Chapter 4: Designing Studies

4.1 Samples and Surveys
4.2 Experiments
4.3 Using Studies Wisely

The Practice of Statistics, 4th edition – For AP*
STARNES, YATES, MOORE

Section 4.1
Samples and Surveys

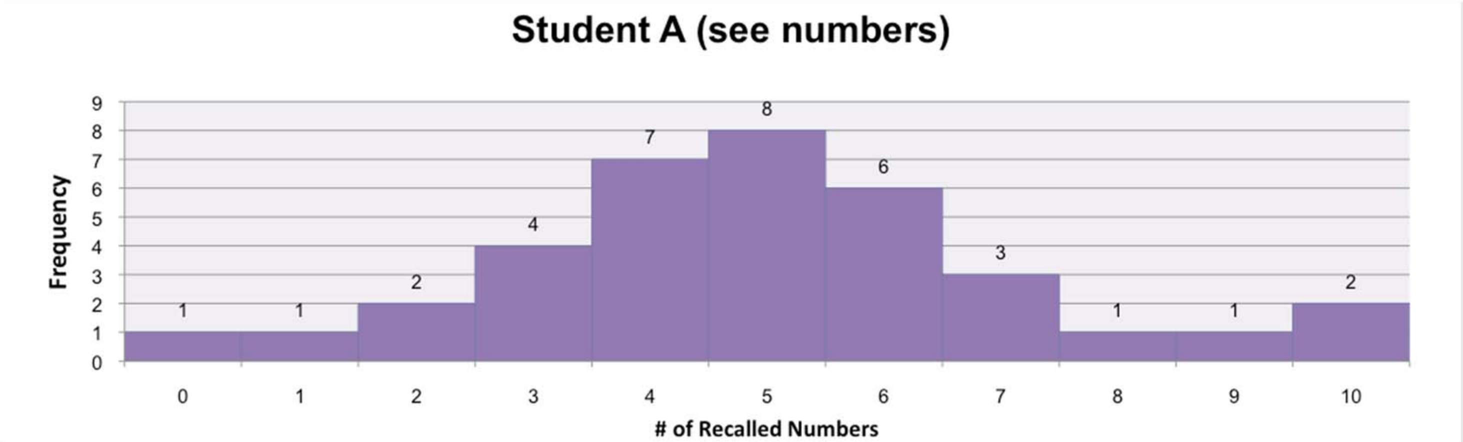
Learning Objectives

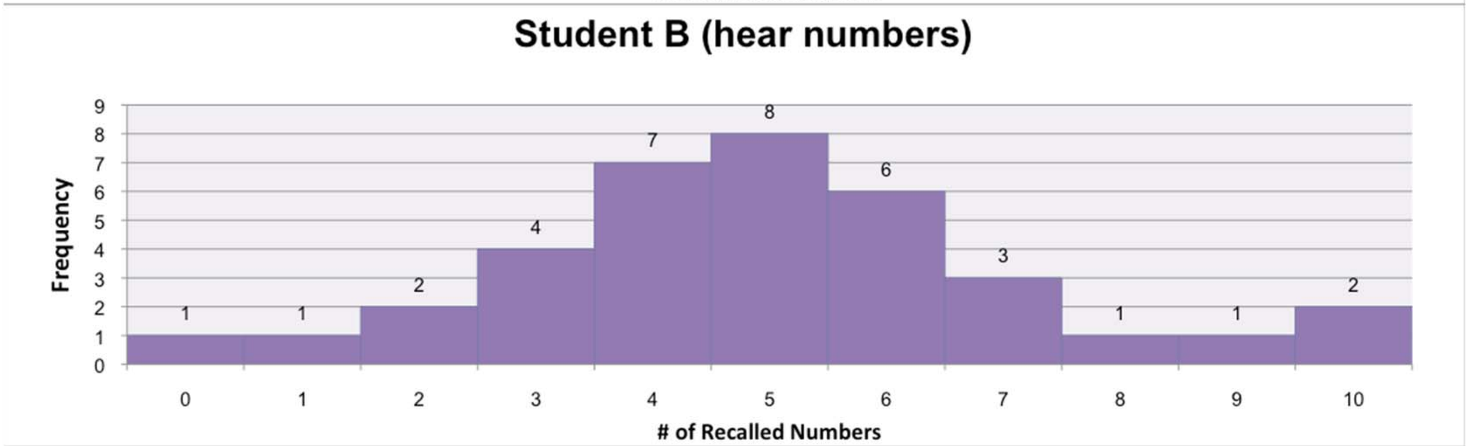
After this section, you should be able to…

- ✓ IDENTIFY the population and sample in a sample survey
- ✓ IDENTIFY voluntary response samples and convenience samples
- ✓ DESCRIBE how to use a table of random digits to select a simple random sample (SRS)
- ✓ DESCRIBE simple random samples, stratified random samples, and cluster samples
- ✓ EXPLAIN how undercoverage, nonresponse, and question wording can lead to bias in a sample survey

Activity: *See no evil, hear no evil?*

- **Follow the directions on Page 206**
- **Turn in your results to your teacher.**
- *Teacher: Right-click (control-click) on the graphs to edit the counts.*





