

CHAPTER 10

SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO SAMPLE MEANS

OBJECTIVES

After completing this chapter, you should

- understand the derivation and properties of the sampling distribution of differences and the related distribution of t scores.
- be able to test the significance of the difference between two independent sample means.
- be able to test the significance of the difference between two dependent sample means.

CHAPTER REVIEW

The chapter begins with the derivation of the sampling distribution of the differences for independent samples. First, pairs of random samples are taken from a population, and means for some characteristic are computed. Next, the difference between the pairs of means is determined. A frequency distribution of the differences is constructed, and the distribution is used to plot a frequency polygon. The distribution is called the *sampling distribution of the mean differences*.

The properties of the sampling distribution of the mean differences are (1) its mean is equal to 0; (2) the larger the size of the samples drawn from the population, the more closely the sampling distribution approximates the normal curve; (3) the larger the size of the samples, the smaller the standard deviation of the sampling distribution. The standard deviation is called the *standard error of the difference between means* and is symbolized by $\sigma_{\bar{x}_1 - \bar{x}_2}$. Both raw-score and defining formulas for estimating the standard error of the differences are given. The estimated standard error is symbolized by $s_{\bar{x}_1 - \bar{x}_2}$.

As in Chapter 9, hypothesis testing involves computing how far from the mean of the sampling distribution our observed difference in sample means lies in estimated standard error units. We first compute a *t* score and then compare the computed *t* score with values known to cut off the deviant 5% (or

1%) of the distribution of t for samples of a given size converted to degrees of freedom. The t score or t ratio is the ratio of the difference between a pair of sample means to the estimated standard error of the differences. The degrees of freedom for the t test for independent samples is $N_1 + N_2 - 2$, and the critical values again are found in Table B.

A *two-tailed* test looks at both ends of the distribution of the test statistic, t in this chapter. Only one end of the distribution is considered in the *one-tailed* test. The end considered in the one-tailed test is the one predicted by the experimenter before conducting the experiment. The one-tailed test is a more powerful test if the outcome of the experiment is in the predicted direction. If you can reasonably predict the outcome's direction before data collection, use of the one-tailed test is warranted because it is more powerful. To determine the one-tailed probabilities, halve the probability values in Table B.

Assumptions required by the two-sample t test for independent samples are that the populations from which the samples are drawn are normal, that the population variances are homogeneous (equal), and that the samples are independent. The first two assumptions apparently can be violated with little effect upon the conclusions made with the test. This property of a statistical test to give valid conclusions even when its assumptions are violated is called *robustness*. If you fear violation of the assumptions, it is recommended that you use fairly large samples of equal size.

The power of the test can be increased by using dependent samples. Dependent samples can be formed by using *matched pairs* of unrelated individuals in which subject pairs, matched as closely as possible on relevant characteristics, are selected. Then, one member of each pair is assigned to one treatment group, and the other member is assigned to the other group. Another procedure for obtaining dependent samples is to use each subject as his or her own control; that is, the same individuals are given each experimental treatment. This type of experiment is called a *repeated measures* design or *within-subjects comparison*.

The power of the test using dependent samples is increased through a decrease in the standard deviation of the sampling distribution of mean differences. The standard deviation is called the *standard error of the mean differences* and is symbolized by $\sigma_{\bar{D}}$. $\sigma_{\bar{D}}$ is estimated by $s_{\bar{D}}$, which is called the *estimated standard error of the mean differences*.

The direct difference method is used to compute a t ratio for dependent or related samples. In the final equation discussed, t is computed by dividing the mean of the differences by the standard error of the mean differences. The computed value of t is then compared with the critical values of t from Table B with $df = N - 1$, where N is the number of pairs of scores.

