AP STAT
SUMMER PACKET

NEED GRAPHING CALC FOR THIS CLASS
TI-84 PLUS IS FINE

Dear AP STAT Student,
How exciting to be taking AP STAT! This is your summer packet. There are two parts. Part 1 is your intro to Statistics. Part 2 contains probability topics that you may or may not have learned. So it is either review for you or it is new learning. This will help when we hit the topics involving probability. In order to cover all topics necessary for the AP test in May, the attached unit P must be completed ahead of time as an introduction to Statistics. More discussions and activities will happen in the first couple weeks as related to this unit. You must come prepared to discuss.

Read, take notes, and complete the questions that are *circled*.
You must submit the circled problems according to the timeline below.
You can find a copy of this on my website: http://sreilly.rsd17.org.

There will be a unit P assessment within two days of the beginning of school (Friday 8/31). All your summer work will be collected on that day (8/31) as well for a notebook/HW grade

- I will send you links to google forms to submit your homework answers.

*You must also sign up for my REMIND text. * This is how:
TO: 81010
In message bar: @2kd443 hit send
Input ***name and period*** when it asks for it.

I will use this to send links and other important info over the summer- including the link to submit your forms for HW answers.

Over the summer, you can also contact me here: sheilareilly@rsd17.org

Due Dates serve to *pace you*. Something will be due about every two weeks. These will be put into Power School as grades.
Enjoy your summer and I look forward to meeting you all in August!

DUE DATES

Due July 8th: *Part 1*, read/notes and turn in circled problems pg 11,19

Due July 22nd: *Part 1*, read/notes and turn in circled problems pgs 31,7,47,48

Due August 5th: *Part 2*, circled problems pgs 705,706,712,713

Due August 19th: *Part 2*, circled problems pgs 719,727-728,734-735,742

BE READY TO BE ASSESSED ON PART 1 CONTENT on 8/31/19

<u>Part 1:</u>

Intro to Stats

8. Do you believe that the students in your class can actually tell the difference between tap water and bottled water? Before you answer, let's perform a brief *simulation*. We'll assume that you and your classmates are guessing at which cup holds the bottled water. Then each student would have a 1-in-3 chance of identifying the correct cup. Roll your die once for each student in your class. Let rolling a 1 or a 2 represent a correct guess. Let rolling a 3 through 6 represent an incorrect guess. (Note that this assignment of numbers gives each individual a 2-in-6 chance of being correct, which is the same as a 1-in-3 chance.) Record the number of times that you get a 1 or a 2. This result simulates the number of correct identifications made by the class.

9. On a number line drawn on the board by your teacher, mark an X above the number of correct identifications in your simulation. Based on the class's simulation results, how many correct identifications would make you doubt that students were just guessing? Why?

10. Look back to your class's actual tasting results in Step 6. What do you conclude about students' abilities to distinguish tap water from bottled water?

Introduction

Do cell phones cause brain cancer? How well do SAT scores predict college success? Should arthritis sufferers take Celebrex to ease their pain, or are the risks too great? What percent of U.S. children are overweight? How strong is the evidence for global warming? These are just a few of the questions that statistics can help answer. But what is statistics? And why should you study it?

statistics

Statistics is the science (and art) of learning from data. Data are usually numbers, but they are not "just numbers." *Data are numbers with a context.* The number 10.5, for example, carries no information by itself. But if we hear that a friend's new baby weighed 10.5 pounds at birth, we congratulate her on the healthy size of the child. The context engages our background knowledge and allows us to make judgments. We know that a baby weighing 10.5 pounds is quite large, and that a human baby is unlikely to weigh 10.5 ounces or 10.5 kilograms. The context makes the number informative.

You can find lots of data in newspapers and magazines and on the Internet. Such data are ripe for exploration. Part I of this book focuses on *exploratory data analysis*. In Chapters 1 through 4, you'll develop tools and strategies for organizing, describing, and analyzing data.

Sometimes data provide insights about questions we have asked. More often, researchers follow a careful plan for *producing data* to answer specific questions. In Part II of the book, you'll discover how they do it. Chapter 5 shows you how to design *surveys*, *experiments*, and *observational studies* correctly. Such well-produced data help us get the most reliable answers to difficult questions.

