

# Contents

9.8	SPSS and MINITAB Tests . . . . .	682
9.8.1	T-TEST procedure of SPSS . . . . .	682
9.8.2	TWOSAMPLE procedure of MINITAB . . . . .	685
9.8.3	Summary of Computer Programs . . . . .	687
9.9	Reporting Results of Hypothesis Test . . . . .	687
9.9.1	Summary . . . . .	691
9.10	Paired t Test . . . . .	692
9.11	Conclusion . . . . .	701

## 9.8 SPSS and MINITAB Tests

Since statistical programs on computers have become widely available, most *t* tests are conducted using a computer program. This section contains results from *t* tests conducted with two of the more commonly used statistical programs, SPSS and MINITAB. This section contains no new theoretical developments. An example of the results from each of these programs is given, along with a short note concerning how these results can be interpreted.

### 9.8.1 T-TEST procedure of SPSS

The test for the difference of two means in SPSS is provided with the **T-TEST** procedure. An example of the output provided by SPSS from this procedure is given in the following example. The program used to obtain these results is not given here. For instructions concerning how to use the SPSS program, you will need to consult an SPSS manual.

#### Example 9.8.1 Variability of Opinions in Edmonton using SPSS

*The results from Example 9.6.4 are given again in this example using the SPSS computer program. This example looks at the mean opinion levels of Liberals and PCs in Edmonton, where the opinions concern whether or not trade unions are partially responsible for unemployment, by pricing their members out of jobs. The scale is a 7 point scale with large numerical values representing agreement and small numerical values representing disagreement. Table 9.19 contains the output obtained from the T-TEST procedure of SPSS.*

*Group 1 is the sample of PCs from the Edmonton Area Study, and Group 2 is the sample of Liberals from this same study. The first panel of Table 9.19 contains the summary data from the two samples. The sample sizes, means and standard deviations are all given in the first panel, and this is the raw material for conducting the test. Note that the results are carried to more decimals than are usually used.*

*The SPSS output contains both the pooled and the separate variance *t* tests, along with the test for the difference of variances. Examining the latter first, it can be seen that the results given in Table 9.19 are the same as those in Example 9.7.2 That is, the *F* value associated with the ratio of the two variances is given as 3.67. Except for a small amount of rounding error, this is the same as the 3.67 calculated in Example 9.7.2. Note that the exact significance of this *F* value is provided, so that you need not*

GROUP 1 - PC  
GROUP 2 - LIBERAL

Variable	Number of Cases	Mean	Standard Deviation	Standard Error
-----				
VAR150	UNEMPL HI - UNIONS PRICED OUT JOB D-Q27H			
GROUP 1	40	5.5750	1.196	.189
GROUP 2	9	3.6667	2.291	.764
-----				
*POOLED VARIANCE ESTIMATE*SEPARATE VARIANCE EST				
* * *				
F	2-TAIL	* T	DEGREES OF 2-TAIL	* T DEGR OF 2-TAIL
VALUE	PROB.	*VALUE	FREEDOM	PROB. * VALUE FREEDOM
-----				
3.67	0.006	* 3.59	47	0.001 * 2.43 9.00 0.038
-----				

Table 9.19: SPSS T-TEST for Difference in Opinions of PCs and Liberals in Edmonton

consult an  $F$  table. That is, if it is assumed that the two variances are equal, the probability of obtaining sample standard deviations as different as 1.196 and 2.291 is 0.006. This is a very small probability, so that the assumption of equal variances can be rejected. Note that this probability of 0.006 is consistent with the result in Example 9.7.2 since this probability is less than 0.01. That is, the  $F$  value is well within the region of rejection of the hypothesis of equal variances.

Given the conclusion that the two variances are unequal, the separate variance  $t$  test should be used. This is a repeat of the  $t$  test conducted in Example 9.6.4. The hypotheses there were

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

$$H_1 : \mu_1 - \mu_2 > 0$$

The output from the SPSS program gives the  $t$  value associated with the difference of sample means. Except for rounding differences, this is 2.43, the same as the  $t$  value of 2.418 from the  $t$  test of Example 9.6.4. The only difference is that the degrees of freedom is 9, rather than 8. The degree of freedom in Table 9.19 is accurately determined from the exact formula for degrees of freedom on page 670. In the earlier example, this formula was not used, and the smaller of the two sample sizes, minus one, was used as the estimate of the degrees of freedom.

