Unit 6: Data Analysis

- How do you decide which data displays would be appropriate?
- How can statistics be misleading?
- Why are mean, median, and mode important to displaying and interpreting data?
- How does additional data change the mean, median, and mode?
- Which measure of central tendency best represents the data for a given situation?

11.1 Stem-and-Leaf Plots and Histograms

Goal: Make stem-and-leaf plots and histograms.

Vocabulary

Stem-and-leaf plot:

Frequency:

Frequency table:

Histogram:

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

The amounts (in pounds) of hay eaten by 16 elephants in one day are listed below. Use a stem-and-leaf plot to display the data.

449, 450, 419, 448, 479, 410, 446, 465,
415, 455, 479, 390, 393, 403, 460, 409

1. The least data value is ☐ and the greatest is ☐. Let the ☐ be the hundreds' and tens' digits of the data values (from 39 to 47). Let the ☐ be the ones' digits.
2. Write the stems first. Then record each amount by writing its ones' digit on the same line as its corresponding stem.
3. Make an ordered plot. Give it a key and a title.

Each stem in a stem-and-leaf plot determines an interval. For the stem-and-leaf plot in Example 1, for instance, the stem 39 determines the interval 390-399.

Hay Eaten by Elephants

Unordered plot

39	
40	
41	
42	
43	
44	
45	
46	
47	

Key: 39 | 0 = ☐

Ordered plot

39	
40	
41	
42	
43	
44	
45	
46	
47	

Key: 39 | 0 = ☐

Checkpoint

1. An elephant eats about 460 pounds of hay per day. How does the amount of hay this elephant eats compare to the amounts in Example 1?

11.2 Box-and-Whisker Plots

Goal: Make and interpret box-and-whisker plots.

Vocabulary

Box-and-whisker plot:	
Lower quartile:	
Upper quartile:	
Lower extreme:	
Upper extreme:	
Interquartile range:	

Example 1 *Making a Box-and-Whisker Plot*

Store Visits A supermarket manager recorded the number of store visits for the last 7 days. The data are given below.

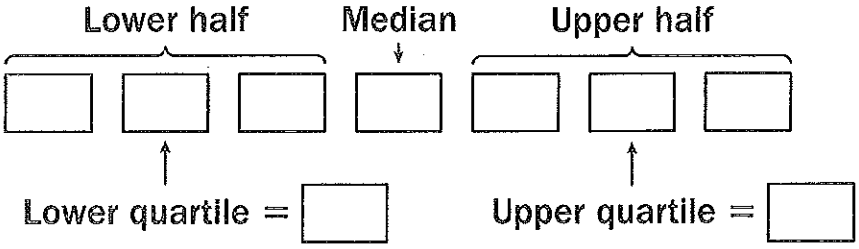
160, 124, 90, 130, 120, 165, 220

Display the data in a box-and-whisker plot.

Solution

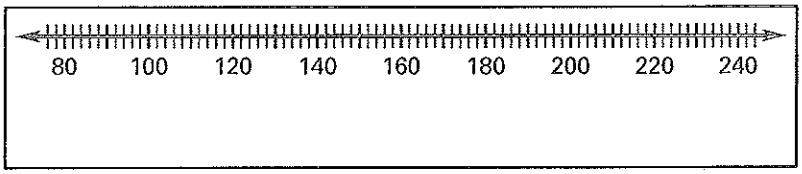
First order the data to find the median, the quartiles, and the extremes.

When a data set has an odd number of values, do not include the median in either half of the data when determining the lower and upper quartiles.



Plot the median, the quartiles, and the extremes below the number line.

Plot the extremes.

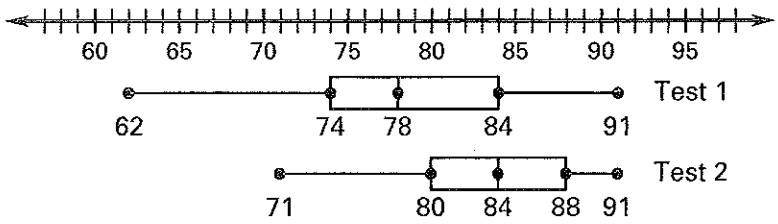


Draw a box from the lower quartile to the upper quartile. Then draw a vertical line through the median.