Probability is the study of chance behavior. When you flip coins, roll dice, deal cards, or play the lottery, the results are uncertain. But the laws of probability can tell you how likely (or unlikely) certain outcomes are. You'll learn how to calculate probabilities in Part III of the book, in Chapters 6 through 9.

population
sample

How can we draw conclusions about a large group (***population***) of *individuals*—people, animals, or things—based on information about a much smaller group (***sample***)? This is the challenge of *statistical inference*. Inference questions ask us to test claims about or provide estimates for unknown population values. Valid inference depends on appropriate data production, skillful data analysis, and careful use of probability. In Part IV of the book, we discuss the logic behind statistical inference and some of its methods. In Chapters 10 through 15, you'll learn some common ways of testing claims and computing estimates.

This Preliminary Chapter is intended to give you a snapshot of what statistics is all about. Where do data come from? What should you do with data once you have them? How can probability help you? What conclusions can you draw? Keep reading for some answers.

In your lifetime, you will be bombarded with data and statistical information. Opinion poll results, television ratings, gas prices, unemployment rates, medical study outcomes, and standardized test scores are discussed daily in the media. People make important decisions based on such data. Statistics will help you make sense of information like this. A solid understanding of statistics will enable you to make sound decisions based on data in your everyday life.

Data Production: Where Do You Get Good Data?

You want data on a question that interests you. Maybe you want to know what causes of death are most common among young adults, or whether the math performance of American schoolchildren is getting better.

It is tempting just to draw conclusions from our own experience, making no use of more representative data. You think (without really thinking) that the students at your school are typical. We hear a lot about AIDS, so we assume it must be a leading cause of death among young people. Or we recall an unusual incident that sticks in our memory exactly because it is unusual. When an airplane crash kills several hundred people, we fear that flying is unsafe, even though data on all flights show that flying is much safer than driving. Here's an example that shows why **data beat personal experiences.**

Example P.1

Power lines and cancer

Got data?

Does living near power lines cause leukemia in children? The National Cancer Institute spent 5 years and $5 million gathering data on this question. The researchers compared 638 children who had leukemia with 620 who did not. They went into the homes and actually measured the magnetic fields in children's bedrooms, in other rooms, and at the front door. They recorded facts about power lines near the family home and also near the mother's residence when she was pregnant. Result: no connection between leukemia and exposure to magnetic fields of the kind produced by power lines was found. The editorial that accompanied the study report in the *New England Journal of Medicine* proclaimed, "It is time to stop wasting our research resources" on the question.[2]

Now consider a devastated mother whose child has leukemia and who happens to live near a power line. In the public mind, the striking story wins every time. A statistically literate person, however, knows that data are more reliable than personal experience because they systematically describe an overall picture rather than focus on a few incidents.

A better tactic is to head for the library or the Internet. There you will find plenty of data, not gathered specifically to answer your questions but available for your use. Recent data can be found online, but locating them can be challenging. Government statistical offices are the primary source for demographic, social, and economic data. Many nations have a single statistical office, like Statistics Canada (www.statcan.ca) or Mexico's INEGI (www.inegi.gob.mx). The United States does not have a national statistical office. More than 70 federal agencies collect data. Fortunately, you can reach most of them through the government's handy FedStats site (www.fedstats.gov).

