Lecture 5

Collecting Data

Data can arise from many sources.

Definition: **Available data** are data that were produced in the past for some other purpose but that may help answer a present question.

However, often we need to produce data to answer specific questions.

Observation vs Experiment:

- In an **observational study** we observe individuals and measure variables of interest but do not attempt to influence the responses.

retrospective   prospective

- In an **experiment** we deliberately impose some treatment on individuals and we observe their responses.

best expt: Comparative, randomized, double-blind, placebo-controlled

The best way to see the effects of a change is to do an **intervention** – where we actually impose the change.

Example: A study of child care enrolled 1364 infants in 1991 and planned to follow them through their sixth year in school. Twelve years later, the researchers published an article finding that «the more time children spent in child care from birth to age four-and-a-half, the more adults tended to rate them, both at age four-and-a-half and at kindergarten, as less likely to get along with others, as more assertive, as disobedient, and as aggressive.»

Question: Is there a cause and effect relationship?

prospective

Example above describes an observational study. Parents made all child care decisions and the study did not attempt to influence them.

The study cannot prove whether spending more time in child care causes children to have more problem behaviors. Perhaps employed parents who use child care are under stress and the children react to their parents' stress. Perhaps single parents are more likely to use child care, etc.

We say that the effect of child care on behaviour is **confounded** with other characteristics of families who use child care.

Experiments help isolate or control the confounding variables.

Design of Experiments

Definition: The individuals on which the experiment is done are the **experimental units**.

When the units are human beings, they are called **subjects**.

A specific experimental condition applied to the units is called a **treatment**.

The explanatory variables in an experiment are often called **factors**. Each treatment is formed by combining a specific value (called a **level**) of each of the factors.

Example: Are smaller classes better? The Tennessee STAR program was an experiment on the effect of class size. It has been called «one of the most important educational investigations ever carried out.» The **subjects** were 6385 students who were beginning kindergarten.

Each student was assigned to one of three **treatments**:

- regular class (22 to 25 students) with one teacher,
- regular class with a teacher and a full-time teacher's aide, and
- small class (13 to 17 students).

These treatments are **levels** of a single **factor**, the *type of class*.

The students stayed in the same type of class for four years, then all returned to regular classes. In later years, students from small classes had higher scores on standard tests, were less likely to fail a grade, had better high school grades, and so on.

Note: Experiments can give good evidence for causation.

Example: What are the effects of repeated exposure to an advertising message? The answer may depend both on the length of the ad and on how often it is repeated. An experiment investigated this question using undergraduate students as subjects. All subjects viewed a 40-min television program that included ads for a digital camera. Some subjects saw a 30-sec commercial; others, a 90-sec version. The same commercial was shown either 1, 3, or 5 times during the programs.

two factors:
- length: 30 sec, 90 sec (levels)
- repetition: 1, 3, 5 (levels)

Factor B (repetition)

Factor A length	1 time	3 times	5 times
30 sec	1st	2nd	3rd
90 sec	4th	5th	6th

treatments are all possible combinations of levels. There are $2 \cdot 3 = 6$ treatments

Comparative Experiments

Example: «Gastric freezing» was a treatment for ulcers in the upper intestine. The patient swallows a deflated balloon with tubes attached, then a refrigerated liquid is pumped through the balloon for an hour. The idea is that cooling the stomach will reduce its production of acid and so relieve ulcers. The treatment was safe and easy and was widely used for several years. The design of the experiment was

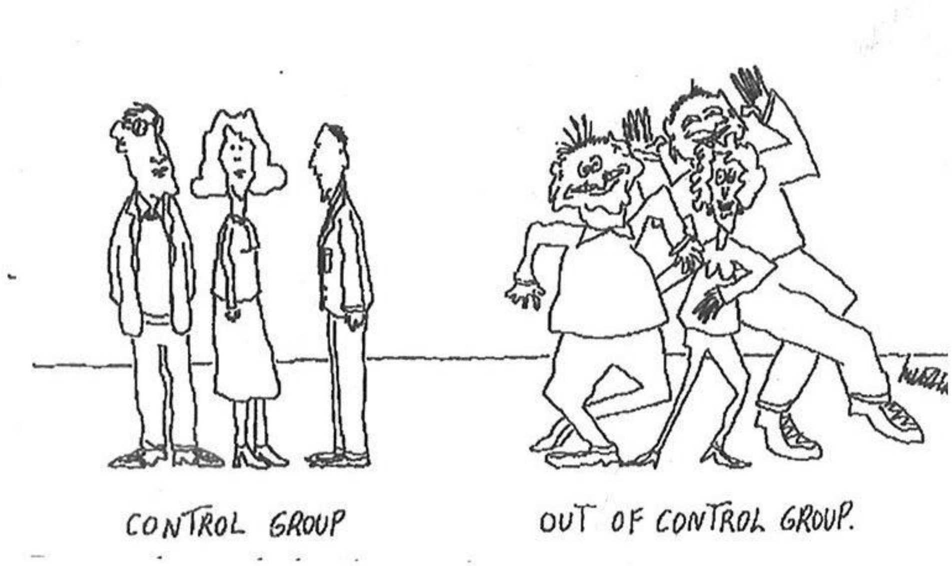
$$\text{Gastric freezing} \rightarrow \text{Observe pain relief}$$

A later experiment divided ulcer patients into two groups:

- One group was treated by gastric freezing as before.
- The other group received a **placebo** (a dummy treatment) in which the liquid in the balloon was at body temperature rather than freezing.