The  $t$  value from the computer output can be checked against the  $t$  table in Appendix I. Pick a significance level, and using 9 degrees of freedom, determine the critical region. If the  $t$  value from the SPSS program is in the critical region, then reject  $H_0$ , but if the  $t$  value is not in the critical region, do not reject the null hypothesis. Once you get used to the computer output, even this is unnecessary. The exact significance of this  $t$  value is computed by the SPSS program. This can be seen to be 0.038. This is a two tailed probability, associated with a two tailed alternative hypothesis. If it is assumed that the means of the two populations are equal, the probability of obtaining a  $t$  value larger than +2.43 or smaller than -2.43 is 0.038. That is,

$$P(t < -2.43 \text{ or } t > +2.43) = 0.038.$$

Since the test here is a one tailed test, the probability given from the SPSS program must be halved in order to determine the exact significance associated with the  $t$  value. That is, the probability associated with a  $t$  value of 2.43 or more is one half of 0.038 or 0.019.

$$P(t > +2.43) = 0.038/2 = 0.019.$$

Since this is an extremely low probability, the null hypothesis of equal mean opinion levels for the PCs and Liberals can be rejected. This hypothesis can be rejected at any level of significance above 0.019. For example, if the 0.02 level of significance had been used, the  $t$  value has a probability less than this (0.019), so this  $t$  value is in the region of rejection of  $H_0$ .

**Additional Comments on the SPSS Program Output.**

The SPSS program provides all the results necessary to conduct the test for the difference between two means. The sample means, standard deviations, and sample sizes are given. The F value and its probability are used merely in order to determine whether the pooled or the separate variance t tests are to be used. If the F value is quite large, with a small probability, the separate variance test is used. If the F value is smaller, with a larger probability, then the pooled variance test is used. Exactly how large the probability associated with the F value should be depends on which significance level the researcher feels is appropriate.

As a rough rule of thumb, if the 2-tail probability associated with the F value is less than 0.10, use the separate variance t test. If the 2-tail probability associated with the F value is greater than 0.10, use the pooled variance t test.

Once you have decided which of the two t values to use, then the t value and the degrees of freedom are given on the computer output. The exact significance associated with this t value are also given. In an article or research report, this result could be reported as it appears on the computer output, with the t value, degrees of freedom and exact significance given. If the test is a two tailed test, the exact significance is just as it appears on the output.

**9.8.2 TWOSAMPLE procedure of MINITAB**

Another popular statistical program is MINITAB. The way in which the data is entered into MINITAB, and the way in which the programming in this statistical package is conducted, is not discussed here. You would need to consult a MINITAB manual to determine these. What this section does provide is an example of the results from the procedure called TWOSAMPLE in MINITAB.

**Example 9.8.2 Test for the Difference in Mean Level of Opinion using MINITAB**

*The data used for this example are the opinions of PCs and Liberals about trade unions being responsible for unemployment. These data were examined in Example 9.6.4 and 9.8.1. In the MINITAB program the PCs were placed in column 1, and are the group labelled C1 here. The Liberals are the group labelled C2. The hypotheses for the t test were*

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

$$H_1 : \mu_1 - \mu_2 > 0$$

The null hypothesis is the hypothesis of no difference in the mean opinion level of PCs and Liberals. The alternative hypothesis is that PCs more strongly agree that trade unions are responsible for unemployment than are Liberals. This produces the alternative hypothesis that  $\mu_1 > \mu_2$ . The MINITAB program that produces these results, along with the results are as follows.

```
MTB > twosample c1 c2;
SUBC> alternative = 1.
```

```
TWOSAMPLE T FOR C1 VS C2
```

	N	MEAN	STDEV	SE MEAN
C1	40	5.57	1.20	0.19
C2	9	3.67	2.29	0.76

```
95 PCT CI FOR MU C1 - MU C2: (0.13, 3.69)
```

```
TTEST MU C1 = MU C2 (VS GT): T= 2.43 P=0.019 DF= 9
```

If you have not used MINITAB before, some of this may appear rather strange. But the essential results for the test of means is given on the last line. There it states that the test is a test for the equality of the means of the data that has been placed in columns C1 (PCs) and C2 (Liberals). The alternative hypothesis is that the mean of C1 is greater than the mean of C2, that is (VS GT). The  $t$  value for the test is 2.43, consistent with the value in Example 9.6.4 and 9.8.1. The exact significance level of this  $t$  value, for a one tailed test with 9 degrees of freedom, is 0.019. This is a low probability, meaning that the null hypothesis can be rejected at any significance level as low as 0.02. Note that the means, standard deviations, standard errors of the mean and the sample sizes are also provided as summary information.

