Producing Data

IPS Chapter 3

- 3.1: Design of Experiments
- 3.2: Sampling Design
- 3.3: Toward Statistical Inference
- 3.4: Ethics

Producing Data

3.1 Design of Experiments

Producing Data

- Design of Experiments
- Sampling Design
- Toward Statistical Inference
- Ethics

Objectives (IPS Chapters 3.1)

Design of experiments

- Anecdotal and available data
- Comparative experiments
- Randomization
- Randomized comparative experiments
- Cautions about experimentation
- Matched pairs designs
- Block designs

Obtaining data

Available data are data that were produced in the past for some other purpose but that may help answer a present question inexpensively. The library and the Internet are sources of available data.

> Government statistical offices are the primary source for demographic, economic, and social data (visit the Fed-Stats site at www.fedstats.gov).

Beware of drawing conclusions from our own experience or hearsay. **Anecdotal evidence** is based on haphazardly selected individual cases, which we tend to remember because they are unusual in some way. They also may not be representative of any larger group of cases.

Some questions require data produced specifically to answer them. This leads to **designing** observational or experimental studies.

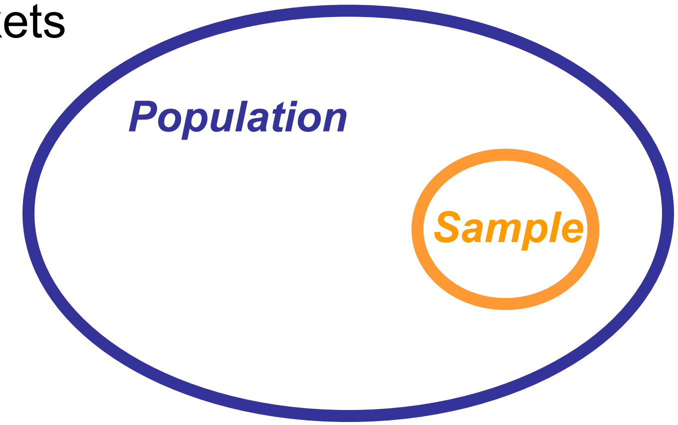
Population versus sample

- **Population:** The entire group of individuals in which we are interested but can't usually assess directly.

  Example: All humans, all working-age people in California, all crickets

- **Sample:** The part of the population we actually examine and for which we do have data.

  How well the sample represents the population depends on the sample design.



- A **parameter** is a number describing a characteristic of the population.

- A **statistic** is a number describing a characteristic of a sample.

Observational study: Record data on individuals without attempting to influence the responses.

Example: Based on observations you make in nature, you suspect that female crickets choose their mates on the basis of their health. → Observe health of male crickets that mated.

Experimental study: Deliberately impose a treatment on individuals and record their responses. Influential factors can be controlled.

Example: Deliberately infect some males with intestinal parasites and see whether females tend to choose healthy rather than ill males.

Observational studies vs. Experiments

- Observational studies are essential sources of data on a variety of topics. However, when our goal is to understand cause and effect, experiments are the only source of fully convincing data.
- Two variables are confounded when their effects on a response variable cannot be distinguished from each other.
- Example: If we simply observe cell phone use and brain cancer, any effect of radiation on the occurrence of brain cancer is confounded with lurking variables such as age, occupation, and place of residence.
- Well designed experiments take steps to defeat confounding.

Terminology

- The individuals in an experiment are the **experimental units**. If they are human, we call them **subjects**.
- In an experiment, we do something to the subject and measure the response. The "something" we do is a called a **treatment**, or **factor**.

  - The factor may be the administration of a drug.
  - One group of people may be placed on a diet/exercise program for six months (treatment), and their blood pressure (response variable) would be compared with that of people who did not diet or exercise.

- If the experiment involves giving two different doses of a drug, we say that we are testing two **levels** of the factor.
- A response to a treatment is **statistically significant** if it is larger than you would expect by chance (due to random variation among the subjects). We will learn how to determine this later.

In a study of sickle cell anemia, 150 patients were given the drug hydroxyurea, and 150 were given a placebo (dummy pill). The researchers counted the episodes of pain in each subject. Identify:

- The subjects • (patients, all 300)
- The factors / treatments • (hydroxyurea and placebo)
- And the response variable • (episodes of pain)

Comparative experiments

Experiments are comparative in nature: We compare the response to a treatment to:

- Another treatment
- No treatment (a control)
- A placebo
- Or any combination of the above

A **control** is a situation where no treatment is administered. It serves as a reference mark for an actual treatment (e.g., a group of subjects does not receive any drug or pill of any kind).

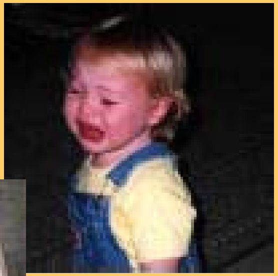
A **placebo** is a fake treatment, such as a sugar pill. This is to test the hypothesis that the response to the actual treatment is due to the actual treatment and not the subject's apparent treatment.

About the placebo effect

The "placebo effect" is an improvement in health not due to any treatment, but only to the patient's belief that he or she will improve.

- The "placebo effect" is not understood, but it is believed to have therapeutic results on up to a whopping 35% of patients.
- It can sometimes ease the symptoms of a variety of ills, from asthma to pain to high blood pressure, and even to heart attacks.
- An opposite, or "negative placebo effect," has been observed when patients believe their health will get worse.

The most famous, and maybe most powerful, placebo is the "kiss," blow, or hug—whatever your technique.