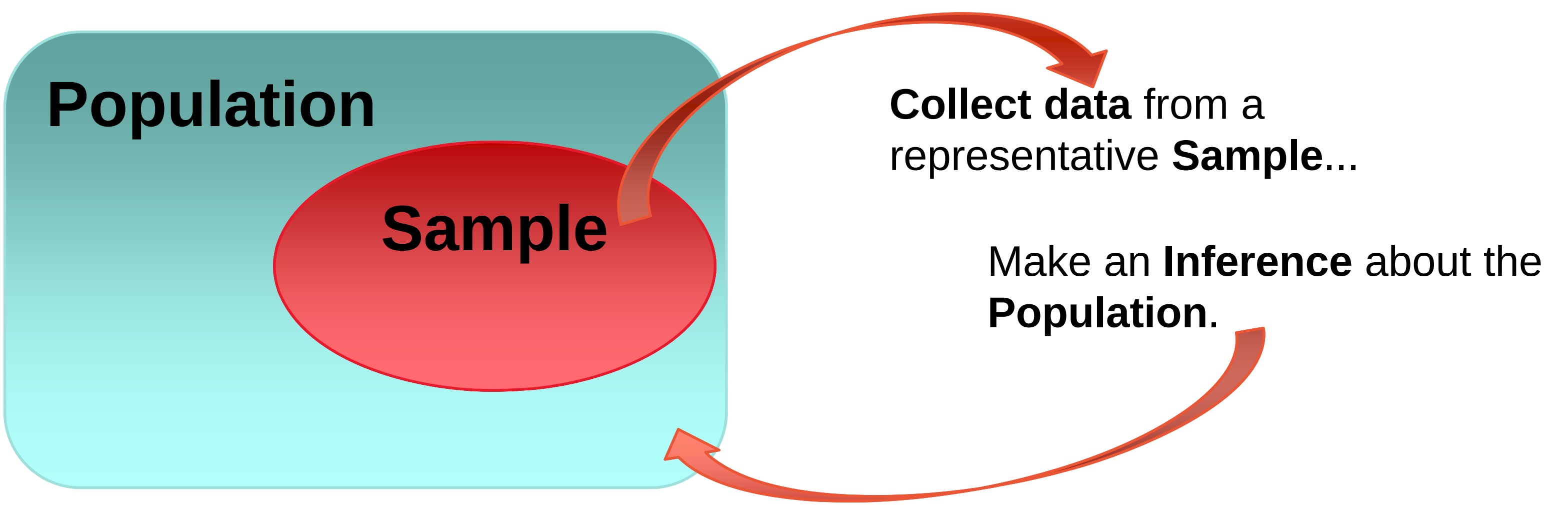
Population and Sample

The distinction between population and sample is basic to statistics. To make sense of any sample result, you must know what population the sample represents

Definition:

The **population** in a statistical study is the entire group of individuals about which we want information.

A **sample** is the part of the population from which we actually collect information. We use information from a sample to draw conclusions about the entire population.



The Idea of a Sample Survey

We often draw conclusions about a whole population on the basis of a sample.

Choosing a sample from a large, varied population is not that easy.

Step 1: Define the *population* we want to describe.

Step 2: Say exactly *what we want to measure.*

A "sample survey" is a study that uses an organized plan to choose a sample that represents some specific population.

Step 3: Decide how to choose a sample from the population.

How to Sample Badly

How can we choose a sample that we can trust to represent the population? There are a number of different methods to select samples.

Definition:

Choosing individuals who are easiest to reach results in a **convenience sample.**

Convenience samples often produce unrepresentative data…why?

Definition:

The design of a statistical study shows **bias** if it systematically favors certain outcomes.

- **How to Sample Badly**
- Convenience samples are almost guaranteed to show bias. So are **voluntary response samples**, in which people decide whether to join the sample in response to an open invitation.

Definition:

A **voluntary response sample** consists of people who choose themselves by responding to a general appeal. Voluntary response samples show bias because people with strong opinions (often in the same direction) are most likely to respond.

How to Sample Well: Random Sampling

- The statistician's remedy is to allow impersonal chance to choose the sample. A sample chosen by chance rules out both favoritism by the sampler and self-selection by respondents.
- **Random sampling**, the use of chance to select a sample, is the central principle of statistical sampling.

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.

In practice, people use random numbers generated by a computer or calculator to choose samples. If you don't have technology handy, you can use a **table of random digits.**

How to Choose an SRS

Definition:

A **table of random digits** is a long string of the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 with these properties:

- Each entry in the table is equally likely to be any of the 10 digits 0 - 9.
- The entries are independent of each other. That is, knowledge of one part of the table gives no information about any other part.

How to Choose an SRS Using Table D

Step 1: Label. Give each member of the population a numerical label of the *same length.*

Step 2: Table. Read consecutive groups of digits of the appropriate length from Table D.

Your sample contains the individuals whose labels you find.

Example: How to Choose an SRS

- Problem: Use Table D at line 130 to choose an SRS of 4 hotels.

