MOREAU
CATHOLIC HIGH SCHOOL
COMPASSIONATE COLLEGE PREP

AP Statistics Summer Review

Welcome to AP Statistics 2026!!

Mrs. Brand's e-mail address: cbrand@moreaucatholic.org

Dear students,

To prepare you for next year's AP Statistics class, the Math Department requires that you complete a summer review assignment. This review will refresh your skills and prepare you for our AP Statistics course. You are expected to spend at least 4 hours on this assignment, but it may require more if you are unfamiliar with the material. ALL problems must be solved, with all work neatly shown and stapled. You may print out the summer review and complete the problems on the document itself, or complete the problems on binder paper. The completed assignment is due on the first day of class. We will quickly review the assignment that day, so bring any questions to the first class. We will have an assessment on this material in the second week of class, so I encourage you to review your work the day before classes start to have it fresh in your mind if you completed the assignment early in the summer. All problems in this review should be material that you have already mastered. If anything does not look familiar, please ensure you get the proper help during the summer.

If you have any questions, you may reach out to the Math Department Chair, Mrs. Williams, at: mwilliams@moreaucatholic.org

See you in August!

- Complete Tasks 1-5 in this packet for review
- This is due on the first day of class, NO EXCEPTIONS!
- Place a star next to any problem you aren't sure how to solve, and be prepared to review it when we have our first class together.
- Knowing and understanding these concepts is KEY to your success in AP Statistics and preparing for the AP Exam.

Disclaimer: I do not recommend waiting until the last minute to complete the summer assignment. This should be completed over one to two weeks. Based on your summer assignment, there will be a test in the first week of school. ***You should do well on the first exam if you have completed the summer assignment. Not completing the summer assignment will result in a failing grade of F (0%) at the start of the semester, which means it will be difficult to earn higher than a B by the end of the semester.***

Remember - doing and turning in some is better than doing and turning in nothing.

Summer Dos

- Find time to enjoy your summer – AP classes require a lot of time commitment during the year, so the challenging work begins in the Fall.
- Work on the assignment in small chunks—30 to 45 minutes per day should get the assignment completed quickly.
- Contact me if you have any questions at all. Over the summer, it may take me a few days to get back to you, but your questions are welcome. Please do not hesitate to reach out! If you plan ahead, we can even set up a Zoom meeting and do one-on-one tutoring over the summer.
- Complete the whole assignment. This class is very rigorous, and you must commit to completing tasks on time. Not doing the homework will limit your outcomes in this course.
- Try! Being successful in an AP class means trying even when assignments are difficult. Do the best you can.
- Find a study buddy to talk with about assignments – this is helpful during the year, too.

Summer Do Nots

- Wait until the last minute to start this assignment – it will be overwhelming, and you most likely will not finish.
- Plagiarize answers—copying and pasting from another student or the internet does not show me what you know, and it will result in a 0 for the summer review and an academic integrity violation (for <u>both</u> people if copied from another student). Let's not start the year this way.

Task 1: Read "To the Student" on pg xxiii (15 points)

Take time to read through the textbook's structure and how it will prepare you for the AP Exam. Then, answer the following questions to assess your understanding.

********** Pages of the textbook are at the end of the packet.**********

1. What is the purpose of the textbook?
2. What four statistical practices make up the statistical problem-solving process? Explain each.
3. How many years of combined experience do the authors have with AP Statistics?
4. What are the five ways the textbook will help you get ready for the AP Statistics Exam?
    a. Briefly describe each in your own words.
5. What is the basic principle of learning statistics?
6. Take Notes on each of the following pages:
    i) Get The Most From This Program
    ii) Examples Are Your Friends
    iii) Exercises: Practice Makes Perfect!
    iv) Learn Statistics By Doing Statistics
    v) Achie√e Your Goals
    vi) Review and Practice For Quizzes And Tests
    vii) … And The AP Statistics Exam
    viii) What Comes Next? Think About Your Future.



Task 2: Send Mrs. Brand an "About Me" e-mail (10 points)

I want the opportunity to get to know you before school begins. To do this, send me an email this summer telling me about the following:

1. Yourself: what are your likes and dislikes, hobbies, personality, etc.?
2. Your family: do you have brothers, sisters, pets, etc.?
3. You, as a student: what are your study habits? What helps you learn best? What doesn't help you learn best? How many APs or Honors classes are you planning on taking this year?
4. How do you face challenges? What inspires you, motivates you, and brings you hope?
5. Your perception of this course: Why did you enroll in AP Statistics? What have you heard from other students, teachers, or counselors about this course? Be honest. I can handle it.
6. Your expectations for me and for this course: Is this your first AP course? If yes, what are your expectations? If not, what are your expectations? How do you reach out for help when you are struggling? This is a challenging course, and all students will struggle and have questions at one or more points throughout the year.

Note: Print the e-mail you send me and include it in your Summer Assignment tab.







A bit about me: Hi! I'm Mrs. Brand, and I am so excited to be your AP Statistics teacher this year. I've been teaching for 15 years, and this will be my 13th year specializing in AP Stats—so you are in experienced hands! A Bay Area native, I earned my B.S. in Mathematics with a concentration in teaching from CSU Monterey Bay. While I am a bit of an introvert who loves staying home with my family, I am deeply inspired by music, art, and nature, and I find so much joy in watching my students perform and get involved here at Moreau.

When I'm not looking at data, I'm likely traveling—I've had a blast exploring places like Japan, Italy, and Germany with students—or leaning into my competitive side with baseball, hockey, basketball, and my fantasy football league. I also love cooking, board games, a good pun, and have recently been getting into D&D! I come from a big family as the youngest of five, and I'm a proud aunt to nine nieces and nephews. However, my biggest highlight lately was welcoming my son, Harrison, last summer; he is truly the joy of my life. I'm looking forward to a fantastic year of learning and growth with all of you!

Task 3: Setup and Organize your AP Stats Binder (15 points)

You will be required to get a 1-inch or 2-inch 3-ring binder used **ONLY** for AP Stats. You will receive points for having your organized binder present on the first day of class. You will keep all notes, homework, and handouts in this binder. **Bonus points if you "decorate" the front of the binder in either a "Statsy" fashion or with your interests.**

1. Write your name on the binder's spine (First and Last) for easy identification.
2. Your binder will need 10 section dividers.
3. Each tab must have a label with the following:

- Summer Assignment / AP Exam Review
- Unit 1 Part 1: Exploring One-Variable Data
- Unit 1 Part 2: Collecting Data
- Unit 2 Part 1: Probability
- Unit 2 Part 2: Random Variables and Probability Distributions
- Unit 3 Part 1: Inference for One Proportion
- Unit 3 Part 2: Inference for Two Proportions and Two-Way Tables
- Unit 4 Part 1: Inference for One Mean
- Unit 4 Part 2: Inference for Two Means
- Unit 5: Regression Analysis







Task 4: Get to Know Your Calculator and Desmos (10 points)

You need to acquire and familiarize yourself with the TI-84+ graphing calculator (or higher) and the Desmos online calculator. We will often use our calculators to perform analysis and tedious calculations. The games below will help you familiarize yourself with the basic operations of both.

A. TI 84 Practice
   Download the worksheet at http://mathbits.com/MathBits/TISection/CachingPage.html to record answers and play the game. Print the worksheet and your completion certificate to include in your binder. **NOTE:** If you use a Ti-Nspire, email me for the link to the required worksheet. It will be your responsibility to learn how to use this calculator. All class demos will be with the Ti-84.

B. Getting Started on Desmos
   Go to https://www.desmos.com/calculator. Click the ➕ button to insert: expressions, tables, folders, inference, or notes. Complete the following tasks on Desmos and print a screenshot of your results.

   1. Creating Lists

   - **From a Table:** Use the table tool by clicking ➕ to create lists of values (Desmos will name them x1, y1, etc.)
   - **Manually Create a List:** Use square brackets [ ] (not curly braces) and do not use x or y. Example: A = [1, 2, 3, 4, 5], B = [2, 3, 3, 4, 5, 5, 5, 6]
   - **Desmos Functions for Lists:** Apply statistical functions directly to lists: mean(A), median(B), stdev(B), min(A), max(A), quartile(B,1) (gives Quartile 1 of set B)

   2. One-Variable Graphs

   - **Dotplot:** dotplot(B)
   - **Boxplot:** boxplot(B)
   - **Histogram:** histogram(B, 1) (bin width of 1)

Task 5: Let's get to the basics of Statistics (50 points)

Although we have until Spring Break to learn all of Units 1-5 and afterward to prepare for the AP Exam, AP Statistics is a dense curriculum. This means we need to start the first Unit now. We will review these Sections and Learning Targets together before you take your first Unit Test, but I trust you can learn the first two sections at home based on the science and math classes you have taken in high school. Check your answers using the back of the book. Use a different colored pen to correct & revise your work. Your grade for homework is based on your depth of practice, and not always getting the right answer. Incomplete or insufficient work will not receive full credit.

Reading Guide & Homework for Unit 1 Part 1: Exploring One-Variable Data

All responses to questions and reading guides must be HAND WRITTEN!

Introduction: pp. 2

1. Statistics is…
2. Describe the statistical problem-solving process.

Section 1A: pp. 3

1. What is a valid statistical investigative question?
2. What is the difference between a population and a sample?
3. How can you tell the difference between a parameter and a statistic?
4. An observational unit is…
5. A variable is...
6. The difference between a categorical variable and a quantitative variable is... (Give an example of each.)
7. Give an example of a number that is not quantitative.
8. Give an example of a discrete and a continuous quantitative variable. What distinguished one from the other?
9. Complete the Check Your Understanding (CYU) on pp. 9. (Check your answers)

☐ HW 1A: pp. 8 #'s 1 - 21 odd (Check your answers)

Section 1B: Displaying & Describing Categorical Data pp. 13

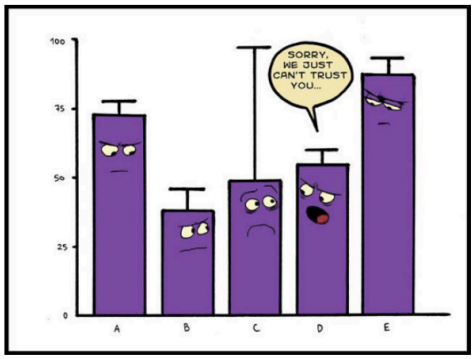
1. Define distribution.
2. What are the similarities and differences between frequency distributions and relative frequency distributions?
3. Why is it sometimes hard to make a frequency distribution for quantitative data?
4. A bar chart is…
5. How do you construct a bar chart? What is the purpose of a bar chart?
6. A pie chart is…
7. How do you construct a pie chart?
8. How many of the individuals must be included in a bar chart or a pie chart?
9. A side-by-side bar chart is…
10. What type of frequencies do we use to compare multiple groups? Why?
11. How can you organize the bars to compare multiple groups?
12. In the "Who buys iMacs?" example, what two important lessons did you learn about what makes a good or bad graph?
13. Complete the CYU on pp. 21. (Check your answers)

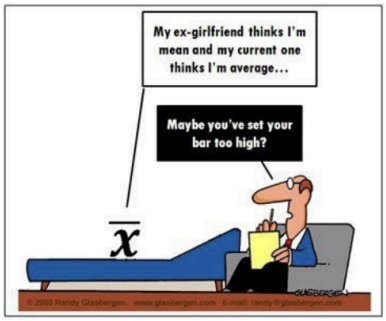
☐ HW 1B: pp. 22 # 1-25 odd (Check your answers)

ASSIGNMENTS ARE DUE ON THE FIRST DAY OF CLASS.

DO NOT drop them off in the office during the summer…they will get lost and you will need to redo the assignment. Late assingments lose 10% per day(including weekends) the assignment is late. Do not start the year trying to dig yourself out of a hole. On the first day of school, I will collect your binder as a 'packet', and our second-class meeting will return it to you.

AP Stats Summer Assignments Checklist ALL ITEMS DUE FIRST DAY OF CLASS	Points
Task 1: Read "To the Student" ☐ Answer the 6 questions	 15
Task 2: Send Mrs. Brand your "About Me" e-mail ☐ Print the e-mail you sent to cbrand@moreaucatholic.org	 10
Task 3: Organize your AP Stats 3-ring binder ☐ Write your name (First and Last) on the spine of the binder ☐ Make Labels for your 10 divider tabs ☐ BONUS: Decorate your binder	 5 10 (+3)
Task 4: Get to Know Your Calculator & Desmos ☐ Download and Print the Worksheet ☐ Complete the activity ☐ Print your Certificate to Verify Completion ☐ Complete the Desmos Activity ☐ Print a Screenshot of your Desmos window	 3 2 3 2
Task 5: Unit 1 Section 1A and 1B ☐ Complete the Reading Guide: Intro, 1A, and 1B ☐ Complete and Correct HW 1A and HW 1B ☐ *Star* any problems you didn't understand for review.	 30 20
Double-Check that you have all of the above items in your binder, ready to turn in. All tasks should be included in the "Summer Assignment / AP Review " tab.	Worth 100 points !!







for the AP® Course

The Practice of Statistics

EIGHTH EDITION

Daren Starnes
Statistics Education Consultant

Josh Tabor
The Potter's School

BFW Publishers
New York Boston

To the Student

Statistics is essential for tracking the economy, analyzing the effectiveness of new medical treatments, measuring changes in climate, gauging people's views about important social issues, and performing research in many fields of study. Companies, governments, and other organizations regularly use statistics to make critical decisions that affect us all. In short, statistics matters!

The Practice of Statistics, Eighth Edition (*TPS 8e*), is full of real-world data that illustrate the usefulness of statistics. Each set of data is presented along with some brief background to help you understand where the data come from. We deliberately chose contexts and data sets in the examples and exercises to pique your interest.

Statistical Thinking and You

The purpose of *TPS 8e* is to build your understanding of the key concepts and methods used in solving statistics problems. We want to help you develop the statistical thinking skills you will need to make informed, data-based decisions in your future studies, career, and daily life. Speaking of careers, make sure to check out the new *See Yourself in Stats* feature at the end of each part and unit to see how statistics is used in a variety of fields.

AP® Statistics focuses on four statistical practices that make up the steps in the statistical problem-solving process:

- *Formulate questions:* Determine an investigative question for a statistical study.
- *Collect data:* Identify and justify methods for collecting data and conducting statistical inference.
- *Analyze data*: Construct representations of data and calculate numerical statistical outputs.
- *Interpret results:* Interpret results and justify conclusions and methods.*

See the table near the inside front cover for a full list of the skills associated with each of these four statistical practices. We also hope that you will tackle some of the *Projects* placed throughout the book for a chance to put the statistical problem-solving process into practice.

The Practice of Statistics and AP® Statistics

The Practice of Statistics was the first book written specifically for the Advanced Placement (AP®) Statistics course. *TPS 8e* is organized to follow the 2026 AP® Statistics *Course and Exam Description* (CED) very closely. Every learning objective and essential knowledge statement in the CED course framework is covered thoroughly in the text. In addition, the statistical practices and skills listed in the CED are emphasized throughout the book.

*AP® *Statistics Course and Exam Description* (effective Fall 2026), the College Board.

Most importantly, *TPS 8e* is designed to prepare you for the AP® Statistics exam. The author team has more than 50 years of combined experience teaching AP® Statistics and more than 45 years of combined experience grading the AP® Statistics exam! Both of us have served on the AP® Reading leadership team for more than 15 years, helping finalize scoring guidelines for free response questions. Including our Content Advisory Board and Program Resources Team (page xiii), we bring extensive knowledge of how the AP® Statistics exam is developed and scored.

The Eighth Edition of *TPS* will help you get ready for the AP® Statistics exam throughout the course by:

- **Using terms, notation, formulas, and tables consistent with those found in the CED and on the AP® Statistics exam.** Key terms are shown in bold in the text, defined in the Glossary/Glosario, and cross-referenced in the Index. The formulas and notation used in *TPS 8e* are exactly the same as those on the AP® Statistics exam formula sheet, which can be found in the back of the book.
- **Following accepted conventions from AP® Statistics exam scoring guidelines in model solutions.** We carefully considered recent scoring guidelines when writing the solutions that appear in *TPS 8e*. For example, the two-step process to answer questions about normal distributions in Unit 2 and the four-step State–Plan–Do–Conclude process that we use to perform inference procedures in Units 3 and 4 were inspired by AP® Statistics exam scoring guidelines.
- **Including specific AP® Statistics exam advice at point of use.**
  - We give numerous **AP® Exam Tips** as "on-the-spot" reminders of common mistakes and ways to avoid them. Specific **Formula Sheet** tips appear whenever we introduce a formula, so you know whether or not it is included on the AP® Statistics exam formula sheet. At the beginning and end of each section, we provide **AP® Classroom** tips that direct you to the relevant topics in AP® Classroom so you can use the available resources to preview and review the content of the section.
  - The **Strategies for Success on the AP® Statistics Exam** resource on Achieve and available from your teacher, summarizes the structure and scoring of the exam, as well as our top tips for maximizing your exam score.
- **Expanding support for a variety of handheld and online graphing calculators.** In addition to integrating the TI-83/84 graphing calculator in the Tech Corners throughout the text, we provide videos with step-by-step instructions for Desmos, NumWorks, TI-Nspire, and the latest TI-84 calculator, all of which can be used on the AP® Statistics exam.
- **Providing about 1500 exercises throughout the book that are designed to prepare you for the types of questions found on the AP® Statistics exam.**
  - **Section Exercises** include paired odd- and even-numbered problems that test the same concept or skill from that section. Every section includes *Multiple Choice* exercises like those found on the AP® Statistics exam.

New *Connecting the Statistical Practices* exercises are multi-focus, free response questions that require you to link skills associated with different steps in the statistical problem-solving process. *For Investigation* exercises ask you to apply the skills from the section in a new context or nonroutine way.

