UNDERSTANDING THE DEPENDENT-SAMPLES t TEST

A dependent-samples t test (a.k.a. matched or paired-samples, matched-pairs, samples, or subjects, simple repeated-measures or within-groups, or correlated groups) assesses whether the mean difference between paired/matched observations is significantly different from zero. That is, the dependent-samples t test procedure evaluates whether there is a significant difference between the means of the two variables (test occasions or events). This design is also referred to as a *correlated groups design* because the participants in the groups are not independently assigned. The participants are either the same individuals tested (assessed) on two occasions or under two conditions on one measure, or there are two groups of participants that are matched (paired) on one or more characteristics (e.g., IQ, age, gender, etc.) and tested on one measure.

HYPOTHESES FOR THE DEPENDENT-SAMPLES t TEST

Null Hypothesis: H_0: $\mu_1 = \mu_2$ *where μ_1 stands for the mean for the first variable/occasion/events and μ_2 stands for the mean for the second variable/occasion/event.*

-or- H_0: $\mu_1 - \mu_2 = 0$

If we think of the data as being the set of difference scores, the null hypothesis becomes the hypothesis that the mean of a population of difference scores (denoted μ_D or δ) equals 0. Because it can be shown that $\mu_D = \mu_1 - \mu_2$, we can write H_0: $\mu_D = \mu_1 - \mu_2 = 0$ or (H_0: $\delta = \mu_1 - \mu_2 = 0$). The hypothesized population parameter, defined by the null hypothesis will be $\delta = 0$, where δ (delta) is defined as the mean of the difference scores across the two measurements.

Alternative (Non-Directional) Hypothesis: H_a: $\mu_1 \neq \mu_2$ -or- H_a: $\mu_1 - \mu_2 \neq 0$

Alternative (Directional) Hypothesis: H_a: $\mu_1 < \mu_2$ -or- H_a: $\mu_1 > \mu_2$

(*depending on direction*)

NOTE: the subscripts (1 and 2) can be substituted with the variable/occasion/event identifiers. For example: H_0: $\mu_{pre} = \mu_{post}$ H_a: $\mu_{pre} \neq \mu_{post}$

ASSUMPTIONS UNDERLYING THE DEPENDENT-SAMPLES t TEST

1. The dependent variable (difference scores) is normally distributed in the two conditions.

2. The independent variable is dichotomous and its levels (groups or occasions) are paired, or matched, in some way (e.g., pre-post, concern for pay-concern for security, etc.).

When there is an extreme violation of the normality assumption or when the data are not of appropriate scaling, the *Wilcoxon Matched-Pairs Signed Ranks Test* should be used.

DEGREES OF FREEDOM

Because we are working with difference (paired) scores, N will be equal to the number of differences (or the number of *pairs* of observations). We will lose (restrict) one df to the mean and have $N - 1$ df. In other words, df = number of pairs minus the 1 restriction.

EFFECT SIZE STATISTICS FOR THE DEPENDENT-SAMPLES t TEST

Cohen's d (which can range in value from negative infinity to positive infinity) evaluates the degree (measured in standard deviation units) that the mean of the difference scores is equal to zero. If the calculated d equals 0, the mean of the difference scores is equal to zero. However, as d deviates from 0, the effect size becomes larger.

The d statistic may be computed using the following equation:

$$d = \frac{Mean}{SD}$$

where the pooled *Mean* and the *Std. Deviation* are reported in the SPSS output under Paired Differences

The d statistic can also be computed from the reported values for t (obtained t value) and N (the number of pairs) as follows:

$$d = \frac{t}{\sqrt{N}}$$

So what does this Cohen's d mean? Statistically, it means that the difference between the two sample means is (e.g., .52) standard deviation units (in absolute value terms) from zero, which is the hypothesized difference between the two population means. Effect sizes provide a measure of the magnitude of the difference expressed in standard deviation units in the original measurement. It is a measure of the practical importance of a significant finding.

SAMPLE APA RESULTS

Using an alpha level of .05, a dependent-samples t test was conducted to evaluate whether students' performance using two methods of mathematics instruction differed significantly. The results indicated that the students' average performance (score out of 10) using the first method of mathematics instruction ($M = 5.67$, $SD = 1.49$) was significantly higher than their average performance using the second method ($M = 4.50$, $SD = 1.83$), with $t(29) = 2.83$, $p < .05$, $d = .52$. The 95% confidence interval for the mean difference between the two methods of instruction was .32 to 2.01.

Note: there are several ways to interpret the results, the key is to indicate that there was a significant difference between the two methods at the .05 alpha level – and include, at a minimum, reference to the group means, effect size, and the statistical strand.