01 Aloha Kai	08 Captiva	15 Palm Tree	22 Sea Shell
02 Anchor Down	09 Casa del Mar	16 Radisson	23 Silver Beach
03 Banana Bay	10 Coconuts	17 Ramada	24 Sunset Beach
04 Banyan Tree	11 Diplomat	18 Sandpiper	25 Tradewinds
05 Beach Castle	12 Holiday Inn	19 Sea Castle	26 Tropical Breeze
06 Best Western	13 Lime Tree	20 Sea Club	27 Tropical Shores
07 Cabana	14 Outrigger	21 Sea Grape	28 Veranda

69051 64817 87174 09517 84534 06489 87201 97245

69 05 16 48 17 87 17 40 95 17 84 53 40 64 89 87 20

Our SRS of 4 hotels for the editors to contact is: 05 Beach Castle, 16 Radisson, 17 Ramada, and 20 Sea Club.

Other Sampling Methods

- The basic idea of sampling is straightforward: take an SRS from the population and use your sample results to gain information about the population. Sometimes there are statistical advantages to using more complex sampling methods.
- One common alternative to an SRS involves sampling important groups (called strata) within the population separately. These "sub-samples" are combined to form one stratified random sample.

Definition:

To select a **stratified random sample,** first classify 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.

Activity: Sampling Sunflowers

- Use Table D or technology to take an SRS of 10 grid squares using the rows as strata. Then, repeat using the columns as strata.

	A	B	C	D	E	F	G	H	I	J
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

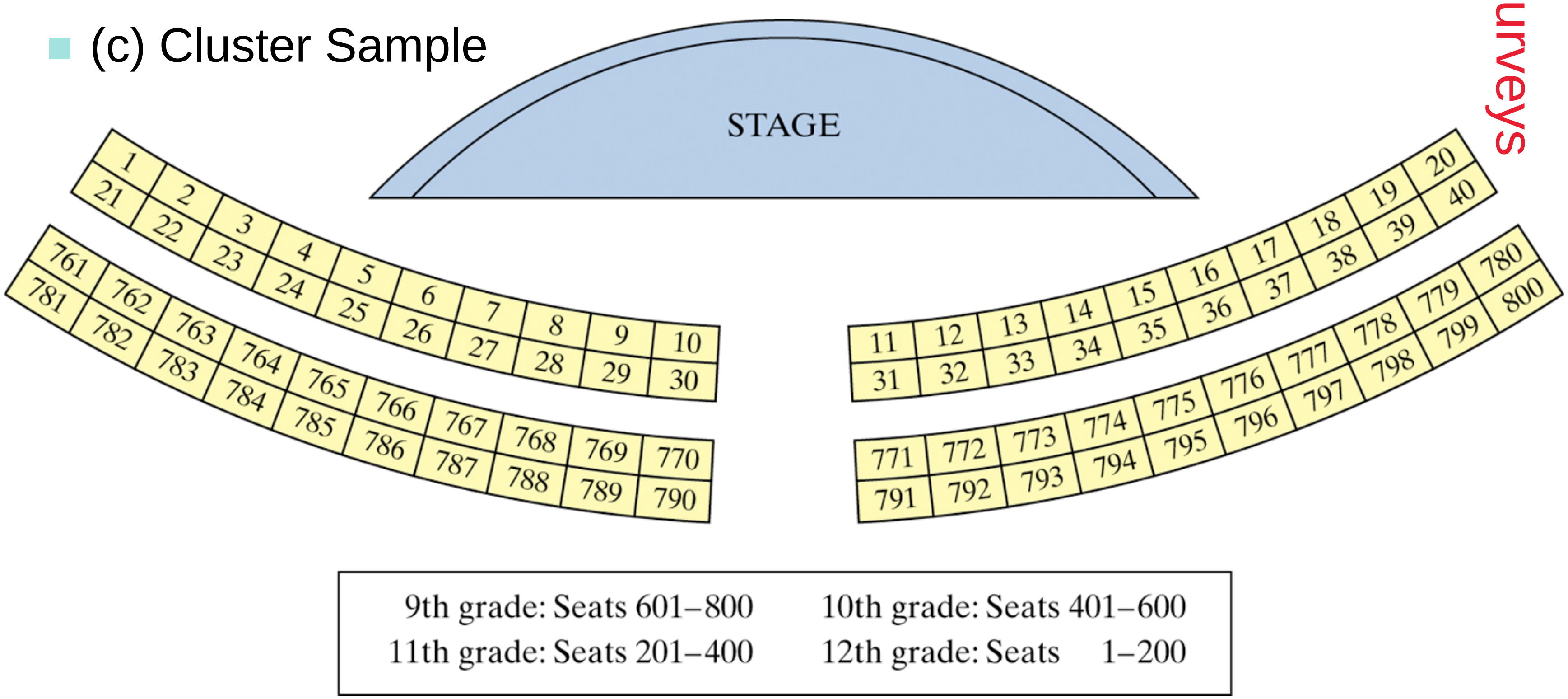
- **Other Sampling Methods**
- Although a stratified random sample can sometimes give more precise information about a population than an SRS, both sampling methods are hard to use when populations are large and spread out over a wide area.
- In that situation, we'd prefer a method that selects groups of individuals that are "near" one another.