Unfortunately, the effect gradually disappears once children figure out that they sometimes get better without help and vice versa.

Designing "controlled" experiments

Sir Ronald Fisher—The "father of statistics"—was sent to Rothamsted Agricultural Station in the United Kingdom to evaluate the success of various fertilizer treatments.

Fisher found that the data from experiments that had been going on for decades was basically worthless because of poor experimental design.

- Fertilizer had been applied to a field one year and not another, in order to compare the yield of grain produced in the two years. BUT
  - It may have rained more or been sunnier during different years.
  - The seeds used may have differed between years as well.
- Or fertilizer was applied to one field and not to a nearby field in the same year. BUT
  - The fields might have had different soil, water, drainage, and history of previous use.

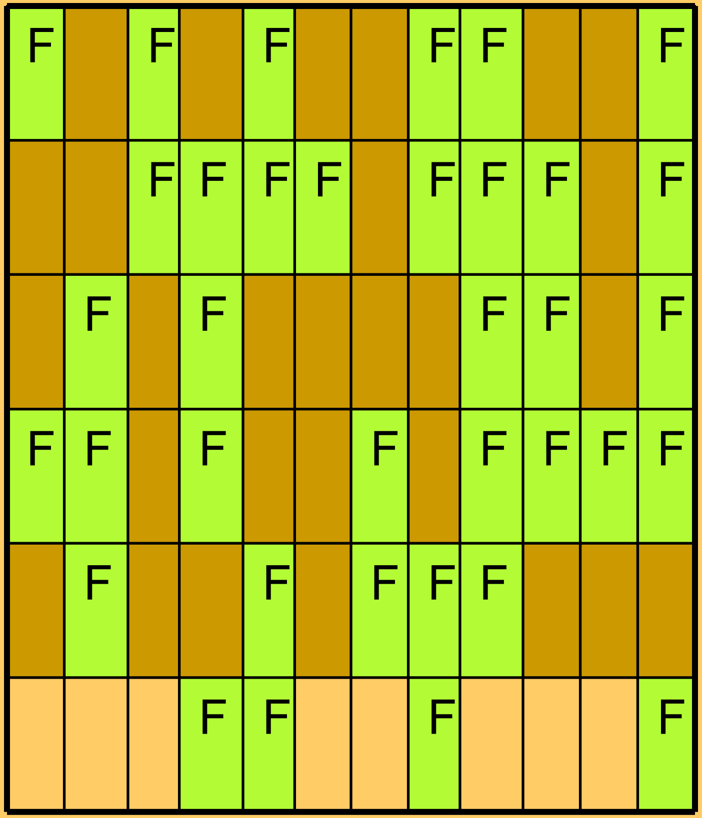
➔ Too many factors affecting the results were "uncontrolled."

Fisher's solution:

"Randomized comparative experiments"

- In the same field and same year, apply fertilizer to randomly spaced plots within the field. Analyze plants from similarly treated plots together.

- This minimizes the effect of variation within the field, in drainage and soil composition on yield, as well as controls for weather.



Randomization

One way to **randomize** an experiment is to rely on **random digits** to make choices in a neutral way. We can use a table of random digits (like Table B) or the random sampling function of a statistical software.

How to randomly choose n individuals from a group of N:

- We first label each of the N individuals with a number (typically from 1 to N, or 0 to $N-1$)
- A list of random digits is parsed into digits the same length as N (if N = 233, then its length is 3; if N = 18, its length is 2).
- The parsed list is read in sequence and the first n digits corresponding to a label in our group of N are selected.
- The n individuals within these labels constitute our selection.

Using Table B

We need to randomly select five students from a class of 20.

1. Since the class is of 20 people, list and number all members as 01,02,…20.

2. The number 20 is two digits long, so parse the list of random digits into numbers that are two digits long. Here we chose to start with line 103 for no particular reason.

TABLE B Random digits

Line								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893

45 46 77 17 09 77 55 80 00 95 32 86 32 94 85 82 22 69 00 56

45 46 77 **17** **09** 77 55 80 00 95 32 86 32 94 85 82 22 69 00 56

52 71 **13** 88 89 93 **07** 46 **02** …

4. Randomly choose five students by reading through the list of two-digit random numbers, starting with line 103 and on.
5. The first five random numbers that match the numbers assigned to students make our selection.

The first individual selected is Ramon, number 17. Then Henry (9, or 09). That's all we can get from line 103.

We then move on to line 104. The next three to be selected are Moe, George, and Amy (13, 07, and 02).

- *Remember that 1 is 01, 2 is 02, etc.*
- *If you were to hit 17 again before getting five people, don't sample Ramon twice—just keep going.*

01 Alison
02 Amy
03 Brigitte
04 Darwin
05 Emily
06 Fernando
07 George
08 Harry
09 Henry
10 John
11 Kate
12 Max
13 Moe
14 Nancy
15 Ned
16 Paul
17 Ramon
18 Rupert
19 Tom
20 Victoria

