

Paired-samples t test
Also known as the Correlated-samples and Dependent-samples t-test

1. Purpose

The correlated t-test allows researchers to consider differences between two groups or sets of scores that are related. Under what conditions will one likely find correlated or dependent samples or groups?

(a) Condition 1: Before/After Studies; Multiple Measures on the Same Subject

These types of data occur most often with pretest-treatment-posttest experimental designs. These designs are used to determine whether some treatment will change posttest scores relative to pretest scores. The pretest and posttest scores are related, or correlated, because the scores are taken from the same individuals, i.e., each person is measured twice.

Examples

(a) A student takes the SAT, enrolls in an SAT enhancement class, and then retakes the SAT. Two scores from the same student exist.

(b) A teacher measured the reading performance of a third-grader, presented some treatment designed to increase reading performance, then re-measured the student's reading performance again (two scores from same individual).

(c) A PE teacher measures the vertical jumping ability of his class, provides his class a weight training program for one month, then re-measures vertical jumping ability of each student (two scores from same students).

(b) Condition 2: Naturally occurring pairs

Natural pairs, such as husbands and wives, twins, brothers, sisters, brothers and sisters, parents and their children, etc. With naturally occurring pairs, one would expect the pairs to hold similar feelings, beliefs, attitudes, etc., so their scores will generally be related.

Examples

(a) Determining whether husbands' attitudes toward politics are similar to their wives' attitudes. Since people tend to marry others like themselves, one would expect that most husbands and wives to hold similar political views.

(b) Determining whether boys' IQ differs from girls' IQ. Since brothers and sisters are similar genetically, one might anticipate the two to have similar IQs, that is, their IQs are likely to be related; therefore, brothers and sisters should be matched.

(c) Condition 3: Matched-pairs Design (Individually Matched Subjects)

Two sets or groups of scores are involved in the study, and sets are matched on some extraneous variable (or perhaps multiple variables) that is related to the dependent variable examined.

Example

A teacher is interested in determining whether "Hooked on Phonics" increases third-grade students' reading performance. Using two groups of students, group A (the experimental group) will use "Hook on Phonics" for one month, and group B (the control) will be exposed to the usual reading lessons during the month. The teacher knows that

IQ influences reading performance, so to control for the effects of IQ on the dependent variable (which is a posttest on reading performance), the researcher matches students in the two groups on their unique IQ levels in a fashion similar to the schematic below:

Group A (treatment)		Group B (control)	
<u>IQ score</u>			<u>IQ score</u>
133	Sue	John	133
126	Susan	Fred	126
		...	
86	Bryan	Walt	86
72	Bill	Josh	72

In this scheme, students from both groups are matched according to their IQ levels. It is important to match on IQ individually student-by-student since one would expect students with higher IQs to perform better on a reading test than students with lower IQs.

2. Steps of Hypothesis Testing

Like with the two independent-samples t-test, the paired-samples t-test follows the same steps for hypothesis testing:

- Define both H_0 and H_1
- Set alpha (α , probability of a Type 1 Error)
- Identify decision rule (either for α , test statistic, or confidence interval)
- Calculate the test statistic (t ratio)
- Find df (degrees of freedom), confidence intervals, and p-values
- Present inferential and interpretation of results (in APA style)

3. Hypotheses

The hypothesis tested with the correlated t-test is the same as in the independent-samples t-test.

(a) Example 1

Is there a difference in mean systolic blood pressure before and after taking 50mg of Losartan?

Null

Written

The mean systolic blood pressure is the same before and after taking Losartan.

Symbolic

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

Alternative Non-directional

Written

The mean systolic blood pressure before and after taking Losartan differs.

Symbolic

$H_1: \mu_1 \neq \mu_2$ or $H_1: \mu_1 - \mu_2 \neq 0$

(b) Example 2

Do identical twins, separated at birth and raised by different families, show similar IQ scores?

Null

Written

There is no difference in mean IQ scores between twins raised by different families.

Symbolic

$$H_0: \mu_1 = \mu_2 \text{ or } H_0: \mu_1 - \mu_2 = 0$$

Alternative Non-directional

Written

There is a difference in mean IQ scores between twins raised by different families.

Symbolic

$$H_1: \mu_1 \neq \mu_2 \text{ or } H_1: \mu_1 - \mu_2 \neq 0$$

4. Decision Rules

Decision rules for the paired-samples t-test are the same as for the two-independent samples t test.

