

Biomedical  
Science

Molecular  
Genetics



HSC 205 | Fall 2026-2027

**Biostatistics**

**Introduction to Statistics**



Free Session

*A+ is more than just a platform*



[Academy Website](#)



[WhatsApp Community](#)



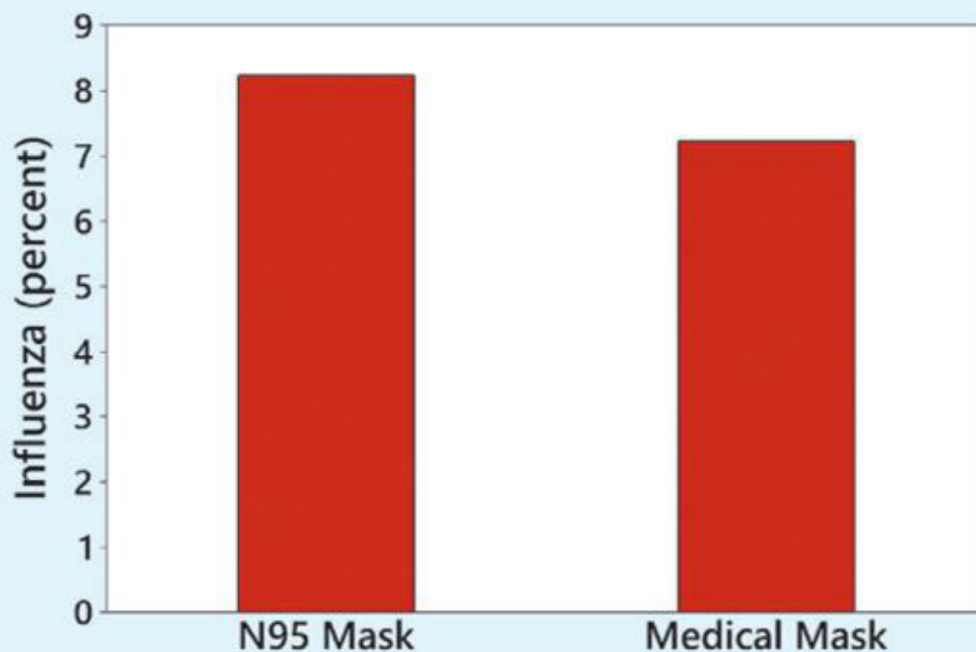
## Statistical & Critical Thinking

### OPENING QUESTION

Which facemasks are better?

**N95 Respirators or Medical Masks!!!**

- Statistical thinking involves **critical thinking** and the ability to make sense of results.
- Statistical thinking demands so much more than the ability to execute **complicated calculations**.



**FIGURE 1-2 Effectiveness of N95 Masks and Medical Masks**



## The Statistical Process

### PROCESS INVOLVED IN CONDUCTING A STATISTICAL STUDY

Prepare



Analyze



Conclude

- 1 **Prepare:** involves consideration of the **context**, **source of data**, and **sampling method**.
- 2 **Analyze:** involves constructing suitable graphs, exploring the data, and executing computations required for the statistical method being used.
- 3 **Conclude:** involves determining whether results have **statistical significance** and **practical significance**.