Principles of Experimental Design

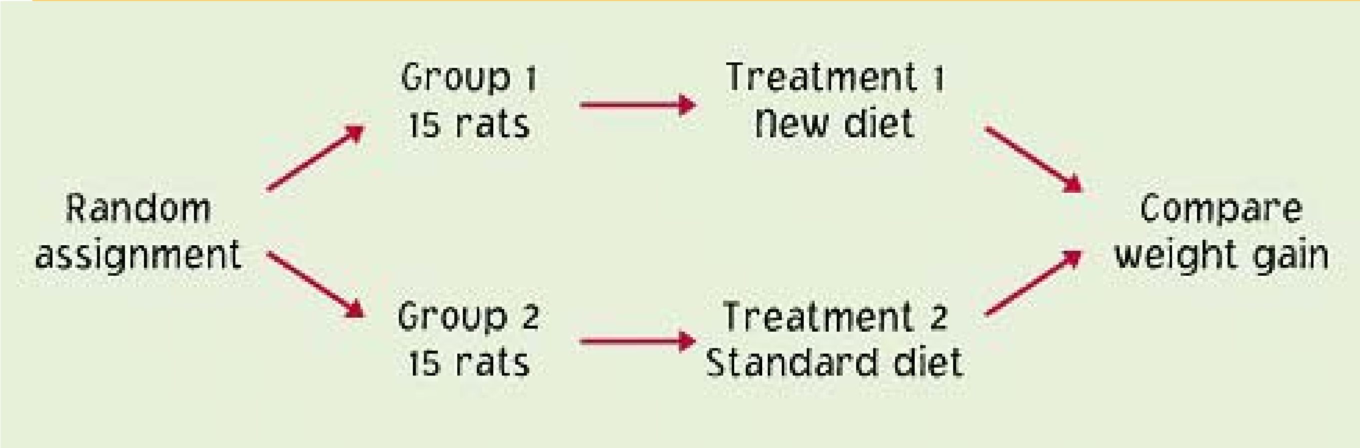
Three big ideas of experimental design:

- Control the effects of lurking variables on the response, simply by comparing two or more treatments.
- Randomize – use impersonal chance to assign subjects to treatments.
- Replicate each treatment on enough subjects to reduce chance variation in the results.

Statistical Significance: An observed effect so large that it would rarely occur by chance is called statistically significant.

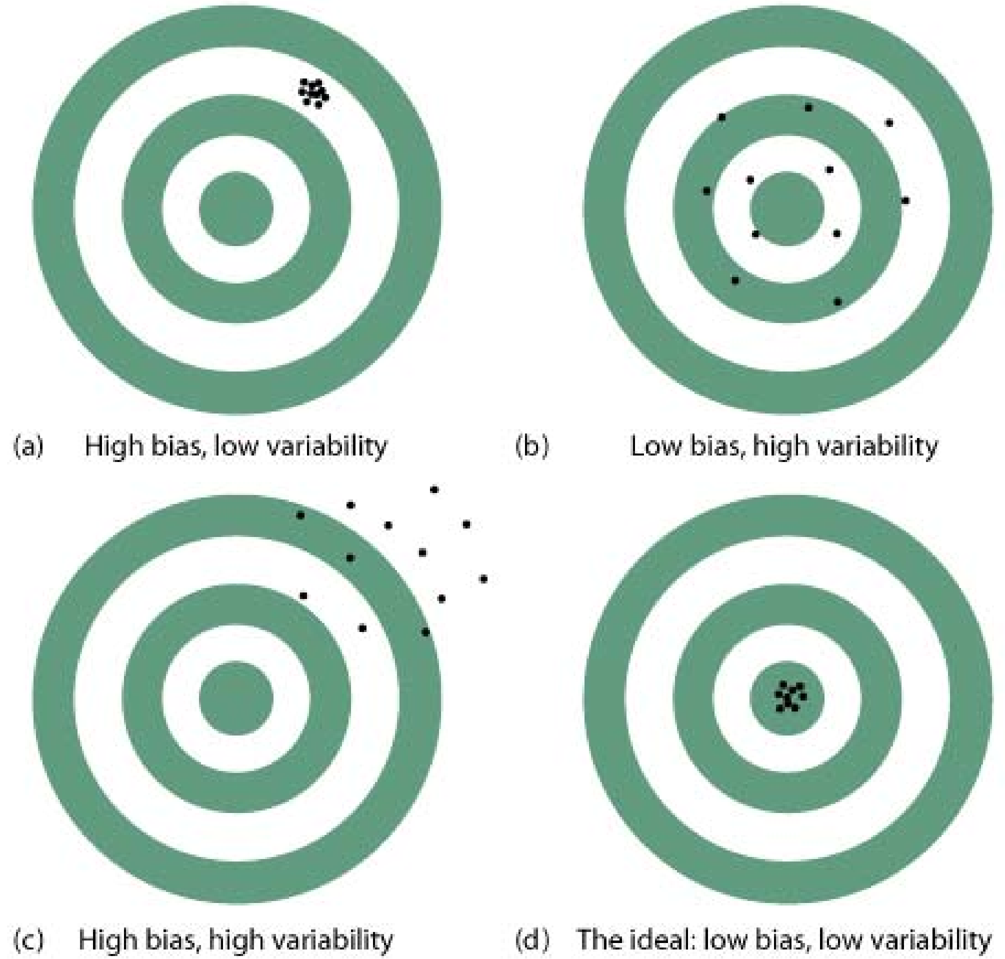
Completely randomized designs

Completely randomized experimental designs: Individuals are randomly assigned to groups, then the groups are randomly assigned to treatments.



Caution about experimentation

The design of a study is **biased** if it systematically favors certain outcomes.



The best way to exclude biases from an experiment is to **randomize** the design. Both the individuals and treatments are assigned randomly.

Other ways to remove bias:

A **double-blind** experiment is one in which neither the subjects nor the experimenter know which individuals got which treatment until the experiment is completed. The goal is to avoid forms of placebo effects and biases based on interpretation.

The best way to make sure your conclusions are robust is to **replicate** your experiment—do it over. Replication ensures that particular results are not due to uncontrolled factors or errors of manipulation.