(a) p-values and α

One normally sets the probability of committing a Type 1 error, α , at .10, .05, or .01.

Decision rule for p-values (this decision rule holds for all statistical tests in which p-values are present) states

$$\textit{If } p \leq \alpha \textit{ reject } H_0, \textit{ otherwise fail to reject } H_0$$

A p-value is calculated for the sampled data assuming the null hypothesis is true. A p-value for the paired-samples t-test holds the same interpretation as for the two-sample t test:

p-value: The probability of obtaining, randomly, two sample means from the population that differ by the amount observed, or more, assuming the null hypothesis is true (i.e., $H_0: \mu_1 - \mu_2 = 0$) given the size of the samples selected.

p-value simplified: The probability of randomly obtaining results this discrepant from the null hypothesis, or more discrepant, assuming the null hypothesis is true.

Assuming there is really no difference between means in the population (i.e., $\mu_1 = \mu_2$), the p-value indicates the probability of obtaining random sample means that will deviate this far or further from each other.

(b) Critical t values (t_{crit}) and calculated t

An equivalent hypothesis testing approach is to compare calculated t-values against critical t-values:

$$\textit{If } -t \leq -t_{crit} \textit{ or } t \geq t_{crit} \textit{ reject } H_0, \textit{ otherwise fail to reject } H_0$$

This can be simplified by using absolute values as follows:

$$\textit{If } |t| \geq |t_{crit}| \textit{ reject } H_0, \textit{ otherwise fail to reject } H_0$$

(c) Confidence Intervals

Since the null hypothesis usually specifies no difference between group means,

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

this indicates that the population difference between means should be 0.00 if the null is true.

One may calculate a confidence interval for the mean difference, and this interval provides a range of possible values for the population mean difference. Given this, one may specify the following decision rule:

If 0.00 lies within the CI for the mean difference, fail to reject H_0 , otherwise reject H_0

The logic here suggests that if 0.00 is within the confidence interval then 0.00 is a possible value for the population mean difference, and if 0.00 is a possible value for the population mean difference then the data collected are consistent with the null hypothesis of no difference, i.e., the data do not contradict $H_0: \mu_1 - \mu_2 = 0.00$. If, however, 0.00 is not within the confidence interval, this suggests 0.00 is not a reasonable value for the population mean difference so the collected data are not consistent with the null hypothesis therefore one may reject the null.

5. t-ratio, SE_d , and df

(a) t-ratio Formula

Like the two-samples t-test, the paired-samples t-test forms a ratio of mean differences divided by the standard error of that difference:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}$$

Since the difference between population means is expected to be 0.00, the component $\mu_1 - \mu_2$ drops from the equation leaving the reduced equation for the t-ratio:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{X}_1 - \bar{X}_2}{SE_d}$$

(b) Standard Error of Difference, SE_d

The t-ratio formula presented above is the same for both the two-independent samples and the paired-samples t-test. The difference between the tests is how the standard error of the mean difference (SE_d) is calculated, and also how degrees of freedom (df) are calculated. For the correlated t-test the standard error of the mean difference is calculated as

$$SE_d = s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) - 2(r_{12})\left(\frac{s_1}{\sqrt{n_1}}\right)\left(\frac{s_2}{\sqrt{n_2}}\right)}$$

The above SE_d formula contains a correlation component where r_{12} represents the Pearson product moment correlation between both sets of scores. The larger this correlation, the smaller will be the SE_d . Smaller SE_d tend to produce larger calculated t-ratios in absolute value, so smaller SE_d result is greater power for the test. Thus, the above formula capitalizes on any dependence, or correlation, among the sets of scores.

In the independent-samples t-test one assumes that the two sets of data are not correlated, so the standard error loses the correlation term, r_{12} , in the formula, i.e.:

$$SE_d = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) - 2(r_{12})\left(\frac{s_1}{\sqrt{n_1}}\right)\left(\frac{s_2}{\sqrt{n_2}}\right)} = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) - 2(\mathbf{0})\left(\frac{s_1}{\sqrt{n_1}}\right)\left(\frac{s_2}{\sqrt{n_2}}\right)}$$

$$= \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) - \mathbf{0}} = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

If there is no correlation, then the SE_d formula for the paired-samples t-test reduces to the SE_d formula given in the independent-samples t-test. In short, the primary difference between the independent-samples and paired-samples t-tests is the calculation of the standard error of the difference, SE_d .