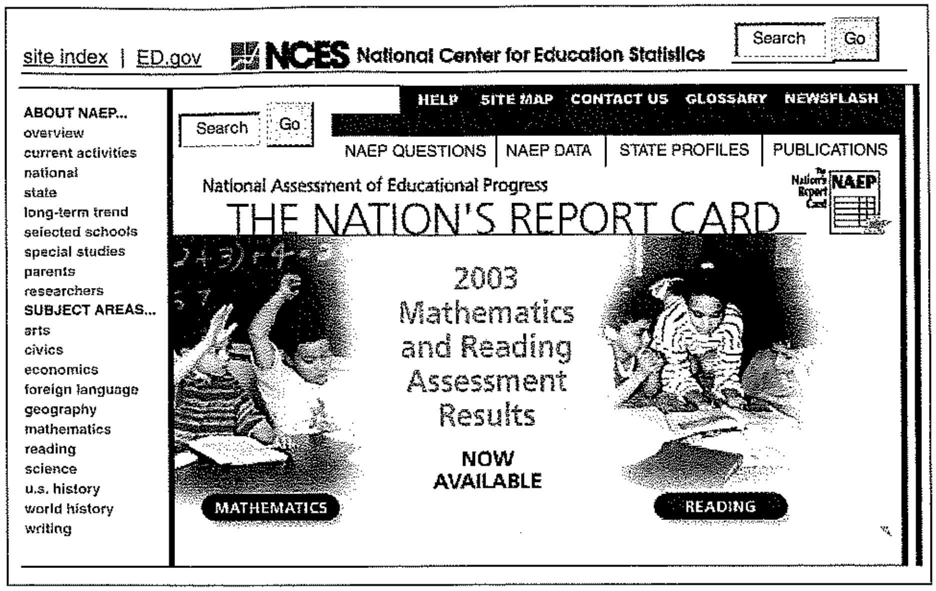
Example P.2

Causes of death and math scores

Finding data on the Internet

If you visit the National Center for Health Statistics Web site, www.cdc.gov/nchs, you will learn that accidents are the most common cause of death for U.S. citizens aged 20 to 24, accounting for over 40% of all deaths. Homicide is next, followed by suicide. AIDS ranks seventh, behind heart disease and cancer, at 1% of all deaths. The data also show that it is dangerous to be a young man: the overall death rate for men aged 20 to 24 is three times that for women, and the death rate from homicide is more than five times higher among men.

Figure P.1 *The Web sites of government statistical offices are prime sources of data. Here is the home page of the National Assessment of Educational Progress.*



If you go to the National Center for Education Statistics Web site, www.nces.ed.gov, you will find the latest National Assessment of Educational Progress (Figure P.1), which provides full details about the math skills of schoolchildren. Math scores have slowly but steadily increased since 1990. All racial/ethnic groups, both girls and boys, and students in most states are getting better in math.

The library and the Internet are sources of ***available data.***

Available Data

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

Available data are the only data used in most student reports. Because producing new data is expensive, we all use available data whenever possible. However, the clearest answers to present questions often require data produced to answer those specific questions. The main statistical designs for producing data are *surveys*, *experiments*, and *observational studies*.

surveys ***Surveys*** are popular ways to gauge public opinion. The idea of a survey is pretty simple:

- Select a *sample* of people to represent a larger *population.*

- Ask the individuals in the sample some questions and record their responses.
- Use sample results to draw some conclusions about the population.

In practice, however, getting valid survey results is not so easy. As the following example shows, **where the data come from is important.**

Example P.3 *Having kids or not?*
Good and bad survey results

The advice columnist Ann Landers once asked her readers, "If you had it to do over again, would you have children?" A few weeks later, her column was headlined "70% OF PARENTS SAY KIDS NOT WORTH IT." Indeed, 70% of the nearly 10,000 parents who wrote in said they would not have children if they could make the choice again. Do you believe that 70% of all parents regret having children?

You shouldn't. The people who took the trouble to write Ann Landers are not representative of all parents. Their letters showed that many of them were angry at their children. All we know from these data is that there are some unhappy parents out there. A statistically designed poll, unlike Ann Landers's appeal, targets specific people chosen in a way that gives all parents the same chance to be asked. Such a poll later showed that 91% of parents would have children again.

The lesson: if you are careless about how you get your data, you may announce 70% "No" when the truth is close to 90% "Yes."

census

You may have wondered: why not survey everyone in the population (a *census*) rather than a sample? Usually, it would take too long and cost too much. Our goal in choosing a sample is a picture of the population, disturbed as little as possible by the act of gathering information. Sample surveys are one kind of ***observational study.***

In other settings, we gather data from an ***experiment.*** In doing an experiment, we don't just observe individuals or ask them questions. We actually do something to people, animals, or objects to observe the response. Experiments can answer questions such as "Does aspirin reduce the chance of a heart attack?" and "Do more college students prefer Pepsi to Coke when they taste both without knowing which they are drinking?" Experiments, like samples, provide useful data only when properly designed. The distinction between experiments and observational studies is one of the most important ideas in statistics.

Observational Study versus Experiment

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

In an **experiment,** we deliberately do something to individuals in order to observe their responses.

The next example illustrates the difference between an observational study and an experiment.

