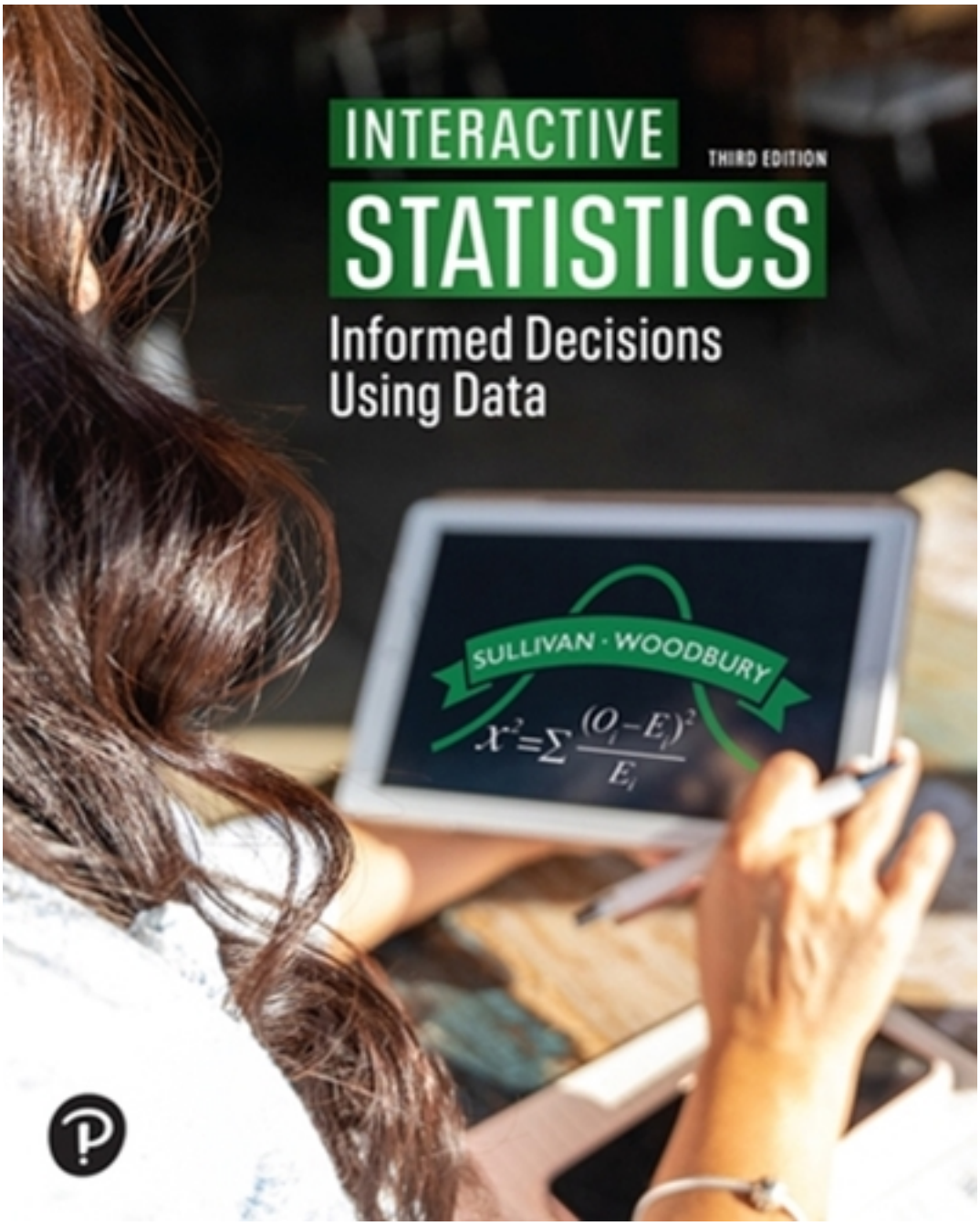
Test Bank for Interactive Statistics Informed Decisions Using Data 3rd Edition by Sullivan

CLICK HERE TO ACCESS COMPLETE Test Bank



Test Bank

INSTRUCTOR'S GUIDED NOTEBOOK

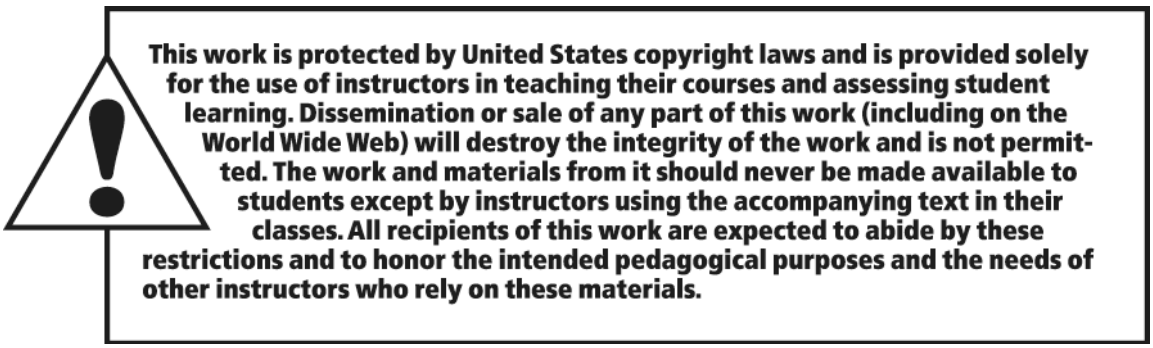
GEORGE WOODBURY

INTERACTIVE STATISTICS: INFORMED DECISIONS USING DATA

THIRD EDITION

Michael Sullivan, III
Joliet Junior College

George Woodbury
College of the Sequoias



Product Manager: Karen Montgomery
Content Analyst: Laura Briskman
Content Producer: Tamela Ambush

PPID: A103000242940

Table of Contents

Chapter 1 – Data Collection

OUTLINE

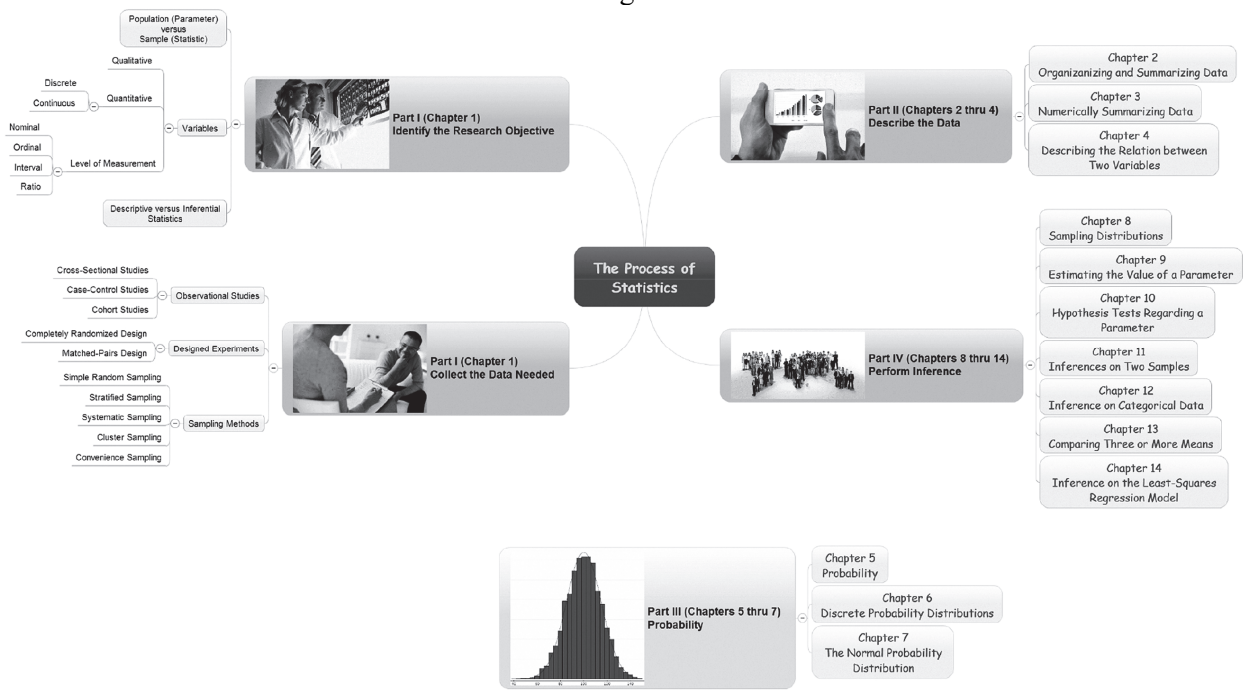
1.1 Introduction to the Practice of Statistics
1.2 Observational Studies versus Designed Experiments
1.3 Simple Random Sampling
1.4 Other Effective Sampling Methods
1.5 Bias in Sampling
1.6 The Design of Experiments

Putting It Together

Statistics plays a major role in many aspects of our lives. It is used in sports, for example, to help a general manager decide which player might be the best fit for a team. It is used in politics to help candidates understand how the public feels about various policies. And statistics is used in medicine to help determine the effectiveness of new drugs.