Focus is always on **critical thinking and not calculations.**

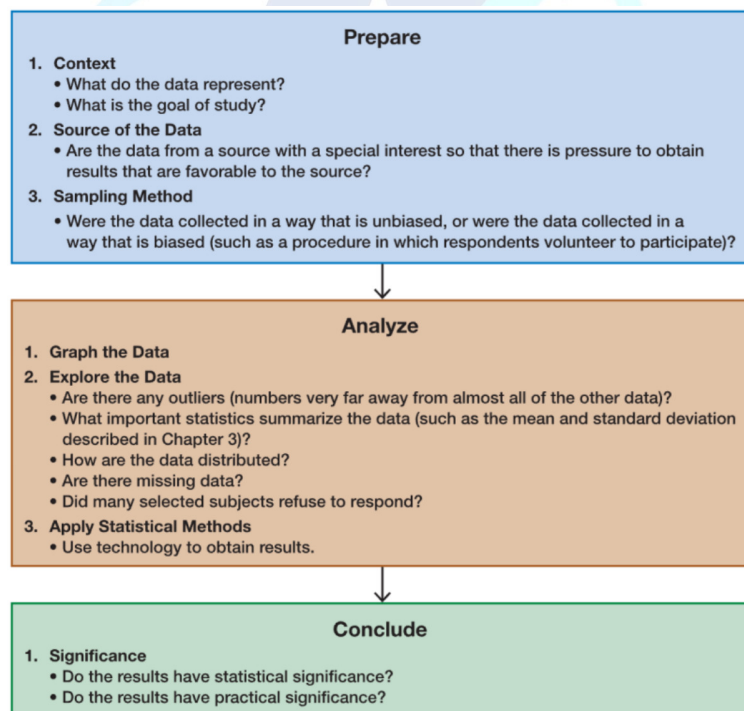


FIGURE 1.3 Statistical and Critical Thinking



## Data, Statistics & Biostatistics

### DATA

Collections of observations — for example measurements, respondent age and sex, or survey responses. **datum vs. data**

#### Counts

Data may be counts.

##### EXAMPLE

Number of patients discharged from hospital.

#### Measurements

Data may be measurements.

##### EXAMPLE

Nurse weighs a baby.

### STATISTICS

The science of planning studies and experiments; obtaining data; and organizing, summarizing, presenting, analyzing, and interpreting those data and then **drawing conclusions based on them.**

### BIostatISTICS

Statistics applied to the collection, analysis, and interpretation of biological data and especially data relating to **human biology, health, and medicine.**



## Population & Sample

### Population

The complete collection of **all** measurements or data that are being considered.

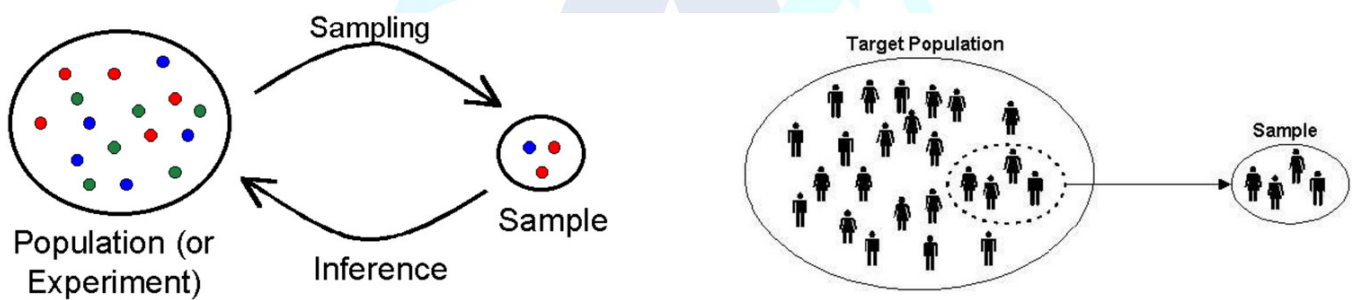
Consists of all subjects that are being studied.

### Sample

A subcollection of members selected from a population.

Is a group of subjects selected from a population.

Because populations are often very large, a common objective of the use of statistics is to obtain data from a sample and then **form a conclusion about the population.**



### Example 1

The community members of City A mandated to wear masks. Among them, there were **2512 subjects wearing N95 masks, 207 presented with influenza.**

**In this case which one is the population, and which one is the sample?**



## Voluntary Response Sample

### DEFINITION

A voluntary response sample (or self-selected sample) is one in which the **respondents themselves decide whether to be included.**

**Is it a good or bad choice?**

### EXAMPLES

- **Internet polls**, in which people online decide whether to respond.
- **Mail-in polls**, in which people decide whether to reply.
- **Telephone call-in polls**, in which newspaper, radio, or television announcements ask that you voluntarily call a special number to register your opinion.

### EXAMPLE 2 Voluntary Response Sample

*USA Today* posted this question on the electronic edition of the newspaper: “Have you ever been bitten by an animal?” Internet users who saw that question then decided themselves whether to respond. Among the 2361 responders, 65% said “yes” and 35% said “no.” Because the 2361 subjects themselves chose to respond, they are a voluntary response sample and the results of the survey are highly questionable. It would be much better to get results through a poll in which the pollster randomly selects the subjects, instead of allowing the subjects to volunteer themselves.



## Census & Bias

### Census

The collection of data from **every member** of a population.

Example: US Census.

### Bias

Built-in inaccuracy of an individual.

- Changes almost from minute to minute, depending on such things as the weather, the individual's blood sugar count, one's emotional state, etc.
- It is impossible to predict what an individual's bias is going to be at any one time, so that one could correct for it. **So how do we correct for this?**



## Prepare

### 1 Context

#### What does the data represent?

The table includes the systolic and diastolic blood pressure measurements (mmHg) of eight females.

