STAT 213, Section 05

Solutions to Assignment No. 1

2.54 The median is the middle number once the data have been arranged in order. If n is even, there is not a single middle number. Thus, to compute the median, we take the average of the middle two numbers. If n is odd, there is a single middle number. The median is this middle number.

A data set with 5 measurements arranged in order is 1, 3, 5, 6, 8. The median is the middle number, which is 5.

A data set with 6 measurements arranged in order is 1, 3, 5, 5, 6, 8. The median is the average of the middle two numbers which is $\frac{5+5}{2}=\frac{10}{2}=5$.

2.56 a. $\bar{x}=\frac{\sum x}{n}=\frac{7+\cdots+4}{6}=\frac{15}{6}=2.5$

Median $=\frac{3+3}{2}=3$ (mean of 3rd and 4th numbers, after ordering)

Mode = 3

b. $\bar{x}=\frac{\sum x}{n}=\frac{2+\cdots+4}{13}=\frac{40}{13}=3.08$

Median = 3 (7th number, after ordering)

Mode = 3

c. $\bar{x}=\frac{\sum x}{n}=\frac{51+\cdots+37}{10}=\frac{496}{10}=49.6$

Median $=\frac{48+50}{2}=49$ (mean of 5th and 6th numbers, after ordering)

Mode = 50

2.66 a. The mean number of ant species discovered is:

$$\bar{x} = \frac{\sum x}{n} = \frac{3+3+...+4}{11} = \frac{141}{11} = 12.82$$

The median is the middle number once the data have been arranged in order:
3, 3, 4, 4, 4, 5, 5, 5, 7, 49, 52.

The median is 5.

The mode is the value with the highest frequency. Since both 4 and 5 occur 3 times, both 4 and 5 are modes.

b. For this case, we would recommend that the median is a better measure of central tendency than the mean. There are 2 very large numbers compared to the rest. The mean is greatly affected by these 2 numbers, while the median is not.

c. The mean total plant cover percentage for the Dry Steppe region is:

$$\bar{x} = \frac{\sum x}{n} = \frac{40+52+...+27}{5} = \frac{202}{5} = 40.4$$

The median is the middle number once the data have been arranged in order:
27, 40, 40, 43, 52.

The median is 40.

The mode is the value with the highest frequency. Since 40 occurs 2 times, 40 is the mode.

d. The mean total plant cover percentage for the Gobi Desert region is:

$$\bar{x} = \frac{\sum x}{n} = \frac{30+16+...+14}{6} = \frac{168}{6} = 28$$

The median is the mean of the middle 2 numbers once the data have been arranged in order: 14, 16, 22, 30, 30, 56.

The median is $\frac{22+30}{2} = \frac{52}{2} = 26$.

The mode is the value with the highest frequency. Since 30 occurs 2 times, 30 is the mode.

e. Yes, the total plant cover percentage distributions appear to be different for the 2 regions. The percentage of plant coverage in the Dry Steppe region is much greater than that in the Gobi Desert region.

2.68 a. The mean number of power plants is:

$$\bar{x} = \frac{\sum x}{n} = \frac{5+3+\ldots+3}{20} = \frac{80}{20} = 4$$

The median is the mean of the middle 2 numbers once the data have been arranged in order: 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 5, 6, 7, 9, 13

The median is $\frac{3+3}{2} = \frac{6}{2} = 3$.

There are 3 numbers that each occur 4 times. They are 1, 2, and 5. Thus, there are 3 modes, 1, 2, and 5.

b. Deleting the largest number, 13, the new mean is:

$$\bar{x} = \frac{\sum x}{n} = \frac{5+3+\ldots+3}{19} = \frac{67}{19} = 3.526$$

The median is the middle number once the data have been arranged in order: 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 5, 6, 7, 9

The median is 3.

There are 3 numbers that each occur 4 times. They are 1, 2, and 5. Thus, there are 3 modes, 1, 2, and 5.

