

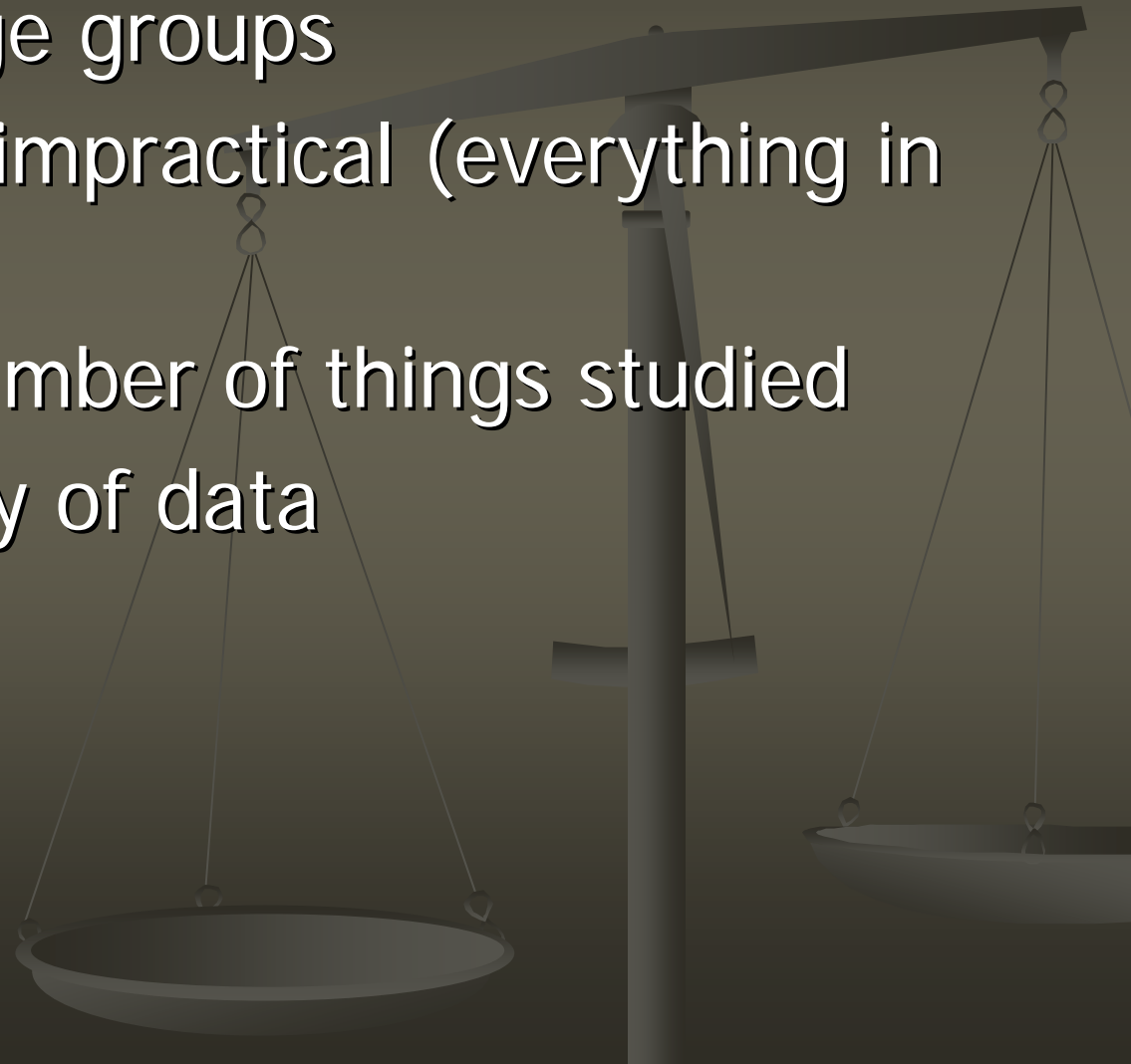


Chapter 7

The Logic Of Sampling

Why Sample?

- Understand large groups
- Census studies impractical (everything in population)
- Minimize the number of things studied
- Maximize quality of data

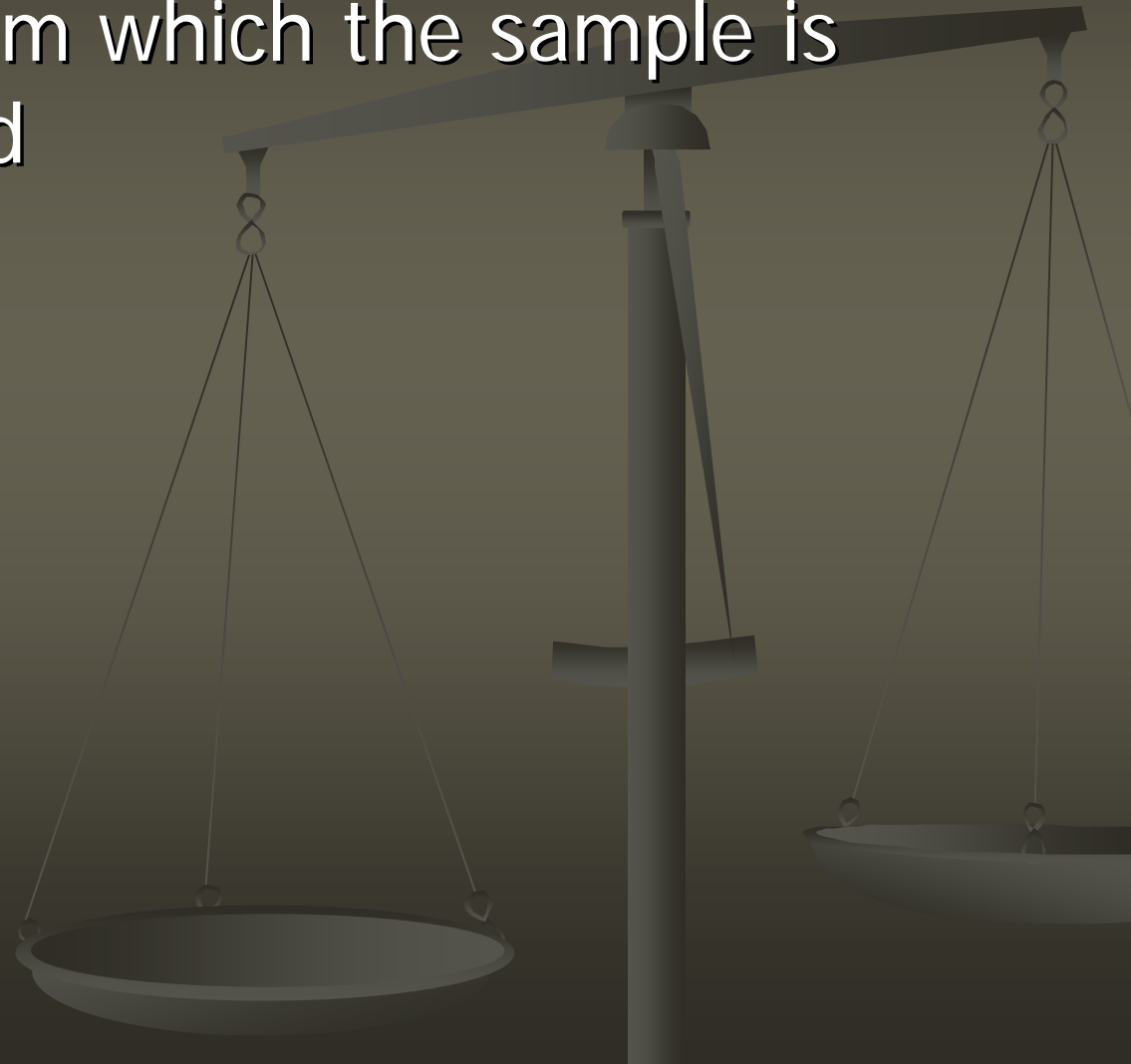


Samples and Populations

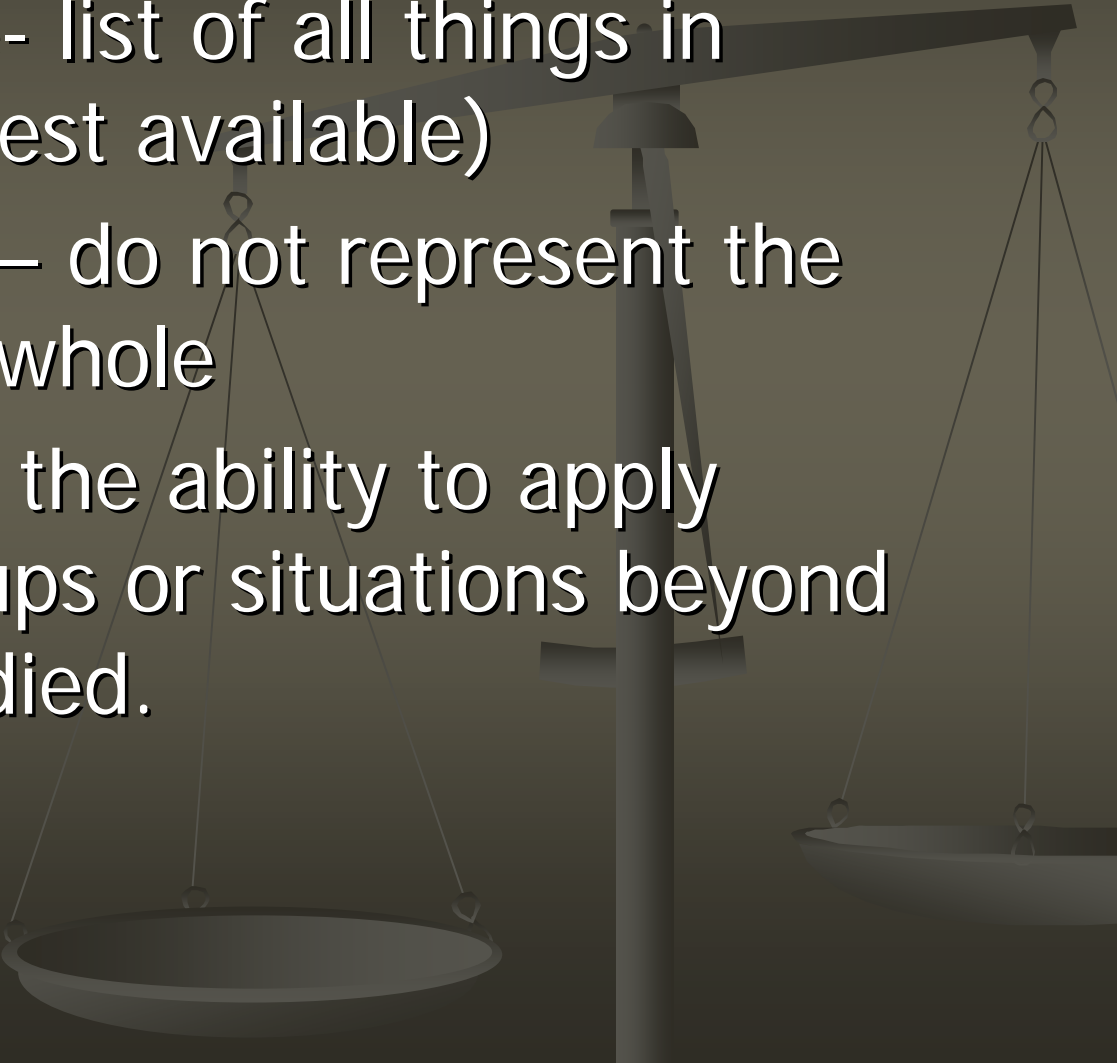
- A **population** includes every case in the group under study. If we are studying Americans, then the population is all Americans. (census) If we are studying libraries then our population is all libraries (in the state? US)?
- A **Sample** includes those cases that we actually observe. A sample is a subset of the population. We rarely have the time or resources to observe every member of the population. We use information from samples of cases to make inferences about the entire population.

