CHAPTER 6

Section 6-1

6-1. No, usually not. For example, if the sample is {2, 3} the mean is 2.5 which is not an observation in the sample.

6-2. No, it is easy to construct a counter example. For example, {1, 2, 3, 1000}.

6-3. No, usually not. For example, the mean of {1, 4, 4} is 3 which is not even an observation in the sample.

6-4. Yes. For example, {9,10,11}, the sample mean = 10, sample standard deviation = 1.

6-5. Yes. For example, {5, 5, 5, 5, 5, 5, 5}, the sample mean = 5, sample standard deviation = 0

6-6. The mean is increased by 5 and the standard deviation is not changed. Try it!

6-7. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{592.037}{8} = 74.0046 \text{ mm}$$

Sample variance:

$$\sum_{i=1}^{8} x_i = 592.037$$

$$\sum_{i=1}^{8} x_i^2 = 43813.47632$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{43813.47632 - \frac{(592.037)^2}{8}}{8-1}$$

$$= \frac{0.0001489}{7} = 0.000021271 \quad (\text{mm})^2$$

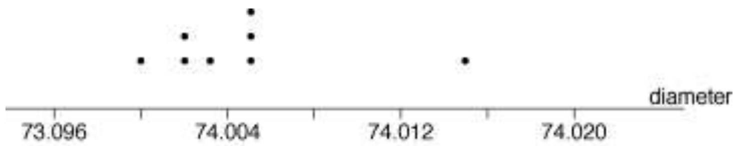
Sample standard deviation:

$$s = \sqrt{0.000021271} = 0.00461 \quad \text{mm}$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \quad \text{where} \quad \sum_{i=1}^{8} (x_i - \bar{x})^2 = 0.0001489$$

Dot Diagram:



There appears to be a possible outlier in the data set.

6-8. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{19} x_i}{19} = \frac{273.16}{19} = 14.377 \text{ min}$$

Sample variance:

$$\sum_{i=1}^{19} x_i = 273.16$$

$$\sum_{i=1}^{19} x_i^2 = 10337.64$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{10337.64 - \frac{(273.16)^2}{19}}{19-1}$$

$$= \frac{6410.46}{18} = 356.14 \text{ (min)}^2$$

Sample standard deviation:

$$s = \sqrt{356.14} = 18.87 \text{ min}$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{19} (x_i - \bar{x})^2 = 6410.46$$

6-9. Sample average:

$$\bar{x} = \frac{84707}{12} = 7058.92 \quad yards$$

Sample variance:

$$\sum_{i=1}^{12} x_i = 84707$$

$$\sum_{i=1}^{19} x_i^2 = 598534033$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{598534033 - \frac{(84707)^2}{12}}{12-1}$$

$$= \frac{594378.9}{11} = 54034.45 \quad (yards)^2$$

Sample standard deviation:

$$s = \sqrt{54034.45} = 232.45 \quad yards$$

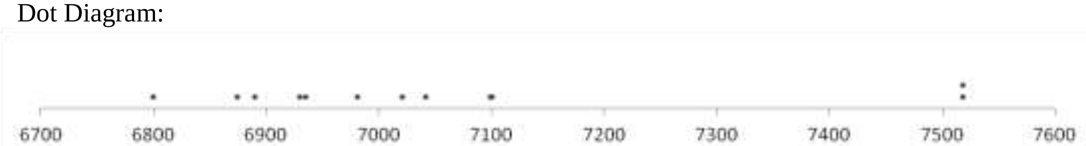
The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{12} (x_i - \bar{x})^2 = 594378.9$$

Dot Diagram:



6-10. Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{18} x_i}{18} = \frac{2276}{18} = 126.44 \quad \text{kN}$$