Lack of realism

Lack of realism is a serious weakness of experimentation. The subjects or treatments or setting of an experiment may not realistically duplicate the conditions we really want to study. In that case, we cannot generalize about the conclusions of the experiment.

Is the treatment appropriate for the response you want to study?

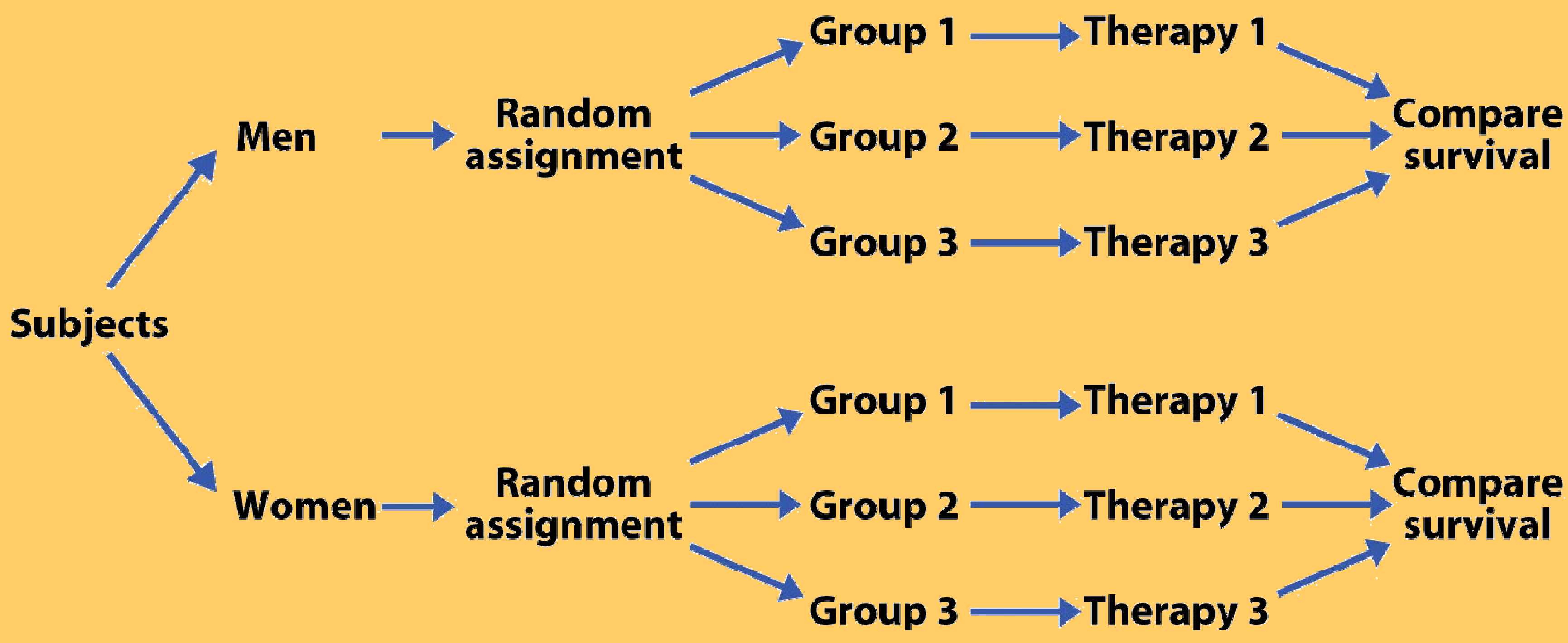
- Is studying the effects of eating red meat on cholesterol values in a group of middle aged men a realistic way to study factors affecting heart disease problems in humans?

- What about studying the effects of hair spray on rats to determine what will happen to women with big hair?

Block designs

In a **block,** or **stratified,** design, subjects are divided into groups, or blocks, prior to experiments, to test hypotheses about differences between the groups.

The blocking, or **stratification,** here is by gender.



Matched pairs designs

Matched pairs: Choose pairs of subjects that are closely matched—e.g., same sex, height, weight, age, and race. Within each pair, randomly assign who will receive which treatment.

It is also possible to just use a single person, and give the two treatments to this person over time in random order. In this case, the “matched pair” is just the same person at different points in time.

The most closely matched pair studies use identical twins.

What experimental design?

A researcher wants to see if there is a significant difference in resting pulse rates between men and women. Twenty-eight men and 24 women had their pulse rate measured at rest in the lab.

- One factor, two levels (male and female)
- Stratified random sample (by gender)

Many dairy cows now receive injections of BST, a hormone intended to spur greater milk production. The milk production of 60 Ayrshire dairy cows was recorded before and after they received a first injection of BST.

- SRS of 60 cows
- Matched pair design (before and after)

Producing Data

3.2 Sampling designs
3.3 Toward statistical inference

Objectives

3.2 and 3.3 Sampling designs; Toward statistical inference

- Sampling methods
- Simple random samples
- Stratified samples
- Caution about sampling surveys
- Population versus sample
- Toward statistical inference
- Sampling variability
- Capture–recapture sampling

Sampling methods

Convenience sampling: Just ask whoever is around.

- Example: “Man on the street” survey (cheap, convenient, often quite opinionated, or emotional => now very popular with TV “journalism”)

- Which men, and on which street?

  - Ask about gun control or legalizing marijuana “on the street” in Berkeley or in some small town in Idaho and you would probably get totally different answers.
  - Even within an area, answers would probably differ if you did the survey outside a high school or a country western bar.

- **Bias:** Opinions limited to individuals present.