The results: 34% of the 82 patients in the treatment group improved, but so did 38% of the 78 patients in the placebo group. This and other properly designed experiments showed that gastric freezing was no better than a placebo, and its use was abandoned.

The group of patients who received a placebo is called **a control group**, because it enables us to control the effects of outside variables on the outcome.



Uncontrolled experiments can be dominated by such influences as the details of the experimental arrangement, the selection of subjects, and the placebo effect (the response to a dummy treatment). The result is often **biased**.

<u>Definition</u>: The design of a study is **biased** if it systematically favors certain outcomes.

Question: How can we assign experimental units to treatments in a way that is fair to all of the treatments?

Answer:

Rely on chance to make an assignment that

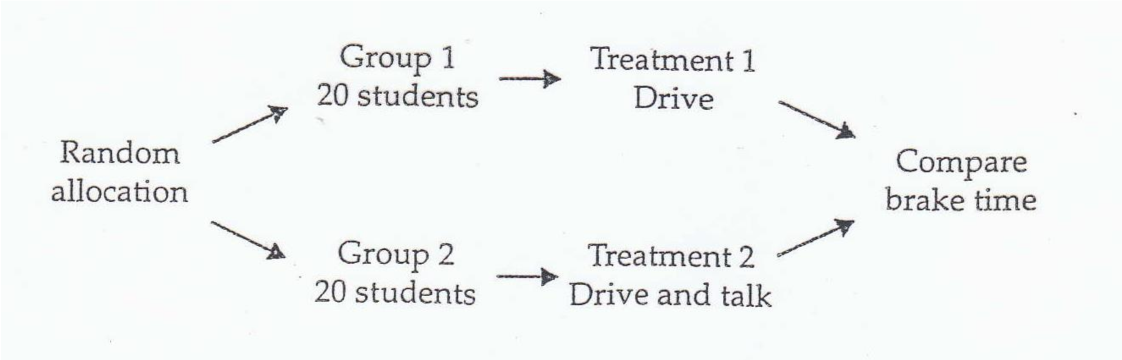
- does not depend on any characteristic of the experimental units
- does not rely on the judgement of the experimenter.

Randomization

<u>Example</u>: Does talking on a hand-free cell phone distract drivers?

Undergraduate students «drove» in a high-fidelity driving simulator equipped with a hand-free cell phone.

The car ahead brakes: how quickly does the subject respond? Twenty students (the **control group**) simply drove. Another 20 (the **experimental group**) talked on the cell phone while driving.



The use of chance to divide experimental units into groups is called **randomization**.

The logic behind the randomized comparative design is as follows:

- Randomization produces two groups of subjects that we expect to be similar in all respects before the treatments are applied.
- Comparative design helps ensure that influences other than the cell phone operate equally on both groups.
- Therefore, differences in average brake reaction time must be due either to talking on the cell or to the play of chance in the random assignment of subjects to the two groups.

How to randomize?

Use Random Digits!!!

A **table of random digits** is a list of the digits: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 that has the following properties:

- The digit in any position in the list has the same chance of being any one of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
- The digits in different positions are independent in the sense that the value of one has no influence on the value of any other.
- Any pair of random digits has the same chance of being any of the 100 possible pairs: 00, 01, 02,..., 98, 99.
- Any triple of random digits has the same chance of being any of the 1000 possible triples: 000, 001, 002,..., 998, 999.
- ... and so on for groups of four and more random digits.

Example: In the cell phone experiment we must divide 40 students at random into two groups of 20 students each.

130	69051	64817	87174	09517	84534	06489	87201	97245
131	05007	16632	81194	14873	04197	85576	45195	96565
132	68732	55259	84292	08796	43165	93739	31685	97150
133	45740	41807	65561	33302	07051	93623	18132	09547
134	27816	78416	18329	21337	35213	37741	04312	68508
135	66925	55658	39100	78458	11206	19876	87151	31260

