

DESCRIPTIVE STATISTICS: HOMEWORK

EXERCISE 1

Twenty-five randomly selected students were asked the number of movies they watched the previous week. The results are as follows:

# of movies	Frequency	Relative Frequency	Cumulative Relative Frequency
0	5		
1	9		
2	6		
3	4		
4	1		

- Find the sample mean, \bar{x} .
- Find the sample standard deviation, s .
- Construct a histogram of the data.
- Complete the columns of the chart.
- Find the first quartile.
- Find the median.
- Find the third quartile.
- Construct a box plot of the data.
- What percent of the students saw fewer than three movies?
- Find the 40th percentile.
- Find the 90th percentile.
- Construct a line graph of the data.
- Construct a stem plot of the data.

EXERCISE 2

The median age for U.S. blacks currently is 30.1 years; for U.S. whites it is 36.6 years. (Source: U.S. Census).

- Based upon this information, give two reasons why the black median age could be lower than the white median age.
- Does the lower median age for blacks necessarily mean that blacks die younger than whites? Why or why not?
- How might it be possible for blacks and whites to die at approximately the same age, but for the median age for whites to be higher?

EXERCISE 3

Forty randomly selected students were asked the number of pairs of sneakers they owned. Let X = the number of pairs of sneakers owned. The results are as follows:

X	Frequency	Relative Frequency	Cumulative Relative Frequency
1	2		
2	5		
3	8		
4	12		
5	12		
7	1		

- Find the sample mean \bar{x}
- Find the sample standard deviation, s
- Construct a histogram of the data.
- Complete the columns of the chart.
- Find the first quartile.
- Find the median.
- Find the third quartile.
- Construct a box plot of the data.
- What percent of the students owned at least five pairs?
- Find the 40th percentile.
- Find the 90th percentile.
- Construct a line graph of the data.
- Construct a stem plot of the data.

EXERCISE 4

600 adult Americans were asked by telephone poll, What do you think constitutes a middle-class income? The results are below. Also, include left endpoint, but not the right endpoint. (Source: Time magazine; survey by Yankelovich Partners, Inc.)

Note: "Not sure" answers were omitted from the results.

Salary (\$)	Relative Frequency
< 20,000	0.02
20,000 - 25,000	0.09
25,000 - 30,000	0.19
30,000 - 40,000	0.26
40,000 - 50,000	0.18
50,000 - 75,000	0.17
75,000 - 99,999	0.02
100,000+	0.01

- What percent of the survey answered "not sure"?
- What percent think that middle-class is from \$25,000 - \$50,000 ?
- Construct a histogram of the data
 - Should all bars have the same width, based on the data? Why or why not?
 - How should the <20,000 and the 100,000+ intervals be handled? Why?
- Find the 40th and 80th percentiles.
- Construct a bar graph of the data.

EXERCISE 5

Following are the published weights (in pounds) of all of the team members of the San Francisco 49ers from a previous year (Source: San Jose Mercury News).

177; 205; 210; 210; 232; 205; 185; 185; 178; 210; 206; 212; 184; 174; 185; 242; 188; 212; 215; 247; 241; 223; 220; 260; 245; 259; 278; 270; 280; 295; 275; 285; 290; 272; 273; 280; 285; 286; 200; 215; 185; 230; 250; 241; 190; 260; 250; 302; 265; 290; 276; 228; 265

- Organize the data from smallest to largest value.
- Find the median.
- Find the first quartile.
- Find the third quartile.
- Construct a box plot of the data.
- The middle 50% of the weights are from _____ to _____.
- If our population were all professional football players, would the above data be a sample of weights or the population of weights? Why?
- If our population were the San Francisco 49ers, would the above data be a sample of weights or the population of weights? Why?
- Assume the population was the San Francisco 49ers. Find:

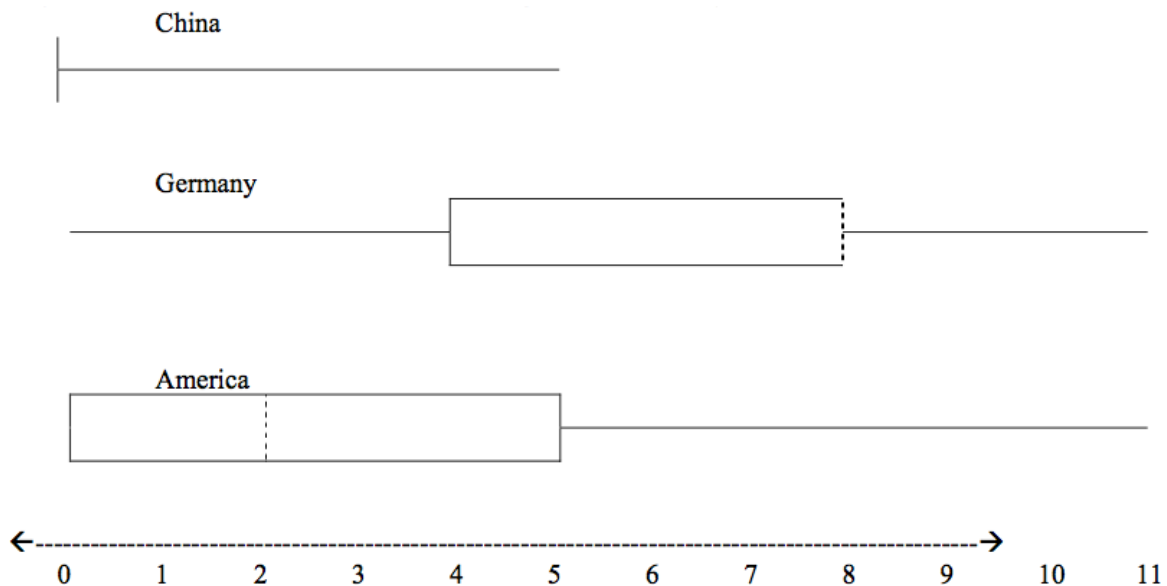
EXERCISE 6

An elementary school class ran 1 mile in an average of 11 minutes with a standard deviation of 3 minutes. Rachel, a student in the class, ran 1 mile in 8 minutes. A junior high school class ran 1 mile in an average of 9 minutes, with a standard deviation of 2 minutes. Kenji, a student in the class, ran 1 mile in 8.5 minutes. A high school class ran 1 mile in an average of 7 minutes with a standard deviation of 4 minutes. Nedda, a student in the class, ran 1 mile in 8 minutes.

- Why is Kenji considered a better runner than Nedda, even though Nedda ran faster than he?
- Who is the fastest runner with respect to his or her class? Explain why.

EXERCISE 7

In a survey of 20 year olds in China, Germany and America, people were asked the number of foreign countries they had visited in their lifetime. The following box plots display the results.



- In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected.
- Explain how it is possible that more Americans than Germans surveyed have been to over eight foreign countries.
- Compare the three box plots. What do they imply about the foreign travel of twenty year old residents of the three countries when compared to each other?

EXERCISE 8

Twelve teachers attended a seminar on mathematical problem solving. Their attitudes were measured before and after the seminar. A positive number change attitude indicates that a teacher's attitude toward math became more positive. The twelve change scores are as follows:

$$\{3; 8; -1; 2; 0; 5; -3; 1; -1; 6; 5; -2\}$$

- What is the average change score?
- What is the standard deviation for this population?
- What is the median change score?
- Find the change score that is 2.2 standard deviations below the mean.

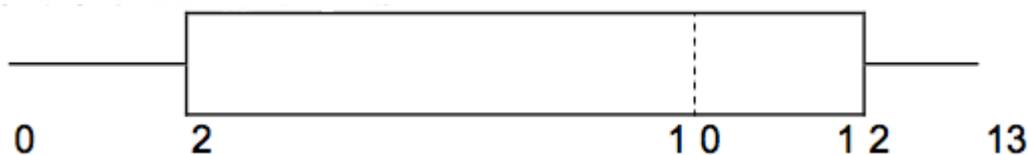
EXERCISE 9

Three students were applying to the same graduate school. They came from schools with different grading systems. Which student had the best G.P.A. when compared to his school? Explain how you determined your answer.

Student	G.P.A.	School Ave. G.P.A.	School Standard Deviation
Thuy	2.7	3.2	0.8
Vichet	87	75	20
Kamala	8.6	8	0.4

EXERCISE 10

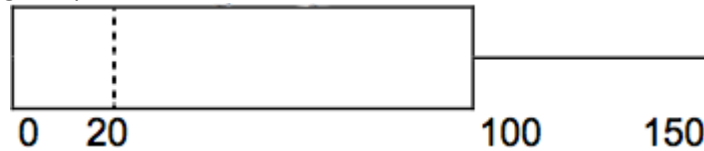
Given the following box plot:



- Which quarter has the smallest spread of data? What is that spread?
- Which quarter has the largest spread of data? What is that spread?
- Find the Inter Quartile Range (IQR).
- Are there more data in the interval 5 - 10 or in the interval 10 - 13? How do you know this?
- Which interval has the fewest data in it? How do you know this?
 - 0-2
 - 2-4
 - 10-12
 - 12-13

EXERCISE 11

Given the following box plot:



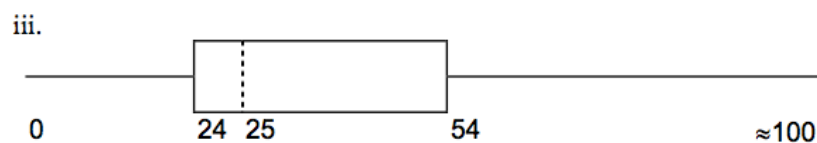
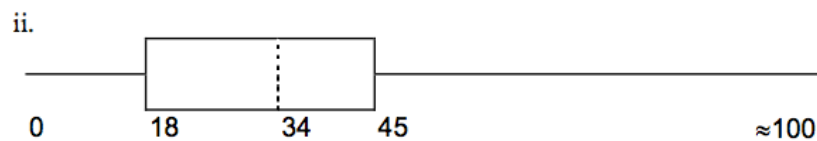
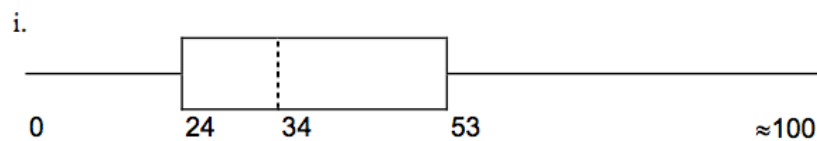
- Think of an example (in words) where the data might fit into the above box plot. In 2-5 sentences, write down the example.
- What does it mean to have the first and second quartiles so close together, while the second to fourth quartiles are far apart?

EXERCISE 12

Santa Clara County, CA, has approximately 27,873 Japanese-Americans. Their ages are as follows. (Source: West magazine)

Age Group	Percent of Community
0-17	18.9
18-24	8.0
25-34	22.8
35-44	15.0
45-54	13.1
55-64	11.9
65+	10.3

- Construct a histogram of the Japanese-American community in Santa Clara. What percent of the community is under age 35?
- Which box plot most resembles the information above?



EXERCISE 13

Suppose that three book publishers were interested in the number of fiction paperbacks adult consumers purchase per month. Each publisher conducted a survey. In the survey, each asked adult consumers the number of fiction paperbacks they had purchased the previous month. The results are below.

Publisher A			Publisher B			Publisher C		
# of books	Freq.	Rel. Freq.	# of books	Freq.	Rel. Freq.	# of books	Freq.	Rel. Freq.
0	10		0	18		0-1	20	
1	12		1	24		2-3	35	
2	16		2	24		4-5	12	
3	12		3	22		6-7	2	
4	8		4	15		8-9	1	
5	6		5	10				
6	2		7	5				
8	2		9	1				

- Find the relative frequencies for each survey. Write them in the charts.
- Using either a graphing calculator, computer, or by hand, use the frequency column to construct a histogram for each publisher's survey. For Publishers A and B, make bar widths of 1. For Publisher C, make bar widths of 2.
- In complete sentences, give two reasons why the graphs for Publishers A and B are not identical.
- Would you have expected the graph for Publisher C to look like the other two graphs? Why or why not?
- Make new histograms for Publisher A and Publisher B. This time, make bar widths of 2.
- Now, compare the graph for Publisher C to the new graphs for Publishers A and B. Are the graphs more similar or more different? Explain your answer.

EXERCISE 14

Often, cruise ships conduct all on-board transactions, with the exception of gambling, on a cashless basis. At the end of the cruise, guests pay one bill that covers all on-board transactions. Suppose that 60 single travelers and 70 couples were surveyed as to their on-board bills for a seven-day cruise from Los Angeles to the Mexican Riviera. Below is a summary of the bills for each group.

