

Chapter 25: Paired Samples and Blocks

The Problem

The two sample procedures (the formulas) require that the samples are independent of one another. This is primarily because of the formula for finding the variance of a combined random variable: $\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$ (which we saw in an earlier chapter). When the samples are not independent, then the preceding formula isn't correct, and as a result the t -statistic that we developed doesn't have an approximate t distribution.

Identifying the Problem

The question is this: when will the samples *not* be independent?

In reality, this is a hard question. In AP Statistics, there is a fairly simple answer: *are the data sets paired in some way?*

If there is some systematic way to connect each datum in one set with a unique datum in the other set, then the data are paired. For example, if we measure the IQ of a set of individuals, and then find the IQ of the birth mother of each of those individuals, then the two sets of data can be paired by family (child paired with mother).

One easy thing to note: pairing can only exist if the two sets of data have the same size. There would be no way to pair 10 objects with a set of 12 objects!

The Solution

When the data are paired, think of the individuals as the *pairs themselves*. This means that the measurement from each individual will be the difference in measurements within the pair! This turns the two sets of data into one set of differences—and you know how to analyze a single set of quantitative data.

Thus, when the data are paired, perform one sample procedures on the set of differences. When the data are not paired (independent), perform two sample procedures on the two sets of data.

Blocking

Pairing is an example of a larger concept called *blocking*. If you're following the textbook in its printed order, then you've already talked about blocking. If you're following my order, then you're not ready to hear about blocking!