Voluntary Response Sampling:

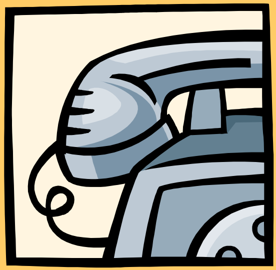
- Individuals choose to be involved. These samples are very susceptible to being biased because different people are motivated to respond or not. Often called “public opinion polls,” these are not considered valid or scientific.

- **Bias:** Sample design systematically favors a particular outcome.

Ann Landers summarizing responses of readers

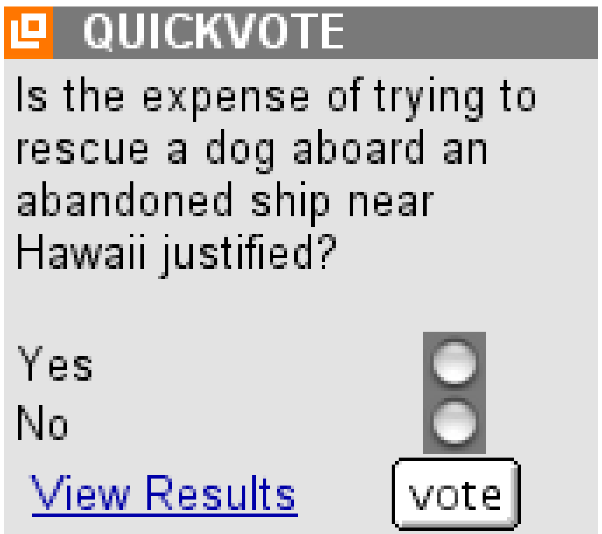
70% of (10,000) parents wrote in to say that having kids was not worth it—if they had to do it over again, they wouldn’t.

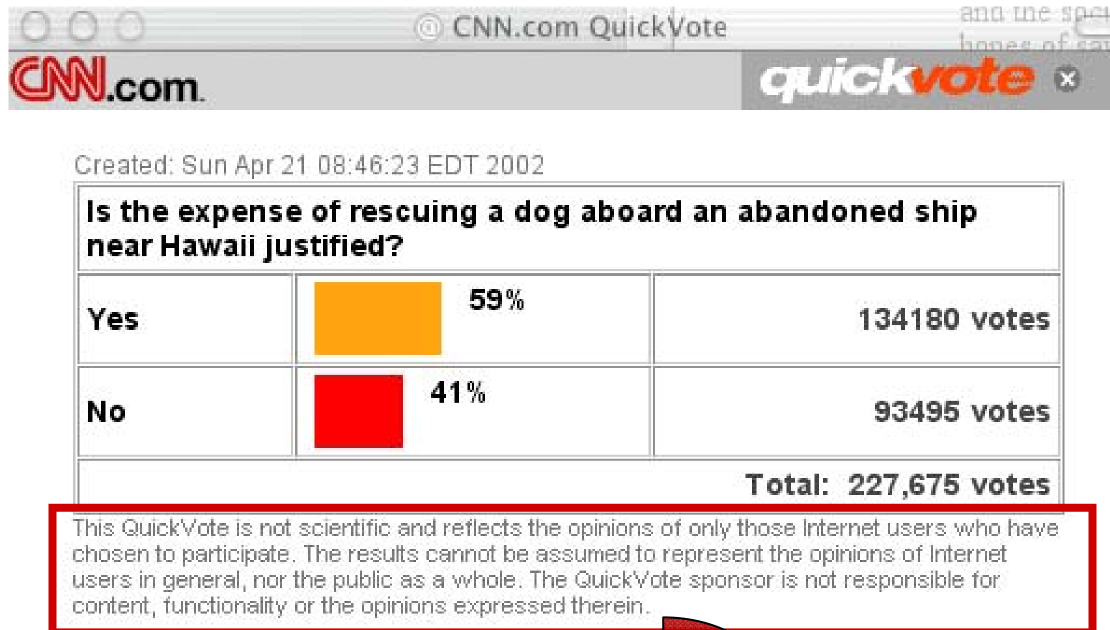
Bias: Most letters to newspapers are written by disgruntled people. A random sample showed that 91% of parents WOULD have kids again.



CNN 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.

In contrast :

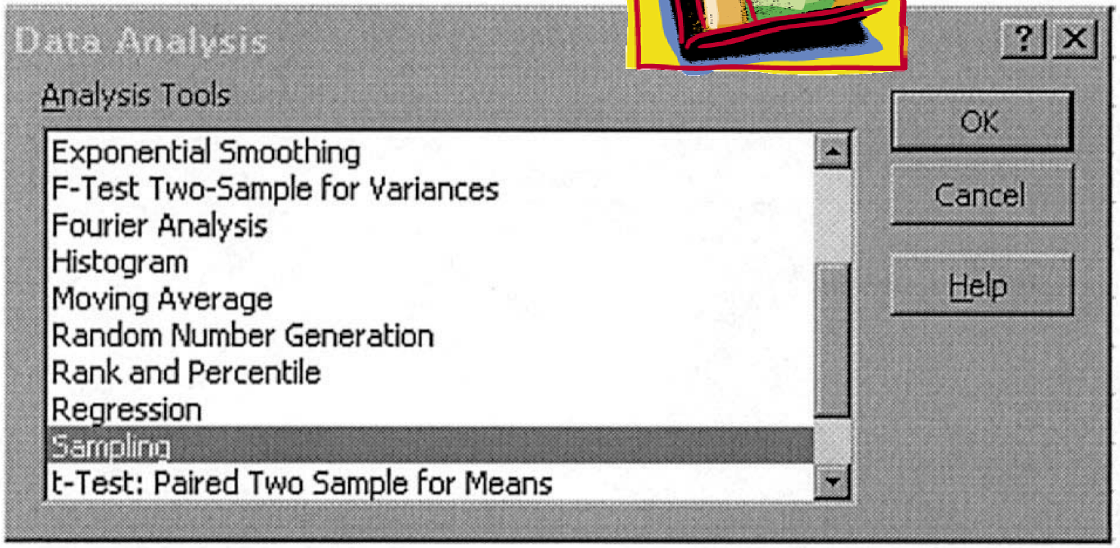
Probability or **random sampling**:

- 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 tables and books of random digits available for random sampling.

