

## Repeated-measures t-test

### Relevant research questions and data requirements

**Research question:** Repeated-measures t-tests (also called dependent-samples t-tests, or paired-samples t-tests) are *mean difference tests*. Specifically, we are seeking to compare if there is a difference between the mean of population 1 (e.g., untreated, or treated with treatment 1) and the mean of population 2 (e.g., treated, or treated with treatment 2). Because we cannot usually get data from everyone in our population of interest, we take a sample from our population and, in this case, expose the same sample *to both treatments*.

Let's say we want to know if participants feel more (or less) Statistics apprehension on the first vs. the last day of the semester. We give a random sample of students enrolled in Statistics a questionnaire on the first and again the last day of classes. This is a *within-subjects design* because the same participants are measured twice.

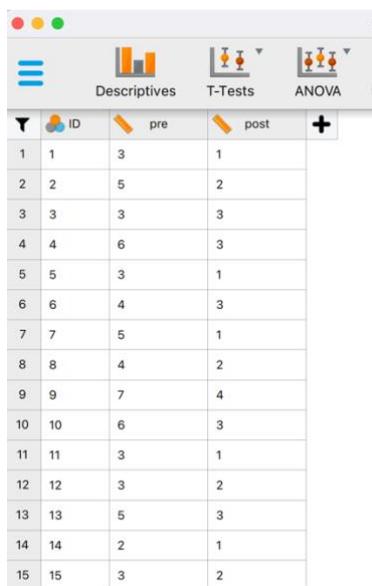
**How many groups?** One group of participants, but there are two conditions/treatments/measurement times for each participant (here: beginning and end of semester).

Fun fact: While a t-test cannot do what ANOVA can do (namely, compare 3+ groups at once), an ANOVA can do whatever a t-test can do (namely, compare two groups). So, you could also analyze these data using a Repeated-Measures ANOVA!

**Data requirements?** Interval/ratio outcome data (here: Statistics apprehension), ideally the difference scores (post-pre) are (roughly) normally distributed and independent (difference scores are not related). Check for outliers.

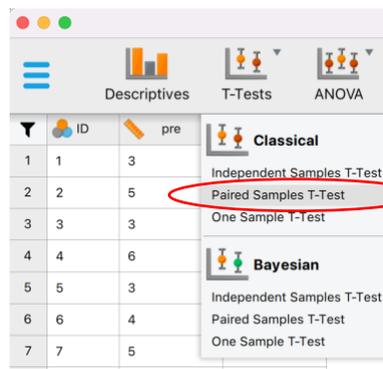
### Running the test in JASP

Note that within-subjects data are sets require you to enter two data points for each participants.

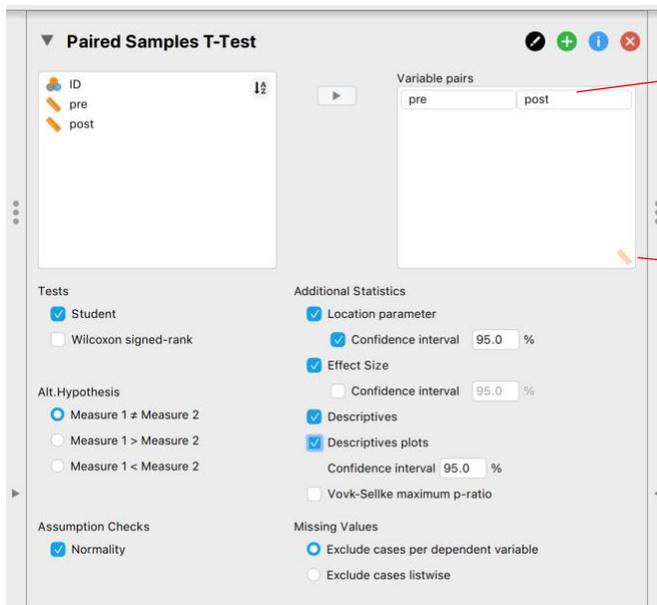


ID	pre	post
1	3	1
2	5	2
3	3	3
4	6	3
5	3	1
6	4	3
7	5	1
8	4	2
9	7	4
10	6	3
11	3	1
12	3	2
13	5	3
14	2	1
15	3	2

This is our data set - each participant gets a row and then *two outcome variable entries* (in separate columns). The pre column contains the data for Statistics apprehension on the first day, the post column contains the data that for Statistics apprehension on the last day



Once the data are entered, you go to "t-tests" and select "paired samples t-test" from the menu



Move your two outcome variables of interest into the “variable” window. Notice that JASP automatically formats this as a variable pair – i.e., two outcome variables for each participant.

And, the little ruler in the bottom right reminds you these variables must be identified as scale variables in your data set.

**Tests:** Select “Student” (Fun fact: This is not a special test for students, but named after a Statistician who published about the t-distribution by the synonym “Student”). Wilcoxon signed-rank is the nonparametric equivalent, i.e., the test to use when your data are not normally distributed (see the output below to tell whether data are normally distributed).

