Computing the t*-test in data analysis*

The *t*-TEST is the most frequently used measure in second language research when comparing mean scores for two groups. Suppose you want to know whether older students versus younger students or males versus females or those with pets versus those without pets performed better when taking the same test. A *t*-test would be the measure you would use to compare the mean scores of the two groups. The *t*-test, with a bit of adjustment, can also be used to compare mean scores of just one group, say between a pre-test and a post-test, when checking to see if some skill had been acquired by the group during training.

The *t*-test is represented (hold on to your hats) with the symbol *t*. It is a very useful measure because it can be used with very large or very small groups. The adjustment for group size is made by using a table that shows different values for various group sizes. Group size is (roughly) adjusted for by *degrees of freedom*. DEGREES OF FREEDOM (*df*) for *t*-tests can be determined by subtracting one from the number of participants in each group and then adding the two resulting numbers together.

Adjustments for differences in types of decisions is made by considering one-tailed and two-tailed decisions separately. These two concepts work very much like directional and non-directional decisions for correlational analyses (see page 186). *One-tailed* decisions for *t*-tests are like directional decisions in correlational analysis in the sense that you should use them when you can reasonably expect one mean to be higher than the other; *two-tailed* decisions for *t*-tests are like non-directional decisions in correlational analysis in the sense that you should use them when you have no reason to expect one or the other of the means to be higher.

For example, let's say we are comparing the mean *remembering* scores of men versus women on word List A. In our sample class, there were 17 women and 17 men, so we would subtract 17 minus 1 (17 – 1 = 16) and 17 minus 1 (17 – 1 = 16) and add these two results together (16 + 16 = 32) to determine that we have 32 degrees of freedom. When we check our *t*-value in the appropriate table, we would check in the row which has 32 degrees of freedom to decide if the difference between the means is significant or not.

So, let's compute a *t*-value for our sample class of word list learners. Table 7.4 once again shows a set of word recall means and standard deviations for Lists A and B for both women and men.

Table 7.4 Recall means and standard deviations in Exercise 7.5 (www)

	List A *M*(*SD*)	List B *M*(*SD*)
Women N = 17	18.5(2.5)	22.5(2.0)
Men N = 17	14.5(3.5)	21.5(2.5)

The difference in recall for List A between women and men looks interesting, so we'll consider that first.

Women ($N_w = 17$)

Mean $M_w = 18.5$

Standard Deviation $S_w = 2.5$

Men ($N_m = 17$)

Mean $M_m = 14.5$

Standard Deviation $S_m = 3.5$

In this case, because the group sizes are the same, the formula for computing the *t*-value would look like this:

$$t = \frac{M_w - M_m}{\sqrt{\frac{S_w^2}{N_w} + \frac{S_m^2}{N_m}}}$$

This equation is read as *t* equals the mean for women minus the mean for men divided by the square root of the standard deviation squared for the women divided by the number of women plus the standard deviation squared for the men divided by the number of men.

Let's take this equation a bit at a time. The top half of this equation is easy. So, $M_w - M_m = 18.5 - 14.5 = 4$. The bottom half of this equation is determined by squaring the women's standard deviation ($S_w^2 = 2.5^2 = 6.25$) and dividing the result by the number of women ($N_w = 17$) and adding to that the squared value of the men's standard deviation ($S_m^2 = 3.5^2 = 12.25$) divided by the number of men ($N_m = 17$) then taking the square root of the result. Computing this out step by step:

$$t = \frac{M_w - M_m}{\sqrt{\frac{S_w^2}{N_w} + \frac{S_m^2}{N_m}}}$$

$$t = \frac{18.5 - 14.5}{\sqrt{\frac{6.25}{17} + \frac{12.25}{17}}}$$

Table 7.5 Critical values for the t-test statistic

One-tailed	*0.05*	*0.025*	*0.01*	*0.005*
Two-tailed	*0.10*	*0.05*	*0.02*	*0.01*
df				
1	6.314	12.706	31.821	63.657
2	2.920	4.303	6.965	9.925
3	2.353	3.182	4.541	5.841
4	2.132	2.776	3.747	4.604
5	2.015	2.571	3.365	4.032
6	1.943	2.447	3.143	3.707
7	1.895	2.365	2.998	3.499
8	1.860	2.306	2.896	3.355
9	1.833	2.262	2.821	3.250
10	1.812	2.228	2.764	3.169
11	1.796	2.201	2.718	3.106
12	1.782	2.179	2.681	3.055
13	1.771	2.160	2.650	3.012
14	1.761	2.145	2.624	2.977
15	1.753	2.131	2.602	2.947
16	1.746	2.120	2.583	2.921
17	1.740	2.110	2.567	2.898
18	1.734	2.101	2.552	2.878
19	1.729	2.093	2.539	2.861
20	1.725	2.086	2.528	2.845
21	1.721	2.080	2.518	2.831
22	1.717	2.074	2.508	2.819
23	1.714	2.069	2.500	2.807
24	1.711	2.064	2.492	2.797
25	1.708	2.060	2.485	2.787
26	1.706	2.056	2.479	2.779
27	1.703	2.052	2.473	2.771
28	1.701	2.048	2.467	2.763
29	1.699	2.045	2.462	2.756
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
60	1.671	2.000	2.390	2.660
120	1.658	1.980	2.358	2.617
∞	1.645	1.960	2.326	2.576

(Adapted f[illegible]

$$t = \frac{4.0}{\sqrt{.3676 + .7206}}$$

$$t = \frac{4.0}{\sqrt{1.0882}} = \frac{4.0}{1.0432} = 3.8344 \approx 3.83$$

Thus, the *t*-value in this case is about 3.83.

