PART I

REVIEW OF PART I

Exploring and Understanding Data

Quick Review

It's time to put it all together. Real data don't come tagged with instructions for use. So let's step back and look at how the key concepts and skills we've seen work together. This brief list and the review exercises that follow should help you check your understanding of Statistics so far.

- We treat data two ways: as categorical and as quantitative.
- To describe categorical data:
  - Make a picture. Bar graphs work well for comparing counts in categories.
  - Summarize the distribution with a table of counts or relative frequencies (percents) in each category.
  - Pie charts and segmented bar charts display divisions of a whole.
  - Compare distributions with plots side by side.
  - Look for associations between variables by comparing marginal and conditional distributions.
- To describe quantitative data:
  - Make a picture. Use histograms, boxplots, stem-and-leaf displays, or dotplots. Stem-and-leafs are great when working by hand and good for small data sets. Histograms are a good way to see the distribution. Boxplots are best for comparing several distributions.
  - Describe distributions in terms of their shape, center, and spread, and note any unusual features such as gaps or outliers.
  - The shape of most distributions you'll see will likely be uniform, unimodal, or bimodal. It may be multimodal. If it is unimodal, then it may be symmetric or skewed.
  - A 5-number summary makes a good numerical description of a distribution: min, Q1, median, Q3, and max.
  - If the distribution is skewed, be sure to include the median and interquartile range (IQR) when you describe its center and spread.
  - A distribution that is severely skewed may benefit from re-expressing the data. If it is skewed to the high end, taking logs often works well.
  - If the distribution is unimodal and symmetric, describe its center and spread with the mean and standard deviation.
  - Use the standard deviation as a ruler to tell how unusual an observed value may be, or to compare or combine measurements made on different scales.
  - Shifting a distribution by adding or subtracting a constant affects measures of position but not measures of spread. Rescaling by multiplying or dividing by a constant affects both.
  - When a distribution is roughly unimodal and symmetric, a Normal model may be useful. For Normal models, the 68–95–99.7 Rule is a good rule of thumb.
  - If the Normal model fits well (check a histogram or Normal probability plot), then Normal percentile tables or functions found in most statistics technology can provide more detailed values.

Need more help with some of this? It never hurts to reread sections of the chapters! And in the following pages we offer you more opportunities[1] to review these concepts and skills.

The exercises that follow use the concepts and skills you've learned in the first six chapters. To be more realistic and more useful for your review, they don't tell you which of the concepts or methods you need. But neither will the exam.

[1] If you doubted that we are teachers, this should convince you. Only a teacher would call additional homework exercises "opportunities."

REVIEW EXERCISES

1. Bananas. Here are the prices (in cents per pound) of bananas reported from 15 markets surveyed by the U.S. Department of Agriculture.

51	52	45
48	53	52
50	49	52
48	43	46
45	42	50

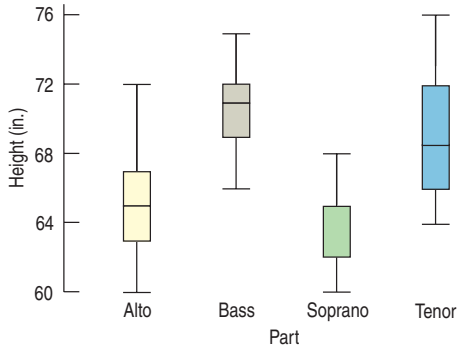
a) Display these data with an appropriate graph.
b) Report appropriate summary statistics.
c) Write a few sentences about this distribution.

2. Prenatal care. Results of a 1996 American Medical Association report about the infant mortality rate for twins carried for the full term of a normal pregnancy are shown on the next page, broken down by the level of prenatal care the mother had received.

Full-Term Pregnancies, Level of Prenatal Care	Infant Mortality Rate Among Twins (deaths per thousand live births)
Intensive	5.4
Adequate	3.9
Inadequate	6.1
Overall	5.1

a) Is the overall rate the average of the other three rates? Should it be? Explain.
b) Do these results indicate that adequate prenatal care is important for pregnant women? Explain.
c) Do these results suggest that a woman pregnant with twins should be wary of seeking too much medical care? Explain.