Used appropriately, statistics can enhance our understanding of the world. Used inappropriately, it can lend support to inaccurate beliefs. Understanding statistical methods will provide you with the ability to analyze and critique studies and the opportunity to become an informed consumer of information.

Understanding statistical methods will also enable you to distinguish solid analysis from bogus "facts."



Section 1.1
Introduction to the Practice of Statistics

Objectives

❶ Define Statistics and Statistical Thinking

❷ Explain the Process of Statistics

❸ Distinguish between Qualitative and Quantitative Variables

❹ Distinguish between Discrete and Continuous Variables

❺ Determine the Level of Measurement of a Variable

Objective 1: Define Statistics and Statistical Thinking

 *Answer the following as you watch the video.*

1) Write the definition of statistics below.

Statistics is the science of collecting, organizing, summarizing, and analyzing information to draw conclusions or answer questions. In addition, statistics is about providing a measure of confidence in any conclusion.

2) Data describes *characteristics* of individuals and can be either *numeric* or *non-numeric*.

Note: Data varies. Consider the students in your class. Is everyone the same height? No. Does everyone have the same color hair? No. So, within groups there is variation. Now consider yourself. Do you eat the same amount of food (as measured by calories) each day? No. Do you sleep the same number of hours each day? No. So, even considering individuals there is variation. One goal of statistics is to describe and understand sources of variation.

Objective 2: Explain the Process of Statistics

 *Answer the following while watching the animation.*

3) What is the entire group to be studied called?

The entire group to be studied is called the population.

4) What do we call a person or object that is a member of the population being studied?

An individual is a person or object that is a member of the population being studied.

5) Give the definition of a sample.

A sample is a subset of the population that is being studied.

6) What do we call a numerical summary of a sample?

A statistic is a numerical summary of a sample.

7) What do we call a numerical summary of a population?

A parameter is a numerical summary of a population.

8) Give the definition of descriptive statistics.

Descriptive statistics consist of organizing and summarizing data. They describe data through numerical summaries, tables, and graphs.

9) Give the definition of inferential statistics.

Inferential statistics uses methods that take a result from a sample, extend it to the population, and measure the reliability of the result.

 *Answer these questions after watching the animation.*

10) In the $100 experiment, what is the population? What is the sample?

Population: *The student body is the population.*

Sample: *The group of 40 students.*

11) Is the statement an example of descriptive statistics or inferential statistics? Circle the correct answer.

A) The percent of students in the survey who would return the money to the owner is 78%.

[Descriptive statistics] Inferential statistics

B) We are 95% confident that between 74% and 82% of all students would return the money.

Descriptive statistics [Inferential statistics]

12) Is the given measure a statistic or a parameter? Circle the correct answer.

A) The percentage of all students on your campus who own a car is 48.2%.

Statistic [Parameter]

B) Suppose a random sample of 100 students is obtained, and from this sample we find that 46% own a car.

[Statistic] Parameter

Watch the video on the process on statistics, and fill in the following steps.

The Process of Statistics

1. *Identify the research objective.*
A researcher must determine the question(s) he or she wants answered. The question(s) must be detailed so that it identifies the population that is to be studied.

2. *Collect the data needed to answer the question(s) posed in (1).*
Conducting research on an entire population is often difficult and expensive, so we typically look at a sample. This step is vital to the statistical process because if the data are not collected correctly, the conclusions drawn are meaningless. Do not overlook the importance of appropriate data collection.

3. *Describe the data.*
Descriptive statistics allow the researcher to obtain an overview of the data and can help determine the type of statistical methods the researcher should use.

4. *Perform inference.*
Apply the appropriate techniques to extend the results obtained from the sample to the population and report a level of reliability of the results.

Example 1 ***The Process of Statistics: Trust Your Neighbors?***

Pew Research conducted a poll and asked, "Do you trust all or most of your neighbors?"
The following statistical process allowed the researchers to conduct their study.

1. Identify the research objective.

The researchers wished to determine the percentage of adult Americans who trust all or most of their neighbors.

2. Collect the information needed to answer the question posed in (1).

The researchers surveyed a sample of 1628 adult Americans. Of those surveyed, 847 stated they trust all or most of their neighbors.

3. Describe the data.

Of the 1628 individuals in the survey, 52% stated that they trust all or most of their neighbors.

4. Perform inference.

The researchers feel 95% certain that the percentage of all adult Americans who trust all or most of their neighbors is somewhere between 49.5% and 54.5%.

Objective 3: Distinguish between Qualitative and Quantitative Variables

Define the following terms.

13) Qualitative variable:

Qualitative, or categorical, variables allow for classification of individuals based on some attribute or characteristic.

14) Quantitative variable:

Quantitative variables provide numerical measures of individuals. The values of a quantitative variable can be added or subtracted and provide meaningful results.

Example 2 ***Distinguishing between Qualitative and Quantitative Variables***

Determine whether the following variables are qualitative or quantitative.

A) Race

Qualitative

B) Temperature

Quantitative

C) Number of days during the past week that a college student studied

Quantitative

D) Zip code

Qualitative

Objective 4: Distinguish between Discrete and Continuous Variables

Define the following terms.

15) Discrete variable:

A discrete variable is a quantitative variable that has either a finite number of possible values or a countable number of possible values. A discrete variable cannot take on every possible value between any two possible values.

16) Continuous variable:

A continuous variable is a quantitative variable that has an infinite number of possible values that are not countable. A continuous variable may take on every possible value between any two values.

Example 3 ***Distinguishing between Discrete and Continuous Variables***

Determine whether the quantitative variables are discrete or continuous.

A) The number of heads obtained after flipping a coin five times.

Discrete

B) The number of cars that arrive at a McDonald's drive-through between 12:00 P.M. and 1:00 P.M.

Discrete

C) The distance a 2011 Toyota Prius can travel in city driving conditions with a full tank of gas.

Continuous

Define the following terms.

17) Data:

The list of observed values for a variable is data.

18) Qualitative data:

Qualitative data are observations corresponding to a qualitative variable.

19) Quantitative data:

Quantitative data are observations corresponding to a quantitative variable.

20) Discrete data:

Discrete data are observations corresponding to a discrete variable.

21) Continuous data:

Continuous data are observations corresponding to a continuous variable.

Example 4 ***Distinguishing between Variables and Data***

The following table presents data from a sample of on-street parking meters within the City of Seattle. (Data are from the City of Seattle Data Portal.)

Car	Payment Method	Amount Paid	Duration in Minutes	Side of Street	Parking Space Number
1	CREDIT CARD	$3.75	30	W	458
2	CREDIT CARD	$2.00	240	E	37
3	CREDIT CARD	$2.00	240	NE	18
4	PHONE	$1.38	225	SW	382
5	PHONE	$0.50	60	W	770
6	CREDIT CARD	$0.25	10	S	75
7	CREDIT CARD	$1.00	120	E	136
8	CREDIT CARD	$0.50	20	S	59
9	PHONE	$0.50	20	S	62
10	PHONE	$0.75	30	S	15
11	PHONE	$1.71	204	SW	382

Identify the individuals, variables, and data.

The individuals are the cars. The qualitative variables are payment method, side of street, and parking space number. The quantitative variables are amount paid (discrete – technically) and duration in minutes (continuous). The data are listed under the variables.

Objective 5: Determine the Level of Measurement of a Variable

List the characteristics used to determine what level of measurement a variable is.

22) Nominal:

A variable is at the nominal level of measurement if the values of the variable name, label, or categorize. In addition, the naming scheme does not allow for the values of the variable to be arranged in a ranked or specific order.

23) Ordinal:

A variable is at the ordinal level of measurement if it has the properties of the nominal level of measurement. However, the naming scheme allows for the values of the variable to be arranged in a ranked or specific order.

24) Interval:

A variable is at the interval level of measurement if it has the properties of the ordinal level of measurement and the differences in the values of the variable have meaning. A value of zero does not mean the absence of the quantity. Arithmetic operations such as addition and subtraction can be performed on the values of the variable.

25) Ratio:

A variable is at the ratio level of measurement if it has the properties of the interval level of measurement and the ratios of the values of the variable have meaning. A value of zero means the absence of the quantity. Arithmetic operations such as multiplication and division can be performed on the values of the variable.

Example 5 ***Determining the Level of Measurement of a Variable***

For each of the following variables, determine the level of measurement.
A) Race

Nominal

B) Temperature

Interval

C) Number of days during the past week that a college student studied

Ratio

D) Letter grade earned in your statistics class

Ordinal

Section 1.2
Observational Studies versus Designed Experiments

Objectives

❶ Distinguish between an Observational Study and a Designed Experiment

❷ Explain the Various Types of Observational Studies

Objective 1: Distinguish between an Observational Study and a Designed Experiment

 *Answer the following as you watch the first video and animation.*

1) Why is the "human" study mentioned in the video an observational study and not a designed experiment?

No attempt was made to influence the value of the explanatory variable, radio-frequency exposure.

2) Why is the "rat" study mentioned in the video a designed experiment and not an observational study?

There was an attempt to influence the individuals in the study because the value of the explanatory variable (exposure to radio frequency) was influenced.

3) What is the response variable in each study, and what is the explanatory variable?

The response variable is whether or not brain cancer was contracted, and the explanatory variable is the level of cell phone usage.

 *Answer the following after watching the first video and animation in this objective.*

4) In research, we wish to determine how varying an explanatory variable affects …

the response variable.

