Name _____ Date _____

For Real?

Sample Surveys, Observational Studies, and Experiments

Vocabulary

Match each definition to its corresponding term.

1. a study that gathers data about a characteristic of the population by simply observing and describing events in their natural settings
   f. observational study
2. a survey that poses one or more questions of interest to a sample of a targeted population
   b. sample survey
3. members of the sample for an experiment
   i. experimental unit
4. the specific question that you are trying to answer or the specific information you are trying to gather
   d. characteristic of interest
5. a situation that occurs when there are other possible reasons for the results to have occurred that were not identified prior to the study
   e. confounding
6. a sample that is not representative of the population
   a. biased sample
7. an experimental condition used on treatment groups
   h. treatment
8. a process that gathers data on the effect of one or more treatments on the characteristic of interest
   g. experiment
9. a sample that is selected from the population in such a way that every member of the population has the same chance of being selected
   c. random sample

a. biased sample

b. sample survey

c. random sample

d. characteristic of interest

e. confounding

f. observational study

g. experiment

h. treatment

i. experimental unit

Problem Set

Identify the population, the sample, and the characteristic of interest for each situation.

1. A manager of a company wanted to know what proportion of employee sick days occurred on Fridays. The manager randomly selects 500 sick days and determines how many of them occurred on Fridays.

   The population is all employee sick days. The sample is 500 randomly selected sick days. The characteristic of interest is whether or each sick day occurred on a Friday.

2. In a survey, 100 randomly selected town residents were asked how many years they have lived there.

   The population is all of the people who live in the town. The sample is the 100 randomly selected town residents. The characteristic of interest is the number of years they have lived in the town.

3. A learning center claims that students can improve SAT scores by taking their prep course. In order to test the claim, an independent organization records the SAT scores of 145 randomly selected students before taking the prep course and their SAT scores after taking the prep course.

   The population is SAT scores of students who take the course at the learning center. The sample is the SAT scores of the 145 randomly selected students. The characteristic of interest is whether or not taking the prep course improves student SAT scores.

4. Every 5000^{th} item that comes off of the assembly line is tested for defects during a 24-hour period.

   The population is all of the items that come off of the assembly line. The sample is every 5000^{th} item that comes off the line in a 24 hour period. The characteristic of interest is whether or not each item is defective.

5. A study is being done to see whether body mass index (BMI) is linked to a higher risk of getting the common cold. A sample of 4565 American adults is surveyed. Their BMI and number of colds in the past year are recorded.

   The population is American adults. The sample is 4565 American adults. The characteristic of interest is the link between BMI and the risk of getting the common cold.

6. A yogurt company wants to know whether the amount of sugar in its yogurt has a significant effect on its taste. The company tests two different yogurts, one with less sugar, on 435 kids.

   The population is all kids. The sample is 435 kids. The characteristic of interest is whether or not the amount of sugar in the company's yogurt has a significant effect on its taste.

Name _____ Date _____

Classify each situation as a sample survey, an observational study, or an experiment. Explain your reasoning. If it is an experiment, identify the treatments.

7. A farmer wants to determine whether a nutritional supplement will have an effect on cows' milk production. For one week, he gives the supplement to half of his cows and then measures their milk output. Then, he compares the milk output of the cows that took the supplement with the milk output of the cows that did not get the supplement.

This is an experiment since the farmer imposed a treatment, the nutritional supplement. There are two treatments. One treatment is giving the supplement to cows. The other treatment is not giving the supplement to cows.

8. A principal at a school wants to determine whether playing in the band has an impact on students' grades. She used the school's database to determine the proportion of students in the band who have a GPA of 3.0 or higher.

This is an observational study because the principal gathered existing data from its natural setting.

9. During lunch in the school cafeteria, students are randomly surveyed about whether they like the school lunch.

This is a sample survey since the students were asked to respond to a specific question.

10. A random sample of registered voters are asked whether they will vote in the midterm elections.

This is a sample survey because the voters were asked to respond to a specific question.

11. A newspaper reporter gathers data on the length of the 40 most recently released independent films.

This is an observational study because the data was gathered from its natural setting.

12. A researcher wants to determine whether listening to classical music while taking a math test helps alleviate student anxiety. The researcher gathered data from two groups of students. One group of students listened to classical music while taking a math test and another group did not listen to classical music while taking a math test.

This is an experiment since the researcher imposed a treatment, classical music. There are two treatments. One treatment is the group of students listening to classical music and the other treatment is the group of students not listening to classical music.

Explain how confounding could occur for each observational study.

13. A researcher wants to know whether there is a link between kids developing less allergies with at least two or more pets in the home.

Confounding could occur because kids have less allergies due to various other reasons, such as having parents who do not have allergies.

14. A company wants to know whether it can claim that an all natural drug will help people with depression. A sample of adults with depression is given the drug during a four month period.

Confounding could occur because the sample of adults with depression could get better during the four month period due to another factor, such as a change of seasons.

15. A researcher wants to know whether there is a link between the amount of coffee adults drink and the frequency of leg cramps.

Confounding could occur because the frequency of leg cramps could be due to another factor, such as vitamin deficiencies or lack of sleep.

16. A factory manager wants to know whether productivity is different between the first and second shift workers due to the different time periods.

Confounding could occur because productivity could be different due to another factor, such as the type of work performed.

17. There are many studies done on whether or not there is a link between violence on TV and aggressive behavior in children.

Confounding could occur because aggressive behavior in children could be related to other factors, such as poor nutrition.

18. A state wants to determine whether there is a link between family income level and educational success for elementary school students.

Confounding could occur because educational success could be affected by other factors, such as the amount of educational capital in the school district.

Name _____ Date _____

Circle Up

Sampling Methods and Randomization

Vocabulary

Choose a term from the box that best completes each statement.

convenience sample	subjective sample	volunteer sample
simple random sample	stratified random sample	cluster sample
cluster	systematic sample	parameter
statistic		

1. A professor divided his class into females and males, then randomly selected a sample from each group. The professor obtained is a stratified random sample.
2. The manager at a discount store determines the mean salary of all of the store workers. The mean salary is an example of a parameter because it describes all of the workers.
3. John is asked to select a sample of his favorite foods from the school cafeteria. This sample is an example of a subjective sample.
4. A quality control specialist tests every 100th tablet that comes off the line. This sample is an example of a systematic sample.
5. In order to get a set of data of girl's heights, Risa uses the heights of all the girls in her class. This is an example of a convenience sample.
6. A college randomly selects 100 out of the 600 students who have taken the GRE exam and records their scores. The mean of these test scores is a statistic because it describes a sample.
7. A city manager randomly selects one block in the city and surveys all of the residents of that block. This type of sample is a cluster sample.
8. An online newspaper asks its readers to answer a question about their satisfaction with the content of the paper. This data collected from the survey results represents a volunteer sample.

9. A theater owner randomly chooses 15 different customers to receive free tickets to the next show. This sample is a simple random sample.

10. A researcher wants to collect data from a state. He divides the state into 16 regions and randomly chooses one of the regions to interview all of its residents. Each of the 16 regions is an example of a cluster.

Problem Set

Select a subjective sample of four items from each data set that best represents the mean of the data set. Explain your method for selecting the sample.

1. Test scores for a math test

70	75	89	60	95	78	54	82	91	76

Answers will vary.

I chose 54, 75, 76, and 95. Fifty-four and 95 are the minimum and maximum values. Seventy-five and 76 are in the middle, between the minimum and maximum values.

2. The weights (kilograms) of wildebeests in a zoo

130	242	227	186	250	192	215	203	232	175

Answers will vary.

I chose 130, 203, 215, and 250. One hundred thirty and 250 are the minimum and maximum values. Two hundred three and 215 are in the middle, between the minimum and maximum values.

3. The number of crimes committed each month during a ten-month period

2	1	0	3	3	4	5	2	0	6

Answers will vary.

I chose 2, 1, 0, and 3 because they are the first four values in the table and seem to be representative of the entire data set.

4. The prices (dollars) of DVDs in an electronics store

5.99	7.95	10.00	14.75	8.35	13.95	21.99	13.27	8.75	11.95

Answers will vary.

I chose 7.95, 10.00, 13.27, and 14.75 because they seem to be representative of the mean of the entire data set.

Name _____ Date _____

5. The batting averages of ten baseball players during a season

0.120	0.215	0.240	0.283	0.175	0.160	0.220	0.302	0.254	0.193

Answers will vary.

I chose 0.193, 0.215, 0.220, and 0.240 because they are the middle four values of the ordered data.

6. The heights (feet) of ten buildings in a city

102	54	76	95	250	37	65	48	27	85

Answers will vary.

I chose 54, 65, 48, and 85 because they are the middle four values of the ordered data.

For each data set, use a calculator to generate four random numbers between 1 and 10. Then use the numbers you generated to create a random sample of four from the data set.

7. Each doctor in a town is given a phone number that ends in the digits 0001 through 0010.

Phone Number	0001	0002	0003	0004	0005	0006	0007	0008	0009	0010
Doctor's Name	Thomas	Easton	Wu	Rodriguez	Pasles	Farris	Shea	Wong	Price	Siegelman

Answers will vary.

The calculator generated the numbers 2, 3, 4, and 10. The random sample consists of Easton, Wu, Rodriguez, and Siegelman.

8. Each contestant participating in a contest is assigned a number.

Contestant Number	1	2	3	4	5	6	7	8	9	10
Contestant's Name	Lia	Joe	Ariel	Jess	Victor	Roberto	Wen	Anita	Mia	Ana

Answers will vary.