$t(29) = 2.83$, $p < .05$, $d = .52$

t	Indicates that we are using a t-Test
(29)	Indicates the degrees of freedom associated with this t-Test
2.83	Indicates the obtained t statistic value (t_{obt})
$p < .05$	Indicates the probability of obtaining the given t value by chance alone
$d = .52$	Indicates the effect size for the significant effect (the magnitude of the effect is measured in standard deviation units)

INTERPRETING THE DEPENDENT-SAMPLES *t* TEST

The first table, (**PAIRED SAMPLES STATISTICS**) shows descriptive statistics that can be used to compare (describe) the alcohol and no alcohol reaction time conditions. Note that the means for the two conditions look somewhat different. This might be due to chance (sampling fluctuation), so we will want to test this with the *t* test to determine if the difference is significant.

The second table, (**PAIRED SAMPLES CORRELATIONS**) provides correlations between the two paired scores. In our example, the correlation between alcohol and no alcohol is $r = .949$, which is a very high positive (imperfect) **relationship**. With a Sig. value of .000, this indicates that the relationship is significantly different from 0 (no relationship) at the .001 alpha level. Note, however, that this **DOES NOT** tell you whether there is a significant **difference** between the alcohol and no alcohol reaction time. That is what the *t* in the third table tells us.

The third table, (**PAIRED SAMPLES TEST**) shows information on paired differences and the paired samples *t* test information. The **Mean**, indicates the mean difference between the two conditions. In our example, we see an 11.57, which indicates the mean difference between alcohol reaction time (42.07) and no alcohol reaction time (30.50). That is, the reaction time for the alcohol condition is 11.57 (hundredths) seconds longer (slower) than the reaction time for the no alcohol condition. The **Std. Deviation** is the pooled standard deviation for the pairs. The **Std. Error Mean** is the pooled standard error of the mean for the pairs. This table also provides the **Lower** and **Upper** values for our confidence interval. For our example, we used an alpha level of .05; therefore, our confidence interval is 95, which results in a lower value of 10.69 and an upper value of 12.45. We also see the obtained **t** value (27.00 for our example) for the test statistic. The degrees of freedom (**df**) for this example is 27, which is $n - 1$ (where n = number of pairs). For our example we had 28 pairs – and when we subtract the one restriction – we get $df = 27$. The **Sig.** provides the actual probability level for our example, which is shown to be .000 (i.e., < .001).

> **Note:** If the Paired Mean Differences and the Obtained Test Statistic (*t*) had been negative – it simply would have meant that the second value (condition) was higher than the first value (condition).

We know that there is a statistically significant difference between the two conditions. That is, the mean difference is significantly different from zero (0). How do we know this?

> **METHOD ONE (**most commonly used**):** comparing the Sig. (probability) value to the *a priori* alpha level. If $p \leq \alpha$ – we reject the null hypothesis of no difference. If $p > \alpha$ – we retain the null hypothesis of no difference. In our example, p is shown to be .000 (i.e., < .001) and $\alpha = .05$ – therefore, $p \leq \alpha$ – indicating that we should reject the null hypothesis of no difference and conclude that the average reaction time for the alcohol condition ($M = 42.07$) was significantly longer (slower) than the average reaction time for the no alcohol condition ($M = 30.50$).

> **METHOD TWO:** comparing the obtained *t* statistic value ($t_{obt} = 27.00$ for our example) to the *t* critical value (t_{cv}). Knowing that we are using a two-tailed (non-directional) *t* test, with an alpha level of .05 ($\alpha = .05$), with $df = 27$, and looking at the Student's *t* Distribution Table – we find the critical value for this example to be 2.052. If $|t_{obt}| \geq |t_{cv}|$ – we reject the null hypothesis of no difference. If $|t_{obt}| < |t_{cv}|$ – we retain the null hypothesis of no difference. For

our example, $t_{obt} = 27.00$ and $t_{cv} = 2.052$, therefore, $t_{obt} \geq t_{cv}$ – so we reject the null hypothesis and conclude that there is a statistically significant difference between the two conditions. That is, the average reaction time for the alcohol condition ($M = 42.07$) was significantly longer (slower) than the average reaction time for the no alcohol condition ($M = 30.50$).