The above test is a one tailed test, testing whether group 1 is greater than group 2. It is possible to conduct the test in various other ways with MINITAB. There are two-tailed tests and one-tailed tests in the negative direction. All of these tests are based on the separate variance  $t$  test. In MINITAB there is also a method of obtaining the pooled variance  $t$  test. These are not provided here, but are easily carried out with MINITAB.

### 9.8.3 Summary of Computer Programs

When working with these statistical programs, the computer carries out all the calculations for you. What you need to know when you work with these programs is how to set up the null and alternative hypotheses, whether to use a one or two tailed test, whether to use the separate or pooled variance *t* tests, and how to interpret the results.

## 9.9 Reporting Results of Hypothesis Test

There are three common ways in which the results of hypothesis tests are reported in research work. Which of these you should use is partly a matter of personal preference. The first two methods select the significance level ahead of time, and report whether the result is statistically significant at that level or not. These are the most commonly used approaches when reporting results in journals and reports. The third method reports the exact significance level associated with the test statistic. This is less common, but has become more widely used in recent years. Each of these procedures is outlined in the following paragraphs. Some of the examples used earlier in this chapter will be used to illustrate the manner in which hypothesis tests are reported.

**1. The Traditional Method.** The method most commonly described in textbooks is to select a particular level of significance and use this throughout the set of hypothesis tests that are to be conducted. The researcher then reports whether the data leads to rejection of the null hypothesis at this significance level or whether there is insufficient evidence to reject the null hypothesis. This is straightforward and easy to understand. The weakness of this approach is that it does not distinguish between tests which are more significant statistically and those which are less significant statistically.

An example of this approach to reporting results of hypotheses tests is the analysis of the explanations of unemployment provided by H. Krahn, G. S. Lowe, T. F. Hartnagel and J. Tanner in "Explanations of Unemployment in Canada," **International Journal of Comparative Sociology**, XXVIII, 3-4, 1987, pages 228-236. Parts of Table I of this article are reproduced here in Table 9.20.

Table 9.20 provides a summary of the responses of those Edmonton respondents who were employed, and of those who were unemployed, concerning two explanations of unemployment. For the PCs and Liberals, the first

Explanation of Unemployment	Mean Score: Labour Force Members Only	
	Employed	Unemployed
Unemployment is high because trade unions have priced their members out of a job	5.00	4.72
Unemployment is high because of the introduction of widespread automation	4.44	3.11*
Sample Size	265	36

\* Differences between means for the employed and the unemployed are statistically significant ( $p < 0.05$ ).

Table 9.20: Report Concerning Explanations of Unemployment

explanation has been examined before in Example 9.6.4. Here the difference in the mean score concerning this explanation of unemployment is given for two other groups, the employed and the unemployed. Another explanation is given in the next row, and the mean opinions for the same two groups are given. Remember that the opinion variable is measured on a 7 point scale with 1 being strongly disagree and 7 being strongly agree. The means reported in the table are means of variables measured on this scale.

Note that in Table 9.20 no standard deviations or  $t$  or  $Z$  values have been given. But the results of the hypothesis test for a difference between two means is provided for each of the two explanations. In each case the authors have selected the 0.05 level of significance as the level that is used for the test. This level is reported at the bottom of the table. An asterik \* denotes those differences between the means of the two groups which are significant at the 0.05 level. At the 0.05 level of significance, for the trade union question, the difference of the means is not statistically significant. For the automation question, there is a significant difference in the means. The authors go on to list the means for the other variables in the study, but these are not reproduced here.



As can be seen, this method provides a single significance level, and some means of identifying those results which are statistically significant at this level, and those which are not significant at this level.