Definition:

To take a **cluster sample**, first divide the population into smaller groups. Ideally, these clusters should mirror the characteristics of the population. Then choose an SRS of the clusters. All individuals in the chosen clusters are included in the sample.

Example: Sampling at a School Assembly

- Describe how you would use the following sampling methods to select 80 students to complete a survey.
- (a) Simple Random Sample
- (b) Stratified Random Sample
- (c) Cluster Sample



Inference for Sampling

- The purpose of a sample is to give us information about a larger population.

- The process of drawing conclusions about a population on the basis of sample data is called **inference**.

Why should we rely on random sampling?

1)To eliminate bias in selecting samples from the list of available individuals.

2)The laws of probability allow trustworthy inference about the population

- Results from random samples come with a **margin of error** that sets bounds on the size of the likely error.
- Larger random samples give better information about the population than smaller samples.

Sample Surveys: What Can Go Wrong?

- Most sample surveys are affected by errors in addition to sampling variability.
- Good sampling technique includes the art of reducing all sources of error.

Definition

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

Nonresponse occurs when an individual chosen for the sample can't be contacted or refuses to participate.

A systematic pattern of incorrect responses in a sample survey leads to **response bias**.

The **wording of questions** is the most important influence on the answers given to a sample survey.

\+

Section 4.1
Samples and Surveys

Summary

In this section, we learned that…

- ✓ A **sample survey** selects a **sample** from the **population** of all individuals about which we desire information.
- ✓ **Random sampling** uses chance to select a sample.
- ✓ The basic random sampling method is a **simple random sample (SRS)**.
- ✓ To choose a **stratified random sample**, divide the population into **strata**, then choose a separate SRS from each stratum.
- ✓ To choose a **cluster sample**, divide the population into groups, or **clusters**. Randomly select some of the clusters for your sample.

Section 4.1
Samples and Surveys

Summary, con't

In this section, we learned that…

- ✓ Failure to use random sampling often results in **bias**, or systematic errors in the way the sample represents the population.
- ✓ **Voluntary response samples** and **convenience samples** are particularly prone to large bias.
- ✓ **Sampling errors** come from the act of choosing a sample. Random sampling error and **undercoverage** are common types of error.
- ✓ The most serious errors are **nonsampling errors.** Common types of sampling error include **nonresponse, response bias,** and **wording of questions.**

Section 4.2
Experiments

Learning Objectives

After this section, you should be able to…

- ✓ DISTINGUISH observational studies from experiments
- ✓ DESCRIBE the language of experiments
- ✓ APPLY the three principles of experimental design
- ✓ DESIGN comparative experiments utilizing completely randomized designs and randomized block designs, including matched pairs design

Observational Study versus Experiment

In contrast to observational studies, experiments don't just observe individuals or ask them questions. They actively impose some treatment in order to measure the response.

Definition:

An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses.

An **experiment** deliberately imposes some treatment on individuals to measure their responses.

When our goal is to understand cause and effect, experiments are the *only* source of fully convincing data.

The distinction between observational study and experiment is one of the most important in statistics.

Observational Study versus Experiment

Observational studies of the effect of one variable on another often fail because of **confounding** between the explanatory variable and one or more **extraneous variables.**

Definition:

~~A **lurking variable** is a variable that is not among the explanatory or response variables in a study but that may influence the response variable~~.

"Do not worry about understanding the term 'lurking variable.' It is simply another extraneous variable that distort your results. -- Ms. Groves"

Confounding occurs when two variables are associated in such a way that their effects on a response variable cannot be distinguished from each other.

Well-designed experiments take steps to avoid confounding.

The Language of Experiments

An experiment is a statistical study in which we actually do something (a **treatment**) to people, animals, or objects (the **experimental units**) to observe the **response**. Here is the basic vocabulary of experiments.

Definition:

A specific condition applied to the individuals in an experiment is called a **treatment**. If an experiment has several explanatory variables, a treatment is a combination of specific values of these variables.

The **experimental units** are the smallest collection of individuals to which treatments are applied. When the units are human beings, they often are called **subjects**.

Sometimes, the explanatory variables in an experiment are called **factors**. Many experiments study the joint effects of several factors. In such an experiment, each treatment is formed by combining a specific value (often called a **level**) of each of the factors.

How to Experiment Badly

- Experiments are the preferred method for examining the effect of one variable on another. By imposing the specific treatment of interest and controlling other influences, we can pin down cause and effect. Good designs are essential for effective experiments, just as they are for sampling.

Example, page 236

A high school regularly offers a review course to prepare students for the SAT. This year, budget cuts will allow the school to offer only an online version of the course. Over the past 10 years, the average SAT score of students in the classroom course was 1620. The online group gets an average score of 1780. That's roughly 10% higher than the long- time average for those who took the classroom review course. Is the online course more effective?

Students -> Online Course -> SAT Scores

How to Experiment Badly

- Many laboratory experiments use a design like the one in the online SAT course example:

In the lab environment, simple designs often work well.