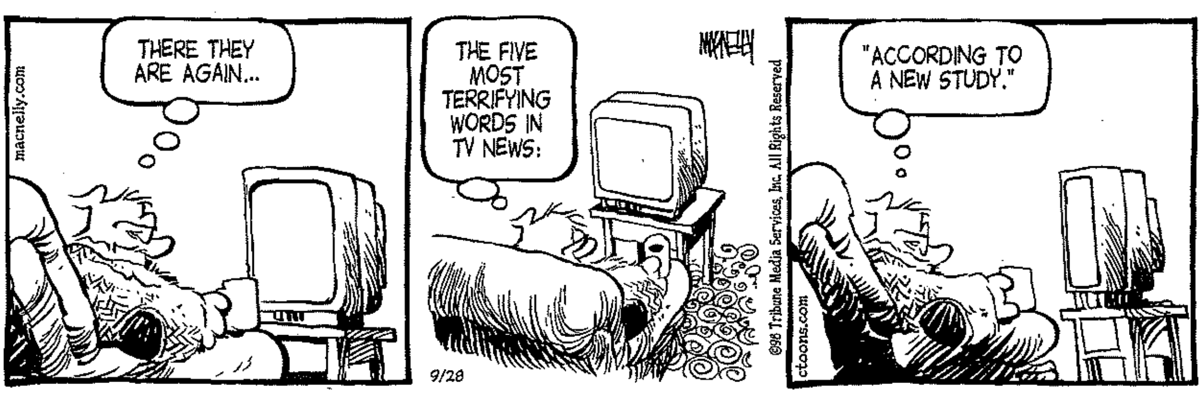
Example P.4 *Estrogen and heart attacks*
Observational study versus experiment

Should women take hormones such as estrogen after menopause, when natural production of these hormones ends? In 1992, several major medical organizations said "Yes." Women who took hormones seemed to reduce their risk of a heart attack by 35% to 50%. The risks of taking hormones appeared small compared with the benefits.

The evidence in favor of hormone replacement came from a number of studies that simply compared women who were taking hormones with others who were not. But women who chose to take hormones were typically richer and better educated, and they saw doctors more often than women who did not take hormones. These women did many things to maintain their health. It isn't surprising that they had fewer heart attacks.

Experiments were needed to get convincing data on the link between hormone replacement and heart attacks. In the experiments, women did not decide what to do. A coin toss assigned each woman to one of two groups. One group took hormone replacement pills; the other took dummy pills that looked and tasted the same as the hormone pills. All kinds of women were equally likely to get either treatment. By 2002, several experiments with women of different ages showed that hormone replacement does *not* reduce the risk of heart attacks. The National Institutes of Health, after reviewing the evidence, concluded that the earlier observational studies were wrong. Taking hormones after menopause fell quickly out of favor.[3]

Observational studies are essential sources of data about topics from the opinions of voters to the behavior of animals in the wild. But an observational study, even one based on a statistical sample, is a poor way to gauge the effect of a change. To see the response to a change, we must actually impose the change. When our goal is to understand cause and effect, experiments are the best source of convincing data.



Exercises

P.1 Need a Jolt? Jamie is a hard-core computer programmer. She and all her friends prefer Jolt cola (caffeine equivalent to two cups of coffee) to either Coke or Pepsi (caffeine equivalent to less than one cup of coffee). Explain why Jamie's preference is not good evidence that most young people prefer Jolt to Coke or Pepsi.

P.2 Cell phones and brain cancer One study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same age, sex, and race who did not have brain cancer, then asked about the use of cell phones.[4] Result: "Our data suggest that the use of handheld cellular phones is not associated with risk of brain cancer."

(a) Is this an observational study or an experiment? Justify your answer.

(b) Based on this study, would you conclude that cell phones do not increase the risk of brain cancer? Why or why not?

P.3 Learning biology with computers An educational software company wants to compare the effectiveness of its computer animation for teaching biology with that of a textbook presentation. The company gives a biology pretest to each of a group of high school juniors, then divides them into two groups. One group uses the animation, and the other studies the text. The company retests all students and compares the increase in biology test scores in the two groups.

(a) Is this an experiment or an observational study? Justify your answer.

(b) If the group using the computer animation has a much higher average increase in test scores than the group using the textbook, what conclusions, if any, could the company draw?

P.4 Survey, experiment, or observational study? What is the best way to answer each of the questions below: a survey, an experiment, or an observational study that is not a survey? Explain your choices. For each question, write a few sentences about how such a study might be carried out.

(a) Are people generally satisfied with how things are going in the country right now?

(b) Do college students learn basic accounting better in a classroom or using an online course?

(c) How long do your teachers wait on the average after they ask the class a question?

P.5 I'll drink to that! In adults, moderate use of alcohol is associated with better health. Some studies suggest that drinking wine rather than beer or spirits yields added health benefits.

(a) Explain the difference between an observational study and an experiment to compare people who drink wine with people who drink beer.