3. Singers. The boxplots shown display the heights (in inches) of 130 members of a choir.

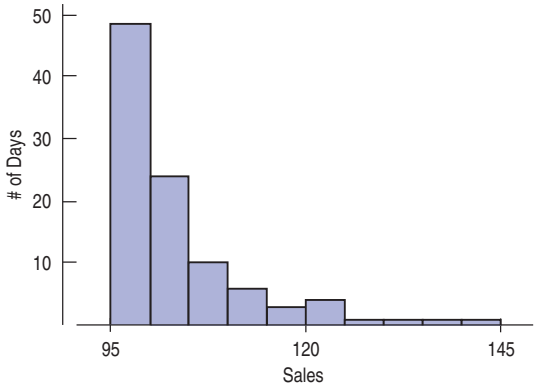


a) It appears that the median height for sopranos is missing, but actually the median and the upper quartile are equal. How could that happen?
b) Write a few sentences describing what you see.

4. Dialysis. In a study of dialysis, researchers found that "of the three patients who were currently on dialysis, 67% had developed blindness and 33% had their toes amputated." What kind of display might be appropriate for these data? Explain.

5. Beanstalks. Beanstalk Clubs are social clubs for very tall people. To join, a man must be over 6′2″ tall, and a woman over 5′10″. The National Health Survey suggests that heights of adults may be Normally distributed, with mean heights of 69.1″ for men and 64.0″ for women. The respective standard deviations are 2.8″ and 2.5″.
a) You are probably not surprised to learn that men are generally taller than women, but what does the greater standard deviation for men's heights indicate?
b) Who are more likely to qualify for Beanstalk membership, men or women? Explain.

6. Bread. Clarksburg Bakery is trying to predict how many loaves to bake. In the last 100 days, they have sold between 95 and 140 loaves per day. Here is a histogram of the number of loaves they sold for the last 100 days.



a) Describe the distribution.
b) Which should be larger, the mean number of sales or the median? Explain.
c) Here are the summary statistics for Clarksburg Bakery's bread sales. Use these statistics and the histogram above to create a boxplot. You may approximate the values of any outliers.

Summary of Sales	
Median	100
Min	95
Max	140
25th %tile	97
75th %tile	105.5

d) For these data, the mean was 103 loaves sold per day, with a standard deviation of 9 loaves. Do these statistics suggest that Clarksburg Bakery should expect to sell between 94 and 112 loaves on about 68% of the days? Explain.

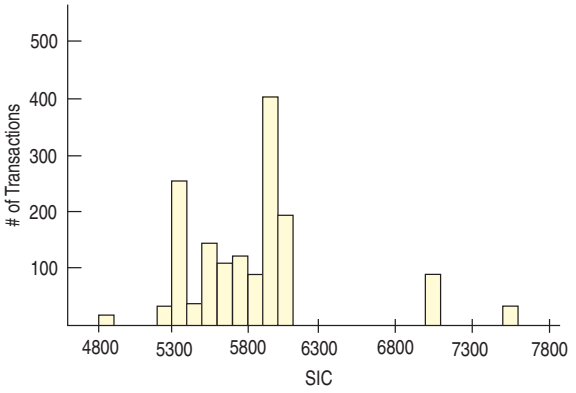
7. State University. Public relations staff at State U. collected data on people's opinions of various colleges and universities in their state. They phoned 850 local residents. After identifying themselves, the callers asked the survey participants their ages, whether they had attended college, and whether they had a favorable opinion of the university. The official report to the university's directors claimed that, in general, people had very favorable opinions about their university.
a) Identify the W's of these data.
b) Identify the variables, classify each as categorical or quantitative, and specify units if relevant.
c) Are you confident about the report's conclusion? Explain.