## 2. Reporting Significance at Several Levels.

The second method is an extension of the first method, with several levels of significance reported, rather than just one. Table 9.21 gives various differences of means for variables in the Teen Mothers Project described in Appendix D. This table is extracted from the report that the researchers in Saskatchewan Social Services made concerning the project.

	Project Group		Control Group		Significance
	Score	<i>n</i>	Score	<i>n</i>	
Bayley Score	109.2	28	98.8	13	**
Initial Knowledge	8.0	28	7.8	24	
Final Knowledge	10.9	28	9.4	13	*
Change in Knowledge	3.0	28#	2.2	13	

\* Significant at the 0.01 level.

\*\* Significant at the 0.001 level.

# Significant at the 0.0001 level.

Table 9.21: Difference of Scores, Teen Mothers Project

Note that in Table 9.21 three different significance levels have been used. Those differences which are significant at the 0.0001 level are reported with three asteriks. Other differences are reported as being significant at the 0.001 and 0.01 level of significance. Where the difference is not significant at any of these levels, there are no asteriks. For the first three rows in the table, the test is the t test for a difference of two means with independent random samples. This is the test described in Section 9.6. Table 9.21 shows that the Bayley scores are significantly different at the 0.001 level, and the difference in final knowledge levels is also significant, but only at the 0.01 level. There is no significant difference in the initial knowledge level.

The last line represents the results from the paired t tests to be examine in Section 9.10. Note that the differences of 3.0 and 2.2 for the change in

knowledge are those which are used in Example 9.10.1, and the results in that example are consistent with what is reported in Table 9.21.

This table does not contain the standard deviations, so that the test results could not be computed from the table itself. But by reporting the three significance levels, the authors allow readers to see that some of the results are more statistically significant, or stonger results, than are other differences. For example, the difference in Bayley scores is a more significant difference than the difference in final knowledge scores, while there is little or no difference in initial knowledge scores.

### 3. Reporting Exact Significance Levels.

The only disadvantage of the second method is that the reader is still restricted to only the significance levels given in the table. If the exact significance or probability associated with each test is given, then the reader can see how these differ, and make a decision concerning which results are statistically significant.

The article conerning elderly women which was used in Example 9.6.3 to examine differences in self-esteem of community and nursing home women contains exact significance levels. A few of the differences and the exact significance associated with these differences is given in Table 9.22. These results are extracted from Table 1 of the article.

		Community	Nursing	t	p
		Group	Home		
			Group		
Age (years)	M	68.9	79.8	5.72	0.000
	SD	3.7	9.2		
No. of health Problems	M	2.5	3.6	2.43	0.019
	SD	1.6	1.9		
Index of Self-Esteem	M	29.4	25.1	1.03	0.310
	SD	15.0	14.4		

Table 9.22: Exact Significance of Differences for Elderly Women

In Table 9.22 the mean (M) and standard deviation (SD) for each of

three variables for each of the two groups is given. The sample sizes are not given in the table, although they are reported elsewhere in the article. Using these means and standard deviations, along with the reported sample sizes, the  $t$  value for the difference between the means of the two groups can be computed. These  $t$  values, and their associated probabilities are given in the table. The probabilities appear to be two tailed probabilities. For example, for the number of health problems,  $t = 2.43$  and the sum of the areas to the right of  $+2.43$  and the area to the left of  $-2.43$  is 0.019. Since this is a  $t$  test, it is not possible to obtain this exact significance from the  $t$  table in this textbook. It is likely that these results come from a computer program such as SPSS which provides both the  $t$  values and the exact significance levels.

Note that the significance of the age difference is 0.000. This is not really zero, but just a very small probability, so small that rounded to 3 decimal places it is 0.000. In this case it might be preferable to report the result as  $< 0.001$ , rather than 0.000. The probability can never be exactly 0, but it is so low that it is less than 0.001. Finally, note that the difference in the index of self-esteem for the two groups is essentially the same as the difference computed in Example 9.6.3. There the difference was determined to be 1.06, rather than the 1.03 reported in Table 9.22. The difference is due to rounding. Note that the null hypothesis of no difference in the levels of self-esteem was not rejected at the 0.10 level of significance in Example 9.6.3. This is consistent with the probability of 0.310 given in Table 9.22. This probability is fairly large, and most researchers would not reject a null hypothesis at the 0.310 level.