#### What is the goal of the study?

The format of the table suggests the goal of determining whether there is a relationship between those two variables.

**TABLE 1-1** Systolic and Diastolic Blood Pressure Measurements of Females

|           |     |     |     |     |     |     |    |     |
|-----------|-----|-----|-----|-----|-----|-----|----|-----|
| Systolic  | 100 | 134 | 138 | 114 | 110 | 100 | 92 | 100 |
| Diastolic | 70  | 94  | 80  | 66  | 72  | 80  | 58 | 50  |

- 2 **Source of the data:** the data in the table are from Data Set 1 “Body Data” in Appendix B, where the source is identified. The data are from the **National Center for Health Statistics**, which is a source that certainly appears to be reputable and unbiased.
- 3 **Sampling method:** the specific sampling method used is fairly complicated, but it is designed to yield a representative sample of the population. Close analysis of the sampling method reveals that it is **sound and unbiased**.

## Prepare

### 1. Context

- What do the data represent?
- What is the goal of study?

### 2. Source of the Data

- Are the data from a source with a special interest so that there is pressure to obtain results that are favorable to the source?

### 3. Sampling Method

- Were the data collected in a way that is unbiased, or were the data collected in a way that is biased (such as a procedure in which respondents volunteer to participate)?



## Analyze

### 1 Graph and explore

- An analysis should begin with appropriate **graphs** and **explorations of the data**.
- This will be discussed in detail in chapter 2 and 3.

### 2 Apply statistical methods

- A good statistical analysis **does not** require strong computational skills.
- A good statistical analysis **does** require using **common sense** and paying careful attention to sound statistical methods.

## Analyze

### 1. Graph the Data

### 2. Explore the Data

- Are there any outliers (numbers very far away from almost all of the other data)?
- What important statistics summarize the data (such as the mean and standard deviation described in Chapter 3)?
- How are the data distributed?
- Are there missing data?
- Did many selected subjects refuse to respond?

### 3. Apply Statistical Methods

- Use technology to obtain results.



## Conclude — Significance

The final step in our statistical process involves conclusions. We should develop an ability to distinguish between **statistical significance and practical significance**.

### Conclude

1. **Significance**
  - Do the results have statistical significance?
  - Do the results have practical significance?

#### 1. STATISTICAL SIGNIFICANCE

Achieved in a study when we get a result that is **very unlikely to occur by chance** — that is, if the likelihood of an event occurring by chance is **5% or less**.

##### 98 females in 100 births

**Statistically significant**, because such an extreme outcome is not likely to result from random chance.

##### 52 females in 100 births

**Not statistically significant**, because that event could easily occur with random chance.



## Practical Significance

#### 2. PRACTICAL SIGNIFICANCE

It is possible that some treatment or finding is effective, but common sense might suggest that the treatment or finding **does not make enough of a difference** to justify its use or to be practical.

#### EXAMPLE 3 Statistical Significance Versus Practical Significance

In a trial of weight loss programs, 21 subjects on the Atkins program lost an average (mean) of 2.1 kg (or 4.6 lb) after one year (based on data from “Comparison of the Atkins, Ornish, Weight Watchers, and Zone Diets for Weight Loss and Heart Disease Risk Reduction,” by Dansinger et al., *Journal of the American Medical Association*, Vol. 293, No. 1). The results show that this loss is *statistically significant* and is not likely to occur by chance. However, many dieters believe that after following this diet for a year, a loss of only 2.1 kg is not worth the time, cost, and effort so that for these people, this diet does not have *practical significance*.



## Analyzing Data: Potential Pitfalls

### Misleading Conclusions

When forming a conclusion based on statistical analysis, we should make statements that are **clear even to those** who have no understanding of statistics and its terminology.

Example: if there is a correlation between pulse rate and height, this does not mean that an increase in pulse rate causes an increase in height.

### Sample Data Reported Instead of Measured

When collecting data from people, it is better to **take measurements yourself** instead of asking subjects to report results.

Example: weight — do not ask people, especially women.

### Loaded Questions

If survey results are not worded carefully, the results of a study can be misleading.

- **97% yes:** "Should the President have the line item veto to eliminate waste?"
- **57% yes:** "Should the President have the line item veto, or not?"

### Order of Questions

Sometimes survey questions are unintentionally loaded by such factors as the **order of the items** being considered.

- "Would you say that traffic contributes more or less to air pollution than industry?" — 45% blamed traffic; 27% blamed industry.
- "Would you say that industry contributes more or less to air pollution than traffic?" — 24% blamed traffic; 57% blamed industry.