Step 1 : Label students

01, 02, 03, ..., 39, 40

Step 2 : Table

05 16 17 40 20 19 32 04 25 29

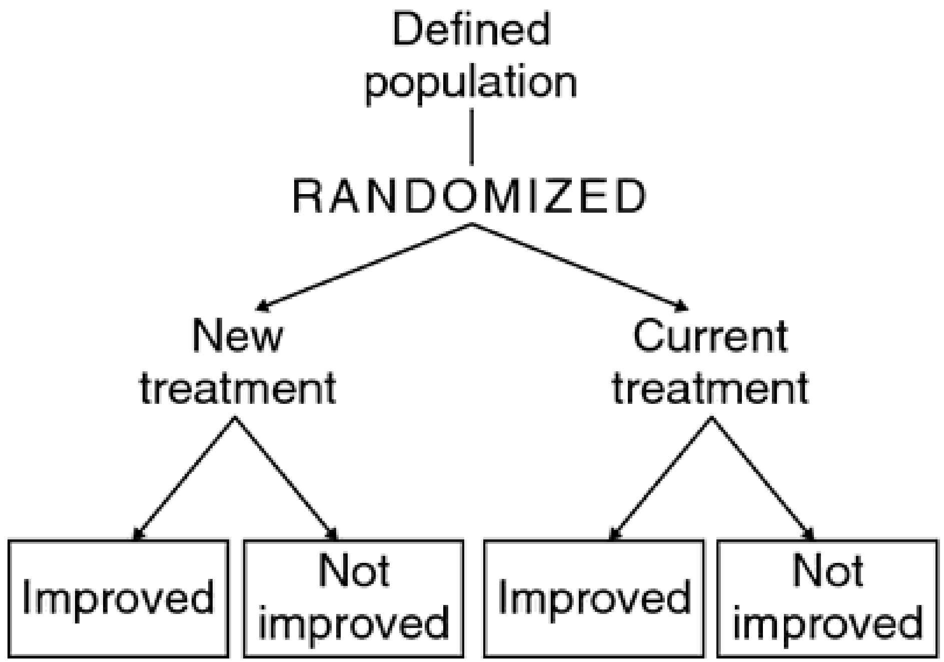
37 39 31 18 07 13 33 02 36 23

experimental group

The remaining 20 will go to the control group

The basic principles of statistical design of experiments are

- **Compare** two or more treatments. This will control the effects of lurking variables on the response.
- **Randomize** – use impersonal chance to assign experimental units to treatments.
- **Repeat** each treatment on many units to reduce chance variation in the results.



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

For example, there was statistically significant evidence that talking on a cell phone increases the mean reaction time of drivers.

Cautions about Experimentation

Many experiments have some weaknesses.

- The **environment** of an experiment can influence the outcomes in unexpected ways.

Example: A study of the effects of marijuana recruited young men who used marijuana.

- Some were randomly assigned to smoke marijuana cigarettes.
- Others were given placebo cigarettes.

This failed: the control group recognized that their cigarettes were phony and complained loudly.

It may be quite common for blindness to fail because the subjects can tell which treatment they are receiving.

- The most serious potential weakness of experiments is **lack of realism**.

Example: How do layoffs at workplace affect the workers who remain on the job?

Psychologists asked student subjects to proofread text for extra course credit, then «let go» some of the workers (who were actually accomplices of the experimenters).

- Some subjects were told that those let go had performed poorly (Treatment 1).
- Others were told that not all could be kept and that it was just luck that they were kept and others let go (Treatment 2).

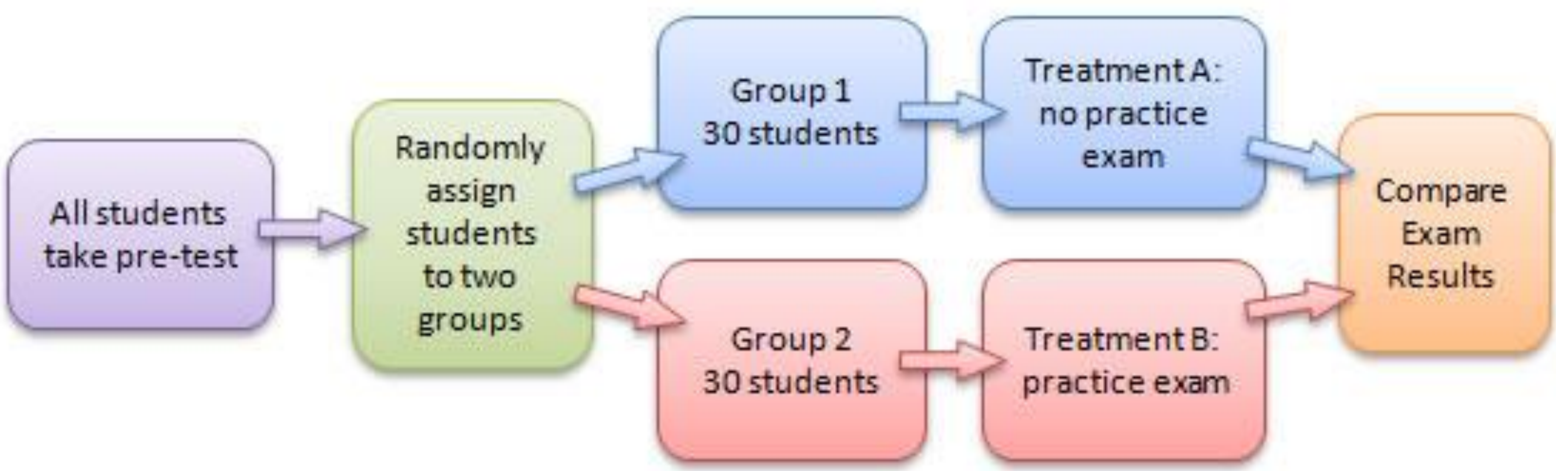
We can't be sure that the reactions of the students are the same as those of workers who survive a layoff in which other workers lose their jobs. Many behavioral science experiments use student subjects in a campus setting. Do the conclusions apply to the real world? No

Matched Pairs Design

- compares just two treatments.

