Pre-lecture Notes II.4 – Hypothesis Testing

First, a brief review of the two main design types. Assuming that there are only two conditions in the experiment, there are two different ways that the experiment can be run. You can have all subjects participate in both conditions, such that the comparison between the conditions is done within each subject; this, logically enough, is known as a *within-subjects design*. Alternatively, you can have two each subject participate in one or the other condition, such that the comparison between the conditions is also a comparison between two groups of subjects; this is known as a *between-subjects design*.

Choosing whether to use a within- or between-subjects design is probably the second-most important decision in planning an experiment (after choosing a method for creating the conditions). As you'll see, there are positives and negatives associated with each of the two design types. This is a great example of why we say that there's no such thing as a perfect experiment. There are almost always trade-offs between different kinds of validity. More on this in lecture.

The second issue to review before class is the way in which hypothesis testing (currently) operates in psychology. Prior to the analysis, there is always a default answer (usually referred to as the *null hypothesis*) and then there is the other possible answer that must prove itself. The default answer is that the independent variable has no effect on the dependent variable in the sampling population. (Note the way that was written. I didn't write that the null hypothesis was no effect of the IV on the DV in the data. We're now talking about inferential statistics – going beyond the data in hand to make a statement about the population from which the sample was take – so the conclusion is in terms of the sampling population, not the actual sample.)

The alternative hypothesis is that the IV does have an effect on the DV. We don't have to say what size the effect will be – in fact, we don't even have to say what direction the effect will be. The alternative is simply the opposite of the hull hypothesis. So, because the null hypothesis is "no effect," the alternative is "some effect."

Finally, note how the notion of a default answer vs an answer that has to "prove itself" works. Prior to the analysis, the conclusion would be that the IV has no effect on the DV. This, in a way, helps to keep things as simple as possible. (The fancy word for this is *parsimony*.) The default answer is that the IV has no effect, so theories do not have to include this factor. In contrast, the alternative must prove itself. We will only feel obligated to devise theories that explain why the IV influences the DV if the IV has "proven itself" to be important.

OK, so how does an IV "prove itself" to be an influence on the DV (in the sampling population)? To see how this works, note that it isn't enough that there be some (small) difference in the data between the two conditions. Small differences could occur by chance. For example, if you take two dice that are absolutely identical (i.e., there is no difference between them) and roll them 20 times each, you will probably get a different mean score for each. (One might have a mean score of 3.4 and the other have a mean score of 3.6, even though the two dice are identical; this is just measurement error.) In other words, even when there is no difference between the two conditions, you will often get different mean scores, all due to the variability/unreliability of the measure being used. So it can't be this simple.

Luckily, if we make a few assumptions, we can calculate the probability of getting the observed data on the assumption that there is no difference between conditions. (For example, we can calculate the probability of getting two means that are 0.2 points or more different when you roll two identical dice twenty times each.) The rule that we use in psychology is this: if the probability of getting the observed data – assuming the null hypothesis that there is no real difference – is less than 5%, then we conclude that this too unlikely to be true, so there must be a real difference between the conditions. This is what we mean when we say that the difference between the conditions is “significant.” A significant difference is one that has less than a 5% probability of happening by mere chance.

As strange as this may sound (and this may help you to see why some people really don’t like this approach to statistical inference), we assume that anything with less than a one-in-twenty chance of occurring does not actually occur.