### 9.9.1 Summary

Any of the three methods can be used to report the results of hypothesis tests. Where reports are being made for readers who are not too familiar with statistics, the first or second method might be best to use. In academic journals, or in reports that are aimed at readers with greater statistical knowledge, giving the exact significance level might be the preferable method. Reporting means, standard deviations and sample sizes, along with the results of the tests, also allows the reader to develop a better understanding of the data.

## 9.10 Paired t Test

The  $t$  tests examined so far have tested whether the true mean of two different populations or groups can be considered to be equal to or different than each other. In these tests, the variable being examined is the same for each of the two groups. The samples from these two populations must be independently selected random samples in order for these tests to be meaningful.

There are other types of sampling where two independent samples may either be inappropriate or may not be possible to obtain. In addition, researchers sometimes study a population and wish to compare the means of two different variables for this population. In either of these situations, there are two variables rather than two populations with one variable. These latter tests are referred to as **paired  $t$  tests**. Two of the types of situation where this test might be used are described in the following paragraphs.

One situation that calls for this type of test might be a standardized test administered to students at the beginning of a school year, and the same standardized test administered again at the end of the year. The test scores of the students at the two different times provide a measure of the change in performance level of the students over the course of a school year. If the students are randomly divided into two different groups, with a different teaching method used for each group, then the method used for the group that shows the greatest increase in mean test score over the school year could be considered to be the superior method.

Another situation which may occur is where individuals in the study are matched with each other. For example, a random sample of husbands could be taken, and some characteristic of the husbands obtained. Then the same characteristic for the wives of these husbands could also be obtained. Although the people are not the same in these two groups, the husbands and wives are not selected independently. The difference in the values of the characteristic for husbands and wives provides a measure of how these husbands and wives differ, but the independent sample  $t$  tests used so far may provide misleading results concerning the significance of the size of the difference. The paired  $t$  test provides a solution to this.

The paired  $t$  test examines the differences of the values of the variable for the single respondent or between the two respondents. Usually the test begins with the hypothesis of no difference, or equivalently, that the mean of the difference is zero. The mean and standard deviation of the differences are computed, and the paired  $t$  test is constructed in much the same manner

as the  $t$  test for a single mean. This procedure is now outlined in more detail, with examples also given.

**Conducting the Test.** Suppose that a random sample of  $n$  members of a population is selected and that these  $n$  cases are matched with another  $n$  cases. The manner in which this matching occurs is a more general methodological problem, rather than a strictly statistical problem. Some examples of matching might be husbands with wives, or parents with children, or matching procedures which attempt to find people with similar characteristics. Alternatively, there may be only one population, and one sample, but the characteristics of these individuals may be measured at two points in time, or using two different variables.

Regardless of which method is used, let the measurements be made on a variable  $X$ . In order to denote that there are two sets of measurements of  $X$ , let  $X_1$  be one measurement on  $X$  and let  $X_2$  be the other measurement on  $X$ . In order to denote that there are  $n$  measurements on each of these, a second subscript is necessary. Let the set of  $n$  measurements of  $X_1$  be  $X_{11}, X_{12}, X_{13}, \dots, X_{1n}$  and the  $n$  measurements of  $X_2$  be  $X_{21}, X_{22}, X_{23}, \dots, X_{2n}$ . This can alternatively be stated as a set of measurement  $X_{1i}$  and  $X_{2i}$ , where  $i = 1, 2, 3, \dots, n$ .

Since the question to be examined here is whether there is a difference in these measurements, let the difference be defined as

$$d_i = X_{1i} - X_{2i}.$$

That is,  $d_i$  represents the difference between the two measurements on variable  $X$  for each of the  $n$  matched observations. This means that  $d_1 = X_{11} - X_{21}$ ,  $d_2 = X_{12} - X_{22}$  and so on until  $d_n = X_{1n} - X_{2n}$ .