9. Ten students record the number of times a coin is flipped until it lands on heads.

Trial Number	1	2	3	4	5	6	7	8	9	10
Number of Flips	1	1	1	2	3	4	1	5	2	1

Answers will vary.

10. A factory worker records the weights, in ounces, of the first ten bolts that he manufactured.

Bolt Number	1	2	3	4	5	6	7	8	9	10
Weight	2.1	2.2	2.3	2.4	2.2	2.1	2.3	2.1	2.2	2.1

Answers will vary.

11. The prices of cars at ten different dealerships are recorded.

Dealership	1	2	3	4	5
Price	24,500	32,000	27,750	23,450	26,875

Dealership	6	7	8	9	10
Price	66,000	22,750	23,650	24,735	37,500

Answers will vary.

12. A coach records the heights of players on her softball team, in meters.

Player	1	2	3	4	5	6	7	8	9	10
Height	1.8	1.5	1.6	1.6	2.2	1.3	1.7	1.8	2.0	2.1

Answers will vary.

Determine whether each study has a source of bias. If so, describe the bias and explain why the bias makes the sample unrepresentative.

13. A survey is mailed to all voters in Albany asking "Will you vote in the upcoming election?"

There is no bias in this study.

14. A survey is mailed to voters in Albany who make more than $100,000 a year asking, "Will you vote in the upcoming election?"

There is bias in this study because the voters in Albany who make less than $100,000 are not represented.

15. A medical company uses healthy patients to test their drugs for side effects.

There is bias in this study because the side effects of healthy patients could be different from the side effects of unhealthy patients.

16. A medical company uses sick patients to test their competitors' drugs for side effects.

There is bias in this study because side effects of sick patients could be different from side effects of healthy patients.

Name _____ Date _____

17. A poll by the department of education is conducted online and asks, "Do you have a computer at home?"

There is bias in this study because only people who visit the department of education's web site have an opportunity to respond.

18. A survey that measures the popularity of a magazine is inserted into the magazine asking, "Do you like this magazine?"

There is bias in this study because the people who read the magazine are more likely to like the magazine.

Use the given data set to select a stratified random sample of the specified size. Explain the method for selecting your sample.

19. The data set below displays the highest temperature recorded for 10 different cities on 4 different continents.

Highest Temperature Recorded			
North America	**Europe**	**Asia**	**Africa**
81	95	94	101
92	81	86	96
90	102	92	103
104	98	97	94
87	87	107	98
111	103	91	107
76	92	102	97
94	97	97	98
95	100	93	112
89	96	88	96

a. Create a stratified random sample that contains 4 data values.

Answers will vary.

My sample consists of the temperatures 81, 103, 97, and 98. I randomly selected one temperature from each continent.

b. Create a stratified random sample that contains 8 data values.

Answers will vary.

My sample consists of the temperatures 92, 95, 81, 103, 94, 97, 101, and 98. I randomly selected two temperatures from each continent.

20. The data set below shows the number of books read by 10 different students in 4 different English classes.

Number of Books Read by Students in Each Teacher's Class			
Mr. James	**Ms. Farley**	**Mr. Nguyen**	**Ms. Razali**
2	0	3	8
4	6	12	0
3	2	8	4
11	13	8	0
2	8	4	15
7	6	4	8
2	3	9	0
11	12	4	14
0	3	1	6
8	2	0	6

a. Create a stratified random sample that contains 12 data values.

Answers will vary.

My sample consists of 4, 11, 7, 6, 8, 12, 8, 4, 9, 8, 0, and 14. I randomly selected the number of books read by 3 different students from each teacher's class.

b. Create a stratified random sample that contains 16 data values.

Answers will vary.

My sample consists of 2, 3, 7, 11, 6, 13, 8, 3, 8, 4, 9, 4, 8, 4, 0, and 8. I randomly selected the number of books read by 4 different students from each teacher's class.

Name _____ Date _____

21. The data set displays the number of cars crossing an intersection at 8 different times during 4 different days.

Number of Cars Crossing an Intersection			
Day 1	**Day 2**	**Day 3**	**Day 4**
124	234	184	192
213	249	253	268
276	281	279	264
302	321	314	319
354	342	349	368
312	324	313	305
297	284	287	279
251	264	255	256

a. Create a stratified random sample that contains 16 data values.

Answers will vary.

My sample consists of 213, 302, 312, 251, 234, 281, 324, 284, 184, 279, 314, 313, 192, 264, 319, and 256. I randomly selected the number of car crossing at 4 different times from each day.

b. Create a stratified random sample that contains 24 data values.

Answers will vary.

My sample consists of 124, 213, 302, 312, 297, 251, 249, 281, 342, 324, 284, 264, 184, 253, 279, 349, 313, 255, 192, 268, 319, 368, 305, and 279. I randomly selected the number of car crossing at 6 different times from each day.

22. A group of five doctors takes turns volunteering at a hospital. Each doctor volunteers for a period of eight days, and then it is the next doctor's turn. The data set below shows the number of patients that each doctor saw during 8 different volunteer periods.

Number of Patients				
Dr. Kim	**Dr. Rodriguez**	**Dr. Lao**	**Dr. Woods**	**Dr. Morena**
16	18	13	19	17
15	16	17	18	16
21	23	24	19	21
26	27	24	29	28
24	23	19	26	23
27	28	26	24	27
13	15	17	16	14
18	16	14	17	18

a. Create a stratified random sample that contains 5 data values.

Answers will vary.

My sample consists of 16, 23, 24, 29, and 28. I randomly selected one volunteer day for each doctor and recorded the number of patient visits for that day.

b. Create a stratified random sample that contains 15 data values.

Answers will vary.

My sample consists of 15, 26, 27, 18, 23, 27, 13, 24, 26, 19, 18, 26, 23, 27, and 18. I randomly selected 3 volunteer days for each doctor and recorded the number of patient visits for each day.

Name _____ Date _____

23. The data set displays the amounts of recycled material collected, in pounds, from six different neighborhoods during 4 different months.

Recycled Material Collected from Six Neighborhoods (pounds)					
Arlington	**Sylvia**	**Burns**	**Beaverton**	**Bayview**	**Hillsdale**
426	482	431	324	274	134
435	324	521	675	184	162
425	398	425	573	234	176
441	436	486	543	246	186

a. Create a stratified random sample that contains 12 data values.

Answers will vary.

My sample consists of 426, 425, 482, 398, 431, 425, 324, 543, 274, 246, 134, and 176. I randomly selected 2 amounts of recycled materials from each of the 6 neighborhoods.

b. Create a stratified random sample that contains 18 data values.

Answers will vary.

My sample consists of 426, 425, 441, 482, 324, 398, 431, 425, 486, 675, 573, 543, 274, 184, 234, 162, 176, and 186. I randomly selected 3 amounts of recycled materials from each of the 6 neighborhoods.

24. The data set displays the number of people who visited a soup kitchen on specific days during the last 8 weeks.

Number of People Who Visit a Soup Kitchen			
Monday	**Wednesday**	**Friday**	**Sunday**
42	56	86	112
43	57	87	124
49	52	92	126
48	61	93	118
38	57	96	116
46	53	87	117
41	48	86	128
48	52	87	127

a. Create a stratified random sample that contains 16 data values.

Answers will vary.

My sample consists of 42, 49, 38, 41, 56, 52, 57, 53, 86, 87, 93, 87, 112, 126, 118, and 116. I randomly selected 4 values from each of the 4 days of the week.

b. Create a stratified random sample that contains 20 data values.

Answers will vary.

My sample consists of 43, 49, 48, 38, 46, 56, 57, 52, 57, 52, 86, 92, 93, 96, 87, 124, 126, 118, 116, and 128. I randomly selected 5 values from each of the 4 days of the week.

Name _____ Date _____

Create two different cluster samples for each data set. Explain the method for selecting each sample.

25.

Weights of Polar Bears at Three Zoos (pounds)		
Bronx Zoo	**San Diego Zoo**	**Philadelphia Zoo**
875	892	884
1295	1216	1237
1416	1375	1384

Answers will vary.

My two samples consisted of the polar bear weights from the San Diego Zoo, {892, 1216, 1375}, and the polar bear weights from the Philadelphia Zoo, {884, 1237, 1384}.

The weights from any one of the zoos are acceptable as a cluster sample.

26.

Movie Theatre Attendance of Four Daily Showings		
Monday	**Tuesday**	**Wednesday**
134	94	112
142	134	141
197	213	206
223	227	216

Answers will vary.

My two samples consisted of the attendance numbers on Monday, {134, 142, 197, 223}, and the attendance numbers on Tuesday, {94, 134, 213, 227}.

The attendance numbers from any one of the weekdays are acceptable as a cluster sample.

27.

Duration of Last Five Baseball Games (minutes)			
Pirates	A's	Rays	Bulldogs
116	123	124	108
124	126	122	123
137	136	142	129
128	126	127	128
153	148	153	149

Answers will vary.

My two samples consisted of the game durations for the A's, {123, 126, 136, 126, 148}, and the game durations for the Bulldogs, {108, 123, 129, 128, 149}.

The times from any one of the teams are acceptable as a cluster sample.

28.

Scores on Last Five Math Tests			
Hugo	Miriam	Anastasia	Nick
85	79	82	83
78	76	72	79
69	72	71	67
82	86	78	84
73	75	72	71

Answers will vary.

