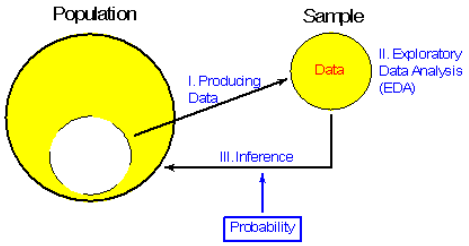
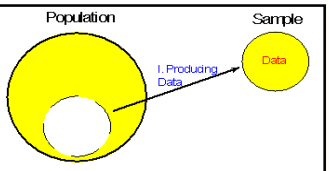
The Big Picture of Statistics



1

Sampling



The ***goal*** is to ***select a representative sample from the population*** for cost-effective data collection so that inferences can be drawn about the population it comes from.

2

Sampling

- But how can we choose a sample that we can trust to represent the population?
- A **sampling design** describes exactly how to choose a sample from the population.

3

Bad sampling methods--Bias

- The sample design is **biased** if it systematically favors certain outcomes.
- Example: consider a research project on attitudes toward sex. Collecting the data by publishing a questionnaire in a magazine and asking people to fill it out and send it in would produce a biased sample. People interested enough to spend their time and energy filling out and sending in the questionnaire are likely to have different attitudes toward sex than those not taking the time to fill out the questionnaire.

4

Bad sampling methods:

Convenience sampling:

Just ask whoever is around.

- "Man on the street" survey (cheap, convenient, often quite opinionated or emotional)
  - Ask about gun control or legalizing marijuana "on the street" in Berkeley, CA and in some small town in Idaho and you would probably get totally different answers.
- **Bias:** Opinions limited to individuals present

5

Bad sampling methods:

- **Voluntary response sampling:** **Individuals choose to be involved.** These samples are very susceptible to being biased because different people are motivated to respond or not. They are often called "public opinion polls" and are not considered valid or scientific.
- **Bias:** Sample design systematically favors a particular outcome.

6

Example of voluntary response sample:

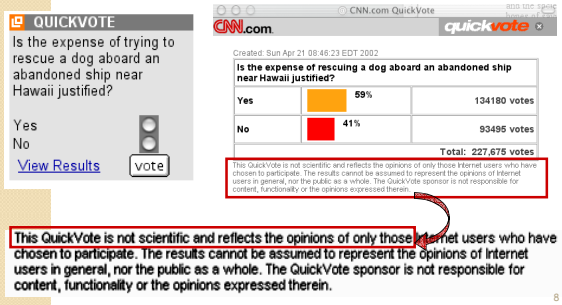
Ann Landers summarizing responses of readers: Seventy percent of (10,000) parents wrote in to say that having kids was not worth it—if they had to do it over again, they wouldn't.

Bias: Most letters to newspapers are written by disgruntled people. A random sample showed that 91% of parents WOULD have kids again.

7

On-line surveys:

Bias: People have to care enough about an issue to bother replying. This sample is probably a combination of people who hate "wasting the taxpayers' money" and "animal lovers."



This QuickVote is not scientific and reflects the opinions of only those internet users who have chosen to participate. The results cannot be assumed to represent the opinions of Internet users in general, nor the public as a whole. The QuickVote sponsor is not responsible for content, functionality or the opinions expressed therein.

8

Good sampling methods:

Probability or **random sampling:** Individuals are randomly selected. No one group should be over-represented.

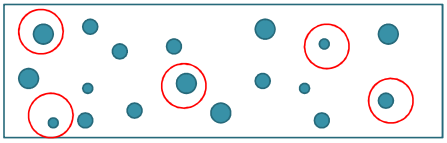
Sampling randomly gets rid of bias.

Random samples rely on the absolute objectivity of random numbers. There are books and tables of **random digits** available for random sampling.

9

Simple Random Sample (SRS)

- In SRS each member of the population has an equal chance to be selected.
- We can use a Random Number Table, or calculator to generate the random numbers to select our SRS.



