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		

a. Find the sample mean, $\overline{x}$.
b. Find the sample standard deviation, s.
c. Construct a histogram of the data.
d. Complete the columns of the chart.
e. Find the first quartile.
f. Find the median.
g. Find the third quartile.
h. Construct a box plot of the data.
i. What percent of the students saw fewer than three movies?
j. Find the 40th percentile.
k. Find the 90th percentile.
l. Construct a line graph of the data.
m. 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).

a. Based upon this information, give two reasons why the black median age could be lower than the white median age.
b. Does the lower median age for blacks necessarily mean that blacks die younger than whites? Why or why not?
c. 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		

a. Find the sample mean $\overline{x}$
b. Find the sample standard deviation, ss
c. Construct a histogram of the data.
d. Complete the columns of the chart.
e. Find the first quartile.
f. Find the median.
g. Find the third quartile.
h. Construct a box plot of the data.
i. What percent of the students owned at least five pairs?
j. Find the 40th percentile.
k. Find the 90th percentile.
l. Construct a line graph of the data.
m. 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

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

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

EXERCISE 6

Given the following box plot:

$$\begin{array}{ccccc} 0 & 20 & & 100 & 150 \end{array}$$
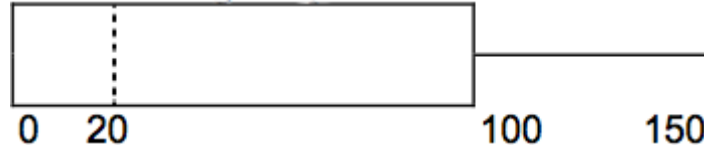
   a. Think of an example (in words) where the data might fit into the above box plot. In 2-5 sentences, write down the example.

   b. What does it mean to have the first and second quartiles so close together, while the second to fourth quartiles are far apart?
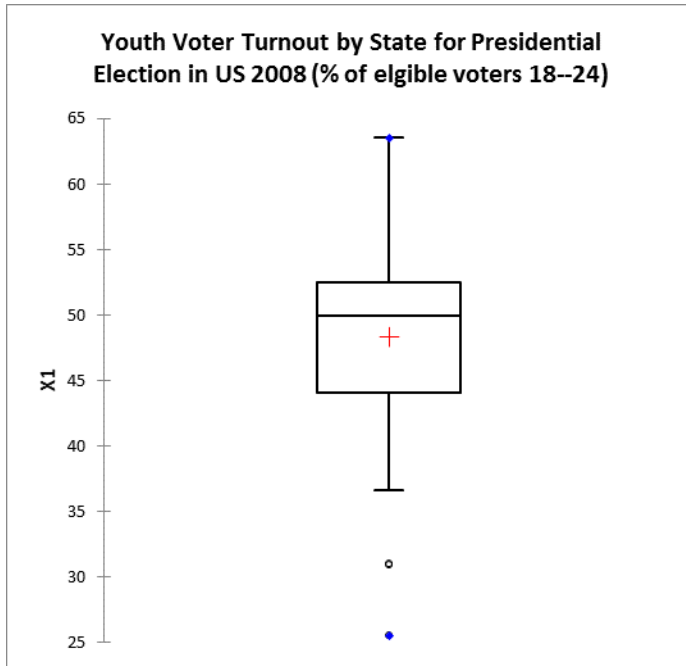
EXERCISE 7

Below are the gross earning of eleven movies in millions of dollars. Construct an outlier boxplot for the data. Include the five number summary.

Movie	Gross earnings in Millions of Dollars
Alone in the Dark	5
Eternal Sunshine	34
Big Fish	66
Collateral	100
Vanilla Sky	101
Last Samurai	111
The Village	114
Break-Up	116
S.W.A.T.	117
DaVinci Code	213
Pirate of the Carribbean (ii)	322

EXERCISE 8

The box plot and descriptive statistics for the United States Youth Voter turnout below (source: US Census Bureau) is from the 2008 presidential election. Based on the data given answer the following questions:

   a. 62.9% of the Minnesota youth (18-24) voted during the 2008 election. Which quartile contains the Minnesota data?

   b. The US average youth turnout was 48.5%, which quartile contains the US overall average percent youth voter turnout?

   c. The 6 lowest and 6 highest youth turnout states were: AR (31.0), GA (25.5), IA (63.5), ME (54.7), MN (62.9), NH (57.7), OH (57), OK (41.5), TN (41.4), TX (36.6), UT (30.9), WI (57.5).

      a. Are all of these States outliers? If so WHY? If not, are any of them outliers? Be specific.