Hypotheses concerning this difference generally are made concerning the mean of the differences. Let

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

The variable  $\bar{d}$  is a statistic, representing sample differences between two values of the variable  $X$ . Let the corresponding parameter for  $\bar{d}$  be  $\mu_d$ , the true mean of the differences if all members of the population were sampled, and if appropriate pairs could be obtained for each of these. If the variable  $d$  has a normal distribution, then it can be shown that  $\bar{d}$ , the sample mean of the differences  $d$  has a  $t$  distribution with mean  $\mu_d$ , standard deviation  $s_d/n$  and  $n - 1$  degrees of freedom. The standard deviation  $s_d$  is the standard

deviation of the  $n$  observed values  $d_i$ . In symbolic terms,

$$\bar{d} \text{ is } t_{n-1} \left( \mu_d, \frac{s_d}{\sqrt{n}} \right).$$

This distribution can then be used to construct a t test for the mean of the differences in the same manner as the t test for a single mean. If the sample size is large, the t distribution approaches the normal distribution, and the  $Z$  values from the normal distribution can be used to obtain the critical value and critical region for the test. If the sample size is small, then the t table must be used to determine these.

Most often the null hypothesis in this test is that there is no difference between the two means. That is, if

$$H_0 : \mu_d = 0$$

is the assumption being tested, this means no change in the values of the characteristic  $X$ . The alternative hypothesis can be a two directional test, or a one directional test in either the positive or negative direction. Two examples of this test for the difference of means in the case of paired observations now follow.

### Example 9.10.1 Change in Knowledge of Teen Parents

*Appendix D contains data from the Teen Parent Pilot Project Data Set. Teen parents who were part of the project and those who were not part of the project group were tested for their knowledge concerning infant development at the beginning of the project, and again six months later. Each unmarried teen mother was given 24 question child development questionnaire. The total number of correct responses is the knowledge score. The scores for the 13 teenage mothers at the start of the project and 6 months later are contained in Table 9.23. The differences  $d_i$  and the means and standard deviations of these differences are also given in the table. At the 0.05 level of significance, test whether there has been any change in the knowledge score of these young women.*

*For the 28 teen mothers who participated in the pilot project the mean of the change in knowledge scores was 3.000, and the standard deviation was 2.388. At the 0.01 level of significance, test whether there was any increase in the mean of the change in the knowledge score for these teen mothers.*

**Solution.** *Let the variable  $X$  represent the knowledge score, with  $X_1$  the score of each teen mother when the program started and  $X_2$  the score of each*

Subject Number	Knowledge Score		Change in Score ( $d_i$ )
	Beginning $X_{1i}$	After 6 Months $X_{2i}$	
29	8	7	-1
30	9	7	-2
31	7	10	3
32	5	11	6
33	5	11	6
34	4	13	9
35	4	9	5
36	9	11	2
37	7	9	2
38	6	6	0
39	10	13	3
40	8	9	1
41	12	6	-6
Mean			2.154
SD			3.934

Table 9.23: Change in Knowledge Scores for 13 Teen Mothers

teen mother 6 months later. The values in column 2 of Table 9.23 represent the 13 observations on  $X_1$  and column 3 gives the scores  $X_2$ . Column 4 contains the change in score for each of the teen mothers. These differences are calculated as  $d_i = X_2 - X_1$ . A positive score implies an increase in knowledge, and a negative score means a decline in the level of knowledge among these mothers.

Since the question asks whether there has been any change in score, the alternative hypothesis is a two direction hypothesis. The null hypothesis is that there is no change in the mean of the difference in test scores. These hypotheses are

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

The test statistic is the  $\bar{d}$ , the mean of the differences in the sample. The

sampling distribution of this statistic is

$$\bar{d} \text{ is } t_{n-1} \left( \mu_d, \frac{s_d}{n} \right).$$

Since there are  $n = 13$  cases in the sample, there are 12 degrees of freedom for this test. From Appendix I, the  $t$  value for a  $t$  distribution with 13 degrees of freedom, and a two tailed test at the 0.05 level of significance is 2.160. The region of rejection is all  $t$  values of less than -2.160 or greater than +2.160.

From Table 9.23, the mean of the differences in knowledge scores is  $\bar{d} = 2.154$  and the standard deviation of the differences is  $s_d = 3.934$ . The standard deviation of the sampling distribution of  $\bar{d}$  is

$$\frac{s_d}{\sqrt{n}} = \frac{3.934}{\sqrt{13}} = \frac{3.934}{3.606} = 1.091.$$

The  $t$  value is the statistic minus its mean, and divided by its standard deviation. From this data, the  $t$  value is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{2.154}{1.091} = 1.974.$$

This  $t$  value is greater than -2.160 and less than +2.160, so that it is not in the region of rejection of  $H_0$ . While the sample mean for this sample of 13 teen mothers does show an increase in knowledge over the 6 months, there is not enough of an increase to reject the hypothesis of no change in the level of knowledge of teen mothers who are not part of a teen mothers project.

