Data Analysis in Experiments

These comments apply to both true and quasi-experiments. As the slide show points out, the logic of the experiment (true or quasi) leads to some sort of data analysis in which the ***average performance of treatment and comparison groups*** is compared for the post-test. The point of an experiment is to see if a treatment has an effect. Therefore, measuring differences between groups (quantitatively or qualitatively) is essentially the only logical way to detect the results of an experiment.

In its simplest form, with just one treatment and one comparison group, the t-test or the Mann-Whitney U test is a logical choice for statistical analysis (see the cheat sheet on statistics to know which of these to use). Either of these will tell you whether there is a ***difference in average effect*** between the two groups. Where there are more than two groups, analysis of variance (ANOVA) or the Kruskal-Wallace one way analysis of variance by ranks can be used. Quite often, probably even more commonly, the researcher wants to know how much performance (blood pressure, reading skills, behavior, etc.) ***changed.*** Therefore, it is the ***change in score between pre-test and post-test*** that is important, not the absolute score on the post-test. In this case, three complications arise.

The first is that pre- and post-test scores are related. Take a very simple example of a pre- and post-test about statistics in college. Student A knows quite a lot about statistics and gets 80 on a pre-test (before any instruction). Student B knows nothing and gets 20 on the pre-test. We want to measure change in knowledge. Student A gets 95 on the post-test and Student B gets 80. Simply comparing scores on the post-test obviously does not tell us who learned the most. Just logically, we can see that Student A showed a 15 point improvement while Student B showed a 60 point improvement. It seems logical that Student B improved the most, but if we simply compare post-test scores, we cannot detect the ***difference in improvement.*** We measure change in score, 15 for Student A and 60 for Student B. In some sense, however, Student A (and others like her) cannot show as much improvement as Student B (and others like him) precisely because she already knew a lot at the pre-test! At most, her score could have improved 20 points on a 100-point exam (from 80 to 100), whereas Student B had a lot of room for improvement, as much as 80 points. To overcome this problem, you must use ***statistical tests for dependent samples.*** This is because the score on the post-test is influenced by the score on the pre-test. This is what we mean by a dependent sample in statistics – that the pre- and post-test scores that we want to measure are related in some way. If you look at your statistics cheat sheet, you will see that there are statistical tests for dependent samples, like the t-test for two dependent samples or the Wilcoxon matched pairs signed ranks test. All of these tests involve calculations that remove the effect of the pre-test score on the post-test score. How they do that is not the concern of this course. Knowing when to use the dependent samples tests is the important point. You must always use these tests when you know that the pre-test score can influence the post-test score. See the Statistics Guide for more information about these tests and when to use them.

The second complication has to do with multiple post-test designs. There are many designs that use more than one post-test (see the cheat sheet on type of experiments). The greater the number of post-tests, the more difficult it is to account for the influence of pre-test scores. Assume that we give these same students a pre-test and three semester exams. Now we have three post-test scores. We expect that there is a cumulative effect from each exam on all the exams that follow. The pre-test will influence post-test 1, post-test 2 and post-test 3. However, both the pre-test and post-test 1 will influence post-tests 2 and 3. And the pre-test, post-test 1 and post-test 2 will influence the score on post-test 3. The ultimate change in score is therefore a product of this cumulative influence. Correct statistical analysis of multiple post-test designs is not simple. I recommend that you consult a statistician. The Statistics Guide will give you some ideas about which tests you can use.

The third complication has to do with how we measure change. In the example of tests of knowledge, this is pretty straightforward. We can use the simple change in raw score as we did in the example. We could also use percentage change in score because we have ratio data (someone could get 0 on the pre and/or post-test). As we will see when we discuss longitudinal designs, simple change in raw score or percentage change does not always work, particularly in designs (including longitudinal designs) where we need to know about within group (between individual participants) variance and change in scores. We will discuss these complications in the module on longitudinal designs because this is an important issue for those designs in particular. Nonetheless, the same kinds of considerations come into play in experiments. For now, simply be aware that there are potential problems.

Another very important consideration in data analysis for experiments is the role of independent variables in the outcome. These are **not the treatment or factors in the experiment.** If you cannot remember the difference between a treatment or factor in experiments and an independent variable, review the materials in the cheat sheets basics of sampling and on characteristics of design groups. No matter what experimental design we use – even the classic true experiment – we often know that we cannot screen for and therefore eliminate all of the characteristics of participants that can influence the outcome of the experiment. We try because the purpose of the experiment is to show that the treatment ***and only the treatment*** produced the outcome effect. “If I poke it, it jumps. If I don’t poke it, it doesn’t jump.” We therefore typically measure characteristics that we think could influence the outcome when ***we cannot screen to eliminate*** these characteristics. In fact, most experiments include several independent variables. We can then statistically remove the influence of these variables on the outcome. One common way to do this is to use the **Single Factor Between-Subjects Analysis of Covariance (ANCOVA)** statistical test. The Statistics Guide explains what this test does. Other procedures that can be used are some of the tests of association. ***Note that Assignment 4 requires that you include at least one independent variable (again, NOT the treatment or factors)*** in your design.