(b) Suggest some characteristics of wine drinkers that might benefit their health. In an observational study, these characteristics are mixed up with the effects of drinking wine on people's health.

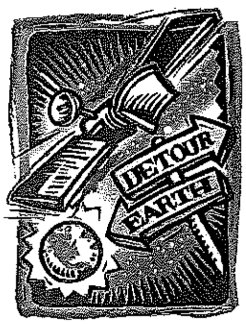
P.6 Get a job! Find some information on this question: what percent of college undergraduates work part-time or full-time while they are taking classes? Start with the National Center for Education Statistics Web site, www.nces.ed.gov. Keep a detailed written record of your search.

Data Analysis: Making Sense of Data

data analysis

The first step in understanding data is to hear what the data say, to "let the statistics speak for themselves." But numbers speak clearly only when we help them speak by organizing, displaying, summarizing, and asking questions. That's ***data analysis.***

Any set of data contains information about some group of ***individuals.*** The characteristics we measure on each individual are called ***variables.***



The importance of data integrity It has been accepted that global warming is a serious ecological problem. But Yale University researchers examined satellite and weather-balloon data collected since 1979 by NOAA (National Oceanic and Atmospheric Administration). They discovered that the satellites had drifted in orbit, throwing off the timing of temperature measurements. Nights looked as warm as days. Corrective action has shown that the pace of global warming over the past 30 years has actually been quite slow, a total increase of about 1 degree Fahrenheit. The lesson: always ask, "How were the data produced?"

Individuals and Variables

Individuals are the objects described by a set of data. Individuals may be people, but they may also be animals or things.

A **variable** is any characteristic of an individual. A variable can take different values for different individuals.

A college's student data base, for example, includes data about every currently enrolled student. The students are the *individuals* described by the data set. For each individual, the data contain the values of *variables* such as age, gender, choice of major, and grade point average. In practice, any set of data is accompanied by background information that helps us understand it.

When you meet a new set of data, ask yourself the following *key questions:*

1. **Who** are the individuals described by the data? How many individuals are there?
2. **What** are the variables? In what units is each variable recorded? Weights, for example, might be recorded in pounds, in thousands of pounds, or in kilograms.
3. **Why** were the data gathered? Do we hope to answer some specific questions? Do we want to draw conclusions about individuals other than the ones we actually have data for?
4. **When, where, how, and by whom** were the data produced? Where did the data come from? Are these available data or new data produced to answer current questions? Are the data from an experiment or an observational study? From a census or a sample? Who directed the data production? Can we trust the data?

Some variables, like gender and college major, simply place individuals into categories. Others, like age and grade point average (GPA), take numerical values for which we can do arithmetic. It makes sense to give an average GPA for a group of students, but it does not make sense to give an "average" gender. We can, however, count the numbers of female and male students and do arithmetic with these counts.

Categorical and Quantitative Variables

A **categorical variable** places an individual into one of several groups or categories.

A **quantitative variable** takes numerical values for which arithmetic operations such as adding and averaging make sense.

Example P.5

Education in the United States

Four key questions

Here is a small part of a data set that describes public education in the United States:

State	Region	Population (1000s)	SAT verbal	SAT math	Percent taking	Percent no HS	Teachers' pay ($1000)
CA	PAC	35,894	499	519	54	18.9	54.3
CO	MTN	4,601	551	553	27	11.3	40.7
CT	NE	3,504	512	514	84	12.5	53.6

Answer the four key questions about these data.

1. **Who?** The *individuals* described are the states. There are 51 of them, the 50 states and the District of Columbia, but we give data for only 3: California (CA), Colorado (CO), and Connecticut (CT). Each row in the table describes one individual.

2. **What?** The rest of the columns each contain the values of one variable for all the individuals. This is the usual arrangement in data tables. Seven *variables* are recorded for each state. The second column lists which region of the country the state is in. Region is a categorical variable. The Census Bureau divides the nation into nine regions. These three are Pacific (PAC), Mountain (MTN), and New England (NE). The third column contains state populations, in thousands of people. Population is a quantitative variable. Be sure to notice that the *units* are thousands of people. California's 35,894 stands for 35,894,000 people.

The remaining five variables are the average scores of the states' high school seniors on the SAT verbal and mathematics exams, the percent of seniors who take the SAT, the percent of students who did not complete high school, and average teachers' salaries in thousands of dollars. These are all quantitative variables. Each of these variables needs more explanation before we can fully understand the data.

