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.

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.

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?
- These are called ***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.

Random Sampling: A sample is random if <u>each and every member of the population</u> as an exactly equal chance of being chosen for the sample.

Simple random sampling (SRS): A sample is simple and random if <u>each and every possible sample from</u> the population as an exactly equal chance of being the sample chosen

One way to do this is to use a random number table and use these numbers to select individuals for your sample (Most calculators have a random number generator for this purpose).

SRS is very difficult to implement - You may think that if you are studying people from Cassadaga you can randomly select phone numbers in the phone book and call them - why is this possibly a poor sample?

- ***Systematic Samples*** (May introduce bias – must be careful) Creating of groups of subjects and selecting from those groups in a systematic pattern. e.g. assigning random ID #'s to subjects and picking a certain number of subjects from equal ranges of ID#'s (1 from 0 – 9, 1 from 10 – 19, …)
- ***Stratified Sample*** – a method to insure that each type of subject is proportionally represented in a sample. e.g. if you have a population that is 10% left-handed and you want to make sure that 10% of your random sample of 20 people is left-handed then you pick 2 left-handed people and 18 right-handed people. Since it is likely that left-handers will not be properly represented in your sample.

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.

Sampling variability: How the values of sample statistics vary from sample to sample, extremely important concept since this variability can be used to infer parametric values if the conditions are right.

Precision: Refers to how much the sample statistic varies from sample to sample. If the values are close together from sample to sample than we can say the samples are precise. ***Do not confuse this with correct***! Statistics may be very precise but completely incorrect. Also There is a relationship between precision and sample size, Very large sample sizes will provide a very precise statistic (tightly spread values), where as a small sample size will yield a much less precise statistic.