t-Tests: A General Approach

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

t-Tests: A General Approach

Introduction

In this module, we'll quickly revisit the general family of t-tests on means that we studied in Psychology 310.

These tests can handle a surprising number of designs, which may be categorized as

- Between-Subjects. All groups are independent.
- Within-Subjects or Matched Samples. Each unit receives all measurements.
- Between-Within, or Hybrid. There are independent groups, but within groups there are also repeated measures.

Introduction

For any number of groups, the null hypothesis is determined by assigning linear weights to each measurement.

The resulting linear combination is the "quantity of interest" in the hypothesis test or confidence interval.

Let's review the designs we examined in Psychology 310, and how we calculate test statistics confidence intervals for these designs.

Between-Subjects Designs

The major between-subjects designs that use the t-test on means include

- One-Sample Design. In this design, we test whether a population mean μ is a particular value or range of values. The typical 2-sided hypothesis is of the form

$$H_0: \quad \mu = \mu_0 \tag{1}$$

  in which μ_0 is the null-hypothesized value of μ.

Between-Subjects Designs

- Two-Sample Design. In this design, we compare two means from two distinct experimental populations (often an experimental and control group). The typical two-sided hypothesis is of the form

$$H_0: \quad \mu_1 - \mu_2 = \kappa_0 \tag{2}$$

  in which κ_0 is the null hypothesized difference between the two means. In the common case in which $\kappa_0 = 0$, the null hypothesis may be written

$$H_0: \quad \mu_1 = \mu_2 \tag{3}$$

- Multiple Sample Designs. The logic of the between-subjects (independent sample) *t*-test may be extended easily to 3 or more groups. We'll consider some examples in the next subsection.

Between-Subjects Designs

Multiple Sample Variants

Many interesting hypotheses can be tested on quantities that are linear combinations of the means of more than two groups.

Let's briefly consider just two such examples, one on 3 groups, and one on 4. After studying these, you should be able to come up with many more examples involving 4 or more groups.

Between-Subjects Designs

A 3-Sample Test

A 3-Sample Test. Suppose we have 2 experimental groups and one control group, and we wish to test whether the average of the two experimental group means deviates from the control group mean. In this case, the null hypothesis would be

$$\frac{\mu_1 + \mu_2}{2} - \mu_3 = 0 \tag{4}$$

or, equivalently

$$\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{5}$$

Between-Subjects Designs

4-Sample Tests

Suppose we have 4 groups, representing, respectively, Male-Experimental (μ_1), Male-Control (μ_2), Female-Experimental (μ_3), Female-Control (μ_4).

In this design, there are several potential hypotheses of interest.

For example:

- Test of No Interaction. Note that the experimental effect for males is $\mu_1 - \mu_2$, and the experimental effect for females is $\mu_3 - \mu_4$. A hypothesis of no difference in experimental effects for males and females is

$$H_0: \quad \mu_1 - \mu_2 = \mu_3 - \mu_4 \tag{6}$$

or, equivalently,

$$H_0: \quad \mu_1 - \mu_2 - \mu_3 + \mu_4 = 0 \tag{7}$$

Between-Subjects Designs

4-Sample Tests

- Test of No Experimental Main Effect. The average of the two experimental group means is $(\mu_1 + \mu_3)/2$. The average of the two control group means is $(\mu_2 + \mu_4)/2$. A hypothesis that the average of the experimental means is equal to the average of the control means is

$$H_0: \quad \frac{\mu_1 + \mu_3}{2} = \frac{\mu_2 + \mu_4}{2} \tag{8}$$

which can be expressed equivalently as

$$H_0: \quad \mu_1 + \mu_3 = \mu_2 + \mu_4 \tag{9}$$

or

$$H_0: \quad \mu_1 - \mu_2 + \mu_3 - \mu_4 = 0 \tag{10}$$

Between-Subjects Designs

Something in Common

As we saw in Psychology 310, all the between-subjects designs and associated hypothesis tests can be processed with the generalized *t*-test formula, because all of the preceding hypotheses can be expressed in the form $\kappa = \kappa_0$, where κ is a linear combination of means, and κ_0 is a numerical constant.

We can write a single, "generalized *t*" routine in R that will process all the examples. All we need input are

- The value of κ_0, which is zero by default.
- The linear weights that define the linear combination of interest.
- The sample means in the correct order.
- The sample standard deviations (or variances).
- A confidence level.
- The direction of the null hypothesis (necessary to compute the *p*-value correctly).

The program will compute a *t* statistic, degrees of freedom, *p*-value, and a confidence interval for the linear combination of interest.