## Potential Pitfalls (continued)

### Nonresponse

A nonresponse occurs when someone either **refuses to respond or is unavailable.**

### Low Response Rates

If a survey has a low response rate, the **reliability of the results decreases**. In addition to having a smaller sample size, there is an increased likelihood of having a bias among those who do respond.

### Straightlining

Straightlining occurs when survey respondents give the **same answers** (such as choice "b") to all or almost all of the questions. Survey responses that result from straightlining should be deleted because it is obvious that they do not truly reflect the opinions of the respondents.

### Misleading Percentages

Some studies cite misleading percentages. Note that **100% of some quantity is all of it**, but if there are references made to percentages that exceed 100%, such references are often not justified.

Example: Migraine.



## Parameter & Statistic

### Parameter

A numerical measurement describing some characteristic of a **population**.

### Statistic

A numerical measurement describing some characteristic of a **sample**.

#### EXAMPLE

There are **258,664,729 adults** in the United States. A sample of **5755 adults** were taken and their measured **mean height is 166.47 cm**.

**Which one is the parameter and which one is the statistic?**



## Quantitative vs. Categorical Data

### Quantitative (numerical)

Consist of numbers representing **counts or measurements**.

Example: weight of models, ages of respondents.

### Categorical (qualitative / attribute)

Consist of **names or labels** (not numbers that represent counts or measurements).

Example: eye color (green, brown), hair color, gender (male, female).

The identification numbers 1, 2, 3, ..., 25 are assigned randomly to the 25 subjects in a clinical trial. Those numbers are **substitutes for names**.



## Discrete & Continuous Data

Quantitative data can be further described by distinguishing between **discrete and continuous types**.

### Discrete data

Result when the data values are quantitative and the number of values is **finite, or "countable."**

Example: number of tosses of a coin before getting tails.

### Continuous (numerical) data

Result from **infinitely many possible quantitative values**, where the collection of values is not countable. That is, it is impossible to count the individual items because at least some of them are on a **continuous scale**.

Example: lengths of distances, weights of individuals, blood pressure.



### Continuous Data

Credit: Suppakij1017/  
Shutterstock



### Discrete Data

Credit: Khamidulin Sergey/  
Shutterstock

What about grades?



## Levels of Measurement

Another way of classifying data is to use **four levels of measurement**:



### WHY IS THIS IMPORTANT?

When we are applying statistics to real problems, the level of measurement of the data helps us **decide which procedure to use.**

### Nominal level of measurement

Characterized by data that consist of **names, labels, or categories only**. The data cannot be arranged in some order (such as low to high).

Example: gender, blood type, marital status, eye color.

### Ordinal level of measurement

Data that **can be arranged in some order**, but differences (obtained by subtraction) between data values either cannot be determined or are meaningless.

Example: grades A, B, C, D, etc.; Likert scale.



## Interval & Ratio Levels

### Interval level of measurement

Data that can be arranged in order, and **differences between data values can be found and are meaningful**. Allows us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them. Data at the interval level **do not have a natural zero starting point** at which none of the quantity is present.

Example: pandemic years; temperature in degrees Fahrenheit or Celsius. A temperature of 40 degrees is higher than 30 degrees; an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees.

### Ratio level of measurement

Data that can be arranged in order, differences can be found and are meaningful, and there is a **natural zero starting point** (where zero indicates that none of the quantity is present). Differences and ratios are both meaningful, allowing statements such as **x is two times more than y**.

Example: measures of time or space, Kelvin temperature, height, weight, length. A temperature of 200 degrees is higher than 100 degrees, and it is twice as high.

**TABLE 1-2** Levels of Measurement

| Level of Measurement | Brief Description                                                                                   | Example                                                         |
|----------------------|-----------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| Ratio                | There is a natural zero starting point and ratios make sense.                                       | Heights, lengths, distances, volumes                            |
| Interval             | Differences are meaningful, but there is no natural zero starting point and ratios are meaningless. | Body temperatures in degrees Fahrenheit or Celsius              |
| Ordinal              | Data can be arranged in order, but differences either can't be found or are meaningless.            | Ranks of medical schools in <i>U.S. News &amp; World Report</i> |
| Nominal              | Categories only. Data cannot be arranged in order.                                                  | Eye colors                                                      |



## Big Data & Data Science

### BIG DATA

Refers to data sets so large and so complex that their analysis is **beyond the capabilities of traditional software tools**. Analysis of big data may require software simultaneously running in parallel on many different computers.