Study population

- Is the group from which the sample is actually selected

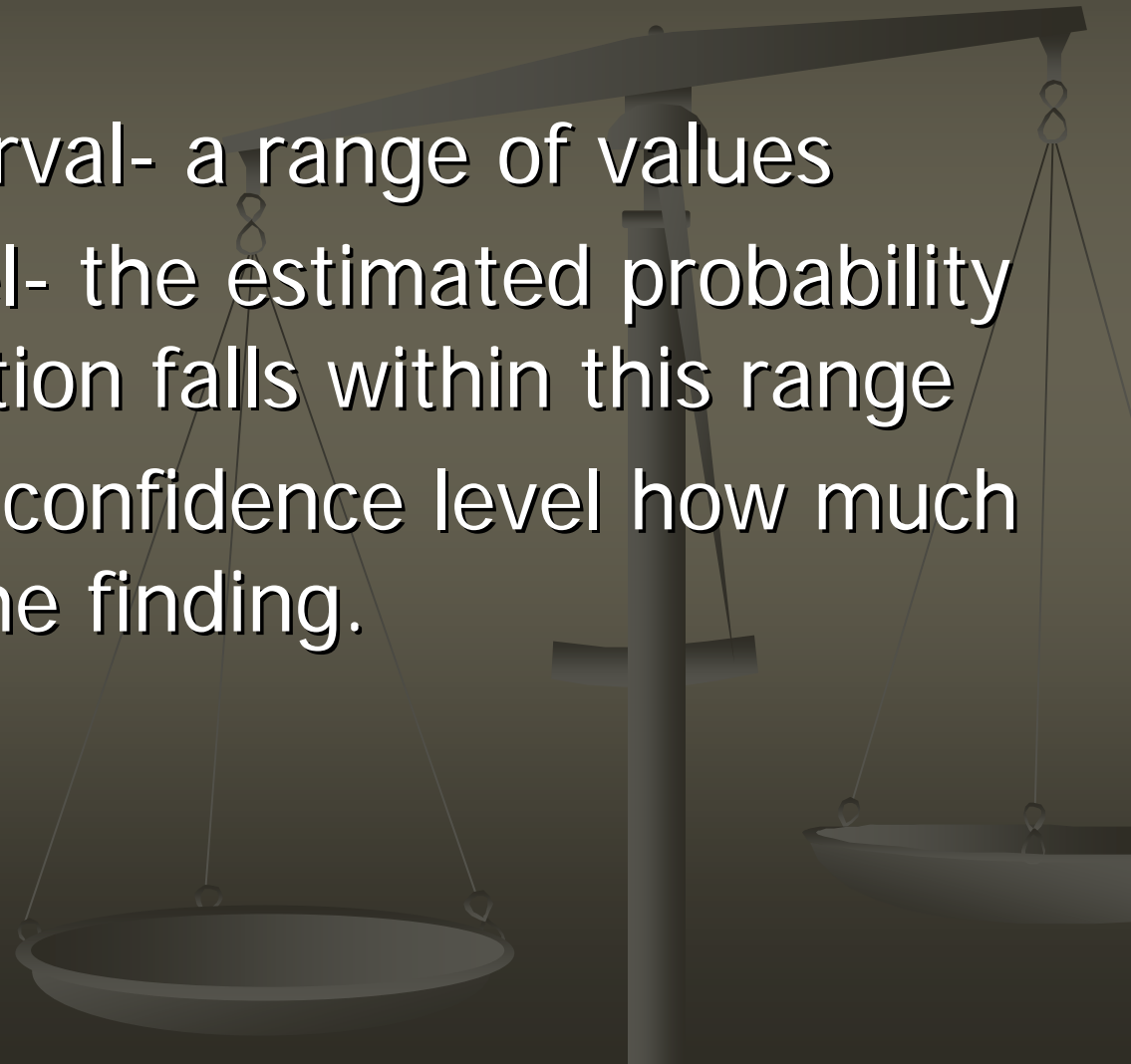


Sampling- some basics

- Sampling Frame- list of all things in population (or best available)
 - Biased samples – do not represent the population as a whole
 - Generalizability- the ability to apply research to groups or situations beyond those being studied.
- 

Sampling distribution

- Bell Curve
- Confidence interval- a range of values
- Confidence level- the estimated probability that the population falls within this range
- Margin of error-confidence level how much range around the finding.

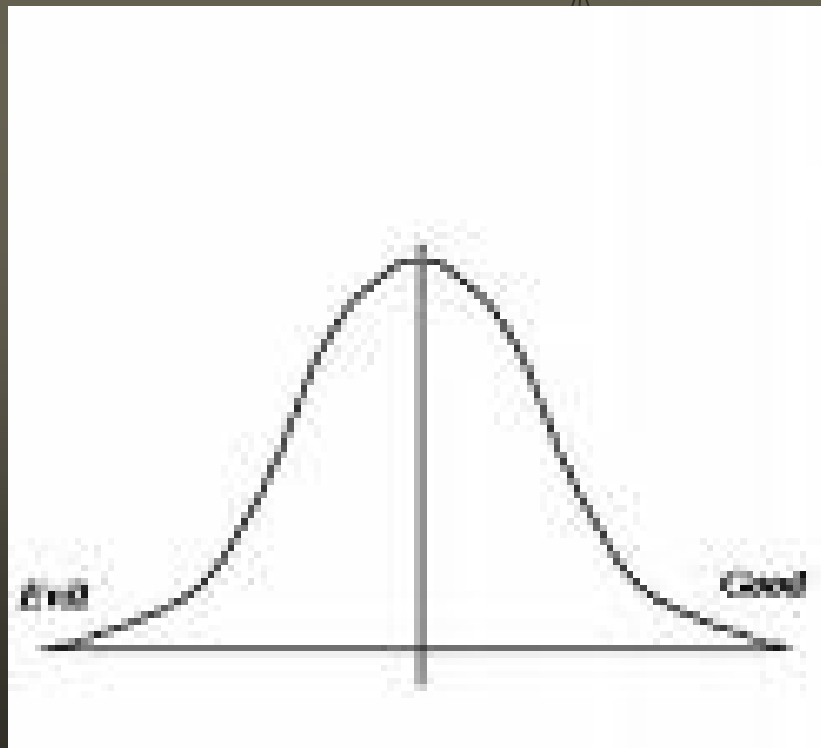


A quick intro to Probability

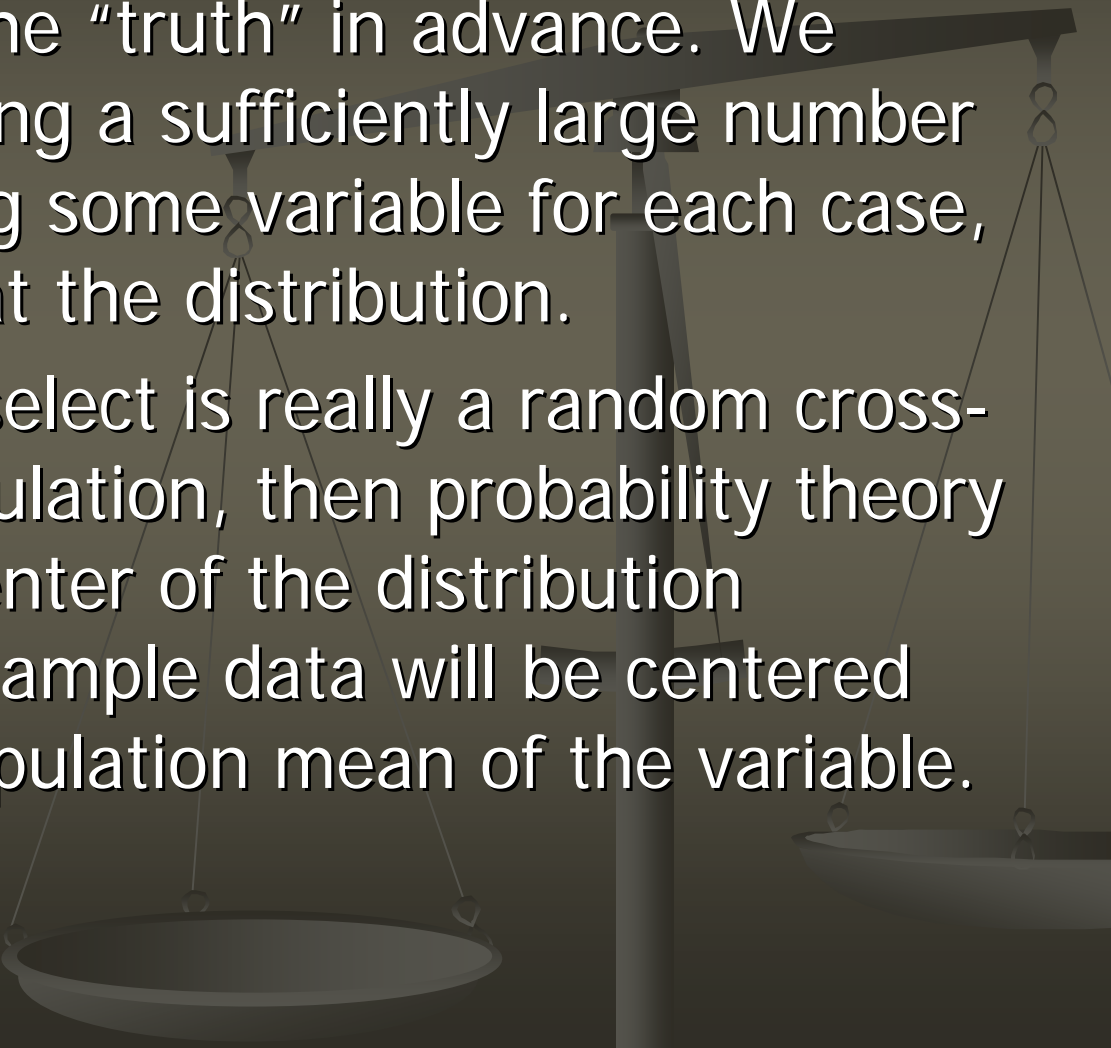
- If you flip a coin ten times, how many heads do you expect to observe?
- Five is the best guess, four and six are also good guesses, but we know that we can get anywhere from 0-10 heads.
- Call the number of heads that we observe X . If we flip a coin ten times, then record X and then repeat N times ($N = \#$ of times we toss the coin), we get N values of X .
- We can then make a histogram chart to show the distribution of X .