<u>Example</u>: Matched pairs for the cell phone experiment above. The experiment compared two treatments, driving in a simulator and driving in the simulator while talking on a hand-free cell phone. The response variable is the time the driver takes to apply the brake when the car in front brakes suddenly.

<u>Example</u>:



Block Design

- extends the use of «similar subjects» from pairs to larger groups.

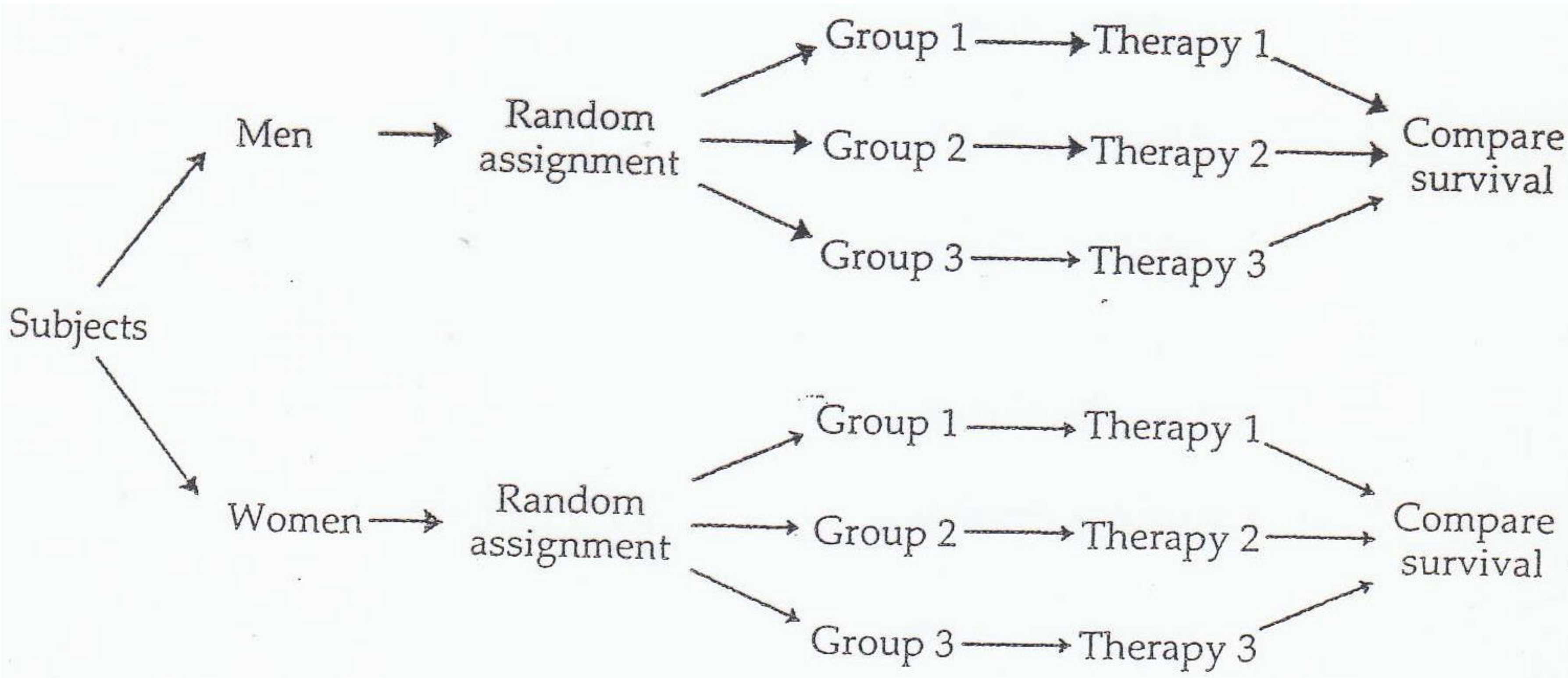
Definition: A **block** is a group of experimental units or subjects that are known before the experiment to be similar in some way that is expected to affect the response to the treatments. In a **block design**, the random assignment of units to treatments is carried out separately within each block.

Note: Blocks are another form of control.

