Chapter Three

THE USE OF THE T-TEST IN LANGUAGE STUDIES*

To test the validity of a hypothesis or inference, researchers need to go beyond simply describing different groups of data. For example, while the calculation of class averages on a test for two different groups of students may describe their respective performance, it provides little insight into the significance of the differences between the two groups. Describing group averages without comparing the differences between them is much like going through the labors of baking a cake and leaving it to sit uneaten. The product of the labor may be a success, yet no one knows for sure since it is left untasted. Alternatively, suppose that two groups of students have respective group averages of 81% and 78%. This may at first glance appear to be an insignificant difference, but a further analysis of the data with an inferential statistical test may reveal that there is a statistically significant difference in the performance of the two groups.

The t-test is a statistical test that allows us to determine to what extent the difference between two arithmetic means is statistically significant or simply a chance occurrence. Unlike descriptive statistical measures (e.g. mean) used to describe a group of data, the t-test is an inferential statistical test. Its results permit a researcher to judge the validity of a hypothesis or inference. The t-test is a relatively simple statistical test, and all the researcher needs to determine whether the difference between two group averages is statistically signifi-

* Marc Weinstein, University of Gdańsk.

cant is a pencil, an eraser, and (preferably) a hand-held calculator.

The t-test is a parametric statistical test, and as such three general conditions must be met before it can be used to analyze data. First, the scores to be analyzed should have a relatively normal distribution. In other words, the more symmetrical the distribution of scores are around the mean, the more valid are the results of the t-test. Secondly, the scores to be compared should have roughly equal variances. If the dispersion of the scores of the different samples are greatly different, the validity of a parametric test (and the t-test is one of these) is called into question. Finally, the data to be analyzed must be measured on an interval scale. On an interval scale, the difference between all points on the scale are equal. The centrigade scale, for example, is interval since the difference between 7 and 9 is the same as the difference between 26 and 28.

If any one of the above three conditions are not met, a comparable non-parametric test should be used. Since the t-test is a fairly robust test to compare two group arithmetic means, the distribution of scores does not have to be perfectly normal and the group variances do not have to be absolutely homogeneous. Yet if the distribution of scores is very skewed, the group variances are widely different, or ordinal rather than interval data are being analyzed, the Mann-Whitney U-Test should be used to compare the group averages (see Chapter Four of this volume).

Once researchers have determined that a parametric test can be used, a decision still has to be made as to which parametric test is most appropriate for the analysis of data. Generally, a t-test can be used when there is one nominal independent variable and one interval dependent variable. When there is more than one nominal variable being examined (e.g. two different treatments and one control group) the analysis of variance (ANOVA) statistical test should be used instead of the t-test. And when it is necessary to analyze more than one dependent variable simultaneously, far more complicated and sophisticated statistical tests requiring the aid of a computer (such as factor analysis or discriminant function analysis), need to be employed.

Most importantly, it should be remembered that the t-test is designed for comparing only two sets of data. The t-test cannot be used to compare Groups 1 and 2, 1 and 3, and then 2 and 3, etc. If such cross comparisons are made, the multiple mathematical calculations artificially increase the size of the sample group. Therefore, by making cross comparisons between more than two groups of data, the likelihood of rejecting the null hypothesis becomes easier and in effect the probability level originally set by the researcher increases.

Once the researcher determines that the t-test is the appropriate statistical test to analyze two arithmetical averages, a final decision has to be made as to which t-test to use. If two groups of subjects (S´s) have been randomly selected, then the formula for the independent t-test should be used (the reader will find a 'T-test Worksheet for Independent Samples' below). If, on the other hand, matching with random assignment is used to compose groups or S´s are used as their own control as in a pretest posttest design, the formula provided for the matched t-test should be used (see the separate 'T-test Worksheet for Matched Pairs).

Before the actual statistical analysis is performed, the researchers have to formulate a null hypothesis. It should be recalled that a hypothesis is a very general expectation about the relationship between variables. As a result, there is an extremely large number of instances under which a hypothesis needs to be tested in order to support it. Alternatively, researchers formulate a null hypothesis which in essence is a negative or 'no difference' hypothesis. Once the null hypothesis is formulated, researchers set out to disprove it at the predetermined significance level. By rejecting a null hypothesis, researchers accept that minor differences can occur due to chance variation and thus are not real differences.