The second part of the question asks whether the larger sample of mothers who were part of the teen mothers project demonstrated enough of an increase in knowledge to show that all such teen mothers would **increase** their knowledge if they participated in such a program. The null and research hypotheses are

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d > 0$$

The test statistic is the  $\bar{d}$ , the mean of the differences in the sample. The sampling distribution of this statistic is

$$\bar{d} \text{ is } t_{n-1} \left( \mu_d, \frac{s_d}{n} \right).$$



Since there are  $n = 28$  cases in the sample, there are 27 degrees of freedom for this test. From Appendix I, the  $t$  value for a  $t$  distribution with 27 degrees of freedom, and a one tailed test at the 0.01 level of significance is 2.473. The region of rejection is all  $t$  values of greater than +2.473.

The mean of the differences in knowledge scores is  $\bar{d} = 3.000$  and the standard deviation of the differences is  $s_d = 2.388$ . The standard deviation of the sampling distribution of  $\bar{d}$  is

$$\frac{s_d}{\sqrt{n}} = \frac{2.388}{\sqrt{28}} = \frac{2.388}{5.292} = 0.451.$$

The  $t$  value is the statistic minus its mean, and divided by its standard deviation. From this data, the  $t$  value is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{3.000}{0.451} = 6.648 > 2.473.$$

This  $t$  value is much greater than +2.473 and is well into the region of rejection of  $H_0$ . This sample of 28 teen mothers who participated in the project provides strong evidence for an increase in knowledge over the 6 months. The  $t$  value of 6.648 is extremely large, and the probability of obtaining a  $t$  value this large, if the null hypothesis is true, is very small. The test allows the researcher to very strongly reject  $H_0$  and be quite certain that the alternative hypothesis of an increase in knowledge is the correct conclusion.

**Additional Comments.** These tests provide support for the argument that a teen mothers project could increase the parenting knowledge of these mothers. That is, the control group of 13 mothers who did not participate in the project showed no significant change in the mean difference in knowledge over the 6 month period. But those who did participate had an increase in knowledge that was very significant statistically.

One weakness of the test is that the sample of teen mothers who did not participate in the project was rather small. These mothers also showed an increase in their level of knowledge, but the increase was less than for those who participated. The difference of the increase for the two groups was  $3 - 2.2 = 0.8$  points. A test for the difference of these two differences might be conducted. This is left as an exercise for the reader. If you use a  $t$  test for the difference between these two means, you should be able to prove at the 0.05 level of significance that the project group shows a larger increase in knowledge than does the non project group.

Based on these test, the researcher cannot be absolutely sure that the greater increase in knowledge for the project group was a result of their participation in the project. However, the evidence does point strongly in that direction, and the project appears to provide promising results.

### Example 9.10.2 Difference in Age of Husbands and Wives

Demographic studies have shown that the average age of first marriage for men in Canada is about two years higher than it is for women. This means that husbands are usually older than their wives, and this is a well known social pattern. In this question a paired  $t$  test is used to see whether this situation is true for a sample of 15 Saskatchewan farm families. The sample of 15 husbands and wives is Table 9.24 is a random sample of husband-wife families from a study of 304 farm families living in four Saskatchewan communities. The study is very briefly described in P. Diaz and P. Gingrich, "Crisis and Community in Rural Saskatchewan," in D. A. Hay and G. S. Basran, **Rural Sociology in Canada** (Toronto, Oxford University Press, 1992), pages 36-50.

Use the paired  $t$  test to determine whether the mean age of husbands exceeds the mean age of wives for all Saskatchewan farm husband-wife families. Use the 0.01 level of significance.

**Solution.** Let  $X$  represent age, with  $X_1$  being the age of the husband in column 1 of Table 9.24 and  $X_2$  the age of the wife in column 2. Column 3 contains the difference in age  $d_i$  for each of the 15 families. Note that in none of the 15 families does the wife's age exceed that of the husband. On average, it can be seen that the mean age of the husband exceeds the mean age of the wife by just over 3 years.

