Hypothesis Testing Using Two Independent Samples - The Independent-Groups t-Test

The Independent t-Test (also called the independent-measures or independent-groups or independent-samples t-test): The independent t-test is used to compare the means of two populations or of two treatment conditions. It is used when the experiment has one independent variable (IV) with two levels. Ideally, subjects are randomly assigned to the two groups (a *true experimental design*), but they may also be self-selected or self-assigned or come pre-assigned "by nature" (a *quasi-experimental design*). This *between-subjects design* is also called between groups, independent measures, or if subjects are randomly assigned a *completely randomized design*. The following assumptions must be met:

1) The samples should be selected by random sampling if inferences are to be made back to the populations. Without random sampling, random assignment is critical. Studies done without either random sampling or random assignment are of dubious value.
2) Each sample (group) consists of independent observations.
3) The two populations represented by the samples must be normally distributed. (Note: the independent t-test is *robust* to violations of this assumption provided the design is balanced and the groups are reasonably large.)
4) The two populations represented by the samples must have equal variances. This assumptions is called *homogeneity of variance*. (Note: there is a "work-around" called the *unpooled variance t-test* when this assumption fails. The version of the t-test presented here is called the *pooled* or *pooled variance t-test*.)

The Dependent t-Test (also called the correlated-groups, correlated-samples, or related-samples t-test): The dependent t-test is used when you wish to compare the performances of a single sample of subjects that has been tested twice under two different levels of an IV. This creates a *within-subjects* or *repeated-measures design*. The dependent t-test is also used when scores in the two groups are correlated due to matching, a so-called *matched-groups* or *matched-subjects design*. The dependent t-test will be covered in detail in a future handout. It makes the following assumptions:

1) The samples should be selected by random sampling, as above.
2) The observations <u>within each treatment</u> must be independent. Obviously the observations are not independent across treatments since they are paired.
3) The population of difference scores represented by the sample must be normally distributed.
4) The two treatments should create populations of scores with the same variance.

The Independent-Groups t-Test

As with all our statistical tests, the basic structure of the independent t-test is:

$$t = \frac{\text{observed difference}}{\text{difference expected due to chance}}$$

In the specific case of the independent t-test, this formula becomes:

$$t = \frac{(M_1 - M_2) - (\mu_1 - \mu_2)}{s_{(M_1 - M_2)}}$$

The numerator is the difference between means, so the denominator must be the *standard error of the difference between means*, which from here on will simply be abbreviated s.e. If all of the numerator is due to random error, then t ≈ 1 (or -1 depending upon which way you subtracted). The formula for the standard error is:

$$\text{s.e.} = s_{(M_1 - M_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$, **caution:** only when $n_1 = n_2$ (called a *balanced design*)

In the pooled variance t-test, the degrees of freedom can be obtained from the simple rule: degrees of freedom equals *total number of subjects minus the number of groups*. In a formula:

$$\text{df} = n_1 + n_2 - 2$$

Effect size: The APA now recommends that all reports of significant effects be accompanied by a measure of effect size. Remember, the fact that a difference is statistically significant does not imply anything about the size of the effect. Even very small p-values do not imply large effects. Thus, a separate measure of effect size is appropriate. One such measure is **Cohen's d**. Another measure is r^2. Both of these are described in the textbook.

$$\text{Cohen's } d = \frac{M_1 - M_2}{\sqrt{s_p^2}}, \quad r^2 = \frac{t^2}{t^2 + df}$$

Power: The power of a test refers to its ability to reject the null hypothesis when the null hypothesis is false, or in other words, the ability of the test not to make a Type II error. The power of the t-test is affected by the following factors:

1) The **alpha level:** the higher the alpha level, the greater the power.
2) The **effect size** or the **size of the difference between means**: the larger the effect, the more powerful the test.
3) The **sample size:** larger sample sizes yield greater power.
4) The **size of the standard error:** the smaller the standard error, the greater the power.
5) **One vs. two-tailed test:** one-tailed tests are more powerful than two-tailed tests.

Example 1. Marijuana smoking is generally believed to impair short term memory functioning. To test this hypothesis, one of our psychology majors conducted the following study as his senior research (Psyc 497) project. He got twenty of volunteer subjects, 10 of whom were marijuana smokers and 10 of whom were not, and gave each of them the digit span subscale of the Wechsler Adult Intelligence Scale, which measures short term memory performance (higher scores indicate better STM). The independent t-test is conducted as follows:

Step 1) State the null and alternative hypotheses.

H_0:
H_1:

Step 2) Establish a decision criterion (set alpha).

Step 3) Go out and collect the data, calculate summary statistics, and calculate the value of the test statistic.

The following data were collected by this student:

smokers:	16	20	14	21	20	18	13	15	17	18
nonsmokers:	18	22	21	17	20	17	23	20	22	21

Step 4) Make a statistical decision concerning the null hypothesis.

Step 5) Write a conclusion describing the result of the statistical analysis.

Question: Was this study a true experiment?

Unbalanced Designs: When $n_1 \neq n_2$, the design is called unbalanced and the formula given above for the standard error will not give the right result for a pooled variance t-test. The pooled variance test is used when the homogeneity of variance assumption has been met. In this case, it should be true that $s_1^2 \approx s_2^2$, since they are both estimating the same thing, i.e., the common variance of the two populations. In fact, the difference between the two sample variances should be due only to chance. Thus, while either one of them alone could be used to estimate the common population variance, a more accurate estimate will be obtained if the two sample variances are combined. This is done by calculating a *pooled variance*, which is a weighted average of the two sample variances.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{SS_1 + SS_2}{df_1 + df_2}$$

There are two things you might want to notice about the pooled variance at this point. First, the pooled variance is equal to $(SS_1 + SS_2) / df$. Second, if the design is balanced, the pooled variance is equal to $(s_1^2 + s_2^2) / 2$. One more thing: the pooled variance is the average of the two sample variances, so as a check on your work, make sure the value you get for it is somewhere between the two sample variances.

Using the pooled variance, the standard error of the difference between means is:

$$s.e. = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

In other words, the pooled variance is used in place of the individual group variances.

Example 2. Siegel (1990) found that elderly people who own dogs visit their doctors less often than those who do not own pets. However, the subjects in this study are self selected (quasi-experimental design), and therefore, it's impossible to say this difference is actually due to dog ownership. After all, pet owners may be different from nonowners in other ways as well, and it might be one of these other differences that makes them differ in number of doctor visits. Suppose we redesigned this study to use a true experimental design. Twelve elderly nursing home residents who do not own pets will be used as subjects in the experiment. We shall randomly select some of these individuals to care for a dog, while others will remain dogless. We want to see if caring for a dog has an effect on the number of visits these people make to their doctors.

Step 1) State the null and alternative hypotheses.

H_0:
H_1:

Step 2) Establish a decision criterion (set alpha).

Step 3) Go out and collect the data, calculate summary statistics, and calculate the value of the test statistic.

The following data (number of doctor visits in one year) were collected:

control group (no dog):	11	9	5	8	15	12	14
treatment group (dog):	7	4	8	3	5		

Step 4) Make a statistical decision concerning the null hypothesis.

Step 5) Write a conclusion describing the result of the statistical analysis.