5) What does an observational study measure? Does an observational study attempt to influence the value of the response variable or explanatory variable?

An observational study measures the value of the response variable without attempting to influence the value of either the response or explanatory variables. That is, in an observational study, the researcher observes the behavior of individuals without trying to influence the outcome of the study.

6) Explain how to determine if a study is a designed experiment.

If a researcher randomly assigns the individuals in a study to a group, intentionally manipulates the value of an explanatory variable, and controls other explanatory variables at fixed values, then records the value of the response variable for each group, the study is a designed experiment.

 *Watch the second video (Which is Better ...) in this objective and answer the following.*

7) Why is the influenza study mentioned in the video an observational study and not a designed experiment?

The seniors decided whether or not to get a flu show. The researchers did not decide who would get a flu shot and did not impose any other conditions on the study.

8) List some changes that could be made to investigate the effectiveness of the flu shot with a designed experiment.

The researchers can take a group of seniors and decide which seniors would get the flu shot and which seniors will get a placebo.

9) List some lurking variables in the influenza study.

Age, heath status, mobility of the senior, ...

 *Answer the following after watching the second video.*

11) Do observational studies allow a researcher to claim causality?

In observational studies, we are not allowed to make statements of causality, meaning we cannot say that changes in the explanatory variable cause changes in the response variable. We can only say changes in the explanatory variable are associated with changes in the response variable.

12) Define confounding in a study.

Confounding in a study occurs when the effects of two or more explanatory variables are not separated. Therefore, any relation that may exist between an explanatory variable and the response variable may be due to some other variable or variables not accounted for in the study.

13) What is a lurking variable?

A lurking variable is an explanatory variable that was not considered in a study, but that affects the value of the response variable.

Define the following term.

14) Confounding variable:

A confounding variable is an explanatory variable that was considered in the study whose effect cannot be distinguished from a second explanatory variable in the study.

Note: The big difference between lurking variables and confounding variables is that lurking variables are not considered in the study whereas confounding variables are measured in the study.

Objective 2: Explain the Various Types of Observational Studies

 *Answer the following while watching the video.*

15) Define: Cross-sectional studies

Cross-sectional studies are observational studies that collect information about individuals at a specific point in time or over a very short period of time.

16) Define: Case-control studies

Case-control studies are observational studies that are retrospective, meaning they require individuals to look back in time or require the researcher to look at existing records. In case-control studies, individuals who have a certain characteristic may be matched with those who do not.

17) List some of the advantages of performing a case-control study over a cross-sectional study.

Case-control studies can be done relatively quickly and inexpensively.

18) List some difficulties that may occur and affect the outcomes of a case-control study.

There may an issue with poor memories when records do not exist.

19) Define: Cohort studies

Cohort Studies are observational studies that follow a group of individuals over a long period of time. Characteristics of the individuals are recorded and some individuals will be exposed to certain factors (not intentionally) and others will not. Because the data are collected over a long time period, these studies are prospective.

20) Explain why the Framingham Heart Study is a cohort study.

More than 10,000 individuals have been monitored since 1948.

Example 1 ***What Type of Study?***

Decide whether the following studies represent a designed experiment or an observational study. For those that represent an observational study, decide which type of observation study was conducted.

A) A survey of 1530 randomly selected individuals was conducted in which participants were asked a variety of questions relating to gambling habits along with satisfaction in basic psychological needs (such as autonomy, a feeling of satisfaction, and a feeling of being competent). The researchers found that weaker mental health was associated with increased gambling problems. Source: Vuorinen, I. et al. Basic psychological needs in gambling and gaming problems. Addictive Behavior Reports Volume 16, December 2022. https://doi.org/10.1016/j.abrep.2022.100445

Observational study, cross-sectional study

B) Researchers wanted to assess the long-term psychological effects on children evacuated during World War II. They obtained a sample of 169 former evacuees and a control group of 43 people who were children during the war but were not evacuated. The subjects' mental states were evaluated using questionnaires. It was determined that the psychological well-being of the individuals was adversely affected by evacuation. Source: Foster D, Davies S, and Steele H (2003) The evacuation of British children during World War II: a preliminary investigation into the long-term psychological effects. Aging & Mental Health (7)5.

Observational study, case-control study

C) Researchers wanted to determine whether there is bias in clinical practice among patients with cardiac chest pain. In this study a doctor was asked to recommend treatment for a white male patient and also recommend treatment for a Black female patient. The patients were of similar age, wearing similar attire, and expressing similar physical complaints. Whether the doctor assessed the white male or Black female first was determined randomly. The results of the study indicated that the Black female patient received recommendations that would have resulted in undertreatment while the white male patient received appropriate treatment recommendations. Source: Centola, D., Guilbeault, D., Sarkar, U. et al. The reduction of race and gender bias in clinical treatment recommendations using clinician peer networks in an experimental setting. Nat Commun 12, 6585 (2021). https://doi.org/10.1038/s41467-021-26905-5

Designed experiment

D) The Cancer Prevention Study-3 (CPS-3) started in 2015. Individuals between the ages of 30 and 65 who had no personal history of cancer enrolled in the study. Study participants completed an initial survey and providing a blood sample. Then, every three years a survey is administered to the participants in which they provide information such as exposure to COVID-19, changes in weight, diet, and physical activity. The blood samples are used to identify genetic risks associated with cancer. The study is expected to last through 2043. Other studies out of CPS-3 include asking participants to wear personal activity monitors to determine how physical activity may be associated with the risk of contracting cancer.

Observational study, cohort study

21) It is not always possible to conduct an experiment. Explain why we could not conduct an experiment to investigate the perceived link between high tension wires and leukemia (on humans).

It would be unethical to deliberately expose a person to something that could cause leukemia.

22) There is no point in reinventing the wheel. List some agencies that regularly collect data that are available to the public.

The Centers for Disease Control and Prevention, the Internal Revenue Service, and the Department of Justice collect data that is available to the public. Also, the University of Chicago's General Social Survey is a useful source of data.

23) What is a census?

A census is a list of individuals in a population along with certain characteristics of each individual.

24) What is web scraping?

Web scraping, or data mining, is the process of extracting data from the Internet.

Section 1.3
Simple Random Sampling

Objective

❶ Obtain a Simple Random Sample

Note: Sampling

Observational studies can be conducted by administering a survey. When administering a survey, the researcher must first identify the population that is to be targeted.

1) Define random sampling:

Random sampling is the process of using chance to select individuals from a population to be included in the sample.

For the results of a survey to be reliable, the characteristics of the individuals in the sample must be representative of the characteristics of the individuals in the population.

The key to obtaining a sample representative of a population is to let chance or randomness play a role in dictating which individuals are in the sample, rather than convenience.

If convenience is used to obtain a sample, the results of the survey are meaningless.

Objective 1: Obtain a Simple Random Sample

2) What is a simple random sample?

A sample of size n from a population of size N is obtained through simple random sampling if every possible sample of size n has an equal chance of occurring. The sample is then called a simple random sample.

The number of individuals in the sample is always less than the number of individuals in the population.

Example 1 **Illustrating Simple Random Sampling**

Sophie has four tickets to a concert. Six of her friends, Yolanda, Michael, Kevin, Marissa, Annie, and Katie, have all expressed an interest in going to the concert. Sophie decides to randomly select three of her six friends to attend the concert.

A) List all possible samples of size $n = 3$ from the population of size $N = 6$. Once an individual is chosen, he/she cannot be chosen again.

Yolanda, Michael, Kevin
Yolanda, Michael, Katie
Yolanda, Kevin, Katie
Yolanda, Annie, Katie
Michael, Kevin, Katie
Michael, Annie, Katie
Kevin, Katie, Annie
Yolanda, Michael, Marissa
Yolanda, Kevin, Marissa
Yolanda, Marissa, Annie
Michael, Kevin, Marissa
Michael, Marissa, Annie
Kevin, Marissa, Annie
Marissa, Annie, Katie
Yolanda, Michael, Annie
Yolanda, Kevin, Annie
Yolanda, Marissa, Katie
Michael, Kevin, Annie
Michael, Marissa, Katie
Kevin, Marissa, Katie

B) Comment on the likelihood of the sample containing Michael, Kevin, and Marissa.

There is a 1 in 20 chance that the simple random sample will contain these 3 friends.

How do we select the individuals in a simple random sample?
Typically, each individual in the population is assigned a unique number between 1 and N, where N is the size of the population. Then n distinct random numbers are selected, where n is the size of the sample. To number the individuals in the population, we need a frame - a list of all the individuals within the population.

Answer the following after watching the animation on simple random sampling.

3) What is the frame in this animation?

The list of the 33 students enrolled in his class are the frame.

7) Explain why a second sample of 5 students will most likely be different than the first sample of 5 students?

The second sample will most likely be different because chance is used to select the individuals.

8) Explain why inferences based on samples vary.

Samples vary because the individuals vary from sample to sample.

Example 2 **Obtaining a Simple Random Sample**

The accounting firm of Senese and Associates has grown. To make sure their clients are still satisfied with the services they are receiving, the company decides to send a survey out to a simple random sample of 5 of its 30 clients.

