different average and variability than the population, but it would likely be close enough to be representative of the population.

Now, say you have a colleague who has an intensive experiential indoor program. You think the colleague's clients might be a sample of people from a whole different population in terms of orientation skills. In other words, you think they are not from one population of clients who present for any type of intensive experiential treatment.



The t -test is a way to help us determine whether two samples are from different populations or if they are so similar that we can only assume that both samples are from the same population. But we must first consider how different everyone was on orientation skills, how variable they are. The t -test actually examines whether the two averages are meaningfully different, given how different scores are from person-to-person. Indeed, the t -test formula gives a ratio of the differences between group averages to the typical difference between individuals and the ratio is converted to a single number, the t -statistic.

This t -statistic, with an average of 0 and a standard deviation of 1 (shown on the right), is compared to a distribution of scores similar to the bell-shaped curve discussed above. The amount of area above any t -distribution score tells us how likely it is, with the highest likelihood above the average, 0. The closer t -statistic is to 0, the less difference there is between the two average scores (given their variability) and thus there is a higher likelihood our two groups are from the same population. The more standard deviations away from 0, the lower the probability on the t -distribution. This means that the greater the difference between the two group averages (taking into account their variability), the less likely they are from the same population.



The t -statistic needs to be different enough from the 0 average of the t -distribution for us to say that the groups are meaningfully different. In many cases, more than 2 standard deviations from 0 difference is big enough to say that it's not likely the two samples are from the same population. However, the actual number that tells us with some certainty that our samples are likely from different populations changes with the number of people in the study. Therefore, we determine instead is the *probability* of our t -statistic.

We know the t -statistic is big enough to say that the two groups are likely not from the same population when it's probability, or p -value, is 5% or less. We can see from the distribution that the bulk of the scores hover between -1 and 1, so any number bigger than 1 or less than -1 will have a lower probability than those closer to 0, and likelihood decreases as numbers get further and further from 0. Lucky for us, excel has made it super simple to get this p -value. Once we know the p -value we can say, with some certainty, whether the groups are from the same, or different, populations.

I'll explain the steps to getting the p -value using my question about self-confidence between hikers and canoers as an example, and we can approach this just like it's a trip we need to plan and execute. Hopefully, you will conclude that doing statistical analyses can be as exciting and rewarding as any great adventure!

- 1. Planning.** This is where the excitement begins and you start to envision the big picture of your adventure. Here, you need to determine the question you want to ask and administer a tool that aligns with that question. My question is whether *confidence* is different between canoers and hikers. We need a numeric score for each person on confidence. Of course, we could look at any scores of physical, mental, behavioral, or relationship health that, ideally, is measured with a valid and reliable standardized tool.

WILDERNESS SURVIVAL GUIDE

- Packing.** Well, not everything is fun. We need to put our scores into excel in a specific way (shown below)

PERSON	GROUP	POST-TRIP CONFIDENCE
Andy	Hike	23
Bob	Hike	27
Candice	Hike	21
David	Hike	19
Eunice	Hike	28
Frank	Hike	22
George	Canoe	18
Harris	Canoe	17
Ian	Canoe	12
Jake	Canoe	15
Kiernan	Canoe	16
Lu	Canoe	13

Each person has all their information on one row. For example, Bob was in the hike group and his confidence score was 27 after the trip.

The top row is for titles

Each cell is filled with a VARIABLE, (something that is different for different people; it varies). People vary across scores, group association, name, etc.

Just by looking at the data, it appears the hikers have higher self-confidence than canoers. But we can also see that almost all people have a different score, there is a lot of variability in the confidence of these trippers. We want to know if self-confidence for the whole group of hikers is meaningfully different than that for the group of canoers, given the person-to-person variability in scores.

- Backpack Weight Test.** Before we set out, we do need to check that things won't go awry because our packs are too heavy. One preliminary test we will do is determine whether the variability for the scores of each group are equal. In any empty cell, type `=var(` and select all the scores for one group, click enter. In another empty cell, get the variability for the other group. The two numbers can be considered fairly similar if their ratio to each other is less than 3:1. In our case, our variances are 12 and 5; we'll call them fairly equal.
- Setting Out.** We are ready for the first few steps of the trip. We can perform the *t*-test. In any empty cell, type `=ttest(`. Excel will automatically prompt us with clues as to what is next needed.
- The First Challenge.** We all know it, solving whatever problem arises, like wet boots and sore muscles, within the first hour. Like these problems, ours is an easy one. The Excel prompt `array1` is telling us to select all the scores for one group. Then we need a comma.
- Physical Obstacle.** You know this, it's like nature planned a nice easy obstacle for the group to tackle. The Excel prompt `array2` is telling us to select all the scores for the other group. Then we again need a comma.
- Decision Time.** Usually around mealtime, the team needs to make a decision, like what or where to eat. Here, we'll have an easy decision as well. The excel prompt `tails` is asking us to determine whether we want to see differences between the two groups regardless of which group is which, in which case we enter the number `2` and a comma. (A 1-tailed test is if we thought one group should have higher scores than another group).
- Finding Camp.** Everyone can breathe a sigh of relief and offload the gear. We are at the finish line. The Excel prompt `type` is asking us whether we are testing two groups or the same group over two times. We want two groups so we enter number `2` and close the bracket.
- Campfire.** My favorite part of any trip. Dinner done, everything cleaned and put away, and time to relax. We have everything we need. We can close our bracket and click `enter`.
- Reflection.** I feel like this is when the real trip begins, the time when we start looking inward and interpreting our experience. In our case, we can think about what excel has provided