      b. Are any of the states "far outliers"? If so state which ones and why you believe they are far outliers.
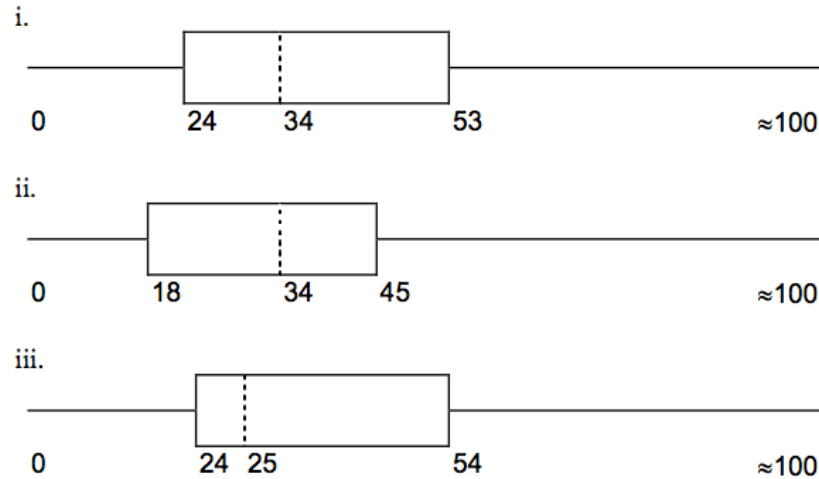
Youth Voter Turnout by State for Presidential Election in US 2008 (% of elgible voters 18--24)

Statistic	X1
No. of observations	41
Minimum	25.5000
Maximum	63.5000
1st Quartile	44.1000
Median	49.9000
3rd Quartile	52.500
Mean	48.3488
Variance (n-1)	61.6996
Standard deviation (n-1)	7.8549

EXERCISE 9

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

a. Construct a histogram of the Japanese-American community in Santa Clara What percent of the community is under age 35?

b. Which box plot most resembles the information above?

i.

```
0        24    34        53              ≈100
```

ii.

```
0     18        34    45                 ≈100
```

iii.

```
0        24  25        54               ≈100
```

EXERCISE 10

The following summary statistics are for the number of pairs of jeans students in your class own.

   Minimum = 0, Q1 = 4, median = 5, Q3 = 6, and maximum = 8

From this we know that
a. there are no outliers in the data.
b. there is at least one low outlier in the data.
c. there is at least one high outlier in the data.
d. None of the above.

EXERCISE 11

Refer to the following box plots.

Data 1

| 0 | 2 | 4 | 7 |

Data 2

| 0 | 2 | 7 |
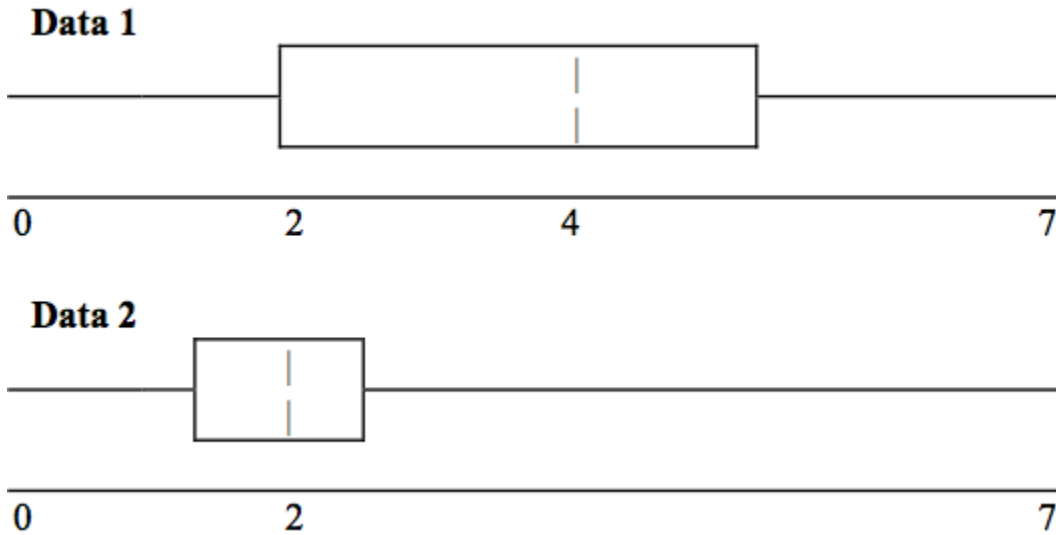
a. In complete sentences, explain why each statement is false.
   i. Data 1 has more data values above 2 than Data 2 has above 2.
   ii. The data sets cannot have the same mode.
   iii. For Data 1, there are more data values below 4 than there are above 4.
b. For which group, Data 1 or Data 2, is the value of "7" more likely to be an outlier? Explain why in complete sentences

EXERCISE 12

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.

a. Organize the data in a chart.
b. Find the median, the first quartile, and the third quartile.
c. Find the 65th percentile.

d. Find the 10th percentile.
   e. Construct a box plot of the data.
   f. The middle 50% of the conferences last from _____ days to _____ days.
   g. Calculate the sample mean of days of engineering conferences.
   h. Calculate the sample standard deviation of days of engineering conferences.
   i. Find the mode.
   j. 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.
   k. Give two reasons why you think that 3 - 5 days seem to be popular lengths of engineering conferences.

