36-201 Spring 1999
Solutions to Homework 2

1. `Siegel and Morgan, 5 (p.51).` The numbers that are 24 or higher are counted in the far right column in the histogram. Reading from the y-axis, we learn that 3 numbers in this distribution are 24 or higher.

   `Siegel and Morgan, 6 (p.51).` There are 3 out of 24 numbers that are greater than 24, that is a 3/24*100 % = 12.5 % of the distribution.

   `Siegel and Morgan, 7 (p.51).` The third column from left to right counts the numbers between 16 and 23. From the y-axis we read that there are 7 numbers in this range, that is a 7/24*100% = 29.17 % of the distribution.

2. `Siegel and Morgan, 27 (p.62).` The distribution looks unimodal and skewed toward low values (skewed to the left). Most of the distribution is concentrated between 12 and 17 years old. The minimum age is at 1 year old while the maximum age is at 19 years old. There seem to be no outliers.

   `Siegel and Morgan, 28 (p.62).` The distribution looks multimodal and skewed toward high values (skewed to the right). Most of the distribution concentrates between living in 1 to 35 different places. There are three large values that are possibly outliers.

   `Siegel and Morgan, 29 (p.62).` The chart reads as follows:

Mother only:	48.28%
Both parents:	24.14%
Relatives:	24.14%
Mother and boyfriend:	3.45%
Group Home:	0.00%

   Almost half of the children with reported problems are children who live only with their Mother. The remaining half is made up mostly by children living either with Both Parents or with Other Relatives, these two percentages being equal to each other, which is somewhat surprising. There is only a 3.45% of children reporting problems who live with their Mother and her Boyfriend and no children with problems living in a Group Home.

3. `Moore, 4.54 (p.257).`

   (a) The mean of this data set is

   $$\frac{4+0+1+4+3+6}{6} = \frac{18}{6} = 3.$$

   The variance is

   $$S^2 = \frac{1^2 + (-3)^2 + (-2)^2 + 1^2 + 0^2 + 3^2}{5} = \frac{24}{5} = 4.8,$$

   so the standard deviation is $S = \sqrt{4.8} = 2.19$.

   (b) The mean of this data set is

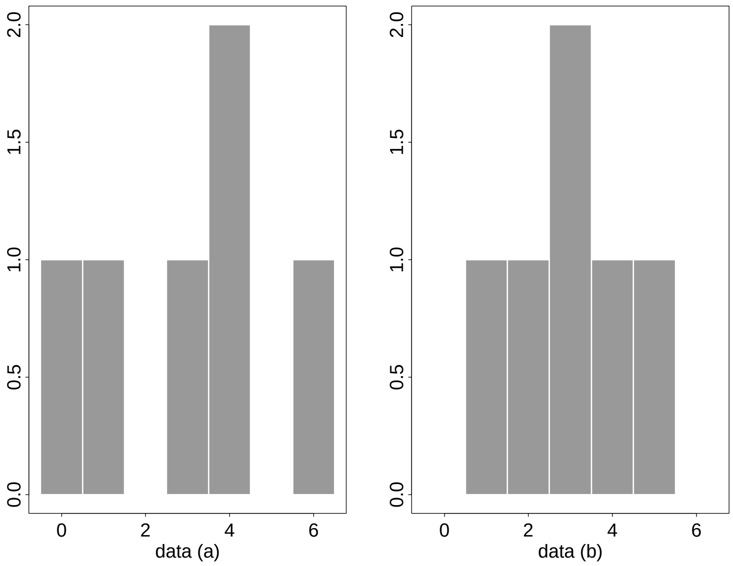
   $$\frac{5+3+1+3+4+2}{6} = \frac{18}{6} = 3.$$

The variance is

$$S^2 = \frac{2^2 + 0^2 + (-2)^2 + 0^2 + 1^2 + (-1)^2}{5} = \frac{10}{5} = 2,$$

so the standard deviation is $S = \sqrt{2} = 1.41.$

Group (a) is more spread out since it has a greater standard deviation. This can also be seen in the histograms below. Note that, however, both groups have the same mean.



`Moore, 4.56 (p.258).`

(a) The mean of the new data set is

$$\frac{6 + 2 + 3 + 6 + 5 + 8}{6} = \frac{30}{6} = 5.$$

The variance is

$$S^2 = \frac{1^2 + (-3)^2 + (-2)^2 + 1^2 + 0^2 + 3^2}{5} = \frac{24}{5} = 4.8,$$

so the standard deviation is $S = \sqrt{4.8} = 2.19.$

(b) Adding 2 to each point in the data set increases the mean by 2. However the standard deviation remains unchanged.

(c) The mean will increase by 10, that is, the new data set will have mean 3 + 10 = 13. The standard deviation will remain the same, 2.19.