Between-Subjects Designs

Something in Common

Recall that the generalized t statistic for testing the null hypothesis $\kappa = \sum_{j=1}^{J} c_j \mu_j = \kappa_0$ may be written in the form

$$t_\nu = \frac{K - \kappa_0}{\sqrt{W \hat{\sigma}^2}} \tag{11}$$

where $W = \sum_j c_j^2 / n_j$, the degrees of freedom are $\nu = \sum_j (n_j - 1)$, and $K = \sum_j c_j \bar{X}_{\bullet j}$.

The pooled unbiased estimator $\hat{\sigma}^2$ is computed from the J sample variances S_j^2 as

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{J} (n_j - 1) S_j^2}{\nu} \tag{12}$$

The corresponding confidence interval is of the form

$$K \pm t^*_{\nu, 1-\alpha/2} \sqrt{W \hat{\sigma}^2} \tag{13}$$

Within-Subjects Designs

The within-subjects design involves units within which more than one observation is taken.

The classic example from Psychology 310 is the correlated sample *t*-test.

Two observations are taken on the same unit. If the unit is an individual subject, the design is called a "repeated measures" design.

If the experimental unit involves natural matching, such as a married couple, one observation might be the wife and the other the husband. The observations are thus correlated.

Suppose one measurement has a mean of μ_1, the other μ_2. The null hypothesis is, once again,

$$H_0 : \quad \mu_1 - \mu_2 = 0 \tag{14}$$

Within-Subjects Designs

Although we have two measurements, we cannot directly employ the independent sample t-test procedures of the preceding section, because the observations are correlated.

However, it turns out that we can employ our generalized t routine anyway!

Since we are interested in the mean difference between measurements, we can utilize the general strategy I call "linear precombination."

If we linearly combine the two observations into 1 column of difference scores using the same linear weights (i.e., in this case, $+1$ and -1) as employed in the null hypothesis, then the one column of numbers (which I'll call D) has a mean of $\mu_D = \mu_1 - \mu_2$. And this is the statistical quantity of interest.

Within-Subjects Designs

Consequently, we can perform a 1-sample t test of the hypothesis that $\mu_D = 0$, and simultaneously be testing the hypothesis that $\mu_1 - \mu_2 = 0$. This is discussed in detail in Case 3 of the *Cases* handout in Psychology 310.

The resulting test is called the "correlated sample t-test" or the "paired sample t test," depending on the circumstances. It is simply a 1-sample t test on the hypothesis that the mean of the difference scores is zero.

Within-Subjects Designs

This strategy used in the correlated sample t generalizes.

For example, suppose I had 3 repeated measures on each unit, and that the first two measures represent pre-treatment assesments, and the third measure represents a post-treament assessment. I wanted to test the hypothesis that

$$\frac{\mu_1 + \mu_2}{2} - \mu_3 = 0 \tag{15}$$

that is, that the average mean of the first two measures is equal to the mean of the third.

Again, I could do this by linearly precombining the three measures to 1, using the formula

$$k_i = \frac{1}{2}X_{i1} + \frac{1}{2}X_{i2} - X_{i3} \tag{16}$$

I would then run a 1-sample t test of the hypothesis $\mu_k = 0$.

Between-Within Designs

So far, we have seen that when the null hypothesis involves only a single linear combination, the t-test is reduced to an independent sample test whether or not the observations are between subjects or within subjects.

This fact remains true in the case of between-within designs, so long as only a single linear combination is of interest.

One classic example should suffice.

Suppose you have 4 columns of data. The first two columns represent two repeated observations on a group of males, and the second two observations represent two repeated observations on a group of females.

Between-Within Designs

You wish to test the hypothesis that the mean change for males is the same as the mean change for females. This hypothesis can be written

$$H_0 \quad (\mu_1 - \mu_2) - (\mu_3 - \mu_4) = 0 \tag{17}$$

or

$$H_0 : \quad \mu_{Dmale} - \mu_{Dfemale} = 0 \tag{18}$$

where $\mu_{Dmale} = \mu_1 - \mu_2$ is the mean difference for males, and $\mu_{Dfemale} = \mu_3 - \mu_4$ is the mean difference for females.

Between-Within Designs

To perform the test, we linearly precombine the first two columns into one column of difference scores (D_{male}), and the last two columns into a second column of difference scores (D_{female}).

These two columns of numbers, computed on independent groups, have means $\mu_1 - \mu_2$ and $\mu_3 - \mu_4$, respectively.

Consequently, we can test the hypothesis of equal change for males and females with a two-sample independent sample t-test of zero mean difference on these two columns of difference scores.

Summary and Conclusions

In the final analysis, we discover that, with the aid of the linear precombination strategy, a broad family of between-subjects, within-subjects, or between-within designs to test a single linear combination hypothesis can be processed with one brief computer routine.

This routine allows us to perform hypothesis tests and confidence intervals on the linear combination quantity of interest.

Subsequently, we will consider issues of power analysis, sample size analysis, and confidence intervals on effect size.

But for now, let's pause to consider some examples.