8. Acid rain. Based on long-term investigation, researchers have suggested that the acidity (pH) of rainfall

in the Shenandoah Mountains can be described by the Normal model $N(4.9, 0.6)$.
a) Draw and carefully label the model.
b) What percent of storms produce rainfall with pH over 6?
c) What percent of storms produce rainfall with pH under 4?
d) The lower the pH, the more acidic the rain. What is the pH level for the most acidic 20% of all storms?
e) What is the pH level for the least acidic 5% of all storms?
f) What is the IQR for the pH of rainfall?

9. Fraud detection. A credit card bank is investigating the incidence of fraudulent card use. The bank suspects that the type of product bought may provide clues to the fraud. To examine this situation, the bank looks at the Standard Industrial Code (SIC) of the business related to the transaction. This is a code that was used by the U.S. Census Bureau and Statistics Canada to identify the type of every registered business in North America.[2] For example, 1011 designates Meat and Meat Products (except Poultry), 1012 is Poultry Products, 1021 is Fish Products, 1031 is Canned and Preserved Fruits and Vegetables, and 1032 is Frozen Fruits and Vegetables.

A company intern produces the following histogram of the SIC codes for 1536 transactions:



He also reports that the mean SIC is 5823.13 with a standard deviation of 488.17.
a) Comment on any problems you see with the use of the mean and standard deviation as summary statistics.
b) How well do you think the Normal model will work on these data? Explain.

10. Streams. As part of the course work, a class at an upstate NY college collects data on streams each year. Students record a number of biological, chemical, and physical variables, including the stream name, the substrate of the stream (*limestone, shale, or mixed*), the pH, the temperature (°C), and the BCI, a measure of biological diversity.

Group	Count	%
Limestone	77	44.8
Mixed	26	15.1
Shale	69	40.1

a) Name each variable, indicating whether it is categorical or quantitative, and giving the units if available.
b) These streams have been classified according to their substrate—the composition of soil and rock over which they flow—as summarized in the table. What kind of graph might be used to display these data?

T **11. Cramming.** One Thursday, researchers gave students enrolled in a section of basic Spanish a set of 50 new vocabulary words to memorize. On Friday the students took a vocabulary test. When they returned to class the following Monday, they were retested—without advance warning. Both sets of test scores for the 28 students are shown below.

Fri	Mon	Fri	Mon
42	36	50	47
44	44	34	34
45	46	38	31
48	38	43	40
44	40	39	41
43	38	46	32
41	37	37	36
35	31	40	31
43	32	41	32
48	37	48	39
43	41	37	31
45	32	36	41
47	44		

a) Create a graphical display to compare the two distributions of scores.
b) Write a few sentences about the scores reported on Friday and Monday.
c) Create a graphical display showing the distribution of the *changes* in student scores.
d) Describe the distribution of changes.

12. Computers and Internet. A U.S. Census Bureau report (August 2000, *Current Population Survey*) found that 51.0% of homes had a personal computer and 41.5% had access to the Internet. A newspaper concluded that 92.5% of homes had either a computer or access to the Internet. Do you agree? Explain.

13. Let's play cards. You pick a card from a deck (see description in Chapter 11) and record its denomination (7, say) and its suit (maybe spades).
a) Is the variable *suit* categorical or quantitative?
b) Name a game you might be playing for which you would consider the variable *denomination* to be categorical. Explain.
c) Name a game you might be playing for which you would consider the variable *denomination* to be quantitative. Explain.

T **14. Accidents.** In 2001, Progressive Insurance asked customers who had been involved in auto accidents how far they were from home when the accident happened. The data are summarized in the table.

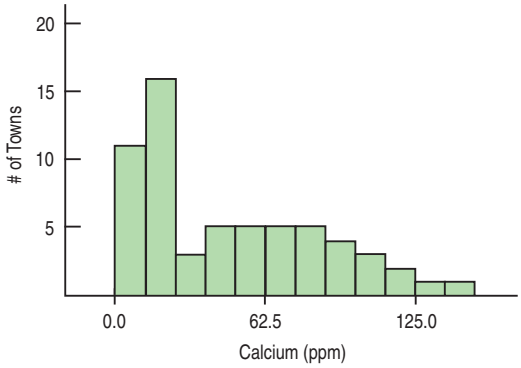
[2] Since 1997 the SIC has been replaced by the NAICS, a code of six letters.