Statistical software can generate random digits (e.g., Excel “=random()”).



Simple random samples

A **Simple Random Sample (SRS)** is made of randomly selected individuals. Each individual in the population has the same probability of being in the sample. All possible samples of size *n* have the same chance of being drawn.

The simplest way to use chance to select a sample is to place names in a hat (the population) and draw out a handful (the sample).

Stratified samples

There is a slightly more complex form of random sampling:

A **stratified random sample** is essentially a series of SRSs performed on subgroups of a given population. The subgroups are chosen to contain all the individuals with a certain characteristic. For example:

- Divide the population of UCI students into males and females.
- Divide the population of California by major ethnic group.
- Divide the counties in America as either urban or rural based on criteria of population density.

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

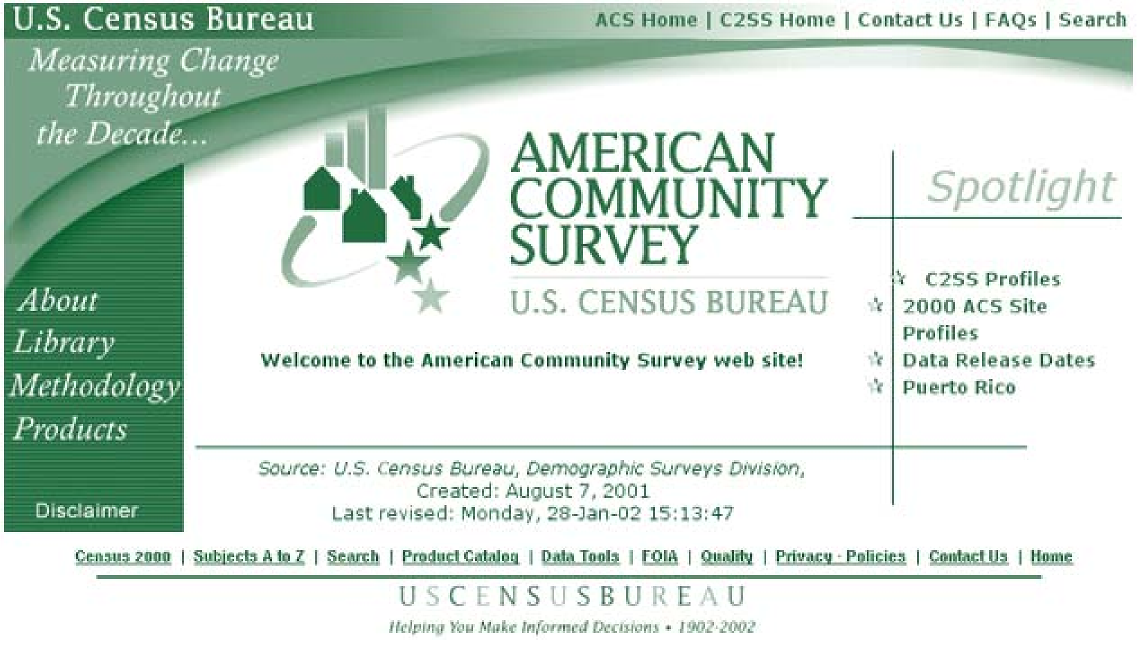
- A stratified random sample of 100 male and 150 female UCI students
- A stratified random sample of a total of 100 Californians, representing proportionately the major ethnic groups

Multistage samples use multiple stages of stratification. They are often used by the government to obtain information about the U.S. population.

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

Data are obtained by taking an SRS for each substrata.

Statistical analysis for multistage samples is more complex than for an SRS.



Caution about sampling surveys

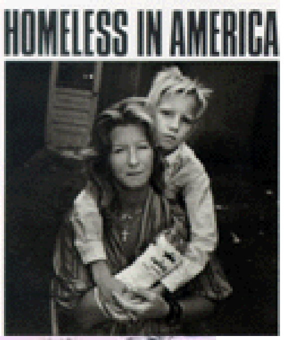
- **Nonresponse:** People who feel they have something to hide or who don't like their privacy being invaded probably won't answer. Yet they are part of the population.

- **Response bias:** Fancy term for lying when you think you should not tell the truth, or forgetting. This is particularly important when the questions are very personal (e.g., "How much do you drink?") or related to the past.

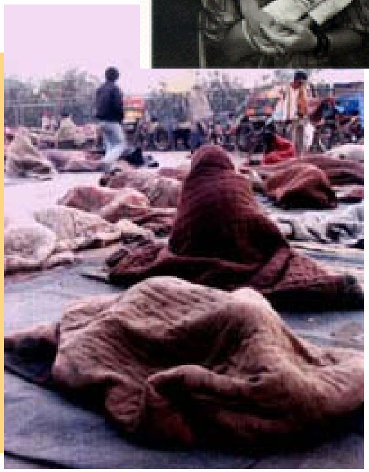
- **Wording effects:** Questions worded like "Do you agree that it is awful that…" are prompting you to give a particular response.

