1.   Basic Definitions

*The **POPULATION** is the entire set of people (or animals, things) we wish to study.*

*A **SAMPLE** is a part of the population.*

*A numerical fact about a sample is a **STATISTIC**.*

*A numerical fact about a population is a **PARAMETER**.*

Example, from the handout -- of the 1,008 adults surveyed as part of a random national sample, 72% said they favor having US forces take military action against Iraq to force Saddam Hussein from power. The 72% is a statistic that describes the sample. Statistic is to sample what PARAMETER is to the population. If 72% is what the survey revealed, the people who conducted the survey hope that it is a close approximation of the true population PARAMETER, that is the percentage of all adults in America who favor forcing Saddam Hussein from power.

2.   Problems

A. Bias -- If a sample is "representative", then a statistic can be a good estimate of the parameter; but if the sample includes or excludes certain people systematically, the sample is BIASED. See examples of non-random samples…results from cnn.com Quick Vote.
B. Selection bias --- you include or exclude certain people
C. Non-response bias --- people don't bother to answer you
D. Response bias --- people answer, but they lie to you or they are manipulated by the way you asked the question
E. Wording of question --- phrasing may not be neutral (e.g. a loaded question).

3. Design Issues

Statisticians are well aware of the problem of bias. Only in the last 50 years have survey organizations used probability methods to draw their samples. These Sampling Designs can help.

a. Simple random sample (SRS): every person in the population has an equal chance of getting into the sample with each draw. In practice this is drawing at random without replacement (because it would not make sense to select the same person or measure the same animal/thing twice).

b. Not every sampling scheme is simple random sampling; other sampling schemes include MULTISTAGE CLUSTER SAMPLING. There is a good example of multistage cluster sampling on p.341, Figure 1

The idea here is that a large population (e.g. the US) is broken down into increasingly smaller areas at and each stage a single unit is drawn randomly until the unit of interest (e.g. households) is reached.

Note: these methods can be applied to things other than households. Examples might be estimating the corn harvest, sampling firms on hiring expectations, etc.

***PRINCIPLE*: Probability methods work well because they are impartial.**