Next, we look in the *t*-test table (see Table 7.5) remembering that there are 32 degrees of freedom (*df*). If the exact *df* is not shown in the table, we take the closest value below it in order to be conservative. There is no row for 32 degrees of freedom in the table, so we take the closest value below 32, which is 30. In that row, the critical value for *t* at the .01 level of significance (two-tailed) is 2.750 (or 2.042 at the .05 level of significance). Since the *t*-value we calculated for the difference between women and men was 3.83 and that value is greater than the critical value found in the table at .01 or .05, we take the more conservative of those two values, which means that we have found a significant difference between women and men at $p < .01$. This indicates that the difference between women's mean and men's mean probably did not happen accidentally. There is only 1 chance in 100 (.01) that the difference in mean scores between women and men occurred by chance. Put another way, this result means that the women's mean is *significantly higher* than the men's mean score for List A recall.

An article describing this experiment might have this partial summary of findings:

> ABSTRACT (partial)
>
> An inquiry was conducted into gender differences on list recall of 25 common English words, listed randomly or in semantic categories. For the random list condition, women participants had a mean word recall of 18.5 ($SD = 2.5$) while men participants had a mean word recall of 14.5 ($SD = 3.5$). A *t*-test analysis of the differences between means yielded a *t* of 3.83. This was significant at the $p < .01$ level (with $df = 32$). It was concluded that, in this experiment, women participants were significantly better than men at remembering random lists of words.

In the above experiment, the number of men and the number of women were equal. Note that another equation should be used if the sample sizes are different for the two groups:

$$t = \frac{M_w - M_m}{\sqrt{\left[\frac{(N_w - 1)S_w^2 + (N_m - 1)S_m^2}{N_w + N_m - 2}\right]\left[\frac{1}{N_w} + \frac{1}{N_m}\right]}}$$

For instance, let's say the number of men in the example above had been nine instead of 17 and the number of women was still 17. The mean recall of words and standard deviations remain as given in the abstract. In that case, the *t*-value would have been calculated as follows:

$$t = \frac{M_w - M_m}{\sqrt{\left[\frac{(N_w - 1)S_w^2 + (N_m - 1)S_m^2}{N_w + N_m - 2}\right]\left[\frac{1}{N_w} + \frac{1}{N_m}\right]}}$$

$$t = \frac{18.5 - 14.5}{\sqrt{\left[\frac{(17-1)2.5^2 + (9-1)3.5^2}{17+9-2}\right]\left[\frac{1}{17} + \frac{1}{9}\right]}}$$

$$t = \frac{4.0}{\sqrt{\left[\frac{(16)6.25 + (8)12.25}{24}\right](.0588 + .1111)}}$$

$$t = \frac{4.0}{\sqrt{\left[\frac{100 + 98}{24}\right](.1699)}}$$

$$t = \frac{4.0}{\sqrt{\left[\frac{198}{24}\right](.1699)}} = \frac{4.0}{\sqrt{[8.25](.1699)}} = \frac{4.0}{\sqrt{1.4017}} = \frac{4.0}{1.1839} = 3.3787$$

This equation could also be used in situations where the group sizes are the same. For instance, in the original example above with 17 in each group, we found a *t*-value of 3.83 using the original formula and applying the formula for here, we get 3.84, which is within one one-hundredth of the same result.

$$t = \frac{M_w - M_m}{\sqrt{\left[\frac{(N_w - 1)S_w^2 + (N_m - 1)S_m^2}{N_w + N_m - 2}\right]\left[\frac{1}{N_w} + \frac{1}{N_m}\right]}}$$

$$t = \frac{18.5 - 14.5}{\sqrt{\left[\frac{(17-1)2.5^2 + (17-1)3.5^2}{17+17-2}\right]\left[\frac{1}{17} + \frac{1}{17}\right]}}$$

$$t = \frac{4.0}{\sqrt{\left[\frac{(16)6.25 + (16)12.25}{32}\right](.0588 + .0588)}}$$

$$t = \frac{4.0}{\sqrt{\left[\frac{100 + 196}{32}\right](.1176)}}$$

$$t = \frac{4.0}{\sqrt{\left[\frac{296}{32}\right](.1176)}} = \frac{4.0}{\sqrt{[9.25](.1176)}} = \frac{4.0}{\sqrt{1.0878}} = \frac{4.0}{1.0430} = 3.8351 \approx 3.84$$

Note that yet another equation must be used if the means being compared are for the same group of participants on two occasions as in a comparison of pre-test and post-test means. (For a full explanation and example, see Hatch and Lazaraton 1991). Notice also that there are now only 26 participants instead of 34, so a *t*-value of 3.84 may no longer be significant. You might check Table 7.5 to find out.