Since the question is whether the age of the husband exceeds the age of the wife, the alternative hypothesis is a one directional test to determined whether the mean of  $d_i = X_1 - X_2$  is greater than zero. The null hypothesis is that there is no difference between the age of husbands and wives. These hypotheses are

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d > 0$$

The test statistic is  $\bar{d}$ , the mean of the differences in the sample. The sampling distribution of this statistic is

$$\bar{d} \text{ is } t_{n-1} \left( \mu_d, \frac{s_d}{n} \right).$$

	Age of Husband	Wife	Difference $d_i$
	60	55	5
	41	34	7
	36	32	4
	32	25	7
	27	25	2
	36	36	0
	60	58	2
	28	28	0
	64	63	1
	53	50	3
	43	37	6
	28	25	3
	42	42	0
	53	49	4
	34	32	2
Mean	42.467	39.400	3.067
SD	12.609	12.721	2.404

Table 9.24: Difference in Age of 15 Farm Husbands and Wives

Since there are  $n = 15$  cases in the sample, there are 14 degrees of freedom for this test. From Appendix I, the  $t$  value for a  $t$  distribution with 14 degrees of freedom, and a one tailed test at the 0.01 level of significance is 2.625. The region of rejection is all  $t$  values greater than 2.625.

From Table 9.24, the mean of the differences in knowledge scores is  $\bar{d} = 3.067$  and the standard deviation of the differences is  $s_d = 2.404$ . The standard deviation of the sampling distribution of  $\bar{d}$  is

$$\frac{s_d}{\sqrt{n}} = \frac{2.404}{\sqrt{15}} = \frac{2.404}{3.873} = 0.621.$$

The  $t$  value is the statistic minus its mean, and divided by its standard

deviation. From this data, the  $t$  value is

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{3.067}{0.621} = 4.941.$$

This  $t$  value is greater than 2.625 and is well into the region of rejection of  $H_0$ . Even though this is a small sample, the difference in ages of husbands and wives in this sample provides quite strong evidence that the mean age of all Saskatchewan farm husbands exceeds the mean age of all Saskatchewan farm wives. This conclusion is made at the 0.01 level of significance.

**Additional Comments.** This conclusion depends on the assumptions of the test being satisfied. That is, the sample of 15 husbands and wives in this sample represent a random sample of all husband-wife farm families. This sample cannot claim to be a random sample of all farm families, but if the four communities surveyed in the study are together fairly representative of all Saskatchewan farm areas, the sample may come close to representing a cross section of all such families.

The assumption of normality for the differences is also subject to question. While this assumption is not likely to be exactly satisfied, it may not be all that far from being satisfied either. Whatever the true mean difference in age of husbands and wives, some families will have an age difference less than this mean, and others will have a larger age difference. It might be expected that there would be a reasonably symmetric distribution of age differences around the true mean difference. This distribution will not be exactly normal, but it may be relatively close.

In summary, these assumptions may be close to being satisfied, the null hypothesis is very decisively rejected, and these results are consistent with age differences for other types of families. As a result, this test provides reasonably good evidence that in Saskatchewan farm husband-wife families, the age difference between husbands and wives is a positive one.

## 9.11 Conclusion

This chapter has examined a considerable variety of hypothesis tests. A test for a single mean with a large sample size was the first test, and Section 9.2 also provided an overview of the various aspects involved in hypothesis testing. Tests of a single mean for a small sample size, a single proportion, two proportions, and two means followed. Each of these used the same basic method for conducting tests of hypotheses, with differences in the notation and formulas for the  $t$  or  $Z$  values. The tests conducted in this chapter are widely used, and account for a large proportion of all hypothesis tests that are conducted.

In Chapter 10, more hypothesis tests are presented. While these will appear somewhat different in terms of the distributions used and the formulas, the principles introduced in this chapter are used there as well. Again in Chapter 11, tests of hypotheses for correlation and regression are discussed, again using the same principles as introduced here. As you become familiar with the tests and the principles of testing outlined in this chapter, you will find that these will allow you to understand the various types of tests used in journals and reports. In addition, these are the same tests and methods which you will be able to use in your own research work.