TABLE 3

01. ABC Electric	11. Fox Studios	21. R&Q Realty
02. Brassil Construction	12. Haynes Hauling	22. Ritter Engineering
03. Bridal Zone	13. House of Hair	23. Simplex Forms
04. Casey's Glass House	14. John's Bakery	24. Spruce Landscaping
05. Chicago Locksmith	15. Logistics Management, Inc.	25. Thors, Robert DDS
06. DeSoto Painting	16. Lucky Larry's Bistro	26. Travel Zone
07. Dino Jump	17. Moe's Exterminating	27. Ultimate Electric
08. Euro Car Care	18. Nick's Tavern	28. Venetian Gardens Restaurant
09. Farrell's Antiques	19. Orion Bowling	29. Walker Insurance
10. First Fifth Bank	20. Precise Plumbing	30. Worldwide Wireless

Answers will vary. The sample in the video is
1: ABC Electric, 7: Dino Jump, 26: Travel Zone, 11: Fox Studios, 23: Simplex Forms

Section 1.4
Other Effective Sampling Methods

Objectives

❶ Obtain a Stratified Sample

❷ Obtain a Systematic Sample

❸ Obtain a Cluster Sample

Objective 1: Obtain a Stratified Sample

1) Explain how to obtain a stratified sample.

A stratified sample is obtained by dividing the population into nonoverlapping groups called strata and then obtaining a simple random sample from each stratum. The individuals within each stratum should be homogenous (similar) in some way.

Example 1 **Obtaining a Stratified Sample**

The president of DePaul University wants to conduct a survey to determine the community's opinion regarding campus safety. The president divides the DePaul community into three groups: resident students, nonresident (commuting) students, and staff (including faculty) so that he can obtain a stratified sample.
Suppose there are 6,204 resident students, 13,304 nonresident students, and 2,401 staff, for a total of 21,909 individuals in the population. What percent of the DePaul community is made up of each group?

Resident students: 6204/21,909 = 28%
Nonresident students: 13,304/21,909 = 61%
Staff: 2401/21,909 = 11%

The president wants to obtain a sample of size 100, with the number of individuals selected from each stratum weighted by the population size. How many individuals should be selected from each stratum?

28 resident students, 61 nonresident students, and 11 staff.

To obtain the stratified sample, construct a simple random sample within each group.

Answers will vary for each student. For resident students, there should be 28 unique numbers from 1 to 6204. For nonresident students, there should be 61 unique numbers from 1 to 13,304. For staff, there should be 11 unique numbers from 1 to 2401.

Objective 2: Obtain a Systematic Sample

2) Explain how to obtain a systematic sample.

A systematic sample is obtained by selecting every kth individual from the population. The first individual selected corresponds to a number between 1 and k.

Note: Because systematic sampling does not require a frame, it is a useful technique when you cannot gather a list of the individuals in the population.

Example 2 **Obtaining a Systematic Sample without a Frame**

The manager of Kroger Food Stores wants to measure the satisfaction of the store's customers. Design a sampling technique that can be used to obtain a sample of 40 customers.

The manager decides to survey every 7^{th} customer. In the example, the manager starts with customer number 3. The customers who are surveyed under those conditions: 3, 10, 17, …, 276.

 *Answer the following after watching the first video after Example 2.*

3) What can result from choosing a value of *k* that is too small?

The sample size may be reached too soon, and the sample will only represent a portion of the population.

4) What can result from choosing a value of *k* that is too large?

The desired sample size may not be achieved.

 *Answer the following after watching the second video after Example 2.*

5) Explain how to determine the value of *k* if the population size *N* is known.

Divide the population size N by the desired sample size n, and round down to the nearest integer.

6) List the five steps in obtaining a systematic sample.

Step 1

If possible, approximate the population size, N.

Step 2

Determine the sample size, n.

Step 3

Divide N by n and round down to the nearest integer. This value is k.

Step 4

Randomly select a number between 1 and k. Call this number p.

Step 5

The sample will consist of the following individuals :p, p + k, p + 2k, …, p + (n – 1)k

---

Objective 3: Obtain a Cluster Sample

7) What is a cluster sample?

A cluster sample is obtained by selecting all individuals within a randomly selected collection or group of individuals.

Example 3 **Obtaining a Cluster Sample**

A sociologist wants to gather data regarding household income within the city of Boston. Obtain a sample using cluster sampling.

The sociologist numbers the blocks from 1 to 10,493, and randomly selects 20 clusters (city blocks) to survey.
Blocks selected in video example:

3951	*6676*	*8408*	*3462*	*10321*
2532	*5585*	*8198*	*8500*	*4025*
1682	*6633*	*4528*	*9887*	*5709*
6917	*7919*	*8200*	*2685*	*8142*

Read the screen "Issues to Consider in Cluster Sampling" and answer the following.

8) If the clusters have homogeneous individuals, is it better to have more clusters with fewer individuals in each cluster or fewer clusters with more individuals in each cluster?

If the clusters have homogeneous individuals, it is better to have more clusters with fewer individuals in each cluster.

9) If the clusters have heterogeneous individuals, is it better to have more clusters with fewer individuals in each cluster or fewer clusters with more individuals in each cluster?

If each cluster is heterogeneous, fewer clusters with more individuals in each cluster are appropriate.

10) Define convenience sampling:

In a convenience sample the individuals are easily obtained and not based on randomness.

Note: The most popular convenience samples are those in which the individuals in the sample are self-selected**,** meaning the individuals themselves decide to participate in the survey. Self-selected surveys are also called voluntary response samples.

11) List an example of a scenario involving multistage sampling.

Nielsen Media Research uses multistage sampling to investigate TV viewing habits.

Note: Sample Size Considerations

Researchers need to know how many individuals they must survey to draw conclusions about the population within some predetermined margin of error. They must find a balance between the reliability of the results and the cost of obtaining these results. The bottom line is that time and money determine the level of confidence researchers will place on the conclusions drawn from the sample data. The more time and money researchers have available, the more accurate the results of the statistical inference.

Section 1.5
Bias in Sampling

Objective

❶ Explain the Sources of Bias in Sampling

Objective 1: Explain the Sources of Bias in Sampling

1) Define bias.

If the results of the sample are not representative of the population, then the sample has bias.

2) List the three sources of bias in sampling:

- *Sampling bias*
- *Nonresponse bias*
- *Response bias*

 *Answer the following after watching the video "Sampling Bias".*

3) What is sampling bias?

Sampling bias means that the technique used to obtain the sample's individuals tends to favor one part of the population over another.

4) Does a convenience sample have sampling bias?

Any convenience sample has sampling bias because the individuals are not chosen through a random sample.

5) What is undercoverage?

Undercoverage is a specific type of sampling bias. This type of bias occurs when the population of one segment of the population is lower in a sample than it is in the population.

 *Answer the following after watching the video "Nonresponse Bias".*

6) When does nonresponse bias exist?

Nonresponse bias exists when individuals selected to be in the sample who do not respond to the survey have different opinions from those who do.

7) List two causes of nonresponse bias.

Nonresponse bias can occur because individuals selected for the sample do not wish to respond or the interviewer was unable to contact them.

8) List one tool that can be used to control nonresponse bias?

Nonresponse bias can be controlled using callbacks.

 *Answer the following after watching the video "Response Bias".*

9) Under what conditions does response bias exist?

Response bias exists when the answers on a survey do not reflect the true feelings of the respondent.

Note: Response bias can occur through interviewer error, misrepresented answers, wording of questions, ordering of questions or words, type of question, or data-entry error.

Note: An open question allows the respondent to choose his or her response (free response).

Note: A closed question requires the respondent to choose from a list of predetermined responses (multiple choice).

Note: Can a Census Have Bias?

A question on a census form could be misunderstood, thereby leading to response bias in the results. It is often difficult to contact each individual in a population. For example, the U.S. Census Bureau is challenged to count each homeless person in the country, so the census data published by the U.S. government likely suffers from nonresponse bias.

Define the following terms.

10) Nonsampling Error:

Nonsampling errors result from undercoverage, nonresponse bias, response bias, or data-entry error. Such errors could also be present in a census.

11) Sampling error:

Sampling error results from using a sample to estimate information about a population. This type of error occurs because a sample gives incomplete information about a population.

Section 1.6
The Design of Experiments

Objectives

❶ Describe the Characteristics of an Experiment

❷ Explain the Steps in Designing an Experiment

❸ Explain the Completely Randomized Design

❹ Explain the Matched-Pairs Design

Objective 1: Describe the Characteristics of an Experiment

Define the following terms after watching the video.

1) Experiment:

An experiment is a controlled study conducted to determine the effect varying one or more explanatory variables or factors has on a response variable.

2) Factor:

Factors are explanatory variables, and affect the response variable.

3) Treatment:

Any combination of the values of the factors.

4) Experimental unit:

An experimental unit (or subject) is a person, object or some other well-defined item upon which a treatment is applied.

5) Control group:

A control group serves as a baseline treatment that can be used to compare it to other treatments.

6) Placebo:

A placebo is an innocuous medication, such as a sugar tablet, that looks, tastes, and smells like the experimental medication.

7) Blinding:

Blinding refers to nondisclosure of the treatment an experimental unit is receiving.

8) Single-blind

A single-blind experiment is one in which the experimental unit (or subject) does not know which treatment they are receiving.