Draw a horizontal line (a "whisker") from the box to each of the extremes.

Example 2 *Comparing Box-and-Whisker Plots*

Test Scores The box-and-whisker plots below show a class's test scores for two tests. What conclusions can you make?



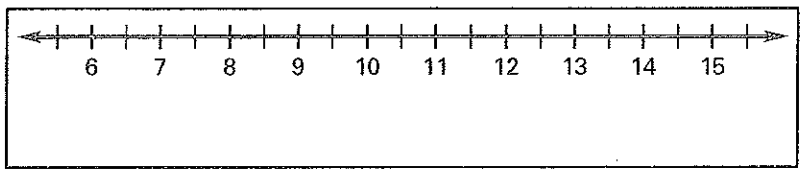
- The _____ are the same for both tests.
- The median for the second test is _____ than the median for the first test.
- The _____ for the first test is the same as the _____ for the second test.
- The scores for the _____ are more spread out than the scores for the _____. Both the range $(91 - 62 =$ ___$)$ and the interquartile range $(84 - 74 =$ ___$)$ of the first test are _____ than the range $(91 - 71 =$ ___$)$ and the interquartile range $(88 - 80 =$ ___$)$ of the second test.

Checkpoint

1. The parking cost (in dollars) at several baseball stadiums are given below.

   6, 12, 7, 8, 6, 11, 10, 9, 15

   Make a box-and-whisker plot of the data. About what percent of the stadiums charge \$9 or less?



11.3 Using Data Displays

Goal: Choose appropriate displays for data sets.

Vocabulary

Categorical data:

Numerical data:

When you choose a data display, one factor you should consider is whether the data are categorical or numerical.

Choosing Appropriate Data Displays

Use a _____ to display numerical data that change over _____.

Use a _____ to see trends in _____ numerical data.

Use a _____ to compare categorical data.

Use a _____ to represent categorical data as parts of a _____.

Use a _____ to organize numerical data based on their _____.

Use a _____ to compare the _____ of numerical data that fall in equal intervals.

Use a _____ to organize numerical data into _____ of approximately equal size.

Example 1 *Choosing an Appropriate Data Display*

The table shows the results of a survey that asked consumers to name the day of the week they usually shop for groceries. Which display(s) can you use to display the data?

When Consumers Shop	
Day	**Percent**
Sunday	16%
Monday	12%
Tuesday	9%
Wednesday	13%
Thursday	12%
Friday	16%
Saturday	22%

Solution

The responses to the survey consist of the days of the week, which are _____ data. Also, the sum of the percents is _____. So, you could use a _____ to display the data. A _____ could also be used.

Checkpoint

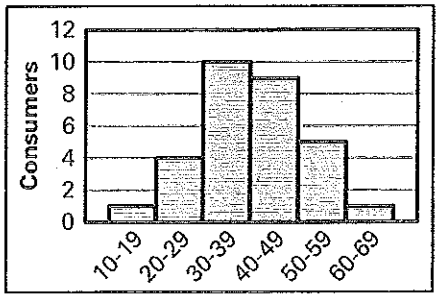
1. The table shows the voting age population and the number who voted in 10 recent federal elections. Both numbers are given in millions. Which display(s) can you use to display the data?

Voting age population (millions)	Voter turnout (millions)
205.8	105.6
200.9	73.1
196.5	96.5
193.7	75.1
189.5	104.4
185.8	67.9
182.8	91.6
178.6	65.0
174.5	92.7
169.9	67.6

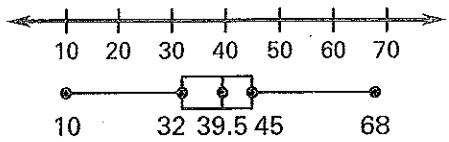
Example 2 *Comparing Data Displays*

Weekly Grocery Expenses An economist uses a histogram and a box-and-whisker plot to display the average weekly grocery expenses of 30 consumers. What are the advantages of each display?

a.



b.