A quick intro to Probability

If we repeat our Test (flipping a coin ten times) 5 times, we will have 5 values of X and the distribution may look like this:



A quick intro to Probability

- We do not know the “truth” in advance. We proceed by sampling a sufficiently large number of cases, observing some variable for each case, and then looking at the distribution.
 - If the sample we select is really a random cross-section of the population, then probability theory tells us that the center of the distribution created from the sample data will be centered over the “true” population mean of the variable.
- 

Generalizability



- When we use sample data to infer population characteristics, we are generalizing our results and findings to the population of interest.
- Valid only if sample is really representative of population.
- To what extent are samples representative of populations
- Often, Non-probability sampling generalizes results and findings to theory.

Generalizability

- Why might these samples not be entirely representative of the populations from which they are drawn?

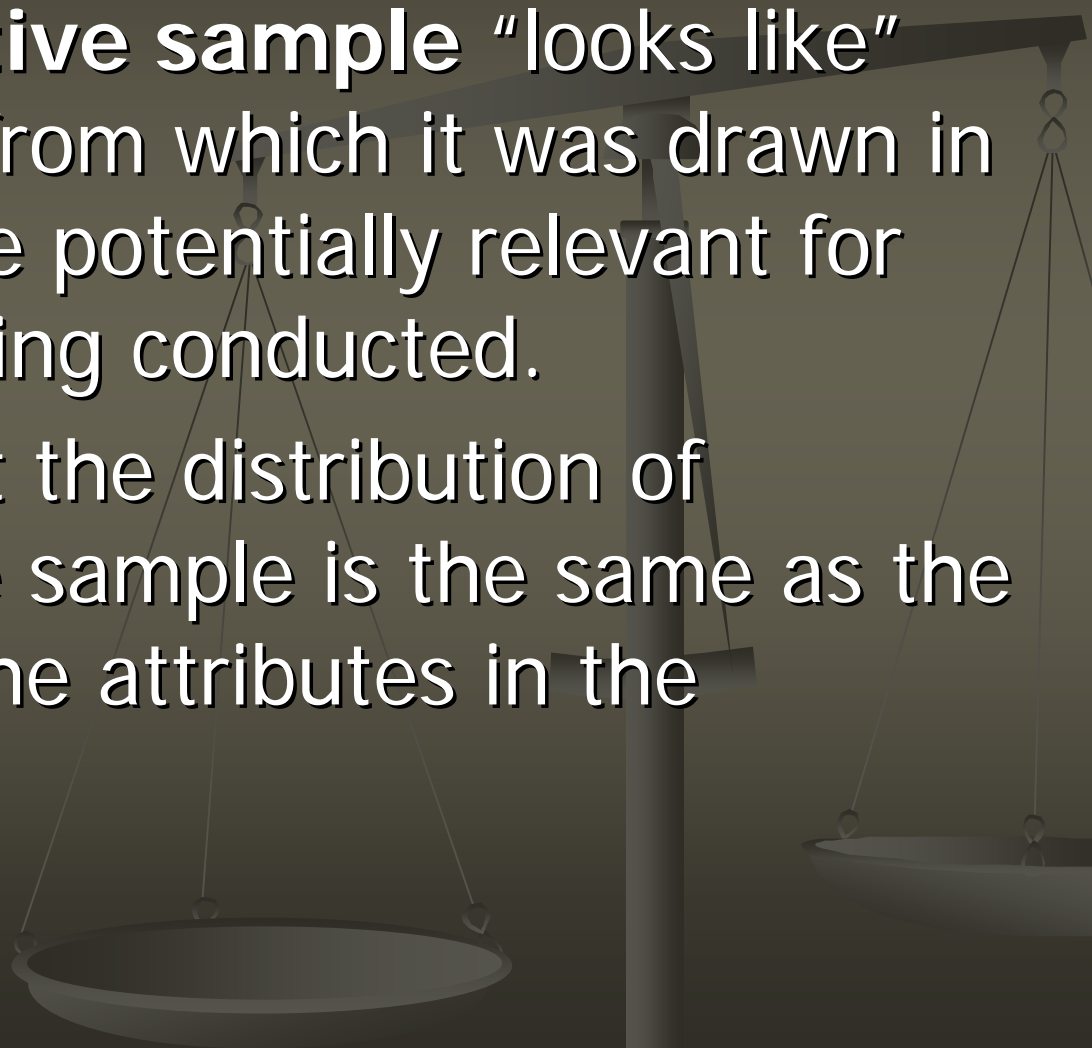
| Population | Randomly selected sample |
|--------------------------------|--|
| New born babies in Minneapolis | 1,000 babies recently born in Minneapolis hospitals |
| Homeless people in St. Paul | 500 homeless people who were on the street or in shelters on a given night |
| Adults in the US | 12,301 adults who could be contacted by phone |

Generalizability

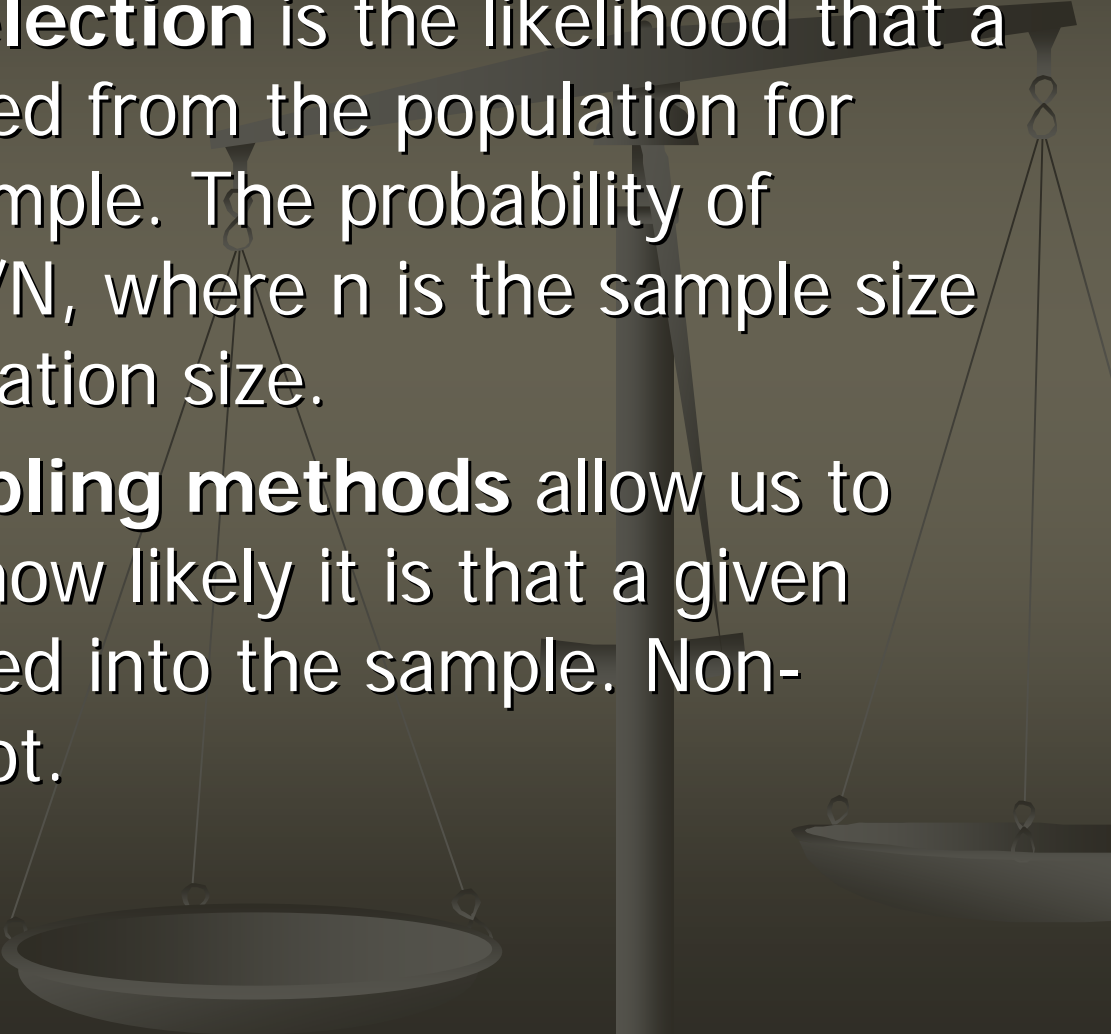


- Sampling error is the difference between the characteristics of a sample and the characteristics of the population from which it was drawn.
- With more sampling error we are less able to validly generalize from the sample to the population