My two samples consisted of Anastasia's scores, {82, 72, 71, 78, 72}, and Hugo's scores, {85, 78, 69, 82, 73}.

The scores from any one of the students are acceptable as a cluster sample.

Name _____ Date _____

29.

Number of Weekday Customers at a Bank				
Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**
123	145	94	113	127
93	88	96	102	94
94	104	93	97	89
154	174	163	164	173
213	197	204	192	216

Answers will vary.

My two samples consisted of the weekday customers from week 2, {145, 88, 104, 174, 197}, and the weekday customers from week 3, {94, 96, 93, 163, 204}.

The numbers of customers from any one of the weeks are acceptable as a cluster sample.

30.

Weekday Prices for a Specific Stock				
Week 1	**Week 2**	**Week 3**	**Week 4**	**Week 5**
2.89	2.77	3.21	2.42	0.69
3.81	4.09	4.66	1.77	1.35
1.48	1.82	1.07	0.96	1.10
3.86	0.98	4.56	3.52	3.33
2.27	3.70	2.59	0.91	1.71

Answers will vary.

My two samples consisted of the weekday stock prices from week 1, {2.89, 3.81, 1.48, 3.86, 2.27}, and the weekday stock prices from week 4, {2.42, 1.77, 0.96, 3.52, 0.91}.

The prices from any one of the weeks are acceptable as a cluster sample.

Estimate each population mean using the data from the samples.

31. The number of home runs hit by baseball players

17, 12, 16, 21, 19, 15, 16, 22, 12, 21, 19, 18, 12, 15, 17

I used the mean number of home runs of the sample, 16.8, as an estimate for the mean number of home runs of the population.

32. The salaries in thousands of dollars of employees in a company

38, 40, 32, 41, 40, 31, 30, 41, 39, 30, 42, 31, 31, 32

I used the mean salary of the sample, $35,600, as an estimate for the mean salary of the population.

33. The heights in inches of people in an aerobics class

70, 69, 65, 60, 62, 64, 73, 65, 66, 60, 66, 65

I used the mean height of the sample, 65.4 inches, as an estimate for the mean height of the population.

34. The number of hours for restaurant employees last week

22, 35, 40, 42, 24, 36, 40, 30, 38, 22, 36, 40, 40, 42, 40, 35, 24

I used the mean number of weekly hours of the sample, 34.5 hours, as an estimate for the mean number of weekly hours of the population.

35. The lengths in inches of fish in an aquarium

20, 22, 20, 19, 14, 12, 18, 20, 14, 21, 20, 15, 19, 14, 19, 19, 21, 12, 20, 21, 22

I used the mean fish length of the sample, 18.2 inches, as an estimate for the mean fish length of the population.

36. The test scores of students in an English class

77, 94, 89, 86, 90, 68, 95, 91, 90, 89, 77, 79, 82, 68, 90, 91, 86, 87, 89, 90, 90

I used the mean test score of the sample, 85.6, as an estimate for the mean test score of the population.

Name _____ Date _____

Sleep Tight

Using Confidence Intervals to Estimate Unknown Population Means

Vocabulary

Write a definition for each term in your own words.

1. population proportion

   A population proportion is the percent of individuals (or items) in a population sharing the same characteristic. Examples of population proportions are the percent of a population who voted for a particular candidate or the percent of a company's products that are defective.

2. sample proportion

   A sample proportion is the percentage of individuals (or items) in a sample sharing the same characteristic. Examples of sample proportions are the percent of a sample who voted for a particular candidate or the percent of a company's sample of products that are defective.

3. sampling distribution

   A sampling distribution is the set of sample means or sample proportions for all possible equal-sized samples. A sampling distribution will be close to a normal distribution and will provide a good estimate for the population mean or population proportion.

4. confidence interval

   A confidence interval is an estimated range of values that will likely include the value of a population parameter.

Problem Set

Determine whether each description represents a 68%, 95%, or 99.7% confidence interval. Explain your reasoning.

1. The confidence interval for a population proportion is 45% ± 3% and the standard deviation of the sampling distribution is 0.015.

   This is a 95% confidence interval because the margin of error is 2 times the standard deviation of the sampling distribution.

   2(0.15) = 0.030 or 3%

2. The confidence interval for a population mean is 245.8 ± 20.4 and the standard deviation of the sampling distribution is 20.4.

   This is a 68% confidence interval because the margin of error is the same as the standard deviation of the sampling distribution.

3. The confidence interval for a population proportion is 22% ± 6% and the standard deviation of the sampling distribution is 0.02.

   This is a 99.7% confidence interval because the margin of error is 3 times the standard deviation of the sampling distribution.

   3(0.02) = 0.06 or 6%

4. The confidence interval for a population mean is 145.7 ± 15.3 and the standard deviation of the sampling distribution is 7.65.

   This is a 95% confidence interval because the margin of error is 2 times the standard deviation of the sampling distribution.

   2(7.65) = 15.3

5. The confidence interval for a population proportion is 6% ± 0.2% and the standard deviation of the sampling distribution is 0.002.

   This is a 68% confidence interval because the margin of error is the same as the standard deviation of the sampling distribution.

6. The confidence interval for a population mean is 7.5 ± 0.58 and the standard deviation of the sampling distribution is 0.193.

   This is a 99.7% confidence interval because the margin of error is three times the standard deviation of the sampling distribution.

   3(0.193) = 0.58

Name _____ Date _____

Determine a range of values for each population proportion. Use a 95% confidence interval.

7. A survey of 200 adults reports that 78% believe in extraterrestrial life.

The interval from 72.2% to 83.8% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.78(1-0.78)}{200}}$$

$$= \sqrt{\frac{0.78(0.22)}{200}}$$

$$\approx 0.029$$

$$78 - 2(2.9) = 72.2$$

$$78 + 2(2.9) = 83.8$$

8. A survey of 500 schools reports that 87% lock all their doors after the first bell.

The interval from 84% to 90% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.87(1-0.87)}{500}}$$

$$= \sqrt{\frac{0.87(0.13)}{500}}$$

$$\approx 0.015$$

$$87 - 2(1.5) = 84$$

$$87 + 2(1.5) = 90$$

9. A random sample of 1500 calculators shipped is checked and 4% are defective.

The interval from 3% to 5% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.04(1-0.04)}{1500}}$$

$$= \sqrt{\frac{0.04(0.96)}{1500}}$$

$$\approx 0.005$$

$$4 - 2(0.5) = 3$$

$$4 + 2(0.5) = 5$$

10. A survey of 850 doctors reports that 75% still have medical school loans.

The interval from 72% to 78% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.75(1-0.04)}{850}}$$

$$= \sqrt{\frac{0.75(0.25)}{850}}$$

$$\approx 0.015$$

$$75 - 2(1.5) = 72$$

$$75 + 2(1.5) = 78$$

11 A survey of 160 teenagers reports that 65% have homework every day of the week.

The interval from 57.4% to 72.6% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.65(1-0.65)}{160}}$$

$$= \sqrt{\frac{0.65(0.35)}{160}}$$

$$\approx 0.038$$

$$65 - 2(3.8) = 57.4$$

$$65 + 2(3.8) = 72.6$$

12. A random sample of 1400 bottles of water is checked and 14% are not filled all the way to the top.

The interval from 12.2% to 15.8% represents a 95% confidence interval for the population proportion.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.14(1-0.14)}{1400}}$$

$$= \sqrt{\frac{0.14(0.86)}{1400}}$$

$$\approx 0.009$$

$$14 - 2(0.9) = 12.2$$

$$14 + 2(0.9) = 15.8$$

Name _____ Date _____

Determine a range of values for each population mean. Use a 95% confidence interval.

13. A sample of 350 adults responded to a survey about the amount of time spent commuting to work each day. The sample mean was 34.5 minutes and the sample standard deviation was 1.7 minutes.

The interval from 34.32 minutes to 34.68 minutes represents a 95% confidence interval for the population mean.

$$\frac{S}{\sqrt{n}} = \frac{1.7}{\sqrt{350}} \approx 0.09$$

$$34.5 - 2(0.09) = 34.32$$

$$34.5 + 2(0.09) = 34.68$$

14. A sample of 300 sockeye salmon was taken and each fish was weighed. The sample mean was 6.10 lbs and the sample standard deviation was 0.15 lbs.

The interval from 6.082 lbs to 6.118 lbs represents a 95% confidence interval for the population mean.

$$\frac{S}{\sqrt{n}} = \frac{0.15}{\sqrt{300}} \approx 0.009$$

$$6.10 - 2(0.009) = 6.082$$

$$6.10 + 2(0.009) = 6.118$$

15. A sample of 150 professionals responded to a survey about the amount of time they spend in meetings each day. The sample mean was 3.5 hours and the sample standard deviation was 0.25 hours.

The interval from 3.46 hours to 3.54 hours represents a 95% confidence interval for the population mean.

$$\frac{S}{\sqrt{n}} = \frac{0.25}{\sqrt{150}} \approx 0.02$$

$$3.5 - 2(0.02) = 3.46$$

$$3.5 + 2(0.02) = 3.54$$

16. A sample of the length of 200 newborn babies was collected. The sample mean was 20.8 inches and the sample standard deviation was 1.8 inches.