Beginning in March 2005, the new SAT consisted of three tests: Critical Reading, Math, and Writing.

3. **Why?** Some people will use these data to evaluate the quality of individual states' educational programs. Others may compare states using one or more of the variables. Future teachers might want to know how much they can expect to earn.

4. **When, where, how, and by whom?** The population data come from the Current Population Survey, conducted by the federal government. They are fairly accurate as of July 1, 2004, but don't show later changes in population. State SAT averages came from the College Board's Web site, www.collegeboard.com, and were based on a census of all test takers that year. The percent of students who did not graduate in each state was determined by the 2003 Current Population Survey. Average teacher salaries were reported in the 2003 *Statistical Abstract of the United States*, using data provided by the National Education Association for 2002. These data are estimates based on samples of teachers from each state.

A variable generally takes values that vary (hence the name "variable"!). Categorical variables sometimes have similar counts in each category and sometimes don't. For example, if you recorded values of the variable "birth month" for the students at your school, you would expect about an equal number of students in each of the categories (January, February, March, . . .). If you measured the variable "favorite type of music," however, you might see very different counts in the categories classical, gospel, rock, rap, and so on. Quantitative variables may take values that are very close together or values that are quite spread out. We call the pattern of variation of a variable its ***distribution.***

Distribution

The **distribution** of a variable tells us what values the variable takes and how often it takes these values.

exploratory data analysis

Statistical tools and ideas can help you examine data in order to describe their main features. This examination is sometimes called ***exploratory data analysis.*** (We prefer data analysis.) Like an explorer crossing unknown lands, we first simply describe what we see. Each example we meet will have some background information to help us, but our emphasis is on examining the data. Here are two basic strategies that help us organize our exploration of a set of data:

- Begin by examining each variable by itself. Then move on to study relationships among the variables.
- Begin with a graph or graphs. Then add numerical summaries of specific aspects of the data.

We will organize our learning the same way. Chapters 1 and 2 examine single-variable data, and Chapters 3 and 4 look at relationships among variables. In both settings, we begin with graphs and then move on to numerical summaries.

Describing Categorical Variables

The values of a categorical variable are labels for the categories, such as "male" and "female." The distribution of a categorical variable lists the categories and gives either the *count* or the *percent* of individuals who fall in each category.

Example P.6 *Do you wear your seat belt?*
Describing categorical variables

Each year, the National Highway and Traffic Safety Administration (NHTSA) conducts an observational study on seat belt use. The table below shows the percent of front-seat passengers who were observed to be wearing their seat belts in each region of the United States in 1998 and 2003.[5]

Region	Percent wearing seat belts, 2003	Percent wearing seat belts, 1998
Northeast	74	66.4
Midwest	75	63.6
South	80	78.9
West	84	80.8

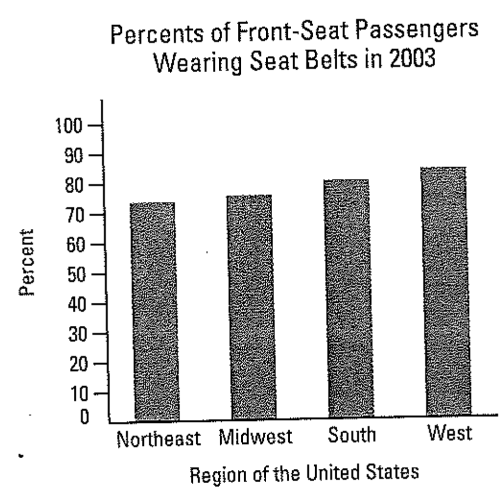
What do these data tell us about seat belt usage by front-seat passengers?

The *individuals* in this observational study are front-seat passengers. For each individual, the values of two *variables* are recorded: region (Northeast, Midwest, South, or West) and seat belt use (yes or no). Both of these variables are categorical.

bar graph

Figure P.2(a) shows a ***bar graph*** for the 2003 data. Notice that the vertical scale is measured in percents.