By dropping the largest measurement from the data set, the mean drops from 4 to 3.526. The median and the modes stay the same. There is no effect on them.

c. Deleting the lowest 2 and highest 2 measurements leaves the following:

1, 1, 2, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 5, 6, 7

The new mean is:

$$\bar{x} = \frac{\sum x}{n} = \frac{1+1+\cdots+7}{16} = \frac{56}{16} = 3.5$$

The trimmed mean has the advantage that some possible outliers have been eliminated.

2.76 a. Range = 4 − 0 = 4

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{22 - \frac{8^2}{5}}{4-1} = 2.3 \qquad s = \sqrt{2.3} = 1.52$$

b. Range = 6 − 0 = 6

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{63 - \frac{17^2}{7}}{7-1} = 3.619 \qquad s = \sqrt{3.619} = 1.90$$

c. Range = 8 − (−2) = 10

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{154 - \frac{30^2}{18}}{10-1} = 7.111 \qquad s = \sqrt{7.111} = 2.67$$

d. Range = 2 − (−3) = 5

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{29 - \frac{(-5)^2}{18}}{18-1} = 1.624 \qquad s = \sqrt{1.624} = 1.274$$

2.80 a. $\sum x = 3 + 1 + 10 + 10 + 4 = 28$

$\sum x^2 = 3^2 + 1^2 + 10^2 + 10^2 + 4^2 = 226$

$$\bar{x} = \frac{\sum x}{n} = \frac{28}{5} = 5.6$$

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{226 - \frac{28^2}{5}}{5-1} = \frac{69.2}{4} = 17.3 \qquad s = \sqrt{17.3} = 4.1593$$

b. $\sum x = 8 + 10 + 32 + 5 = 55$

$\sum x^2 = 8^2 + 10^2 + 32^2 + 5^2 = 1213$

$$\bar{x} = \frac{\sum x}{n} = \frac{55}{4} = 13.75 \text{ feet}$$

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{1213 - \frac{55^2}{4}}{4-1} = \frac{456.75}{3} = 152.25 \text{ square feet}$$

$s = \sqrt{152.25} = 12.339$ feet

c. $\sum x = -1 + (-4) + (-3) + 1 + (-4) + (-4) = -15$

$\sum x^2 = (-1)^2 + (-4)^2 + (-3)^2 + 1^2 + (-4)^2 + (-4)^2 = 59$

$$\bar{x} = \frac{\sum x}{n} = \frac{-15}{6} = -2.5$$

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{59 - \frac{(-15)^2}{6}}{6-1} = \frac{21.5}{5} = 4.3 \qquad s = \sqrt{4.3} = 2.0736$$

(continued)

2.80 (continued)

d. $\sum x = \frac{1}{5}+\frac{1}{5}+\frac{1}{5}+\frac{2}{5}+\frac{1}{5}+\frac{4}{5} = \frac{10}{5} = 2$

$\sum x^2 = \left(\frac{1}{5}\right)^2+\left(\frac{1}{5}\right)^2+\left(\frac{1}{5}\right)^2+\left(\frac{2}{5}\right)^2+\left(\frac{1}{5}\right)^2+\left(\frac{4}{5}\right)^2 = \frac{24}{25} = .96$

$\bar{x} = \frac{\sum x}{n} = \frac{2}{6} = \frac{1}{3} = .33$ ounce

$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{\frac{24}{25} - \frac{2^2}{6}}{6-1} = \frac{.2933}{5} = .0587$ square ounce

$s = \sqrt{.0587} = .2422$ ounce

2.126 For Sample A, the 21 ordered observations are:

85, 100, 121, 142, 145, 157, 158, 159, 161, 163, 165,
166, 170, 171, 171, 172, 172, 173, 184, 187, 196