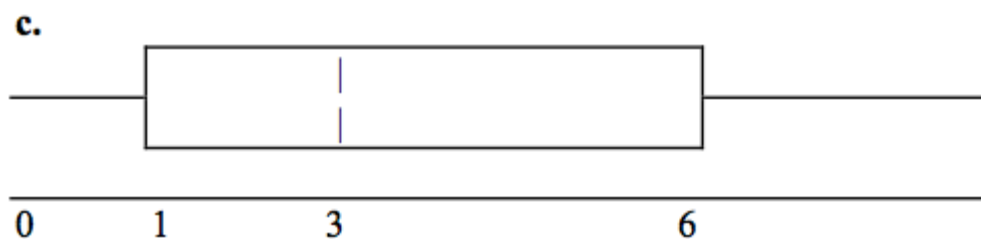
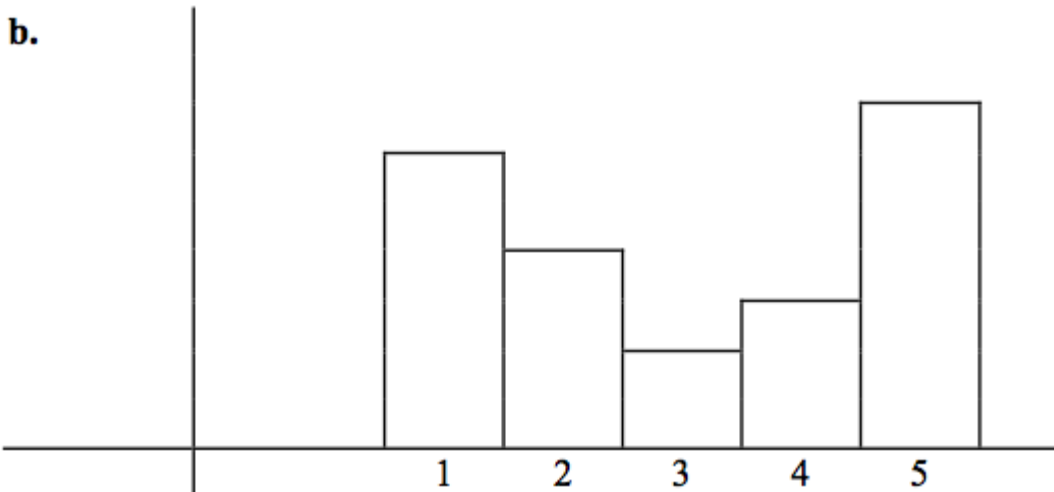
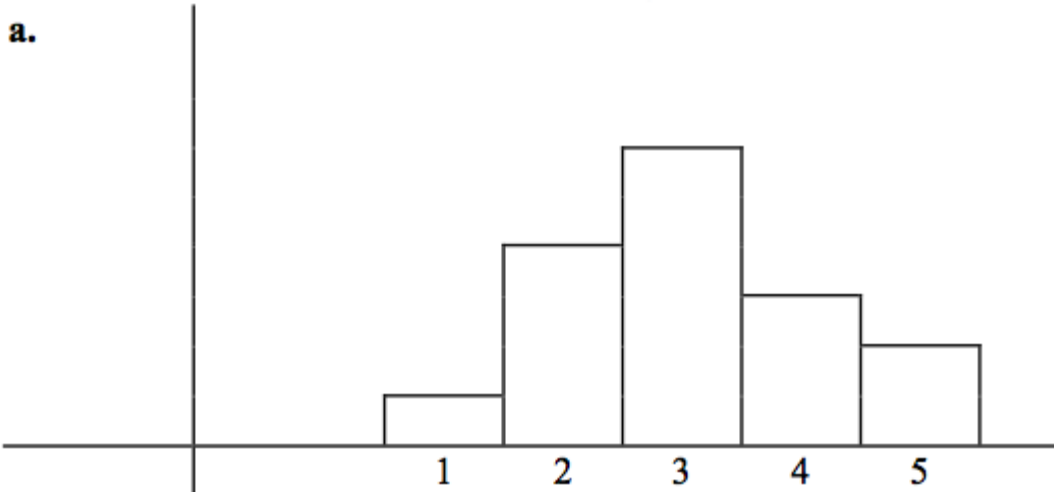
Singles			Couples		
Amount(\$)	Frequency	Rel. Frequency	Amount(\$)	Frequency	Rel. Frequency
51-100	5		100-150	5	
101-150	10		201-250	5	
151-200	15		251-300	5	
201-250	15		301-350	5	
251-300	10		351-400	10	
301-350	5		401-450	10	
			451-500	10	
			501-550	10	
			551-600	5	
			601-650	5	

- Fill in the relative frequency for each group.
- Construct a histogram for the Singles group. Scale the x-axis by \$50. widths. Use relative frequency on the y-axis.
- Construct a histogram for the Couples group. Scale the x-axis by \$50. Use relative frequency on the y-axis.
- Compare the two graphs:
 - List two similarities between the graphs.
 - List two differences between the graphs.
 - Overall, are the graphs more similar or different?
- Construct a new graph for the Couples by hand. Since each couple is paying for two individuals, instead of scaling the x-axis by \$50, scale it by \$100. Use relative frequency on the y-axis.
- Compare the graph for the Singles with the new graph for the Couples:
 - List two similarities between the graphs.
 - Overall, are the graphs more similar or different?
- By scaling the Couples graph differently, how did it change the way you compared it to the Singles?