Generalizability

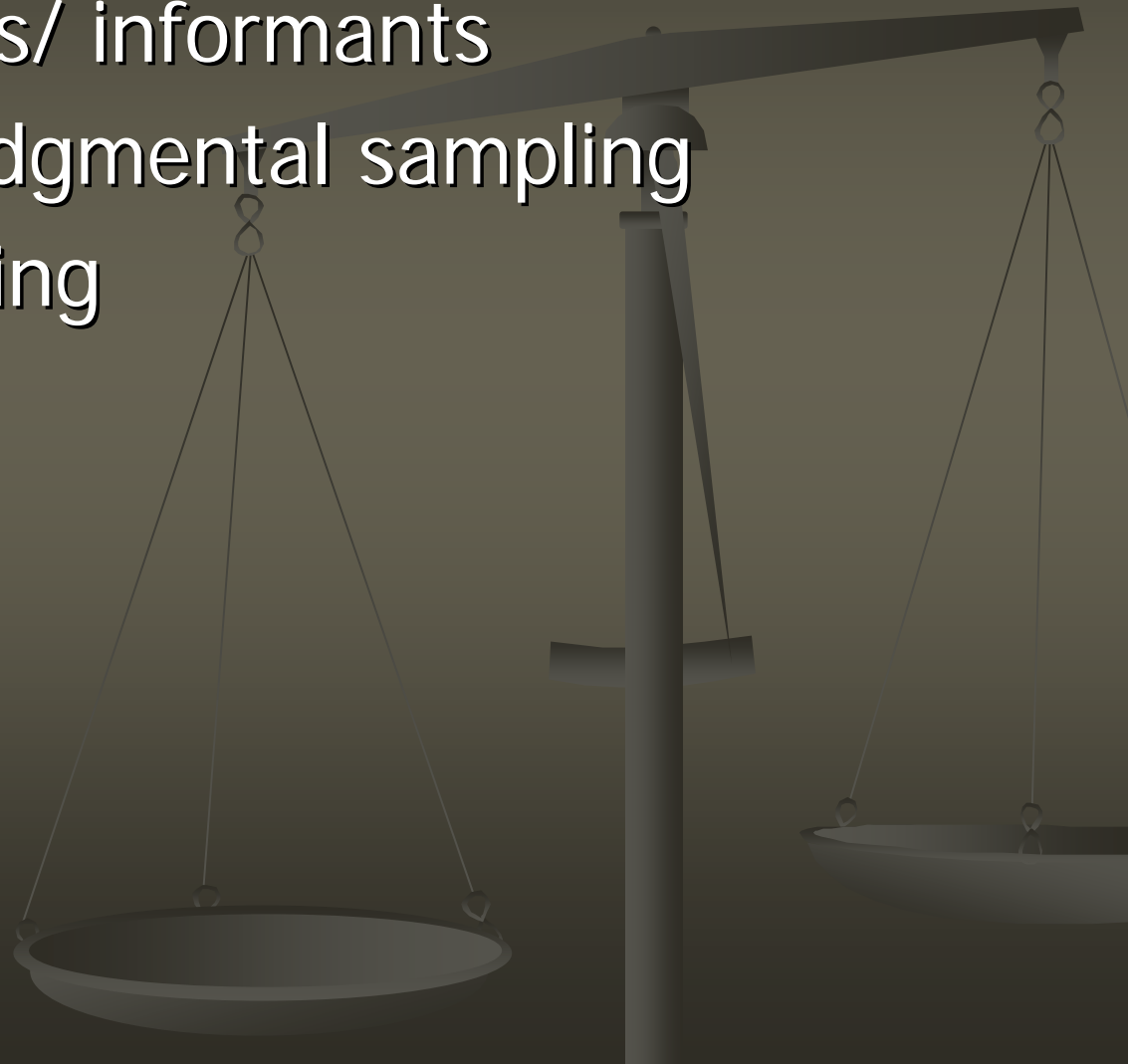
- A **Representative sample** “looks like” the population from which it was drawn in all ways that are potentially relevant for the research being conducted.
 - This means that the distribution of attributes in the sample is the same as the distribution of the attributes in the population.
- 

Sampling: some definitions

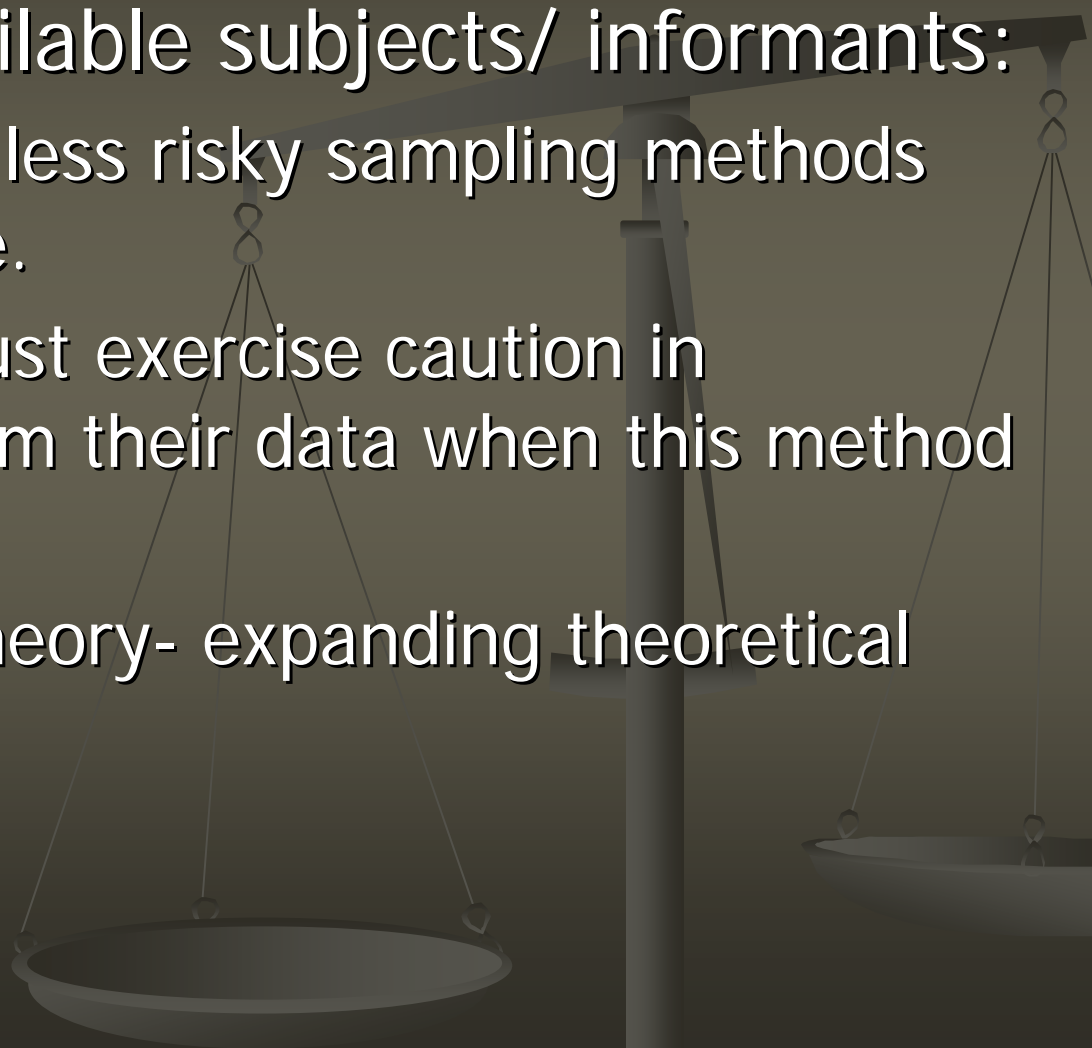
- **Probability of selection** is the likelihood that a case will be selected from the population for inclusion in the sample. The probability of selection equals n/N , where n is the sample size and N is the population size.
 - **Probability sampling methods** allow us to know in advance how likely it is that a given case will be selected into the sample. Non-probability does not.
- 

Nonprobability sampling

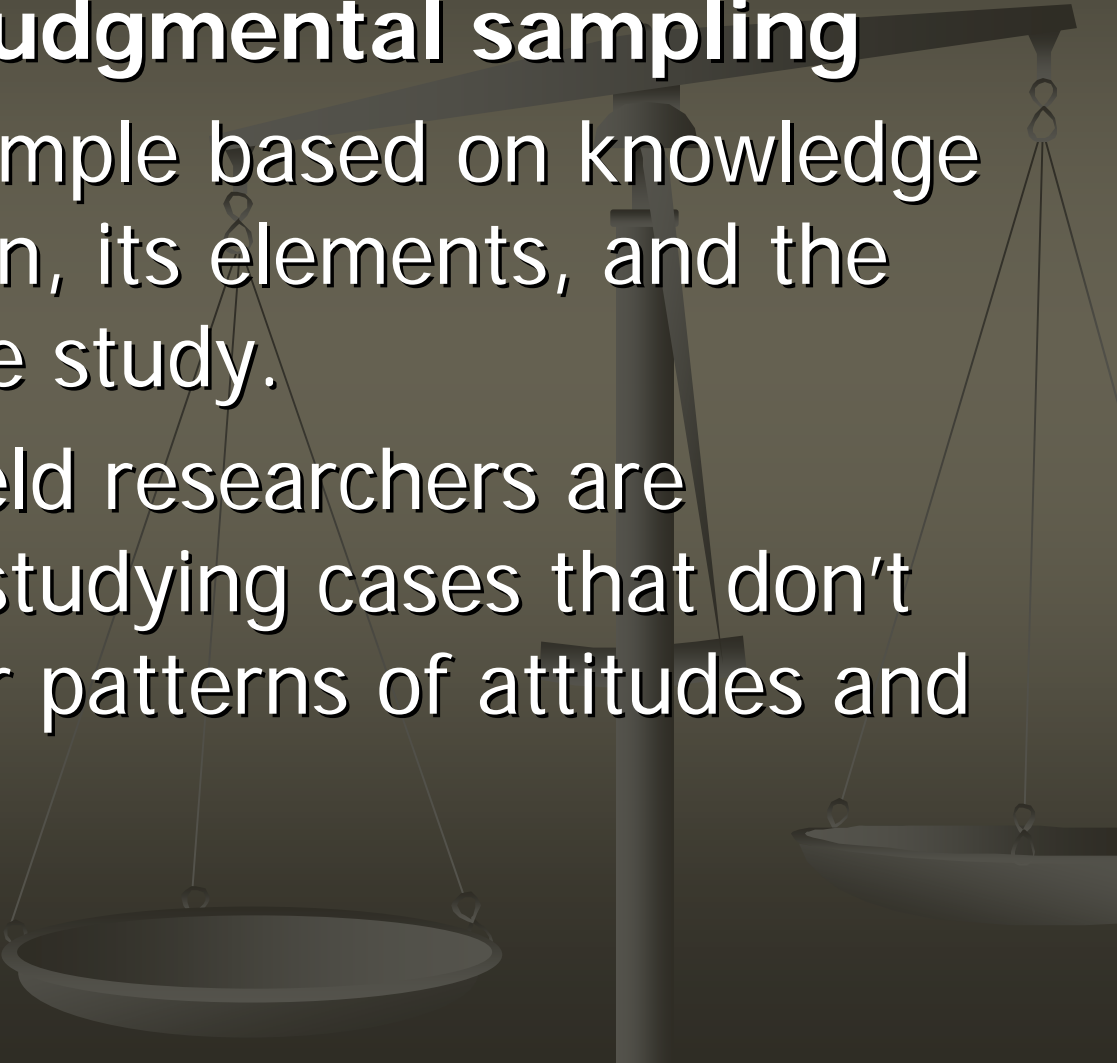
- Reliable subjects/ informants
- Purposive or Judgmental sampling
- Snowball sampling
- Quota sampling