**Alt. hypothesis:** Here, the two-tailed option is chosen automatically as the alternative hypothesis, i.e. “Statistics apprehension is different at the beginning of the semester than at the end of the semester”. (Recall a one-tailed test would specify a direction, e.g. we might hypothesize “students experience *less* Statistics apprehension at the end of the semester than at the beginning.” Unless you have good reason to run a one-tailed test, however, you should always run a two-tailed test because it requires slightly more extreme cut-offs for the critical region, i.e. it is more stringent.)

#### Assumption Checks:

Check “Normality” to test the hypothesis that the difference scores are, in fact, (roughly) normally distributed.

#### Additional Statistics:

- Check “Location Parameter” – this is the difference between the pre and post sample means; then click “confidence interval”
- Check “effect size” - this will tell you the distance between the sample means in SD units
- Descriptives will calculate M and SD as well as standard error (SE) for your two measurements
- Descriptive Plots will present a visual overview of the two measurements with 95% CI around each mean

Reading and understanding the output

**Results** ▾

**Paired Samples T-Test** ▾

Paired Samples T-Test

Measure 1	Measure 2	t	df	p	Mean Difference	SE Difference	95% CI for Mean Difference		Cohen's d
							Lower	Upper	
pre	- post	7.246	14	< .001	2.000	0.276	1.408	2.592	1.871

Note. Student's t-test.

This is our hypothesis test. Shown in order are the t-value, the df (for a dependent samples t this is n-1), the p-value, the mean difference between sample and test value (sample mean-test value), the 95% CI around the mean difference, and the effect size Cohen's d

Here we learn that the **difference between our sample means** is **2.00** with a **standard error** of **0.276**. Statistics apprehension scores decreased by 2 points from the first to the last day of the semester. As the **p-value** suggests, this is a statistically significant effect, because **p < .001**. We therefore have enough evidence to reject our null hypothesis that  $\mu_{\text{difference}} = 0$  (or  $\mu_{\text{pre}} - \mu_{\text{post}} = 0$ ).

We also see that 0 is not contained in the **95% CI [1.41; 2.59] for the mean difference**, so 0 is not a likely value of the true difference between Statistics apprehension scores before and after a Statistics class. This means we are left to conclude that Statistics classes have an effect on Statistics apprehension, based on this one study.

The effect size **Cohen's d is 1.87** which tells us our sample means are 1.87 SDs apart. This makes it a huge effect (d=0.20 is small, d=0.50 is medium, d=0.80 is a large effect). In other words, we may have evidence of a strong effect here.

**Assumption Checks**

Test of Normality (Shapiro-Wilk)

			W	p
pre	-	post	0.934	0.316

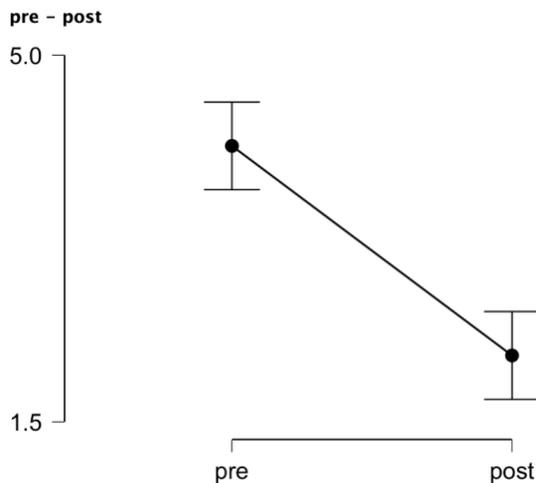
Note. Significant results suggest a deviation from normality.

In our assumption checks, we tested for normal distribution of differences scores. The test of normality is not significant (**p = .316**, which is **> .05**). This means the null hypothesis that normality can be assumed is not rejected, so there are no serious deviations from normality.

## Descriptives

Descriptives				
	N	Mean	SD	SE
pre	15	4.133	1.457	0.376
post	15	2.133	0.990	0.256

## Descriptives Plots



Lastly, our descriptive statistics and descriptive plots.

Our plots also show that students showed much less Statistics apprehension at the end than at the beginning of the semester. Also notice that the 95% CIs do not overlap, which means that the true population means likely do not overlap, i.e. there is a difference.

Also notice that the y-axis is scaled from 1.5-5, not 1-7 (or whatever the Likert scale labels were), so differences between the two groups appear even more pronounced.

## Writing up results in APA style

A dependent samples t-test indicated that students reported less Statistics apprehension at the end of the semester ( $n = 15$ ,  $M = 2.13$ ,  $SD = 0.99$ ) than at the beginning ( $M = 4.13$ ,  $SD = 1.46$ )  $t(14) = 7.25$ ,  $p < .001$ , 95% CI [1.41; 2.59], Cohen's  $d = 1.87$ .

\*Note that there are spaces before and after equal signs;  $M$ ,  $SD$ ,  $t$ ,  $p$  are italicized; everything is rounded to two digits *except* for  $p$ -values, which should be reported exactly as given in the output. Only report leading zeros for values that can exceed 1 (hence,  $p$ -values should *not* be reported with leading 0s)