- At the end of each part of a unit, your teacher can provide you with access to a **FRAPPY!**—Free Response AP® Problem, Yay! Each FRAPPY! gives you the chance to answer an AP®-style free response question based on the content and skills you should have learned by that point in the text. After you finish, your teacher can provide two student solutions for you to view and critique. Then you can score your own response using scoring guidelines provided by your teacher.
- **Review Exercises** at the end of each part of a unit consist of free response questions aligned to specific learning targets from that part of the unit.
- The **AP® Statistics Practice Test** at the end of each part of a unit will help you prepare for in-class exams. These tests include both multiple-choice and free response questions assessing the skills required for the AP® Statistics exam.
- Your teacher can provide you with access to **Cumulative AP® Practice Tests.** These tests, which are available after each unit, provide challenging, cumulative multiple-choice and free response questions like those you might find on a midterm, a final, or the AP® Statistics exam.

Learning with *The Practice of Statistics, Eighth Edition*

TPS 8e is designed to be easy to read and easy to use. This book is written by experienced high school AP® Statistics teachers, for high school students. We aimed to provide clear, concise explanations and to use a conversational approach that encourages you to read the book. We also tried to enhance both the visual appeal and the book's clear organization in the layout of the pages.

Be sure to take advantage of all that *TPS 8e* has to offer. You can gain a lot of knowledge by reading the text, but you learn statistics best by *doing* statistics. To deepen your understanding of key concepts and skills, do the Activities, answer the Check Your Understanding questions along the way, work lots of exercises, and watch the Example and Exercise videos. Use the answers at the back of the book to help identify strengths and weaknesses in your solutions.

The main ideas of statistics took a long time to discover—so it isn't surprising that they take some time to master. The best way to learn them is to be persistent.

Turn the page for a tour of the text. See how the book can help you realize success in the course and on the AP® Statistics exam.

Get the Most from This Program

Your AP® Statistics adventure begins here! This program was created with YOU in mind. It is packed with features and resources to help you learn effectively and to do well on the AP® Statistics Exam.

Written to support your AP® Statistics journey

Your text matches the **AP® STATISTICS COURSE** both in organization and in scope. Units, Parts, and Sections focus on the content and skills that are central to the course. Daily practice is built in to support you at every step.

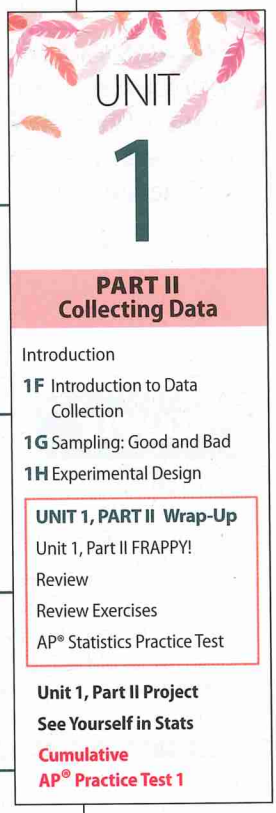
UNIT 1 Exploring One-Variable Data and Collecting Data
- PART I Exploring One-Variable Data
- PART II Collecting Data
- Achieve Cumulative AP® Practice Test 1

UNIT 2 Probability, Random Variables, and Probability Distributions
- PART I Probability
- PART II Random Variables and Probability Distributions
- Achieve Cumulative AP® Practice Test 2

UNIT 3 Inference for Categorical Data: Proportions
- PART I Inference for One Proportion
- PART II Inference for Two Proportions and Two-Way Tables
- Achieve Cumulative AP® Practice Test 3

UNIT 4 Inference for Quantitative Data: Means
- PART I Inference for One Mean
- PART II Inference for Two Means
- Achieve Cumulative AP® Practice Test 4

UNIT 5 Regression Analysis
- Achieve Cumulative AP® Practice Test 5



Pay attention to the Integrated AP® EXAM TIPS

AP® EXAM TIP: AP® Classroom
Preview the content of this section with the resources in AP® Classroom for Topics 1.7 and 1.8.

AP® EXAM TIP
Formula sheet
The formula for the standard deviation σ_X of a discrete random variable is included on the AP® Statistics exam formula sheet.

AP® EXAM TIP
You may be asked to determine whether a quantitative data set has any outliers. Be prepared to state and use the $1.5 \times IQR$ rule and the $2 \times$ SD rule for identifying outliers.

AP® EXAM TIP
Formula sheet
The formula for the $1.5 \times IQR$ rule is NOT included on the AP® Statistics exam formula sheet.

Pay attention to the Integrated **AP® EXAM TIPS** that support you in three ways:

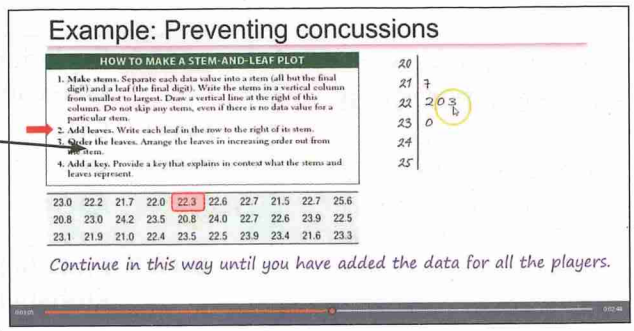
- Guiding you to the relevant AP® Classroom resources to preview and/or review the content presented in the corresponding sections of the text
- Providing advice from experts on how to be successful on the AP® Statistics exam
- Offering insights so you know which formulas ***are*** and ***are NOT*** included on the AP® Statistics exam formula sheet—which lets you know where to focus your study

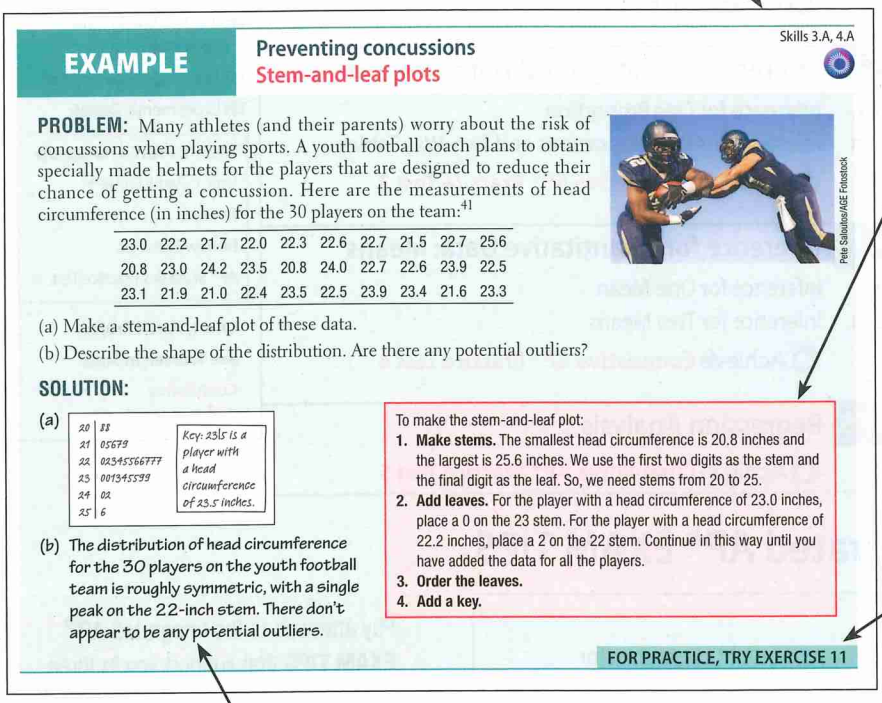
Examples Are Your Friends

Study the examples for detailed instruction, with text and video guidance, for every concept and skill in the course.

Examples with paired exercises are the heart of program.

EXAMPLES Each example title is followed by the content and skills being taught, and an icon links to a paired **Example Video**. Each video was prepared by an experienced AP® Statistics teacher to help you study, review, and get extra support when needed.





THE VOICE OF THE TEACHER Pay special attention to the carefully placed "teacher talk" boxes that guide you through the solution. These comments offer lots of good advice—as if your teacher is working directly with you to solve a problem.

PRACTICE! Take the time to work the paired problem in the Section Exercises. Practicing a similar problem after studying the worked example helps you know whether or not you've "got it."

Note that the **SOLUTION** is presented in a special font to model the style, steps, and language you should use to earn full credit on the AP® Statistics exam.

Exercises: Practice Makes Perfect!

Work the exercises assigned by your teacher. Watch the videos when you need extra help.

Time spent working problems is key to success in this course.

THE *TPS* "BACKWARDS/FORWARDS" NAVIGATION SYSTEM

Do you read the text first? Or do you start with the assigned exercises? Either way, *TPS* is designed to support you.

Look for the Icons that appear next to selected exercises. They will guide you:

- Back to the **example** that models the exercise so that you can review the worked solution
- To a short **video** that provides step-by-step instruction for solving the exercise

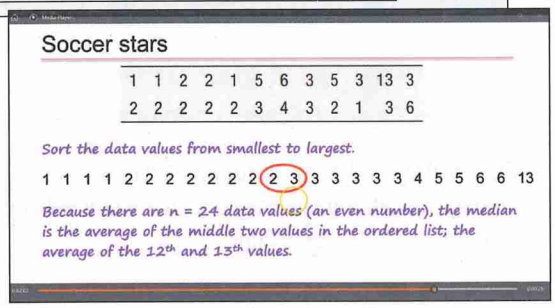
SECTION EXERCISES ORGANIZED BY TOPIC

Use the handy headings to guide you to topics on which you need practice.

Measuring Center: The Median

1. pg 50 Soccer stars How good was the 2019 U.S. women's soccer team? With players like Carli Lloyd, Alex Morgan, and Megan Rapinoe, the team put on an impressive showing en route to winning the 2019 Women's World Cup. Here are data on the number of goals scored by the team in 24 games played in the 2019 season:[69]

1	1	2	2	1	5	6	3	5	3	13	3
2	2	2	2	2	3	4	3	2	1	3	6



Connecting the Statistical Practices *Multi-focus exercises that connect two or more statistical practices.*

52. Something fishy As part of a study on salmon health, researchers measured the pH of a random sample of 25 salmon fillets at a fish processing plant.[86] Here are the data:

6.34	6.39	6.53	6.36	6.39	6.25	6.45	6.38	6.33
6.26	6.24							
6.36	6.44							

(a) Identify the p
cal study.

(b) Make a stem-a

Multiple Choice *Select the best answer for each question.*

53. If a distribution of quantitative data is strongly skewed to the right with no
relationships is mos

(A) Mean < media

(B) Mean ≈ media

For Investigation *Apply the skills from the section in a new context or nonroutine way.*

60. Trimmed mean Another measure of center for a quantitative data set is the *trimmed mean*. To calculate the trimmed mean, order the data set from lowest to highest, remove the same number of data values from each end, and calculate the mean of the remaining values.

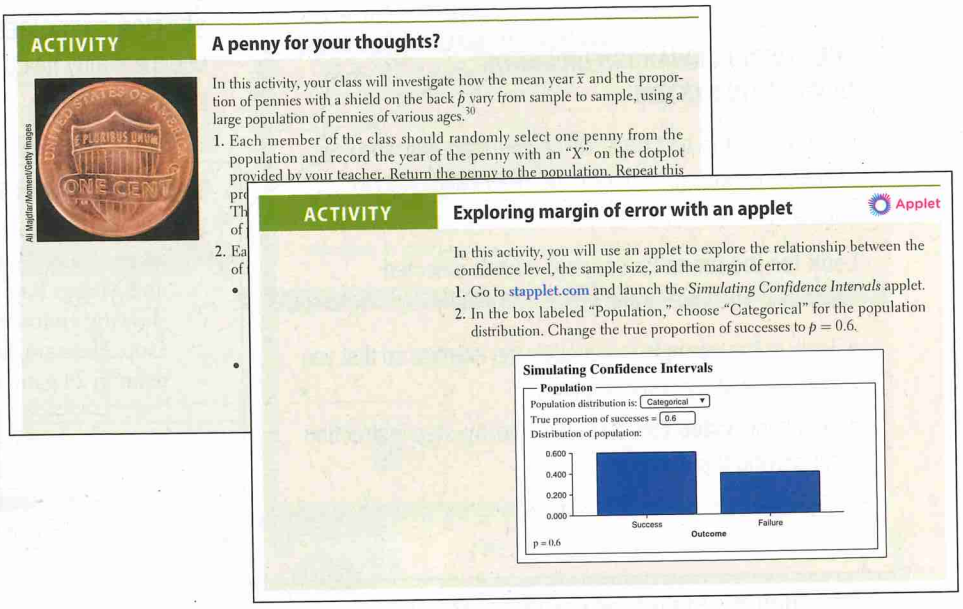
Additional problem types in the Section Exercises let you practice solving many different types of problems, including:

- **Connecting the Statistical Practices,** which are new multi-focus exercises that connect two or more statistical practices
- **AP®-style Multiple Choice** questions with four answer choices to give you regular practice for the AP® Statistics exam
- **For Investigation** exercises that assess multiple skills and concepts, and let you apply your knowledge in new contexts and nonroutine ways

Learn Statistics by Doing Statistics

Experience statistical principles with activities and technology tools.

ACTIVITIES Every Unit includes a hands-on activity in the first few pages to introduce the content, while other activities appear later in the Unit. The Activities Appendix identifies all of the activities in the book and notes where to find them. Many of the 29 activities use dynamic **applets** that help you experience the process of collecting data and drawing conclusions from those data. All of the applets are available to you whether you do the activity in class or on your own.



ACTIVITY A penny for your thoughts?

All Majdar/Moment/Getty Images

In this activity, your class will investigate how the mean year $\bar{x}$ and the proportion of pennies with a shield on the back $\hat{p}$ vary from sample to sample, using a large population of pennies of various ages.[30]

1. Each member of the class should randomly select one penny from the population and record the year of the penny with an "X" on the dotplot provided by your teacher. Return the penny to the population. Repeat this pr
Th
of
2. Ea
of
•
•

ACTIVITY Exploring margin of error with an applet

Applet

In this activity, you will use an applet to explore the relationship between the confidence level, the sample size, and the margin of error.

1. Go to stapplet.com and launch the *Simulating Confidence Intervals* applet.
2. In the box labeled "Population," choose "Categorical" for the population distribution. Change the true proportion of successes to $p = 0.6$.

TECHNOLOGY Use technology as a tool for discovery and analysis. The 26 **Tech Corners** are placed strategically throughout the book at the optimal point of use. They give step-by-step instructions for using the TI-83/84 calculator.

Supporting **Tech Corner Videos** are available to walk you through the keystrokes needed to perform each analysis using the TI-83/84 (including the latest TI-84 calculator), as well as **Desmos, Numworks, Nspire,** and other calculators. Find these videos on Achieve, or your teacher can share them with you.

2. Tech Corner CALCULATING SUMMARY STATISTICS

Instructions for other technology, including Desmos, NumWorks, and the latest TI-84 calculator, can be found on Achieve or your teacher can share this content with you.

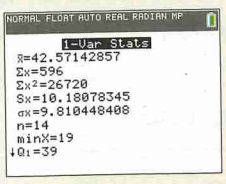
You can use technology to calculate measures of center and variability for a distribution of quantitative data. We'll illustrate using the data on the percent air in a sample of 14 bags of chips from previous examples.

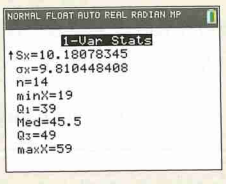
1. Type the values into list L1.
2. Press STAT, arrow over to the CALC menu, and choose 1-Var Stats.

   OS 2.55 or later: In the dialog box, press 2nd 1 (L1) and ENTER to specify L1 as the List. Leave FreqList blank. Arrow down to Calculate and press ENTER.

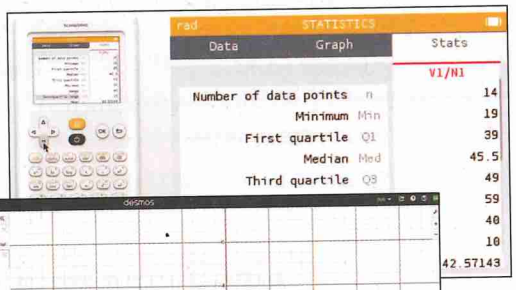
   Older OS: Press 2nd 1 (L1) and ENTER.

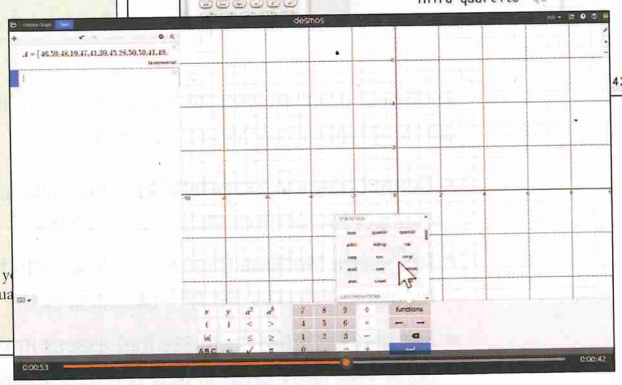
Press the down arrow to see the rest of the one-variable statistics.





Note: The TI-83/84 does not give the range or interquartile range (IQR) directly when y
one-variable statistics. However, the calculator provides the minimum, maximum, and the qua
can calculate range = max − min and $IQR = Q_3 - Q_1$.





Achieve Your Goals

BFW Publishers

Access everything you need for success through the ACHIEVE platform.

Read the text, watch videos, and receive guided feedback on your homework.

EASY ACCESS TO RESOURCES WITH THE EBOOK
Interactive and mobile-ready, the ebook allows you to read and reference the text both online and offline. All offline highlights and notes sync when you connect to the Internet. What's more, clicking on gives you instant access to each video while you are online.