The interval from 20.54 inches to 21.06 inches represents a 95% confidence interval for the population mean.

$$\frac{S}{\sqrt{n}} = \frac{1.8}{\sqrt{200}} \approx 0.13$$

$$20.8 - 2(0.13) = 20.54$$

$$20.8 + 2(0.13) = 21.06$$

17. A sample of 150 households was selected and their monthly electricity usage (kWh) was recorded. The sample mean was 94 kWh and the standard deviation was 2.8 kWh.

The interval from 93.54 kWh to 94.46 kWh represents a 95% confidence interval for the population mean.

$$\frac{S}{\sqrt{n}} = \frac{2.8}{\sqrt{150}} \approx 0.23$$

$$94 - 2(0.23) = 93.54$$

$$94 + 2(0.23) = 94.64$$

18. A sample of 80 college students responded to a survey about the amount of money they spend on outside food vendors each week. The sample mean was \$22 and the sample standard deviation was \$1.75.

The interval from \$21.60 to \$22.40 represents a 95% confidence interval for the population mean.

$$\frac{S}{\sqrt{n}} = \frac{1.75}{\sqrt{80}} \approx 0.20$$

$$22 - 2(0.20) = 21.60$$

$$22 + 2(0.20) = 22.40$$

Name _____ Date _____

How Much Different?
Using Statistical Significance to Make Inferences About Populations

Vocabulary

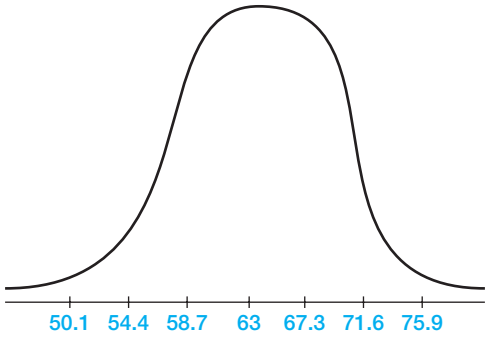
Define statistical significance in your own words.

If a result is statistically significant, then it is a result that is unlikely to occur and instead must have a reason as to why it occurred. Any result that is more than 2 standard deviations above the mean or more than two standard deviations below the mean is typically considered to be statistically significant.

Problem Set

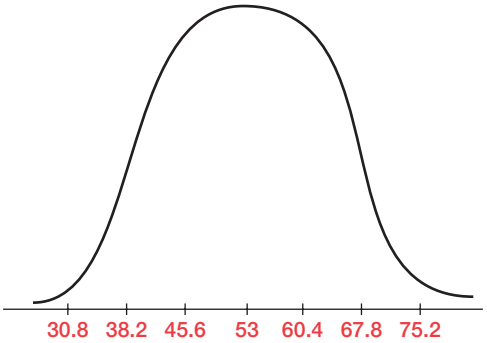
Label the horizontal axis of the normal curve using the sample proportion and standard deviation of the sampling distribution. Then, determine what sample proportions would be statistically significant.

1. A sample proportion of teenagers who attended amusement parks over the summer is 0.63 and the standard deviation of the sampling distribution is 0.043.



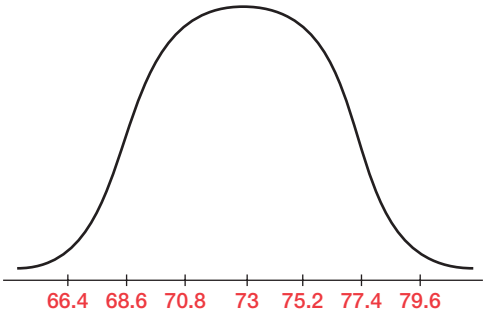
Sample proportion values less than 54.4% and greater than 71.6% are statistically significant because those values are outside of the 95% confidence interval.

2. A sample proportion of students who ate breakfast before coming to school is 0.53 and the standard deviation of the sampling distribution is 0.074.



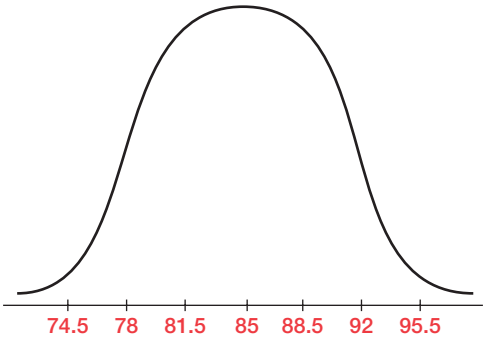
Sample proportion values less than 38.2% and greater than 67.8% are statistically significant because those values are outside of the 95% confidence interval.

3. A sample proportion of college students who want the campus to be smoke free is 0.73 and the standard deviation of the sampling distribution is 0.022.



Sample proportion values less than 68.6% and greater than 77.4% are statistically significant because those values are outside of the 95% confidence interval.

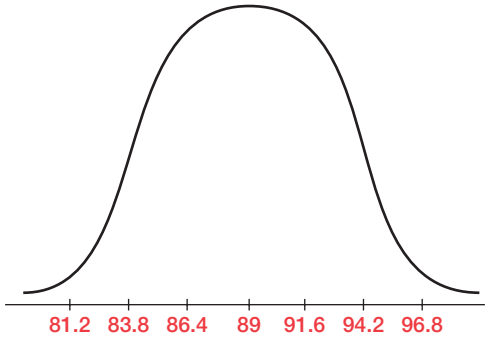
4. A sample of adult drivers who are opposed to a red-light camera in the town is 0.85 and the standard deviation of the sampling distribution is 0.035.



Sample proportion values less than 78% and greater than 92% are statistically significant because those values are outside of the 95% confidence interval.

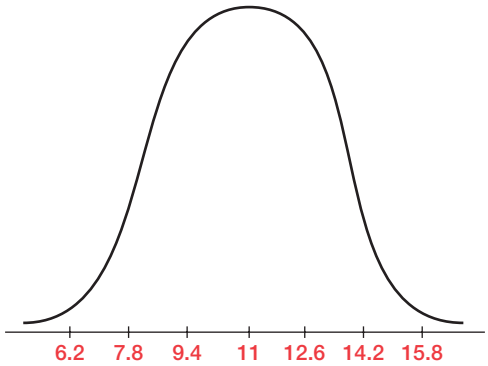
Name _____ Date _____

5. A sample of teen drivers who wear seat belts is 0.89 and the standard deviation of the sampling distribution is 0.026.



Sample proportion values less than 83.8% and greater than 94.2% are statistically significant because those values are outside of the 95% confidence interval.

6. A sample of adults who are vegetarian is 0.11 and the standard deviation of the sampling distribution is 0.016.



Sample proportion values less than 7.8% and greater than 14.2% are statistically significant because those values are outside of the 95% confidence interval.

Determine whether the difference between each pair of population proportion estimates is statistically significant. Use a 95% confidence interval. Explain your reasoning.

7. A nose spray manufacturer claims that less than 5% of users experience constant headaches. In a sample of 450 users of the nose spray, 39 of the users experienced constant headaches.

The manufacturer's claim is not within the 95% confidence interval of the population proportion estimate. The results show statistical significance.

Thirty-nine out of 450 means that approximately 9% of the users in the sample experienced constant headaches.

$\frac{39}{450} \approx 0.09$

The interval from 6.4% to 11.6% represents a 95% confidence interval for the population proportion of users who experienced constant headaches when using the nose spray.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.09(1-0.09)}{450}}$$

$$= \sqrt{\frac{0.09(0.91)}{450}}$$

$$\approx 0.013$$

$0.09 + 2(0.013) = 0.064$

$0.09 + 2(0.013) = 0.116$

Because the manufacturer's claim, 5%, is not within the 95% confidence interval, the results are statistically significant and their claim is unlikely to occur.

8. A TV show is watched on a Tuesday night by 450 out of 1000 people surveyed. It is moved to a Thursday night and it is watched by 650 out of 1250 people surveyed.

The difference between the two population proportion estimates of viewers is statistically significant because the two 95% confidence intervals do not overlap.

The 95% confidence interval for the population proportion of the Tuesday night viewership ranges from 41.8% to 48.2%. The margin of error is approximately ±0.032.

The 95% confidence interval for the population proportion of the Thursday night viewership ranges from 49.2% to 54.8%. The margin of error is approximately ±0.028.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.45(1-0.45)}{1000}}$$

$$= \sqrt{\frac{0.45(0.55)}{1000}}$$

$$\approx 0.016$$

$45 - 2(1.6) = 41.8$

$45 + 2(1.6) = 48.2$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.52(1-0.52)}{1250}}$$

$$= \sqrt{\frac{0.52(0.48)}{1250}}$$

$$\approx 0.014$$

$52 - 2(1.4) = 49.2$

$52 + 2(1.4) = 54.8$

Name _____ Date _____

9. Fifteen hundred people were polled and 36% of them support John for mayor. After a commercial ad was run for a week, fourteen hundred people were polled and 38% of them support John for mayor.

The difference between the two population proportion estimates of voters is not statistically significant because the two 95% confidence intervals overlap.

The 95% confidence interval for the population proportion of voters supporting John before the ad is run ranges from 33.6% to 38.4%. The margin of error is approximately ±0.024.