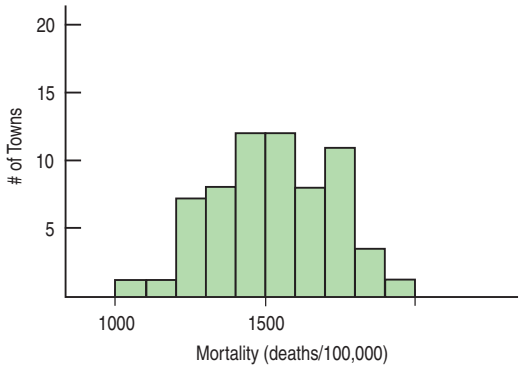
Miles from Home	% of Accidents
Less than 1	23
1 to 5	29
6 to 10	17
11 to 15	8
16 to 20	6
Over 20	17

a) Create an appropriate graph of these data.
b) Do these data indicate that driving near home is particularly dangerous? Explain.

15. Hard water. In an investigation of environmental causes of disease, data were collected on the annual mortality rate (deaths per 100,000) for males in 61 large towns in England and Wales. In addition, the water hardness was recorded as the calcium concentration (parts per million, ppm) in the drinking water.

a) What are the variables in this study? For each, indicate whether it is quantitative or categorical and what the units are.
b) Here are histograms of calcium concentration and mortality. Describe the distributions of the two variables.

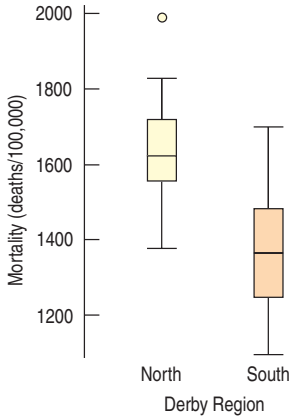




16. Hard water II. The data set from England and Wales also notes for each town whether it was south or north of Derby. Here are some summary statistics and a comparative boxplot for the two regions.

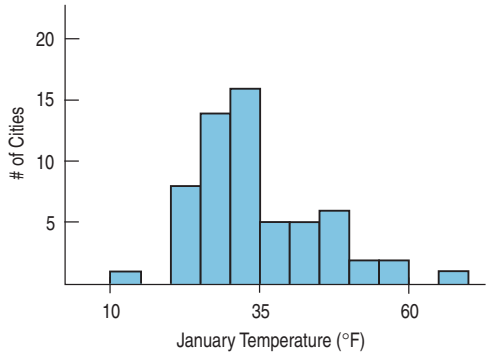
Summary of Mortality

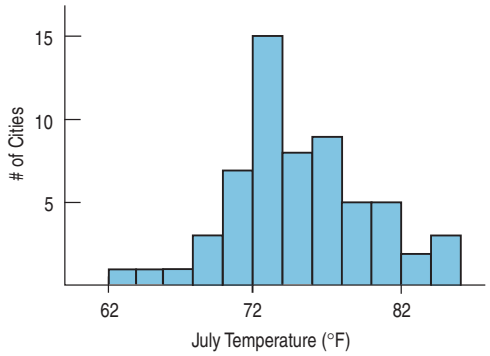
Group	Count	Mean	Median	StdDev
North	34	1631.59	1631	138.470
South	27	1388.85	1369	151.114



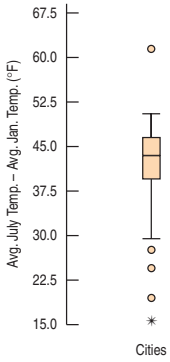
a) What is the overall mean mortality rate for the two regions?
b) Do you see evidence of a difference in mortality rates? Explain.

17. Seasons. Average daily temperatures in January and July for 60 large U.S. cities are graphed in the histograms below.



