Objectives:

- To understand the importance of proper data collection designs in order to draw meaningful conclusions from a study
- To appreciate the fundamental distinctions between ***populations*** and ***samples*** and between ***parameters*** and ***statistics***
- To recognize ***biased*** sampling methods and to be wary of conclusions drawn from studies which employ them
- To discover the principle of ***simple random sampling*** and to be able to implement it using a table of ***random digits***
- To begin to develop an informal sense for some of the properties of randomness, particularly with regard to sample size

Sampling: Sampling must be used in most cases to get an idea of the characteristics of a population is often impossible to study an entire population (would it feasible to give EVERYONE a "Neilson Box"?)By using the Neilson Box networks obtain information on what viewers want to see (quite accurately) by studying a very small sample of the American population.

Population: Entire group being studied NO MORE NO LESS. If we are studying females from 30 - 40 years old than we mean each every female between these ages and we cannot include a female that is 29 years old - have seen health studies that have to refuse subjects? They must refuse subjects because they may not perfectly fit the population that needs to be studied.

Representative: A sample is representative if the characteristics of the sample are similar to that of the population. A sample is useful only if it is selected in a manner that the population is represented properly.

Sample: a portion (subset) of the population that is studied to find characteristics about a population. A sample of 1500 people may be used to predict a national election in November and done so quite accurately if sampled correctly

Sample Size: The number of observational units studied in a sample (usually represented by ***n***.

Poor examples of samples:

- **Convenience sample** - your there so I ask you
- **Voluntary response** - call in your opinion about Elvis being dead or alive - what kind of person would call for this?

Sampling Bias: When a sampling method tends to over-represent a portion of the population, the above methods are ***biased*** sampling techniques since the subjects may be in a particular place at a particular time for a particular reason (NOT RANDOM) or because a person if they get the gumption will respond - this may only happen if opinion is strongly held - somebody who doesn't care won't respond.

Non-response Bias: A bias created by the fact that one who does/is able to respond is different than someone that can't/won't respond.

Voluntary Response Bias: A bias created by the simple fact that people who respond to surveys are different those that do not. People who are rather apathetic will tend to not respond, where as people that are passionate will respond.

Parameter vs. statistic A **statistic** is a value calculated using a sample - and this statistic is *inferred* onto the population. A **parameter** is a value derived from the entire population. It is a value that exactly represents the population. Information from a study of a sample is a statistic. Information given in the census is a parameter.

Explanatory Variable: The variable that "***seems"*** to cause the affect we are looking for. KEYWORD "***seems***" We do not assume in any way that this variable causes the other – it simply appears to or may have an effect.

Response Variable: The variable that "***seems***" to be affected by the explanatory variable.

Observational studies: The investigator passively studies variables already in place and observes and records information based on this. Susceptible to confounding variables since there may be reasons we are not accounting for that the variables are in place. Only an association can be assumed between variables in an observational study, a cause and effect relationship cannot be assumed.

Lurking variable: An unmonitored variable that has an effect on the measurement of the response variable. These lurking variables have a confounding effect on the response.

Confounding variable: A variable that's affect on the response is indistinguishable from the affects of the explanatory variable. This prevents the investigator from isolating the effects of each variable.

Cause-and-effect: An experiment is necessary to establish and/or assess a ***cause-and-effect*** relationship between the ***explanatory*** and ***response*** variables. In the experiment, the experimenter actively imposes the treatment on the subjects to see if there is a direct effect on the response variable from the explanatory.