The 95% confidence interval for the population proportion of voters supporting John after the ad is run ranges from 35.4% to 40.6%. The margin of error is approximately ±0.026.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.36(1-0.36)}{1500}}$$

$$= \sqrt{\frac{0.36(0.64)}{1500}}$$

$$\approx 0.012$$

$$36 - 2(1.2) = 33.6$$

$$36 + 2(1.2) = 38.4$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.38(1-0.38)}{1400}}$$

$$= \sqrt{\frac{0.38(0.62)}{1400}}$$

$$\approx 0.013$$

$$38 - 2(1.3) = 35.4$$

$$38 + 2(1.3) = 40.6$$

10. In a poll of 330 girls at a school, 225 watch more than 2 hours of TV a day. In a poll of 314 boys at the same school, 235 of them watch more than 2 hours of TV a day.

The difference between the two population proportion estimates of TV viewers is not statistically significant because the two 95% confidence intervals overlap.

The 95% confidence interval for the population proportion of girls who watch more than 2 hours of TV a day ranges from 62.8% to 73.2%. The margin of error is approximately ±0.052.

The 95% confidence interval for the population proportion boys who watch more than 2 hours of TV a day ranges from 70.2% to 79.8%. The margin of error is approximately ±0.048.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.68(1-0.68)}{330}}$$

$$= \sqrt{\frac{0.68(0.32)}{330}}$$

$$\approx 0.026$$

$$68 - 2(2.6) = 62.8$$

$$68 + 2(2.6) = 73.2$$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.75(1-0.75)}{314}}$$

$$= \sqrt{\frac{0.75(0.25)}{314}}$$

$$\approx 0.024$$

$$75 - 2(2.4) = 70.2$$

$$75 + 2(2.4) = 79.8$$

11. Twenty-two percent of students attending college in a city are commuters. A college in a neighboring rural town polls a sample of 2700 of its students and determines that 425 of them are commuters.

The percent of commuting students to each of the colleges is statistically significant because the percent of students who commute to the city college, 22%, is not within the 95% confidence interval of the population proportion of students who commute to the rural college.

Four hundred twenty-five out of 2700 means that approximately 16% of the rural college students are commuters.

$\frac{425}{2700} \approx 0.16$

The interval from 14.6% to 17.4% represents a 95% confidence interval for the population proportion of rural college students who commute.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.16(1-0.16)}{2700}}$$

$$= \sqrt{\frac{0.16(0.84)}{2700}}$$

$$\approx 0.007$$

$0.16 - 2(0.007) = 0.146$

$0.16 + 2(0.007) = 0.174$

Because the percent of city college students who commute, 22%, is not within the 95% confidence interval, the results are statistically significant.

12. A state educator asks teachers to try two different teaching methods for a particular math topic. A sample of test scores from both methods shows that 400 out of 550 students passed with the first method and 515 out of 700 students passed with the second method.

The difference between the two population proportion estimates of passing test scores is not statistically significant because the two 95% confidence intervals overlap.

The 95% confidence interval for the population proportion of passing test scores for the first method ranges from 69.2% to 76.8%. The margin of error is approximately ±0.038.

The 95% confidence interval for the population proportion of passing test scores for the second method ranges from 70.6% to 77.4%. The margin of error is approximately ±0.034.

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.73(1-0.73)}{550}}$$

$$= \sqrt{\frac{0.73(0.27)}{550}}$$

$$\approx 0.019$$

$73 - 2(1.9) = 69.2$

$73 + 2(1.9) = 76.8$

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.74(1-0.74)}{700}}$$

$$= \sqrt{\frac{0.74(0.26)}{700}}$$

$$\approx 0.017$$

$74 - 2(1.7) = 70.6$

$74 + 2(1.7) = 77.4$

Name _____ Date _____

Determine a range of values for the population mean using a 95% confidence interval. Explain your work.

13. A sample of 75 taxi riders were asked how much their trip cost. The sample mean was $23.75 and the sample standard deviation was $1.20.

The interval from $23.47 to $24.03 represents a 95% confidence interval for the population mean.

The margin of error is approximately ±0.28.

$\frac{S}{\sqrt{n}} = \frac{1.20}{\sqrt{75}} \approx 0.14$

$23.75 - 2(0.14) = 23.47$

$23.75 + 2(0.14) = 24.03$

14. A grocery store management team wants to study the employment duration of their workers. They takes a sample of 38 cashiers from their stores and determine how long they have worked at the store. The sample mean was 2.3 years and the sample standard deviation was 1.5 years.

The interval from 1.82 to 2.78 years represents a 95% confidence interval for the population mean.

The margin of error is approximately ±0.48.

$\frac{S}{\sqrt{n}} = \frac{1.50}{\sqrt{38}} \approx 0.24$

$2.3 - 2(0.24) = 1.82$

$2.3 + 2(0.24) = 2.78$

15. During one month, a sample of 85 hardware store transactions was collected in order to determine average monthly sales for the store. The sample mean was $3550 and the sample standard deviation was $800.

The interval from $3376.40 to $3723.60 represents a 95% confidence interval for the population mean.

The margin of error is approximately ±173.6.

$\frac{S}{\sqrt{n}} = \frac{800}{\sqrt{85}} \approx 86.8$

$3550 - 2(86.8) = 3376.40$

$3550 + 2(86.8) = 3723.60$

16. A sample of 30 students taking an early morning class responded to a survey about the amount of sleep they got the previous night. The sample mean was 7.5 hours and the sample standard deviation was 1.4 hours.

The interval from 6.98 to 8.02 hours represents a 95% confidence interval for the population mean.

The margin of error is approximately ±0.52.

$\frac{S}{\sqrt{n}} = \frac{1.4}{\sqrt{30}} \approx 0.26$

$7.5 - 2(0.26) = 6.98$

$7.5 + 2(0.26) = 8.02$

17. A realtor sampled 108 apartment owners to determine an average rental amount for their properties. The sample mean was $850 and the sample standard deviation was $125.

The interval from $825.94 to $874.06 represents a 95% confidence interval for the population mean.

The margin of error is approximately ±24.06.

$\frac{S}{\sqrt{n}} = \frac{125}{\sqrt{108}} \approx 12.03$

$850 - 2(12.03) = 825.94$

$850 + 2(12.03) = 874.06$

18. In order to look at health costs, the prescription co-payment amount was recorded for fifty customers. The sample mean was $18 and the sample standard deviation was $8.

The interval from $15.74 to $20.26 represents a 95% confidence interval for the population mean.

The margin of error is approximately ±2.26.

$\frac{S}{\sqrt{n}} = \frac{8}{\sqrt{50}} \approx 1.13$

$18 - 2(1.13) = 15.74$

$18 + 2(1.13) = 20.26$

Name _____ Date _____

Use the given confidence intervals to make inferences about populations.

19. A product tester randomly selected 100 cars. Fifty of the cars were randomly selected to have an additive put into their fuel tanks while the other fifty did not get an additive. For the cars that got the additive, the interval from 1.7 to 4.4 represents a 95% confidence interval for the population mean difference in miles per gallon. For the cars that did not get the additive, the interval from −0.3 to 2.3 represents a 95% confidence interval for the population mean difference in miles per gallon. What conclusion can you state about whether or not the additive effectively increases the miles per gallon that a car gets?

The range of values for the population mean of each difference overlap so there is not any statistical significance to the results. It is not likely that there is a correlation between using the additive and increasing gas mileage.

20. A large company is trying to decide between two different shipping methods for their products. Method A uses longer hours and less drivers, and Method B uses shorter hours but more drivers. After sampling method A, the interval from 17.5 to 22 represents a 95% confidence interval for the population mean time of deliveries. After method B, the interval from 22.5 to 24.8 represents a 95% confidence interval for the population mean time of deliveries. Is there a statistically significant difference between the two shipping methods?

The data suggests that there is a statistically significant difference between the two shipping methods because the two confidence intervals overlap. It is likely that Method A is linked to shorter delivery times.

21. A professor took samples of test scores from his 8 am classes and his 9 am classes. The interval from 70.8 to 77.5 represents a 95% confidence interval for the population mean test score for the 8 am classes. The interval from 78.1 to 83.9 represents a 95% confidence interval for the population mean test score for the 9 am classes. Does attending the earlier class cause lower test scores than attending the later class?

The data does suggest a possible link between the time of the class and the test scores because the two confidence intervals do not overlap. It cannot be stated that having to get up earlier will cause the students to do worse on the test, but there is statistical evidence of a link between the test score and time of the class.

22. An amusement park wants to test a different method of loading riders on a popular ride to see whether average wait time decreases. The interval from 45.7 to 51.8 represents a 95% confidence interval for the population mean wait time using the original method. The interval from 44.4 to 46.0 represents a 95% confidence interval for the population mean wait time using the new method. Does the new method seem to decrease the wait time for the riders?

The new method does not seem to decrease the wait time for riders because the two confidence intervals overlap. The type of method does not appear to be linked to the wait time for the riders.

23. Some adults are turning to health coaches to help them to maintain a better lifestyle and lose weight. Two-hundred adults were tested for the difference in their weights over a 3 month period. One-hundred of them were asked to diet on their own, while the other 100 worked with a health coach. For the adults who did not work with a health coach, the interval from -3.5 to 8.7 represents a 95% confidence interval for the population mean weight loss. For the adults who worked with a health coach, the interval from -12.7 to -4.9 represents a 95% confidence interval for the population mean weight loss. Is there a link between weight loss and working with a health coach?

The data does suggest a possible link between weight loss and working with a health coach because the two confidence intervals do not overlap. The adults who worked with a health coach lost more weight than those who did not work with a health coach.

24. The sleep times of a random sample of 63 men and 74 women are recorded. The interval from 6.5 to 8.7 hours represents a 95% confidence interval for the population mean amount of sleep for men. The interval from 6.9 to 7.5 represents a 95% confidence interval for the population mean amount of sleep for women. Is gender linked to the amount of sleep for adults?