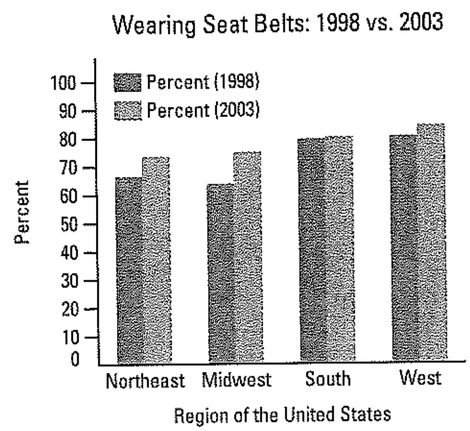
Figure P.2a (a) *A bar graph showing the percent of front-seat passengers who wore their seat belts in each of four U.S. regions in 2003.*



Front seat passengers in the South and West seem more concerned about wearing seat belts than those in the Northeast and Midwest. In all four regions, a high percent of front-seat passengers were wearing seat belts. Figure P.2(b) (on the next page) shows a ***side-by-side bar graph*** comparing seat belt usage in 1998 and 2003. Seat belt usage increased in all four regions over the five-year period.

side-by-side bar graph

Figure P.2b *(b) A side-by-side bar graph comparing the percent of front-seat passengers who wore their seat belts in the four U.S. regions in 1998 and 2003.*



Describing Quantitative Variables

dotplot

Several types of graphs can be used to display quantitative data. One of the simplest to construct is a ***dotplot.***

Example P.7 *GOOOOAAAAALLLLLL!*
Describing quantitative variables

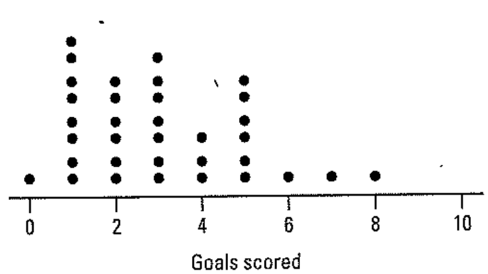
The number of goals scored by the U.S. women's soccer team in 34 games played during the 2004 season is shown below:[6]

3 0 2 7 8 2 4 3 5 1 1 4 5 3 1 1 3 3 3 2 1
2 2 2 4 3 5 6 1 5 5 1 1 5

What do these data tell us about the performance of the U.S. women's team in 2004?

A *dotplot* of the data is shown in Figure P.3. Each dot represents the goals scored in a single game. From this graph, we can see that the team scored between 0 and 8 goals per game. Most of the time, they scored between 1 and 5 goals. Their most frequent number of goals scored (the ***mode***) was 1. They averaged 3.059 goals per game. (Check our calculation of the ***mean*** on your calculator.)

Figure P.3 *A dotplot of goals scored by the U.S. women's soccer team in 2004.*



Making a statistical graph is not an end in itself. After all, a computer or graphing calculator can make graphs faster than we can. The purpose of a graph is to help us understand the data. After you (or your calculator) make a graph, always ask, "What do I see?"

Exploring Relationships between Variables

Quite often in statistics, we are interested in examining the relationship between two variables. For instance, we may want to know how the percent of students taking the SAT in U.S. states is related to those states' average SAT math scores, or perhaps how seat belt usage is related to region of the country. As the next example illustrates, **many relationships between two variables are influenced by other variables lurking in the background.**

CAUTION

Example P.8 *On-time flights*
Describing relationships between variables

Air travelers would like their flights to arrive on time. Airlines collect data about on-time arrivals and report them to the Department of Transportation. Here are one month's data for flights from several western cities for two airlines:

	On time	Delayed
Alaska Airlines	3274	501
America West	6438	787

You can see that the percents of late flights were

$$\text{Alaska Airlines} \quad \frac{501}{3775} = 13.3\%$$

$$\text{America West} \quad \frac{787}{7225} = 10.9\%$$

It appears that America West does better.

This isn't the whole story, however. For each flight (individual), we have data on two categorical variables: the airline and whether or not the flight was late. Let's add data on a third categorical variable, departure city.[7] The following table summarizes the results.

	Alaska Airlines		America West	
Departure city	**On time**	**Delayed**	**On time**	**Delayed**
Los Angeles	497	62	694	117
Phoenix	221	12	4840	415
San Diego	212	20	383	65
San Francisco	503	102	320	129
Seattle	1841	305	201	61
Total	3274	501	6438	787