Example: The progress of a type of cancer differs in women and men. A clinical experiment to compare three therapies for this cancer therefore treats sex as a blocking variable. Two separate randomizations are done:

- one assigning the female subjects to the treatments
- the other assigning the male subjects.

Figure below outlines the design of this experiment. Note that there is no randomization involved in making up the blocks.



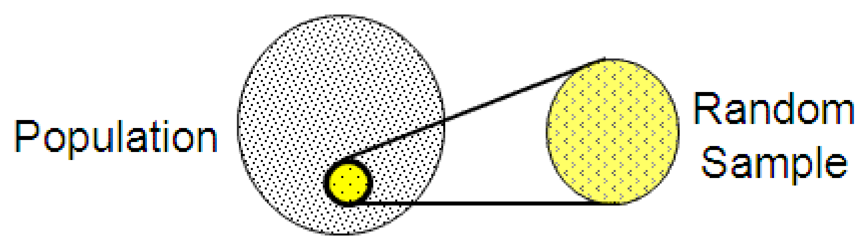
Sampling Design (Sample Surveys)

 What if we want ask questions about a large population of people, items, etc.?

Example: One of the most important sample surveys in the US is the General Social Survey (GSS) conducted by the NORC, a national organization for research and computing affiliated with the University of Chicago.

The GSS interviews about 3000 adult residents of the United States every second year.

Note: The GSS selects a **sample** of adults to represent the large **population** of all English-speaking adults living in the country.



How should we select a representative sample?

1. Carefully select individuals to match the population in every way we can think of:

- Males and females
- the right mix of ages
- right number of people living in each city/rural area
- right mix of political opinions
- etc, etc.

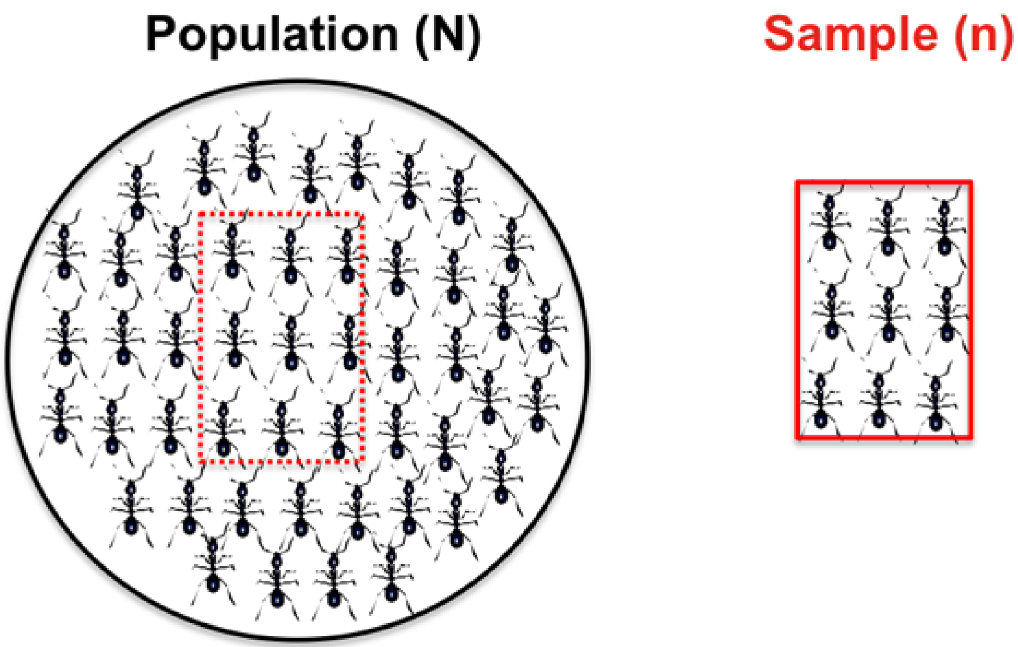
Sometimes it is difficult to do. Might miss something important.

2. Select individuals at random.

Easy to do. Approximately represents population in all ways, including ones you didn't think of.

Definition: The entire group of individuals that we want information about is called the **population**.

A **sample** is a part of the population that we actually examine in order to gather information.



Three keys for sampling:

- We study a part to gain information about the whole.
- Randomize (to obtain the sample)
- Sample size

How large a random sample do we need for the sample to be reasonably representative of the population?

It's the size of the sample, not the size of the population, that makes the difference in sampling.

Exception: If the population is small enough and the sample is more than 10% of the whole population, the population size can matter.

Sample size $\leq$ 10% of population

Does a Census Make Sense?

Definition: A survey with all individuals in the population is called a **census**.

Wouldn't it be better to just include everyone and "sample" the entire population?