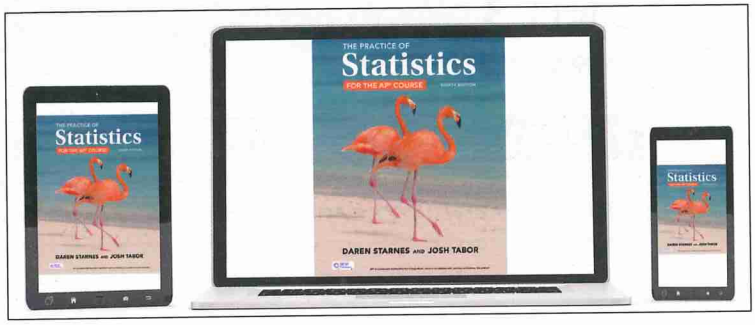
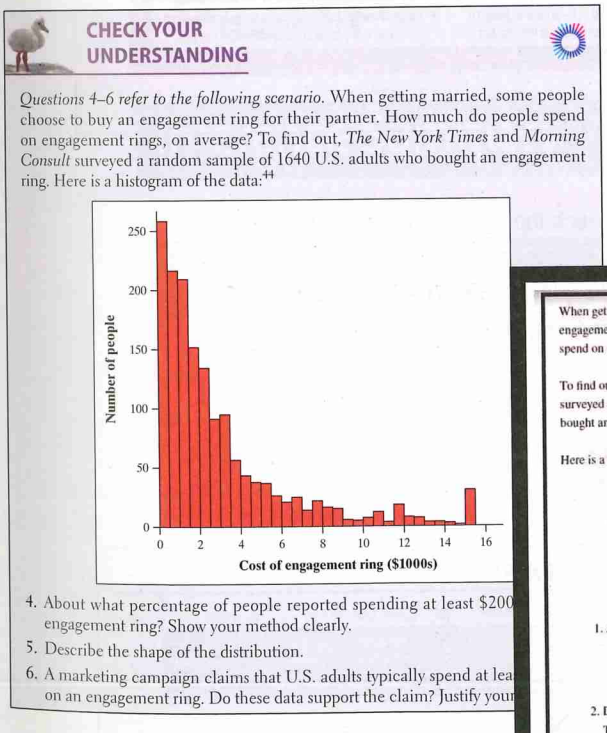


Photo: Judy M. Starnes

The **online homework** system helps you learn with targeted feedback based on common misconceptions, so you learn even if you get the answer wrong! All of the feedback was developed by experienced AP® Statistics teachers who write questions for the AP® Statistics exam.

Thousands of practice problems are available, including the **Check Your Understanding** problems, selected **Section Exercises,** all **Unit/Part Review exercises,** each **Unit/Part AP® Statistics Practice Test**, and all **Cumulative AP® Practice Tests** from the text.

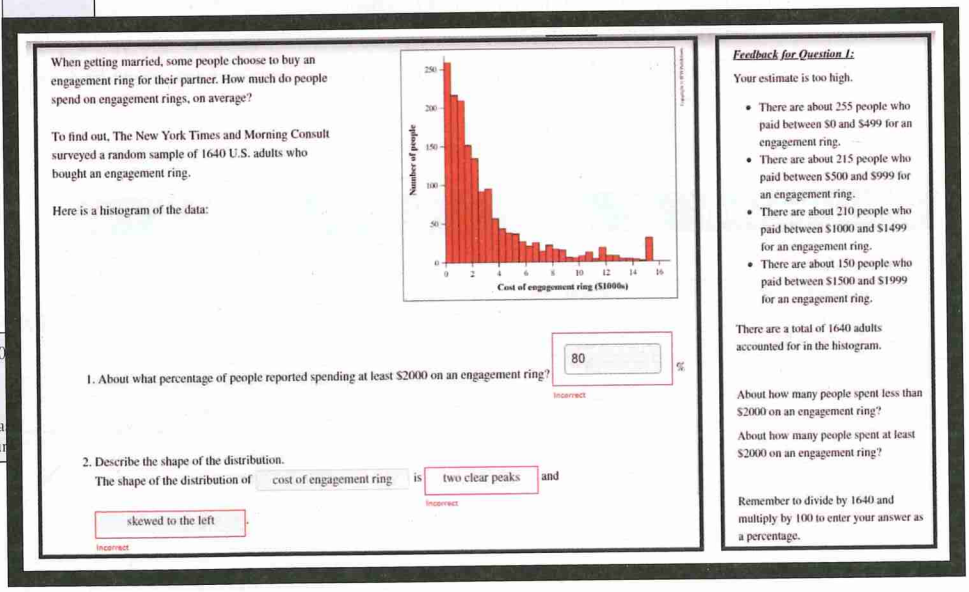


CHECK YOUR UNDERSTANDING

Questions 4–6 refer to the following scenario. When getting married, some people choose to buy an engagement ring for their partner. How much do people spend on engagement rings, on average? To find out, *The New York Times* and *Morning Consult* surveyed a random sample of 1640 U.S. adults who bought an engagement ring. Here is a histogram of the data:[44]

4. About what percentage of people reported spending at least $200 engagement ring? Show your method clearly.
5. Describe the shape of the distribution.
6. A marketing campaign claims that U.S. adults typically spend at lea on an engagement ring. Do these data support the claim? Justify your

Sprinkled throughout the text after key concepts and skills are presented, **Check Your Understanding** questions do just that. Working them in Achieve or on paper will help you stay on track.



Review and Practice for Quizzes and Tests

Take advantage of the review tools included in every Unit and Part.

UNIT 3 Wrap-Up
PART I

FRAPPY! Free Response AP® Problem, Yay!

Find the Unit 3, Part I FRAPPY! on Achieve or your teacher can share this content with you.

DO THE FRAPPY! Learn how to answer FRQs successfully by working the FRAPPY!—the Free Response AP® Problem, Yay!—that begins each Unit/Part Wrap-up. You can access these problems on Achieve or get them from your teacher.

SNAPSHOT REVIEW Each Unit or Part Review provides a short summary of each section, to be sure you understand the key concepts presented.

SUMMARY TABLES in Units 3 and 4 review the important details of each inference procedure, including conditions and formulas.

UNIT 3, PART I REVIEW

SECTION 3A The Sampling Distribution of a Sample Proportion

In this section, you learned about the shape, center, and variability of the sampling distribution of a sample proportion $\hat{p}$.

Confidence intervals are formed by adding and subtracting the **margin of error** from the point estimate (value of the statistic):

CI = point estimate ± margin of error

Inference for a Population Proportion

	Confidence interval for p	Hypothesis test for p
Name (TI-83/84)	One-sample z interval for a population proportion (1-PropZInt)	One-sample z test for a population proportion (1-PropZTest)
Null Hypothesis	*Not applicable*	$H_0: p = p_0$
Conditions	• **Random:** The data come from a random sample from the population of interest. ○ **10%:** When sampling without replacement, $n \leq 10\%N$. • **Large Counts:** Both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10. That is, the number of successes and the number of failures in the sample are both at least 10.	• **Random:** The data come from a random sample from the population of interest. ○ **10%:** When sampling without replacement, $n \leq 10\%N$. • **Large Counts:** Both np_0 and $n(1-p_0)$ are at least 10, where p_0 is the proportion specified by the null hypothesis.
Formula	$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Use the **WHAT DID YOU LEARN?** table to verify your mastery of each topic and to find help when needed. All of the individual **LEARNING TARGETS** are listed with references to the section in which they are introduced, as well as related examples and 5–12 relevant **UNIT/PART REVIEW EXERCISES**. And, of course, watch the Unit Review Exercise videos to get tips on solving these multifaceted problems.

What Did You Learn?

Learning Target	Section	Related Example on Page(s)	Relevant Unit 3, Part I Review Exercise(s)
Calculate and interpret the mean and standard deviation of the sampling distribution of a sample proportion $\hat{p}$.	3A		
Determine whether the sampling distribution of $\hat{p}$ is approximately normal.	3A		

UNIT 3, PART I REVIEW EXERCISES

These exercises are designed to help you review the important concepts and skills of the unit.

R1 Do you jog? (3A) A Gallup poll asked a random sample of 1540 adults, "Do you regularly jog?" Suppose that the population proportion of all adults who jog is $p = 0.15$.

(a) Calculate and interpret the mean of the sampling distribution of $\hat{p}$.

(b) Verify that the 10% condition is met. Then calculate and interpret the standard deviation of the sampling distribution of $\hat{p}$.

(c) Is the sampling distribution of $\hat{p}$ approximately normal? Justify your answer.

R2 Sports fans (3B) Are you a sports fan? That's the question the Gallup polling organization asked a random sample of 1527 U.S. adults.[40] Gallup reported that a 95% confidence interval for the proportion of all U.S. adults who would say they are sports fans is 0.565 to 0.615.

(a) Calculate the point estimate and margin of error used to create this interval.

(b) Interpret the confidence interval.

(c) Interpret the confidence level.

(d) Based on the interval, is there convincing evidence that a majority of U.S. adults are sports fans? Explain your answer.

. . . and for the AP® Statistics Exam

Consistent daily practice will help you avoid cramming before the AP® exam.

Each Unit/Part concludes with an **AP® STATISTICS PRACTICE TEST**, which typically consists of 10–12 multiple-choice questions and 3 free response questions.

UNIT 3, PART I AP® STATISTICS PRACTICE TEST

Section I: Multiple Choice *Select the best answer for each question.*

T1 Suppose we select an SRS of size $n = 100$ from a large population having proportion p of successes. Let $\hat{p}$ be the proportion of successes in the sample. For which value of p would it be safe to use a normal distribution to approximate the sampling distribution of $\hat{p}$?

(A) 0.01 (C) 0.85
(B) 0.09 (D) 0.975

T2 The student newspaper at a large university plans to ask an SRS of 250 undergraduates, "Do you favor eliminating the carnival from the term-end celebration?" Suppose that 55% of all undergraduates favor eliminating the carnival. How far do you expect the sample proportion who favor the proposal to typically vary from the population proportion?

(A) 0.001 (C) 0.035
(B) 0.031 (D) 0.062

Section II: Free Response *Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.*

T13 The U.S. Forest Service is considering additional restrictions on the number of vehicles allowed to enter Yellowstone National Park. To assess public reaction, the service asks a random sample of 150 visitors if they favor the proposal. Of these, 89 say "Yes."

(a) Construct and interpret a 99% confidence interval for the proportion of all visitors to Yellowstone National Park who favor the restrictions.

(b) Interpret the confidence level.

T14 According to the Pew Research Center, "More than half of Americans live within an hour of extended family." In this study, "extended family" was defined as children, parents, grandparents, grandchildren,

(a) Do these data provide convincing evidence for the Pew Research Center's claim?

(b) Which type of error, Type I or Type II, could you have made in part (a)? Explain your answer.

T15 A mattress manufacturer has designed a new production process that is intended to reduce p = the proportion of all mattresses made with the new process that have defects in the stitching. Before the change, 15% of the mattresses had this type of defect. After the new production process is implemented, 100 mattresses will be randomly selected and inspected for stitching defects. Then the manufacturer will test $H_0: p = 0.15$ versus $H_a: p < 0.15$.

Cumulative AP® Practice Test 5

Find Cumulative AP® Practice Test 5 on Achieve or your teacher can share this content with you.

Five **CUMULATIVE AP® PRACTICE TESTS** simulate the real exam. Placed at the end of each Unit, these tests expand in coverage of content and skills as you work through the book. The last test models a full AP® Statistics exam. These valuable practice tests may be accessed on Achieve or obtained from your teacher.

Inference Summary

How to Organize an Inference Problem: The Four-Step Process

	Confidence intervals	Hypothesis tests
STATE:	State the type of study and the variable(s) in the study (including classification). Then, state the confidence level and the parameter you want to estimate.	State the type of study and the variable(s) in the study (including classification). Then, state the hypotheses, parameter(s), and significance level.
PLAN:	Identify the appropriate inference method and check the conditions.	Identify the appropriate inference method and check the conditions.
DO:	If the conditions are met, perform calculations.	If the conditions are met, perform calculations. • Calculate the test statistic. • Find the P-value.
CONCLUDE:	Interpret your interval in the context of the problem.	Make a conclusion about the hypotheses in the context of the problem.

$$\text{confidence interval} = \text{statistic} \pm (\text{critical value})(\text{standard error of the statistic})$$

$$\text{standardized test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$$

Inference about	Number of samples/ groups	Interval or test (Section)	Name of procedure (TI-83/84 name) Formula	Conditions
	1	Interval (3B)	One-sample z interval for a population proportion (1-PropZInt) $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	• **Random:** The data come from a random sample from the population of interest. ○ **10%:** When sampling without replacement, $n \leq 10\%N$. • **Large Counts:** *Interval:* Both $n\hat{p}$ and $n(1-\hat{p}) \geq 10$. *Test:* Both np_0 and $n(1-p_0) \geq 10$, where p_0 is the value of p specified by the null hypothesis.
		Test (3C)	One-sample z test for a population proportion (1-PropZTest) $z = \frac{\hat{p} - p_0}{}$	

Consult the **INFERENCE SUMMARY** in the back of the book for a helpful way to organize the key inference procedures, formulas, and conditions from Units 3 and 4. The summary is organized by what the inference is about, the number of samples/groups, and the type of inference.

What comes next? . . . Think about your future.

Unit 1, Part I Project

American Community Survey

Each month, the U.S. Census Bureau selects a random sample of U.S. households to participate in the American Community Survey (ACS). The chosen households are notified by mail and invited to complete the survey online. The Census Bureau follows up on any uncompleted surveys by phone or in person. Data from the ACS are used to determine how the federal government allocates more than $675 billion in funding for local communities.[125]

The file TPS8U1P1project_data, which can be accessed on Achieve or provided by your teacher, contains data for 3000 randomly selected households from the ACS in a recent month.[126] Download the file to a computer for further analysis using the application specified by your teacher.

Each row in the spreadsheet describes a household. A serial number that identifies the household is in the first column. The other columns contain values of several variables. See the TPS8U1P1project_code sheet on Achieve or provided by your teacher, for details on how each variable is recorded. Note that all the categorical variables have been coded to have numerical values in the spreadsheet.

Use the files provided to answer the following questions:

1. How many variables are recorded? Classify each variable as categorical or quantitative.
2. Examine the distribution of household language (HHL). Make a bar chart to display the data. Then calculate numerical summaries (counts, percentages, or proportions). Describe what you see.
3. Does the distribution of household language differ among the four U.S. regions (REGION)? Make an appropriate graph to display the data. Then calculate numerical summaries. Write a few sentences comparing the distributions.
4. Analyze the distribution of household income (HINCP) using appropriate graphs and numerical summaries.
5. Does the distribution of household income differ among the four U.S. regions? Make an appropriate graph to display the data. Then calculate numerical summaries. Write a few sentences comparing the distributions.

Extension: Use what you have learned in Unit 1, Part I to do a complete analysis of another variable in the data set that interests you.

See Yourself in Stats

With a solid grasp of statistics, you could become . . . a market research analyst.

Market research analysts play a central and growing role in the modern workplace. They work in almost every area, including government, business, and scientific fields. These analysts help companies interpret data using statistical techniques and software to draw conclusions about the products and services people want and the price consumers will pay for them. They often rely on demographic data to identify and refine plans for new product development and to forecast future trends in a given market. To communicate trends and opportunities clearly, market research analysts rely on well-constructed graphs to provide compelling visual displays of the data.

To learn more about this fascinating field, check out **See Yourself in Stats** by clicking the link in your ebook, visiting the See Yourself in Stats spreadsheet on Achieve, or obtaining it from your teacher.

PROJECTS GIVE REAL-WORLD EXPERIENCE Explore the kind of robust research projects that you are likely to experience in future STEM, Social Studies, Business/Finance, or Psych courses in college or in the workplace by diving into one of the Unit/Part Projects. These projects are featured at the end of seven of the nine Parts in the book.

WHAT CAREERS MAY AWAIT? Learn more about various careers that require knowledge of statistics by reading the **SEE YOURSELF IN STATS** feature that concludes each Part/Unit. Then explore the spreadsheet with links to sites where you can learn more about possible career opportunities.

UNIT

1

Exploring One-Variable Data and Collecting Data

PART I Exploring One-Variable Data

Introduction

1A Statistics: Learning from Data

1B Displaying and Describing Categorical Data

1C Displaying and Describing Quantitative Data with Graphs

1D Summarizing Quantitative Data with Numbers

1E Comparing Distributions of Quantitative Data

UNIT 1, PART I Wrap-Up

Unit 1, Part I FRAPPY!

Review

Review Exercises

AP® Statistics Practice Test

Unit 1, Part I Project

See Yourself in Stats



Manuela Schewe-Behnisch/EyeEm/GettyImages

Introduction

We live in a world of *data*. Every day, the media report poll results, outcomes of medical studies, and analyses of data on everything from gasoline prices to elections to weather to the latest technology. These data are trying to tell us a story. To understand what the data are saying, we use **statistics**.

DEFINITION Statistics

Statistics is the science and art of collecting, analyzing, and drawing conclusions from data.

A solid understanding of statistics will help you make informed decisions based on data in your daily life. The following activity illustrates one of the many uses of statistics in the real world.

ACTIVITY Smelling Parkinson's

Applet

Joy Milne, a retired nurse from Scotland, noticed a "subtle musky odor" on her husband Les that she had never encountered before. At first, Joy thought the smell might be from Les's sweat after long hours of work. But when Les was diagnosed with Parkinson's disease 6 years later, Joy suspected the odor might be a result of the disease.