Finally, researchers have to determine the probability level at which they will consider their results to be statistically significant. In the behavioral sciences, the five percent level is usually considered an acceptable benchmark. This is usually represented as $p \leq .05$ (read as p less than or equal to .05). At this level, the researchers are accepting that there is a five percent chance that the distribution of scores is a function of chance.

EXERCISE

A group of researchers would like to test whether providing background knowledge to students before a reading task raises their reading comprehension. In other words, the researchers hypothesize that:

> There will be a statistically significant difference on the performance of a reading comprehension test between students who receive background information on a reading passage and those who do not.

Thus, a group of forty second-year English philology students (the subjects of the study) are randomly divided into two groups. A reading comprehension test is piloted and its validity and reliability are determined to be satisfactorily high. The researchers have decided to consider their results statistically significant at the $p \leq .05$ level. At this level, the researchers are accepting that there is a five percent chance that the distribution of scores is a function of chance. The experimental group, which received the background information, is arbitrarily designated group A, and the control group, which received no background information, is designated group B. After the administration of the test, the following scores were obtained:

Group A: 75, 88, 80, 85, 78, 90, 82, 88, 76, 82, 81, 89, 77, 84, 88, 93, 91, 81, 85, 87.

Group B: 72, 81, 70, 80, 70, 85, 73, 82, 72, 78, 79, 83, 73, 85, 82, 90, 80, 78, 81, 74.

Determine whether the results of study are statistically significant by: a) formulating a null hypothesis; b) determining whether a matched t-test or independent t-test should be used; c) calculating the t-value using the appropriate formula; d) determining whether a one-tailed or two-tailed test should be used; and e) looking up the t-value in the abridged table 'Critical Values of t' found near the end of this chapter.

Before beginning the exercise, the reader is cautioned about the algebraic terms X^2 and $(X)^2$. The term X^2 means to first square each score and then add the squares. The term $(X)^2$ includes parentheses and means first to add all the scores together, then square the sum.

Also, below is a short glossary of abbreviations used in the worksheets:

For the independent t-test:

N = number of subjects

X = mean = $\frac{X}{N}$

s^2 = variance

For the matched pairs t-test:

N_p = number of pairs

D = difference between pairs

S_d = standard error of differences

$S_{\bar{d}}$ = standard deviation of differences

Numerical subscripts specify groups. For example, N_1 specifies the number of subjects in group 1.

SOLUTION

The following is the null hypothesis: There will not be a statistically significant difference on the performance of a reading comprehension test between students who receive background information on a reading passage and those who do not.

An independent t-test should be used since the S´s are randomly divided. The S´s are not their own control nor are they match paired.

See *T-test Problem Solution* at the end of this chapter.

A two-tailed test should be used. The hypothesis is non-directional. The researchers are stating that there will be a difference on the performance of the reading comprehension test, but they are not speculating whether the experimental group or the control group will perform better on the reading comprehension test. A non-directional hypothesis requires the use of a two-tailed test. A directional hypothesis, which states one group will do better than the other, would require the use of a one-tailed test.

The difference between the averages is statistically significant at the $p \leq .01$ level. The t value in step six of the independent t-test worksheet is 3.27. To determine the p level at which the null hypothesis can be rejected, we need to go to the far left column of the table 'Critical Value of t' (below) and find the appropriate degree of freedom (df) figure. In our example the df is 38. Thirty-eight is not on the table, so we go to the next *lower* number, since we prefer to err on the side of caution and, if anything, make it more difficult to prove statistical significance. (The smaller the number of S's, the smaller our degree of freedom will be, and the more difficult it will be to prove statistical significance). In this case, the next lower df will be 30.

At this point, we read across the df = 30 row until we find the largest t-value which our observed t-value exceeds. In our example, the observed t-value is 3,27 which exceeds 2.750, but is less than 3.646. Therefore, we will want to look up the level of significance at the 2.750 level. In part d of the problem we decided that this is a two-tailed test, so we read up from the 2.750 column to row b, and we see that the level of significance is less than or equal to .01. Therefore, we can conclude that there is a one percent chance or less that our ability to reject the null hypothesis was a chance occurrence.