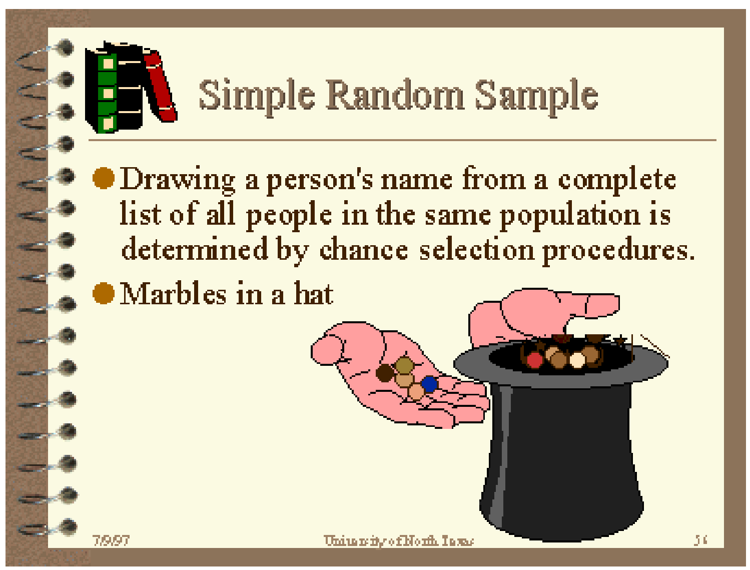
It can be difficult:

- There are always some individuals who are hard to locate or hard to measure, high cost etc.
- Sometimes the population changes while you work, and your findings might not be relevant to the current population by the time you complete your census.

Simple Random Sample

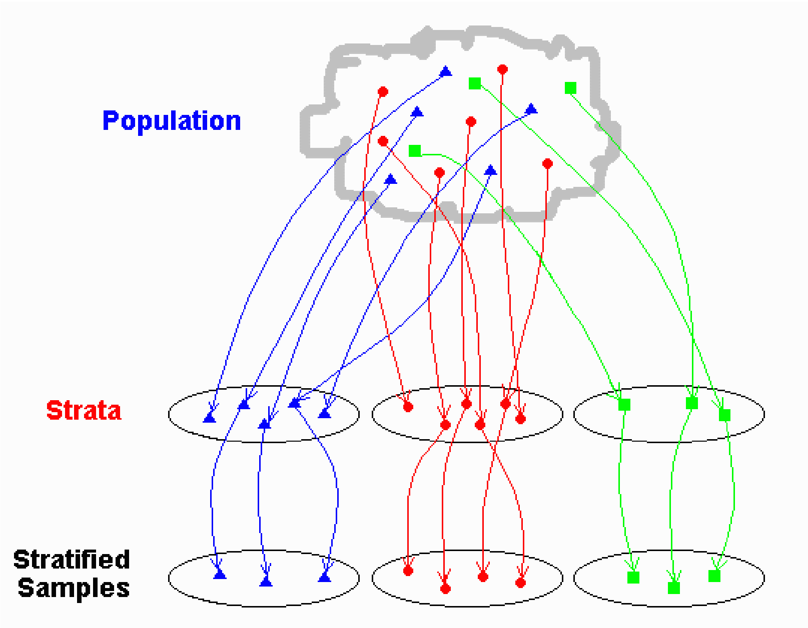
Definition: A **simple random sample (SRS)** of size *n* consists of *n* individuals from the population chosen in such a way that every set of *n* individuals has an equal chance to be the sample actually selected.

Design: place names in a hat (population) and then draw a handful (sample).



We can also use the table of random digits to draw an SRS.

Stratified Random Sample



To select a stratified random sample:

- Divide the population into groups of similar individuals, called strata.
- Then choose a separate SRS in each stratum and combine these SRSs to form the full sample.

Why is this good?

- Stratified random sampling can reduce bias.

- Stratifying can also reduce the variability of our results.

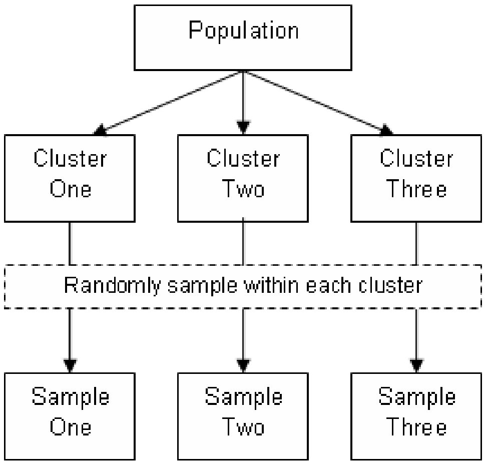
<u>Example</u>: A dentist is suspected of defrauding insurance companies by describing some dental procedures incorrectly on claim forms and overcharging for them.
An investigation begins by examining a sample of his bills for the past three years. Because there are five suspicious types of procedures, the investigators take a stratified sample. That is, they randomly select bills for each of the five types of procedures separately.

Cluster sampling

Sometimes stratifying isn't practical and simple random sampling is difficult.

Splitting the population into similar parts or clusters can make sampling more practical.

Then we could select one or a few clusters at random and perform a census (or take a sample if the clusters are large) within each of them.



This sampling design is called **cluster sampling**.