Scientists were intrigued by Joy's claim and designed an experiment to test her ability to "smell Parkinson's." Joy was presented with 12 different shirts, each worn by a different person, some of whom had Parkinson's disease and some of whom did not. The shirts were given to Joy in a random order, and she had to decide whether or not each shirt was worn by a patient with Parkinson's disease. Joy identified 11 of the 12 shirts correctly.[1]

Although the researchers wanted to believe that Joy could detect Parkinson's disease by smell, it is possible that she was just a lucky guesser. You and your classmates will perform a simulation with cards (or using the *Can You Smell Parkinson's?* applet at stapplet.com) to determine which explanation is more believable. Here are the directions for the simulation with cards:



© Chris Watt/Telegraph Media Group Limited 2017

1. Your teacher will hand each pair of students a set of 12 cards (shirts). On the back of some cards is "Parkinson's" and on the back of others is "No Parkinson's." Shuffle the cards thoroughly.
2. Decide who will guess first and have your partner act as the researcher. For each card, guess "Parkinson's" or "No Parkinson's." The researcher will not reveal whether each guess is right or wrong, but will record the number of correct guesses. Now switch roles and repeat the process.
3. Your teacher will draw and label a number line for a class dotplot. Plot the number of correct guesses you made in Step 2 on the graph.
4. Repeat the process until you have a total of at least 50 trials of the simulation for your class.
5. How often were 11 or more shirts correctly identified by chance alone? Based on this result, which seems more believable: Joy is just a lucky guesser, or Joy really can smell Parkinson's disease? Explain your reasoning.

The Smelling Parkinson's activity outlines the steps in the **statistical problem-solving process.**[2]

DEFINITION Statistical problem-solving process

- **Formulate questions:** Clarify the research problem and ask one or more valid statistical investigative questions.
- **Collect data:** Design and carry out a plan to collect appropriate data.
- **Analyze data:** Use appropriate graphical and numerical methods to analyze the data.
- **Interpret results:** Draw conclusions based on the data analysis. Be sure to answer the investigative question(s)!

Researchers began by identifying the *investigative question* to be answered: "Can Joy Milne correctly identify Parkinson's disease status by smell for more than 50% of all shirts like the ones in this experiment?" To answer this question, the researchers designed and conducted a study to gather appropriate data. The resulting data consisted of Joy's 11 correct and 1 incorrect shirt identifications. When analyzing the data, researchers had to consider the possibility that Joy was just guessing and correctly identified 11 of the 12 shirts by chance alone. After careful analysis, the researchers concluded that Joy Milne could actually smell Parkinson's disease. To make the case even stronger, the researchers later discovered that Joy had correctly identified all 12 shirts. Her one "mistake" was a person who was diagnosed with Parkinson's disease a few months later. That's pretty amazing!

In AP® Statistics, you will be asked to solve a variety of statistical problems. Your success will depend on mastering the course content and on developing *skills* related to the four *statistical practices* highlighted in the statistical problem-solving process: formulate questions, collect data, analyze data, and interpret results. See the table near the inside front cover for a listing of the statistical practices and skills in AP® Statistics.[3] We'll emphasize content as well as the statistical practices and skills throughout this book.

SECTION 1A Statistics: Learning from Data

LEARNING TARGETS *By the end of the section, you should be able to:*

- Determine a valid investigative question in a statistical study.
- Identify the population and sample in a statistical study.
- Identify the observational units and variables in a statistical study or data set, and classify the variables as categorical or quantitative.

AP® EXAM TIP: AP® Classroom
Preview the content of this section with the resources in AP® Classroom for Topics 1.1 and 1.2.

This section begins with a closer look at the first step in the statistical problem-solving process: formulate questions. Next, you will learn some basics about gathering and considering data. Unit 1, Part II provides additional details about the second step in the statistical problem-solving process: collect data.

Determining Investigative Questions

A statistical study starts with an investigative question. But not just any question will do. Unlike most mathematical questions, *a valid statistical investigative question is based on data that vary*. For instance, "Can Joy Milne correctly identify Parkinson's disease status by smell for more than 50% of all shirts like the ones in this experiment?" is a valid investigative question because Joy may identify some shirts correctly and other shirts incorrectly. However, "How much screen time did you have yesterday?" is not a valid investigative question because the question can be answered with a single data value.

We'll examine the components of statistical investigative questions in more detail in Section 1F. For now, you should be able to determine a valid investigative question in a statistical study.

Skill 1.A

EXAMPLE

Out to lunch
Determining investigative questions

PROBLEM: Two friends go off campus every day to eat lunch. Because the lunch period is short, they wonder whether it would be faster to order inside their favorite fast-food restaurant or to use the drive-thru. Each day, they flip a coin to determine which method (inside or drive-thru) to use and record the total length of time it takes from the moment they enter the parking lot to the moment they receive their food. After several weeks of collecting data, they analyze their results and determine that ordering inside took about 2 minutes and 34 seconds less than using the drive-thru, on average. Their conclusion: it's faster to go inside.

Determine the investigative question in this statistical study.



Erik Isakson/Tetra images/Getty Images

SOLUTION:

During all possible lunch visits to this restaurant, is the average time to receive food when ordering inside different than the average time to receive food when ordering at the drive-thru?

This is a valid investigative question because the length of time it takes these friends to enter the parking lot, place their order, and receive their food will vary from day to day.

FOR PRACTICE, TRY EXERCISE 3

Every statistical study or data set involves a real-world context. On the AP® Statistics exam, you will be expected to answer questions "in context."

Populations and Samples

Suppose we want to find out what percentage of young drivers in the United States text while driving. To answer this question, we will survey 16- to 20-year-old drivers who live in the United States. Ideally, we would ask them all by conducting a **census**. Of course, contacting every driver in this age group wouldn't be practical—it would take too much time and cost too much money. Instead, we pose the question to a **sample** chosen to represent the entire **population** of young drivers.

DEFINITION Population, Census, Sample

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

A **census** collects data from every item or individual in the population.

A **sample** is a subset of items or individuals in the population from which we collect data.

Skill 2.A

EXAMPLE Sampling monitors and voters
Populations and samples



Vladimir Vladimirov/E+/Getty Images

PROBLEM: Identify the population and the sample in each of the following settings.

(a) The quality control manager at a factory selects 10 computer monitors from the 50 monitors produced during a particular hour and inspects each monitor for defects in construction and performance.

(b) Prior to an election, a news organization surveys 1000 registered voters to predict the percentage of voters who prefer candidate A for president.

SOLUTION:

(a) The population is all 50 computer monitors produced in this factory during that hour. The sample is the 10 monitors selected and inspected for defects.

(b) The population is all registered voters. The sample is the 1000 registered voters surveyed.

To identify the population, consider which items or individuals could have been selected for the sample. In part (a), the sample was selected from just computer monitors produced during that hour, so the population is limited to all monitors produced in this factory during that hour.

FOR PRACTICE, TRY EXERCISE 7

In a statistical study, the sample size is represented by n and the population size is represented by N. In part (a) of the example, the quality control inspector selected a sample of $n = 10$ computer monitors from the population of $N = 50$ monitors produced that hour.

Most statistical studies collect data from samples to answer investigative questions about larger populations. In part (b) of the example, the news organization wants to answer the investigative question "What are the plausible (believable) values for the percentage of all registered voters who prefer candidate A for president?" We refer to this unknown percentage as the population **parameter**. Suppose that 528 of the 1000 registered voters surveyed prefer candidate A for president. That's 52.8%, which we refer to as the sample **statistic**. Remember p and s: parameters come from populations and statistics come from samples.

DEFINITION Parameter, Statistic

A **parameter** is a number that describes some characteristic of a population.

A **statistic** is a number that describes some characteristic of a sample.

We use sample statistics to estimate population parameters. In the news organization survey, 52.8% of the sample of registered voters prefer candidate A for president. That's our best guess for the unknown population percentage. Should we conclude that exactly 52.8% of the population of registered voters prefer candidate A for president? No. If another sample of 1000 registered voters was surveyed, the percentage who prefer candidate A for president would probably yield a different sample statistic. Can we at least say that the actual population parameter is "close" to 52.8%? As you will learn in future units, that depends on what we mean by "close" and how the sample was selected.

Observational Units and Variables

Every year, the U.S. Census Bureau collects data from more than 3.5 million randomly selected U.S. households as part of the American Community Survey (ACS). The table displays some data from 10 households included in the ACS in a recent year.[4]

Household	Region	Number of people in household	Time in dwelling (years)	Response mode	Household income ($)	Internet access?
1	South	3	20–29	Phone	272,000	Yes
2	Midwest	3	30+	Internet	54,600	Yes
5	South	1	10–19	Mail	49,900	Yes
6	South	2	2–4	Phone	297,000	Yes
7	Northeast	4	5–9	Internet	130,000	Yes
11	Midwest	3	2–4	Internet	82,000	Yes
14	Midwest	3	10–19	Internet	57,000	Yes
16	West	3	2–4	Mail	36,800	Yes
17	South	2	2–4	Mail	133,000	Yes
25	Midwest	1	20–29	Mail	80,000	No


Bruce Leighty/Photodisc/Getty Images

Most data tables follow this format: each row describes an **observational unit**, and each column contains the values of a **variable**. The entry in a single cell of the data table, representing the value of one variable for a specific observational unit, is called a *datum*—the singular form of the word *data*. Sometimes the observational units in a data set are called *individuals* or *cases*. Note that observational units can be people, animals, or things, such as photographs, sounds, or videos.

DEFINITION Observational unit, Variable

An **observational unit** is an item or individual described in a data set or statistical study.

A **variable** is a characteristic that can take different values for different observational units.

For the ACS data set, the *observational units* are households. The *variables* recorded for each household are region, number of people in the household,

time in their current dwelling, survey response mode, household income (in dollars), and whether the dwelling has internet access. It is important to note that household is *not* a variable in this data set—the numbers in that column of the data table are just labels for the observational units. Region, time in dwelling, response mode, and internet access status are **categorical variables**. Number of people in the household and household income (in dollars) are **quantitative variables**.

DEFINITION Categorical variable, Quantitative variable

A **categorical variable** takes values that are labels, which place each item or individual into a particular group, called a category.

A **quantitative variable** takes number values that are quantities—counts or measurements.

Sometimes categorical variables are called *qualitative variables* and quantitative variables are called *numerical variables*. It can also be helpful to think of a quantitative variable as one that takes number values for which it makes sense to find the average.

Not every variable that takes number values is quantitative. Zip code is one example. Although zip codes are numbers, they are neither counts nor measurements of anything. They are simply labels for a regional location, making zip code a categorical variable. Time in dwelling from the ACS data set is also a categorical variable because the values are recorded as intervals of time, such as 2–4 years. If time in dwelling had been recorded to the nearest year for each household, this variable would be quantitative.

In many data sets, the variable "year" is treated as categorical. But it depends on how the data are being used. Consider a data set about cars, in which one of the variables recorded is model year. If we want to know what percentage of cars on the road are 2026 models, we treat year as categorical. If we want to know the average age of cars on the road, we would convert model year to age (in years) and treat this variable as quantitative.

EXAMPLE Census at School
Observational units and variables

Skill 2.A

PROBLEM: Census at School is an international project that collects data about primary and secondary school students using online surveys. Since its launch in 2000, students from Australia, Canada, Ireland, Japan, New Zealand, South Africa, the United Kingdom, and the United States have taken part in the project.[5] We selected a random sample of 50 U.S. high school students who completed the survey in a recent year. The table displays data from some of the survey questions for the first 10 students in the sample.



EF Volart/Moment/Getty Images

State	Birth month	Age (years)	Handedness	Height (cm)	Number of home occupants	Allergies	Preferred communication method
WI	11	17	Right	175	4	Yes	Internet chat/IM
IN	6	16	Right	175.5	5	No	In person
NY	6	17	Right	157	5	Yes	In person
NC	6	17	Right	169	3	No	Internet chat/IM
MA	6	18	Right	169	3	Yes	Phone call
MO	10	18	Right	170	5	No	Text messaging
PA	5	14	Right	170	6	No	Text messaging
IA	1	17	Left	176	2	No	Text messaging
NC	5	17	Right	175	5	No	Social media
CA	2	17	Right	158	8	Yes	Social media
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(a) Identify the observational units and variables in this data set.

(b) Classify each variable as categorical or quantitative.

SOLUTION:

(a) Observational units: 50 randomly selected U.S. high school students who completed the Census at School survey. Variables: State where student lives, birth month, age (years), handedness, height (cm), number of home occupants, whether the student has allergies, preferred communication method.

We'll see in Unit 1, Part II why selecting at random, as we did in this example, is a good idea.

(b) Categorical: State where student lives, birth month, handedness, whether the student has allergies, preferred communication method. Quantitative: Age (years), height (cm), and number of home occupants.

Note that birth month is categorical, even though the values listed are numbers. It wouldn't make sense to find the average for this variable!

FOR PRACTICE, TRY EXERCISE 13

As you will learn, the proper method of data analysis depends on whether a variable is categorical or quantitative. For that reason, it is important to distinguish between these two types of variables. Be sure to include any units of measurement for a quantitative variable (such as centimeters for height). To make life simpler, we sometimes refer to *categorical data* or *quantitative data* instead of identifying the variable as categorical or quantitative.

AP® EXAM TIP

If you learn to distinguish categorical from quantitative variables now, it will pay big rewards later. You will be expected to analyze categorical and quantitative data appropriately on the AP® Statistics exam.

There are two types of quantitative variables: *discrete* and *continuous*. Most **discrete quantitative variables** result from counting something, such as the number of people in a household or the number of lottery tickets a person buys until they win the jackpot. **Continuous quantitative variables** typically result

from measuring something, such as height (in inches) or time to run a 100-meter dash (in seconds). Age is technically a continuous quantitative variable—a high school student's age might be 17.30162... years. Even so, it is often treated as a discrete quantitative variable—for example, age = 17 years.

DEFINITION Discrete quantitative variable, Continuous quantitative variable

A **discrete quantitative variable** can take a set of possible values with gaps between them on the number line.

A **continuous quantitative variable** can take any value in an interval on the number line.

Note that the number of possible values of a discrete quantitative variable can be finite or infinite, whereas a continuous quantitative variable always has an infinite number of possible values. Some people say that the values of discrete quantitative variables are *countable*, which means that we can number the possible values of the variable using whole numbers. As an example, the number of lottery tickets a person buys until they win the jackpot can take any of the values 1, 2, 3, and so on. The values of a continuous quantitative variable, however, are not countable. For instance, the height of a fully grown giraffe (in feet) could be *any* value between about 14 feet and 18 feet.[6] There is no way to number all of these possible heights using whole numbers.



CHECK YOUR UNDERSTANDING

Malena is a car buff who wants to find out more about the cars that high school students drive. The principal of a large local high school gives Malena permission to go to the student parking lot and collect some data. Malena selects a random sample of 50 cars from the lot and creates a spreadsheet with each car's license plate, model, year, number of stickers on the car, color, weight (in kilograms), whether it has a navigation system, and highway gas mileage (in miles per gallon).

1. Identify the sample and the population in this statistical study.
2. Identify the observational units and variables. Classify each variable as categorical or quantitative.
3. Formulate a valid investigative question that Malena could ask about the number of stickers on students' cars.

SECTION 1A Summary

- **Statistics** is the science and art of collecting, analyzing, and drawing conclusions from data.
- The **statistical problem-solving process** involves four steps: formulate questions, collect data, analyze data, and interpret results.
- Most statistical studies collect data from samples to answer investigative questions about larger populations. A valid investigative question is based on data that vary.

- The **population** in a statistical study is the entire group of items or individuals we want information about. A **census** collects data from every item or individual in the population.
- A **sample** is a subset of items or individuals in the population from which we collect data.
- A **statistic** is a number that describes some characteristic of a sample. A **parameter** is a number that describes some characteristic of a population.
- An **observational unit** is an item or individual described in a data set or statistical study. Observational units may be people, animals, or things.
- A **variable** is a characteristic that can take different values for different observational units.
- A **categorical variable** takes values that are labels, which place each individual into a particular group, called a category. A **quantitative variable** takes number values that are quantities—counts or measurements. Sometimes categorical variables are called *qualitative variables* and quantitative variables are called *numerical variables*.
- There are two types of quantitative variables: discrete and continuous. A **discrete quantitative variable** can take a set of possible values with gaps between them on the number line. A **continuous quantitative variable** can take any value in an interval on the number line. Discrete quantitative variables usually result from counting something; continuous quantitative variables usually result from measuring something.

AP® EXAM TIP
AP® Classroom

Review the content of this section with the resources in AP® Classroom for Topics 1.1 and 1.2.

SECTION 1A Exercises

The solutions to all exercises numbered in red can be found in the Solutions Appendix.

Determining Investigative Questions

1. **Regular exercise** Determine whether each of the following is a valid statistical investigative question. Explain your answer.

(a) What are the plausible values for the proportion of all U.S. adults who engage in vigorous exercise at least once per week?

(b) On how many days in the past week did you engage in vigorous exercise?

2. **Carrying cash** Determine whether each of the following is a valid statistical investigative question. Explain your answer.

(a) How much money are you carrying right now?

(b) What are the plausible values for the average amount of money that all high school students are carrying on the first day of school?

3. **Boarding time** Airlines are interested in finding ways for passengers to board their flights more quickly and efficiently. Researchers tested different boarding methods using a group of 72 volunteer passengers of varying ages. The researchers compared the "back-to-front" boarding method used by many airlines at the time with a modified version of this method proposed by astrophysicist Jason Steffen. In the modified method, passengers lined up in advance at the gate in a predetermined order, so that all passengers with an even-numbered window seat would board first from back to front, followed by those with an odd-numbered window seat, then those with an even-numbered middle seat, and so on. On average, the Steffen method was twice as fast as the traditional boarding method.[7] Determine the investigative question in this statistical study.

pg 4

4. **Paid to quit** In an effort to reduce health care costs, General Motors sponsored a study to help its employees stop smoking. In the study, 439 volunteer subjects were assigned at random to receive up to $750 for quitting for a year, while the other 439 volunteer subjects were simply encouraged to use traditional methods to stop smoking. After one year, people who had the financial incentive were 3 times more likely to have quit smoking.[8] Determine the investigative question in this statistical study.