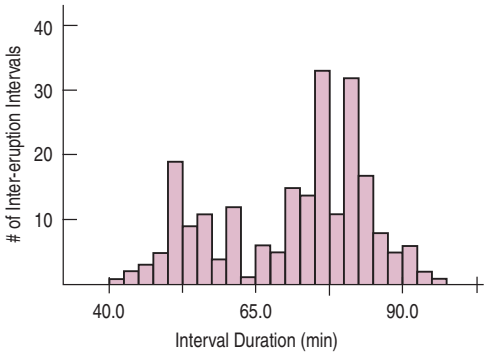


a) What aspect of these histograms makes it difficult to compare the distributions?
b) What differences do you see between the distributions of January and July average temperatures?



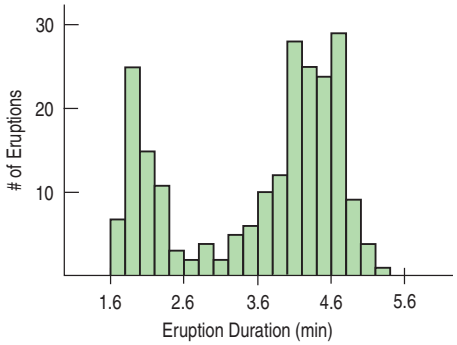
c) Differences in temperatures (July–January) for each of the cities are displayed in the boxplot above. Write a few sentences describing what you see.

18. Old Faithful. It is a common belief that Yellowstone's most famous geyser erupts once an hour at very predictable intervals. The histogram below shows the time gaps (in minutes) between 222 successive eruptions. Describe this distribution.

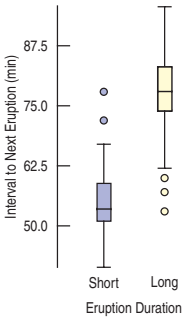


19. Old Faithful? Does the duration of an eruption have an effect on the length of time that elapses before the next eruption?

a) The histogram below shows the duration (in minutes) of those 222 eruptions. Describe this distribution.



b) Explain why it is not appropriate to find summary statistics for this distribution.
c) Let's classify the eruptions as "long" or "short," depending upon whether or not they last at least 3 minutes. Describe what you see in the comparative boxplots.



20. Teen drivers. In its *Traffic Safety Facts 2005*, the National Highway Traffic Safety Administration reported that 6.3% of licensed drivers were between the ages of 15 and 20, yet this age group was behind the wheel in 15.9% of all fatal crashes. Use these statistics to explain the concept of independence.

T **21. Liberty's nose.** Is the Statue of Liberty's nose too long? Her nose measures, 4′6″, but she is a large statue, after all. Her arm is 42 feet long. That means her arm is 42/45 = 9.3 times as long as her nose. Is that a reasonable ratio? Shown in the table are arm and nose lengths of 18 girls in a Statistics class, and the ratio of arm-to-nose length for each.

Arm (cm)	Nose (cm)	Arm/Nose Ratio
73.8	5.0	14.8
74.0	4.5	16.4
69.5	4.5	15.4
62.5	4.7	13.3
68.6	4.4	15.6
64.5	4.8	13.4
68.2	4.8	14.2
63.5	4.4	14.4
63.5	5.4	11.8
67.0	4.6	14.6
67.4	4.4	15.3
70.7	4.3	16.4
69.4	4.1	16.9
71.7	4.5	15.9
69.0	4.4	15.7
69.8	4.5	15.5
71.0	4.8	14.8
71.3	4.7	15.2

a) Make an appropriate plot and describe the distribution of the ratios.
b) Summarize the ratios numerically, choosing appropriate measures of center and spread.
c) Is the ratio of 9.3 for the Statue of Liberty unrealistically low? Explain.