- h. Based on the graphs, do you think that individuals spend the same amount, more or less, as singles as they do person by person in a couple? Explain why in one or two complete sentences.

EXERCISE 15

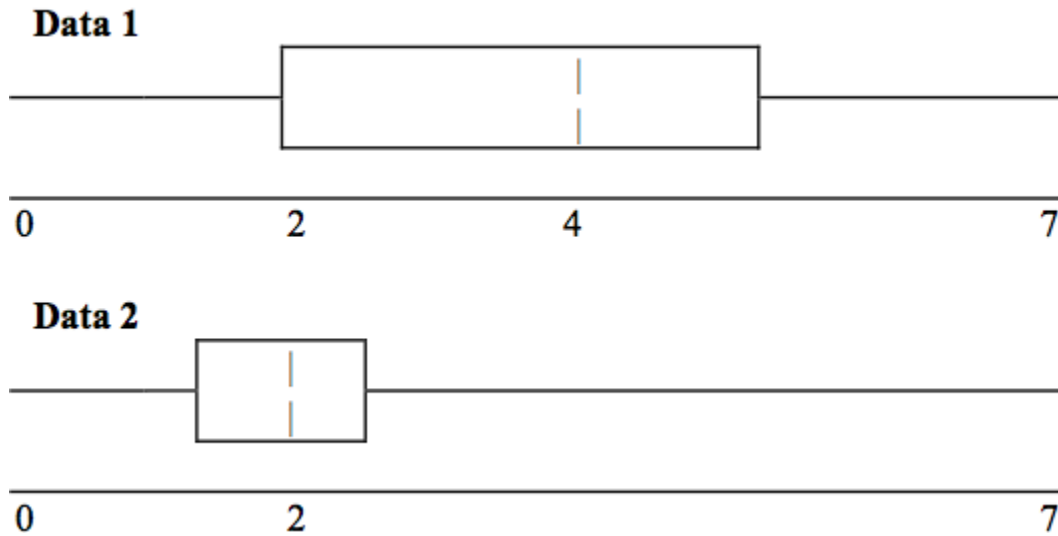
Refer to the following histograms and box plot. Determine which of the following are true and which are false. Explain your solution to each part in complete sentences.



- The medians for all three graphs are the same.
- We cannot determine if any of the means for the three graphs is different.
- The standard deviation for (b) is larger than the standard deviation for (a).
- We cannot determine if any of the third quartiles for the three graphs is different.

EXERCISE 16

Refer to the following box plots.



- In complete sentences, explain why each statement is false.
 - Data 1 has more data values above 2 than Data 2 has above 2.
 - The data sets cannot have the same mode.
 - For Data 1, there are more data values below 4 than there are above 4.
- For which group, Data 1 or Data 2, is the value of "7" more likely to be an outlier? Explain why in complete sentences

EXERCISE 17

In a recent issue of the IEEE Spectrum, 84 engineering conferences were announced. Four conferences lasted two days. Thirty-six lasted three days. Eighteen lasted four days. Nineteen lasted five days. Four lasted six days. One lasted seven days. One lasted eight days. One lasted nine days. Let X = the length (in days) of an engineering conference.

- Organize the data in a chart.
- Find the median, the first quartile, and the third quartile.
- Find the 65th percentile.
- Find the 10th percentile.
- Construct a box plot of the data.
- The middle 50% of the conferences last from _____ days to _____ days.
- Calculate the sample mean of days of engineering conferences.
- Calculate the sample standard deviation of days of engineering conferences.
- Find the mode.
- If you were planning an engineering conference, which would you choose as the length of the conference: mean; median; or mode? Explain why you made that choice.
- Give two reasons why you think that 3 - 5 days seem to be popular lengths of engineering conferences.

EXERCISE 18

A survey of enrollment at 35 community colleges across the United States yielded the following figures (source: Microsoft Bookshelf):

6414; 1550; 2109; 9350; 21828; 4300; 5944; 5722; 2825; 2044; 5481; 5200; 5853; 2750; 10012; 6357; 27000; 9414; 7681; 3200; 17500; 9200; 7380; 18314; 6557; 13713; 17768; 7493; 2771; 2861; 1263; 7285; 28165; 5080; 11622

- Organize the data into a chart with five intervals of equal width. Label the two columns "Enrollment" and "Frequency."
- Construct a histogram of the data.
- If you were to build a new community college, which piece of information would be more valuable: the mode or the average size?
- Calculate the sample average.
- Calculate the sample standard deviation.
- A school with an enrollment of 8000 would be how many standard deviations away from the mean?