Sample variance:

$$\sum_{i=1}^{18} x_i = 2276$$

$$\sum_{i=1}^{18} x_i^2 = 299432$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{299432 - \frac{(2276)^2}{18}}{18-1}$$

$$= \frac{11644.44}{17} = 684.97 \quad (\text{kN})^2$$

Sample standard deviation:

$$s = \sqrt{684.97} = 26.17 \quad \text{kN}$$

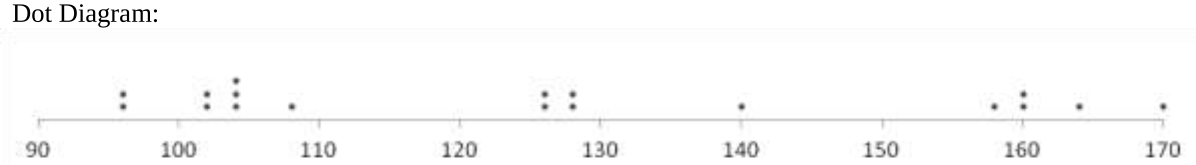
The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{18} (x_i - \bar{x})^2 = 11644.44$$

Dot Diagram:



6-11. Sample average:

$$\bar{x} = \frac{351.27}{8} = 43.91$$

Sample variance:

$$\sum_{i=1}^{8} x_i = 351.27$$

$$\sum_{i=1}^{19} x_i^2 = 16536.95$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{16536.95 - \frac{(351.27)^2}{8}}{8-1}$$

$$= \frac{1113.12}{7} = 159.02$$

Sample standard deviation:

$$s = \sqrt{159.02} = 12.61$$

The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = 1113.12$$

Dot Diagram:

6-12. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{35} x_i}{35} = \frac{28367}{35} = 810.486 \text{ watts/m}^2$$

Sample variance:

$$\sum_{i=1}^{35} x_i = 28368$$
$$\sum_{i=1}^{35} x_i^2 = 23550565$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{23550565 - \frac{(28367)^2}{35}}{35-1} = \frac{559516.743}{34}$$
$$= 16456.37 \quad (\text{watts/m}^2)^2$$

Sample standard deviation:

$$s = \sqrt{16456.37} = 128.28 \quad \text{watts/m}^2$$

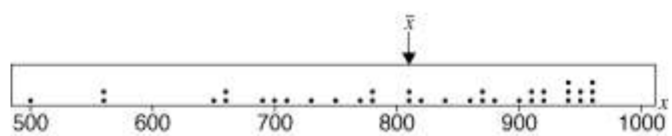
The sample standard deviation could also be found using

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where

$$\sum_{i=1}^{35}(x_i - \bar{x})^2 = 559516.743$$

Dot Diagram (rounding of the data is used to create the dot diagram)



The sample mean is the point at which the data would balance if it were on a scale.

6-13.

$$\mu = \frac{6978}{1255} = 5.56$$

The value 5.56 is the population mean because the actual physical population of all flight times during the operation is available.

6-14. Sample average:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{19.57}{9} = 2.174 \quad \text{mm}$$

Sample variance:

$$\sum_{i=1}^{9} x_i = 19.57$$

$$\sum_{i=1}^{9} x_i^2 = 45.958$$

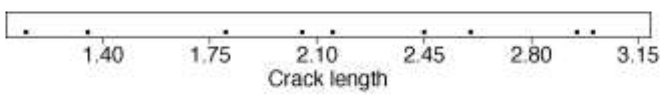
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{45.958 - \frac{(19.57)^2}{9}}{9-1} = \frac{3.404}{8}$$

$$= 0.4255 \quad (\text{mm})^2$$

Sample standard deviation:

$$s = \sqrt{0.4255} = 0.6523 \quad \text{mm}$$

Dot Diagram



6-15. Sample average of exercise group:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3455.68}{12} = 287.97$$