(c) Degrees of Freedom, df

The df for the correlated t-test is calculated as:

$$df = n - 1$$

where n represents the number of pairs across the two sets of scores.

6. Computational Formula for Paired-samples t-test

Calculation of the SE_d for the correlated-samples t-test requires finding the Pearson product moment correlation, r_{12} , between the two sets of scores. While finding Pearson's r is not difficult, especially with use of software, it does introduce one additional computational step.

An easier approach to calculating the paired-samples t-test can be found with this formula

$$t = \frac{\bar{d} - \mu_d}{\sqrt{\frac{s_d^2}{n}}} = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}} = \frac{\bar{d}}{SE_d}$$

where \bar{d} is the mean of the differences between pairs of scores

$$\bar{d} = \frac{\sum d}{n}$$

and SE_d is the standard error of the mean difference

$$SE_d = \sqrt{\frac{s_d^2}{n}}$$

where s_d^2 is the variance of the difference scores, and is calculated like a regular variance, i.e.,

$$s_d^2 = \frac{\sum(d - \bar{d})^2}{n - 1}$$

In short, the correlated-samples t-test may be viewed as the mean of the differences, \bar{d} , divided by the standard error of the differences, SE_d

$$t = \frac{\bar{d}}{SE_d}$$

This computational approach for the t-test is illustrated below.

7. Confidence Intervals about the Mean Difference

For the independent-samples t-test the confidence interval is for the mean difference is

$$(1 - \alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (\text{critical } t \text{ value})(\text{standard error of mean difference})$$

and the same formula may be used to construct confidence intervals for the paired-samples t-test.

8. Example t-tests with SPSS and the Computational Formula

(a) Example 1

Is there a difference in mean systolic blood pressure before and after taking 50mg of Losartan? Raw data for a sample of eight participants is provided below.

Table 1
Systolic Blood Pressure by Losartan Usage

Participant	Systolic Blood Pressure	
	Before Losartan	After Losartan
Bill	145	133
John	133	119
Peggy	175	151
Bertha	139	133
Dot	143	146
Kenneth	132	126
Scott	149	141
Marshall	162	153

If calculating by hand, follow these steps:

(a) Determine difference scores for each set of matched scores:

Table 2
Systolic Blood Pressure by Losartan Usage Difference Scores

Participant	Systolic Blood Pressure		Difference Score
	Before Losartan	After Losartan	
Bill	145	133	12
John	133	119	14
Peggy	175	151	24
Bertha	139	133	6
Dot	143	146	-3
Kenneth	132	126	6
Scott	149	141	8
Marshall	162	153	9
		Sum =	76

i.e., 145 – 133 = 12

(b) Find the mean of the difference scores, \bar{d}

$$\bar{d} = \frac{\sum d}{n} = \frac{76}{8} = 9.5$$

(c) Find variance of the difference scores, s_d^2

$$s_d^2 = \frac{\sum (d - \bar{d})^2}{n - 1}$$

Table 3
Calculating Variance of Difference Scores

Difference Score, d	\bar{d}	$d - \bar{d}$	$(d - \bar{d})^2$
12	9.5	2.5	6.25
14	9.5	4.5	20.25
24	9.5	14.5	210.25
6	9.5	-3.5	12.25
-3	9.5	-12.5	156.25
6	9.5	-3.5	12.25
8	9.5	-1.5	2.25
9	9.5	-0.5	0.25
		Sum = $\sum (d - \bar{d})^2 =$	420.00

$$n - 1 = 8 - 1 = 7$$

$$s_d^2 = \frac{\sum (d - \bar{d})^2}{n - 1} = \frac{420}{7} = 60.00$$

(d) Find standard error of the difference scores, SE_d

$$SE_d = \sqrt{\frac{s_d^2}{n}} = \sqrt{\frac{60.00}{8}} = \sqrt{7.50} = 2.739$$

(e) Calculate the t-ratio

$$t = \frac{\bar{d}}{SE_d} = \frac{9.50}{2.739} = 3.468$$

(f) Find df and critical values for whichever significance level (α) is specified if using t-values for hypothesis testing:

$$df = n - 1 = 8 - 1 = 7$$

$$\alpha = .05; \text{ critical } t = \pm 2.36$$

$$\alpha = .01; \text{ critical } t = \pm 3.50$$

(g) Apply decision rule

If $|t| \geq |t_{crit}|$ reject H_0 , otherwise fail to reject H_0

*If **3.468** \geq **2.36** reject H_0 , otherwise fail to reject H_0*

Since the calculated t-value of 3.468 is greater than the critical t-value of 2.36 one would reject H_0 and conclude that mean systolic blood pressure appears to differ before and after taking Losartan.

