STATISTICS FOR MANAGEMENTS(IDS 570)

HOMEWORK 1 SOLUTION

Problem 1

(a) Quantitative
(b) Categorical
(c) Categorical
(d) Quantitative
(e) Categorical

Problem 2

The stem-and-leaves display of median pay is:

Stem	Leaf
6	6 7 7
7	2 4 6 7 7 8 9
8	0 0 1 3 7
9	9
10	0 6
11	0
12	1

The median pay for these careers is generally in the \$70 and \$80 thousands. Only four careers have a median pay of \$100 thousand or more. The highest median pay is \$121 thousand for a finance director. Similarly, the stem-and-leaves display of top pay is:

Stem	Leaf
10	0 6 9
11	1 6 9
12	2 5 6
13	0 5 8 8
14	0 6
15	2 5 7
16	
17	
18	
19	
20	
21	4
22	1

The most frequent top pay is in the \$130 thousand range. However, the top pay is rather evenly distributed between \$100 and \$160 thousand. Two unusually high top pay values occur at \$214 thousand for a finance director and \$221 thousand for an investment banker. Also, note that the top pay has more variability than the median pay.

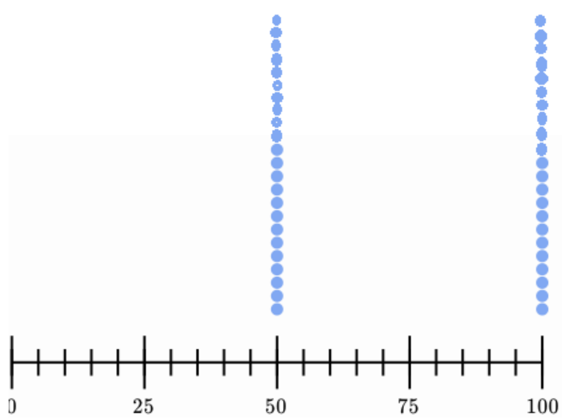
R code:
```
> prob2 ← read.xlsx(file =“Assignment 1 excel file.xlsx”, sheetIndex = 1, header = T)
> stem(prob2$Median_Pay, scale = 2)
> stem(prob2$Top_Pay, scale = 4)
```

Problem 3

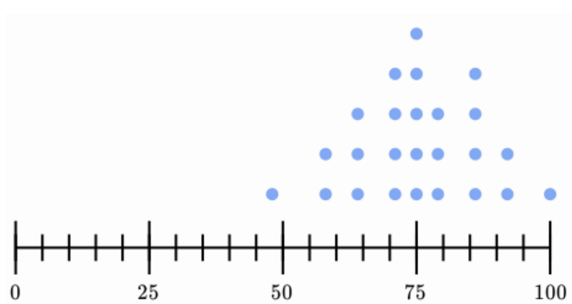
(a) No. In Yale's class, 23 students got 50% and 22 students got 100%. The mean is $\dfrac{(23 \times 50) + (22 \times 100)}{45} \approx 75\%$.
But there were 23 scores of 50% and 22 scores of 100%, so the middle, or median score was 50%. Therefore, the mean is greater than the median.

(b) No. In a normal distribution, the median is typically close to the mean, so we expect both the mean and the median score in UIC's class to be around 75%.

(c) No. The standard deviation measures how spread our the data are from the mean. This plot shows the distribution of Yale's scores.



UIC's class scores might look like this:



Therefore, the UIC's score are much more tightly packed around the mean than Yale's scores.

Problem 4

(a) $\bar{x} = \frac{\sum_i x_i}{n} = \frac{350}{19} = 18.42$

(b) $\bar{x} = \frac{\sum_i x_i}{n} = \frac{120}{19} = 6.31$

(c) $\frac{120}{350}(100) = 34.3\%$ of 3-point shots were made from the 20 feet, 9 inch line during the 19 games.