### EXAMPLES OF DATA SET MAGNITUDES

Terabytes  $10^{12}$



Petabytes  $10^{15}$



Exabytes  $10^{18}$



Zettabytes  $10^{21}$



Yottabytes  $10^{24}$

### DATA SCIENCE

Involves applications of statistics, computer science, and software engineering, along with some other relevant fields (such as genetics).

#### Example: social media platforms

- **Scenario:** platforms like Facebook, Twitter, and Instagram generate massive amounts of data every second, including text, images, videos, likes, shares, and comments from billions of users worldwide.
- **Big data characteristics:** high in **volume** (terabytes or petabytes), **variety** (text, images, videos), **velocity** (real-time data streams), and **veracity** (data quality and trustworthiness).
- **Use case:** companies analyze user behavior, sentiment, and trends — for instance analyzing millions of tweets in real time to detect emerging trends, monitor brand reputation, or understand public sentiment.

#### Kaiser Permanente

A healthcare network based in California, has roughly **40,000,000,000,000,000 bytes** of data from electronic health records.

#### Flatiron Health

Analyzes **billions of data points** from cancer patients so that care can be enhanced.



## Missing Data

When collecting sample data, it is quite common to find that some values are **missing**. Ignoring missing data can sometimes **create misleading results**.

Body Temperatures (in degrees Fahrenheit) of Healthy Adults

| Subject | Sex | Smoke | Temperature Day 1 |       | Temperature Day 2 |       |
|---------|-----|-------|-------------------|-------|-------------------|-------|
|         |     |       | 8 AM              | 12 AM | 8 AM              | 12 AM |
| 1       | M   | Y     | 98.0              | 98.0  | 98.0              | 98.6  |
| 2       | M   | Y     | 97.0              | 97.6  | 97.4              | ----  |
| 3       | M   | Y     | 98.6              | 98.8  | 97.8              | 98.6  |

### Missing completely at random

If the likelihood of its being missing is **independent of its value** or any of the other values in the data set. That is, any data value is just as likely to be missing as any other data value.

Example: when using a keyboard to manually enter ages of survey respondents, the operator is distracted by a colleague singing "Fever" and makes the mistake of failing to enter the age of 37 years.

### Missing not at random

A data value is missing not at random if the missing value is **related to the reason that it is missing**.

Example: a survey question asks respondents to enter their annual income, but respondents with very low incomes skip this question because they find it embarrassing.



## Correcting for Missing Data

### How do we correct for missing data?

- 1 **Delete Cases:** one very common method for dealing with missing data is to delete all subjects having any missing values.

#### If missing completely at random

The remaining values are **not likely to be biased** and good results can be obtained, but with a smaller sample size.

#### If missing not at random

Deleting subjects that have any missing values can easily result in a **bias** among the remaining values, so results can be misleading.

- 2 **Impute Missing Values:** we "impute" missing data values when we **substitute values** for them.



## Collecting Sample Data

- When analyzing sample data, it is essential to use an **appropriate method** for collecting those sample data.
- Of particular importance is the **simple random sample.**
- If sample data are not collected in an appropriate way, the data may be so utterly useless that no amount of statistical torturing can salvage them.



## Placebo

### DEFINITION

A **placebo** is a harmless and ineffective pill, medicine, or procedure sometimes used for psychological benefit or sometimes used by researchers for comparison to other treatments.

- A placebo such as a **sugar pill** has no medicinal effect.
- Randomness with placebo/treatment groups is sometimes called the **"gold standard"** because it is so effective.



## Observational Study vs. Experiment

Statistical methods are driven by the data that we collect. We typically obtain data from **two distinct sources**.

### Observational study

Observe and measure specific characteristics **without attempting to modify** the individuals being studied.

Example: to understand the relationship between dietary habits and the risk of developing heart disease.

### Experiment

**Apply some treatment** and then proceed to observe its effects on the individuals. The individuals in experiments are called **experimental units**, and they are often called **subjects** when they are people.

Example: to determine whether a new drug is effective in reducing blood pressure compared to a placebo.



## Variables & Lurking Variables

Experiments are often better than observational studies because well-planned experiments typically **reduce the chance** of having the results affected by some variable that is not part of a study.

### Variable

A characteristic that takes on **different values** in different persons, places, or things.

Example: age, weight, height, marital status, heart rate, blood pressure.

### Lurking variable

One that **affects the variables** included in the study, but it is **not included** in the study.