9) Double-blind

A double-blind experiment is one in which neither the experimental unit nor the researcher in contact with the experimental unit knows which treatment the experimental unit is receiving.

Example 1 **The Characteristics of an Experiment**

Lipitor is a cholesterol-lowering drug made by Pfizer. In the Collaborative Atorvastatin Diabetes Study (CARDS), the effect of Lipitor on cardiovascular disease was assessed in 2838 subjects, ages 40 to 75, with type 2 diabetes, without prior history of cardiovascular disease. In this placebo-controlled, double-blind experiment, subjects were randomly allocated to either Lipitor 10 mg daily (1428) or placebo (1410) and were followed for 4 years. The response variable whether there was an occurrence of any major cardiovascular event or not.

Lipitor significantly reduced the rate of major cardiovascular events (83 events in the Lipitor group versus 127 events in the placebo group). There were 61 deaths in the Lipitor group versus 82 deaths in the placebo group.

A) What does it mean for the experiment to be placebo-controlled?

The placebo control group serves as a baseline against which to compare the results from the group receiving Lipitor.

B) What does it mean for the experiment to be double-blind?

The subjects, as well as the individual monitoring the subjects, do not know whether the subjects are receiving Lipitor or the placebo.

C) What is the population for which this study applies? What is the sample?

Population: Individuals from 40 to 75 years of age with type 2 diabetes without a prior history of cardiovascular disease.
Sample: The 2838 subjects in the study.

D) What are the treatments?

The treatments are 10 mg of Lipitor or a placebo daily.

E) What is the response variable? Is it qualitative or quantitative?

Response variable: Whether the subject had any major cardiovascular event, such as a stroke, or not. The response variable is qualitative.

Objective 2: Explain the Steps in Designing an Experiment

Steps in Conducting a Designed Experiment
Fill in each step.

Step 1: *Identify the Problem to Be Solved.*

The statement of the problem should be as explicit as possible and should provide the experimenter with direction. The statement must also identify the response variable and the population to be studied. Often, the statement is referred to as the *claim.*

Step 2: *Determine the Factors That Affect the Response Variable.*

The factors are usually identified by an expert in the field of study. In identifying the factors, ask, "What things affect the value of the response variable?" After the factors are identified, determine which factors to fix at some predetermined level, which to manipulate, and which to leave uncontrolled.

Step 3: *Determine the Number of Experimental Units.*

As a general rule, choose as many experimental units as time and money allow. Techniques exist for determining sample size, provided certain information is available.

Step 4: *Determine the level of each factor.*

Factors can be dealt with in two ways - control or randomize.
Control means to either set the factor at one value throughout the experiment or set the level of the factor at various levels).
Randomize means to randomly assign the experimental units to various treatment groups.

Step 5: *Conduct the experiment.*

Replication occurs when each treatment is applied to more than one experimental unit.

Step 6: *Test the claim.*

Inferential statistics is a process in which generalizations about a population are made on the basis of results obtained from a sample.

Objective 3: Explain the Completely Randomized Design

10) What is a completely randomized design?

A completely randomized design is one in which each experimental unit is randomly assigned to a treatment.

Example 2 **A Completely Randomized Design**

A farmer wishes to determine the optimal level of a new fertilizer on his soybean crop. Design an experiment that will assist the farmer.

Step 1: So the response variable will be crop yield.

Step 2: Factors that affect crop yield include fertilizer, precipitation, sunlight, method of tilling the soil, type of soil, plant, and temperature.

Step 3: 60 soybean plants

Step 4:
Treatment A: 20 soybean plants receive no fertilizer.
Treatment B: 20 soybean plants receive 2 teaspoons of fertilizer per gallon of water every 2 weeks.
Treatment C: 20 soybean plants receive 4 teaspoons of fertilizer per gallon of water every 2 weeks.

Step 5: Assign each plant to a treatment group. At the end of the growing season, we determine the crop yield for each plant.

Step 6: We determine any differences in yield among the three treatment groups.

Objective 4: Explain the Matched-Pairs Design

13) What is a matched-pairs design?

A matched-pairs design is an experimental design in which the experimental units are paired up.

The pairs are selected so that they are related in some way.

There are only two levels of treatment in a matched-pairs design.

Example 3 **A Matched-Pairs Design**

An educational psychologist wants to determine whether listening to music has an effect on a student's ability to learn. Design an experiment to help the psychologist answer the question.

Match students according to IQ and gender.

Flip a coin to determine which student gets the quiet room and which student gets the room with music playing in the background.

Each student will be given a statistics textbook and asked to study Section 1.1.
After 2 hours, the students will enter a testing center and take a short quiz on the material in the section.
We compute the difference in the scores of each matched pair.
Any differences in scores will be attributed to the treatment.

Chapter 2 – Organizing and Summarizing Data

OUTLINE

Putting It Together

Chapter 1 discussed how to identify the research objective and collect data. We learned that data can be obtained from either observational studies or designed experiments. When data are obtained, they are referred to as **raw data**.

The purpose of this chapter is to learn how to organize raw data into a meaningful form so that we can understand what the data are telling us. The first step in determining how to organize raw data is to determine whether the data is qualitative or quantitative.

Section 2.1
Organizing Qualitative Data

Objectives

❶ Organize Qualitative Data in Tables

❷ Construct Bar Graphs

❸ Construct Pie Charts

Objective 1: Organize Qualitative Data in Tables

1) What is used to list each category of data and the number of occurrences for each category of data?

A frequency distribution lists each category of data and the number of occurrences for each category of data

Example 1 ***Organizing Qualitative Data into a Frequency Distribution***

A physical therapist wants to determine types of rehabilitation required by her patients. To do so, she obtains a simple random sample of 30 of her patients and records the body part requiring rehabilitation. (See Table 1.) Construct a frequency distribution of location of injury.

Table 1

Back	Back	Hand	Wrist	Back	Back
Groin	Elbow	Back	Back	Back	Groin
Shoulder	Shoulder	Hip	Knee	Hip	Shoulder
Neck	Knee	Knee	Shoulder	Shoulder	Neck
Back	Back	Back	Back	Knee	Back

Data from Krystal Catton, student at Joliet Junior College

Answer:

Body Part	Frequency
Back	12
Hand	2
Wrist	2
Groin	1
Elbow	1
Shoulder	4
Hip	2
Knee	5
Neck	1

2) In any frequency distribution, it is a good idea to add up the frequency column. What should the total be equal to?

The total should equal the number of observations.

3) Define the relative frequency of a category.

The relative frequency is the proportion (or percent) of observations within a category.

4) What is a relative frequency distribution?

A relative frequency distribution lists each category of data together with the relative frequency.

Example 2 ***Constructing a Relative Frequency Distribution of Qualitative Data***

Using the summarized data in Table 2, construct a relative frequency distribution.

Table 2

Body Part	**Frequency**
Back	12
Hand	2
Wrist	2
Groin	1
Elbow	1
Shoulder	4
Hip	2
Knee	5
Neck	1

Answer

Body Part	***Relative Frequency***
Back	*0.4*
Hand	*0.0667*
Wrist	*0.0667*
Groin	*0.0333*
Elbow	*0.0333*
Shoulder	*0.1333*
Hip	*0.0667*
Knee	*0.1667*
Neck	*0.0333*

5) When working with a relative frequency distribution, what should the total of the relative frequencies be equal to? Why?

The total should be 1, which represents 100% of the data.

Objective 2: Construct Bar Graphs

6) Explain how a bar graph is constructed? What do the heights of each rectangle represent?

A bar graph is constructed by labeling each category of data on either the horizontal or vertical axis and the frequency or relative frequency of the category on the other axis. Rectangles of equal width are drawn for each category. The height of each rectangle represents the category's frequency or relative frequency.

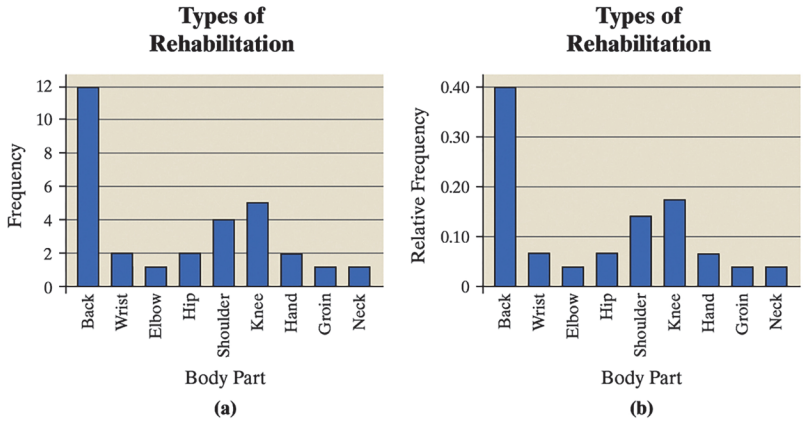
Example 3 ***Constructing a Frequency and Relative Frequency Bar Graph***

Use the data summarized in Table 3 to construct a frequency bar graph and relative frequency bar graph.

Table 3

Body Part	Frequency	Relative Frequency
Back	12	0.4
Hand	2	0.0667
Wrist	2	0.0667
Groin	1	0.0333
Elbow	1	0.0333
Shoulder	4	0.1333
Hip	2	0.0667
Knee	5	0.1667
Neck	1	0.0333

Answer:



7) What is a Pareto chart?