(d) Moving the 3-point line back to 20 feet, 9 inches has reduced the number of 3-point shots taken per game from 19.07 to 18.42, or 19.07 - 18.42 = .65 shots per game. The percentage of 3-points made per game has been reduced from 35.2% to 34.3%, or only .9%. The move has reduced both the number of shots taken per game and the percentage of shots made per game, but the differences are small. The data support the Associated Press Sports conclusion that the move has not changed the game dramatically.
The 2008-09 sample data shows 120 3-point baskets in the 19 games. Thus, the mean number of points scored from the 3-point line is 120(3)/19 = 18.95 points per game. With the previous 3-point line at 19 feet, 9 inches, 19.07 shots per game and a 35.2% success rate indicate that the mean number of points scored from the 3-point line was 19.07(.352)(3) = 20.14 points per game. There is only a mean of 20.14 - 18.95 = 1.19 points per game less being scored from the 20 feet, 9 inch 3-point line.

R code:
```
> prob4 ← read.xlsx(file =“Assignment 1 excel file.xlsx”, sheetIndex = 2, header = T)
> mean(x = prob4$3PointShots)
> mean(x = prob4$ShotsMade)
> sum(prob4$ShotsMade)/sum(prob4$3PointShots) * 100
```

Problem 5

(a) $\bar{x} = \frac{\sum_i x_i}{n} = \frac{148}{10} = 14.8$

(b) Order the data from low 6.7 t high 36.6. Using 5^{th} and 6^{th} positions, we have: Median = $\frac{10.1 + 16.1}{2} = 13.1$.

(c) Mode = 7.2 (occurs 2 times)

(d) Use 3^{th} position, $Q_1 = 72$ and using 8^{th} position, $Q_3 = 172$

(e) $\sum x_i = \$148$ billion. The percentage of total endowments held by these 2.3% of colleges and universities is (148/413)(100) = 35.8%.

R code:
```
> prob5 ← read.xlsx(file =“Assignment 1 excel file.xlsx”, sheetIndex = 3, header = T)
> mean(prob5$Endowment)
> median(prob5$Endowment)
> count ← table(prob5$Endowment)
> names(count)[which.max(count)]
> quantile(x = prob5$Endowment, c(0.25, 0.75))
> sum(prob5$Endowment)/413 *100
```

Problem 6

(a) n = 20; So to find Median we look at the 10^{th} and 11^{th} positions of the sorted data

$$\text{Median} = \frac{73+74}{2} = 73.5$$

(b) We have

$$\text{samllest value: } 68$$

$$Q_1 = \frac{\overbrace{71+72}^{5^{th} \text{ and } 6^{th} \text{ positions}}}{2} = 71.5$$

$$Q_2 = 73.5$$

$$Q_3 = \frac{\overbrace{74+75}^{15^{th} \text{ and } 16^{th} \text{ positions}}}{2} = 74.5$$

$$\text{largest} = 77$$

5- number summary: 68, 71.5, 73.5, 74.5, 77

(c) IQR $= Q_3 - Q_1 = 74.5 - 71.5 = 3$, Lower Limit $= Q_1 - 1.5\cdot$ (IQR) $= 71.5 \;\; 1.5 \cdot (3) = 67$, Upper Limit $= Q_3 + 1.5\cdot$ (IQR) $= 74.5 + 1.5 \cdot (3) = 79$

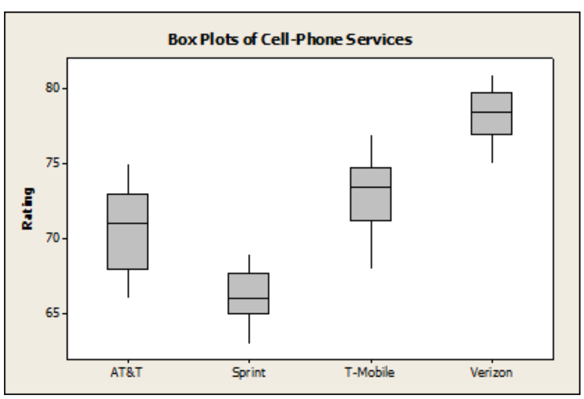
All ratings are between 67 and 79. There are no outliers for the T-Mobile service.