5. **Online bullying** The Pew Research Center surveyed a random sample of 1316 U.S. teens aged 13 to 17. The survey asked about whether the person had ever experienced online bullying and also recorded data about the respondents' age, race/ethnicity, area where they live (urban/suburban/rural), and household income.[9] Formulate a valid investigative question that researchers could ask using the data from this study.

6. **Beach living?** Researchers surveyed more than 15,000 people in Europe and Australia about their health and where they live. One question on the survey was: "In general, would you say your health is: 1 – Excellent, 2 – Very good, 3 – Good, 4 – Fair, or 5 – Poor?" Another question asked, "Approximately how far do you live from the coast in kilometers (km): <1 km, 1–<2 km, 2–<5 km, 5–<10 km, 10–<20 km, 20–<50 km, 50–<100 km, or 100+ km?"[10] Formulate a valid investigative question that researchers could ask using the data from this study.

Populations and Samples

7. **Sampling stuffed envelopes** A large retailer prepares its customers' monthly credit card bills using an automatic machine that folds the bills, stuffs them into envelopes, and seals the envelopes for mailing. To ensure that the envelopes are completely sealed, inspectors choose 40 envelopes at random from the 1000 envelopes stuffed in one hour for visual inspection. Identify the population and the sample.

pg 5

8. **Student archaeologists** An archaeological dig turns up large numbers of pottery shards, broken stone tools, and other artifacts. Students working on the project classify each artifact and assign a number to it. The counts in different categories are important for understanding the site, so the project director chooses 2% of the artifacts at random and checks the students' work. Identify the population and the sample.

9. **Students as customers** A high school's student newspaper plans to survey businesses in their large city about the importance of students as customers. From an alphabetical list of all local businesses, the newspaper staff chooses 150 businesses at random. Of these, 73 return the questionnaire mailed by the staff. Identify the population and the sample.

10. **Customer satisfaction** A department store mails a customer satisfaction survey to people who make credit card purchases at the store. This month, 45,000 people made credit card purchases. Surveys are mailed to 1000 of these people, chosen at random, and 137 people return the survey. Identify the population and the sample.

11. **Dairy cows** A farmer has 1000 dairy cows and wants to estimate the minimum daily milk production for dairy cows in the herd. For 20 randomly chosen dairy cows, the minimum production was 6.3 gallons. Identify the population, the parameter, the sample, and the statistic in this setting.

12. **On sale now?** Advertisements for a local clothing store claim that 60% of the items are on sale this week. A consumer group randomly selected 20 items and found that only 9 were on sale. Identify the population, the parameter, the sample, and the statistic in this setting.

Observational Units and Variables

13. **A class survey** Here is a small part of a data set that describes the students in a math class. The data come from anonymous responses to a questionnaire filled out on the second day of class.

pg 7

Grade level	Dominant hand	GPA	Number of children in family	Homework time last night (min)	Type of phone
9	L	2.3	3	0–14	iPhone
11	R	3.8	6	15–29	Android
10	R	3.1	2	15–29	Android
10	R	4.0	1	45–59	iPhone
10	R	3.4	4	0–14	iPhone
10	L	3.0	3	30–44	Android
9	R	3.9	2	15–29	iPhone
12	R	3.5	2	0–14	iPhone

(a) Identify the observational units and variables in this data set.

(b) Classify each variable as categorical or quantitative.

14. **Coaster craze** Many people like to ride roller coasters. Amusement parks try to increase attendance by building exciting new coasters. The following table displays data on several roller coasters from around the world.[11]

Roller coaster	Type	Height (ft)	Design	Speed (mph)	Duration (sec)
Copperhead Strike	Steel	82.0	Sit down	50.0	144
Eurostar	Steel	98.9	Inverted	50.2	140
Jungle Trailblazer	Wood	108.3	Sit down	54.1	150
Falcon	Steel	197.5	Wing	73.3	156
Olympia Looping	Steel	106.7	Sit down	49.7	105
Time Traveler	Steel	100.0	Sit down	50.3	117

(a) Identify the observational units and variables in this data set.

(b) Classify each variable as categorical or quantitative.

15. **Hit movies** *Avatar* was the top-earning movie based on box-office receipts worldwide as of February 2025. The following table displays data on several popular movies.[12] Identify the observational units and variables in this data set. Then classify each variable as categorical or quantitative.

Movie	Year	Rating	Time (min)	Genre	Box office ($)
Avatar	2009	PG-13	162	Action	2,923,706,026
Avengers: Endgame	2019	PG-13	181	Action	2,748,242,781
Avatar: The Way of Water	2022	PG-13	190	Action	2,313,161,020
Titanic	1997	PG-13	194	Drama	2,223,048,786
Star Wars: The Force Awakens	2015	PG-13	136	Adventure	2,056,046,835
Avengers: Infinity War	2018	PG-13	156	Action	2,048,359,754
Spider Man: No Way Home	2021	PG-13	148	Action	1,921,206,586
Inside Out 2	2024	PG	100	Adventure	1,698,863,816
Jurassic World	2015	PG-13	124	Action	1,671,063,641
The Lion King	2019	PG	118	Adventure	1,661,454,403

16. **Skyscrapers** Here is some information about the tallest buildings in the world as of 2025.[13] Identify the observational units and variables in this data set. Then classify each variable as categorical or quantitative.

Building	Country	Height (m)	Floors	Use	Year completed
Burj Khalifa	United Arab Emirates	828.0	163	Mixed	2010
Merdeka PNB 118	Malaysia	678.9	118	Mixed	2023
Shanghai Tower	China	632.0	128	Mixed	2015
Abraj Al Bait	Saudi Arabia	601.0	120	Hotel	2012
Ping An Finance Center	China	599.1	115	Office	2017
Lotte World Tower	South Korea	554.5	123	Mixed	2017
One World Trade Center	United States	541.3	94	Office	2014
Tianjin CTF Finance Center	China	530.4	97	Mixed	2019
Guangzhou CTF Finance Centre	China	530.0	111	Mixed	2016
CITIC Tower	China	527.7	109	Mixed	2018

17. **Protecting wood** How can we help wood surfaces resist weathering, especially when restoring historic wooden buildings? In a study that attempted to answer this question, researchers prepared wooden panels and then exposed them to the weather. Here are some of the variables the researchers recorded: type of wood, paint thickness, paint color, weathering time (1, 2, or 3 months), and number of blemishes.

(a) Identify the observational units.

(b) Classify each variable as categorical, quantitative (discrete), or quantitative (continuous).

18. **Social media** A social media company records data on each of its users for several variables: internet provider, age, how many times they visited the site, total time spent on the site, country where they live, and how long since they created a member profile.

(a) Identify the observational units.

(b) Classify each variable as categorical, quantitative (discrete), or quantitative (continuous).

Multiple Choice *Select the best answer for each question.*

Exercises 19 and 20 refer to the following setting. A realtor in Hilton Head, South Carolina, collected data on all homes sold in the town during 2025. The resulting data set lists each home's address, along with the following information: zip code, number of bedrooms, primary building material (wood, brick, etc.), square footage, whether it has a pool, age of home, and sales price (in dollars).

19. The observational units in this data set are

(A) dollars.

(B) all Hilton Head homeowners who sold their homes during 2025.

(C) all homes sold in Hilton Head during 2025.

(D) all homes in Hilton Head during 2025.

20. This data set contains

(A) 7 variables, 2 of which are categorical.

(B) 7 variables, 3 of which are categorical.

(C) 8 variables, 3 of which are categorical.

(D) 8 variables, 4 of which are categorical..

21. The manager of a town's athletic complex is trying to decide whether to convert some of the existing tennis courts to pickleball courts. To help with the decision, the manager will survey a random sample of 200 adult town residents about whether they are in favor of this conversion. If the data provide convincing evidence that more than half of all adult town residents favor the conversion, the manager will submit a budget proposal to the town council. Which of the following is the investigative question for this statistical study?

(A) Is the proportion of adult town residents in the sample who use the town's athletic complex greater than 0.5?

(B) Is the proportion of all adult town residents who use the town's athletic complex greater than 0.5?

(C) Is the proportion of adult town residents in the sample who favor converting some existing tennis courts to pickleball courts greater than 0.5?

(D) Is the proportion of all adult town residents who favor converting some existing tennis courts to pickleball courts greater than 0.5?

SECTION 1B Displaying and Describing Categorical Data

LEARNING TARGETS *By the end of the section, you should be able to:*

- Make and interpret a frequency table or a relative frequency table for a distribution of categorical data.
- Make and interpret bar charts and pie charts of categorical data.
- Compare distributions of categorical data.
- Identify what makes some graphs of categorical data misleading.

AP® EXAM TIP: AP® Classroom
Preview the content of this section with the resources in AP® Classroom for Topics 1.3 and 1.4.

A variable generally takes values that vary from one observational unit to another. That's why we call it a variable! The **distribution** of a variable describes the pattern of variation of the values.

DEFINITION Distribution
The **distribution** of a variable tells us what values the variable takes and how often it takes each value.

In this section, you will learn how to display and describe the distribution of one categorical variable. You will also learn how to compare distributions of a categorical variable in two or more groups. Sections 1C, 1D, and 1E focus on displaying and describing the distribution of one quantitative variable, and on comparing distributions of a quantitative variable in multiple groups. Section 2A examines relationships between two categorical variables, and Unit 5 examines relationships between two quantitative variables. This process of exploratory data analysis is known as *descriptive statistics*.

Summarizing Categorical Data with Tables

We can summarize the distribution of a categorical variable with a **frequency table** or a **relative frequency table.** To make either kind of table, start by tallying the number of times that the variable takes each value.

DEFINITION Frequency table, Relative frequency table

A **frequency table** shows the number of observational units having each value of a variable.

A **relative frequency table** shows the proportion or percentage of observational units having each value of a variable.

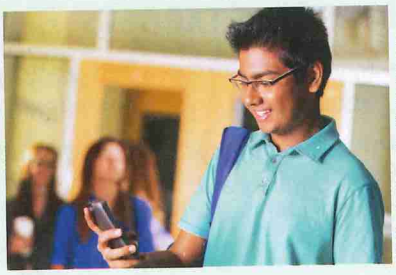
We can use a frequency table or a relative frequency table to describe the distribution of a categorical variable or to help justify a claim about the variable in context.

Skills 3.A, 4.B

EXAMPLE Call me, maybe? Summarizing categorical data with tables

PROBLEM: Here are the data on preferred method of communicating with friends for all 50 students in the Census at School sample from Section 1A:

Internet chat/IM	In person	In person	Internet chat/IM	Phone call
Text messaging	Text messaging	Text messaging	Social media	Social media
Text messaging	Text messaging	Text messaging	In person	In person
Text messaging	Text messaging	Text messaging	In person	In person
Text messaging	Text messaging	Internet chat/IM	Social media	Text messaging
Text messaging	Text messaging	In person	Text messaging	Internet chat/IM
Text messaging	Social media	Text messaging	Social media	Text messaging
Text messaging	In person	Social media	In person	Text messaging
Text messaging	Text messaging	Text messaging	Internet chat/IM	In person
Text messaging	In person	Internet chat/IM	Text messaging	In person



SDI Productions/E+/Getty Images

(a) Make a frequency table and a relative frequency table to summarize the distribution of preferred communication method.

(b) Do these data support the claim that a majority of students prefer to communicate with their friends using text messaging? Justify your answer.

SOLUTION:

(a)

Frequency table

Preferred method	Frequency
In person	12
Internet chat/IM	6
Phone call	1
Social media	6
Text messaging	25
Total	50

The frequency table shows the *number* of students who prefer each communication method. To create the frequency table, count how many students said "In person," how many said "Internet chat or instant messenger," and so on.

Relative frequency table

Preferred method	Relative frequency
In person	12/50 = 0.24 or 24%
Internet chat/IM	6/50 = 0.12 or 12%
Phone call	1/50 = 0.02 or 2%
Social media	6/50 = 0.12 or 12%
Text messaging	25/50 = 0.50 or 50%
Total	50/50 = 1.00 or 100%

The relative frequency table shows the *ratio* (as a fraction), *proportion*, and *percentage* of students who prefer each communication method. Note that in statistics, a proportion is a value between 0 and 1 that is equivalent to a percentage.

(b) No. Exactly 50% of the students in the sample said that they prefer to communicate with their friends via text messaging, but a majority is more than half.

FOR PRACTICE, TRY EXERCISE 1

Note that the frequencies and relative frequencies listed in these tables are not data. The frequency and relative frequency tables summarize the data by telling us how many, or what proportion or percentage of, students in the Census at School sample prefer each method of communicating with friends.

The same process could be used to summarize the distribution of a quantitative variable. However, it would be hard to make a frequency table or a relative frequency table for quantitative data that take many different values, such as height (cm) in the Census at School data set. We'll look at a better option for summarizing the distribution of a quantitative variable in Section 1C.

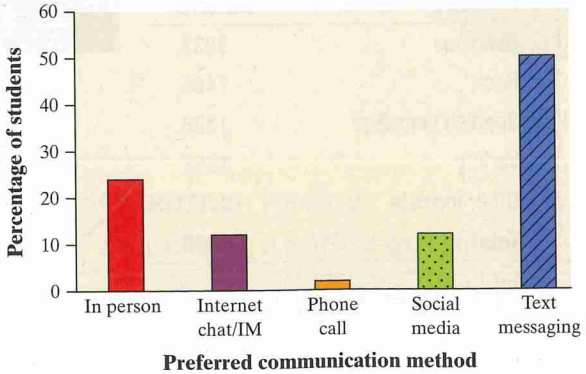
Displaying Categorical Data: Bar Charts and Pie Charts

A frequency table or relative frequency table summarizes a categorical variable's distribution with numbers. To display the distribution more clearly, use a graph. The most common way to display categorical data is with a **bar chart** (also known as a *bar graph*).

DEFINITION Bar chart

A **bar chart** is a graph of data for one categorical variable that displays each category as a bar. The height of each bar shows the category frequency or relative frequency.

Figure 1.1 shows a bar chart of the data on preferred communication method for the random sample of 50 U.S. high school students who completed the Census at School survey. Note that the percentages for each category come from the relative frequency table.



Relative frequency table

Preferred method	Relative frequency
In person	12/50 = 0.24 or 24%
Internet chat/IM	6/50 = 0.12 or 12%
Phone call	1/50 = 0.02 or 2%
Social media	6/50 = 0.12 or 12%
Text messaging	25/50 = 0.50 or 50%

FIGURE 1.1 Bar chart and relative frequency table of the distribution of preferred communication method for a random sample of 50 U.S. high school students.

It is fairly straightforward to make a bar chart by hand. Here's how you do it.

HOW TO MAKE A BAR CHART

1. **Draw and label the axes.** Put the name of the categorical variable under the horizontal axis. To the left of the vertical axis, indicate whether the graph shows the frequency (count) or relative frequency (percentage or proportion) of observational units in each category.
2. **"Scale" the axes.** Write the names of the categories in a logical order at equally spaced intervals under the horizontal axis. On the vertical axis, start at 0 and place tick marks at equal intervals until you equal or exceed the largest frequency or relative frequency in any category.
3. **Draw bars** above the category names. Make the bars equal in width and leave gaps between them. Be sure that the height of each bar corresponds to the frequency or relative frequency of observational units in that category.

Note: You can also make a bar chart with the axes reversed, so that values of the categorical variable are on the vertical axis and frequencies or relative frequencies are on the horizontal axis.

Making a graph is not an end in itself. The real purpose of a graph is to help us interpret the data. When you look at a graph, always ask, "What do I see?" We can use a graph to describe the distribution of a categorical variable or to help justify a claim about the variable in context.

In Figure 1.1, the bar chart reveals that the most preferred method of communicating with friends for these high school students is text messaging (50%). The next most preferred method is talking with their friends in person (24%). Social media and internet chat/instant messenger are somewhat less popular (12% each) methods of communicating with friends. By far the least preferred method of communication for these students is a phone call (2%).

Skills 3.A, 4.A

EXAMPLE What's on the radio? Bar charts

PROBLEM: Nielsen Audio, the rating service for radio audiences, places U.S. radio stations into categories that describe the kinds of programs they broadcast. The frequency table summarizes the distribution of station format in a recent year.[14]

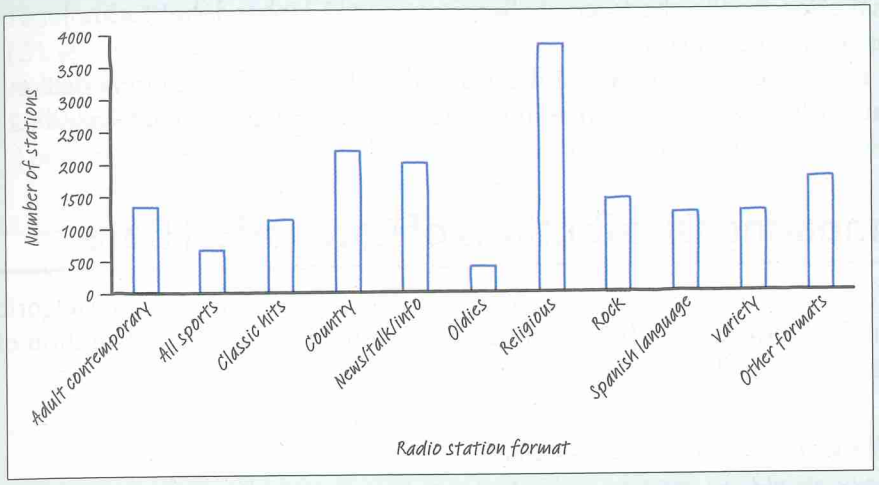


ACORN 1/Alamy Stock Photo

Format	Number of stations	Format	Number of stations
Adult contemporary	1357	Religious	3837
All sports	669	Rock	1466
Classic hits	1140	Spanish language	1228
Country	2200	Variety	1257
News/talk/information	2002	Other formats	1769
Oldies	405	**Total**	**17,330**

Make a frequency bar chart to display the data. Describe what you see.