The "Total" row shows that the new table describes the same flights as the earlier table. Look again at the percents of late flights, first for Los Angeles:

$$\text{Alaska Airlines} \quad \frac{62}{559} = 11.1\%$$

$$\text{America West} \quad \frac{117}{811} = 14.4\%$$

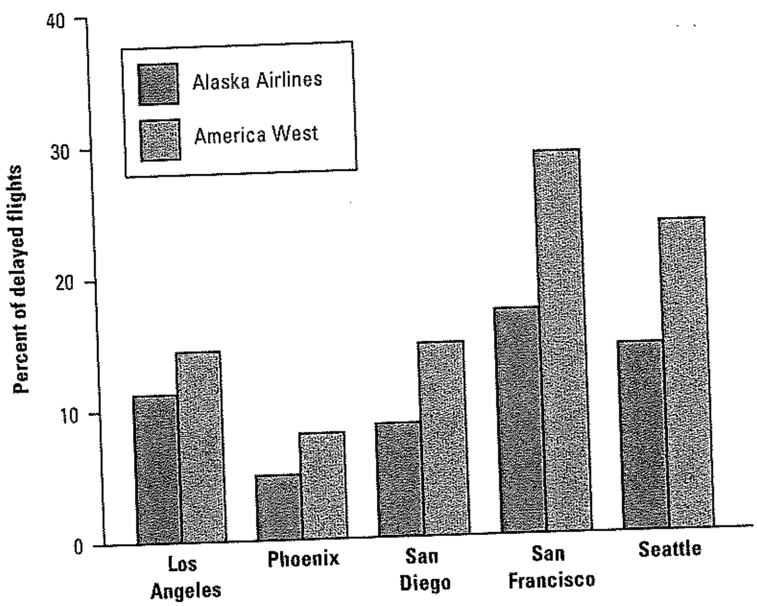
Alaska Airlines is better. The percents of late flights for Phoenix are

$$\text{Alaska Airlines} \quad \frac{12}{233} = 5.2\%$$

$$\text{America West} \quad \frac{415}{5255} = 7.9\%$$

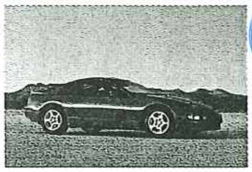
Alaska Airlines is better again. In fact, as Figure P.4 shows, Alaska Airlines has a lower percent of late flights at every one of these cities.

Figure P.4 *Comparing the percents of delayed flights for two airlines at five airports.*



How can it happen that Alaska Airlines wins at every city but America West wins when we combine all the cities? Look at the data: America West flies most often from sunny Phoenix, where there are few delays. Alaska Airlines flies most often from Seattle, where fog and rain cause frequent delays. What city we fly from has a major influence on the chance of a delay, so including the city data reverses our conclusion. (We'll see other examples like this one in Chapter 4 when we examine *Simpson's paradox.*) The message is worth repeating: many relationships between two variables (like airline and whether the flight was late) are influenced by other variables lurking in the background (like departure city).

Exercises



P.7 Cool car colors Here are data on the most popular car colors for vehicles made in North America during the 2003 model year.[8]

Color	Percent of vehicles
Silver	20.1
White	18.4
Black	11.6
Medium/dark gray	11.5
Light brown	8.8
Medium/dark blue	8.5
Medium red	6.9

(a) Display these data in a bar graph. Be sure to label your axes and title your graph.

(b) Describe what you see in a few sentences. What percent of vehicles had other colors?

P.8 Comparing car colors Favorite vehicle colors may differ among types of vehicle. Here are data on the most popular colors in 2003 for luxury cars and SUVs, trucks, and vans. The entry "—" means "less than 1%."

Color	Luxury car percent	SUV/truck/van percent
Black	10.9	11.6
Light brown	—	6.3
Medium/dark blue	3.8	9.3
Medium/dark gray	23.3	8.8
Medium/dark green	—	7.0
Medium red	3.9	6.2
White	30.4	22.3
Silver	18.8	17.0

(a) Make a side-by-side bar graph to compare colors by vehicle type.

(b) Write a few sentences describing what you see.

P.9 U.S. women's soccer scores In Example P.7 (page 16), we examined the number of goals scored by the U.S. women's soccer team in games during the 2004 season. Here are data on the goal differential for those same games, computed as U.S. score minus opponent's score.

3 0 2 7 8 2 4 1 4 1 −2 3 4 3 0 1 2 2 3 2 0
1 1 1 1 3 5 6 1 4 5 0 −2 5

(a) Make a dotplot of these data.

(b) Describe what you see in a few sentences.

P.10 Olympic gold! Olympic athletes like Michael Phelps, Natalie Coughlin, Amanda Beard, and Paul Hamm captured public attention by winning gold medals in the 2004 (a)