T-TEST WORKSHEET FOR INDEPENDENT SAMPLES

GROUP	1	2
N =		
ΣX =		
ΣX^2 =		
$\bar{X}$		

1. Calculation of group variances

$$s^2 = \frac{N_1\ \Sigma X^2 - (\Sigma X_1)^2}{N_1\ (N_1 - 1)} = \_\_\_\_\_\_\_\_ \qquad s_1^2 = \_\_\_\_\_\_\_\_$$

$$s^2 = \frac{N_2\ \Sigma X^2 - (\Sigma X_2)^2}{N_2\ (N_2 - 1)} = \_\_\_\_\_\_\_\_ \qquad s_2^2 = \_\_\_\_\_\_\_\_$$

2. Calculation of t-value.

Steps

1. $$\frac{(N_1 - 1)s_1^{\ 2} + (N_2 - 1)s_2^{\ 2}}{N_1 + N_2 - 2} = \_\_\_\_\_\_\_\_$$

2. $\frac{N_1 + N_2}{N_1 N_2}$ = ____________

3. (Step 1 X Step 2) = ____________

4. $\sqrt{\text{Step 3}}$ = ____________

5. $\bar{X}_1 - \bar{X}_2$ = ____________

6. $t = \frac{\text{Step 5}}{\text{Step 4}}$ = ____________ $df = N_1 + N_2 - 2$ =

7. Look up t value in table, p < ____________

(Worksheet adapted from Tuckman, Bruce *Conducting Educational Research*, Harcourt, Brace, Jovanovich, San Diego 1978).

T-TEST WORKSHEET FOR MATCHED PAIRS

N_p = ____________

ΣD = ____________

(ΣD^2) = ____________

$\bar{X}_1$ = ____________

$\bar{X}_2$ = ____________

Calculation of standard deviation of differences

1. $Sd = \sqrt{\frac{\Sigma(D^2) - (1/N_p)(\Sigma D)^2}{N_p - 1}}$ = ____________

Calculation of standard error of differences

2. $S\bar{d} = \frac{\text{Step 1}}{\sqrt{N_p}}$

Calculation of t

3. $t = \frac{\bar{X}_1 - \bar{X}_2}{\text{Step 2}}$ = ____________, $df = N_p - 1$

4. Look up t-value in table, $p \leq$ ____________

The unmanipulated formula for Matched pair t-test is:

$$\frac{\bar{X}_1 - \bar{X}_2}{S\bar{d}}$$

CRITICAL VALUES OF T

	Level of Significance for one-tailed test				
a	.05	.025	.01	.005	.0005
df	Level of Significance for two-tailed test				
b	.10	.05	.02	.01	.001
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.767
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.690
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.659
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460

T-TEST PROBLEM SOLUTION - PART C

GROUP	1	2
N =	20	20

$X =$	1,680	1,568
$X^2 =$	141,642	123,520
$\overline{X}$	84	78,4

1. Calculation of group variances

$$s^2 = \frac{(20)\ (141,642) - (1680)^2}{(20)\ (20 - 1)} = \frac{10,440}{380} \quad s^2 = \underline{27.47}$$

$$s^2 = \frac{(20)\ (123,520) - (1568)^2}{(20)\ (20 - 1)} = \frac{11,776}{380} \quad s^2 = \underline{30.99}$$

2. Calculation of t-value.
   Steps

   1. $\frac{(20 - 1)\ (27.47) + (20 - 1)\ (30.99)}{20 + 20 - 2} = \underline{29.23}$

   2. $\frac{20 + 20}{(20)\ (20)} = \underline{.1}$

   3. $(29.23)\ (.1) = \underline{2.923}$

   4. $2.923 = 1.7096$

   5. $\overline{X}_1 - \overline{X}_2 = \underline{5.6}$

   6. $t = \frac{5.6}{1.7096} = \underline{3.27}$ $df = N_1 + N_2 - 2 = \underline{38}$

   7. Look up t-value in table, $p < \underline{.01}$

(Worksheet adapted from Tuckman, Bruce *Conducting Educational Research*, Harcourt, Brace, Jovanovich, San Diego 1978).