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. Geographical districts with a lack of coverage tend to be poor. Representatives from wealthy areas typically oppose statistical adjustment of the census.



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.

1. To assess the opinion of students at the Ohio State University about campus safety, a reporter interviews 15 students he meets walking on the campus late at night who are willing to give their opinion.

→ What is the sample here? What is the population? Why?

- All those students walking on campus late at night
- All students at universities with safety issues
- The 15 students interviewed
- All students approached by the reporter

2. An SRS of 1,200 adult Americans is selected and asked: “In light of the huge national deficit, should the government at this time spend additional money to establish a national system of health insurance?“ Thirty-nine percent of those responding answered yes.

→ What can you say about this survey?

- The sampling process is sound, but the wording is biased. The results probably understate the percentage of people who do favor a system of national health insurance.

Should you trust the results of the first survey? Of the second? Why?

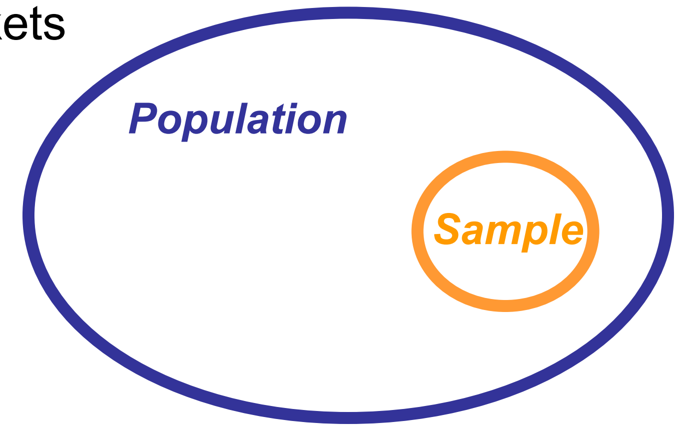
Population versus sample

- **Population:** The entire group of individuals in which we are interested but can't usually assess directly.

    Example: All humans, all working-age people in California, all crickets

- **Sample:** The part of the population we actually examine and for which we do have data.

    How well the sample represents the population depends on the sample design.

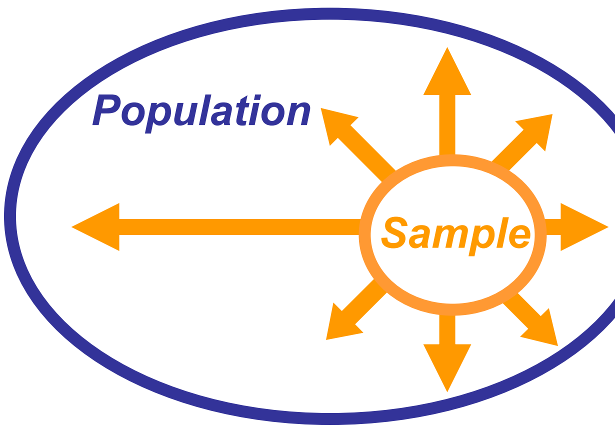


- A **parameter** is a number describing a characteristic of the population.
- A **statistic** is a number describing a characteristic of a sample.

Toward statistical inference

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

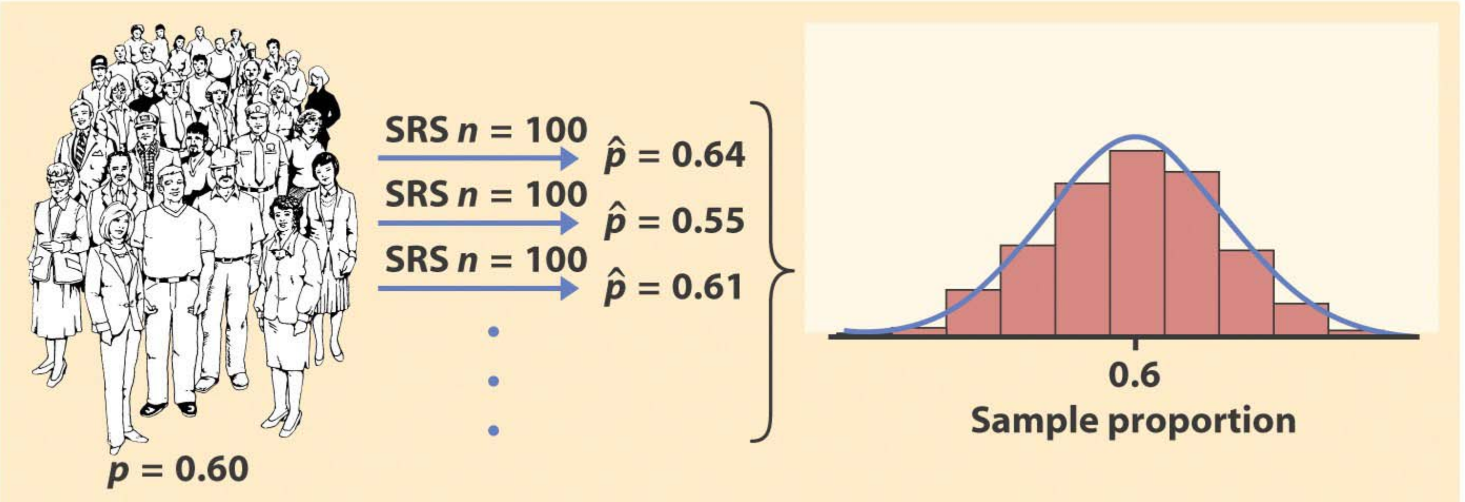
- Your estimate of the population is only as good as your sampling design. → Work hard to eliminate biases.
- Your sample is only an estimate—and if you randomly sampled again you would probably get a somewhat different result.
- The bigger the sample the better.