SYMBOLS

<i>Symbol</i>	<i>Stands For</i>
$\sigma_{\bar{x}_1 - \bar{x}_2}$	standard error of the difference between means
$s_{\bar{x}_1 - \bar{x}_2}$	estimated standard error of the mean differences
$\bar{X}_1 - \bar{X}_2$	score in the sampling distribution of the differences
$\mu_1 - \mu_2$	mean of the sampling distribution of the differences
$z_{\bar{x}_1 - \bar{x}_2}$	z score based on the sampling distribution of the differences
$t_{\bar{x}_1 - \bar{x}_2}$	t score based on the sampling distribution of the differences
s^2_{pooled}	pooled variance; an estimate of the common variance of the parent populations
$\sigma_{\bar{D}}$	standard error of the mean differences
$s_{\bar{D}}$	estimated standard error of the mean differences
$\bar{X}_{\bar{D}}$	mean of the differences
$s_{\bar{D}}$	standard error of the differences
$s_{\bar{D}}$	standard deviation of the differences
D	difference between a pair of scores

FORMULAS

Formula 10-3. Computational formula for the estimated standard error of the mean differences for independent samples

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \right) \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}$$

This formula is used to compute the estimated standard error of the mean differences. N_1 and N_2 are the numbers of subjects in the first and second samples, respectively. s_1^2 and s_2^2 are the variances of the two samples.

Formula 10-6. Equation for the two-sample t test for independent samples

$$t_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

The t ratio for the two-sample t test for independent samples is the difference in sample means divided by the estimated standard error of the mean differences. When the computational formula for $s_{\bar{X}_1 - \bar{X}_2}$ is substituted in the denominator, the t ratio becomes

Formula 10-7. Computational equation for the two-sample t test for independent samples

$$t_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \right) \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

Formula 10-8. Computational formulas for the t test for dependent samples

$$t = \frac{\bar{X}_D}{s_{\bar{X}_D}} \text{ or } \frac{\bar{X}_D \sqrt{N}}{s_D}$$

TERMS TO DEFINE AND/OR IDENTIFY

independent t test

control group

independent sample

sampling distribution of the mean differences

standard error of the difference between means

estimated standard error of the mean differences

two-tailed test

one-tailed test

homogeneity of variance

robustness of a test

matched pairs

within-subjects comparison

repeated measures design

counterbalancing

double-blind

FILL-IN-THE-BLANK ITEMS

Introduction

In Chapter 9 we looked at a distribution composed of means of single samples drawn from a population, whereas in this chapter we considered (1) _____ samples taken simultaneously.

The Sampling Distribution of the Differences Between Sample Means

If the behavior of the members of one sample is not related to the behavior of subjects in the other sample, the samples are (2) _____. The assumption of independence is often made as long as subjects are selected at (3) _____ and are (4) _____ assigned to the different treatment conditions.

To construct the sampling distribution of the differences, we first draw (5) _____ of random samples from the same population. For each sample of the pair, a (6) _____ is computed and the (7) _____ between the means is found. A frequency (8) _____ based on the differences is then made, and from this a frequency (9) _____ is plotted. The standard deviation of the sampling distribution of differences is called the (10) _____ of the differences.

Three properties of the sampling distribution of the differences are: the mean of the distribution is equal to (11) _____; the larger the size of the samples taken from the population, the more closely the distribution approximates the (12) _____ curve; and the larger the size of the samples, the (13) _____ the standard error of the difference between means.

In the sampling distribution of the differences, a score is symbolized by (14) _____. The mean is symbolized by (15) _____, and the standard error is symbolized by (16) _____. Putting these together, the formula for a z score is (17) _____ = _____. If we divide by the estimated standard error, we obtain the formula for t , which is usually written as (18) _____ = _____.

Computing t : Independent Samples

To compute t for independent samples, we need three things: the mean of each sample, the (19) _____ of subjects in each sample, and the (20) _____ of each sample. The null hypothesis is that both samples were drawn from the (21) _____ and that the mean

of the sampling distribution is (22) _____. The degrees of freedom for the test are given by (23) _____.