A Pareto chart is a bar graph whose bars are drawn in decreasing order of frequency or relative frequency.

8) Explain why it is best to use relative frequencies when comparing data sets.

Because different sample or population sizes make comparisons using frequencies difficult or misleading.

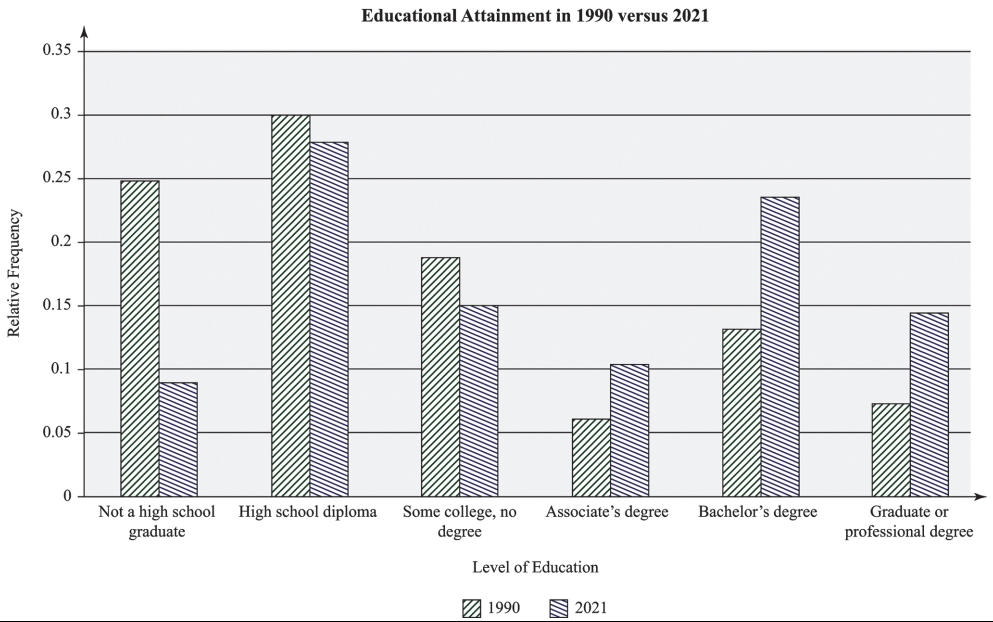
Example 4 ***Comparing Two Data Sets***

The frequency data in Table 4 represent the educational attainment (level of education) in 1990 and 2021 of adults 25 years and older who are U.S. residents. The data are in thousands. So 39,344 represents 39,344,000.

Table 4

Educational Attainment	**1990**	**2021**
Not a high school graduate	39,344	20,054
High school diploma	47,643	62,547
Some college, no degree	29,780	33,455
Associate's degree	9792	23,487
Bachelor's degree	20,833	52,805
Graduate or professional degree	11,478	32,232
Totals	158,870	224,580

A) Draw a side-by-side relative frequency bar graph of the data.



B) The side-by-side relative frequency bar graph shows additional information that was not easy to identify from the frequency table in Table 4. Comment on the interesting features of the side-by-side relative frequency bar graph.

The side-by-side bar graph illustrates that the proportion of Americans 25 years and older who had some college but no degree was higher in 1990. This information is not clear from the frequency table because the total population sizes are different. The increase in the number of Americans who did not complete a college degree is due partly to the increases in the size of the population. In addition, the number of individuals with a high school diploma increased by approximately 15 million individuals (62,547 thousand versus 37,643 thousand). However, from the side-by-side bar graph, we see that the proportion of Americans 25 years and older who had a high school diploma is actually higher in 1990. It is also clear that adult Americans have more education in 2021 than in 1990 with a much higher percentage of the population having at least a bachelor's degree.

9) Explain when it would be preferable to use horizontal bars rather than vertical bars when constructing a bar graph.

Horizontal bars are preferable when category names are lengthy.

Objective 3: Construct Pie Charts

10) What is a pie chart?

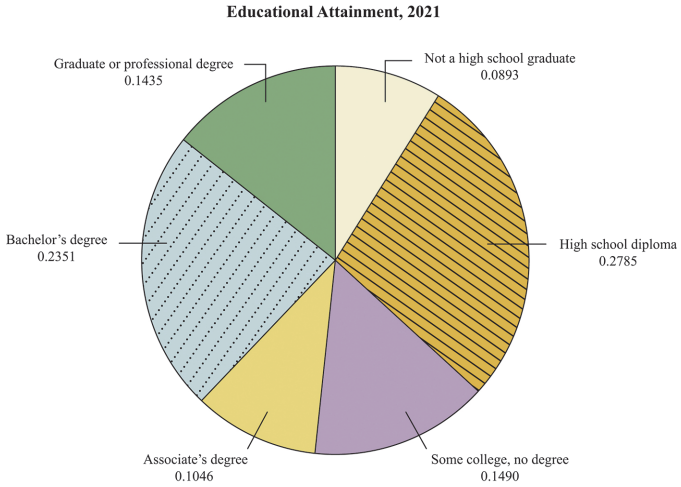
A pie chart is a circle divided into sectors. Each sector represents a category of data. The area of each sector is proportional to the frequency of the category.

Example 5 ***Constructing a Pie Chart***

The frequency data presented in Table 6 represent the educational attainment of U.S. residents 25 years and older in 2021. The data are in thousands so 20,054 represents 20,054,000. Construct a pie chart of the data.

Table 6

Educational Attainment	**2021**
Not a high school graduate	20,054
High school diploma	62,547
Some college, no degree	33,455
Associate's degree	23,487
Bachelor's degree	52,805
Graduate or professional degree	32,232
Totals	224,580



Answer the following after watching the video.

11) Which graph, a pie chart or a bar graph, is better at comparing one category to another category?

A bar graph is better at comparing one category to another category.

12) Which graph, a pie chart or a bar graph, is better at comparing one category to the whole?

A pie chart is better at comparing one category to the whole.

Section 2.2
Organizing Quantitative Data: The Popular Displays

Objectives

❶ Organize Discrete Data in Tables

❷ Construct Histograms of Discrete Data

❸ Organize Continuous Data in Tables

❹ Construct Histograms of Continuous Data

❺ Draw Dot Plots

❻ Identify the Shape of a Distribution

Objective 1: Organize Discrete Data in Tables

1) What do we use to create the classes when the number of distinct data values of a discrete variable is small?

We use the values of a discrete variable to create the classes when the number of distinct data values is small.

Example 1 ***Constructing Frequency and Relative Frequency Distributions from Discrete Data***

The manager of a Wendy's® fast-food restaurant wants to know the typical number of customers who arrive during the lunch hour. The data represent the number of customers who arrive at Wendy's for 40 randomly selected 15-minute intervals of time during lunch.
Construct a frequency and relative frequency distribution.

Number of Arrivals at Wendy's

7	6	6	6	4	6	2	6
5	6	6	11	4	5	7	6
2	7	1	2	4	8	2	6
6	5	5	3	7	5	4	6
2	2	9	7	5	9	8	5

Answer:

Number of Customers	Frequency	Relative Frequency
1	1	0.025
2	6	0.15
3	1	0.025
4	4	0.1
5	7	0.175
6	11	0.275
7	5	0.125
8	2	0.05
9	2	0.05
10	0	0.0
11	1	0.025

Objective 2: Construct Histograms of Discrete Data

2) Explain how a histogram is constructed.

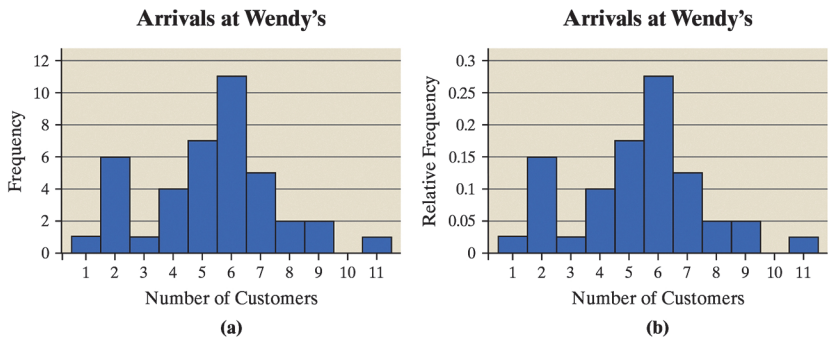
A histogram is constructed by drawing rectangles for each class of data. The height of each rectangle is the frequency or relative frequency of the class. The width of each rectangle is the same, and the rectangles touch each other.

Example 2 ***Drawing a Histogram of Discrete Data***

Construct a frequency histogram and a relative frequency histogram using the data in Table 9. Recall that this table summarizes the data for the number of customers who arrive at Wendy's for 40 randomly selected 15-minute intervals of time during lunch.

Table 9

Number of Customers	Frequency	Relative Frequency
1	1	0.025
2	6	0.15
3	1	0.025
4	4	0.1
5	7	0.175
6	11	0.275
7	5	0.125
8	2	0.05
9	2	0.05
10	0	0.0
11	1	0.025



Objective 3: Organize Continuous Data in Tables

Note: When a data set consists of a large number of different discrete data values or when a data set consists of continuous data, we must create classes by using intervals of numbers.