Sampling variability

Each time we take a random sample from a population, we are likely to get a different set of individuals and a calculate a different statistic. This is called **sampling variability**.

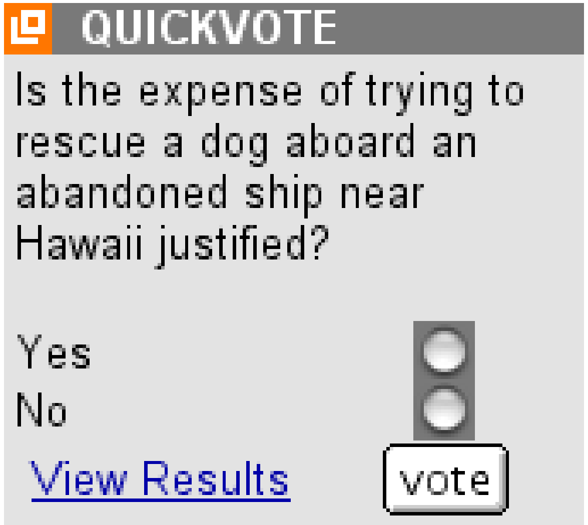
The good news is that, if we take lots of random samples of the same size from a given population, the variation from sample to sample—the **sampling distribution**—will follow a predictable pattern. All of statistical inference is based on this knowledge.



The **variability of a statistic** is described by the spread of its sampling distribution. This spread depends on the sampling design and the sample size *n,* with larger sample sizes leading to lower variability.

→ Statistics from large samples are almost always close estimates of the true population parameter. However, this only applies to random samples.

Remember the "QuickVote" online surveys. They are worthless no matter how many people participate because they use a voluntary sampling design and not random sampling.



Practical note

Large samples are not always attainable.

- Sometimes the cost, difficulty, or preciousness of what is studied limits drastically any possible sample size.
- Blood samples/biopsies: No more than a handful of repetitions acceptable. We often even make do with just one.
- Opinion polls have a limited sample size due to time and cost of operation. During election times though, sample sizes are increased for better accuracy.

Capture–recapture sampling

Repeated sampling can be used to estimate the size N of a population (e.g., animals). Here is an example of **capture-recapture sampling**:

What is the number of a bird species (least flycatcher) migrating along a major route? Least flycatchers are caught in nets, tagged, and released. The following year, the birds are caught again and the numbers tagged versus not tagged recorded. The proportion of tagged birds in the sample should be a reasonable estimate of the proportion of tagged birds in the population.

	Year 1	Year 2
Sample size	200	120
Number tagged		12

If N is the unknown total number of least flycatchers, we should have approximately

$$12/120 = 200/N$$

$$\Rightarrow N = 200 \times 120/12 = 2000$$

This works well if both samples are SRSs from the population and the population remains unchanged between samples. In practice, however, some of the birds tagged last year died before this year's migration.

Producing Data

3.4 Ethics

Objectives

3.4 Ethics

- Institutional review boards
- Informed consent
- Confidentiality
- Clinical trials
- Behavioral and social science experiments

Institutional review boards

- The organization that carries out the study must have an institutional review board that reviews all planned studies in advance in order to protect the subjects from possible harm.

- The purpose of an institutional review board is "to protect the rights and welfare of human subjects (including patients) recruited to participate in research activities"

- The institutional review board:
    - reviews the plan of study
    - can require changes
    - reviews the consent form
    - monitors progress at least once a year

Informed consent

- All subjects must give their informed consent before data are collected.
- Subjects must be informed in advance about the nature of a study and any risk of harm it might bring.
- Subjects must then consent in writing.
- Who can't give informed consent?
  - prison inmates
  - very young children
  - people with mental disorders

Confidentiality

- All individual data must be kept confidential. Only statistical summaries may be made public.
- Confidentiality is not the same as anonymity. Anonymity prevents follow-ups to improve non-response or inform subjects of results.
- Separate the identity of the subjects from the rest of the data immediately!
- Example: Citizens are required to give information to the government (tax returns, social security contributions). Some people feel that individuals should be able to forbid any other use of their data, even with all identification removed.

Clinical trials

- Clinical trials study the effectiveness of medical treatments on actual patients—these treatments can harm as well as heal.

- Points for a discussion:
  - Randomized comparative experiments are the only way to see the true effects of new treatments.
  - Most benefits of clinical trials go to future patients. We must balance future benefits against present risks.
  - The interests of the subject must always prevail over the interests of science and society.

- In the 1930s, the Public Health Service Tuskegee study recruited 399 poor blacks with syphilis and 201 without the disease in order to observe how syphilis progressed without treatment. The Public Health Service prevented any treatment until word leaked out and forced an end to the study in the 1970s.

Behavioral and social science experiments

- Many behavioral experiments rely on hiding the true purpose of the study.
- Subjects would change their behavior if told in advance what investigators were looking for.
- The “Ethical Principles” of the American Psychological Association require consent unless a study merely observes behavior in a public space.

Alternate Slide

The following slide offers alternate software output data and examples for this presentation.

In contrast:

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 tables and books of random digits available for random sampling.

Statistical software can generate random digits
(e.g., *JMP SE*:

Cols ➔ Formula ➔ Random ➔ Col Shuffle).