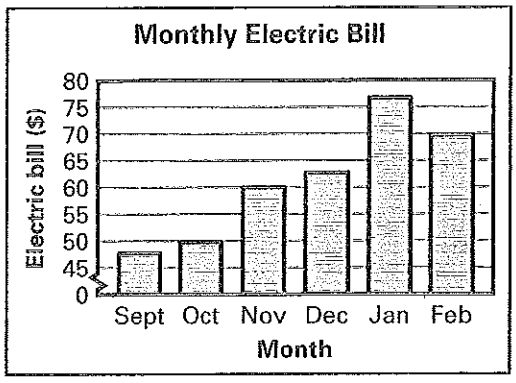
Solution

a. Using the histogram, the economist can compare the number of [ ] in each interval. For example, the economist can see that the average weekly grocery expenses for 5 of the consumers is from [ ], while the average weekly grocery expenses for [ ] of the consumers is \$30 to \$39.

b. Using the box-and-whisker plot, the economist can easily divide the average weekly grocery expenses into low, low-middle, high-middle, and high groups of approximately equal size. For example, the economist can conclude that about [ ] of the consumers spend \$32 or less per week on groceries, and about [ ] spend from \$32 to \$45 per week.

Example 3 *Identifying Misleading Data Displays*

Electric Bill The bar graph displays a household's monthly electric bill from September to February. What is misleading about the display?



Solution

The broken vertical axis exaggerates the differences in the bar lengths. For instance, it appears that February's electric bill is [ ] times that of October's, but it is actually [ ] times larger.

11.4 Collecting Data

Goal: Identify populations and sampling methods.

Vocabulary

Population:

Sample:

Biased sample:

Biased questions:

Sampling Methods

In a _____ sample, every member of the population has an _____ chance of being selected.

In a _____ sample, a _____ is used to select members of the population.

In a _____ sample, the population is divided into _____ _____. Members are selected from each _____.

In a _____ sample, only members of the population who are _____ are selected.

In a _____ sample, members of the population can _____ by volunteering.

Example 1 ***Identifying Populations and Sampling Methods***

For each survey, describe the population and the sampling method.

a. A reporter wants to determine whether residents of your county agree or disagree with budget cuts that will close some branches of the public library. The reporter asks each customer at one branch of the public library whether they agree or disagree with the budget cuts.

b. A company wants to determine whether people who eat yogurt will like the company's new brand of yogurt. The company sends a researcher to a mall and the researcher asks for volunteers to participate in a taste test comparing the new brand of yogurt to other brands.

Solution

a. The population consists of [ ]. Because the reporter only asks customers at one library branch, the sample is a [ ] sample.

b. The population consists of [ ]. Because the researcher asks for volunteers, the sample is a [ ] sample.

Checkpoint

1. A time management researcher wants to know if Americans feel that they have enough time to do what they want to do. The researcher goes to a shopping mall and asks the shoppers if they have enough time to do what they want to do. Describe the survey's population and sampling method.

Example 2 *Identifying Potentially Biased Samples*

Government A county council wants residents to give their opinions about a potential increase in property taxes. Residents can say either yes or no to the increase. The council needs the tax increase to keep every public library branch in the county open. Because the council cannot survey every resident, it decides to survey a sample. Tell whether the method could result in a biased sample. Explain.

County Council

Are you willing to pay more in property taxes to keep all county public library branches open?

☐ Yes
☐ No

a. Survey patrons at one of the library branches.

b. Give a phone number for county residents to call.

c. Visit a random sample of the county's residents at home and record their opinions.

Solution

a. This method could result in a [ ] sample because the library patrons have an interest in keeping the library open. They may favor an increase in taxes.

b. This method could result in a [ ] sample because the sample is [ ].

c. This method is less likely to result in a biased sample because a wide range of people will be surveyed. However, people who are not at home when the surveyor visits will not have their opinion recorded. As a result, the sample may be [ ].

Checkpoint

2. Refer to the survey in Checkpoint 1. Do you think the survey could result in a biased sample? Explain.

Example 3 ***Identifying Potentially Biased Questions***

Tell whether the question is potentially biased. Explain your answer. If the question is biased, rewrite it so that it is not.

a. Would you rather spend a lot of money watching a movie in a crowded movie theater or spend less money and watch a movie at home?

b. Don't you think that watching a football game is more interesting than watching a baseball game?