SOLUTION:



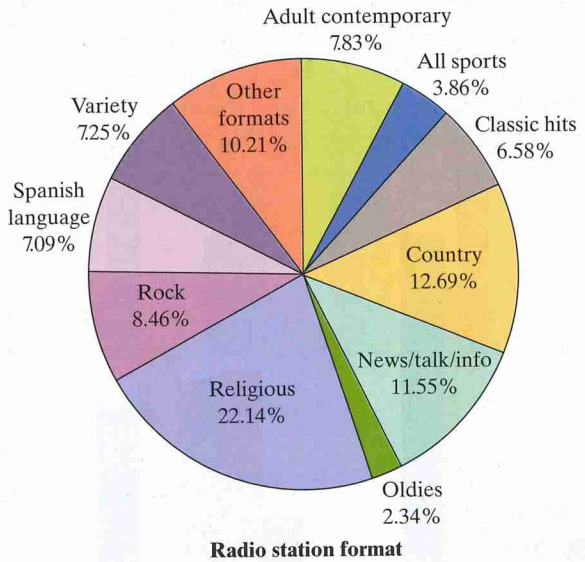
To make the bar chart:
1. **Draw and label the axes.**
2. **"Scale" the axes.** The largest frequency is 3837, so we choose a vertical scale from 0 to 4000, with tick marks 500 units apart.
3. **Draw bars** above the category names.

For U.S. radio stations, the most common formats are Religious (3837), Country (2200), and News/talk/information (2002), while the least common formats are Oldies (405) and All sports (669). Moderately common formats offered by a similar number of stations include Classic hits (1140), Spanish language (1228), Variety (1257), Adult contemporary (1357), and Rock (1466). There are 1769 stations with Other formats.

Note that the observational units in this statistical study are radio stations. The variable recorded for each observational unit is station format, which is categorical.

FOR PRACTICE, TRY EXERCISE 5

Here is a **pie chart** of the radio station format data from the preceding example. Notice that this graph shows the *relative frequency* for each response category. For instance, the "Spanish language" slice makes up 7.09% of the graph because the relative frequency for this category is $1228 / 17{,}330 = 0.0709 = 7.09\%$.



DEFINITION Pie chart

A **pie chart** is a graph of data for one categorical variable that displays each category as a slice of the "pie." The area of each slice is proportional to the category frequency or relative frequency.

Use a pie chart when you want to emphasize each category's relation to the whole. Each slice of the pie shows the count or percentage of observational units in that category. Pie charts are challenging to make by hand, but technology can easily do the job for you.

A bar chart or pie chart that displays a distribution of categorical data must include *all* observational units in the data set. This might require including an "Other" category, as in the radio station example.

Comparing Distributions of Categorical Data

You can use a bar chart or a pie chart to display the distribution of a categorical variable. A **side-by-side bar chart** can be used to compare the distribution of a categorical variable in two or more groups.

DEFINITION Side-by-side bar chart

A **side-by-side bar chart** is a graph of data for one categorical variable in each of two or more groups that displays a separate bar corresponding to each group for every category. The height of each bar shows the category frequency or relative frequency within that group.

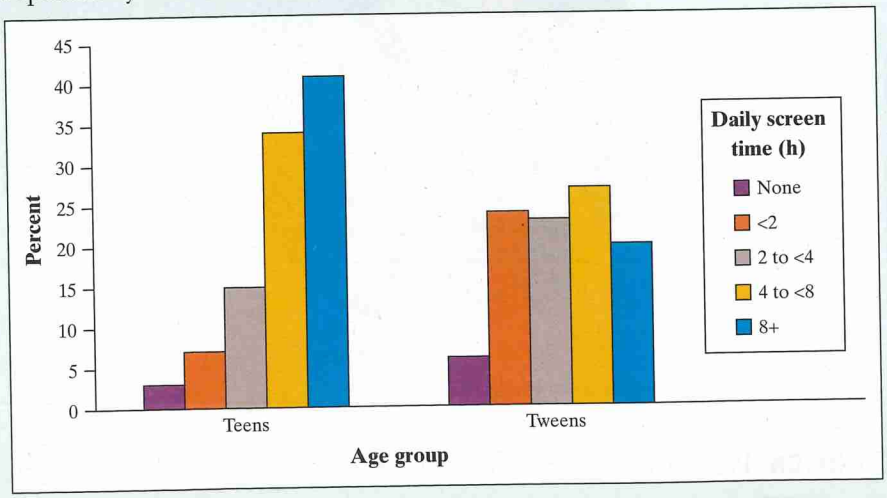
It's a good idea to use relative frequencies when comparing data for multiple groups, especially if the groups have different sizes.

Skill 4.A

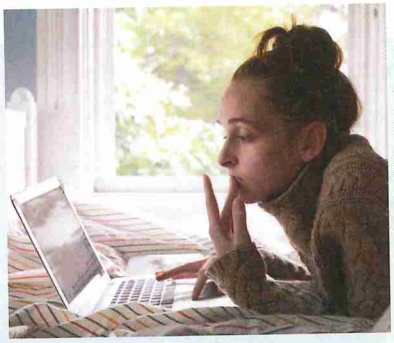
EXAMPLE

Screen time for teens and tweens
Comparing distributions of categorical data

PROBLEM: How much time do tweens (ages 8–12) and teens (ages 13–18) spend using digital devices each day? Researchers surveyed a random sample of more than 1300 U.S. 8- to 18-year-olds to find out. The side-by-side bar chart summarizes the data on daily screen time reported by the 560 tweens and 746 teens in the sample.[15]



Compare the distributions of daily screen time for teens and tweens.



Melanie Acevedo/DigitalVision/Getty Images

SOLUTION:

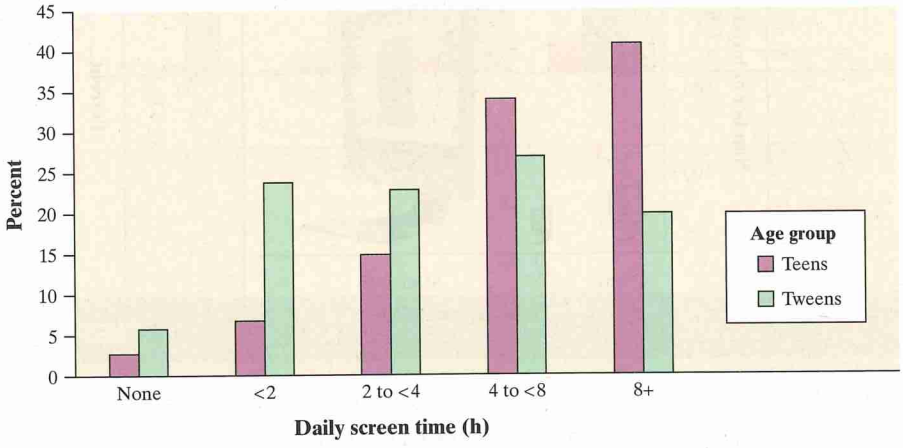
Tweens tend to report less daily screen time than teens. Tweens are more likely than teens to spend no time (6% versus 3%), less than 2 hours (24% versus 7%), and from 2 to less than 4 hours (23% versus 15%) on digital devices per day. Teens are more likely than tweens to spend from 4 to less than 8 hours (34% versus 27%) and 8 or more hours (41% versus 20%) on digital devices per day. Fewer than half (47%) of tweens report 4 or more hours of screen time per day, while 75% of teens report 4 or more hours of screen time per day.

FOR PRACTICE, TRY EXERCISE 11

The following relative frequency table summarizes the data on daily screen time from the example by age group. As a result, the percentages in each row add to 100%. Note that we could have also compared the distributions using only this relative frequency table.

		Daily time spent on digital devices				
		None	< 2 hours	2 to <4 hours	4 to <8 hours	8+ hours
Age group	Teens	3%	7%	15%	34%	41%
	Tweens	6%	24%	23%	27%	20%

In the example, we grouped the bars by age group in the side-by-side bar chart. This arrangement clearly shows the two distributions of daily screen time—one for the teens and one for the tweens in the sample. It is also possible to group the bars by daily screen time, as in the following graph. This arrangement makes category-by-category comparisons for the daily time spent on digital devices easier—but now it's harder to see the individual distributions of daily screen time for the teens and for the tweens. Whichever way you organize the bars, be sure to include a key that describes what each color or type of shading on the side-by-side bar chart represents.



It is possible, but more challenging, to compare distributions of categorical data with pie charts.

Misleading Graphs

Bar charts can be a bit dull to look at. It is tempting to replace the bars with pictures or to use special three-dimensional (3D) effects to make the graphs seem more interesting. Don't do it! Our eyes react to the area of the bars as well as to

their height. When all bars have the same width, the area (width × height) varies in proportion to the height, and our eyes receive the right impression about the quantities being compared.

Skill 3.A

EXAMPLE Who buys iMacs? Misleading graphs



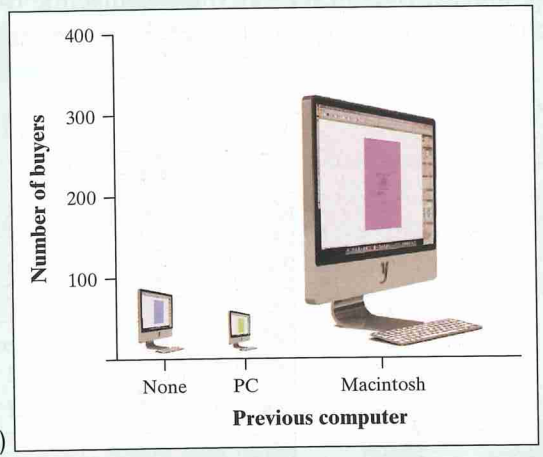
Getty Images/Getty Images

PROBLEM: When Apple, Inc., first introduced the iMac in 1998, the company wanted to know whether this new computer was expanding its market share. (The iMac has enjoyed great success ever since!) Was the iMac mainly being bought by previous Macintosh owners, or was it being purchased by first-time computer buyers and by previous PC users who were switching over? To find out, Apple hired a firm to conduct a survey of 500 randomly selected iMac customers. The firm categorized each customer as a new computer purchaser, a previous PC owner, or a previous Macintosh owner. The table summarizes the survey data.[16]

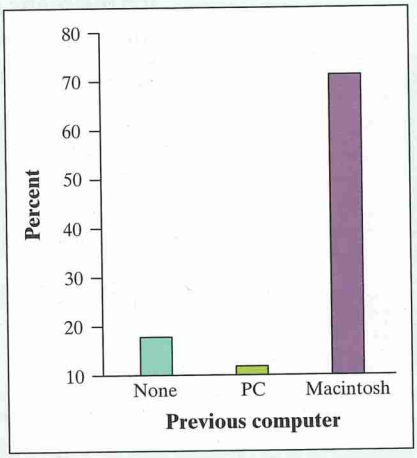
Previous ownership	Count	Percentage (%)
None	85	17.0
PC	60	12.0
Macintosh	355	71.0
Total	**500**	**100.0**

(a) A graph of the data that uses pictures instead of the more traditional bars is shown in figure (a). How is this pictograph misleading?

(b) A bar graph of the data is shown in figure (b). Explain why this graph could be considered deceptive.



(a)



(b)

SOLUTION:

(a) The pictograph is misleading because the areas of the computers make it look like the number of iMac buyers who are former Mac owners (355) is at least 10 times as large as the number of buyers in either of the other two categories (None: 85, PC owner: 60), which isn't true.

In part (a), the *heights* of the images are correct, but the *areas* of the images are misleading. The Macintosh image is about 6 times as tall as the PC image, but its area is about 36 times as large!

(b) The bar graph is misleading because starting the vertical scale at 10 instead of 0 makes it look like the percentage of iMac buyers who previously owned a PC (12.0%) is less than half the percentage who are first-time computer buyers (17.0%), which isn't true.

FOR PRACTICE, TRY EXERCISE 15

There are two important lessons about misleading graphs to be learned from this example: **(1) beware the pictograph and (2) watch those scales.**

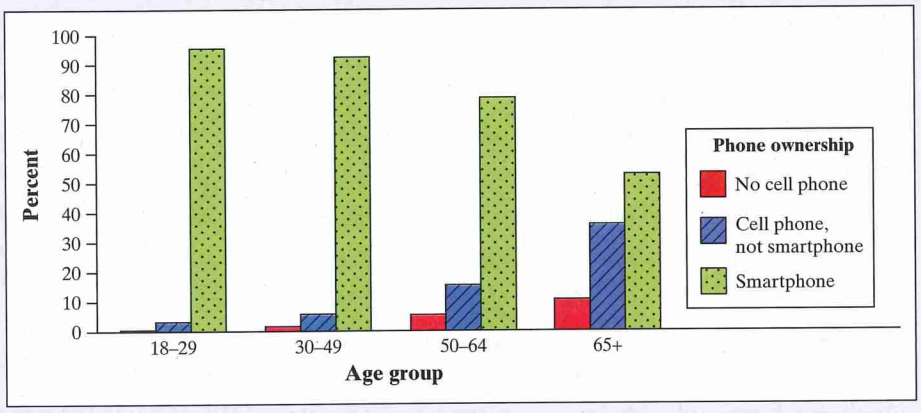


CHECK YOUR UNDERSTANDING

The Pew Research Center asked a random sample of 1502 U.S. adults about their cell phone ownership. The frequency table summarizes their responses.[17]

Type of cell phone	Frequency
None	73
Cell phone, not smartphone	212
Smartphone	1217
Total	**1502**

1. Make a relative frequency table for these data.
2. Make a relative frequency bar chart to display the distribution of cell phone ownership for the 1502 people in the sample. Describe what you see.
3. The side-by-side bar chart displays the distribution of cell phone ownership for each of four age groups in the sample. Use the graph to justify the claim that cell phone ownership differs across age groups.



SECTION 1B Summary

- The **distribution** of a variable describes what values the variable takes and how often it takes each value.
- To summarize the distribution of a variable, you can use a **frequency table** that shows the number of observational units having each value or a **relative frequency table** that shows the proportion or percentage of observational units having each value.
- A **bar chart** or **pie chart** can be used to display the distribution of a categorical variable.
- You can use a table or graph of categorical data to describe the distribution of a categorical variable or to justify a claim about the variable in context.

AP® EXAM TIP
AP® Classroom

Review the content of this section with the resources in AP® Classroom for Topics 1.3 and 1.4.

- You can use a **side-by-side bar chart** to compare the distribution of one categorical variable for two or more groups. Be sure to use relative frequencies when comparing groups of different sizes.
- Beware of graphs that mislead the eye. Look at the scales to see if they have been distorted to create a particular impression. Avoid making graphs that replace the bars of a bar chart with pictures whose height and width vary.

SECTION 1B Exercises

Summarizing Categorical Data with Tables

1. **New York squirrels** How do squirrels in New York City's Central Park respond to humans? The Squirrel Census project enlisted volunteers to record data on all squirrels observed in Central Park in a recent year. One of the variables measures how the squirrels responded to humans. We selected a random sample of 75 of the more than 3000 squirrels observed. Here are the data on their reactions to humans:[18]

pg 14

Approach	Run from	Indifferent	Indifferent	Run from
Indifferent	Approach	Indifferent	Indifferent	Indifferent
Approach	Indifferent	Approach	Run from	Run from
Indifferent	Indifferent	Indifferent	Run from	Indifferent
Indifferent	Indifferent	Run from	Indifferent	Indifferent
Indifferent	Run from	Run from	Indifferent	Indifferent
Run from	Run from	Approach	Approach	Run from
Indifferent	Approach	Indifferent	Indifferent	Indifferent
Indifferent	Approach	Indifferent	Indifferent	Indifferent
Indifferent	Run from	Approach	Approach	Indifferent
Indifferent	Indifferent	Run from	Approach	Indifferent
Indifferent	Approach	Indifferent	Run from	Approach
Run from	Approach	Run from	Indifferent	Indifferent
Indifferent	Approach	Indifferent	Run from	Indifferent
Indifferent	Indifferent	Approach	Approach	Indifferent

(a) Make a frequency table and a relative frequency table to summarize the distribution of response to humans.

(b) Were these Central Park squirrels more likely to approach humans, run from humans, or be indifferent to humans? Justify your answer.

2. **Disc dogs** Here is a list of the breeds of dogs that won the World Canine Disc Championships over a 45-year period.[19]

Whippet	Australian shepherd	Australian shepherd
Whippet	Border collie	Border collie
Whippet	Australian shepherd	Border collie
Mixed breed	Mixed breed	Australian shepherd
Mixed breed	Mixed breed	Border collie
Other purebred	Mixed breed	Border collie
Labrador retriever	Border collie	Other purebred
Mixed breed	Border collie	Border collie
Mixed breed	Australian shepherd	Border collie
Border collie	Border collie	Border collie
Mixed breed	Australian shepherd	Mixed breed
Mixed breed	Border collie	Australian shepherd
Labrador retriever	Mixed breed	Australian shepherd
Labrador retriever	Australian shepherd	Mixed breed
Mixed breed	Australian shepherd	Other purebred

(a) Make a frequency table and a relative frequency table to summarize the distribution of dog breed among World Canine Disc Championship winners.

(b) Which breeds of dog were the most frequent winners? Justify your answer.

3. **Don't worry, be happy!** College admissions departments are continually honing their message to appeal to high school students. Knowing about potential students' long-term goals and priorities helps colleges tailor information about the benefits of attending their institutions. Several colleges use an online survey to collect data about secondary school students. Here is one of the survey questions:

Which would you prefer to be? Select one.
_____ Rich _____ Happy _____ Famous _____ Healthy

The relative frequency table summarizes data on preferred status for a random sample of students who completed the survey in a recent year.

Preferred status	Proportion
Famous	0.040
Happy	0.520
Healthy	0.133
Rich	0.307

(a) There were a total of 75 students in the sample. How many said that their preferred status was rich?

(b) A college admissions officer suggests a recruiting campaign for prospective students that focuses on future happiness. Explain why this suggestion makes sense based on the data.

