t-Tests – Part 1

The *t*-test is the most basic inferential statistic. It answers one of the following questions: 1. *do we have sufficient reason to conclude that the mean of this population (μ) [from where this sample was taken] is <u>not</u> equal to some value V?* or 2. *do we have sufficient reason to conclude that the difference between the means of these two populations (i.e., $\mu_1 - \mu_2$) is <u>not</u> equal to V?* Note what might appear to be an inherent bias in each of these questions. We have to have "sufficient reason" to conclude that the true mean or true difference between means is not equal to some pre-specified value; the "default" conclusion is that it is.

In technical terms, the idea that $\mu = V$ or that $\mu_1 - \mu_2 = V$ is called the **null hypothesis**. The null hypothesis is usually written as follows: $H_0: \mu = V$ or $H_0: \mu_1 - \mu_2 = V$ (where *V* can be any value, but is almost always zero in the latter case). The *t*-test is best thought of as being part of a search for evidence *against* the null hypothesis; it's a test of whether the evidence against the null hypothesis is sufficient to get rid of it. The null hypothesis wins by default.

Because a *t*-test concerns whether we should get rid of (or reject) the null hypothesis, that's all we really need to specify before we get started. And, yet, some people also like to write down what is usually called the **alternative hypothesis** (H_1), which is just the complement to the null -- e.g., if $H_0: \mu = V$, then $H_1: \mu \neq V$. I try to avoid this because it can give the impression that H_0 and H_1 are on equal footing when they definitely are not. Remember: the null hypothesis wins by default. If you can't produce evidence against it, you must retain it. Also, under the "bass-ackwards" approach of hypothesis testing, the way that the statistical test is actually conducted is by calculating the probability of the data under the assumption that the null hypothesis is true; the alternative hypothesis is not used for any of this. Finally, specifying an alternative hypothesis opens the door to having multiple alternatives, which is how you end up in the world of unjustified one-tailed tests (and you don't want to go there).

Types of *t*-Test

Depending on how deeply you dig, you could easily develop a strong defense for the idea that there is only one type of *t*-test ... or two types ... or three types ... or four types. Arguing about this would not be useful. My reason for bringing this up at all is to prepare you for some issues that you'll need to think about, as these go a bit deeper than just clicking an option in SPSS.

On the surface (as well as in the **Analyze ... Compare Means** sub-menu in SPSS), there are three kinds of *t*-test: the one-sample *t*-test, the independent-samples *t*-test, and the paired-samples *t*-test. The one-sample (or univariate) *t*-test is for when you use a single set of values to test $H_0: \mu = V$. For example, you want to know if people (in general) can perform a task without simply guessing. If it's a two-choice task, you test $H_0: \pi = .50$.

Both the independent-samples and paired-samples *t*-test are for testing $H_0: \mu_1 - \mu_2 = V$, where *V* is almost always zero (so that's all I'll discuss). Because of this, many people prefer to write the null hypothesis of a zero difference as $H_0: \mu_1 = \mu_2$, instead. I half agree. My preference is to let the way that you write the null hypothesis act as a reminder of what kind of data you have. If the values of X1 and X2 came from two different groups of subjects -- as they would if this was an

experiment using a between-subjects design (or a pseudo-experiment involving, for example, males and females) -- then I agree that we should use $H_0: \mu_1 = \mu_2$, because separating the two μs with the equal sign will act as a reminder that the two sets of values came from different sources. But if the values of X1 and X2 all came from the same subjects -- as they would if this was an experiment using a within-subjects design -- then I prefer to use $H_0: \mu_1 - \mu_2 = 0$ with both μs on the same side.   And as you'll see very soon, I have an even better reason for preferring $H_0: \mu_1 - \mu_2 = 0$ for within-subject designs.

The Univariate *t*-test

The one-sample or univariate *t*-test starts with four numbers -- an assumed value for population mean (according to the null hypothesis, μ), the sample mean ($\overline{X}$), an estimate of the spread of the sampling distribution for the mean (s$\overline{x}$ or se) and a measure of quality (df) -- plus the Assumption of Normality.   From this, you (or SPSS) can calculate the *t*-statistic and then look up the *p*-value for this particular *t*-statistic for the df that you have.   If the *p*-value is less than .05, you reject the null hypothesis.

The population mean according to the null hypothesis does not require any calculation.   It is usually set by theory.   (In the example above involving performance that is different from chance, the null-hypo value was .50.)

$\overline{X}$ is the sample mean.   End of story.

s$\overline{x}$ or se is calculated from N and the version of standard deviation, s, that uses N–1 in the denominator (so the value provided by **Analyze … Descriptives … Explore** is correct).   Here are all the related formulae in one place:

First, one calculates the best estimate of population standard deviation:

$$s^2 \;=\; \Sigma\,(\,X_i - \overline{X}\,)^2 \,/\, (\,N - 1\,)$$

$$s \;=\; \sqrt{s^2}$$

Second, one converts this into a standard error (via the "square-root of N" rule):

$$s_{\overline{x}} \;=\; s \,/\, \sqrt{N}$$

Note the important implication of this second step (which wasn't mentioned before).   That one divides by the square-root of N to convert an estimated standard deviation into a standard error tells you that larger samples have less error.   That's the core of the "square-root of N" rule. This is one of the two places that increasing N helps to give a *t*-test more power.   The other place is in the degrees of freedom (df) associated with this standard error, which is N – 1 (as will be repeated in a moment).   Because the shape of the *t* distribution becomes more compact (i.e., more like a normal) as df increases, the larger the N, the more power one has from this, as well.   Finally, please don't say (or think) that increasing N helps a *t*-test by doing something to $\overline{X}$, because it doesn't; $\overline{X}$ is an unbiased estimate of μ for all sample sizes.