One-tailed versus two-tailed tests

A (24) _____ test of significance is one considering both ends of the distribution. To use it, we don't have to make any (25) _____ about the experiment's outcome. The (26) _____ test, on the other hand, looks only at the tail of the distribution predicted by the experimenter before the experiment. However, with a one-tailed test, we must make our prediction (27) _____ doing the study.

The (28) _____ test is a more powerful test if the prediction comes true. A more powerful test is one with which it will be (29) _____ to reject the null hypothesis.

Assumptions of the two-sample t test

The t test assumes that the dependent variable is (30) _____ distributed in the population from which the samples are drawn. Another assumption is that the population (31) _____ are homogeneous. If you have reason to suspect that the assumptions will be violated, you should use (32) _____ samples with the same number of subjects in each. Both assumptions can apparently be violated with (33) _____ effect upon the conclusions reached with the t test, which means that the t test is a (34) _____ test.

Computing t : Dependent Samples

In testing H_0 , the most desired outcome is (35) _____ of the null hypothesis. One way to increase the (36) _____ of the t test is to use dependent samples.

One way to obtain dependent samples is to form (37) _____ of unrelated individuals with one member of a pair assigned to one treatment group and the other member assigned to the other group. In the (38) _____ design, each subject is given both treatments; that is, each subject is his or her own (39) _____. Sometimes, this

type of design is called the (40) _____ design. The dependent-samples design increases the power of the test by (41) _____ the standard deviation of the sampling distribution of differences between related samples.

The direct difference method

(42) _____ is a procedure used to control for the effects of the order of presentation of the treatment in experiments. A (43) _____ method of drug presentation is one in which neither the administrator nor the subject knows which drug is being given. With the direct difference method, all calculations are based on the (44) _____ between each pair of scores rather than on the scores themselves. The t ratio is the mean of the differences divided by the estimated (45) _____ of the differences. The mean of the differences is defined as the (46) _____ sum of the differences divided by N , which is the number of (47) _____ of scores. For the test, $df =$ (48) _____.

Troubleshooting Your Computations

The (49) _____ used must be appropriate to your data; that is, when the samples are assumed to be independent of each other, the t test for (50) _____ samples should be used. If the samples are related or matched in some way, the t test for (51) _____ samples should be used.

The estimated standard error of the differences should always have a (52) _____ sign. Also, be sure to retain the appropriate (53) _____ when computing the final value for t . In the t test for dependent samples, all computations are made on the (54) _____ scores rather than on the actual scores themselves. Be careful to add the difference scores (55) _____ — that is, taking the signs into account.

Remember the decision rule for a nondirectional test: If the absolute value of the computed t is equal to or larger than the critical value of t from Table B, (56) _____ H_0 .

PROBLEMS

1. For each of the following, compute $s_{\bar{X}_1 - \bar{X}_2}$ using both the raw-score and the defining formulas. Assume that the samples are independent.

a. $N_1 = N_2 = 12, s_1 = 3.6, s_2 = 4.3$

b. $N_1 = 22, \Sigma X_1 = 112.2, \Sigma X_1^2 = 643.5; N_2 = 22, \Sigma X_2 = 138.6, \Sigma X_2^2 = 1,010.68$

c. $N_1 = N_2 = 25, s_1^2 = 2.25, s_2^2 = 1.96$

2. In a military training program, the complex reaction-time ability of pilots and navigators is compared. Determine whether there is a significant difference between the groups in the number of errors made (failure to respond) in 100 stimulus presentations.

<u>Pilots</u>	<u>Navigators</u>
$N = 15$	$N = 20$
$\bar{X} = 23.5$	$\bar{X} = 41.3$
$s = 10.5$	$s = 12.7$

3. Experimental evidence indicates that persons with a family history of alcoholism are more likely to become alcoholic than persons with no such history. A large group of adults with and without family histories of alcoholism is selected, and from this large group, eight pairs of participants are matched in terms of race, age, marital status, and drinking history. Each participant is given a drink of alcohol; 30 minutes later, a blood sample is taken, and a metabolite of alcohol is measured. The results are shown here.