METHOD THREE: examining the confidence interval and determining whether zero (the hypothesized mean difference) is contained within the lower and upper boundaries. If the confidence intervals do not contain zero – we reject the null hypothesis of no difference. If the confidence intervals do contain zero – we retain the null hypothesis of no difference. In our example, the lower boundary is 10.69 and the upper boundary is 12.45, which does not contain zero – therefore, we reject the null hypothesis of no difference. That is, the reaction time for the alcohol condition ($M = 42.07$) was significantly longer (slower) than the reaction time for the no alcohol condition ($M = 30.50$).

> **Note** that if the upper and lower bounds of the confidence intervals have the same sign (+ and + or – and –), we know that the difference is statistically significant because this means that the null finding of zero difference lies *outside* of the confidence interval.

As shown above, we reject the null hypothesis in favor of the alternative hypothesis. This indicates that the participant's reaction time for the alcohol condition ($M = 42.07$) was, on average, significantly higher (slower) than their reaction time for the no alcohol condition ($M = 30.50$). Further this indicates that the mean difference (11.57) between the two conditions was significantly different from zero.

CALCULATING AN EFFECT SIZE: Since we concluded that there was a significant difference – we will need to calculate an effect size. <u>Had we not found a significant difference</u> – typically, no effect size would have to be calculated. A non-significant finding would have indicated that the participant's reaction time in the two conditions only differed due to random fluctuation or chance – or that the mean difference was not significantly different from zero.

To calculate the effect size for our example, we can use either of two formulas:

$$\text{For: } d = \frac{Mean}{SD} = \frac{11.571}{2.268} = 5.1018519 = \mathbf{5.10}$$

> *where* the pooled *Mean* and the *Std. Deviation* are reported in the SPSS output under Paired Differences

> The *d* statistic can also be computed from the reported values for *t* (obtained *t* value) and *N* (the number of pairs) as follows:

$$\text{For: } d = \frac{t}{\sqrt{N}} = \frac{27.00}{\sqrt{28}} = \frac{27.00}{5.2915026} = 5.1025204 = \mathbf{5.10}$$

Dependent-Samples T-Test Example

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	alcohol	42.07	28	7.081	1.338
	no_alcohol	30.50	28	7.172	1.355

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	alcohol & no_alcohol	28	.949	.000

Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	alcohol - no_alcohol	11.571	2.268	.429	10.692	12.451	27.000	27	.000

Drop-Down Syntax for Dependent-Samples T-Test Example

```
T-TEST
 PAIRS = alcohol WITH no_alcohol (PAIRED)
 /CRITERIA = CI(.95)
 /MISSING = ANALYSIS.
```

Alternative Syntax for Dependent-Samples T-Test Example

```
t-test / pairs = alcohol no_alcohol.
```

UNDERSTANDING THE DEPENDENT-SAMPLES *t* TEST

SAMPLE APA RESULTS

Using an alpha level of .05, a dependent-samples *t* test was conducted to evaluate whether students' reaction time differed significantly as a function of whether they had received no alcohol or received 12 oz. of alcohol prior to a timed task. The results indicated that the students' average reaction time when given no alcohol (*M* = 30.50, *SD* = 7.17) was significantly lower (i.e., faster) than their average reaction time when they received 12 oz. alcohol (*M* = 42.07, *SD* = 7.08), with *t*(27) = 27.00, *p* < .05, *d* = 5.10. The 95% confidence interval for the mean difference between the two conditions was 10.69 to 12.45.

Note: there are several ways to interpret the results, the key is to indicate that there was a significant difference between the two methods at the .05 alpha level – and include, at a minimum, reference to the group means, effect size, and the statistical strand.

$t(27) = 27.00, p < .05, d = 5.10$

t	Indicates that we are using a *t*-Test
(27)	Indicates the degrees of freedom associated with this *t*-Test
27.00	Indicates the obtained *t* statistic value (t_{obt})
p < .05	Indicates the probability of obtaining the given *t* value by chance alone
d = 5.10	Indicates the effect size for the significant effect (the magnitude of the effect is measured in standard deviation units)

REFERENCES

Cochran, W. G., & Cox, G. M. (1957). *Experimental Designs*. New York: John Wiley & Sons.

Green, S. B., & Salkind, N. J. (2003). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). New York: Houghton Mifflin Company.

Howell, D. C. (2007). *Statistical Methods in Psychology* (6th ed.). Belmont, CA: Thomson Wadsworth.

Huck, S. W. (2004). *Reading Statistics and Research* (4th ed.). New York: Pearson Education Inc.

Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2004). *SPSS for Introductory Statistics: Use and Interpretation* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Pagano, R. R. (2007). *Understanding Statistics in the Behavioral Sciences* (8th ed.). Belmont, CA: Thomson Wadsworth.

Satterthwaite, F. W. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110-114.