(h) Calculate confidence interval about mean difference

$$(1 - \alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (\text{critical } t \text{ value})(\text{standard error of mean difference})$$

So for a 95% CI one would use the following numbers to construct the CI:

- $\alpha = .05$
- $\bar{X}_1 - \bar{X}_2 = \bar{d} = 9.50$
- *critical t value* = 2.36
- *standard error of mean difference* = $SE_d = 2.739$

$$(.95)CI = (9.50) \pm (2.36)(2.739)$$

$$\text{Upper Limit} = 9.50 + (2.36)(2.739) = 9.50 + 6.464 = 15.964$$

$$\text{Lower Limit} = 9.50 - (2.36)(2.739) = 9.50 - 6.464 = 3.036$$

The 99% CI, .99CI, would require changing the critical t value to 3.50, so the limits would be:

$$\text{Upper Limit} = 9.50 + (\mathbf{3.50})(2.739) = 9.50 + \mathbf{9.587} = 19.0877$$

$$\text{Lower Limit} = 9.50 - (\mathbf{3.50})(2.739) = 9.50 - \mathbf{9.587} = -0.087$$

Interpretation of the 95% CI:

One may be 95% confident that the interval 3.036 to 15.964 includes the population mean difference of systolic blood pressure before and after taking the medication Losartan.

Since 0.00 does not appear within this interval, one would reject H_0 since 0.00 is an unlikely value for the mean difference in blood pressure.

Results from SPSS are presented below.

Figure 1
SPSS Output for Paired-samples t-test of Blood Pressure by Losartan Usage

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	blood_pressure_before	147.2500	8	14.72364	5.20559
	blood_pressure_after	137.7500	8	12.08009	4.27096

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	blood_pressure_before & blood_pressure_after	8	.851	.007

Paired Samples Test									
		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	blood_pressure_before - blood_pressure_after	9.50000	7.74597	2.73861	3.02421	15.97579	3.469	7	.010

Questions

(a) What does the middle table represent?

(b) How would one obtain the 99% CI with SPSS; does the 99% CI change the inference and interpretation of results?

(b) Example 2

Do identical twins, separated at birth and raised by different families, show similar IQ scores? Below are IQ scores for twins.

Table 4
IQ Scores by Twins

Twin Pairs, Last Name	IQ Scores	
	Twin A	Twin B
Johnson	103	107
Murphy	92	90
Baggett	108	105
Sykes	129	128
Coleman	114	112
McNeil	121	118
Frasier	86	90

If calculating by hand, follow these steps:

(a) Determine difference scores for each set of matched scores:

Table 5
IQ Scores by Twins Difference Scores

Twin Pairs, Last Name	IQ Scores		Difference Score
	Twin A	Twin B	
Johnson	103	107	-4
Murphy	92	90	2
Baggett	108	105	3
Sykes	129	128	1
Coleman	114	112	2
McNeil	121	118	3
Frasier	86	90	-4
		Sum =	3

i.e., $103 - 107 = -4$

(b) Find the mean of the difference scores, \bar{d}

$$\bar{d} = \frac{\sum d}{n} = \frac{3}{7} = 0.4286$$

(c) Find variance of the difference scores, s_d^2

$$s_d^2 = \frac{\sum (d - \bar{d})^2}{n - 1}$$

Table 6
Calculating Variance of Difference Scores

Difference Score, d	\bar{d}	$d - \bar{d}$	$(d - \bar{d})^2$
-4	0.4286	-4.4286	19.6125
2	0.4286	1.5714	2.469298
3	0.4286	2.5714	6.612098
1	0.4286	0.5714	0.326498
2	0.4286	1.5714	2.469298
3	0.4286	2.5714	6.612098
-4	0.4286	-4.4286	19.6125
Sum = $\sum(d - \bar{d})^2 =$			57.7143

$$n - 1 = 7 - 1 = 6$$

$$s_d^2 = \frac{\sum(d - \bar{d})^2}{n - 1} = \frac{57.7143}{6} = 9.6191$$

(d) Find standard error of the difference scores, SE_d

$$SE_d = \sqrt{\frac{s_d^2}{n}} = \sqrt{\frac{9.6191}{7}} = \sqrt{1.3742} = 1.1723$$