4. **CubeSat missions** A CubeSat is a miniature satellite used for space research that can be easily deployed from a launch vehicle or from the International Space Station. More than 2500 CubeSats have been launched since specifications for these satellites were jointly developed by Cal Poly San Luis Obispo and Stanford University in 1999. The relative frequency table summarizes data on the type of mission for each CubeSat launched in November and December 2019.[20]

Mission type	Proportion
Science	0.18
Technology	0.28
Imaging	0.34
Communications	0.12
Education	0.08

(a) There were a total of 50 CubeSats launched during these two months. How many had a communications mission type?

(b) A report states that a majority of these CubeSat launches had an imaging mission type. Explain why this statement is incorrect.

Displaying Categorical Data: Bar Charts and Pie Charts

5. **Birth days** The frequency table summarizes data on the numbers of babies born on each day of a single week in the United States.[21]

pg 16

Day	Births
Sunday	7374
Monday	11,704
Tuesday	13,169
Wednesday	13,038
Thursday	13,013
Friday	12,664
Saturday	8459

Make a frequency bar chart to display the data. Describe what you see.

6. **Going up?** Oliver Roeder wrote an interesting article about the more than 75,000 elevators in New York City. The frequency table summarizes data on the number of elevators of each type at that time.[22]

Type	Count	Type	Count
Passenger elevator	66,602	Private elevator	252
Freight elevator	4140	Handicap lift	227
Escalator	2663	Manlift	73
Dumbwaiter	1143	Public elevator	45
Sidewalk elevator	943		

Make a frequency bar chart to display the data. Describe what you see.

7. **Cool car colors** The popularity of colors for cars and light trucks changes over time. Silver passed green in 2000 to become the most popular color worldwide, then gave way to shades of white in 2007. The relative frequency table summarizes data on the colors of vehicles sold worldwide in 2020.[23]

Color	Percentage of vehicles	Color	Percentage of vehicles
Black	19	Red	5
Blue	7	Silver	9
Brown	3	White	38
Gray	15	Yellow	2
Green	1	Other	??

(a) What percentage of vehicles would fall in the "Other" category?

(b) Make a bar chart to display the data. Describe the distribution.

(c) Would it be appropriate to make a pie chart of these data? Explain your answer.

8. **Spam** Email spam is very annoying. The relative frequency table summarizes data on the most common types of spam.[24]

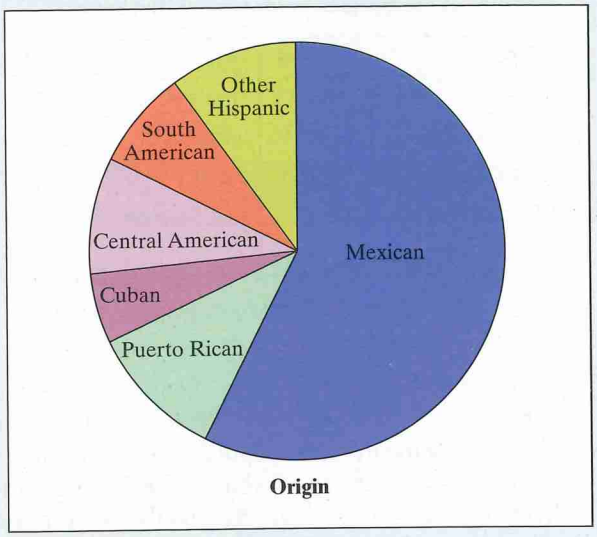
Type of spam	Percentage
Adult	19
Financial	20
Health	7
Internet	7
Leisure	6
Products	25
Scams	9
Other	??

(a) What percentage of spam would fall in the "Other" category?

(b) Make a bar chart to display the data. Describe the distribution.

(c) Would it be appropriate to make a pie chart of these data? Explain your answer.

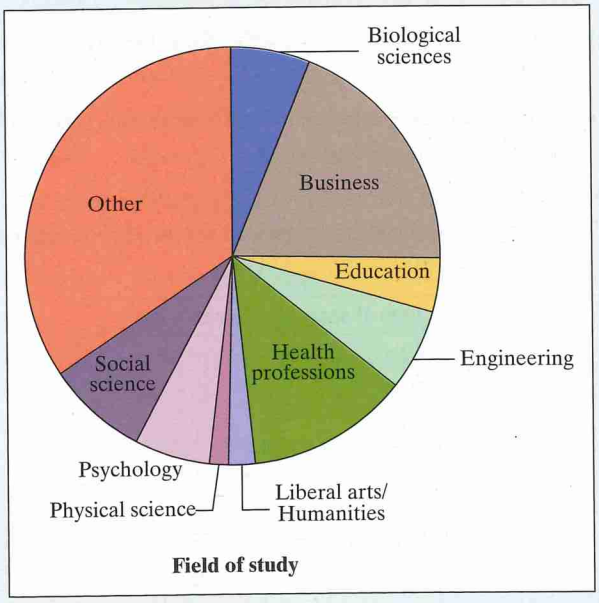
9. **Family origins** Here is a pie chart of U.S. Census Bureau data showing the origin of more than 62 million Hispanic people in the United States.[25]



(a) About what percentage of Hispanic people in the United States are of Mexican origin? Puerto Rican origin?

(b) Can we use this graph to justify the claim that people of South American origin make up the smallest proportion of Hispanic people in the United States? Explain your answer.

10. **Which major?** More than 2 million students earn bachelor's degrees in U.S. colleges and universities each year. The pie chart displays data on bachelor's degrees earned by field of study in a recent year.[26]



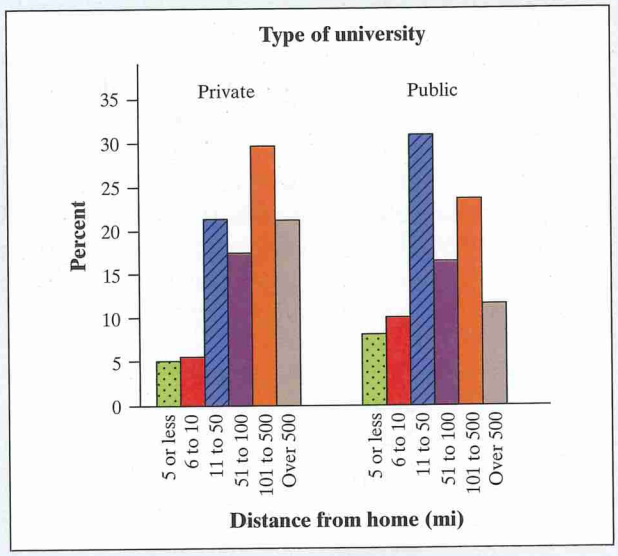
(a) About what percentage of students earned their bachelor's degrees in business? In health professions? In social science?

(b) Can we use this graph to justify the claim that the proportions of bachelor's degrees earned in psychology and engineering were about the same in this recent year in the United States? Explain your answer.

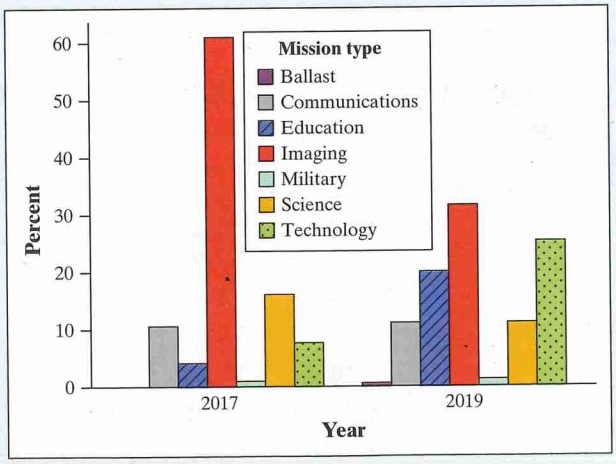
Comparing Distributions of Categorical Data

11. **Far from home** A survey asked first-year college students, "How many miles is this college from your permanent home?" Students selected from the following options: 5 or fewer, 6 to 10, 11 to 50, 51 to 100, 101 to 500, or more than 500. The side-by-side bar chart shows the percentage of students at public and private 4-year colleges who chose each option.[27] Compare the distributions of distance from home for students from private and public 4-year colleges who completed the survey.

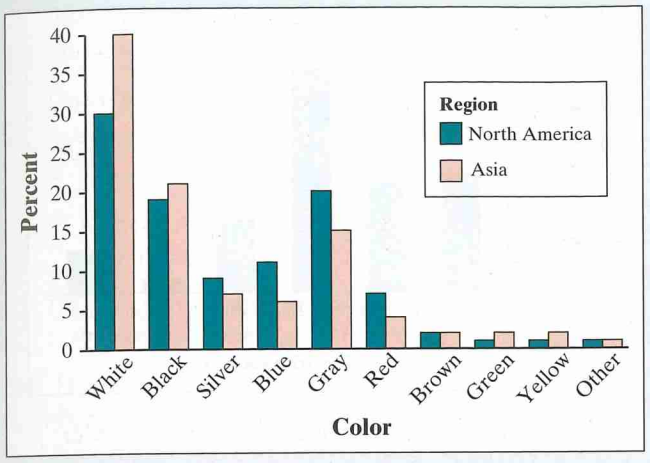
pg 18



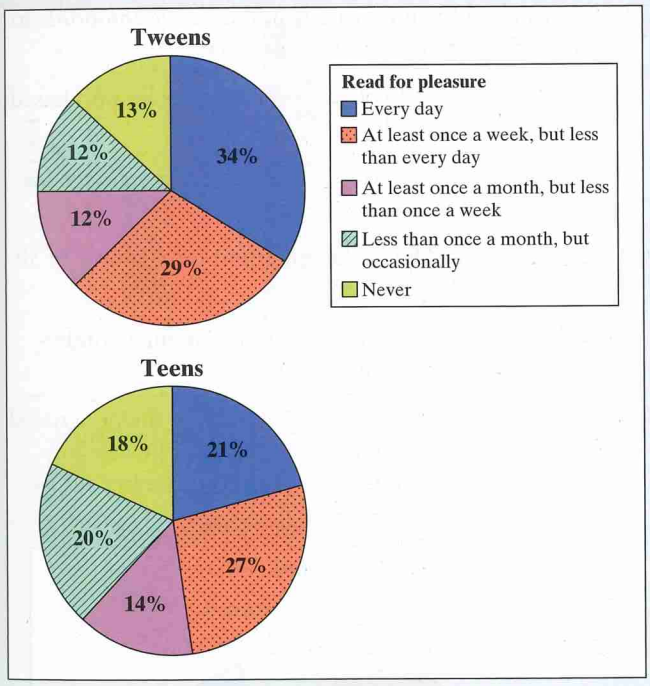
12. **More CubeSat missions** Refer to Exercise 4. The side-by-side bar chart displays data on the types of missions undertaken by CubeSats launched in 2017 and 2019. Compare the distributions of mission type for these two years.



13. **Popular car colors** Favorite car colors may differ among regions of the world. The side-by-side bar chart displays data on the most popular car colors in a recent year for North America and Asia.[28] Describe similarities and differences in the distributions for these two regions.



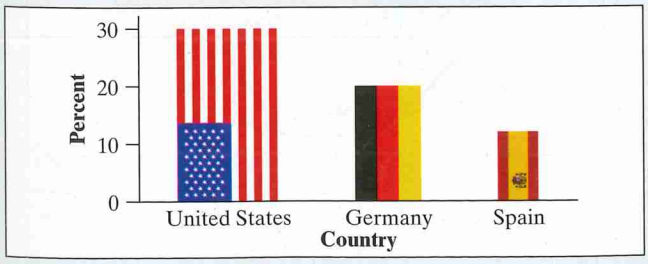
14. **Reading for pleasure** Researchers surveyed separate random samples of U.S. tweens (ages 8–12) and teens (ages 13–18) about how often they read for pleasure. The pie charts display the data.[29] Describe similarities and differences in the distributions for tweens and teens.



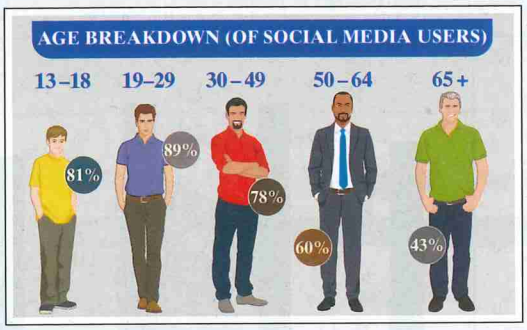
Misleading Graphs

15. **Who dislikes shopping?** Harris Interactive asked adults from several countries if they like or dislike shopping for clothes. The following pictograph displays the percentage of adults in the United States, Germany, and Spain who said that they dislike shopping for clothes. Explain how this graph is misleading.[30]

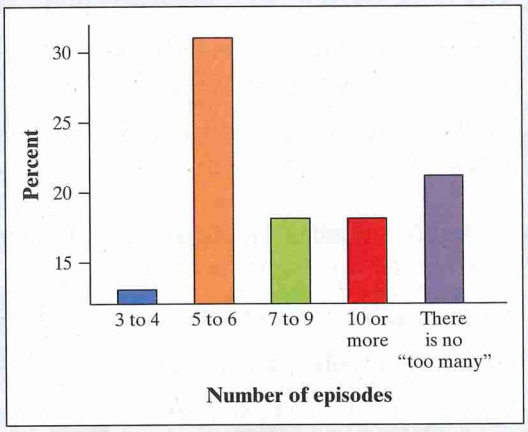
pg 20



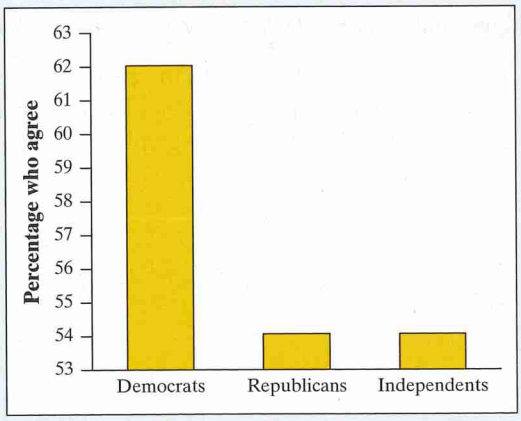
16. **Social media** The Pew Research Center surveyed a random sample of U.S. teens and adults about their use of social media. The following pictograph (not created by the Pew Research Center!) displays the percentage of people in various age groups who report using social media. Explain how this graph is misleading.



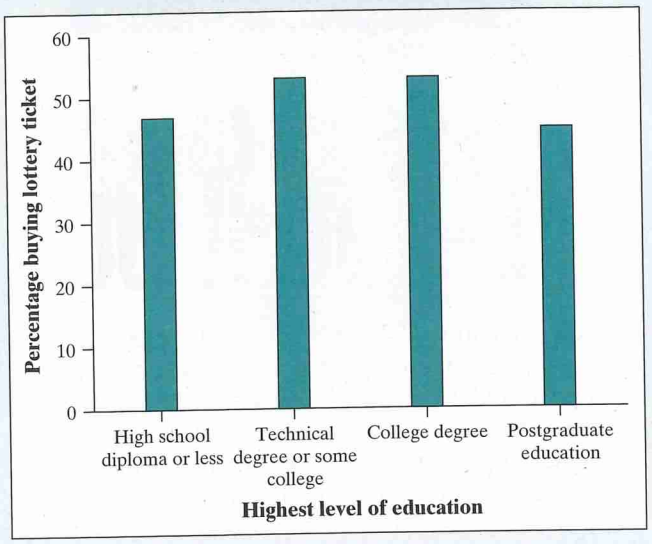
17. **Binge-watching** Do you "binge-watch" television series by viewing multiple episodes of a series at one sitting? A survey of 800 people who binge-watch were asked how many episodes is too many to watch in one viewing session. The results are displayed in the bar chart.[31] Explain how this graph is misleading.



18. **Support the court?** A news network reported the results of a survey about a controversial court decision. The network initially posted on its website a bar chart of the data similar to the one shown here. Explain how this graph is misleading. (*Note:* When notified about the misleading nature of its graph, the network posted a corrected version.)



19. **Lotteries and education** A Gallup Poll asked respondents about their highest level of education and whether they had bought a state lottery ticket in the last 12 months.[32] Here is a modified bar chart of the data that displays only the percentage of respondents who bought a lottery ticket in each category of highest educational level.



(a) Describe what this graph reveals about lottery ticket buying habits among the different education groups.

(b) Would it be appropriate to make a pie chart for this data set? Explain your answer.

Connecting the Statistical Practices *Multi-focus exercises that connect two or more statistical practices.*

20. **Teen versus tween reading** Refer to Exercise 14.

(a) Identify the populations and the samples.

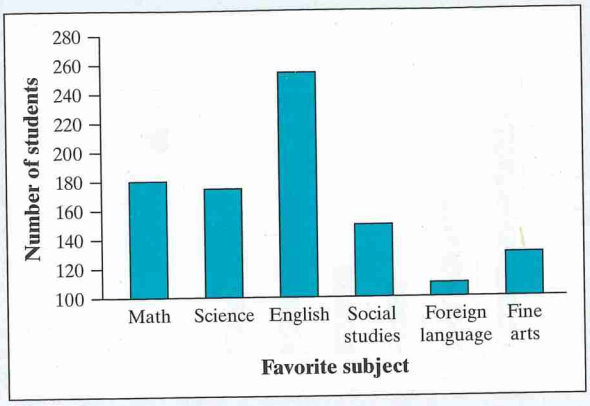
(b) Identify the variable of interest. Classify the variable as categorical or quantitative.

(c) Make a relative frequency table that compares the survey responses for teens and tweens.

(d) Formulate a possible investigative question for this statistical study.

Multiple Choice *Select the best answer for each question.*

21. The following bar chart shows the distribution of favorite subject for a sample of 1000 students. What is the most serious problem with the graph?