Solution

a. This question is [ ] because it suggests that watching a movie at home is better. An unbiased question is "[ ] [ ]?"

b. This question is [ ] because it suggests that watching a football game is more interesting. An unbiased question is "[ ]?"

Checkpoint **Tell whether the question is potentially biased. Explain your answer. If the question is biased, rewrite it so that it is not.**

3. Do you agree with the unfair policy that requires students to wear uniforms to school?

4. Do you like the new frozen yogurt flavor?

11.5 Interpreting Data

Goal: Make conclusions about populations using surveys.

Vocabulary

Margin of error:

A margin of error accounts only for errors due to the nature of random sampling. It does not account for errors that result from biased questions or biased sampling.

Example 1 *Making a Population Prediction*

A tennis industry association wants to estimate the number of Americans 12 years old and older that played tennis at least 4 times in the last 12 months. The association conducts a survey of 1000 Americans 12 years old and older and finds that 62 of them played tennis at least 4 times in the past 12 months. Of all Americans 12 years old and older, estimate how many played tennis at least 4 times in the last 12 months. (There are about 237 million Americans 12 years old and older.)

Solution

To estimate the number of Americans 12 years old and older that played tennis at least 4 times in the past 12 months, follow these steps:

1. Find the percent of Americans in the sample that played tennis at least 4 times in the past 12 months.

   $\square = \square = \square\%$

2. Find $\square$% of the total number of Americans 12 years old and older.

   $\square\% \cdot 237$ million Americans $\approx \square$ Americans

Answer: About $\square$ Americans 12 years old and older played tennis at least 4 times in the last 12 months.

Example 2 *Interpreting a Margin of Error*

Election A survey of a random sample of voters predicts that candidate A will receive 53% of the votes and that candidate B will receive 47% of the votes. The margin of error is ±4%. Can you predict who will win the election?

Solution

Use the margin of error to find intervals in which each candidate's actual percent is most likely to lie.

Candidate A: 53% − 4% = ☐ 53% + 4% = ☐

Candidate B: 47% − 4% = ☐ 47% + 4% = ☐

Using the margin of error, you can conclude that candidate A is likely to receive between ☐ and ☐ of the votes. Candidate B is likely to receive between ☐ and ☐ of the votes.

Answer: Because the intervals ☐ for the two candidates, you ☐ predict which candidate will win the election.

Checkpoint

1. A survey of a random sample of voters predicts that candidate A will receive 54% of the votes and that candidate B will receive 46% of the votes. The margin of error is ±3%. Can you predict who will win the election?

Summary of Data Analysis

When reading the results of a survey, consider the following.

- Identify the population and the sampling method.
- Determine whether the _____ represents the population.
- Determine whether the survey questions are _____.
- Identify the margin of error.
- Determine whether any data displays are potentially _____.
- Decide if the conclusions are supported by the _____.

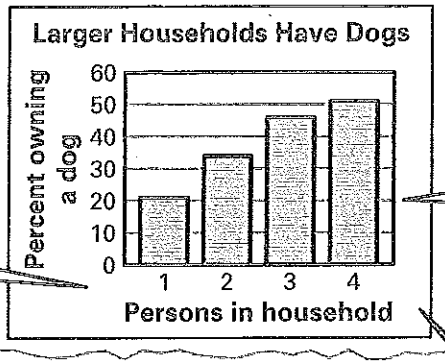
Example 3 *Interpreting a Newspaper Survey*

Tell what conclusions you can make from the following newspaper article.

Sample is not truly _____. Only American households with phones could be surveyed.

Larger families more likely to own a dog.

A recent survey of 1012 American households by telephone found that the larger the family living in the household, the more likely that the household has a dog as a pet. Each household was asked if they owned a dog. The results are shown below.



Question is _____.

Graph _____ appear misleading. There is no break in the frequency axis and all bars have the same width.

List of household sizes is _____. A larger household size might be less likely to own a dog.

Margin of error _____ given.