<u>Example</u>: How would you randomly sample 100 words from the textbook?

- SRS: number every single word and then sample from them.

- easier: randomly sample 10 pages first, then randomly sample 10 words on each page.

Multistage sampling

Often we work with hierarchy of clusters, e.g. chapter – section – sentence – word, and could choose:

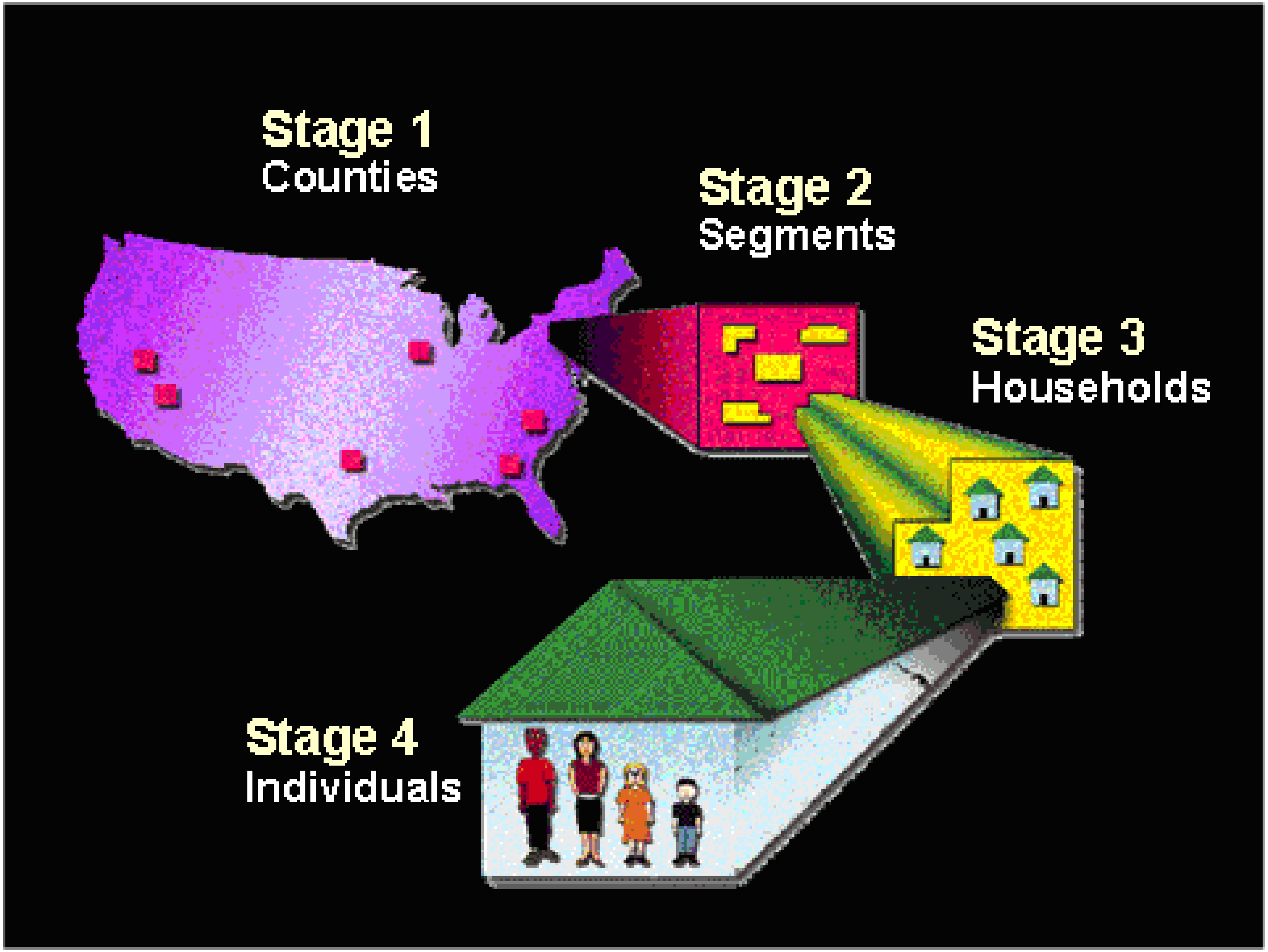
- chapters
- section within chosen chapter
- sentence within chosen section
- word within chosen sentence

It is called **multistage sampling**.

- Stage 1: Randomly select primary sampling units (specific communities, special health facilities)

- Stage 2: Within the primary sampling units, randomly select the final sampling units (households, patients)

- Sometimes in complex samples, additional stages are needed.


Stage 1
Counties
Stage 2
Segments
Stage 3
Households
Stage 4
Individuals

Systematic sampling

Some samples select individuals systematically.

Systematic sampling is a statistical method involving the selection of elements from an ordered sampling frame.

(A **sampling frame** is the source material from which a sample is drawn. It is a list of all those within a population who can be sampled, and may include individuals, households or institutions.)