Pair Number	Family History ($\mu\text{g/ml}$)	No Family History ($\mu\text{g/ml}$)
1	3.1	2.5
2	3.5	2.1
3	3.2	1.8
4	2.8	2.3
5	2.6	2.2
6	3.0	1.9
7	1.9	2.3
8	2.7	2.5

Compare the groups with and without a family history of alcoholism.

4. Thirty-two students are randomly selected from a large introductory class and are randomly and evenly assigned to one of two groups. Each participant is given four 15-second exposures to a list of 15 nouns. After 30 minutes, each participant in one group is given a free recall test, whereas each participant in the other group is given a recognition test. The results in terms of the number of words correctly retained are shown here. Compare the groups.

<u>Recall Test</u>		<u>Recognition Test</u>	
X	f	X	f
4	1	15	6
2	5	14	1
1	7	13	4
0	3	12	2
		10	1
		8	1
		6	1

5. In an attempted replication of the Rosenthal effect (experimenter bias can influence the outcome of an experiment), two groups of 15 randomly selected students each are given some rats to train. One group is told that the rats are retarded, whereas the other group is told that the rats are intelligent. Errors are recorded during training, and the data for each group are shown here.

<u>Group "Stupid"</u>	<u>Group "Intelligent"</u>
$\Sigma X = 379.5$	$\Sigma X = 252$
$\Sigma X^2 = 9,853.5$	$\Sigma X^2 = 4,450.2$
$N = 15$	$N = 15$

- a. Before you analyze the data, determine whether there is a rationale for using the one-tailed test of significance. If so, what is it?
- b. Compare the groups using the appropriate statistical test.

6. Ten politically active individuals are selected, all of whom label themselves as liberal. Each is attached to a physiograph, and heart rate is recorded while a sequence of 20 slides is projected on a screen at which the participant is looking. Half of the slides are pictures of famous individuals without any expressed political philosophy; the other half are known conservatives. The arrangement of the slides is random. Each participant receives two scores: The first is the average heart rate during exposure to the neutral slides, and the second is the average heart rate during exposure to the conservative slides. The data are shown here. Does reaction to the slides differ?

<u>Participant</u>	<u>Neutral</u>	<u>Conservative</u>
1	65.3	71.8
2	75.7	73.5
3	85.6	99.3
4	73.7	81.7
5	69.5	75.7
6	68.2	73.5
7	70.1	79.8
8	72.5	70.3
9	71.0	85.3
10	83.5	107.1

7. An experiment is performed to compare the performance of students taught statistics by a traditional lecture approach with students taught by group discussion sessions. Fifty students are randomly and equally assigned to one of the two treatment conditions. In each case, the same instructor is present; he lectures to one group and leads the discussion in the other. The final class averages based on scores on standardized, machine-scored tests are presented here. Is there a significant difference between the groups?

<u>Lecture</u>	<u>Discussion</u>
$N = 25$	$N = 25$
$\bar{X} = 81.7$	$\bar{X} = 74.1$
$s = 8.3$	$s = 10.1$

8. The average ACT score for 937 freshmen at State University is 21.4, with a variance of 24.1. For Private University in the same city, the average ACT score for 421 freshmen is 22.1, with a variance of 14.5. Is there a significant difference between the schools in average freshman ACT scores? If so, how can you account for it in terms of factors affecting the power of a test?

9. Ten students are tested for accuracy of distance estimation using either one or both eyes. The test object is 24 inches from the viewer. Is there a significant difference in the amount of error (in inches) between the two conditions?