Define the following terms.

3) Lower class limit

The lower class limit is the smallest value within the class.

4) Upper class limit

The upper class limit is the largest value within the class.

5) Class width

The upper class limit is the largest value within the class.

6) When creating classes for a frequency distribution, the classes must not overlap.

7) What is an open-ended table?

A table is open-ended if the first class has no lower class limit or the last class has no upper class limit.

Example 3 ***Organizing Continuous Data into a Frequency and Relative Frequency Distribution***

The data in Table 12 represent the total fine including late penalties, in dollars, for a simple random sample of 50 parking and camera violations in the City of New York. Construct a frequency and relative frequency distribution of the data.

Table 12
Total Fine Including Late Penalties (In Dollars)

211.09	209.29	150.56	271.20	210.20	125.76	190.72	187.23	229.18	125.00
105.00	256.69	262.99	256.70	322.61	243.80	236.91	260.00	162.46	65.00
227.59	224.35	223.99	155.40	193.42	151.46	127.37	140.06	216.19	210.25
105.23	125.21	208.40	208.39	207.79	124.85	124.67	148.02	206.29	183.16
147.70	206.01	123.41	204.49	134.66	167.65	121.25	131.76	120.53	143.52

Source: NYC Open Data

Answer:

Class (Amount of Fine)	Frequency	Relative Frequency
50-74.99	1	0.02
75-99.99	0	0
100-124.99	7	0.14
125-149.99	10	0.20
150-174.99	5	0.10
175-199.99	4	0.08
200-224.99	13	0.26
225-249.99	4	0.08
250-274.99	5	0.10
275-299.99	0	0
300-324.99	1	0.02

Objective 4: Construct Histograms of Continuous Data

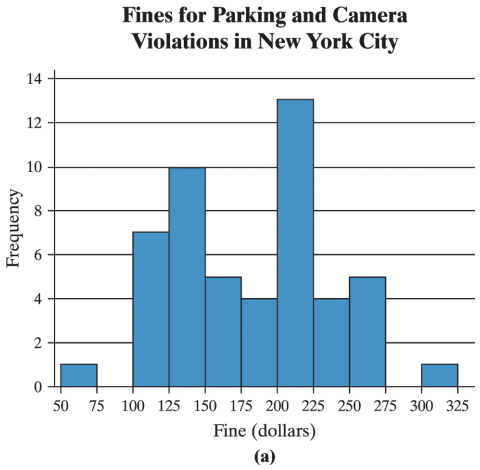
Example 4 ***Drawing a Histogram of Continuous Data***

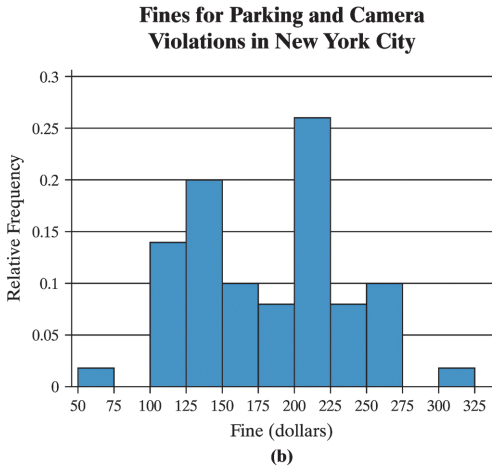
Construct a frequency and relative frequency histogram of the City of New York data discussed in Example 3.

Table 12
Total Fine Including Late Penalties (In Dollars)

211.09	209.29	150.56	271.20	210.20	125.76	190.72	187.23	229.18	125.00
105.00	256.69	262.99	256.70	322.61	243.80	236.91	260.00	162.46	65.00
227.59	224.35	223.99	155.40	193.42	151.46	127.37	140.06	216.19	210.25
105.23	125.21	208.40	208.39	207.79	124.85	124.67	148.02	206.29	183.16
147.70	206.01	123.41	204.49	134.66	167.65	121.25	131.76	120.53	143.52

Source: NYC Open Data





There is no one correct frequency distribution for a particular set of data. However, some frequency distributions better illustrate patterns within the data than others. So constructing frequency distributions is somewhat of an art form. Use the distribution that seems to provide the best overall summary of the data.

Answer the following after using the applet in Activity 1: Choosing Class Width.

8) What happens to the number of classes as the bin width increases?

The number of classes decreases as the bin width increases.

9) The number of classes in a frequency distribution is typically between what two numbers?

Typically between 5 and 20 classes.

10) Explain how to choose the lower class limit of the first class in a frequency distribution.

Choose the smallest observation in the data set or a convenient number slightly smaller than the smallest observation in the data set.

11) Once you decide on the number of classes, explain how to determine the class width.

Find the difference between the largest data value and the smallest data value. Divide that difference by the number of classes.

Objective 5: Draw Dot Plots

12) Explain how to draw a dot plot.

We draw a dot plot by placing each observation horizontally in increasing order and placing a dot above the observation each time it is observed.

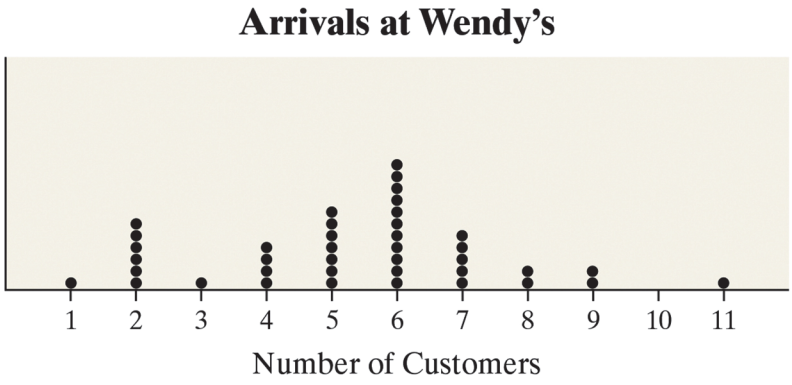
Example 5 ***Drawing a Dot Plot***

Draw a dot plot for the data from Table 8.

Table 8
Number of Arrivals at Wendy's

7	6	6	6	4	6	2	6
5	6	6	11	4	5	7	6
2	7	1	2	4	8	2	6
6	5	5	3	7	5	4	6
2	2	9	7	5	9	8	5

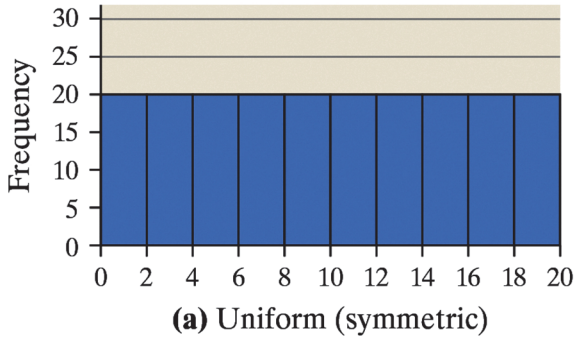
Answer:



Objective 6: Identify the Shape of a Distribution

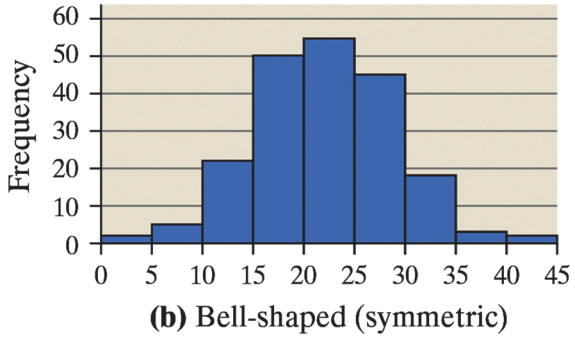
13) Draw an example of a uniform distribution.

Answer:



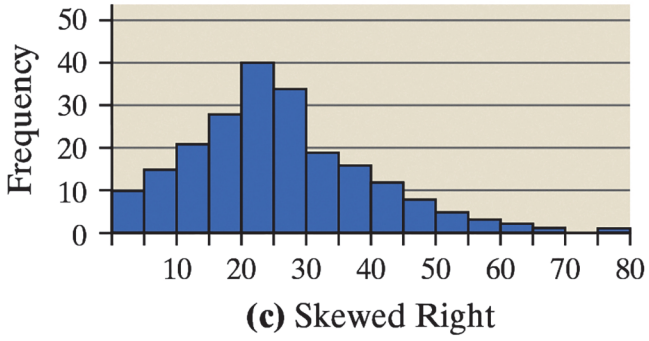
14) Draw an example of a bell-shaped distribution.

Answer:



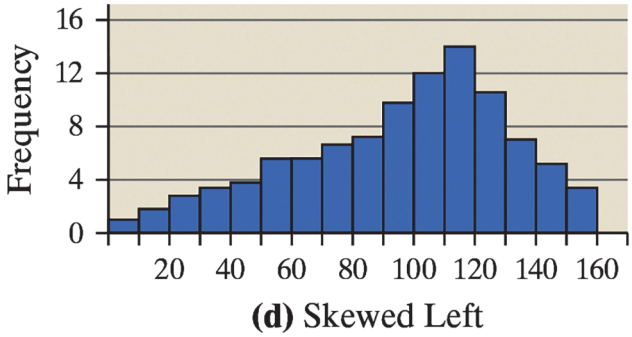
15) Draw an example of a distribution that is skewed right.