(d) Using the solution procedures shown in parts a, b, and c, the five number summaries and outlier limits for the other three cell-phone services are as follows.

AT&T	66, 68, 71, 73, 75	Limits: 60.5 and 80.5
Sprint	63, 65, 66, 67.5, 69	Limits: 61.25 and 71.25
Verizon	75, 77, 78.5, 79.5, 81	Limits: 73.25 and 83.25

There are no outliers for any of the cell-phone services.

(e) The box plots show that Verizon is the best cell-phone service provider in terms of overall customer satisfaction. Verizons lowest rating is better than the highest AT&T and Sprint ratings and is better than 75% of the T-Mobile ratings. Sprint shows the lowest customer satisfaction ratings among the four services.



R code:

```
> prob6 ← read.xlsx(file = "Assignment 1 excel file.xlsx", sheetIndex = 4, header = T)
> median(prob6$TMobile)
> summary(prob6$TMobile)
> summary(prob6$ATT)
> summary(prob6$Sprint)
> boxplot(prob6[,2:5], main= "Box Plots of Cell-Phone Services", ylab = "Rating")
```

Problem 7

(a) Using empirical rule, we know 95% of data are between $\mu - 2\sigma = 30 - 2(5) = 20$ and $\mu + 2\sigma = 30 + 2(5) = 40$. In addition, 99.7% of data are between $\mu - 3\sigma = 30 - 3(5) = 15$ and $\mu + 3\sigma = 30 + 3(5) = 45$

(b) $P(x \leq 820) = P(\frac{x-1000}{210} \leq \frac{820-1000}{210}) = P(z \leq -0.86) = 0.1949$

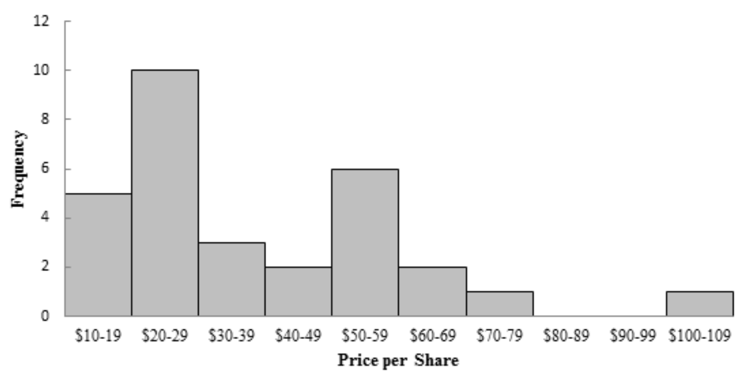
Problem 8

(a) The highest price stock is for IBM with a price of \$109 per share. The lowest price stock is for Alcoa with a price of \$11 per share.

(b) A class size of 10 results in 10 classes.

Price per Share	Count	Frequency (%)
\$11–20	5	16.67
\$ 21–30	10	33.33
\$ 31–40	3	10.00
\$ 41–50	2	6.67
\$ 51–60	6	20.00
\$ 61–70	2	6.67
\$ 71–80	1	3.33
\$ 81–90	0	0
\$ 91–100	0	0
\$ 101–110	1	3.33

(c) The general shape of the distribution is skewed to the right. Half of the companies (15) have a price per share less than \$30. A mid-priced stock appears to be in the \$30 to\$49 range, while the most frequently priced stock is in the \$20 to \$29 range. Five stocks are less than \$20 per share (Alcoa, Bank of America, General Electric, Intel and Pfizer). Four stocks are \$60 or more per share (3M, Chevron, ExxonMobil and IBM).



R code:

```
> prob8 ← read.xlsx(file = "Assignment 1 excel file.xlsx", sheetIndex = 5, header = T)
> hist(prob8$share, xlab = "Price per Share", ylab = "Frequency", col = "gray")
```