2 – Introduction to Statistics, Part II

In a study, data is collected and then conclusions are drawn or conjectures tested from the data.

There are two types of studies:

- An observational study draws conclusions based upon observations. Researchers make no changes.
- An experimental study draws conclusions based upon the manipulation of one or more variables and uses control & experimental groups to compare.

Example: In a study to determine the effect of homework frequency on course performance, the control group is given traditional homework assignments. The experimental group is given more frequent homework. The course grades of the two groups are compared to identify whether or not the difference had any effect.

There are three types of variables in an experimental study:

- The independent variable is the one that is manipulated (homework frequency in the above example).
- The dependent variable is the variable being observed for change (course grades in the above example)
- All other variables are confounding variables. (the problems on the assignments themselves, instructor, time of day, male:female ratio, student majors, student prior math experience, etc. in the above example)

In an observational study, if there is a dependency between variables it should be indicated in the study description. For example, if a study description has the language, "in order to determine if classroom attendance affects course performance," then the implied dependency is that attendance is the independent variable and performance is the dependent variable.

Controlling for confounding variables, or working to limit their influence, is an important part of designing a study. This can be done in several ways:

- Sample size – the larger, the better (promotes diversity among subjects)
- Uniformity – in the example above, we could control for the variable *instructor* by making sure the same instructor taught all sections of the class involved in the study.

Consumers of statistics should be skeptical of misuses. There is no conventional way to define this, but some common misuses are exemplified below:

Improper sampling techniques – A college Calculus III class is surveyed and the mean SAT math score of the students is 680. The claim is made that the average SAT of college bound students is 680. (this is a biased sample)

Detachment – A new version of a food product claims to have 20% less fat. (less than what?)

False cause – Students in my Statistics courses held in the morning do better than those in the evening. (it would be improper to attribute the difference to time of day; it's probably due to some other variable, like the scheduling constraints of various majors)

Poorly constructed surveys – A survey given to Math 175 students in the past: How often do you get sick: all the time, sometimes, rarely, or never? (This is a personally relative question and a specific frequency could elicit different responses from different people).