Types of Nonprobability Sampling

- Reliance on available subjects/ informants:
 - Only justified if less risky sampling methods are not possible.
 - Researchers must exercise caution in generalizing from their data when this method is used.
 - Generalize to theory- expanding theoretical ideas.
- 

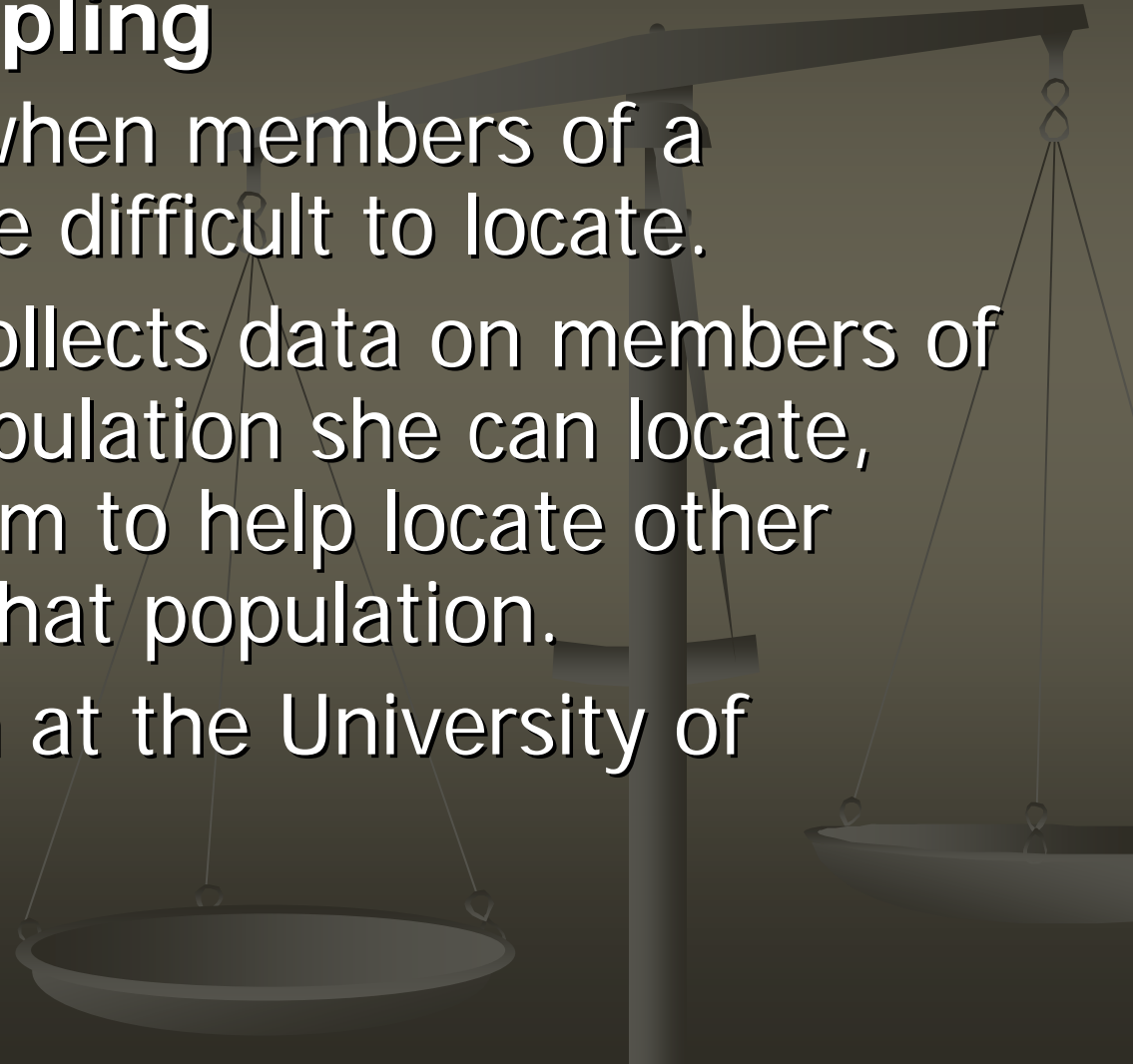
Types of Nonprobability Sampling

- **Purposive or judgmental sampling**
 - Selecting a sample based on knowledge of a population, its elements, and the purpose of the study.
 - Used when field researchers are interested in studying cases that don't fit into regular patterns of attitudes and behaviors
- 

Types of Nonprobability Sampling

■ Snowball sampling

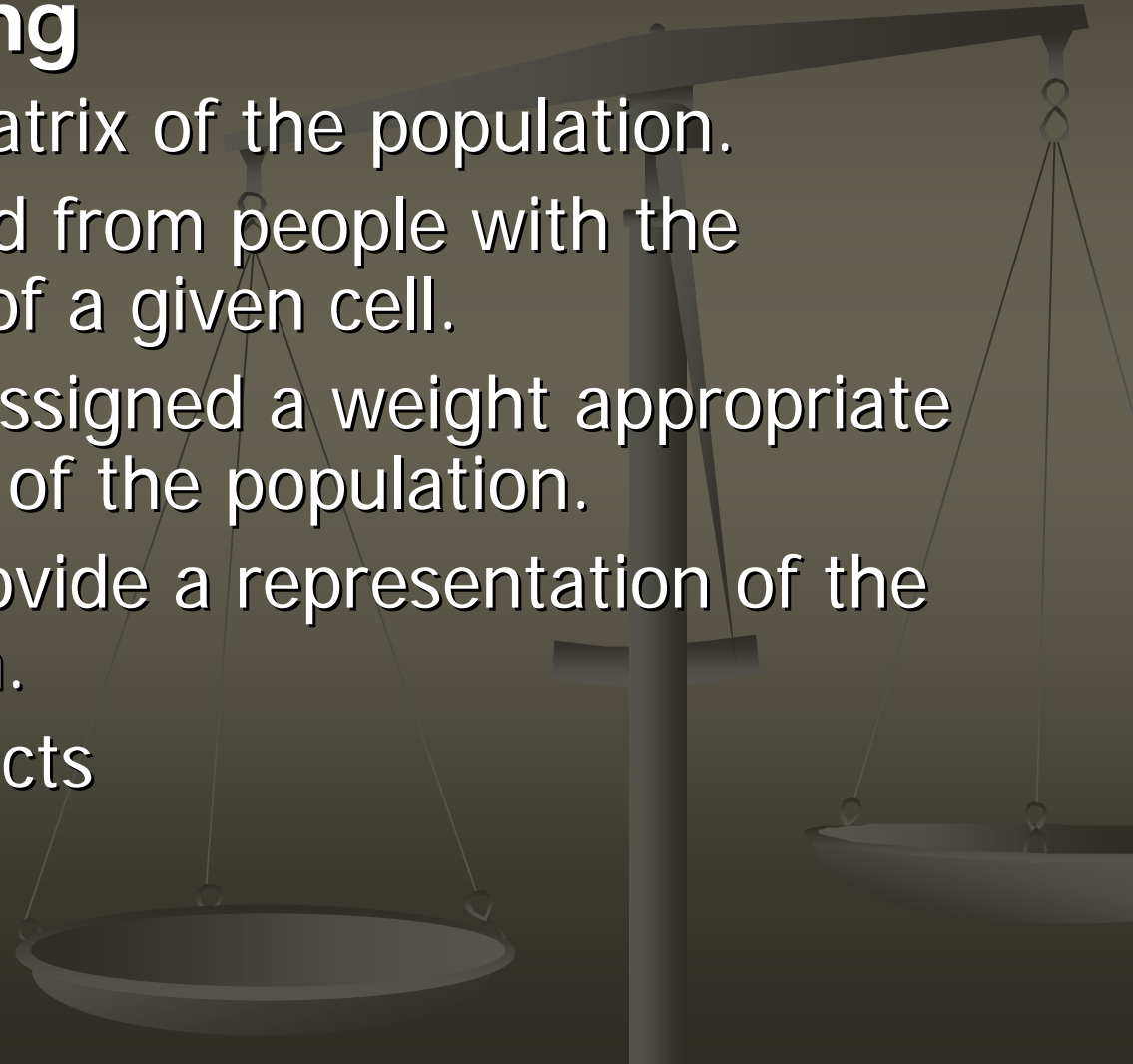
- Appropriate when members of a population are difficult to locate.
- Researcher collects data on members of the target population she can locate, then asks them to help locate other members of that population.
- Not an option at the University of Minnesota