<u>Student</u>	<u>One Eye</u>	<u>Two Eyes</u>
A	3	2
B	11	8
C	5	7
D	4	2
E	6	4
F	3	1
G	15	10
H	4	5
I	8	7
J	9	7

10. A study is done to see whether children are more likely to have psychic powers than young adults. Thirteen randomly selected 8-year-old children each make one run through a standard PSI deck trying to guess the identity of each target card by “reading” the mind of the tester. There are four different targets, and each child receives a score indicating proportion of “hits.” The same procedure is used on 15 randomly selected college students. Compare the groups.

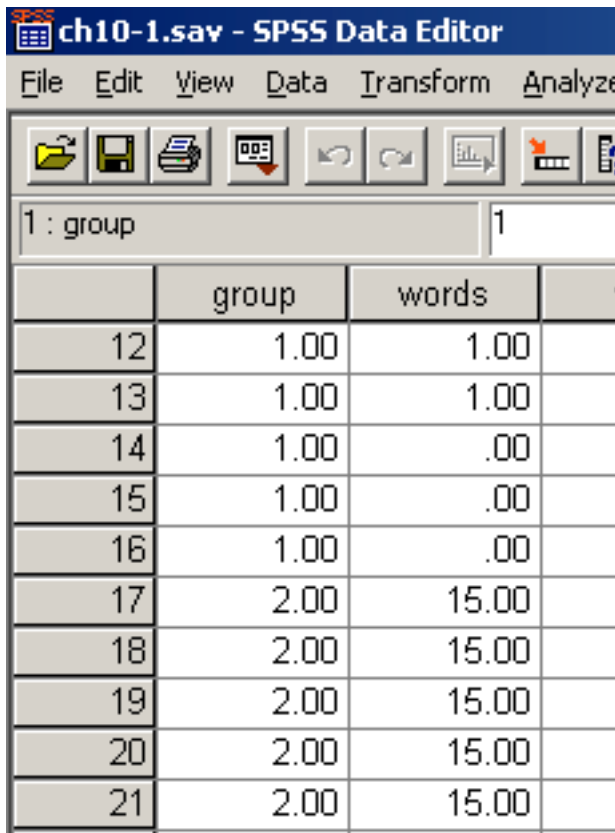
<u>Children</u>	<u>Young Adults</u>
$N = 13$	$N = 15$
$\bar{X} = .28$	$\bar{X} = .24$
$s = .067$	$s = .073$

USING SPSS—EXAMPLES AND EXERCISES

The SPSS procedure for computing t tests has options that will compute both the t test for independent samples and the t test for dependent samples.

Example: First, we will use SPSS to work an exercise for the t test for *independent* samples—specifically, Problem 4. The steps are as follows:

1. Start SPSS. Data entry is slightly different for this type of problem. Name the first variable **group** and the second, **words**. Enter the data for the Recall Test group under **words** with a 1 in the **group** column, and enter the data for the Recognition Test group under **words** but with a 2 in the **group** column in front of the data. Part of the Data Editor showing the pattern of data entry follows:



The screenshot shows the SPSS Data Editor window titled 'ch10-1.sav - SPSS Data Editor'. The menu bar includes File, Edit, View, Data, Transform, and Analyze. Below the menu bar is a toolbar with icons for file operations and data manipulation. The data grid shows a table with the following data:

	group	words	
12	1.00	1.00	
13	1.00	1.00	
14	1.00	.00	
15	1.00	.00	
16	1.00	.00	
17	2.00	15.00	
18	2.00	15.00	
19	2.00	15.00	
20	2.00	15.00	
21	2.00	15.00	

2. Choose *Analyze>Compare Means>Independent Samples T Test*.
3. Move **words** into the Test Variables box and **group** into the Grouping Variable box.
4. Click on *Define Groups* and enter the code values for Group 1 and Group 2. The code values will be 1 and 2 because these are the numbers we entered to identify the groups. Click *Continue>OK*. The results should appear in the output Viewer window.

Notes on Reading the Output

1. The t value we will use is the one in the first row labeled “equal variances assumed,” because this is one of the assumptions we have made. (Levene’s Test is a test of the homogeneity of variances assumption.)
2. “Sig. (2-tailed)” is the exact p value for the computed t . We reject the null hypothesis if this value is $\leq .05$ and conclude that there is a significant difference between the groups.

