Stat 145 Homework Solutions: Chapter 2

Problem 2.1

(a) The highway mileages sum to 570, so the mean is:

$$\bar{x} = \frac{570}{22} = 25.91$$

(b) Without the outlier, the mean is:

$$\bar{x} = \frac{514}{22} = 24.48$$

Including the outlier increases the value of the mean.

Problem 2.4

Including all seasons, the mean is 36.06 home runs and the median is 34 home runs. Dropping the 73-homer season, the mean drops to 33.75 home runs and the median remains 34 home runs. The general fact that is illustrated is that the mean is sensitive to outliers while the median is resistant to them.

Problem 2.7

(a) The stemplot appears below:

4	23
4	667899
5	0111112244444
5	555566677778
6	0111144
6	589

The distribution is roughly symmetric, so the median should be about the same as the mean.

(b) The mean is 54.8 years and the five-number summary is:

$$42 \quad 51 \quad 55 \quad 58 \quad 69$$

The median and mean are very close in value, as was expected.

(c) The middle half is between $Q_1 = 51$ years and $Q_3 = 58$ years. Bill Clinton is in the youngest 25% percent because $46 < Q_1$.

Problem 2.8

(a) The mean is:

$$\bar{x} = \frac{5.6 + 5.2 + 4.6 + 4.9 + 5.7 + 6.4}{6} = \frac{32.4}{6} = 5.4$$

(b) The standard deviation is:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
5.6	0.2	0.04
5.2	-0.2	0.04
4.6	-0.8	0.64
4.9	-0.5	0.25
5.7	0.3	0.09
6.4	1	1
32.4	0	2.06

$$\begin{aligned} s &= \sqrt{\frac{2.06}{5}} \\ &= 0.6419 \end{aligned}$$

Problem 2.9

For Data A, $\bar{x}_A = 7.501$ and $s_A = 2.032$, while for Data B $\bar{x}_B = 7.501$ and $s_B = 2.031$. The stemplots appear below:

Data A

3	1
4	7
5	
6	1
7	2
8	1177
9	112

Data B

5	257
6	58
7	079
8	48
9	
10	
11	
12	5

Data A is skewed to the left while Data B has a high outlier.

Problem 2.11

Income distributions will be skewed to the right, so the mean should be larger than the median. Thus the mean is \$72,674 and the median is \$53,054.

Problem 2.14

Five-number summaries:

Player	Min	Q_1	M	Q_3	Max
Ruth	22	35	46	54	60
McGwire	9	29	39	52	70

The boxplots hide the fact that McGwire's minimum seasons are outliers from the rest of his distribution.

Problem 2.29

(a) Choose four numbers that are exactly the same (e.g. 1, 1, 1, 1).

(b) 0, 0, 10, 10

(c) Part (a) has more than once choice, as any set of four identical numbers will give $s = 0$. Part (b) has only one choice. We want to make the numbers as spread out as possible, so 0 and 10 seem to be reasonable choices. We also want to make the squared differences from the mean – $(x_i - \bar{x})^2$ – as large as possible. If we choose 0, 10, 10 , 10 or 10, 0, 0, 0 we make one of the squared deviations 7.5^2 but the other three are only 2.5^2. Choosing 0, 0, 10, 10 makes all four squared deviations equal to 5^2.