Types of Nonprobability Sampling

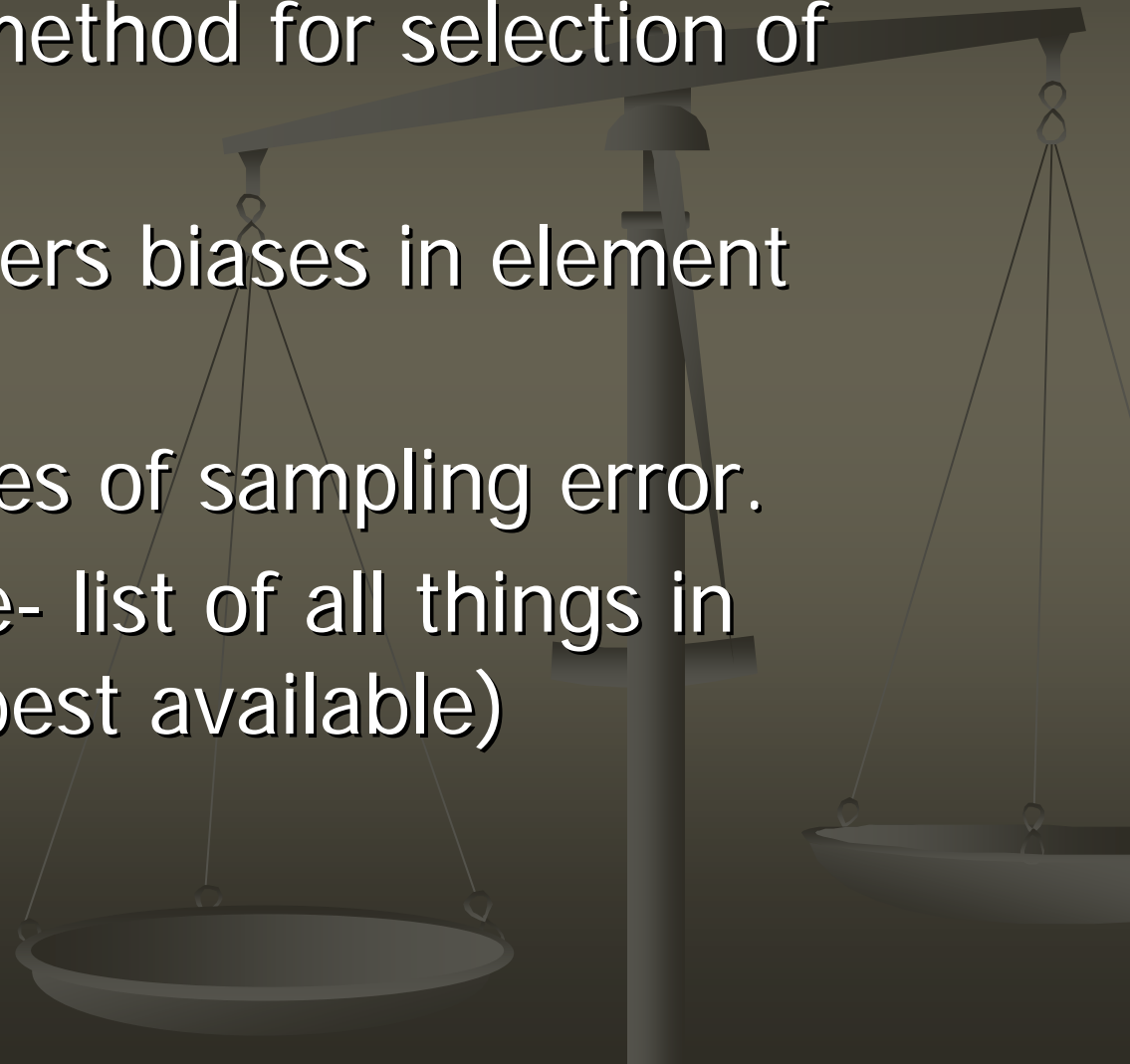
■ Quota sampling

- Begin with a matrix of the population.
- Data is collected from people with the characteristics of a given cell.
- Each group is assigned a weight appropriate to their portion of the population.
- Data should provide a representation of the total population.
- Hair care products



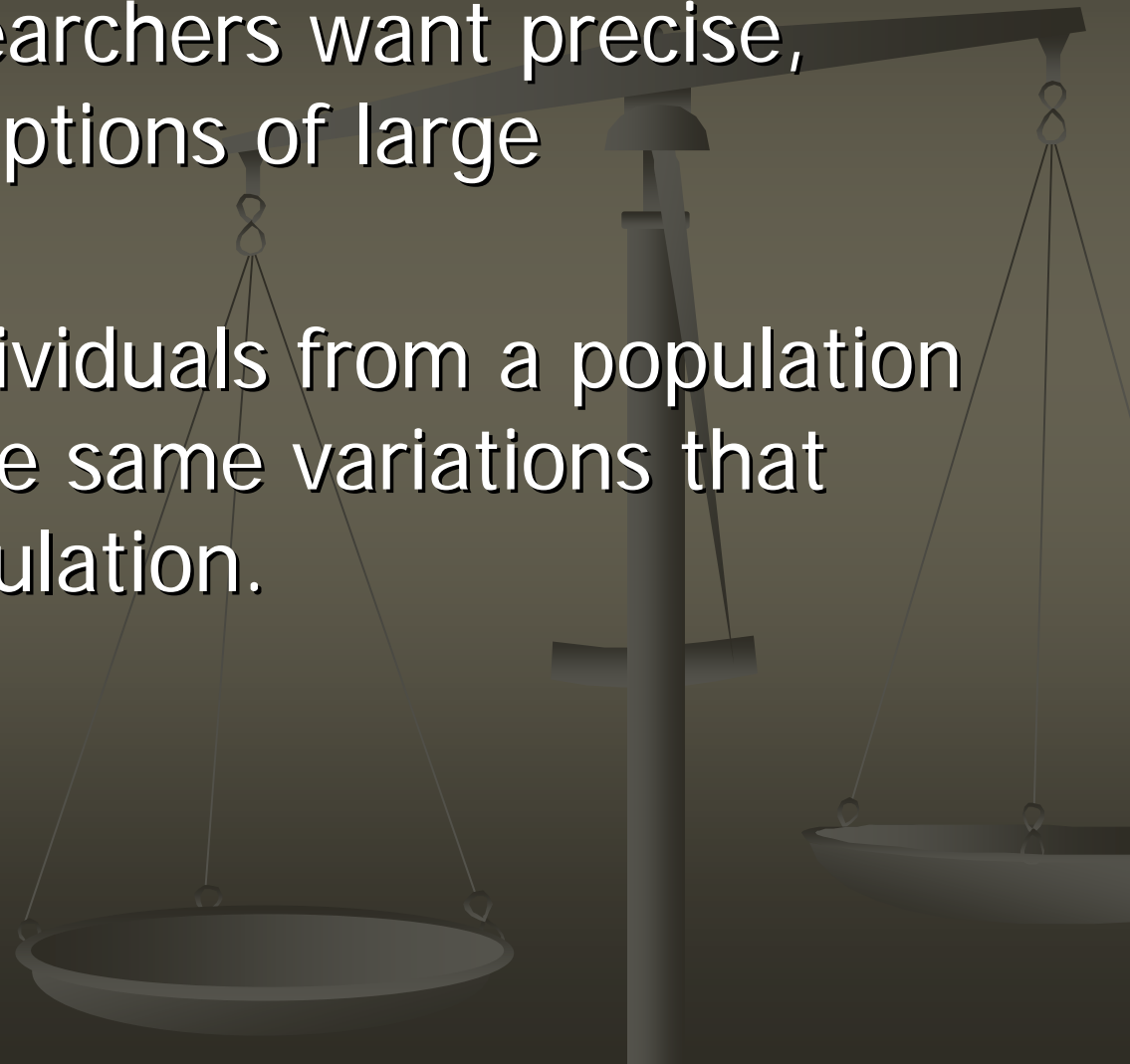
Probability Sampling

- Most effective method for selection of study elements.
- Avoids researchers biases in element selection.
- Permits estimates of sampling error.
- Sampling Frame- list of all things in population (or best available)



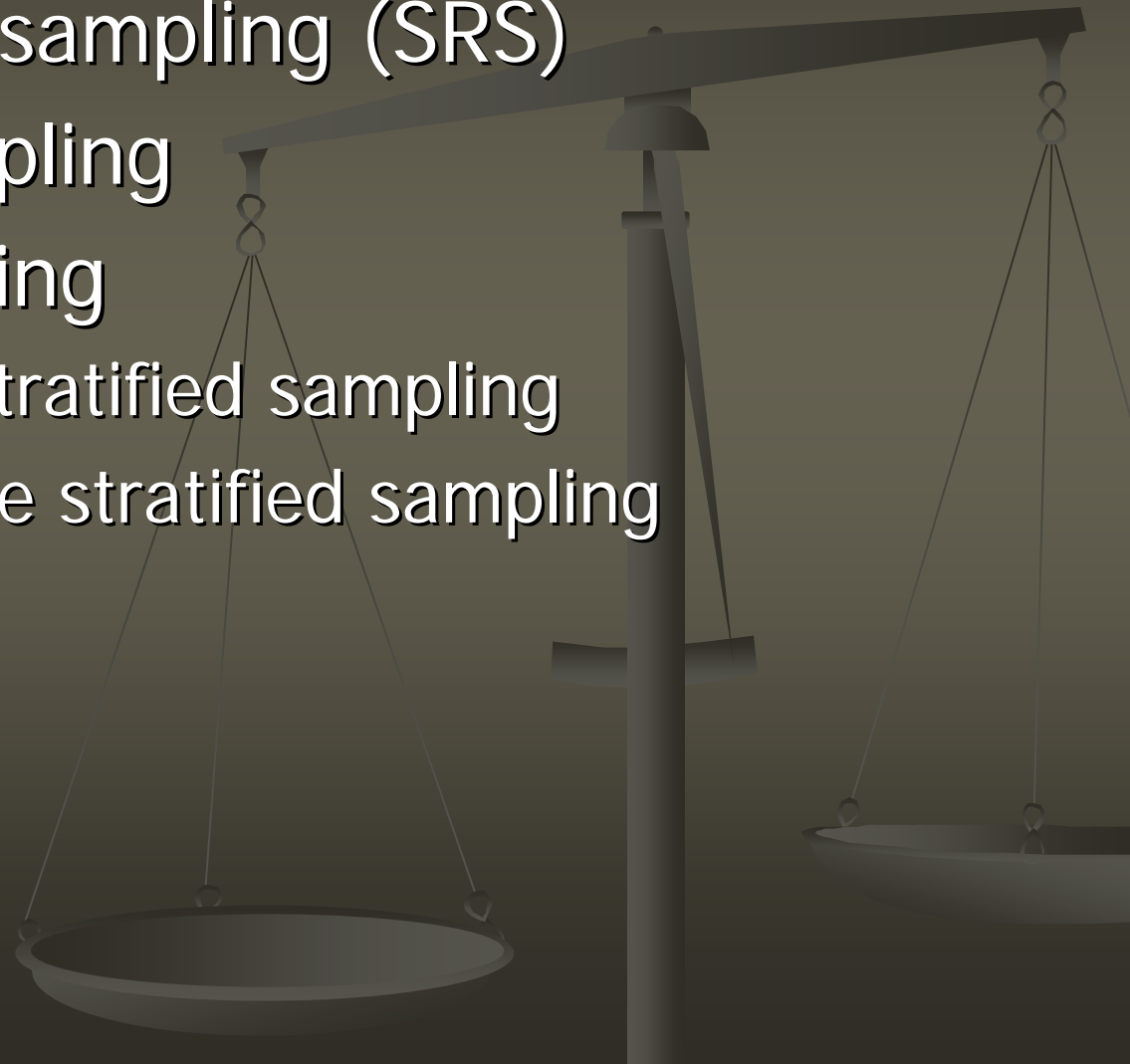
Probability Sampling

- Used when researchers want precise, statistical descriptions of large populations.
- A sample of individuals from a population must contain the same variations that exist in the population.

