Correlated-Scores *t* test

The correlated scores t test, which is also known as the *within-subjects t test*, *paired t test* (the SPSS term), the *dependent-samples t test*, or the *repeated measures t test*, is used for comparisons with a continuous dependent variable. The reason the test has all of these names is because it is used in several different situations: the same person is in the study twice (longitudinal or repeated measures design), pairs of individuals are linked together or "yoked" (e.g., twins, or married couples), or because they are naturally linked or the experimenter linked them as when they are 'matched' on some score (e.g., matched on age).

Example. We could have conducted the retrieval practice experiment in a different way—by comparing students' recall after using each type of studying strategy. This design could be classified as a repeated measures experiment or single-group pretest-posttest design (Note that there are some methodological problems with this design that can be addressed with some changes, but let us just assume this is the design for now).[1]

I have used the same numbers as in the first between-subjects example given in class to illustrate a point, but this is completely different example where we have two scores for each of 5 students. Notice that in this design we only are using half the number of cases. Each student has two scores.

Student	Reading Only	Retrieval Practice	Y_{change}	$Y_{change}-\bar{Y}_{change}$	$\left(Y_{change}-\bar{Y}_{change}\right)^2$
1	2	7	-5	-2	4
2	4	8	-4	-1	1
3	6	10	-4	-1	1
4	8	8	0	3	9
5	10	12	-2	1	1
			$\bar{Y}_{change}=-3$		$\sum\left(Y_{change}-\bar{Y}_{change}\right)^2=16$

Formulas:[2]

$$s^2_{Y_{change}}=\frac{\sum\left(Y_{change}-\bar{Y}_{change}\right)^2}{n-1}=\frac{16}{4}=4$$

$$s_{\bar{Y}_{change}}=\sqrt{\frac{s^2_{Y_{change}}}{n}}=\sqrt{\frac{4}{5}}=.89$$

$$t=\frac{\bar{Y}_{change}}{s_{\bar{Y}_{change}}}=\frac{-3}{.89}=-3.35$$

df = n-1 = 5-1 = 4. $t_{df,\alpha/2}$ at α = .05 with df = 4 is 2.776. The difference is significant, because the (absolute value of the) calculated value, 3.35, exceeds the critical value of 2.776. The 95% confidence limits are constructed in the same manner as with the independent-samples t test:

$$CI=\left(\bar{Y}_{change}\right)\pm\left(t_{df,\alpha/2}s_{\bar{Y}change}\right)=-3\pm\left[(-2.776)(.89)\right]=-3\pm(-2.47)=-5.47,-.53$$

---

[1] Visit this site for a brief review of research study designs: http://sphweb.bumc.bu.edu/otlt/mph-modules/programevaluation/ProgramEvaluation7.html

[2] Note that your text does not present the full computational formula for the repeated-measures *t* test, and, to be clearer, I use a notation that differs somewhat from the text.

Compute Cohen's d for the correlated samples case (the negative sign is usually omitted because direction is arbitrary):

$$\hat{d} = \frac{t}{\sqrt{n}} = \frac{3.35}{\sqrt{5}} = \frac{3.35}{2.24} = 1.50$$

Statistical comment. Notice that the same data in the very first between-subjects example (presented in class) yielded a non-significant difference (with twice as many cases!!). The reason the within-subjects test is more powerful is that variation due to individual differences is eliminated in the within-subjects design. Each student serves as that student's own comparison or control.

Syntax

`t-test pairs=retrieval` ~~`practice`~~ `reading` ~~`only`~~`.`

Menu Steps

1. Analyze→ Compare Means→ Paired Samples t-test
2. Move over the two variables (e.g., pretest score is Variable1, posttest score is Variable2)
3. Click "Ok."

The Output will look something like this:

SPSS Output

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	reading recall after reading only	6.0000	5	3.16228	1.41421
	retrieval recall after retrieval practice	9.0000	5	2.00000	.89443

Paired Samples Test

		Paired Differences							Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	One-Sided p	Two-Sided p
Pair 1	reading recall after reading only - retrieval recall after retrieval practice	-3.00000	2.00000	.89443	-5.48333	-.51667	-3.354	4	.014	.028

Paired Samples Effect Sizes

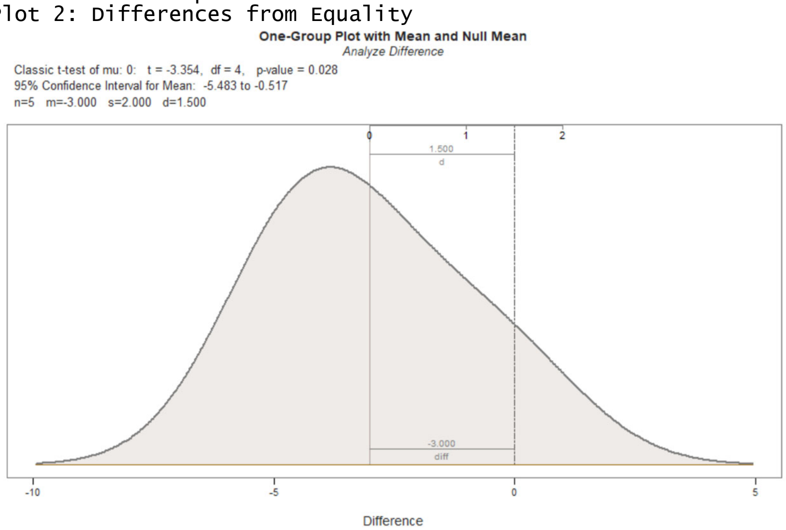
					95% Confidence Interval	
			Standardizer[a]	Point Estimate	Lower	Upper
Pair 1	reading recall after reading only - retrieval recall after retrieval practice	Cohen's d	2.00000	-1.500	-2.797	-.137
		Hedges' correction	2.50663	-1.197	-2.231	-.109

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Use the "Point Estimate" of 1.50 for Cohen's d.

R

```
> #clear active frame from previous analyses
> rm(d)

> library(haven)
> d = read_sav("c:/jason/spsswin/uvclass/ws ttest.sav")

> library(lessR)
> #obtain paired/dependent t-test/correlated samples
> ttest(retrieval, reading, paired = TRUE)


------ Describe ------

Difference: n.miss = 0, n = 5,  mean = -3.000, sd = 2.000


------ Normality Assumption ------

Null hypothesis is a normal distribution of Difference.
Shapiro-Wilk normality test: W = 0.9053, p-value = 0.44


------ Infer ------

t-cutoff for 95% range of variation: tcut = 2.776
Standard Error of Mean: SE = 0.894

Hypothesized Value H0: mu = 0
Hypothesis Test of Mean: t-value = -3.354, df = 4, p-value = 0.028

Margin of Error for 95% Confidence Level: 2.483
95% Confidence Interval for Mean: -5.483 to -0.517


------ Effect Size ------

Distance of sample mean from hypothesized: -3.000
Standardized Distance, Cohen's d: 1.500


------ Graphics Smoothing Parameter ------

Density bandwidth for 1.650

--------------------------------
Plot 1: One-Group Plot
Plot 2: Differences from Equality
```



Example write-up. Using a repeated measures *t* test, students' recall after reading only was compared with students' recall after retrieval practice. Recall was significantly higher after retrieval practice (*M* = 9) than after reading only (*M* = 6), as indicated by a significant *t* test, *t*(4) = -3.35, p = .03, *d* = 1.50, 95% CI[-5.48,-.52]. This finding indicates that there was a large increase in recall over time that was not likely to be due to chance.