Sample variance of exercise group:

$$\sum_{i=1}^{n} x_i = 3455.68$$

$$\sum_{i=1}^{n} x_i^2 = 1119122$$

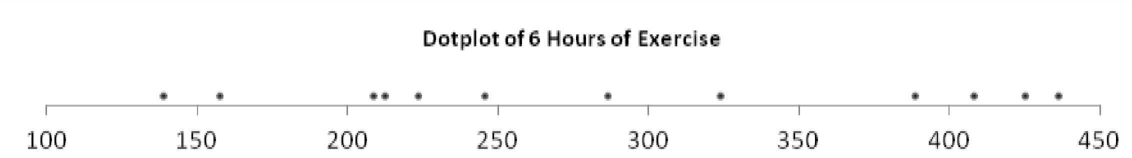
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{1119122 - \frac{(3455.68)^2}{12}}{12-1}$$

$$= \frac{123978.9}{11} = 11270.81$$

Sample standard deviation of exercise group:

$$s = \sqrt{11270.81} = 106.16$$

Dot Diagram of exercise group:



Sample average of no exercise group:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2601.44}{8} = 325.18$$

Sample variance of no exercise group:

$$\sum_{i=1}^{n} x_i = 2601.44$$

$$\sum_{i=1}^{n} x_i^2 = 948309.4$$

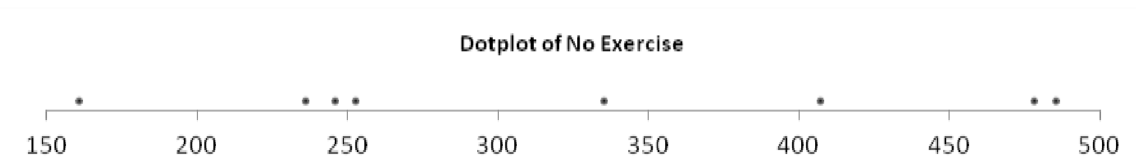
$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{948309.4 - \frac{(2601.44)^2}{8}}{8-1}$$

$$= \frac{102373.7}{7} = 14624.82$$

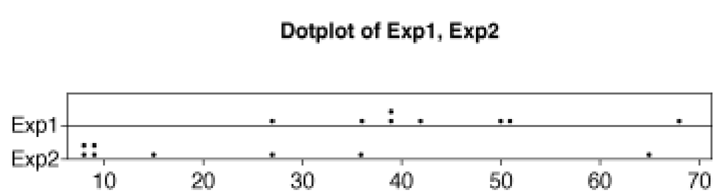
Sample standard deviation of no exercise group:

$$s = \sqrt{14624.82} = 120.93$$

Dot Diagram of no exercise group:



6-16. Dot Diagram of CRT data in exercise 6-11 (Data were rounded for the plot)



The data are centered a lot lower in the second experiment. The lower CRT resolution reduces the visual accommodation.

6-17. $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{57.5}{8} = 7.188$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{413.284 - \frac{(57.5)^2}{8}}{8-1} = \frac{0.00275}{7} = 0.000393$$

$$s = \sqrt{0.000393} = 0.0198$$

Examples: repeatability of the test equipment, time lag between samples, during which the pH of the solution could change, and operator skill in drawing the sample or using the instrument.

6-18.

sample mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{753.0}{9} = 83.67$ drag counts

sample variance $s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{63343 - \frac{(753.0)^2}{9}}{9-1}$

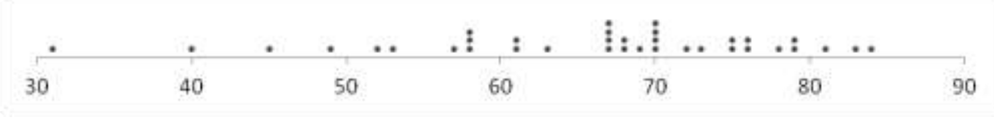
$$= \frac{342.0}{8} = 42.75 \text{drag counts}^2$$

sample standard deviation $s = \sqrt{42.75} = 6.54$ drag counts

Dot Diagram

6-19. a) $\bar{x} = 65.83$ °F
$s = 12.16$ °F

Dot Diagram



b) Removing the smallest observation (31), the sample mean and standard deviation become
$\bar{x} = 66.83$ °F
$s = 10.75$ °F

Section 6-2

6-20. The median will be equal to the mean when the sample is symmetric about the mean value.