(e) Calculate the t-ratio

$$t = \frac{\bar{d}}{SE_d} = \frac{0.4286}{1.1723} = 0.3656$$

(f) Find df and critical values for whichever significance level (α) is specified if using t-values for hypothesis testing:

$$df = n - 1 = 7 - 1 = 6$$

$$\alpha = .05; \text{ critical } t = \pm 2.45$$

$$\alpha = .01; \text{ critical } t = \pm 3.71$$

(g) Apply decision rule

If $|t| \geq |t_{crit}|$ reject H_0 , otherwise fail to reject H_0

*If **$|0.3656| \geq |2.45|$** reject H_0 , otherwise fail to reject H_0*

Since the calculated t-value of 0.3656 is less than the critical t-value of 2.45 one would fail to reject H_0 and conclude that mean IQ between twins is similar.

(h) Calculate confidence interval about mean difference

$$(1 - \alpha)CI = (\bar{X}_1 - \bar{X}_2) \pm (\text{critical } t \text{ value})(\text{standard error of mean difference})$$

So for a 95% CI one would use the following numbers to construct the CI:

- $\alpha = .05$
- $\bar{X}_1 - \bar{X}_2 = \bar{d} = 0.4286$
- *critical t value* = 2.45
- *standard error of mean difference* = $SE_d = 1.1723$

$$(.95)CI = (0.4286) \pm (2.45)(1.1723)$$

$$\text{Upper Limit} = 0.4286 + (2.45)(1.1723) = 0.4286 + 2.8721 = 3.30$$

$$\text{Lower Limit} = 0.4286 - (2.45)(1.1723) = 0.4286 - 2.8721 = -2.44$$

The 99% CI, .99CI, would require changing the critical t value to 3.71, so the limits would be:

$$\text{Upper Limit} = 0.4286 + (3.71)(1.1723) = 0.4286 + 4.3492 = 4.778$$

$$\text{Lower Limit} = 0.4286 - (3.71)(1.1723) = 0.4286 - 4.3492 = -3.92$$

Interpretation of the 95% CI:

One may be 95% confident that the interval -2.44 to 3.30 includes the population mean difference of IQ scores between twins.

Since 0.00 appears within this interval, one would fail to reject H_0 since 0.00 is one likely value for the mean difference in IQ scores for twins.

Results from SPSS are presented below.

Figure 2
SPSS Output for Paired-samples t-test of IQ Scores by Twins

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair	Twin_A	107.5714	7	15.32816	5.79350
1	Twin_B	107.1429	7	13.95742	5.27541

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	Twin_A & Twin_B	7	.982	.000

Paired Samples Test									
		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Twin_A - Twin_B	.42857	3.10146	1.17224	-2.43980	3.29694	.366	6	.727

9. APA Style Results Presentation

It is possible to report paired-samples t-test results in either textual or tabular format, but the tabular format is preferred. Both options are provided below.

(a) Tabular Format

Two examples are provided:

- Table 7 shows a t-test for blood pressure data, and
- Table 8 provides results from three t-tests reported simultaneously in one table.

As with the two-independent samples t-test, the written presentation following a table should include first inferential information (whether H_0 was rejected) and this should be followed with interpretational information (what the results mean in simple language).

Table 7 demonstrates APA style for the t-test with blood pressure data.

Table 7

Results of t-test and Descriptive Statistics for Systolic Blood Pressure by Losartan Usage

Outcome	Before Losartan		After Losartan		n	95% CI for Mean Difference	r	t	df
	M	SD	M	SD					
	147.25	14.72	137.75	12.08	8	3.02, 15.98	.85*	3.47*	7

* $p < .05$.

Results show a statistically significant difference in mean systolic blood pressure before and after Losartan usage. Blood pressure appears to drop an average of about 9.5 points following the usage of Losartan.

Table 8 shows how to combine results from several paired-samples t-tests performed for the same groups of participants.

Table 8

Descriptive Statistics and t-test Results for Anxiety, Efficacy, and Persistence

Outcome	Pretest		Posttest		n	95% CI for Mean Difference	r	t	df
	M	SD	M	SD					
Anxiety	5.80	1.89	6.12	1.23	45	-0.48, -0.16	.75*	-2.59*	44
Efficacy	5.24	1.56	5.31	1.97	44	-0.16, 0.02	.86*	1.98	43
Persistence	6.13	1.89	5.56	1.21	45	0.55, 0.83	.90*	4.89*	44

* $p < .05$.