The data does not suggest a possible link between gender and amount of sleep for adults because the two confidence intervals overlap. A correlation between gender and amount of sleep for adults is unlikely.

Name _____ Date _____

DIY
Designing a Study and Analyzing the Results

Problem Set

Decide whether a sample survey, an observational study, or an experiment is the best method for each scenario. Describe how to obtain a random sample.

1. You want to determine the average income of public school teachers in a certain city.

   Answers will vary.

   I chose an observational study because teacher salaries are public records in my city. To obtain a random sample, I could assign an identification number to each teacher and use a computer to randomly generate a sample of teachers.

2. You want to compare the average income of male and female public school teachers in a certain city.

   Answers will vary.

   I chose an observational study because teacher salaries and teacher genders are public records in my city. To obtain a random sample, I could assign an identification number to each teacher, divide the population of teachers by gender, then use a computer to randomly select a sample from each gender group.

3. A high school principal wants to learn how much time students in each grade of the school spend on homework each week.

   Answers will vary.

   I chose a sample survey to gather data because students are the best source for accurate information about their study habits. To obtain a random sample, I could randomly select one boy and one girl from every homeroom of every grade.

4. A company wants to determine whether their drugs are harmful to people in a certain town.

   Answers will vary.

   I chose a sample survey to gather data because health records are private and the individuals have to consent to sharing information. To obtain a random sample, I could systematically select every 20th address in the town's directory and send the sample surveys those addresses.

5. A university wants to study the population trends of small towns in the state, but they only have enough resources to analyze three towns.

Answers will vary.

I chose a cluster sample to gather data. To obtain a random sample, I could arrange all of the state's towns in clusters of three, then randomly select one cluster for analysis.

6. A teacher wants to determine whether students learn better with shorter, more frequent study sessions or longer, less frequent study sessions.

Answers will vary.

I chose an experiment to gather data because two treatments can be used. One treatment is using shorter, more frequent study sessions and the other treatment is using longer, less frequent study sessions. To obtain a random samples, I could assign identification numbers to students and use a random number generator.

Decide whether random sampling, stratified random sampling, or cluster sampling is the best method for each scenario. Describe how to obtain a sample.

7. A newspaper conducts an observational study to determine which areas of town have the least number of subscriptions.

Answers will vary.

I chose a stratified random sampling to collect data. I could divide the population of the town into groups according to their geographic location (possibly north, south, east, west), and randomly select members from each group.

8. You conduct a sample survey to estimate the number of people in your school who are vegetarians.

Answers will vary.

I chose cluster sampling to collect data. During lunch, I could randomly choose one boy and one girl from each table of the lunch room.

9. You conduct a sample survey to determine whether 9th graders or 11th graders are more likely to be vegetarians.

Answers will vary.

I chose a stratified random sampling to collect data. During the 9th grade lunch, I could randomly choose one boy and one girl from each table of the lunch room. I could repeat this method during the 11th grade lunch.

Name _____ Date _____

10. A political campaign conducts an observational study to determine voting trends in preparation for the upcoming election. However, they can only analyze three neighborhoods.

Answers will vary.

I chose cluster sampling to collect data. I could divide the neighborhoods into groups of three, then randomly select one of the groups.

11. Health and Human Services conducts a survey to determine the number of city residents who do not currently have health insurance.

Answers will vary.

I chose stratified random sampling to collect data. I could divide the city into neighborhoods, then randomly survey a small sample from each neighborhood.

12. A manufacturer conducts an experiment to determine whether customers prefer their frozen vegetables to a competitors' frozen vegetables.

Answers will vary.

I chose random sampling to collect data. I could randomly select people that walk into the supermarket and ask them to participate in a taste test.

Identify and explain possible sources of bias in each situation.

13. A cell phone company wants to know how many text messages their subscribers typically send during one month. The company asks a random sample of adults that subscribe to their cell phone service, "How many text messages did you send last month?"

Answers will vary.

The sample could be biased because they only asked adults.

14. A principal wants to know if students should be allowed to use cell phones while in the classroom. She surveys one math class to see how many students have cell phones and asks them how they could use their cell phones to improve instruction.

Answers will vary.

The sample could be biased because only one class is surveyed. Also, math students may have different classroom needs for the cell phone than other subject-matter classes.

15. A principal wants to know if students should be allowed to use graphing calculators in the classroom. She surveys one English class and asks, "How do you use a graphing calculator to improve instruction?"

Answers will vary.

The sample could be biased because only one class is surveyed. Also, English students probably do not need a graphing calculator.

16. A teacher wants to know how students use graphing calculators while in the classroom. She asks volunteers for suggestions.

Answers will vary.

The sample could be biased because students who did not volunteer may be underrepresented.

17. A principal wants to know which topics students need to discuss with her. She sends out a survey to 1000 students. Only 25 respond.

Answers will vary.

The sample could be biased because only 25 students responded. They may not be a representative sample of all students.

18. The teacher wants to know how many hours students are reading each week. She asks the first 10 students who pass her room, "How many hours do you spend reading each week?"

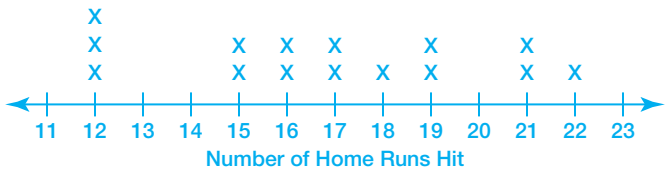
Answers will vary.

The sample could be biased because only students passing her room responded. All students did not have the same opportunity to respond.

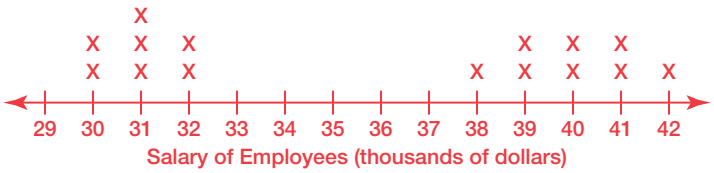
Name _____ Date _____

Create a dot plot for each set of data.

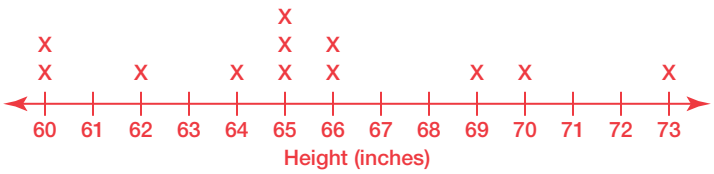
19. The following data represent the number of home runs hit by baseball players.
17, 12, 16, 21, 19, 15, 16, 22, 12, 21, 19, 18, 12, 15, 17



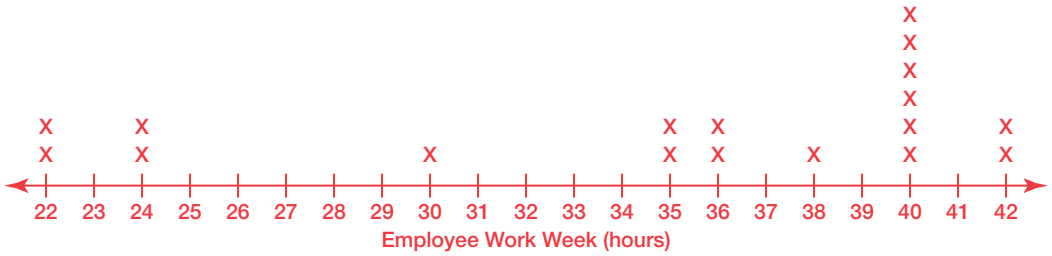
20. The following data represent the salaries in thousands of dollars of employees in a company.
38, 40, 32, 41, 40, 31, 30, 41, 39, 39, 30, 42, 31, 31, 32



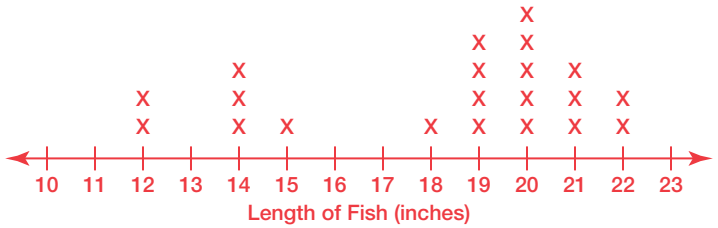
21. The following data represent the heights in inches of people in an aerobics class.
70, 69, 65, 60, 62, 64, 73, 65, 66, 60, 66, 65



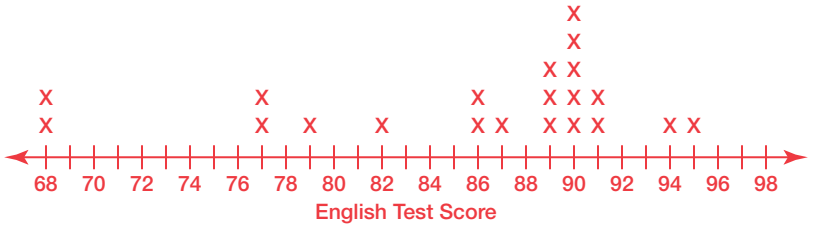
22. The following data represent the number of hours that restaurant employees worked last week.
22, 35, 40, 42, 24, 36, 40, 30, 40, 38, 22, 36, 40, 40, 42, 40, 35, 24