## EXAMPLE 2 Ice Cream and Drownings

**Observational Study:** Observe past data to incorrectly conclude that ice cream causes drownings (based on data showing that increases in ice cream sales are associated with increases in drownings). The mistake is to miss the lurking variable of temperature and the failure to see that as the temperature increases, ice cream sales increase and drownings increase because more people swim.

**Experiment:** Conduct an *experiment* with one group treated with ice cream while another group gets no ice cream. We would see that the rate of drowning victims is about the same in both groups, so ice cream consumption has no effect on drownings.

Here, the experiment is clearly better than the observational study.



## Design of Experiments

Good design of experiments includes:

Replication



Blinding



Randomness

### Replication

The **repetition of an experiment** on more than one individual. Good use of replication requires sample sizes that are **large enough** so that we can see effects of treatments.

### Blinding

Used when study subjects **or** the investigators do not know who is in the control group and who is in the experimental group, with the purpose of **reducing bias**. Blinding is a way to get around the **placebo effect**, which occurs when an untreated subject reports an improvement in symptoms.

Double-blinding is used when study subjects **and** the investigators do not know who is in the control group and who is in the experimental group.

### Randomness

Used when individuals are assigned to different groups through a **process of random selection**. The logic behind randomness is to use chance as a way to create two groups that are similar. A **simple random sample** of  $n$  subjects is selected in such a way that every possible sample of the same size  $n$  has the same chance of being chosen — often called a **random sample**.



## Common Sampling Methods

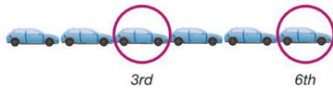
In addition to simple random sampling, here are some other sampling methods **commonly used for surveys.**



555-867-5309  
555-606-0842  
555-777-9311

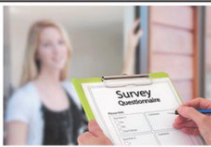
### Simple Random Sample

Select a sample of  $n$  subjects so that every sample of the same size  $n$  has the same chance of being selected.



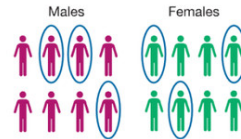
### Systematic Sample

Select every  $k$ th subject.



### Convenience Sample

Use data that are very easy to get.



### Stratified Sample

Subdivide population into strata (groups) with the same characteristics, then randomly sample within those strata.



### Cluster Sample

Partition the population in clusters (groups), then randomly select some clusters, then select all members of the selected clusters.

FIGURE 1-4 Common Sampling Methods

Credit: Pixsooz/Shutterstock



## Multistage Sampling

### EXAMPLE 3 Multistage Sample Design

The U.S. government's unemployment statistics are based on surveys of households. It is impractical to personally survey each household in a simple random sample, because they would be scattered all over the country, making it nearly impossible to contact each of them. Instead, the U.S. Census Bureau and the Bureau of Labor Statistics collaborate to conduct a survey called the Current Population Survey. A recent survey incorporates a multistage sample design, roughly following these steps:

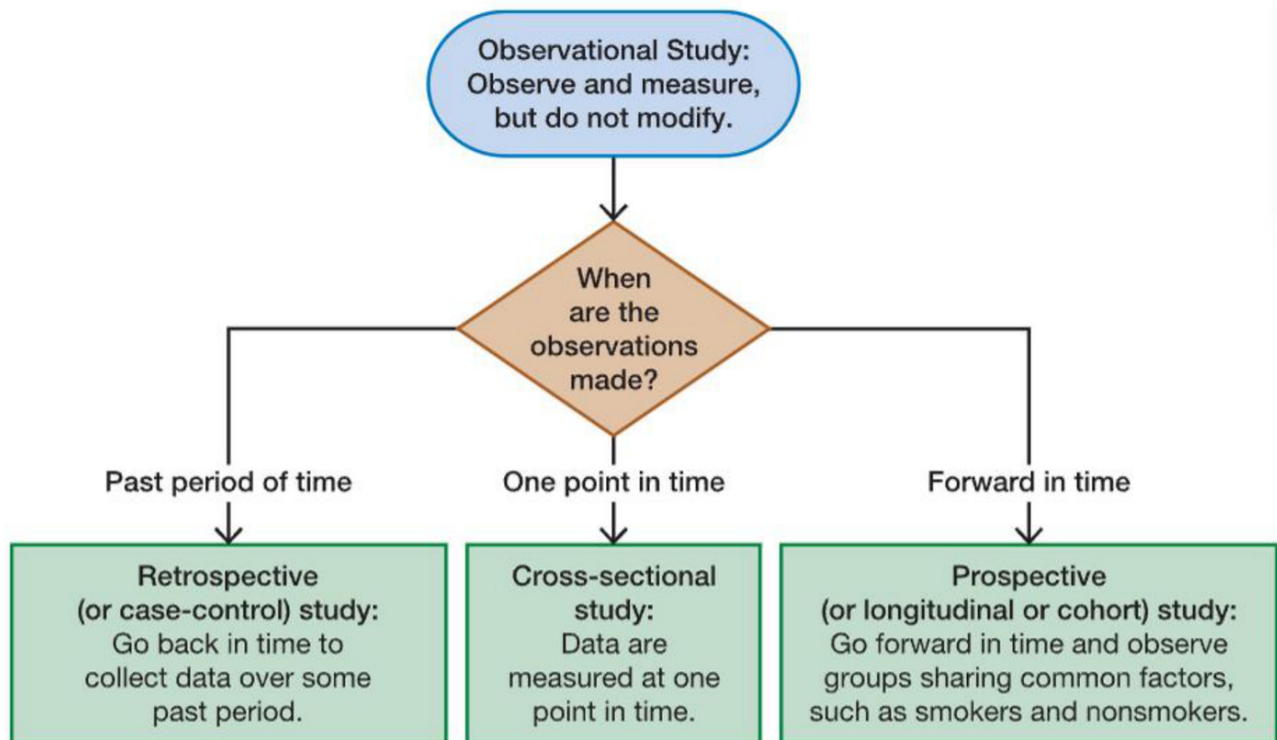
1. The entire United States is partitioned into 2025 different regions called *primary sampling units* (PSUs). The primary sampling units are metropolitan areas, large counties, or combinations of smaller counties. The 2025 primary sampling units are then grouped into 824 different strata.
2. In each of the 824 different strata, one of the primary sampling units is selected so that the probability of selection is proportional to the size of the population in each primary sampling unit.
3. Among the 824 selected primary sampling units, census data are used to randomly select about 60,000 households.
4. A responsible person in each of the 60,000 selected households is interviewed about the employment status of each household member of age 16 or older.

This multistage sample design includes a combination of random, stratified, and cluster sampling at different stages. The end result is a very complicated sampling design, but it is much more practical, less expensive, and faster than using a simpler design, such as a simple random sample. (Using a simple random sample would result in households that are far apart and difficult to contact.)

- Collect data by using **some combination** of the basic sampling methods.
- In a multistage sample design, pollsters select a sample in **different stages**, and each stage might use different methods of sampling.



## Types of Observational Studies



**FIGURE 1-5** Types of Observational Studies



### Confounding

#### DEFINITION

Confounding occurs when we see some effect, but we **can't identify the specific factor** that caused it.

- Try to plan the experiment so that confounding **does not occur**. How????



## Completely Randomized & Block Design

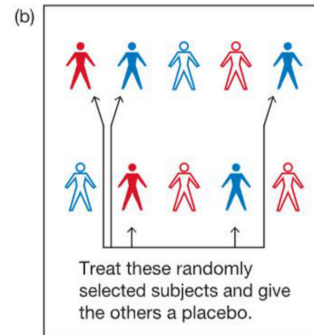
### Completely randomized experimental design

Assign subjects to different treatment groups through a **process of random selection.**

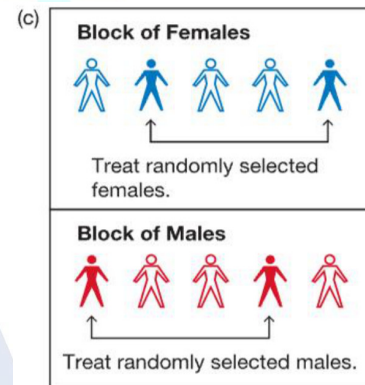
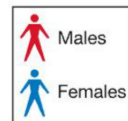
### Randomized block design

A **block** is a group of subjects that are similar, but blocks differ in ways that might affect the outcome of the experiment.

- Form blocks (or groups) of subjects with **similar characteristics.**
- Randomly assign treatments to subjects **within each block.**



**Completely randomized experimental design:**  
Use randomness to determine who gets the treatment and who gets the placebo.



**Randomized block design:**

1. Form a block of females and a block of males.
2. Within each block, randomly select subjects to be treated.

### EXAMPLE — ASPIRIN & HEART DISEASE

Form a block of males and a block of females, because it is known that the hearts of males and females can behave differently. By controlling for sex, this randomized block design **eliminates sex as a possible source of confounding.**



## Matched Pairs Design

### DEFINITION

Compare two treatment groups (such as treatment and placebo) by using **subjects matched in pairs** that are somehow related or have similar characteristics.