As displayed in Table 8, there are statistically significant differences, at the .05 significance level, in pretest to posttest scores for anxiety and persistence, but not for efficacy. Results show that anxiety increased, persistence decreased, and efficacy remained unchanged after exposure to the treatment.

(b) Textual Format

Results of the paired-samples t-test show that mean systolic blood pressure differs before taking Losartan ($M = 147.25$, $SD = 14.72$) and after taking Losartan ($M = 137.75$, $SD = 12.08$) at the .05 level of significance ($t = 3.47$, $df = 7$, $n = 8$, $p < .05$, 95% CI for mean difference 3.02 to 15.98, $r = .85$). On average systolic blood was about 9.5 points lower after taking Losartan.

10. Matched individually vs. Sub-group Matching

Often data suitable for a paired-samples t-test form pairs naturally such as with before-after type studies (e.g., weight from same individuals before and after diet) or studies of naturally occurring pairs (e.g., attitudinal differences of husbands and wives; comparison of scores between siblings).

Sometimes data may be paired through use of matched pairings such as with case-control studies (popular in the medical field). Matched data can pose problems for the paired-samples t-test if subgroups rather than individuals are matched. The two types of matching are illustrated in Tables 9 and 10.

Note that when subgroups are matched, often there can be unequal numbers per subgroup for a given match (e.g., 4 individuals with IQ 110+ for Group A vs. 2 individuals for Group B), whereas uniquely matched individuals will form 1-to-1 pairs. The latter type, illustrated in Table 9, are suitable for paired-samples t-test. If one has the former type, illustrated in Table 10, then a more suitable analysis may be two-way ANOVA with the matching variable treated as a blocking factor.

Table 9
Individuals Matched Unique on IQ Scores

	Group A (treatment)		Group B (control)	
<u>IQ score</u>				<u>IQ score</u>
133	Sue		John	133
126	Susan		Fred	126
		...		
86	Bryan		Walt	86
72	Bill		Josh	72

Table 10
Subgroups of Individuals Matched on IQ Scores

	Group A (treatment)		Group B (control)	
<u>IQ score</u>				<u>IQ score</u>
High (110+)	Beth, Sue, John, Greg		John and Ann	High (110+)
Middle (90-110)	Bob and Susan		Fred, Bill, and Erin	Middle (90-110)
Low (<90)	Bryan and Bill		Josh, Walt, and Jon	Low (<90)

Why is the paired-samples t-test ill-suited for analysis of matches like those found in Table 10? It is possible to randomly form pairs within matched subgroups so t-test results could vary according to which pairs were matched – change pairings and the t-test results could change.

11. Topics not Covered

(a) Directional tests

Directional (group 1 has higher mean than group 2)

The experimental group will show a higher level of achievement.

$$H_0: \mu_1 \leq \mu_2$$

$$H_1: \mu_1 > \mu_2$$

or

$$H_0: \mu_1 - \mu_2 \leq 0.00$$

$$H_1: \mu_1 - \mu_2 > 0.00$$

Directional (group 2 has higher mean than group 1)

The experimental group will show a lower level of achievement.

$$H_0: \mu_1 \geq \mu_2$$

$$H_1: \mu_1 < \mu_2$$

or

$$H_0: \mu_1 - \mu_2 \geq 0.00$$

$$H_1: \mu_1 - \mu_2 < 0.00$$

These are not covered because I think non-directional is better since it less likely one will “cheat” to obtain statistical significance (i.e., adopts in a post-hoc fashion directional tests since they offer more power compared with non-directional tests). Also statistical software tends to provide p-values only for non-directional tests.

(b) Effect Size d

The effect size d represents the mean difference divided by the common or pooled standard deviation:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s}$$

Like a Z score, this value indicates the size of the mean difference in standard deviation units. The d formula may appear to be the same as the t formula, but note the different denominator. In the d formula the denominator is s, the sample standard deviation, and in the t formula the denominator is $s_{\bar{x}}$, standard error of the mean.

(c) Sample Size

One may calculate required sample size for a two-sample t test to meet specifications for Type 1 error rate (α), power ($1-\beta$), and effect size (d). This requires use of tabled values or software.

12. Exercises

See the course web page and the course text to find exercises for correlated-samples t test.