Now you can calculate the *t*-statistic:

$$t = (\overline{X} - \mu_{H0}) / s\overline{x}$$

where μ_{H0} is short-hand for the value of μ according to the null hypothesis.

Combined with the value of df -- which is N–1 for a univariate *t*-test (because the error term was based on N pieces of data, minus one for the mean that was needed calculate s) -- you can then use a table or a web-based gizmo to "look up" the probability of getting a value of *t* that is at least this large. Note: if the *t* was negative, drop the sign; this is a two-sided test -- both positive and negative deviations from the μ_{H0} count as evidence against the null.

☞ Note that most journals now require that you report *p*-values to three or more decimal places, while other journals don't allow more than two decimal places when N is small.. Other journals only ask that you say "significant" or not (i.e., below .05 or not). Be ready for all of these by writing down *p* to at least three places.

Assumptions of the univariate t-test

There is only one assumption of the univariate *t*-test: the data are normal. This can and should be tested -- <u>prior</u> to the running the *t*-test, please! – using, e.g., the Shapiro-Wilk test. (Reminder: you can get this by clicking an option in the **Plots** sub-menu of the **Analyze... Descriptive Stats... Explore** command.)

How-to: Univariate t-test using SPSS

Ask for **Analyze... Compare Means... One-Sample T Test**... then highlight and "push" the data variable that contains the DV over to the **Test Variable** window. The default comparison value is zero (i.e., SPSS defaults to testing $H_0: \mu = 0$), but you can change this by changing the entry inside **Test Value** near the bottom of the sub-menu. Click **OK** to finish.

(I know that what comes might seem out of order, but there's a reason for this….)

The Paired-samples *t*-Test

The paired-samples *t*-test is actually the exact same thing as a univariate *t*-test, although many stats packages, including SPSS, have a separate command for this test. This, in my opinion, is not a good thing. I'll try to show why.

But before getting into all this, it is important to explain the name. The reason that we call it the "paired-samples *t*-test" and not the "within-subjects *t*-test" is that the values being paired are not always from the same subject. For example, if every subject in the "experimental group" is paired with a specific subject in the "control group" in terms of psychologically-meaningful subject variables (e.g., IQ, age, sex, etc.), then one can use the paired-samples *t*-test (instead of an independent-samples *t*-test) and gain some advantage in statistical power. Note, however, a danger here: if the cross-group matching procedure is insufficient, such that members of each pair

are importantly different, then the use of a paired-samples t-test can actually hurt the analysis. This could well explain why matched-group designs are rather rare these days. But done right, they're much more powerful than between-subject, independent-sample tests.

The typical null hypothesis in a paired-samples t-test is that the two population means are the same. (Note that the word "population" isn't only used to refer to people; in an experiment, the set of all values for each of the conditions is also a population, so you can get two populations from one set of people.) As mentioned earlier, a typical way to write the null hypothesis for any t-test comparing two conditions is $H_0: \mu_1 = \mu_2$. However, I would prefer that, when the data are paired, you wrote this as $H_0: \mu_1 - \mu_2 = 0$. Not only does this remind you that the data are paired, but now you can make one little change and the relationship between paired t-tests and univariate t-test should become clear:

$$H_0: \mu_1 - \mu_2 = 0 \rightarrow H_0: \mu_D = 0, \text{ where } \mu_D = \mu_1 - \mu_2$$

which is to say that, if the null hypothesis is that the difference between the means is zero, then the null hypothesis also says that the mean difference is zero. Note that this is only possible when the values are paired, such that you can get a set of difference scores from your data, as well.

The reasons that I prefer to think of paired t-tests as actually being tests of the mean difference against zero are these: 1. that's how it's actually done and 2. if you do it this way, you'll actually be able to test the Assumption of Normality (correctly).

Thus, I urge you to avoid using SPSS's **Analyze... Compare Means... Paired-Samples T Test**. Instead, use **Transform … Compute** to get the within-subject or within-pair differences and then use **Analyze... Compare Means... One-Sample T Test** to test these differences against zero. Oh, and don't forget to test the differences for deviations from normality before running the t-test.

That's nice, but why did you go off about how it's "not a good thing" that SPSS saves you from having to use **Transform … Compute** *to get the paired differences? Isn't that a good thing? Do you like wasting time?*

The reason that I say it's not good that SPSS has an option for **Analyze... Compare Means... Paired-Samples T Test** is that you won't have the correct values for the Shapiro-Wilk test. First, keep in mind that both the paired-samples t-test and univariate t-test actually operate on a single set of numbers (*viz.*, the set of differences or the single set of values, respectively). All that SPSS's paired-samples option does is save you from having to run **Transform … Compute**. But now note this: because the paired-samples test uses the difference scores, it's the difference scores that need to be normal. The two columns of original data that you'll have in a paired-samples spreadsheet can be any shape that their little hearts desire; those two shapes are totally irrelevant. Running S-W on these two columns is, therefore, inappropraite. Again, what needs to be normal are the difference scores because these are what enter the analysis. So you need to run **Transform … Compute** to get them, so you can test the assumption before running the analysis. So SPSS isn't saving you any time by doing this for you when you ask for a paired-samples t-test. Even worse, it could be seen as encouraging you to either run S-W on the wrong thing(s) or not run a test of normality at all.