T **22. Winter Olympics 2006 speed skating.** The top 25 women's 500-m speed skating times are listed in the table below:

Skater	Country	Time
Svetlana Zhurova	Russia	76.57
Wang Manli	China	76.78
Hui Ren	China	76.87
Tomomi Okazaki	Japan	76.92
Lee Sang-Hwa	South Korea	77.04
Jenny Wolf	Germany	77.25
Wang Beixing	China	77.27
Sayuri Osuga	Japan	77.39
Sayuri Yoshii	Japan	77.43
Chiara Simionato	Italy	77.68
Jennifer Rodriguez	United States	77.70
Annette Gerritsen	Netherlands	78.09
Xing Aihua	China	78.35
Sanne van der Star	Netherlands	78.59
Yukari Watanabe	Japan	78.65
Shannon Rempel	Canada	78.85
Amy Sannes	United States	78.89
Choi Seung-Yong	South Korea	79.02
Judith Hesse	Germany	79.03
Kim You-Lim	South Korea	79.25
Kerry Simpson	Canada	79.34
Krisy Myers	Canada	79.43
Elli Ochowicz	United States	79.48
Pamela Zoellner	Germany	79.56
Lee Bo-Ra	South Korea	79.73

a) The mean finishing time was 78.21 seconds, with a standard deviation of 1.03 second. If the Normal model is appropriate, what percent of the times should be within 0.5 second of 78.21?
b) What percent of the times actually fall within this interval?
c) Explain the discrepancy between a and b.

23. Sample. A study in South Africa focusing on the impact of health insurance identified 1590 children at birth and then sought to conduct follow-up health studies 5 years later. Only 416 of the original group participated in the 5-year follow-up study. This made researchers concerned that the follow-up group might not accurately resemble the total group in terms of health insurance. The table in the next column summarizes the two groups by race and by presence of medical insurance when the child was born. Carefully explain how this study demonstrates Simpson's paradox. (*Birth to Ten Study*, Medical Research Council, South Africa)

		Number (%) Insured	
		Follow-up	**Not traced**
Race	**Black**	36 of 404 (8.9%)	91 of 1048 (8.7%)
	White	10 of 12 (83.3%)	104 of 126 (82.5%)
	Overall	46 of 416 (11.1%)	195 of 1174 (16.6%)

24. Sluggers. Roger Maris's 1961 home run record stood until Mark McGwire hit 70 in 1998. Listed below are the home run totals for each season McGwire played. Also listed are Babe Ruth's home run totals.

McGwire: 3*, 49, 32, 33, 39, 22, 42, 9*, 9*, 39, 52, 58, 70, 65, 32*, 29*

Ruth: 54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22

a) Find the 5-number summary for McGwire's career.
b) Do any of his seasons appear to be outliers? Explain.
c) McGwire played in only 18 games at the end of his first big league season, and missed major portions of some other seasons because of injuries to his back and knees. Those seasons might not be representative of his abilities. They are marked with asterisks in the list above. Omit these values and make parallel boxplots comparing McGwire's career to Babe Ruth's.
d) Write a few sentences comparing the two sluggers.
e) Create a side-by-side stem-and-leaf display comparing the careers of the two players.
f) What aspects of the distributions are apparent in the stem-and-leaf displays that did not clearly show in the boxplots?

25. Be quick! Avoiding an accident when driving can depend on reaction time. That time, measured from the moment the driver first sees the danger until he or she steps on the brake pedal, is thought to follow a Normal model with a mean of 1.5 seconds and a standard deviation of 0.18 seconds.
a) Use the 68–95–99.7 Rule to draw the Normal model.
b) Write a few sentences describing driver reaction times.
c) What percent of drivers have a reaction time less than 1.25 seconds?
d) What percent of drivers have reaction times between 1.6 and 1.8 seconds?
e) What is the interquartile range of reaction times?
f) Describe the reaction times of the slowest 1/3 of all drivers.

26. Music and memory. Is it a good idea to listen to music when studying for a big test? In a study conducted by some Statistics students, 62 people were randomly assigned to listen to rap music, Mozart, or no music

while attempting to memorize objects pictured on a page. They were then asked to list all the objects they could remember. Here are the 5-number summaries for each group:

	n	Min	Q1	Median	Q3	Max
Rap	29	5	8	10	12	25
Mozart	20	4	7	10	12	27
None	13	8	9.5	13	17	24

a) Describe the W's for these data: *Who, What, Where, Why, When, How.*
b) Name the variables and classify each as categorical or quantitative.
c) Create parallel boxplots as best you can from these summary statistics to display these results.
d) Write a few sentences comparing the performances of the three groups.