Field experiments and experiments with animals or people deal with more variable conditions.

Outside the lab, badly designed experiments often yield worthless results because of confounding.

- **How to Experiment Well: The Randomized Comparative Experiment**

- The remedy for confounding is to perform a *comparative experiment* in which some units receive one treatment and similar units receive another. Most well designed experiments compare two or more treatments.

- Comparison alone isn't enough, if the treatments are given to groups that differ greatly, *bias* will result. The solution to the problem of bias is **random assignment.**

Definition:

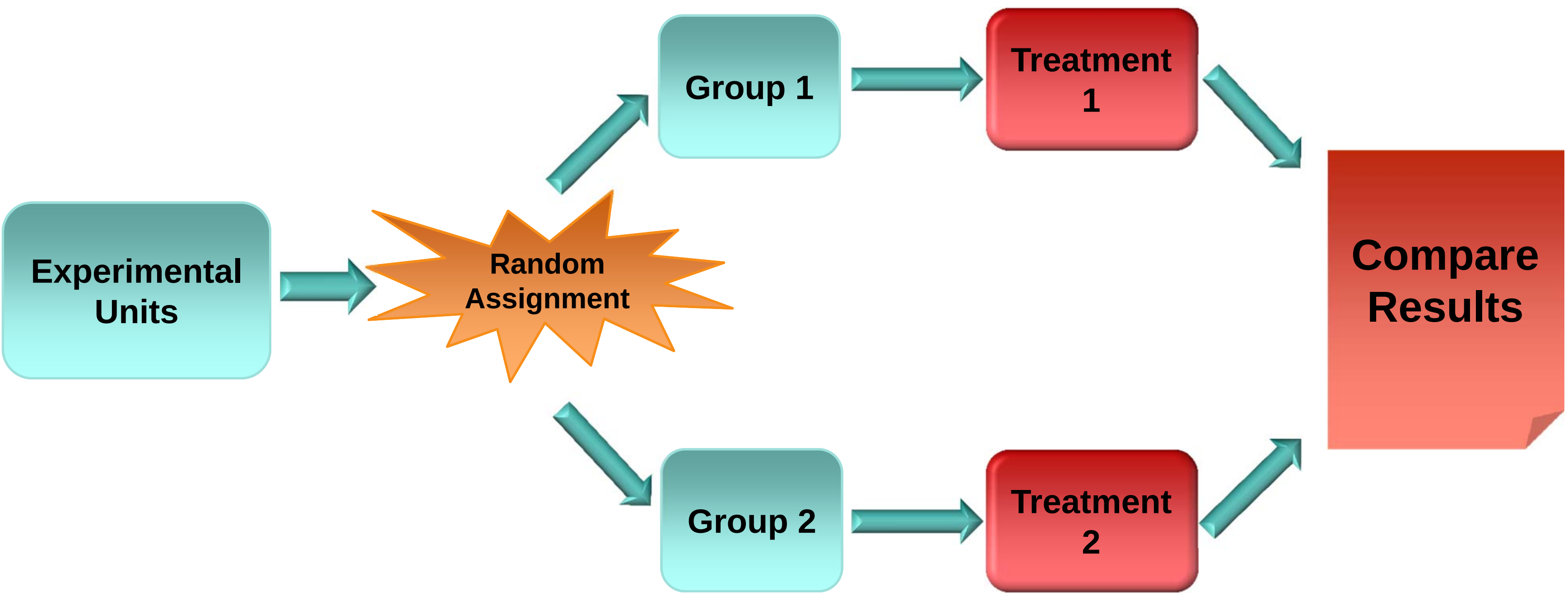
In an experiment, **random assignment** means that experimental units are assigned to treatments at random, that is, using some sort of chance process.

The Randomized Comparative Experiment

Definition:

In a **completely randomized design**, the treatments are assigned to all the experimental units completely by chance.

Some experiments may include a **control group** that receives an inactive treatment or an existing baseline treatment.