Answer:



16) Draw an example of a distribution that is skewed left.

Answer:

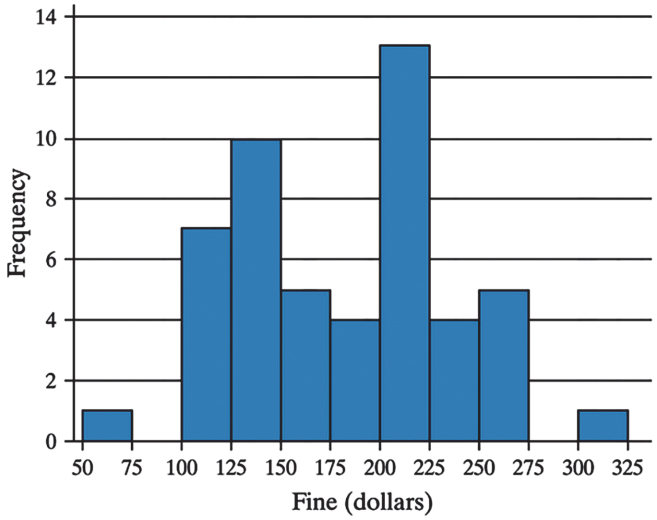


(d) Skewed Left

Example 6 ***Identifying the Shape of a Distribution***

Figure 10 displays the histogram obtained in Example 4 for the fine paid for parking and camera violations in New York City. Describe the shape of the distribution.

Figure 10
Fines for Parking and Camera Violations in New York City



The distribution is fairly symmetric.

Section 2.3
Additional Displays of Quantitative Data

Objectives

❶ Draw Stem-and-Leaf Plots

❷ Construct Frequency Polygons

❸ Create Cumulative Frequency and Relative Frequency Distributions

❹ Construct Frequency and Relative Frequency Ogives

❺ Draw Time-Series Graphs

Objective 1: Draw Stem-and-Leaf Plots

1) In a stem-and-leaf plot, how are the stem and leaf identified?

The digits to the left of the rightmost digit form the stem and the rightmost digit forms a leaf.

Example 1 ***Constructing a Stem-and-Leaf Plot***

The data in Table 14 represent the percentage of persons living in poverty, by state, in 2020. Draw a stem-and-leaf plot of the data.

Table 14
Percentage of People in Poverty by State, 2020

State	Percent	State	Percent	State	Percent
Alabama	13.9	Kentucky	13.8	North Dakota	9.4
Alaska	11.8	Louisiana	16.7	Ohio	12.5
Arizona	10.4	Maine	9.2	Oklahoma	13.2
Arkansas	14.1	Maryland	8.1	Oregon	8.8
California	10.6	Massachusetts	8.0	Pennsylvania	9.7
Colorado	9.4	Michigan	10.6	Rhode Island	8.8
Connecticut	9.8	Minnesota	7.0	South Carolina	14.2
Delaware	8.5	Mississippi	18.4	South Dakota	11.1
District of Columbia	14.6	Missouri	10.0	Tennessee	13.1
Florida	12.4	Montana	10.3	Texas	12.5
Georgia	12.7	Nebraska	8.5	Utah	7.4
Hawaii	9.7	Nevada	11.7	Vermont	8.6
Idaho	8.1	New Hampshire	4.9	Virginia	8.3
Illinois	8.6	New Jersey	7.2	Washington	7.6
Indiana	11.1	New Mexico	15.9	West Virginia	14.0
Iowa	9.3	New York	12.1	Wisconsin	8.2
Kansas	9.3	North Carolina	13.2	Wyoming	9.5

Data from United States Census Bureau

Answer:

Stem	Leaf
4	9
5	
6	
7	0246
8	01123556688
9	233445778
10	03466
11	1178
12	14557
13	12289
14	0126
15	9
16	7
17	
18	4

2) List the four steps for constructing a stem-and-leaf plot.

Step 1 The stem of a data value will consist of the digits to the left of the rightmost digit. The leaf of a data value will be the rightmost digit.

Step 2 Write the stems in a vertical column in increasing order. Draw a vertical line to the right of the stems.

Step 3 Write each leaf corresponding to the stems to the right of the vertical line.

Step 4 Within each stem, rearrange the leaves in ascending order, title the plot, and include a legend to indicate what the values represent.

3) List an advantage that a stem-and-leaf plot has over frequency distributions and histograms.

One advantage of the stem-and-leaf plot over frequency distributions and histograms is that the raw data can be retrieved from the stem-and-leaf plot.

4) Under what conditions do stem-and-leaf plots lose their usefulness?

Stem-and-leaf plots lose their usefulness when data sets are large or consist of a large range of values.

5) When constructing a stem-and-leaf plot, under what conditions is it advisable to use split stems?

We can use split stems when the data appear to be bunched.

Objective 2: Construct Frequency Polygons

6) Explain how to construct a frequency polygon.

A frequency polygon is a graph that uses points, connected by line segments, to represent the frequencies for the classes. It is constructed by plotting a point above each class midpoint (the sum of consecutive lower class limits divided by 2) on a horizontal axis at a height equal to the frequency of the class.

Example 2 ***Constructing a Frequency Polygon***

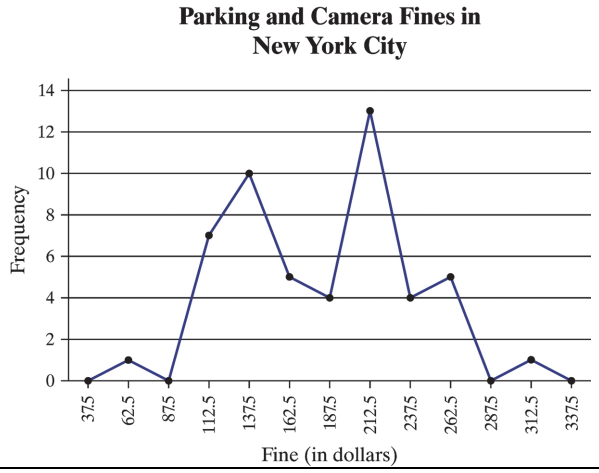
Draw a frequency polygon of the City of New York fine data listed in Table 16.

Table 16
Total Fine Including Late Penalties (In Dollars)

211.09	209.29	150.56	271.20	210.20	125.76	190.72	187.23	229.18	125.00
105.00	256.69	262.99	256.70	322.61	243.80	236.91	260.00	162.46	65.00
227.59	224.35	223.99	155.40	193.42	151.46	127.37	140.06	216.19	210.25
105.23	125.21	208.40	208.39	207.79	124.85	124.67	148.02	206.29	183.16
147.70	206.01	123.41	204.49	134.66	167.65	121.25	131.76	120.53	143.52

Source: NYC Open Data

Answer:



Objective 3: Create Cumulative Frequency and Relative Frequency Distributions

7) What does a cumulative frequency distribution display?

A cumulative frequency distribution displays the aggregate frequency of the category. In other words, it displays the total number of observations less than or equal to the upper class limit of the class.

8) What does a cumulative relative frequency distribution display?

A cumulative relative frequency distribution displays the proportion (or percentage) of observations less than or equal to the upper class limit of the class.

9) Explain how to find the cumulative frequency for the fifth class in a cumulative frequency distribution.

The cumulative frequency for the fifth class is the sum of the frequencies of classes 1, 2, 3, 4, and 5.

Example 3 ***Constructing a Cumulative and Cumulative Relative Frequency Distribution***

Obtain a cumulative frequency distribution and cumulative relative frequency distribution for the New York City fine data listed in Table 13.

Table 13

Class (Amount of Fine)	Frequency	Relative Frequency
50-74.99	1	0.02
75-99.99	0	0
100-124.99	7	0.14
125-149.99	10	0.20
150-174.99	5	0.10
175-199.99	4	0.08
200-224.99	13	0.26
225-249.99	4	0.08
250-274.99	5	0.10
275-299.99	0	0
300-324.99	1	0.02

Answer:

Class (Amount of Fine)	Cumulative Frequency	Cumulative Relative Frequency
50-74.99	1	0.02
75-99.99	1	0.02
100-124.99	8	0.16
125-149.99	18	0.36
150-174.99	23	0.46
175-199.99	27	0.54
200-224.99	40	0.80
225-249.99	44	0.88
250-274.99	49	0.98
275-299.99	49	0.98
300-324.99	50	1

Objective 4: Construct Frequency and Relative Frequency Ogives

10) What does an ogive represent?

An ogive is a graph that represents the cumulative frequency or cumulative relative frequency for the class.

11) Explain the difference between *x*-coordinates for a frequency polygon and a frequency ogive.

The x-coordinates for a frequency polygon are the class midpoints for each class, while the x-coordinates for an ogive are the upper class limits for each class.

12) Explain the difference between *y*-coordinates for a frequency polygon and a frequency ogive.

The y-coordinates for a frequency polygon are the class frequencies for each class, while the y-coordinates for an ogive are the cumulative frequencies for each class.