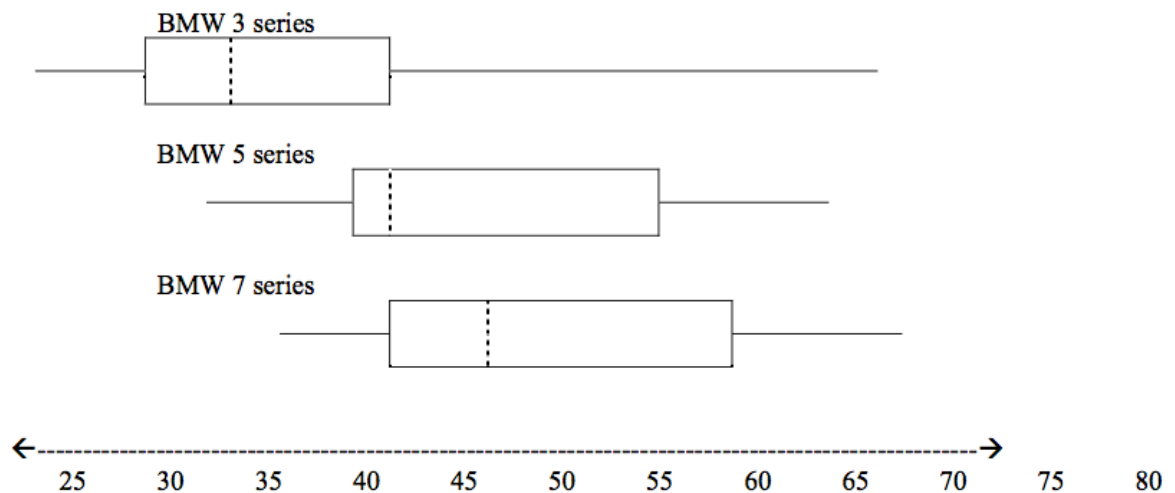
EXERCISE 19

The median age of the U.S. population in 1980 was 30.0 years. In 1991, the median age was 33.1 years. (Source: Bureau of the Census)

- What does it mean for the median age to rise?
- Give two reasons why the median age could rise.
- For the median age to rise, is the actual number of children less in 1991 than it was in 1980? Why or why not?

EXERCISE 20

A survey was conducted of 130 purchasers of new BMW 3 series cars, 130 purchasers of new BMW 5 series cars, and 130 purchasers of new BMW 7 series cars. In it, people were asked the age they were when they purchased their car. The following box plots display the results.

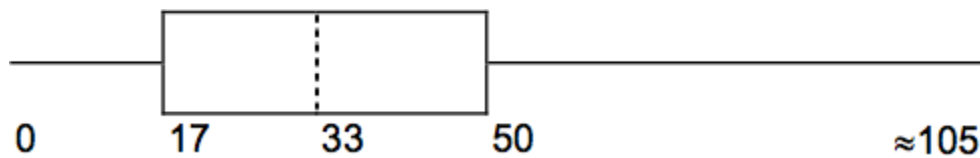


- In complete sentences, describe what the shape of each box plot implies about the distribution of the data collected for that car series.
- Which group is most likely to have an outlier? Explain how you determined that.
- Compare the three box plots. What do they imply about the age of purchasing a BMW from the series when compared to each other?
- Look at the BMW 5 series. Which quarter has the smallest spread of data? What is that spread?
- Look at the BMW 5 series. Which quarter has the largest spread of data? What is that spread?
- Look at the BMW 5 series. Find the Inter Quartile Range (IQR).
- Look at the BMW 5 series. Are there more data in the interval 31-38 or in the interval 45-55? How do you know this?
- Look at the BMW 5 series. Which interval has the fewest data in it? How do you know this?
- 31-35

- ii. 38-41
- iii. 41-64

EXERCISE 21

The following box plot shows the U.S. population for 1990, the latest available year. (Source: Bureau of the Census, 1990 Census)



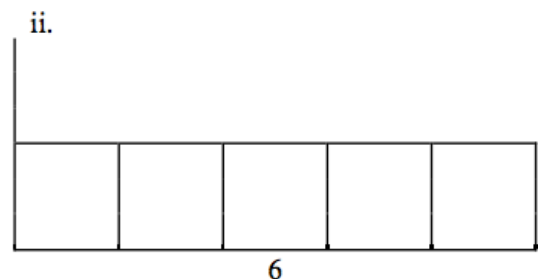
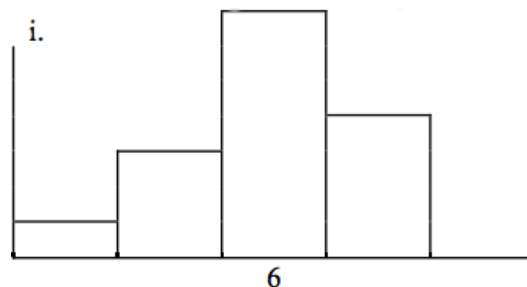
- a. Are there fewer or more children (age 17 and under) than senior citizens (age 65 and over)? How do you know?
- b. 12.6% are age 65 and over. Approximately what percent of the population are of working age adults (above age 17 to age 65)?