10

Stratified Random Samples

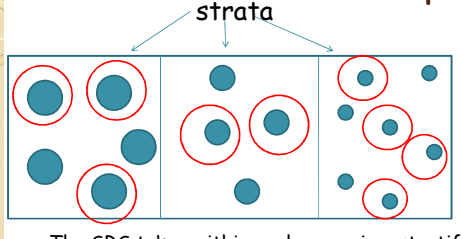
A **stratified random sample** is a series of SRS performed on subgroups of a given population. The subgroups (strata) are chosen to contain all the individuals with a certain characteristic.

For example:

- Divide the population of CSUN students into males and females.
- Divide the population of California by major ethnic group.

11

Stratified Random Sampling

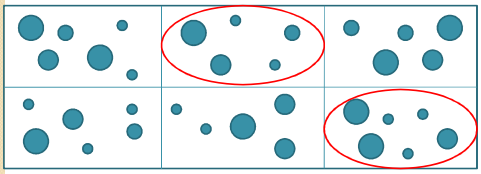


The SRS taken within each group in a stratified random sample **need not be of the same size.**

12

Other methods

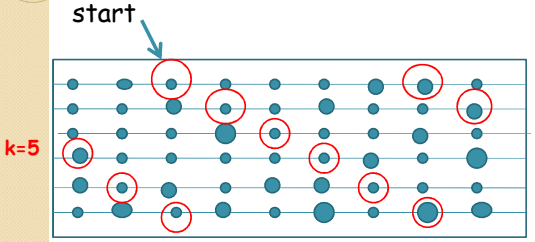
- **Cluster sample:** first select clusters at random, and then use **all** the individuals in the clusters as your sample.



13

Other methods

- **Systematic sampling:** "count off" your population by a number *k* with a random start

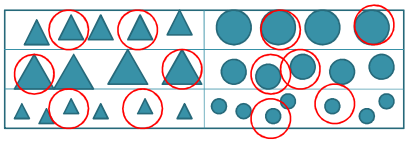


14

Other methods

Multistage samples use multiple stages of stratification. Data are obtained by taking an SRS for each substrata.

Example: Sampling both urban and rural areas, people in different ethnic groups within the urban and rural areas, and then individuals of different ethnicities within those strata.



15

Summary of good sampling methods:

- SRS
- Stratified Random Sampling
- Cluster Sampling
- Multistage Sampling
- Systematic Sampling with Random Start

16

Caution about sampling surveys

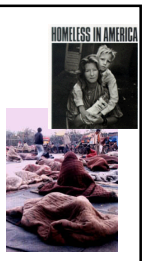
- **Nonresponse:** People who feel they have something to hide or who don't like their privacy being invaded probably won't answer. Yet they are part of the population.
- **Response bias:** Fancy term for lying when you think you should not tell the truth. Like if your family doctor asks: "How much do you drink?" Or a survey of female students asking: "How many men do you date per week?" People also simply forget and often give erroneous answers to questions about the past.
- **Wording effects:** Questions worded like "Do you agree that it is awful that…" are prompting you to give a particular response.

17

Undercoverage

Undercoverage occurs when parts of the population are left out in the process of choosing the sample.

Because the U.S. Census goes "house to house," homeless people are not represented. Illegal immigrants also avoid being counted.



Historically, clinical trials have avoided including women in their studies because of their periods and the chance of pregnancy. This means that medical treatments were not appropriately tested for women. This problem is slowly being recognized and addressed.

18

Learning about populations from samples

The techniques of inferential statistics allow us to draw inferences or conclusions about a population from a sample.

- Your estimate of the population is only as good as your sampling design → Work hard to eliminate biases.
- Your result from the sample is only an estimate—and if you randomly sampled again, you would probably get a somewhat different result.
- The bigger the sample the better.

19