T **27. Mail.** Here are the number of pieces of mail received at a school office for 36 days.

123	70	90	151	115	97
80	78	72	100	128	130
52	103	138	66	135	76
112	92	93	143	100	88
118	118	106	110	75	60
95	131	59	115	105	85

a) Plot these data.
b) Find appropriate summary statistics.
c) Write a brief description of the school's mail deliveries.
d) What percent of the days actually lie within one standard deviation of the mean? Comment.

T **28. Birth order.** Is your birth order related to your choice of major? A Statistics professor at a large university polled his students to find out what their majors were and what position they held in the family birth order. The results are summarized in the table.

a) What percent of these students are oldest or only children?
b) What percent of Humanities majors are oldest children?
c) What percent of oldest children are Humanities students?
d) What percent of the students are oldest children majoring in the Humanities?

	Birth Order*				
Major	**1**	**2**	**3**	**4+**	Total
Math/Science	34	14	6	3	**57**
Agriculture	52	27	5	9	**93**
Humanities	15	17	8	3	**43**
Other	12	11	1	6	**30**
Total	**113**	**69**	**20**	**21**	**223**

* 1 = oldest or only child

29. Herbal medicine. Researchers for the Herbal Medicine Council collected information on people's experiences with a new herbal remedy for colds. They went to a store that sold natural health products. There they asked 100 customers whether they had taken the cold remedy and, if so, to rate its effectiveness (on a scale from 1 to 10) in curing their symptoms. The Council concluded that this product was highly effective in treating the common cold.
a) Identify the W's of these data.
b) Identify the variables, classify each as categorical or quantitative, and specify units if relevant.
c) Are you confident about the Council's conclusion? Explain.

T **30. Birth order revisited.** Consider again the data on birth order and college majors in Exercise 28.
a) What is the marginal distribution of majors?
b) What is the conditional distribution of majors for the oldest children?
c) What is the conditional distribution of majors for the children born second?
d) Do you think that college major appears to be independent of birth order? Explain.

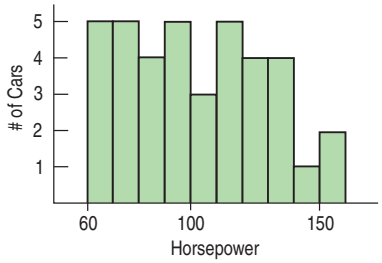
31. Engines. One measure of the size of an automobile engine is its "displacement," the total volume (in liters or cubic inches) of its cylinders. Summary statistics for several models of new cars are shown. These displacements were measured in cubic inches.

Summary of Displacement	
Count	38
Mean	177.29
Median	148.5
StdDev	88.88
Range	275
25th %tile	105
75th %tile	231

a) How many cars were measured?
b) Why might the mean be so much larger than the median?
c) Describe the center and spread of this distribution with appropriate statistics.
d) Your neighbor is bragging about the 227-cubic-inch engine he bought in his new car. Is that engine unusually large? Explain.
e) Are there any engines in this data set that you would consider to be outliers? Explain.
f) Is it reasonable to expect that about 68% of car engines measure between 88 and 266 cubic inches? (That's 177.289 ± 88.8767.) Explain.
g) We can convert all the data from cubic inches to cubic centimeters (cc) by multiplying by 16.4. For example, a 200-cubic-inch engine has a displacement of 3280 cc. How would such a conversion affect each of the summary statistics?

32. Engines, again. Horsepower is another measure commonly used to describe auto engines. Here are the summary statistics and histogram displaying horsepowers of the same group of 38 cars discussed in Exercise 31.

Summary of Horsepower	
Count	38
Mean	101.7
Median	100
StdDev	26.4
Range	90
25th %tile	78
75th %tile	125



a) Describe the shape, center, and spread of this distribution.
b) What is the interquartile range?
c) Are any of these engines outliers in terms of horsepower? Explain.
d) Do you think the 68–95–99.7 Rule applies to the horsepower of auto engines? Explain.
e) From the histogram, make a rough estimate of the percentage of these engines whose horsepower is within one standard deviation of the mean.
f) A fuel additive boasts in its advertising that it can "add 10 horsepower to any car." Assuming that is true, what would happen to each of these summary statistics if this additive were used in all the cars?