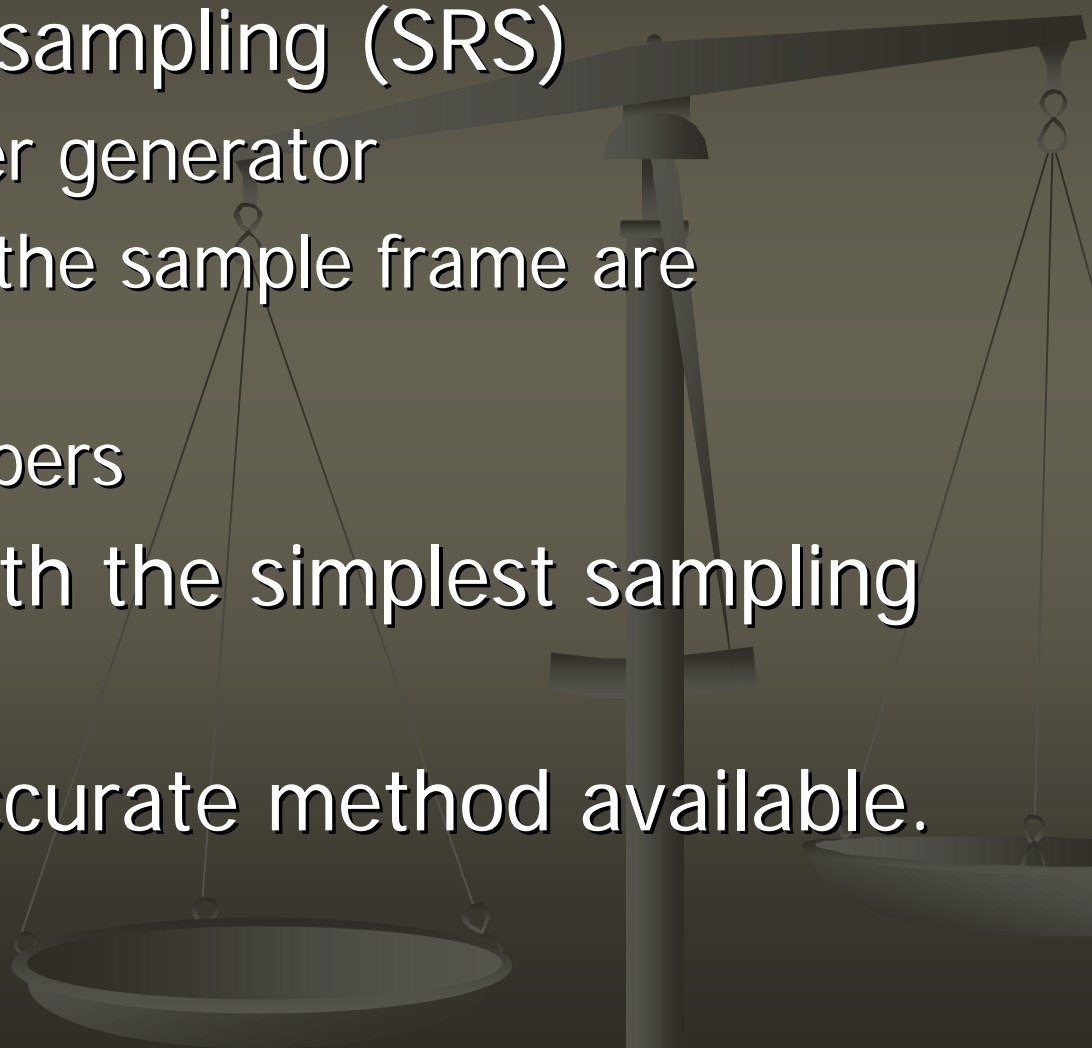


Types of Probability Sampling

- Simple random sampling (SRS)
- Systematic sampling
- Stratified sampling
 - Proportionate stratified sampling
 - Disproportionate stratified sampling



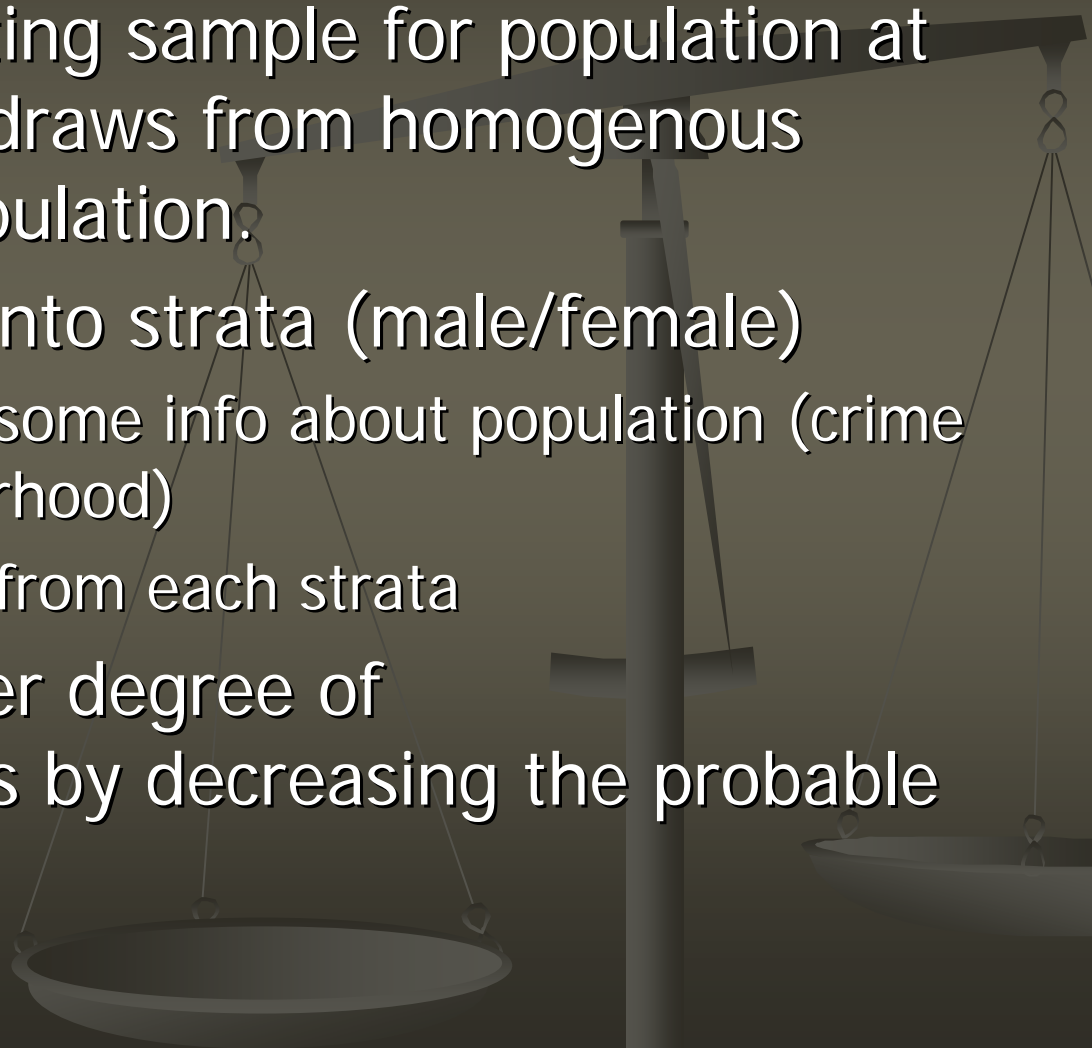
Types of Sampling Designs

- Simple random sampling (SRS)
 - Random number generator
 - All elements in the sample frame are numbered
 - Chose the numbers
 - Feasible only with the simplest sampling frame.
 - Not the most accurate method available.
- 

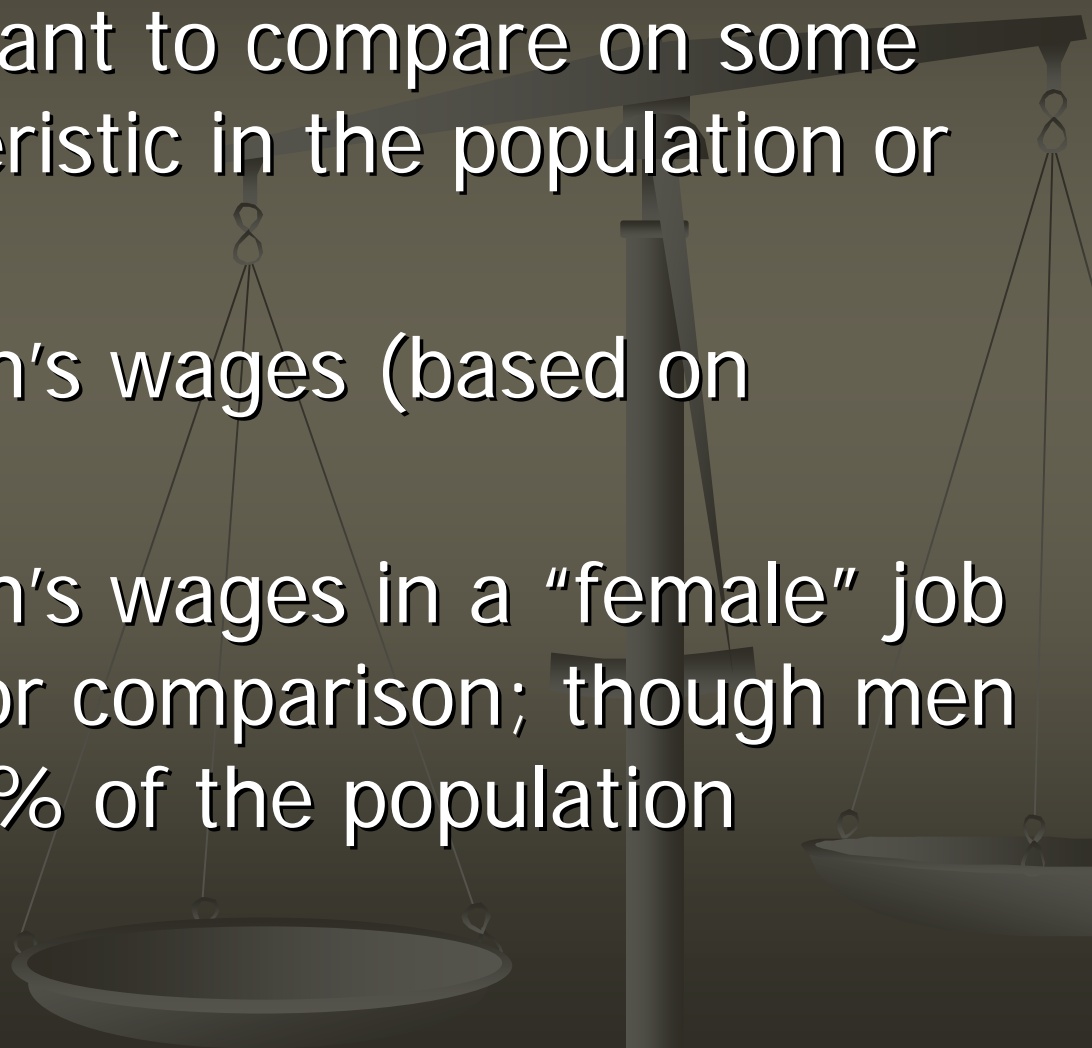
Systematic Sampling

- Three steps:
 - Compute sampling interval, N/n . Sampling interval = Z .
 - Ten Interviews with 70 Sociologists.
 - $Z=7$
 - If the sampling interval is Z , then randomly select one of the first Z cases.
 - Select one of the first 7 sociologists randomly.
 - After observing the first case, select every Z th case that follows that first case.
 - Every 7th sociologist until we get to 10 cases.