Three Principles of Experimental Design

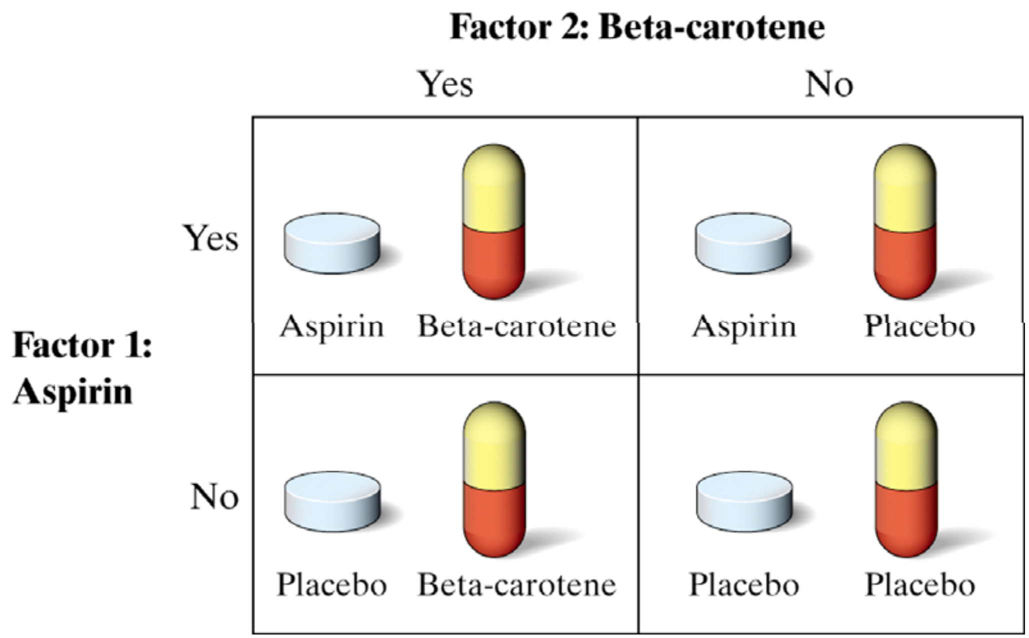
- Randomized comparative experiments are designed to give good evidence that differences in the treatments actually cause the differences we see in the response.

Principles of Experimental Design

1. **Control** for ~~lurking~~ **extraneous** variables that might affect the response: Use a comparative design and ensure that the only systematic difference between the groups is the treatment administered.
2. **Random assignment**: Use impersonal chance to assign experimental units to treatments. This helps create roughly equivalent groups of experimental units by balancing the effects of lurking variables that aren't controlled on the treatment groups.
3. **Replication**: Use enough experimental units in each group so that any differences in the effects of the treatments can be distinguished from chance differences between the groups.

Example: The Physicians' Health Study

- Read the description of the Physicians' Health Study on page 241. Explain how each of the three principles of experimental design was used in the study.



A **placebo** is a "dummy pill" or inactive treatment that is indistinguishable from the real treatment.

Experiments: What Can Go Wrong?

- The logic of a randomized comparative experiment depends on our ability to treat all the subjects the same in every way except for the actual treatments being compared.
- Good experiments, therefore, require careful attention to details to ensure that all subjects really are treated identically.

A response to a dummy treatment is called a **placebo effect**. The strength of the placebo effect is a strong argument for randomized comparative experiments.

Whenever possible, experiments with human subjects should be **double-blind**.

Definition:
In a **double-blind experiment,** neither the subjects nor those who interact with them and measure the response variable know which treatment a subject received.

Experiments

Inference for Experiments

- In an experiment, researchers usually hope to see a difference in the responses so large that it is unlikely to happen just because of chance variation.
- We can use the laws of probability, which describe chance behavior, to learn whether the treatment effects are larger than we would expect to see if only chance were operating.
- If they are, we call them **statistically significant**.

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

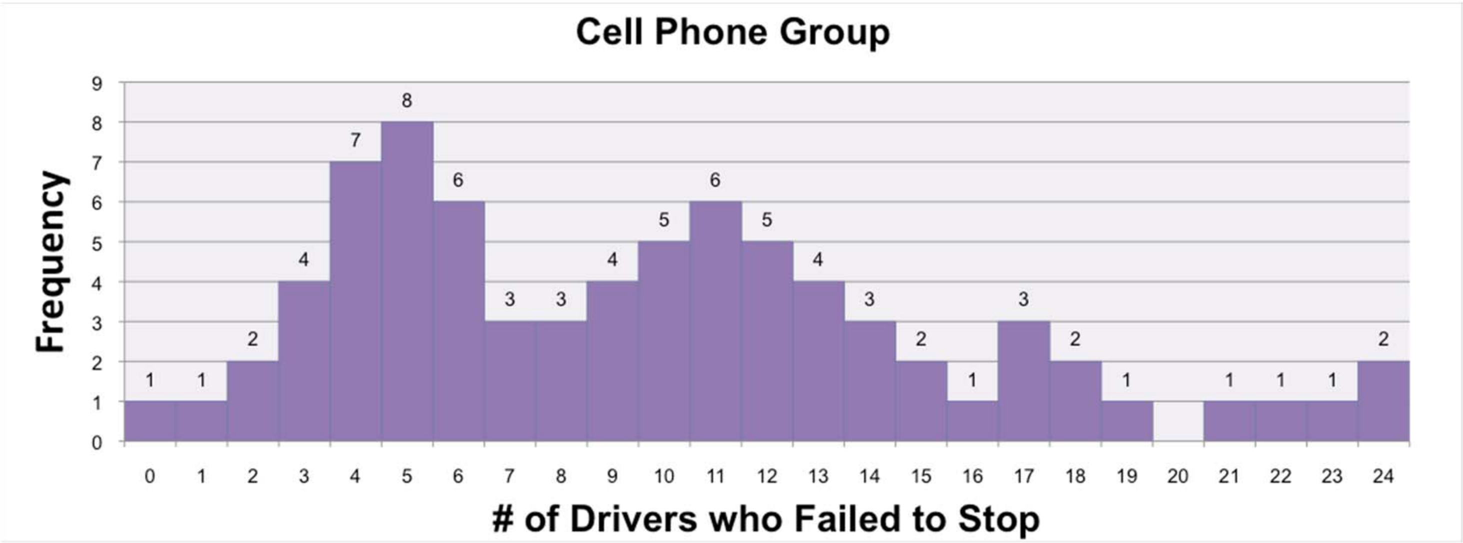
A statistically significant association in data from a well-designed experiment does imply causation.