33. Age and party 2007. The Pew Research Center conducts surveys regularly asking respondents which political party they identify with. Among their results is the following table relating preferred political party and age. (http://people-press.org/reports/)

		Party			
		Republican	Democrat	Others	Total
Age	18–29	2636	2738	4765	**10139**
	30–49	6871	6442	8160	**21473**
	50–64	3896	4286	4806	**12988**
	65+	3131	3718	2934	**9784**
	Total	**16535**	**17183**	**20666**	**54384**

a) What percent of people surveyed were Republicans?
b) Do you think this might be a reasonable estimate of the percentage of all voters who are Republicans? Explain.
c) What percent of people surveyed were under 30 or over 65?
d) What percent of people were classified as "Other" and under the age of 30?
e) What percent of the people classified as "Other" were under 30?
f) What percent of people under 30 were classified as "Other"?

34. Pay. According to the *2006 National Occupational Employment and Wage Estimates for Management Occupations*, the mean hourly wage for Chief Executives was $69.52 and the median hourly wage was "over $70.00." By contrast, for General and Operations Managers, the mean hourly wage was $47.73 and the median was $40.97. Are these wage distributions likely to be symmetric, skewed left, or skewed right? Explain.

35. Age and party II. Consider again the Pew Research Center results on age and political party in Exercise 33.
a) What is the marginal distribution of party affiliation?
b) Create segmented bar graphs displaying the conditional distribution of party affiliation for each age group.
c) Summarize these poll results in a few sentences that might appear in a newspaper article about party affiliation in the United States.
d) Do you think party affiliation is independent of the voter's age? Explain.

T **36. Bike safety 2003.** The Bicycle Helmet Safety Institute website includes a report on the number of bicycle fatalities per year in the United States. The table below shows the counts for the years 1994–2003.

Year	Bicycle fatalities
1994	796
1995	828
1996	761
1997	811
1998	757
1999	750
2000	689
2001	729
2002	663
2003	619

a) What are the W's for these data?
b) Display the data in a stem-and-leaf display.
c) Display the data in a timeplot.
d) What is apparent in the stem-and-leaf display that is hard to see in the timeplot?
e) What is apparent in the timeplot that is hard to see in the stem-and-leaf display?
f) Write a few sentences about bicycle fatalities in the United States.

37. Some assembly required. A company that markets build-it-yourself furniture sells a computer desk that is advertised with the claim "less than an hour to assemble." However, through postpurchase surveys the company has learned that only 25% of its customers succeeded in building the desk in under an hour. The mean time was 1.29 hours. The company assumes that consumer assembly time follows a Normal model.

a) Find the standard deviation of the assembly time model.
b) One way the company could solve this problem would be to change the advertising claim. What assembly time should the company quote in order that 60% of customers succeed in finishing the desk by then?
c) Wishing to maintain the "less than an hour" claim, the company hopes that revising the instructions and labeling the parts more clearly can improve the 1-hour success rate to 60%. If the standard deviation stays the same, what new lower mean time does the company need to achieve?
d) Months later, another postpurchase survey shows that new instructions and part labeling did lower the mean assembly time, but only to 55 minutes. Nonetheless, the company did achieve the 60%-in-an-hour goal, too. How was that possible?

T **38. Profits.** Here is a stem-and-leaf display showing profits as a percent of sales for 29 of the *Forbes* 500 largest U.S. corporations. The stems are split; each stem represents a span of 5%, from a loss of 9% to a profit of 25%.

−0	99
−0	1234
0	111123444
0	5555679
1	00113
1	
2	2
2	5

Profits (% of sales)
(−0|3 means a loss of 3%)

a) Find the 5-number summary.
b) Draw a boxplot for these data.
c) Find the mean and standard deviation.
d) Describe the distribution of profits for these corporations.