Stat 151.003 Fall 2006 (Kozdron)
Solutions to Assignment #2

1. (a) The measures of central tendency (mean, median, mode) are easily seen to be as follows:

	Mean	Median	Mode	Minimum	Maximum	Q1	Q3
Stock I	7	7	5,9	3	11	4.75	9.25
Stock II	7	7	8,13	1	13	3.5	10.75
Stock III	7	7	none	-3	16	1.0	13.25

All three stocks have the same centre of data—each has mean 7 and median 7. However, the data for Stock I are clustered more tightly around the mean/median. This is evident because it has the narrowest IQR. Stock II is a little more spread out around the centre, while Stock III has the greatest range and IQR.

1. (b) We calculate the standard deviations as follows:

- Stock I:
$$s^2 = \frac{1}{9}\left[(4-7)^2 + (5-7)^2 + \cdots + (10-7)^2\right] = 2.75,$$
- Stock II:
$$s^2 = \frac{1}{9}\left[(4-7)^2 + (10-7)^2 + \cdots + (13-7)^2\right] = 4.19,$$
- Stock III:
$$s^2 = \frac{1}{9}\left[(5-7)^2 + (8-7)^2 + \cdots + (-3-7)^2\right] = 6.62.$$

These standard deviations show that Stock I has less fluctuations in rate-of-return than Stocks II or III. Stock I data is clustered more tightly, on the average, around the centre of data ($\bar{X} = MD = 7$) then either Stock II or Stock III. Stock II is more tightly clustered than Stock III. This observation agrees with the observations of central tendency noted in part **(a)**.

2. (a) There are 47 observations, and so the median is the 24th observation when the data are arranged in increasing order. Thus, $MD = 79$.

The first quartile is the median of the observations strictly less than 79. Since there are 23 such observations, the first quartile is the 12th observation when the data are arranged in increasing order. Thus, $Q1 = 68$.

The third quartile is the median of the observations strictly greater than 79. Since there are 23 such observations, the third quartile is the $(23+1+12) = 36$th observation when the data are arranged in increasing order. Thus, $Q3 = 86$.

The interquartile range is thus given by $IQR = Q3 - Q1 = 86 - 68 = 18$.

2. (b) The frequency table for this data is as follows:

Score	Frequency	Relative Frequency
$10 \leq x < 20$	4	$4/47 = 0.085$
$20 \leq x < 30$	2	$2/47 = 0.043$
$30 \leq x < 40$	3	$3/47 = 0.064$
$40 \leq x < 50$	1	$1/47 = 0.021$
$50 \leq x < 60$	0	$0/47 = 0$
$60 \leq x < 70$	3	$3/47 = 0.064$
$70 \leq x < 80$	12	$12/47 = 0.255$
$80 \leq x < 90$	12	$12/47 = 0.255$
$90 \leq x < 100$	10	$10/47 = 0.213$

2. (c) We find the mean for this data set is

$$\bar{X} = \frac{1}{47}(12 + 14 + 18 + \cdots + 2(95) + 97 + 98) = \frac{3304}{47} = 70.30.$$

The variance is given by

$$S^2 = \frac{1}{46}\left[(12 - 70.3)^2 + (14 - 70.3)^2 + \cdots + (98.70.3)^2\right] = 608.12$$

and so the standard deviation is

$$S = \sqrt{S^2} = \sqrt{608.12} = 24.66.$$

2. (d) The stem-and-leaf plot is given by

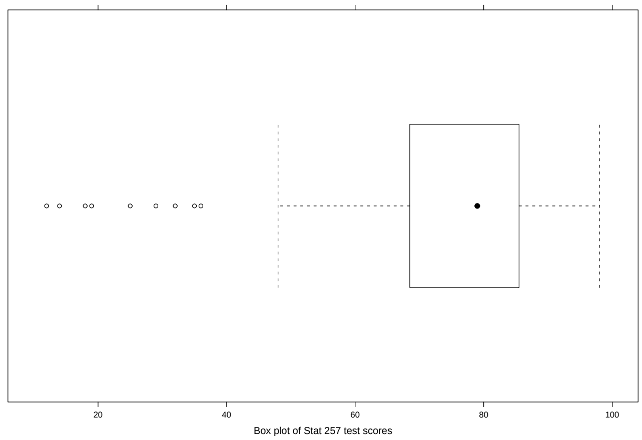
1	2 4 8 9
2	5 9
3	2 5 6
4	8
5	
6	7 8 9
7	0 1 2 2 3 5 5 6 8 8 **9** 9
8	0 0 1 1 3 4 4 4 5 5 6 6
9	0 0 2 2 3 3 5 5 7 8

The median is indicated in bold face.

2. (e) The five-number summary for this data set is:

$$(\text{min}, Q1, \text{median}, Q3, \text{max}) = (12, 68, 79, 86, 98).$$

2. (f) The box plot is as follows:



2. (g) From looking at the frequency table, the stem-and-leaf plot, and the box-plot, the distribution of the test scores is skewed to the left. The data is not "nicely and evenly" distributed around the centre of the data. It has a long, straggly tail to the left which can be easily seen from the box-plot and the five-number summary.

3. Note that 16 oz = 1 lb = 0.4563 kg = 453.6 g. Therefore, 1 oz = 28.35 g. Now, if we wish to express our package weights in grams, all we need to do is scale them by the factor 28.35. However, such a scale is a linear transformation. Thus, under a multiplicative transformation, the effect on S is natural; that is, it undergoes the same scaling. Therefore,

$$S = 2.5\text{oz} \times 28.35\text{g/oz} = 70.87\text{g}.$$

4. If the median is \$48,500 and each employee receives a \$500 increment plus 3% of current salary, then since each employee receives a comparable increase, the order of the data points will not change, and thus neither will the median's location. That is, if $X_n < MD < X_{n+1}$, then $1.03X_n + 500 < 1.03MD + 500 < 1.03X_{n+1} + 500$. Hence, the new median is

$$MD_{\text{new}} = 50,455.$$

5. **(Page 133 #41)** Recall that when data are normally distributed, we expect 68% of data to lie within one standard deviation of the mean. That is, we expect 68% Americans to consume between 22.6 pounds and 31 pounds of citrus fruit per year. Since the normal distribution is symmetric about its mean, we expect the same amount to lie below 22.6 as above 31, namely $(100 - 68)/2 = 16\%$. (Draw a picture if this is confusing.)

(Page 133 #43) For this data set we find the mean is $\bar{X} = 214.97$ and the standard deviation is $S = 20.76$. Notice that **all 30** of the observations lie within 2 standard deviations of the mean. (That is, all 30 observations lie between 173.45 and 256.48.) Recall that Chebychev's theorem tells us that in the worst case scenario, at least 75% of data points will lie within 2 standard deviations of the mean. This result is therefore consistent with Chebychev's result. Indeed, at least 75% **do** lie within 2 SDs of the mean!