T-TEST

```
GROUPS=group(1 2)
/MISSING=ANALYSIS
/VARIABLES=words
/CRITERIA=CIN(.95) .
```

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
WORDS	1.00	15	1.2667	1.0328	.2667
	2.00	16	12.7500	2.6957	.6739

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
WORDS	Equal variances assumed	6.193	.019	-15.456	29	.000	-11.4833	.7430	-13.0029	-9.9638
	Equal variances not assumed			-15.844	19.551	.000	-11.4833	.7248	-12.9974	-9.9693

Example: Now we will use SPSS to work an example of the *t* test for *dependent* samples—Problem 6. The steps are as follows:

1. Start SPSS, name variables **neutral** and **conserv**, and enter the data.
2. *Analyze>Compare Means>Paired-Samples T Test.*
3. Move both variables into the Paired Variables box, then click *OK*. The results should appear in the output Viewer window.

Notes on Reading the Output

1. You should find the results reasonably easy to identify and interpret in this output. The critical results are found in the Paired Samples Test box. The results indicate that $t(9) = -3.389$ and $p = .008$. As before, SPSS labels the p value as “Sig. (2-tailed).” Because the obtained p value is less than our alpha level of .05, we reject the null hypothesis and conclude that there is a significant difference in the heart rates, which increase when a liberal views the slide of a known conservative.

T-TEST

```
PAIRS= neutral WITH conserv (PAIRED)
/CRITERIA=CIN(.95)
/MISSING=ANALYSIS.
```

T-Test

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 NEUTRAL	73.5100	10	6.5101	2.0587
CONSERV	81.8000	10	12.3371	3.9013

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 NEUTRAL & CONSERV	10	.839	.002

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	NEUTRAL - CONSERV	-8.2900	7.7348	2.4460	-13.8232	-2.7568	-3.389	9	.008

Exercises Using SPSS

1. Use SPSS to work Self-Test Exercise 6, assuming the data are from independent groups; that is, assume that 20 participants were randomly assigned to attend or not to attend a speed reading course. At the conclusion of the course, all participants were given an evaluation to measure their reading speed in words per minute. Was the class effective? (*Hint:* Don’t forget the differences in the data input arrangement between the independent and dependent *t* tests.) Write a conclusion for this problem.
2. Use SPSS to work Self-Test Exercise 6 just as stated—as a dependent *t* test problem. You will have to re-enter your data. Compare your results with those from SPSS Exercise 1. Given the same data, which design is more powerful?

CHECKING YOUR PROGRESS: A SELF-TEST

1. What are the properties of the sampling distribution of the mean differences?
2. Describe the research conditions in which a t test for independent samples is appropriate.
3. Describe the research conditions in which a t test for dependent samples is appropriate.
4. A t ratio is another version of which of the following?
 - a. standard score
 - b. z score
 - c. two-tailed test
 - d. standard error
 - e. a and b
5. A social psychologist studied the effect of leadership style on the productivity of workers at an assembly plant for fax machines. Nine teams of workers were supervised by democratic leaders, and 15 teams worked for autocratic leaders. The average number of fax machines assembled in an hour by each team was recorded. The groups working for democratic leaders completed an average of 28.6 machines per hour, with $s^2 = 62.41$. The groups working for autocratic leaders completed an average of 26.2 machines per hour, with $s^2 = 45.32$. Did leadership style affect performance?

6. A group of 10 students completed a class to improve reading speed. Their average reading speed, in words per minute, was measured before and after taking the class. Was the class effective?

<u>Student</u>	<u>Speed Before</u>	<u>Speed After</u>
A	476	519
B	342	460
C	527	499
D	261	450
E	439	610
F	517	547
G	391	519
H	422	662
I	321	456
J	410	420