Example: You want to sample 8 houses from a street of 120 houses. 120/8=15, so every 15th house is chosen after a random starting point between 1 and 15. If the random starting point is 11, then the houses selected are 11, 26, 41, 56, 71, 86, 101, and 116.

This is random sampling with a system. From the sampling frame, a starting point is chosen at random, and choices thereafter are at regular intervals.

Cautions About Sample Surveys:

- Nonresponse (not getting who you want)
- Undercoverage (some groups in the populations are left out)
- The behavior of the respondent can cause bias in sample results

- The wording of questions might be misleading, or have hints
- Sampling volunteers (don't rely on people who choose to respond, e.g. callers to radio show)

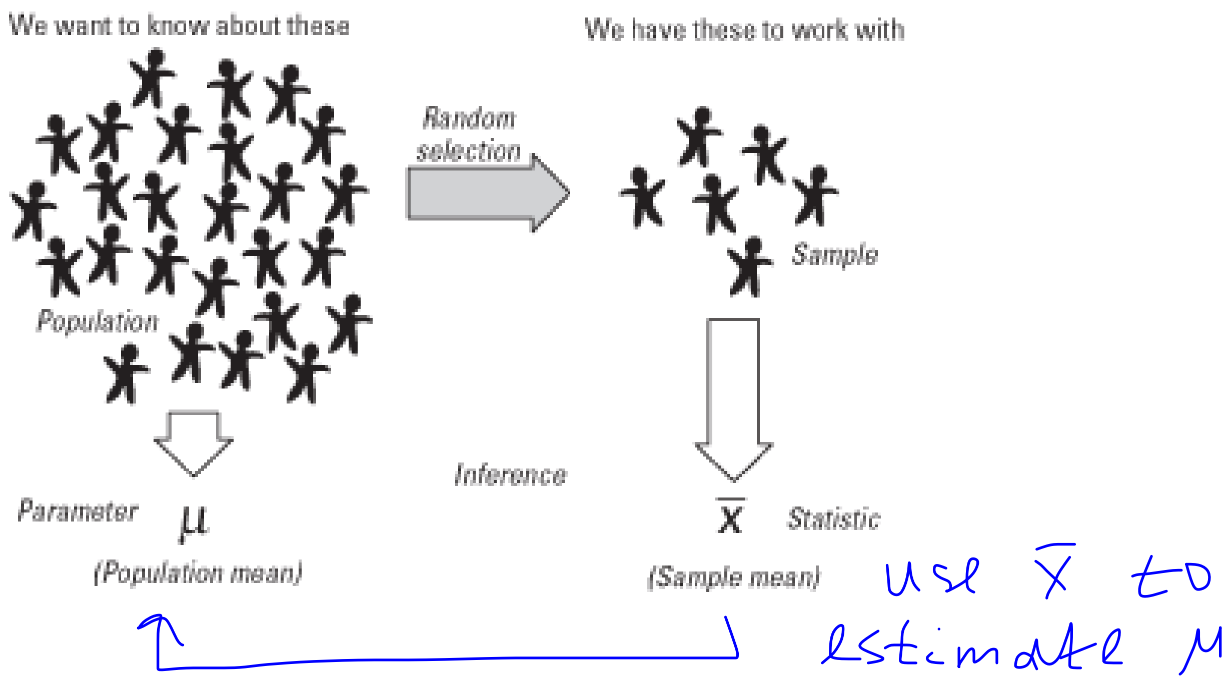
Statistical Inference

We infer conclusions about the wider population from data collected from a sample.

<u>Definition</u>: A **parameter** is a number that describes the population. A parameter is a fixed number, but in practice we do not know its value.

A **statistic** is a number that describes a sample. The value of a statistic is known when we have taken a sample, but it can change from sample to sample.

We often use a statistic to estimate an unknown parameter.



Example: Are attitudes toward shopping changing?

Sample surveys show that fewer people enjoy shopping than in the past. A survey by the market research firm Yankelovich Clancy Shulman asked a nationwide random sample of 2500 adults if they agreed or disagreed that «I like buying new clothes, but shopping is often frustrating and time-consuming.» Of the respondents, 1650, or 66%, said they agreed.

$\hat{p} = \frac{1650}{2500} = 0.66 = 66\%$ → sample proportion (proportion of people in the sample who would agree)

$\hat{p}$ is a statistics

$n = 2500$ → sample size

True parameter (p) is proportion of all adults who would agree. Use $\hat{p}$ to estimate p.

Sampling Variability

- The value of a statistic varies in repeated random sampling.

All of statistical inference is based on one idea: to see how trustworthy a procedure is, ask what would happen if we repeated it many times:

- Take a large number of samples from the same population.
- Calculate the sample proportion $\hat{p}$ for each sample.
- Make a histogram of the values of $\hat{p}$.
- Examine the distribution displayed in the histogram for shape, center, and spread, as well as outliers or other deviations.

Definition: The **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same sample size from the same population.