23. The following data represent the lengths in inches of fish in an aquarium.
20, 22, 20, 19, 14, 12, 18, 20, 14, 21, 20, 15, 19, 14, 19, 19, 21, 12, 20, 21, 22

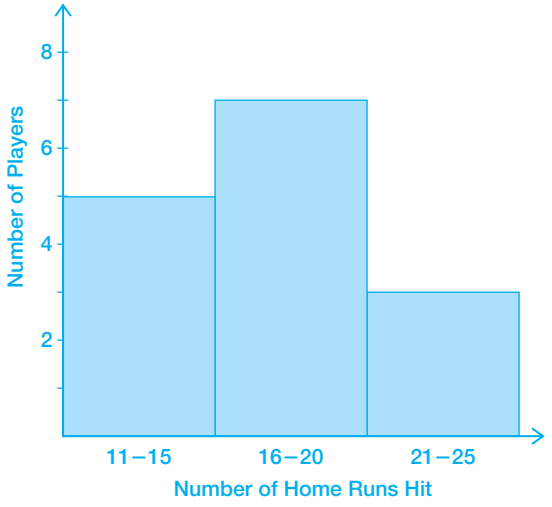


24. The following data represent students' English test scores.
77, 94, 89, 86, 90, 68, 95, 91, 90, 89, 77, 79, 82, 68, 90, 91, 86, 87, 89, 90, 90



Create a histogram for each set of data.

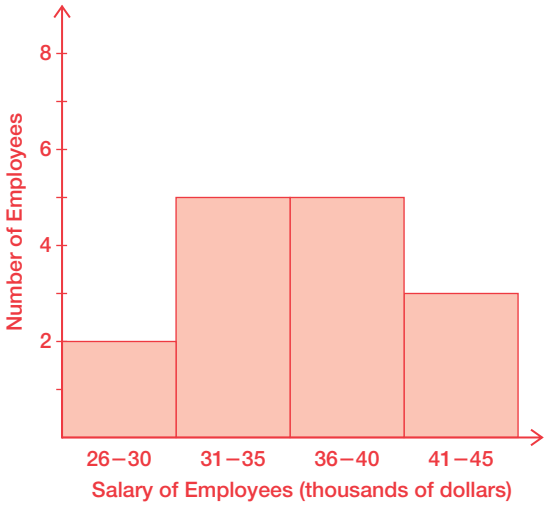
25. The following data represent the number of home runs hit by baseball players.
17, 12, 16, 21, 19, 15, 16, 22, 12, 21, 19, 18, 12, 15, 17



Name _____ Date _____

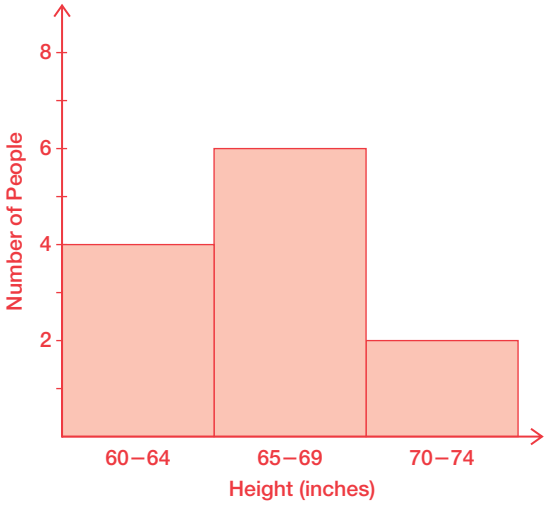
26. The following data represent the salaries in thousands of dollars of employees in a company.

38, 40, 32, 41, 40, 31, 30, 41, 39, 39, 30, 42, 31, 31, 32



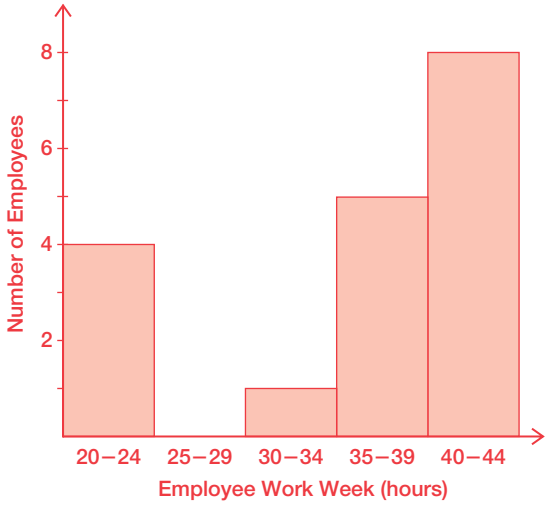
27. The following data represent the heights in inches of people in an aerobics class.

70, 69, 65, 60, 62, 64, 73, 65, 66, 60, 66, 65



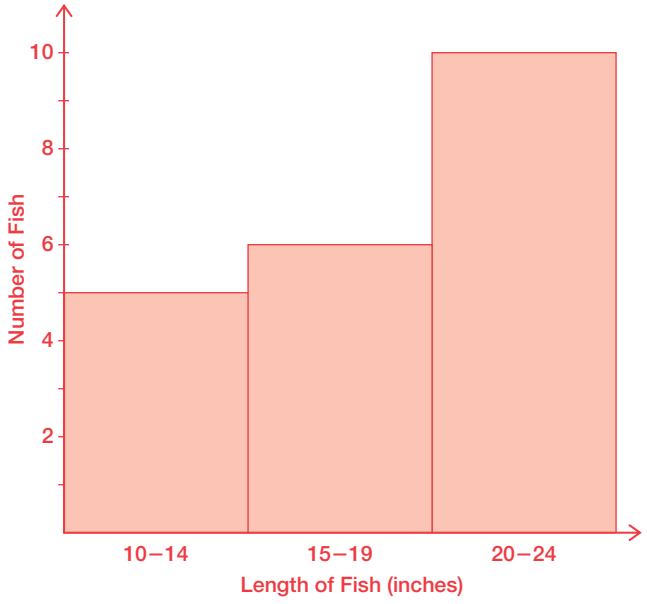
28. The following data represent the number of hours that restaurant employees worked last week.

22, 35, 40, 42, 24, 36, 40, 30, 40, 38, 22, 36, 40, 40, 42, 40, 35, 24



29. The following data represent the lengths in inches of fish in an aquarium.

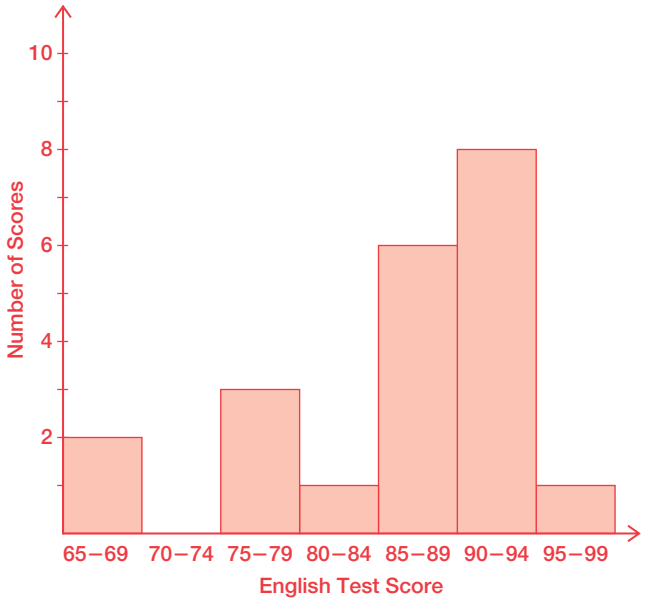
20, 22, 20, 19, 14, 12, 18, 20, 14, 21, 20, 15, 19, 14, 19, 19, 21, 12, 20, 21, 22



Name _____ Date _____

30. The following data represent students' English test scores.

77, 94, 89, 86, 90, 68, 95, 91, 90, 89, 77, 79, 82, 68, 90, 91, 86, 87, 89, 90, 90



Create a stem-and-leaf plot for each set of data.