Stratified Sampling

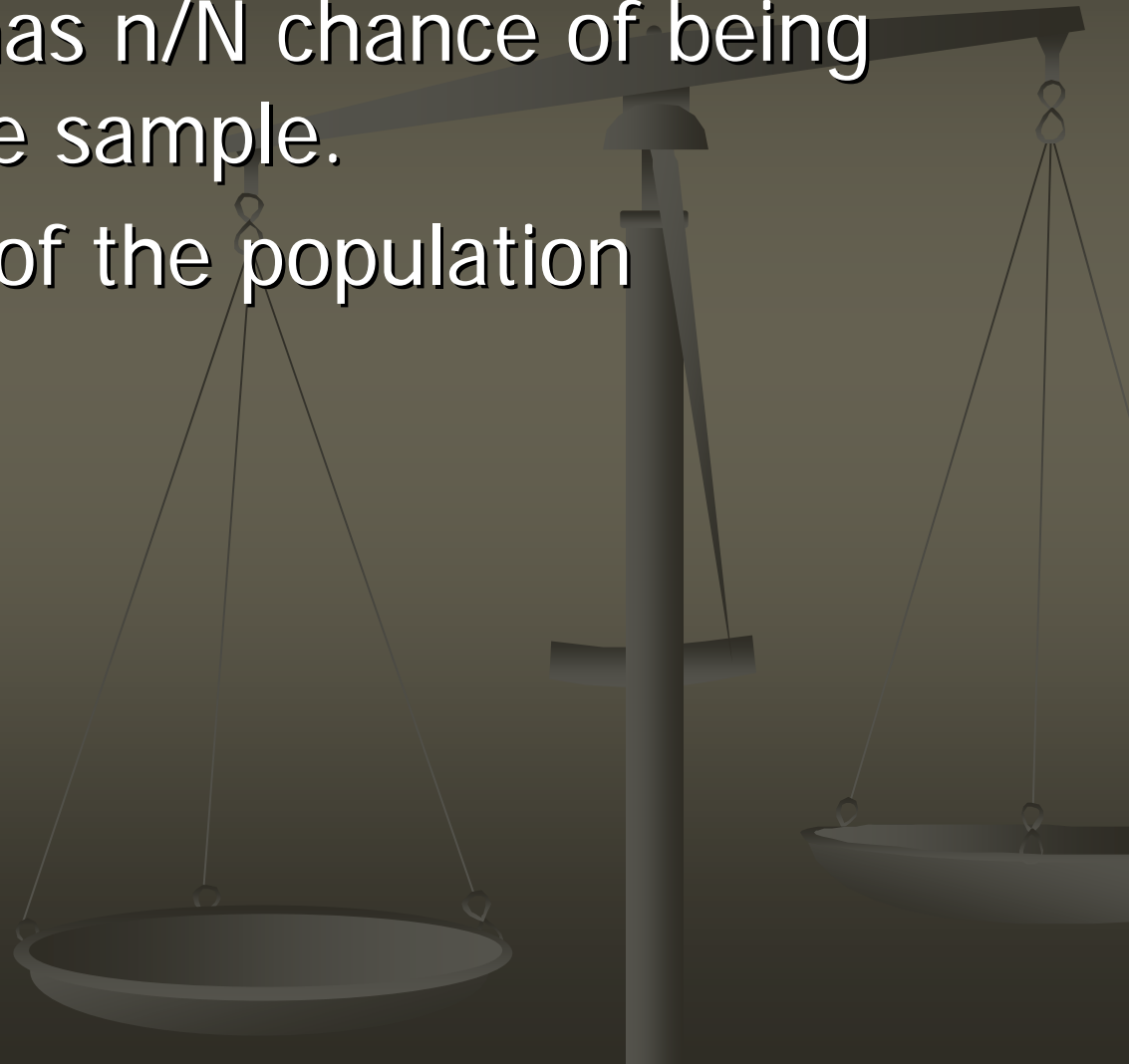
- Rather than selecting sample for population at large, researcher draws from homogenous subsets of the population.
 - Break population into strata (male/female)
 - We already know some info about population (crime report by neighborhood)
 - Sample randomly from each strata
 - Results in a greater degree of representativeness by decreasing the probable sampling error.
- 

Stratified sampling

- Use when we want to compare on some “basic” characteristic in the population or theoretically.
 - Men and women’s wages (based on population size)
 - Men and women’s wages in a “female” job sample 50/50 for comparison; though men may only be 10% of the population
- 

Proportionate stratified random sampling

- Each case still has n/N chance of being selected into the sample.
- Representative of the population



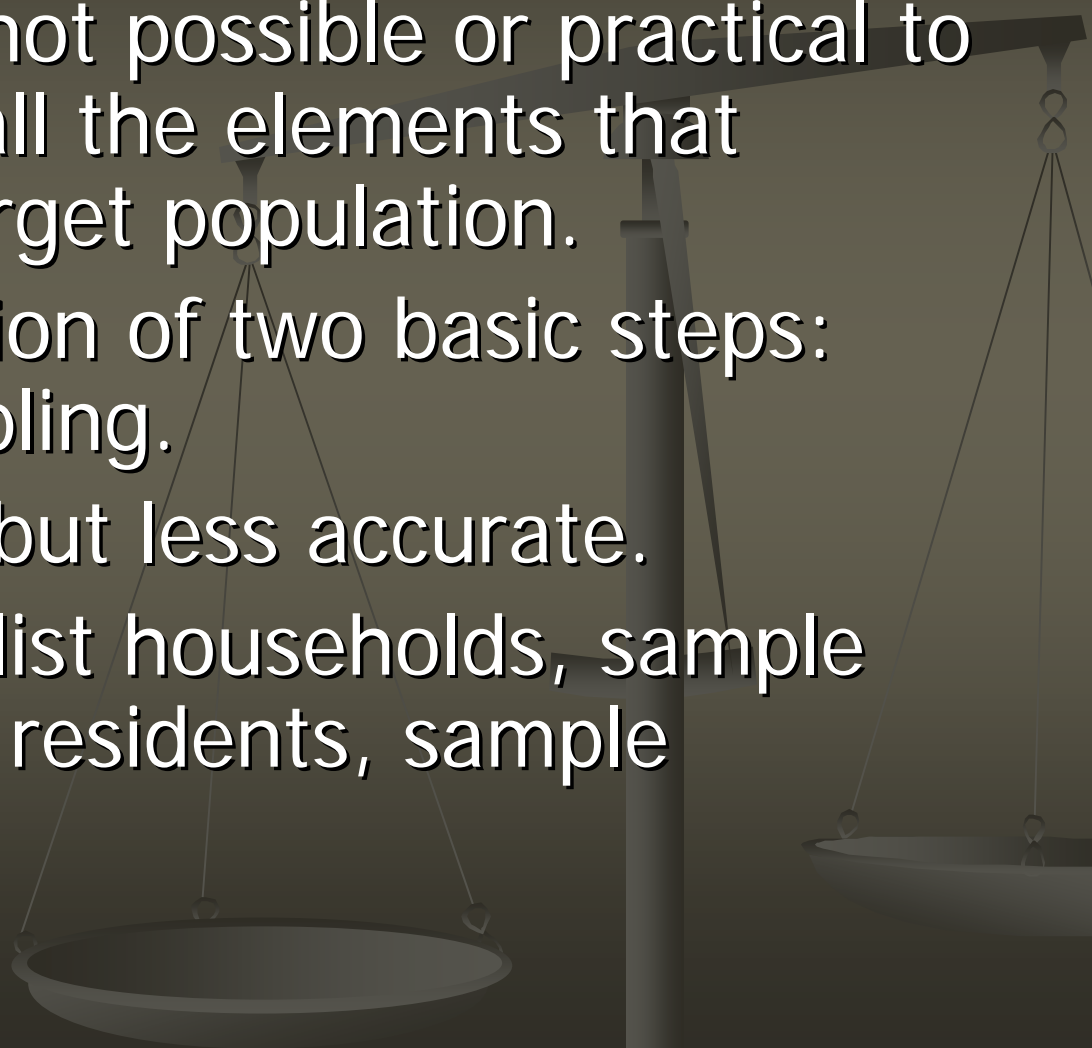
Proportionate stratified random sampling

- If we wanted to know something about race and education in the population. We are interested in comparing Whites/ African-Americans/ Asian-Americans and Native Americans.
- Lets say that Whites are 75% of the population, African-Americans 15%, Asian-Americans 8% and Native Americans 2%
- You would include this percent of each group in your study.

Disproportionate stratified random sample

- The proportion of each stratum that is included in the sample is intentionally different from what is in the population.
- Varies from the population
- Useful if we want information on specific subpopulations
 - American Indians are 1% of total population but 50% of your sample.

Multistage Cluster Sampling

- Used when it's not possible or practical to create a list of all the elements that compose the target population.
 - Involves repetition of two basic steps: listing and sampling.
 - Highly efficient but less accurate.
 - Sample blocks, list households, sample households, list residents, sample residents etc.
- 

Sources of bias and error



- Nonresponse error- who chooses to participate is different than the rest of the population
 - Distribute survey in school to study truancy
- Coverage error- the sampling frame does not accurately reflect the population
 - Telephone book to talk to college students

Sources of bias and error

- Sampling error-
Methodology/frame/sample done correctly still doesn't represent sample. Random error
- Probability sampling lets us know how "correct" we are