Activity: Distracted Drivers

Is talking on a cell phone while driving more distracting than talking to a passenger? Read the Activity on page 245.

Perform 10 repetitions of your simulation and report the number of drivers in the cell phone group who failed to stop

Teacher: Right-click (control-click) on the graph to edit the counts.



In what percent of the class' trials did 12 or more people in the cell phone group fail to stop?

Based on these results, how surprising would it be to get a result this large or larger simply due to chance involved in random assignment? Is this result statistically significant?

Blocking

- Completely randomized designs are the simplest statistical designs for experiments. But just as with sampling, there are times when the simplest method doesn't yield the most precise results.

Definition

A **block** is a group of experimental units that are known before the experiment to be similar in some way that is expected to affect the response to the treatments.

In a **randomized block design**, the random assignment of experimental units to treatments is carried out separately within each block.

Form blocks based on the most important unavoidable sources of variability (lurking variables) among the experimental units.

Randomization will average out the effects of the remaining lurking variables and allow an unbiased comparison of the treatments.

Control what you can, block on what you can't control, and randomize to create comparable groups.

Matched-Pairs Design

- A common type of randomized block design for comparing two treatments is a matched pairs design. The idea is to create blocks by matching pairs of similar experimental units.

Definition

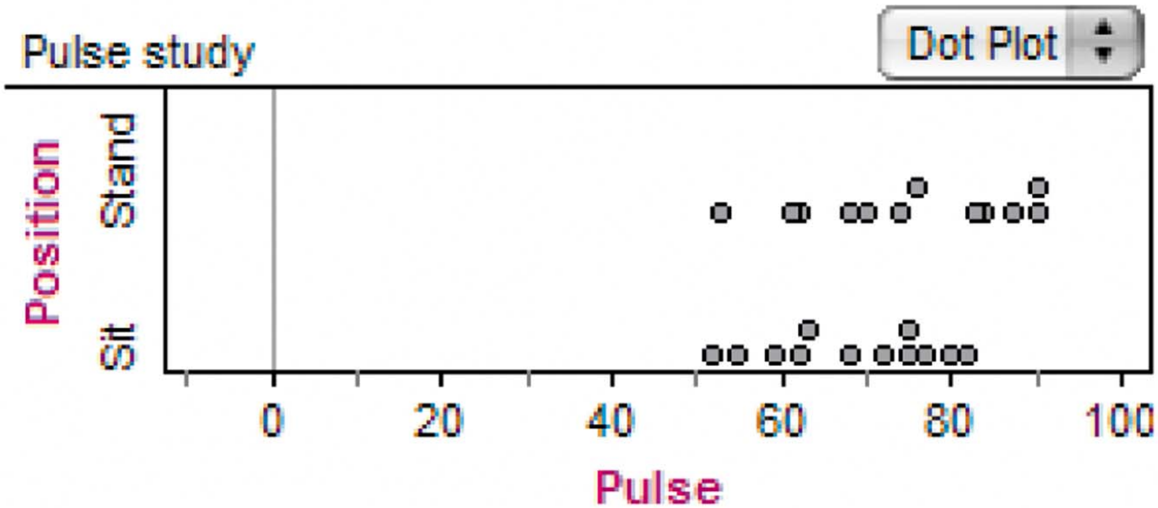
A **matched-pairs design** is a randomized blocked experiment in which each block consists of a matching pair of similar experimental units.

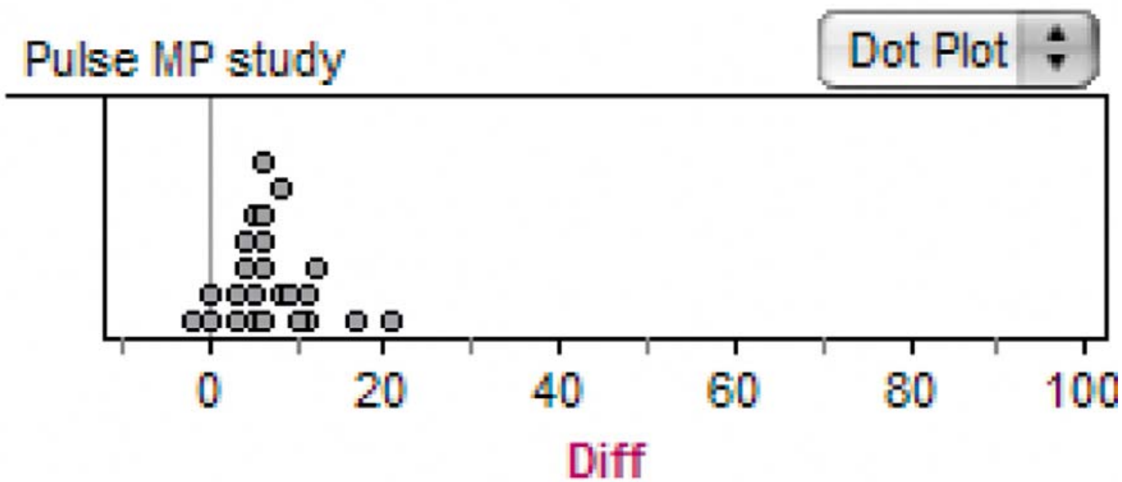
Chance is used to determine which unit in each pair gets each treatment.