(A) The subjects are not listed in the correct order.

(B) This distribution should be displayed with a pie chart.

(C) The vertical axis should show the percentage of students.

(D) The vertical axis should start at 0 rather than 100.

22. For which of the following would it be *inappropriate* to display the data with a single pie chart?

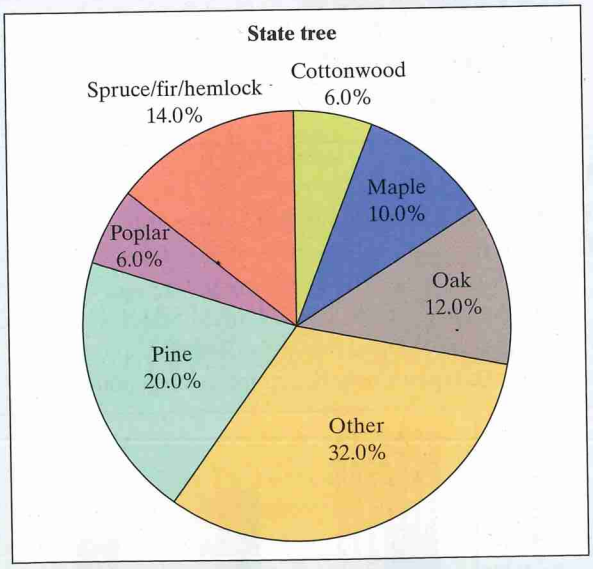
(A) The distribution of car color for vehicles purchased in the last month

(B) The distribution of unemployment percentage for each of the 50 states

(C) The distribution of favorite sport for a sample of 30 middle school students

(D) The distribution of presidential candidate preference for voters in a state

23. The pie chart displays the distribution of the designated state tree by type for the 50 U.S. states. The category "Other" includes all trees that are the state tree for two or fewer states.



Which of the following conclusions can we justify from this graph?

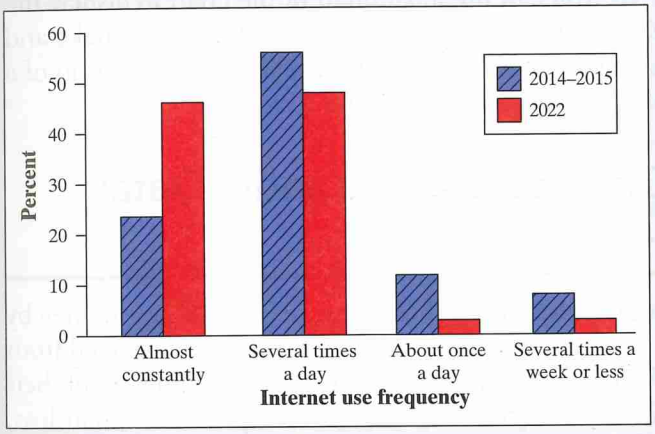
(A) The cottonwood is the state tree for 12 states.

(B) Taken together, oak, pine, and maple are the state trees for more than half the states.

(C) There are 10 states that have designated a pine as their state tree.

(D) There is no state that has designated the Eastern Red Cedar as its state tree.

24. How has teen internet use changed over time? The Pew Research Center surveyed separate random samples of U.S. teens aged 13 to 17 in 2014–2015 and in 2022. The side-by-side bar chart summarizes the teens' responses to the question, "About how often do you use the internet, either on a computer or a cellphone?"[33]



Which of the following is a correct statement?

(A) A majority of teens said "almost constantly" in 2022.

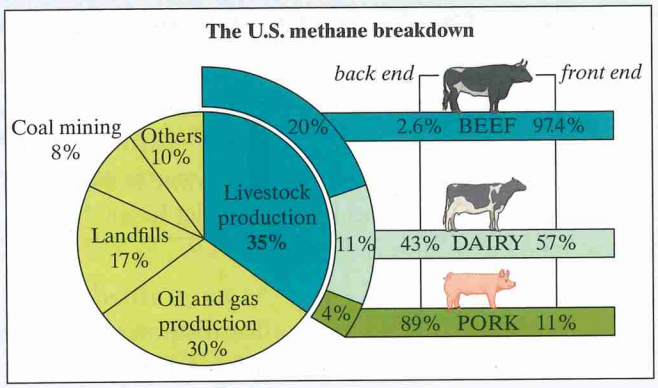
(B) The percentage of teens who said "almost constantly" approximately doubled from 2014–2015 to 2022.

(C) A majority of teens said "several times a day" in both 2014–2015 and 2022.

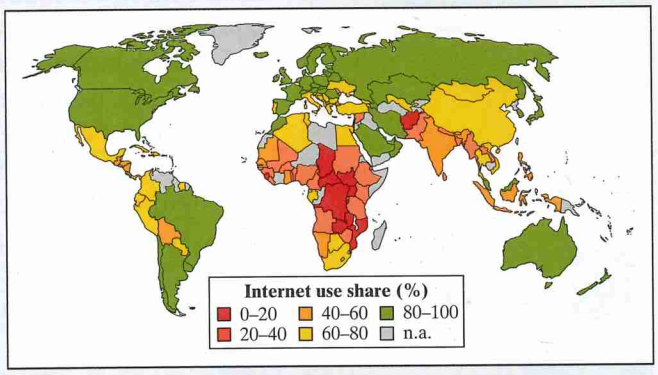
(D) The percentage of teens who said "about once a day" doubled from 2014–2015 to 2022.

For Investigation *Apply the skills from the section in a new context or nonroutine way.*

25. **Cow pie chart?** Methane is a greenhouse gas that may contribute to climate change. The modified pie chart shown here displays data on the sources of methane emissions in the United States.[34] Write a thorough analysis of what the graph shows. Be sure to include a comparison of methane emissions by beef cows, dairy cows, and pigs.



26. **Choropleth maps** A *choropleth map* is a graphical representation of data by geographic region in which values are depicted by color. For instance, the choropleth map presented here shows the percentage of people who were internet users in each country in 2020.[35] Write a thorough analysis of what the graph shows. Be sure to include a comparison of internet use in different regions of the world.



Unit 5 Project

Dissolved Oxygen in Lake Water

The concentration of dissolved oxygen (DO) in a lake is very important for the health of the lake and the organisms that depend on it. The file tps8u5project_data, which can be accessed on Achieve or obtained from your teacher, contains a random sample of 500 measurements of DO from water samples taken at various depths over the past 30 years in Lake Champlain. In addition to measurements of DO (mg/l) and depth (m), the data file includes the temperature of the water when the measurements were obtained (°C).[57] Download the file to a computer for further analysis using the application specified by your teacher.

1. Make a scatterplot to investigate the relationship between $x =$ temperature and $y =$ DO.
2. Calculate the equation of the least-squares regression line and make a residual plot. What does the residual plot indicate about the linear regression model?
3. Calculate and interpret the value of r^2.
4. Make a scatterplot to investigate the relationship between $x =$ depth and $y =$ DO.
5. Calculate the equation of the least-squares regression line and make a residual plot. What does the residual plot indicate about the linear regression model?
6. Calculate and interpret the value of r^2.
7. Compare the direction, form, and strength of the two associations.
8. Based on your previous answers, is temperature or depth a better predictor of DO? Explain your reasoning.

Extension: Is there evidence of a change in DO over time? Use the data to investigate.

See Yourself in Stats

With a solid grasp of statistics, you could become . . . an econometrician.

An econometrician is an individual who is trained in economics, statistics, and mathematics and uses those skills to study, model, and predict economic principles and outcomes. This specialized branch of economics is critical to the modern world and is in great demand to help governments and businesses. Many econometricians work for the Federal Reserve or in the financial services industry, where they are often called "quants." Relying heavily on least-squares regression analysis and other advanced statistical techniques, econometricians analyze data and current economic indicators to make forecasts and predictions that guide economic policy and future decisions.

To learn more about this fascinating field, check out **See Yourself in Stats** by clicking the link in your ebook, visiting the See Yourself in Stats spreadsheet on Achieve, or obtaining it from your teacher.

Cumulative AP® Practice Test 5

Find Cumulative AP® Practice Test 5 on Achieve or your teacher can share this content with you.

Solutions

UNIT 1, PART I

Section 1A Statistics: Learning from Data

Answers to Check Your Understanding

page 9: 1. *Sample*: the 50 randomly selected cars for which Malena recorded data. *Population*: all the cars in the student parking lot at this large high school

2. *Observational units*: the 50 cars for which Malena recorded data. Note: the license plates identify the observational units. *Variables*: model, year, number of stickers, color, weight, whether it has a navigation system, and highway gas mileage. *Categorical*: model, year, color, and whether it has a navigation system. *Quantitative*: number of stickers, weight (kg), and highway gas mileage (mpg). *Note:* Year might be considered quantitative if we want to know the average year.

3. Answers will vary. Sample answer: What are the plausible values for the mean number of stickers per car for the population of all cars in the student parking lot?

Answers to Odd-Numbered 1A Exercises

1. (a) Valid because whether a U.S. adult engages in vigorous exercise at least once a week varies from person to person. (b) Not valid because the number of days "you" engaged in vigorous exercise in the past week can be answered with a single data value that does not vary.

3. For all passengers like the volunteers in this study, does the Steffen boarding method result in a faster average boarding time for an airline flight than the traditional boarding method?

5. Answers will vary. Sample answer: Is there a difference in the proportions of all urban, suburban, and rural U.S. teens aged 13 to 17 who have ever experienced online bullying?

7. *Pop.:* the 1000 envelopes stuffed during this hour. *Sample:* the 40 randomly selected envelopes.

9. *Pop.:* all businesses in this large city. *Sample:* the 73 businesses that return the questionnaire.

11. *Pop.:* all 1000 dairy cows. *Par.:* the minimum daily milk production for the 1000 dairy cows. *Sample*: the 20 randomly chosen dairy cows. *Stat:* the minimum daily milk production for the 20 randomly chosen dairy cows = 6.3 gallons.

13. (a) *Obs. units*: students who completed the questionnaire. *Var.:* grade level, dominant hand, GPA, number of children in family, time spent on homework last night (min), type of phone. (b) *Cat.:* grade level, dominant hand, amount of time spent on homework last night (in an interval of minutes), type of phone. *Quant.:* GPA, number of children in family.

15. *Obs. units:* 10 movies. *Var.:* year, rating, time (min), genre, box office receipts ($). *Cat.:* year, rating, genre. *Quant.:* time (min), box office receipts ($). *Note:* Year might be considered quantitative if we want to know the average year.

17. (a) The wooden panels that were prepared and then exposed to the weather. (b) *Cat.:* type of wood, paint color. *Quant. (discrete):* number of blemishes, weathering time. *Quant. (continuous):* paint thickness

19. C

21. D

Section 1B Displaying and Describing Categorical Data

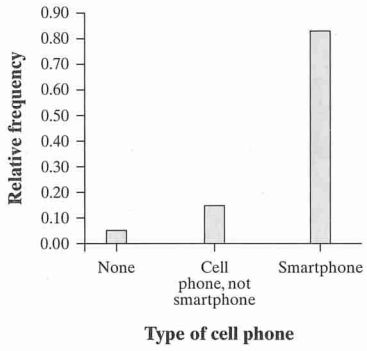
Answers to Check Your Understanding

page 21: 1.

Relative frequency table

Type of cell phone	Relative frequency
None	73/1502 = 0.049 or 4.9%
Cell phone, not smartphone	212/1502 = 0.141 or 14.1%
Smartphone	1217/1502 = 0.810 or 81.0%
Total	**1502/1502 = 1 or 100%**

2. Most of the adults (81.0%) who were surveyed own a smartphone. Only 14.1% of adults who were surveyed own a cell phone that is not a smart phone and only 4.9% of adults who were surveyed do not own a cell phone.



3. The graph provides some evidence that cell phone ownership differs across age groups. For the individuals in this sample, as age group increases: The percentage of individuals who own a smartphone decreases, the percentage of individuals who own a cell phone but not a smartphone increases, and the percentage of individuals who do not own a cell phone increases.

Answers to Odd-Numbered 1B Exercises

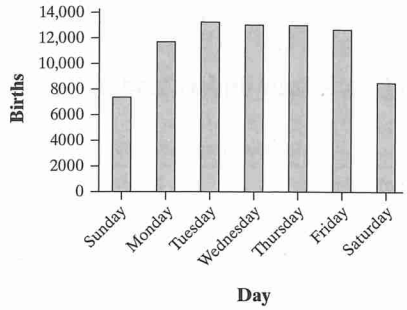
1. (a)

Frequency table		Relative frequency table	
Response	**Frequency**	**Response**	**Relative frequency**
Approach	17	Approach	17/75 = 0.227 or 22.7%
Run from	17	Run from	17/75 = 0.227 or 22.7%
Indifferent	41	Indifferent	41/75 = 0.547 or 54.7%
Total	**75**	**Total**	**75/75 = 1 or 100%**

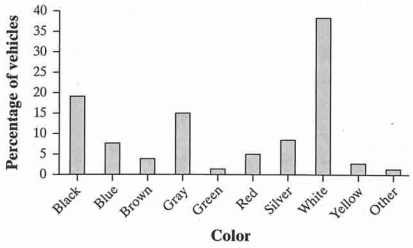
(b) The squirrels were most likely to be indifferent (54.7%) and were equally likely to approach (22.7%) or run from humans (22.7%).

3. (a) $75 \times 0.307 = 23.025$, so 23 students said their preferred status was rich. (b) Because happiness is the preferred status that was most often selected (proportion = 0.52) by the students in the sample. A campaign focused on future happiness may appeal to the greatest number of students.

5. It appears that births occurred with similar frequencies on weekdays, but with noticeably smaller frequencies on the weekend days.



7. (a) $100-19-7-3-15-1-5-9-38-2=1\%$. (b) The most popular color of vehicles sold that year was white, followed by black, gray, and silver. It appears that a majority of car buyers that year preferred vehicles that were shades of black or white.



(c) Yes, because the numbers in the table refer to all the parts of a single whole.

9. (a) Estimates may vary. About 58% Mexican and 10.5% Puerto Rican. (b) No, because the pie chart reveals that people of Cuban origin make up the smallest proportion of Hispanic people in the U.S.

11. Regardless of whether a student went to a private or public college, most students chose a school that was at least 11 miles from home. Those who went to a public university were most likely to choose a school that was 11 to 50 miles from home (about 30%), while those who went to a private university were most likely to choose a school that was 101 to 500 miles from home (about 29%).

13. On both continents, the most popular vehicle colors are white (North America, 30%; Asia, 40%), black (North America, 18%; Asia, 20%), and gray (North America, 20%; Asia, 15%). The most popular color that is NOT a shade of black or white is blue (North America, 10%; Asia, 5%). The biggest difference is that the preference for white is much greater in Asia (40%) than in North America (30%).

15. The areas of the flags should be proportional to the percentage of adults they represent. As drawn, it appears that adults from the United States are more than twice as likely to dislike shopping for clothes than adults in Germany and about 4 times as likely as adults in Spain — when, in fact, the percentages are approximately U.S. 30%, Germany 20%, and Spain 12%.

17. By starting the vertical scale at about 12 instead of 0, it looks like the percentage of binge-watchers who think that 5 to 6 episodes is too many to watch in one viewing session is almost 20 times higher than the percentage of binge-watchers who think that 3 to 4 episodes is too many to watch in one viewing session. In truth, the percentage is less than 3 times higher (31% to 13%). Similar arguments can be made for the relative sizes of the other categories represented in the bar chart.

19. (a) The graph reveals that those who have a technical degree or some college and those who have a college degree are more likely to have purchased a lottery ticket (just over 50%) than are those who have a high school diploma or less (about 47%) or those who have a postgraduate education (about 45%). (b) No, because the data do not represent parts of a whole. The sum of all these percentages is greater than 100%. This is because we have the percentage of each education group that has bought a lottery ticket, rather than the percentage of all lottery ticket buyers who are in each education group.

21. D

23. C

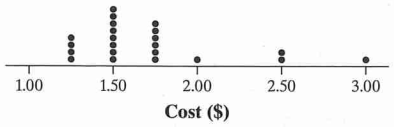
25. The pie chart shows that 35% of methane emissions come from livestock production, 30% come from oil and gas production, 17% come from landfills, 8% come from coal mining, and 10% come from other sources. The graph further breaks down the sources of the methane emissions from livestock production: 20% come from beef cows, 11% from dairy cows, and 4% from pigs. The graph also provides the distribution of methane emission from the front and back end of beef cows, dairy cows, and pork. It is interesting to note that beef cows produce only 2.6% of their methane emissions out the back end, whereas dairy cows produce 43% of their emissions out the back end, and pigs produce 89% of their emissions out the back end. Life lesson: Don't stand downwind of the back of a pig. Also keep in mind the back end of a beef cow isn't nearly as toxic as you might think.

Section 1C Displaying and Describing Quantitative Data with Graphs

Answers to Check Your Understanding

page 34:

1.



2. The distribution of cost for rides at Knoebels Amusement Park is skewed to the right with a single peak at \$1.50. There are two small gaps at \$2.25 and \$2.75. The median cost is \$1.50, and the cost varies from \$1.25 to \$3.

3. Yes, the graph supports the claim that a family of 4 with a \$100 budget can enjoy at least 12 different rides together. The least expensive 12 rides would cost one person $4(1.25)+8(1.50)=\$17$. Therefore, a family of 4 would spend $4(17)=\$68$ on these 12 rides. The family can ride 12 different rides and still have \$32 left over for ice cream!

page 42: 1. 20.6% of Maine's residents are aged 65 and older.

2. *Shape*: The distribution of percentage of residents aged 65 and older for the 50 states is roughly symmetric with a single peak in the 15.0% to 15.9% stem. *Outliers*: There are no apparent outliers. *Center*: The median is $\frac{16.4+16.5}{2}=16.45\%$ of residents aged 65 and older. *Variability*: The percentage of residents aged 65 and older in a state varies from 11.1% to 20.6%.