$\frac{1}{4}(21) = 5.25 \uparrow 6$, so $Q_1 = 157$

$\frac{1}{2}(21) = 10.5 \uparrow 11$, so $Q_2 = \text{med} = 165$

$\frac{3}{4}(21) = 15.75 \uparrow 16$, so $Q_3 = 172$

$IQR = Q_3 - Q_1 = 172 - 157 = 15$

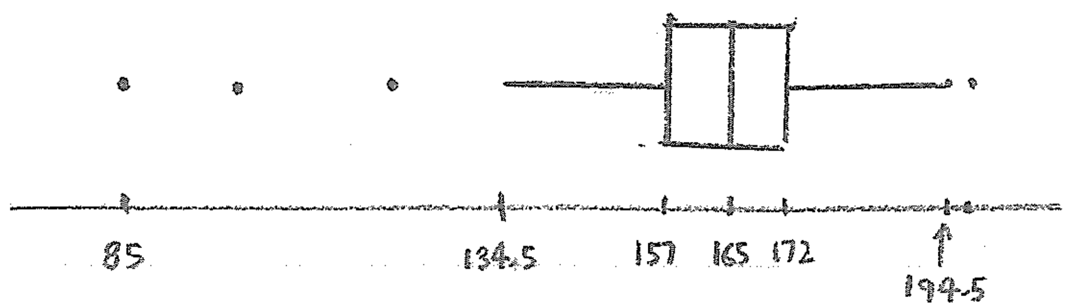
$Q_3 + 1.5\, IQR = 172 + 22.5 = 194.5$

$Q_1 - 1.5\, IQR = 157 - 22.5 = 134.5$

(continued)

2.126 (continued)

So the boxplot for Sample A is:



(b) Defining an "outlier" as any observation more than 1.5 IQR further beyond Q_1 or Q_3, the outliers in Sample A are 85, 100, 121, 196.

(The text may use a different definition of "outlier".)

Additional statistics for Sample A:

(.10) 21 = 2.1 ↑ 3, so 10^{th} percentile is 121
(.40) 21 = 8.4 ↑ 9, so 40^{th} " is 161
.90 (21) = 18.9 ↑ 19, so 90^{th} " is 184.

I also obtained (using a program, not a hand calculation) that, for Sample A, $\bar{x} = 158.0$, $s = 27.11$.

(I did not have time to do any calculations for Sample B).

2.130 (c) The ordered observations are:

2.25, 2.55, 2.73, 2.73, 2.95, 3.06, 3.13, 3.21

3.23, 3.27, 3.30, 3.32, 3.37, 3.38, 3.60, 3.75

3.81, 3.85, 3.88, 3.90, 4.05, 4.06, 4.09, 4.09

4.56, 5.06

$26(\frac{1}{4}) = 6.5 \uparrow 7$, so $Q_1 = \underline{3.13}$

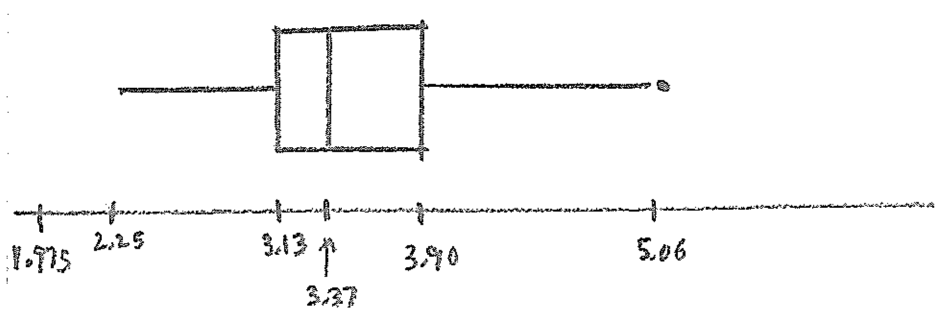
$26(\frac{1}{2}) = 13$, so med $= (3.37+3.38)/2 = \underline{3.375}$

$26(\frac{3}{4}) = 19.5 \uparrow 20$, so $Q_3 = \underline{3.90}$.

$IQR = Q_3 - Q_1 = 3.90 - 3.13 = \underline{0.77}$

$Q_3 + 1.5\ IQR = 3.90 + 1.155 = \underline{5.055}$

$Q_1 - 1.5\ IQR = 3.13 - 1.155 = 1.\underline{975}$



(continued)

2.130 (continued)

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{26} = 3.507 \qquad s = 0.634$$

$.3(26) = 7.8 \uparrow 8$, so 30^{th} percentile $= 3.21$

$.8(26) = 16.8 \uparrow 17$, so 80^{th} percentile $= 3.81$.