### Matched pairs design:

Get measurements from the same subjects before and after some treatment.



## Rigorously Controlled Design

### Rigorously controlled design

Carefully assign subjects to different treatment groups, so that those given each treatment are **similar in ways that are important to the experiment**.

- This can be **extremely difficult to implement**, and often we can never be sure that we have accounted for all of the relevant factors.



## Errors in Sample Collection

No matter how well you plan and execute the sample collection process, there is likely to be **some error in the results.**

Not in the results.

### DEFINITIONS

A **sampling error** (or **random sampling error**) occurs when the sample has been selected with a random method, but there is a discrepancy between a sample result and the true population result; such an error results from chance sample fluctuations.

A **nonsampling error** is the result of human error, including such factors as wrong data entries, computing errors, questions with biased wording, false data provided by respondents, forming biased conclusions, or applying statistical methods that are not appropriate for the circumstances.

A **nonrandom sampling error** is the result of using a sampling method that is not random, such as using a convenience sample or a voluntary response sample.



## Clinical Trials

### DEFINITION

A **clinical trial** is a prospective experiment designed to test the **safety and effectiveness** of a new drug, device, surgery technique, counseling method, and so on.



## Clinical Trials — Phases & Checkpoints

### Phase 1: Test for Safety

**Goal:** The main goal of this first phase is to assess the safety of the treatment and to identify the correct dosage.

**Sample Size:** This phase usually involves a small number of subjects, such as 20 to 100.

**Time and Cost:** Phase 1 can take several months and cost around \$4 million.

**Success Rate:** About 70% of medications proceed to phase 2.

### Phase 2: Test for Effectiveness

**Goal:** The main goal of phase 2 is to assess the effectiveness of the treatment.

**Sample Size:** Phase 2 uses a larger group, such as 100 to 300 subjects.

**Time and Cost:** Phase 2 can take several months up to two years at a cost of \$13 million.

**Success Rate:** About 50% of medications proceed to phase 3.

### Phase 3: Test for Effectiveness with Different Populations

**Goal:** The main goal of phase 3 is to collect data about safety and effectiveness with a larger and more diverse group of subjects.

**Sample Size:** Phase 3 uses a much larger group, such as several hundred to 3000 subjects.

**Time and Cost:** Phase 3 can last years and can cost around \$20 million.

**Success Rate:** About 60% of medications successfully complete phase 3.

### Phase 4: Monitor Results from Widespread Use

Phase 4 occurs when the medication receives widespread use. The effectiveness of the drug is monitored along with its safety. The safety is reflected through observations of adverse reactions.

## Checkpoints

- Before human testing begins with a medication, prior research must be conducted with animals and/or human cells. Based on the results, the FDA may or may not grant approval.
- An Institutional Review Board reviews the data, monitors the trials, and ensures that the trials are ethical and that they protect the rights and safety of the subjects.
- Upon completion of phase 3, FDA approval is required before the medication can be made available to the general public.
- In phase 4, the effectiveness and safety are monitored as the medication experiences widespread use.



## To Summarise...

### PROCESS 3

#### Conducting a statistical study

- **Prepare:** context, source of data, sampling method
- **Analyze:** graph, explore, apply statistical methods
- **Conclude:** statistical and practical significance

### SIGNIFICANCE 5%

#### Statistical vs. practical

- Statistically significant when a result is very unlikely by chance
- Achieved if the likelihood by chance is 5% or less
- Practical significance asks whether it justifies its use

### DATA 2

#### Basic types of data

- Quantitative: numbers representing counts or measurements
- Categorical: names or labels, not counts or measurements
- Quantitative splits into discrete and continuous

### LEVELS 4

#### Levels of measurement

- Nominal: categories only, cannot be ordered
- Ordinal: can be ordered, differences meaningless
- Interval: no natural zero · Ratio: natural zero

### SAMPLING 5

#### Common sampling methods

- Simple random, systematic, convenience
- Stratified (strata) and cluster (clusters)
- Multistage combines the basic methods in stages

### DESIGN 3

#### Good design of experiments

- Replication: repeat on more than one individual
- Blinding: subjects or investigators don't know the groups
- Randomness: assign to groups by random selection

## EXERCISES

**Chapter Quick Quiz:** Q1 to 10, p. 56 · **Review Exercise:** Q1, 4, 7, 9, p. 57–58