Sometimes, a “pair” in a matched-pairs design consists of a single unit that receives both treatments. Since the order of the treatments can influence the response, chance is used to determine with treatment is applied first for each unit.

Example: Standing and Sitting Pulse Rate

- Consider the Fathom dotplots from a completely randomized design and a matched-pairs design. What do the dotplots suggest about standing vs. sitting pulse rates?





Section 4.2 Experiments

Summary

In this section, we learned that…

- ✓ We can produce data intended to answer specific questions by **observational studies** or **experiments.**
- ✓ In an experiment, we impose one or more **treatments** on a group of **experimental units** (sometimes called **subjects** if they are human).
- ✓ The **design** of an experiment describes the choice of treatments and the manner in which the subjects are assigned to the treatments.
- ✓ The basic principles of experimental design are **control** for lurking variables, **random assignment** of treatments, and **replication** (using enough experimental units).
- ✓ Many behavioral and medical experiments are **double-blind**.

Section 4.2 Experiments

Summary, con't

In this section, we learned that…

- ✓ Some experiments give a **placebo** (fake treatment) to a control group that helps confounding due to the **placebo-effect.**
- ✓ In addition to comparison, a second form of control is to form **blocks** of individuals that are similar in some way that is important to the response. Randomization is carried out within each block.
- ✓ **Matched pairs** are a common form of blocking for comparing just two treatments. In some matched pairs designs, each subject receives both treatments in a random order.

Section 4.3
Using Studies Wisely

Learning Objectives

After this section, you should be able to…

- ✓ DESCRIBE the challenges of establishing causation
- ✓ DEFINE the scope of inference
- ✓ DESCRIBE data ethics in designing studies

Scope of Inference

What type of inference can be made from a particular study? The answer depends on the design of the study.

Well-designed experiments randomly assign individuals to treatment groups. However, most experiments don't select experimental units at random from the larger population. That limits such experiments to inference about cause and effect.

Observational studies don't randomly assign individuals to groups, which rules out inference about cause and effect. Observational studies that use random sampling can make inferences about the population.

		Were individuals randomly assigned to groups?	
		Yes	**No**
Were individuals randomly selected?	**Yes**	Inference about the population: YES Inference about cause and effect: YES	Inference about the population: YES Inference about cause and effect: NO
	No	Inference about the population: NO Inference about cause and effect: YES	Inference about the population: NO Inference about cause and effect: NO

The Challenges of Establishing Causation

A well-designed experiment tells us that changes in the explanatory variable cause changes in the response variable.

Lack of realism can limit our ability to apply the conclusions of an experiment to the settings of greatest interest.

In some cases it isn't practical or ethical to do an experiment. Consider these questions:

- Does texting while driving increase the risk of having an accident?
- Does going to church regularly help people live longer?
- Does smoking cause lung cancer?

It is sometimes possible to build a strong case for causation in the absence of experiments by considering data from observational studies.

The Challenges of Establishing Causation

When we can't do an experiment, we can use the following criteria for establishing causation.

- *The association is strong.*
- *The association is consistent.*
- *Larger values of the explanatory variable are associated with stronger responses.*
- *The alleged cause precedes the effect in time.*
- *The alleged cause is plausible.*

Discuss how each of these criteria apply to the observational studies of the relationship between smoking and lung cancer.

Data Ethics

Complex issues of data ethics arise when we collect data from people. Here are some basic standards of data ethics that must be obeyed by all studies that gather data from human subjects, both observational studies and experiments.

Basic Data Ethics

- All planned studies must be reviewed in advance by an **institutional review board** charged with protecting the safety and well-being of the subjects.
- All individuals who are subjects in a study must give their **informed consent** before data are collected.
- All individual data must be kept **confidential**. Only statistical summaries for groups of subjects may be made public.

Section 4.3
Using Studies Wisely

Summary

In this section, we learned that…

- ✓ **Inference about the population** requires that the individuals taking part in a study be randomly selected from the larger population. A well-designed experiment that randomly assigns treatments to experimental units allows **inference about cause-and-effect**.
- ✓ **Lack of realism** in an experiment can prevent us from generalizing its results.
- ✓ In the absence of an experiment, good evidence of causation requires a strong association that appears consistently in many studies, a clear explanation for the alleged causal link, and careful examination of possible lurking variables.
- ✓ Studies involving humans must be screened in advance by an **institutional review board**. All participants must give their **informed consent,** and any information about the individuals must be kept **confidential**.