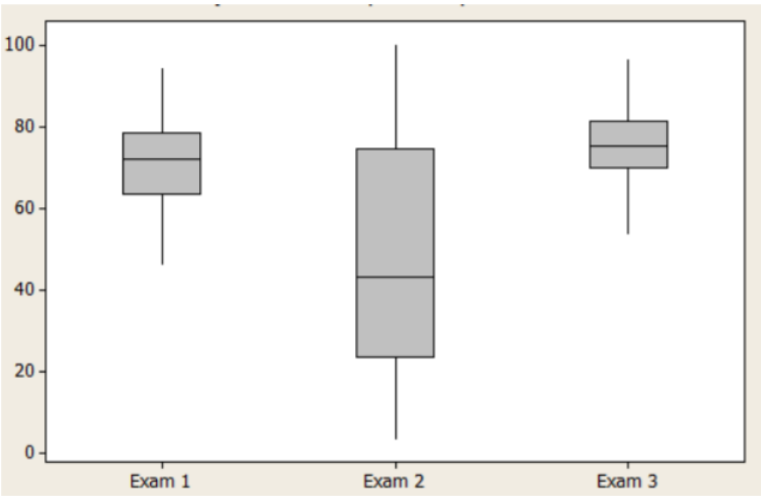
15.2.2: Chapter 3 Homework

1. A poll was taken of 150 students at De Anza College. Students were asked how many hours they work outside of college. The students were interviewed in the morning between 8 AM and 11 AM on a Thursday. The sample mean for these 150 students was 9.2 hours.
   a. What is the Population?
   b. What is the Sample?
   c. Does the 9.2 hours represent a statistic or parameter? Explain.
   d. Is the sample mean of 9.2 a reasonable estimate of the mean number of hours worked for all students at De Anza? Explain any possible bias.
2. The box plots represent the results of three exams for 40 students in a Math course.



   a. Which exam has the highest median?
   b. Which exam has the highest standard deviation?
   c. For Exam 2, how does the median compare to the mean?
   d. In your own words, compare the exams.
3. Examine the following average daily commute time (minutes) for residents of two cities.

City A	2	4	4	4	4	5	7	9	13	14	16	16	16	18	19	19	**Sample mean = 29.06**
	21	21	21	27	30	35	37	38	47	48	50	59	70	72	87	97	**Sample Std Dev = 25.35**
City B	29	38	38	40	40	48	48	50	52	52	54	55	56	57	57	58	**Sample mean = 57.00**
	58	58	59	59	59	62	62	63	66	66	67	69	69	71	75	89	**Sample Std Dev = 12.12**

   a. Compute and interpret the z-score for a 75-minute commute for City A.
   b. Compute and interpret the z-score for a 75-minute commute for City B.
   c. For which group would a 75 -minute commute be more unusual? Explain.
4. The February 10, 2017 Nielsen ratings of 20 TV programs shown on commercial television, all starting between 8 PM and 10 PM, are given below:

2.1	2.3	2.5	2.8	2.8	3.6	4.4	4.5	5.7	7.6
7.6	8.1	8.7	10.0	10.2	10.7	11.8	13.0	13.6	17.3

   a. Obtain the sample mean and median. Do you believe that the data is symmetric, right-skewed or left skewed?
   b. Determine the sample variance and standard deviation.
   c. Assuming the data are bell shaped, between which two numbers would you expect to find 68% of the data?

5. The following data represents recovery time for 16 patients (arranged in a table to help you out).

count	Days (X)	$X - \overline{X}$	$(X - \overline{X})^2$	Z Score
#1	2			
#2	3			
#3	4			
#4	4			
#5	5			
#6	5			
#7	5			
#8	5			
#9	5			
#10	6			
#11	6			
#12	7			
#13	7			
#14	8			
#15	8			
#16	16			
Totals				

a. Calculate the sample mean and median
b. Use the table to calculate the variance and standard deviation.
c. Use the range of the data to see if the standard deviation makes sense. (Range should be between 3 and 6 standard deviations).
d. Using the empirical rule between which two numbers should you expect to see 68% of the data? 95% of the data? 99.7% of the data?
e. Calculate the Z-score for observation. Do you think any of these data are outliers?

6. The following data represents the heights (in feet) of 20 almond trees in an orchard.

14	14	14	14	15	18	18	20	21	21
22	24	25	25	25	27	27	29	31	45

a. Construct a box plot of the data.
b. Do you think the tree with the height of 45 feet is an outlier? Use the box plot method to justify your answer.

7. The following average daily commute time (in minutes) for residents of 2 cities are shown in the table.

City A	2	4	4	4	4	5	7	9	13	14	16	16	16	18	19	19
	21	21	21	27	30	35	37	38	47	48	50	59	70	72	87	97
City B	29	38	38	40	40	48	48	50	52	52	54	55	56	57	57	58
	58	58	59	59	59	62	62	63	66	66	67	69	69	71	75	89

a. Find the quartiles and interquartile range for each group.
b. Calculate the 80th percentile for each group.
c. Construct side-by-side box plots, and compare the two groups

8. Rank the following correlation coefficients from weakest to strongest.

.343, -.318, .214, -.765, 0, .998, -.932, .445

9. If you were trying to think of factors that affect health care costs:
   a. Choose a variable you believe would be positively correlated with health care costs.
   b. Choose a variable you believe would be negatively correlated with health care costs.
   c. Choose a variable you believe would be uncorrelated with health care costs.