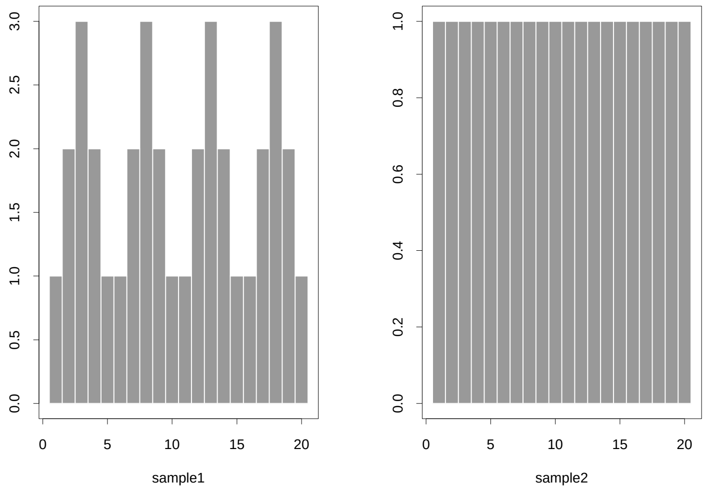
4. `Moore, 4.58 (p.258).` The variance is the square of the standard deviation, so the variance of the IQ test is $SD^2 = 15^2 = 225$

`Moore, 4.59 (p.258).` No collection of descriptive statistics of a data set can completely describe its shape. So in both cases, the histograms can have different shapes. In the following example, the two data sets have the same five number summary statistics:

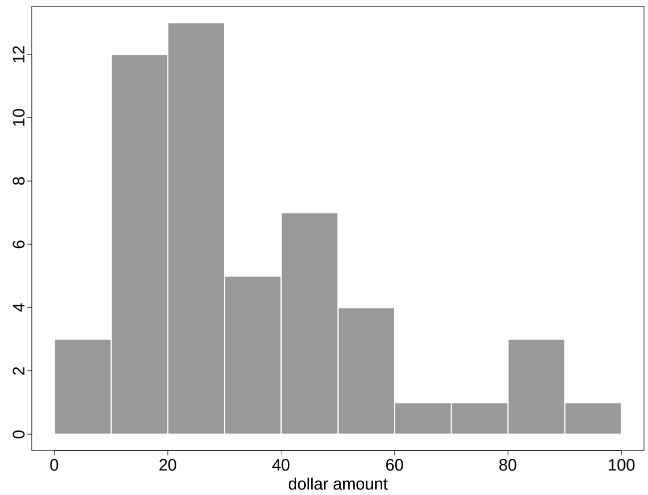
```
Sample 1:  1  5  6 10 11 15 16 20  2  4  7  9 12 14 17 19  2  4  7
           9 12 14 17 19  3  8 13 18  3  8 13 18  3  8 13 18

Sample 2:  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
```

Both data sets have `Min = 1`, `Q1 = 5.5`, `Median = 10.5`, `Q3 = 15.5` `Max = 20`. However, the shapes of their histograms are quite different from each other.



`Moore, 4.62 (p.259).` The histogram below shows that the distribution is skewed (to the right), so we prefer the five number summary as a better description of the data.



Note that the data in the book are already ordered from smallest to largest, so it will be easy to find the five number summary. There are 50 observations in the data set.

Median: There is an even number of data values, so the median is the average of the two middle numbers. That is, the average of the 25-th and 26-th observations, $(27.65 + 28.06)/2 = 27.85$

Divide the data into two groups, one of them including observations 1 to 25 (lower half) and the other one including observations 26 to 50 (upper half). Both halves have an odd number of data points. So,

Q1: The first quartile is the middle observation of the lower half, that is, observation 13. Q1 =19.27.

Q3: The third quartile is the middle observation of the upper half, that is, observation 38. Q3 = 45.40.

Min: The minimum value is 3.11.

Max: The maximum value is 93.34.

5. `Moore, 4.64 (p.272).`

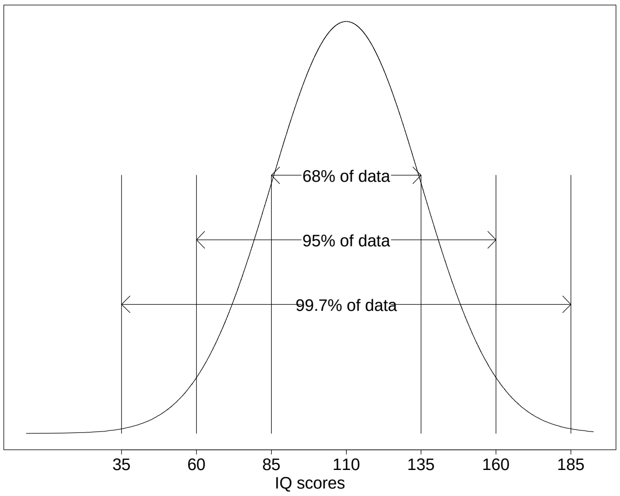
   (a) The Mode is at A, since that is the maximum height of the density. The distribution is skewed right so we know that the mean should be greater than the median. So the Median is at B and the Mean is at C.

   (b) This distribution is unimodal and symmetric, so the Mean, the Median and the Mode are at the center point A.

   (c) The Mode is at C, since that is the maximum height of the density. The distribution is skewed left so we know that the mean should be smaller than the median. So the Median is at B and the Mean is at A.

   `Moore, 4.65 (p.272).` The Mean is at the maximum height, 28. The standard deviation can be calculated looking for the distance from the mean to the change of curvature points, those are 27 and 29. So the standard deviation is 1.

6. `Moore, 4.71 (p.274).` The 68-95-99.7 rule says that 68% of the distribution is between $110 \pm 25 = (85,\ 135)$, 95% of the distribution is between $110 \pm 2 \times 25 = (60,\ 160)$ and 99.7% of the distribution is between $110 \pm 3 \times 25 = (35,\ 185)$. This can be seen in the figure below.



(a) In the normal case, the Mean 110 is also the Median. So a 50% of the people have scores above 110.

(b) 5% of the data are below 60 or above 160. The normal distribution is symmetric, then we know that 2.5% of the people are above 160.

(c) 32% of the data are below 85 or above 135. The normal distribution is symmetric, then we know that 16% of the people are below 85.