31. The following data represent the number of home runs hit by baseball players.

17, 12, 16, 21, 19, 15, 16, 22, 12, 21, 19, 18, 12, 15, 17

Number of Home Runs Hit

1	2 2 2 5 5 6 6 7 7 8 9 9
2	1 1 2

2 | 1 = 21

32. The following data represent the salaries in thousands of dollars of employees in a company.

38, 40, 32, 41, 40, 31, 30, 41, 39, 39, 30, 42, 31, 31, 32

Salary of Employees (thousands of dollars)

3	0 0 1 1 1 2 2 8 9 9
4	0 0 1 1 2

3 | 0 = 30

33. The following data represent the heights in inches of people in an aerobics class.
70, 69, 65, 60, 62, 64, 73, 65, 66, 60, 66, 65

Height (inches)

6	0 0 2 4 5 5 5 6 6 9
7	0 3

6 | 4 = 64

34. The following data represent the number of hours that restaurant employees worked last week.
22, 35, 40, 42, 24, 36, 40, 30, 40, 38, 22, 36, 40, 40, 42, 40, 35, 24

Employee Work Week (hours)

2	2 2 4 4
3	0 5 5 6 6 8
4	0 0 0 0 0 0 2 2

2 | 2 = 22

35. The following data represent the lengths in inches of fish in an aquarium.
20, 22, 20, 19, 14, 12, 18, 20, 14, 21, 20, 15, 19, 14, 19, 19, 21, 12, 20, 21, 22

Length of Fish (inches)

1	2 2 4 4 4 5 8 9 9 9 9
2	0 0 0 0 0 1 1 1 2 2

1 | 2 = 12

36. The following data represent students' English test scores.
77, 94, 89, 86, 90, 68, 95, 91, 90, 89, 77, 79, 82, 68, 90, 91, 86, 87, 89, 90, 90

English Test Score

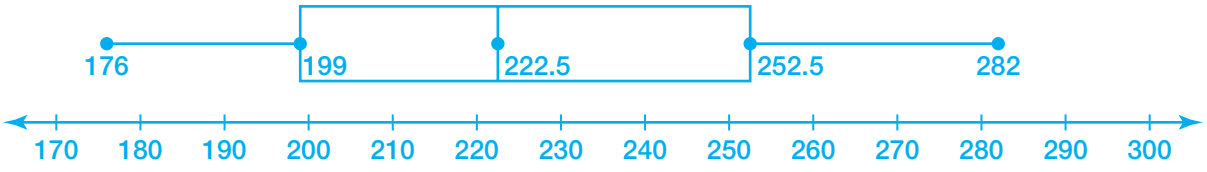
6	8 8
7	7 7 9
8	2 6 6 7 9 9 9
9	0 0 0 0 0 1 1 4 5

8 | 2 = 82

Create a box-and-whisker plot for each set of data.

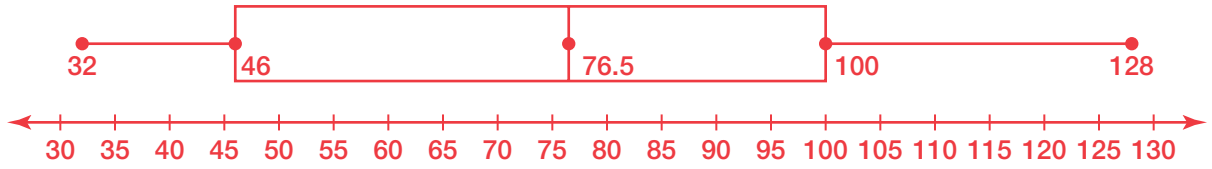
37. The following data represent the weights in pounds of 12 players on a football team.
220, 244, 185, 261, 243, 225, 273, 282, 176, 190, 208, 214

minimum: 176, first quartile: 199, median: 222.5, third quartile: 252.5, maximum: 282



38. The following data shows the amounts of annual snowfall in inches for 10 northeastern U.S. cities.
43, 32, 56, 128, 85, 96, 46, 122, 100, 68

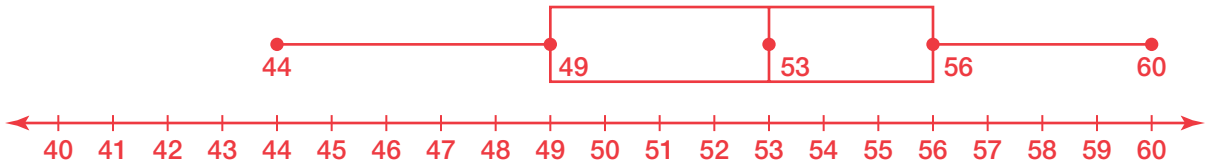
minimum: 32, first quartile: 46, median: 76.5, third quartile: 100, maximum: 128



Name _____ Date _____

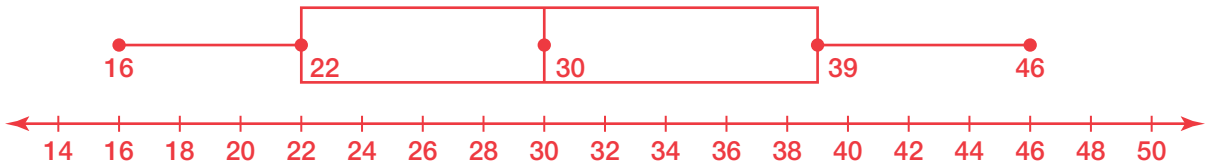
39. The following data shows the average daily low temperatures in degrees Fahrenheit for 15 U.S. cities.
60, 44, 54, 50, 59, 53, 48, 46, 55, 56, 59, 52, 49, 55, 51

minimum: 44, first quartile: 49, median: 53, third quartile: 56, maximum: 60



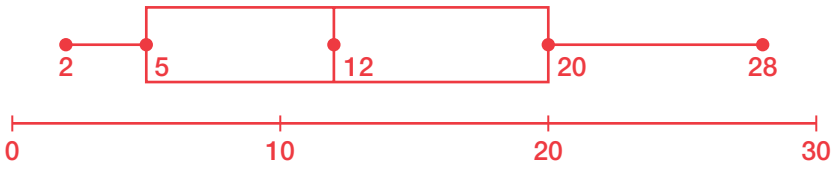
40. The following data shows the heights in feet of 11 trees in a forest.
41, 30, 22, 16, 28, 32, 27, 46, 39, 35, 20

minimum: 16, first quartile: 22, median: 30, third quartile: 39, maximum: 46



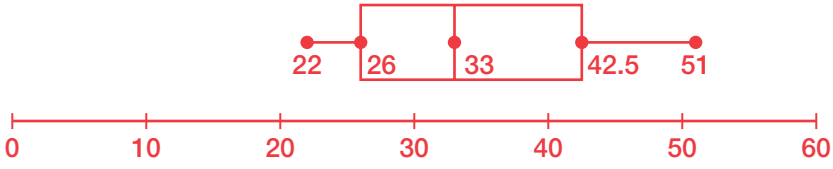
41. The following data represent the total number of points scored by the 11 players on a basketball team.
12, 28, 11, 2, 17, 5, 20, 26, 7, 2, 18

minimum: 2, first quartile: 5, median: 12, third quartile: 20, maximum: 28



42. The following data represent the ages of 13 students in an adult karate class.
26, 35, 45, 22, 23, 30, 26, 47, 33, 35, 40, 32, 51

minimum: 22, first quartile: 26, median: 33, third quartile: 42.5, maximum: 51



Determine the mean, median, and mode of each data set. Tell if the data is symmetric.

43. The following data represent the amounts of time in minutes that it takes nine different people to get to work.

15, 20, 30, 10, 55, 45, 25, 20, 30

Mean = 27.78

Median = 25

Modes = 20, 30

Symmetry: not symmetric

44. The following data represent the long jump distances in feet of 12 participants in a track meet.

19.9, 22.1, 18.5, 20.4, 20.8, 21.3, 19.2, 20.1, 21.0, 15.2, 20.0, 21.6

Mean = 20.00

Median = 20.25

Mode = none

Symmetry: not symmetric

45. The following data represent the fuel efficiencies in miles per gallon of 8 cars at a dealer.

20, 16, 32, 30, 28, 25, 24, 32

Mean = 25.88

Median = 26.5

Mode = 32

Symmetry: not symmetric

46. The following data represent the prices in dollars of 11 shirts in a store.

20, 35, 40, 20, 15, 20, 30, 25, 20, 40, 35

Mean = 27.27

Median = 25

Mode = 20

Symmetry: not symmetric

Name _____ Date _____

47. The following data represent the lengths in inches of 10 fish in an aquarium.

3, 6, 8, 11, 9, 5, 7, 6, 8, 10

Mean = 7.3

Median = 7.5

Modes = 6, 8

Symmetry: not symmetric

48. The following data represent the number of extra-credit points earned by 14 students in a science class.

10, 15, 0, 15, 0, 5, 10, 20, 15, 0, 5, 10, 0, 0

Mean = 7.5

Median = 7.5

Mode = 0

Symmetry: not symmetric

Determine the standard deviation and the quartiles of each data set.

49. The following data represent the amount of time in minutes that it takes nine different people to get to work.

15, 20, 30, 10, 55, 45, 25, 20, 30

Standard deviation: 14.39

Quartile 1: 17.5

Quartile 2: 25

Quartile 3: 37.5

50. The following data represent the long jump distances in feet of 12 participants in a track meet.

19.9, 22.1, 18.5, 20.4, 20.8, 21.3, 19.2, 20.1, 21.0, 15.2, 20.0, 21.6

Standard deviation: 1.82

Quartile 1: 19.55

Quartile 2: 20.25

Quartile 3: 21.15

51. The following data represent the fuel efficiencies in miles per gallon of 8 cars at a dealer.
20, 16, 32, 30, 28, 25, 24, 32

Standard deviation: 5.77

Quartile 1: 22

Quartile 2: 26.5

Quartile 3: 31

52. The following data represent the prices in dollars of 11 shirts in a store.
20, 35, 40, 20, 15, 20, 30, 25, 20, 40, 35

Standard deviation: 9.05

Quartile 1: 20

Quartile 2: 25

Quartile 3: 35

53. The following data represent the lengths in inches of 10 fish in an aquarium.
3, 6, 8, 11, 9, 5, 7, 6, 8, 10

Standard deviation: 2.41

Quartile 1: 6

Quartile 2: 7.5

Quartile 3: 9

54. The following data represent the number of extra-credit points earned by 14 students in a science class.
10, 15, 0, 15, 0, 5, 10, 20, 15, 0, 5, 10, 0, 0

Standard deviation: 7.0

Quartile 1: 0

Quartile 2: 7.5

Quartile 3: 15