EXERCISE 13

Construct the outlier boxplot for the exam scores of 13 students in your statistics class.
48, 55, 63, 68, 75, 76, 76, 78, 78, 81, 87, 89, 93.

Write a complete description of the analysis of the exam score data. Remember that a complete description of numerical data includes comments on shape, center, and spread.

EXERCISE 14

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

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

EXERCISE 15

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)

    a. What does it mean for the median age to rise?
    b. Give two reasons why the median age could rise.
    c. 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 16

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

| 0 | 17 | 33 | 50 | ≈105 |
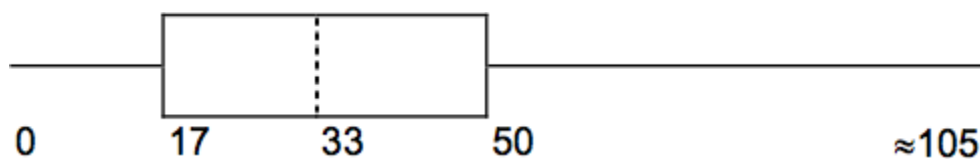
    a. Are there fewer or more children (age 17 and under) than senior citizens (age 65 and over)? How do you know?
    b. 12.6% are age 65 and over. Approximately what percent of the population are of working age adults (above age 17 to age 65)?

Try these multiple choice questions (Exercises 17 – 23).

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 17

What is the IQR?

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

EXERCISE 18

What is the mode?

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

EXERCISE 19

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 20
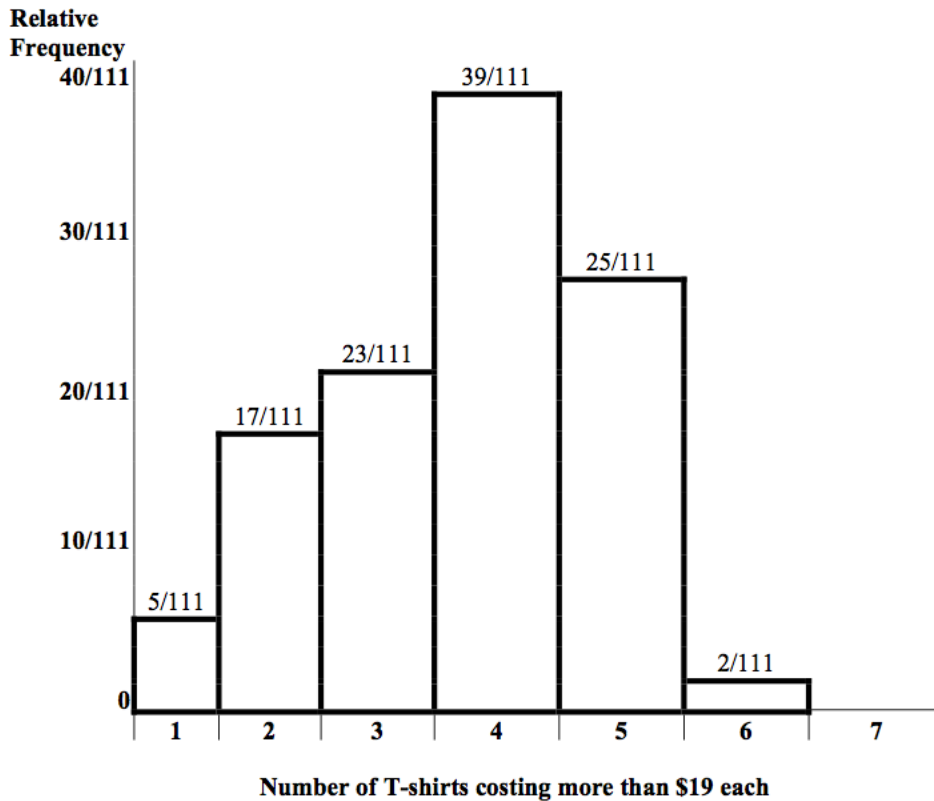
The 80th percentile is:

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

EXERCISE 21

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.

Relative Frequency

Number of T-shirts costing more than $19 each

(Histogram values: 1 → 5/111, 2 → 17/111, 3 → 23/111, 4 → 39/111, 5 → 25/111, 6 → 2/111; y-axis gridlines at 10/111, 20/111, 30/111, 40/111)

EXERCISE 22

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 23

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 24

Below are the 2008 obesity rates by U.S. states and Washington, DC. (Source:

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		

a. 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.

b. Use a random number generator to randomly pick 8 states. Construct a bar graph of the obesity rates of those 8 states.

c. Construct a bar graph for all the states beginning with the letter "A."

d. Construct a bar graph for all the states beginning with the letter "M."