EXERCISE 22

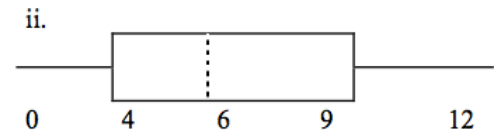
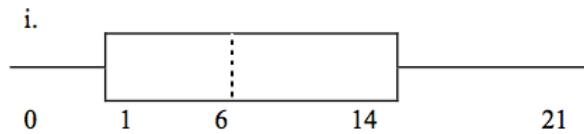
Javier and Ercilia are supervisors at a shopping mall. Each was given the task of estimating the mean distance that shoppers live from the mall. They each randomly surveyed 100 shoppers. The samples yielded the following information:

	Javier	Ercilia
\bar{x}	6.0 miles	6.0 miles
s	4.0 miles	7.0 miles

- a. How can you determine which survey was correct?
- b. Explain what the difference in the results of the surveys implies about the data.
- c. If the two histograms depict the distribution of values for each supervisor, which one depicts Ercilia's sample? How do you know?



- d. If the two box plots depict the distribution of values for each supervisor, which one depicts Ercilia's sample? How do you know?



EXERCISE 23

Student grades on a chemistry exam were:

$\{ 77, 78, 76, 81, 86, 51, 79, 82, 84, 99 \}$

- Construct a stem-and-leaf plot of the data.
- Are there any potential outliers? If so, which scores are they? Why do you consider them outliers?

Try these multiple choice questions (Exercises 24 – 30).

The next three questions refer to the following information. We are interested in the number of years students in a particular elementary statistics class have lived in California.

Number of years	Frequency
7	1
14	3
15	1
18	1
19	4
20	3
22	1
23	1
26	1
40	2
42	2
Total = 20	

EXERCISE 24

What is the IQR?

- a. 8
- b. 11
- c. 15
- d. 35

EXERCISE 25

What is the mode?

- a. 19
- b. 19.5
- c. 14 and 20
- d. 22.65

EXERCISE 26

Is this a sample or the entire population?

- a. Sample
- b. Entire population
- c. Neither

The next two questions refer to the following table. X = the number of days per week that 100 clients use a particular exercise facility.

X	Frequency
0	3
1	12
2	33
3	28
4	11
5	9
6	4

EXERCISE 27

The 80th percentile is:

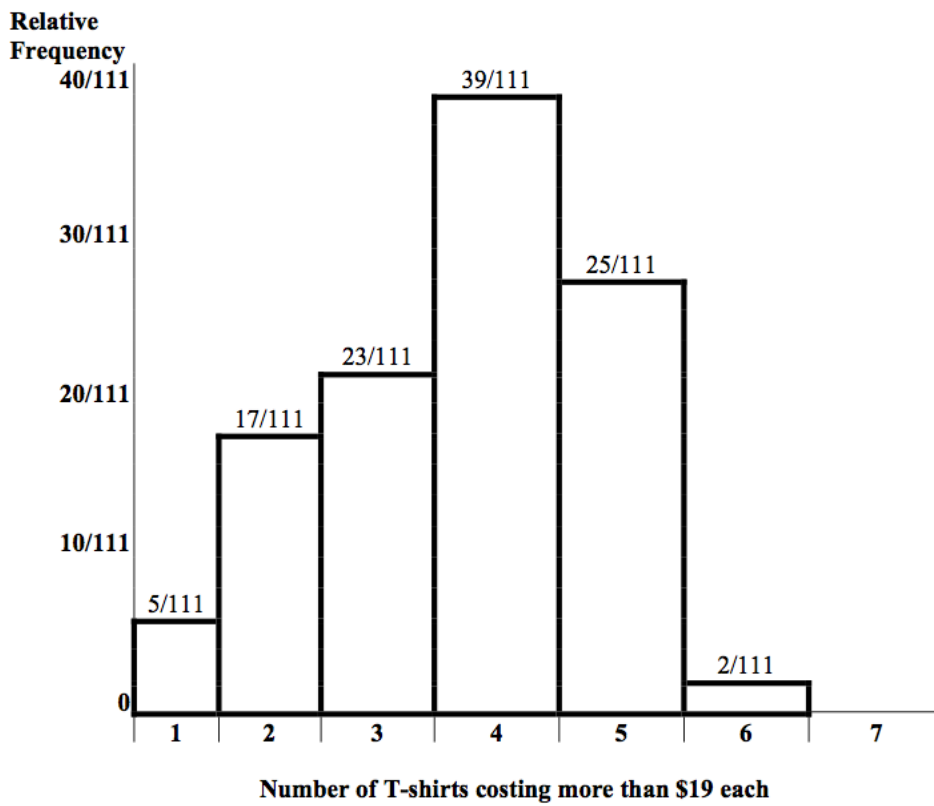
- a. 5
- b. 80
- c. 3
- d. 4

EXERCISE 28

The number that is 1.5 standard deviations BELOW the mean is approximately:

- a. 0.7
- b. 4.8
- c. -2.8
- d. Cannot be determined

The next two questions refer to the following histogram. Suppose one hundred eleven people who shopped in a special T-shirt store were asked the number of T-shirts they own costing more than \$19 each.



EXERCISE 29

The percent of people that own at most three (3) T-shirts costing more than \$19 each is approximately:

- a. 21
- b. 59
- c. 41
- d. Cannot be determined

EXERCISE 30

If the data were collected by asking the first 111 people who entered the store, then the type of sampling is:

- a. Cluster
- b. Simple random
- c. Stratified
- d. Convenience

EXERCISE 31

Below are the 2008 obesity rates by U.S. states and Washington, DC. (Source: <http://www.cdc.gov/obesity/data/trends.html#State>)

State	Percent	State	Percent
Alabama	31.4	Montana	23.9
Alaska	26.1	Nebraska	26.6
Arizona	24.8	Nevada	25
Arkansas	28.7	New Hampshire	24
California	23.7	New Jersey	22.9
Colorado	18.5	New Mexico	25.2
Connecticut	21	New York	24.4
Delaware	27	North Carolina	29
Washington DC	21.8	North Dakota	27.1
Florida	24.4	Ohio	28.7
Georgia	27.3	Oklahoma	30.3
Hawaii	22.6	Oregon	24.2
Idaho	24.5	Pennsylvania	27.7
Illinois	26.4	Rhode Island	21.5
Indiana	26.3	South Carolina	30.1
Iowa	26	South Dakota	27.5
Kansas	27.4	Tennessee	30.6

Kentucky	29.8	Texas	28.3
Louisiana	28.3	Utah	22.5
Maine	25.2	Vermont	22.7
Maryland	26	Virginia	25
Massachusetts	20.9	Washington	25.4
Michigan	28.9	West Virginia	31.2
Minnesota	24.3	Wisconsin	25.4
Mississippi	32.8	Wyoming	24.6
Missouri	28.5		

- Construct a bar graph of obesity rates of your state and the four states closest to your state. Hint: The x-axis is labeled with the state names.
- Use a random number generator to randomly pick 8 states. Construct a bar graph of the obesity rates of those 8 states.
- Construct a bar graph for all the states beginning with the letter "A."
- Construct a bar graph for all the states beginning with the letter "M."

EXERCISE 32

A music school has budgeted to purchase 3 musical instruments. They plan to purchase a piano costing \$3000, a guitar costing \$550, and a drum set costing \$600. The average cost for a piano is \$4,000 with a standard deviation of \$2,500. The average cost for a guitar is \$500 with a standard deviation of \$200. The average cost for drums is \$700 with a standard deviation of \$100. Which cost is the lowest, when compared to other instruments of the same type? Which cost is the highest when compared to other instruments of the same type. Justify your answer numerically.

EXERCISE 33

Suppose that a publisher conducted a survey asking adult consumers the number of fiction paperback books they had purchased in the previous month. The results are summarized in the table below. (Note that this is the data presented for publisher B in homework exercise 13).

# of books	Freq.	Rel. Freq.
0	18	
1	24	
2	24	
3	22	
4	15	
5	10	
7	5	
9	1	

- Are there any outliers in the data? Use an appropriate numerical test involving the IQR to identify outliers, if any, and clearly state your conclusion.

- b. If a data value is identified as an outlier, what should be done about it?
- c. Are any data values further than 2 standard deviations away from the mean? In some situations, statisticians may use this criteria to identify data values that are unusual, compared to the other data values. (Note that this criteria is most appropriate to use for data that is mound-shaped and symmetric, rather than for skewed data.)
- d. Do parts (a) and (c) of this problem give the same answer?
- e. Examine the shape of the data. Which part, (a) or (c), of this question gives a more appropriate result for this data?
- f. Based on the shape of the data which is the most appropriate measure of center for this data: mean, median or mode?