	A	B	C	D	E
1	PERSON	GROUP	POST-TRIP CONFIDENCE		
2	Andy	Hike	23		
3	Bob	Hike	27		
4	Candice	Hike	21		
5	David	Hike	19		
6	Eunice	Hike	28		
7	Frank	Hike	22		
8	George	Canoe	18		
9	Harris	Canoe	17		
10	Ian	Canoe	12		
11	Jake	Canoe	15		
12	Kiernan	Canoe	16		
13	Lu	Canoe	13		
14					
15					
16					
17					
18					

12.26666667

5

12.26666667

5

12.26666667

5

	A	B	C	D	E
1	PERSON	GROUP	POST-TRIP CONFIDENCE		
2	Andy	Hike	23		
3	Bob	Hike	27		
4	Candice	Hike	21		
5	David	Hike	19		
6	Eunice	Hike	28		
7	Frank	Hike	22		
8	George	Canoe	18		
9	Harris	Canoe	17		
10	Ian	Canoe	12		
11	Jake	Canoe	15		
12	Kiernan	Canoe	16		
13	Lu	Canoe	13		
14					
15					
16					
17					
18					

=ttest(array1,array2,tails,type)

	A	B	C	D	E
1	PERSON	GROUP	POST-TRIP CONFIDENCE		
2	Andy	Hike	23		
3	Bob	Hike	27		
4	Candice	Hike	21		
5	David	Hike	19		
6	Eunice	Hike	28		
7	Frank	Hike	22		
8	George	Canoe	18		
9	Harris	Canoe	17		
10	Ian	Canoe	12		
11	Jake	Canoe	15		
12	Kiernan	Canoe	16		
13	Lu	Canoe	13		
14					
15					
16					
17					
18					

=ttest(C2:C7,C8:C13,2,2)

WILDERNESS SURVIVAL GUIDE

for us, in this case, a number. For my example, the value is **0.00076**. This is our p -value and it's less than .05, so we know there is little chance that these two group scores are from the same population in terms of self-confidence, and the average for hikers was higher than that of canoers. My interpretation of this is that the two groups are different in terms of confidence. We can't say that this difference is caused by the trip, we just know that at the end of the trip the groups were not the same on this measure.

The other question we wanted to ask is whether our trippers increased their self-confidence over the course of five days. This requires more *pre-treatment* measures of confidence. Many of us are accustomed to administering various tools at program entry, which we will insert in a new column. Once entered, performing a test to see if scores changed over time is quite simple. I'll spare you the tripping analogy, now that you are comfortable with the excel steps.

1. Enter =ttest(
2. Select the set of scores taken before the trip, then put a comma.
3. Select the set of scores taken after the trip, and a comma.
4. The Excel prompt **tails** is asking if you want to test for increases or decreases in scores or both (we choose both, **#2**). Now a comma.
5. The excel prompt type asks what kind of t-test to perform and this time we want to choose paired, so enter **#1**. Now close the bracket.

Excel feeds us the p -value for the t -test, which in this case is .03. Since it's less than .05, we can assert that our scores from pre- to post-trip were not the same. My interpretation here is that confidence increased from pre- to post-trip for all students.

	PERSON	GROUP	PRE-TRIP CONFIDENCE	POST-TRIP CONFIDENCE
1				
2	Andy	Hike	17	23
3	Bob	Hike	22	27
4	Candice	Hike	25	21
5	David	Hike	16	19
6	Eunice	Hike	25	28
7	Frank	Hike	12	22
8	George	Canoe	15	18
9	Harris	Canoe	16	17
10	Ian	Canoe	10	12
11	Jake	Canoe	11	15
12	Kiernan	Canoe	17	16
13	Lu	Canoe	14	13
14			=TTEST(C2:C13,D2:D13,2,1)	
15			TTEST(array1, array2, tails, type)	

We have walked through two types of t -tests, a between-groups test and a two-time-point test. One told us that two groups, one hikers and one canoers, were not the same in confidence at the end of their trips. The other told us that confidence increased over the course of a trip for all students. I hope you now have an idea of what it means to test differences between groups and will try it out in excel with your own program data.

Some amazing work is being done to show the benefits of our types of programs across North America. Our results give us amazing tools for treatment planning, therapeutic dialogue, and assessing outcomes. It seems a great time to build on this work by engaging with agency-level data and addressing the questions that spark the curiosity of clients, staff, and your other stakeholders. For example, you might want to know if depression scores for males are different than those of females at program entry, if trauma symptoms decrease over the course of a wilderness program, or if family functioning is different for girls than boys after a therapeutic boarding school experience. These questions, and many others, can be addressed for internal learning and dialogue, program improvement, and more broadly advancing knowledge.