6-21. The median will equal the mode when the sample is symmetric with a single mode. The symmetry implies the mode is at the median of the sample.

6-22.

Stem	Leaves	Counts
83	4	1
84	33	2
85	3	1
86	777	3
87	456789	6
88	23334556679	11
89	0233677889	10
90	01113444567889	14
91	001122356688	12
92	22236777	8
93	033457	6
94	2247	4
95		0
96	15	2
97		0
98	8	1
99		0
100	3	1

Stem units: 1

Q1	Median	Q3
88.6	90.4	92.2

6-23. Stem-and-leaf display for cycles to failure: unit = 100   1|2 represents 1200

```
  1     0T|3
  1     0F|
  5     0S|7777
 10     0o|88899
 22     1*|000000011111
 33     1T|22222223333
(15)    1F|444445555555555
 22     1S|66667777777
 11     1o|888899
  5     2*|011
  2     2T|22
```

Median = 1436.5, Q_1 = 1097.8, and Q_3 = 1735.0

Yes. 22 out if 70 coupons survived beyond 1500 cycles.

6-24. Stem-and-leaf display of percentage of cotton   N = 64
Leaf Unit = 0.10   32|1 represents 32.1%

```
  1    32    1
  5    32    5569
  9    33    1144
 17    33    56666688
 25    34    01112334
(13)   34    5566667777779
 26    35    00111234
 18    35    55678
 13    36    1234
  9    36    8888
```

```
5   37   13
3   37   568
```

Median = 34.7, Q_1 = 33.800, and Q_3 = 35.6

6-25.

```
Stem | Leaves              Counts
   7 | 8                       1
   8 |                         0
   8 | 3333                    4
   8 | 4444444455555555       16
   8 | 66666666777777         14
   8 | 88888899999            11
   9 | 000000000111111        15
   9 | 222223333               9
   9 | 444444455555           12
   9 | 666777                  6
   9 | 88                      2
Stem units: 10
```

Median = 89.25, Q_1 = 86.1, and Q_3 = 93.1

6-26. The data in the 42nd is 90.4 which is median.

The mode is the most frequently occurring data value. There are several data values that occur 3 times. These are: 86.7, 88.3, 90.1, 90.4, 92.2 and 92.7, so this data set has a multimodal distribution.

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{7423.2}{83} = 90.53$

6-27. Sample median is at $\dfrac{(70+1)}{2}$ = 35.5th observation, the median is 1436.5.

Modes are 1102, 1315, and 1750 which are the most frequent data.

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{98259}{70} = 1403.7$

6-28. Sample median is at $\dfrac{(64+1)}{2}$ = 32.5th

The 32nd is 34.7 and the 33rd is 34.7, so the median is 34.7.

Mode is 34.7 which is the most frequent data.

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{2227.5}{64} = 34.805$

6-29. Do not use the total as an observation. There are 23 observations.

```
Stem-and-leaf of Billion of kilowatt hours  N  = 23
Leaf Unit = 100

(18)  0  000000000000000111
 5    0  23
 3    0  5
 2    0
 2    0  9
 1    1
 1    1
 1    1
 1    1  6
```

Sample median is at 12th = 38.43.

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{4398.8}{23} = 191.0$

Sample variance: s^2 = 150673.8

Sample standard deviation: s = 388.2

6-30. Sample mean: $\bar{x} =$ 166.78 cm, standard deviation $s = 5.329$ cm, and sample median: $\tilde{x} = 167.5$ cm

Stem-and-leaf display of female engineering student heights N = 37
Leaf Unit = 0.10 154|0 represents 154.0 centimeters

```
  1   154|0
  3   157|00
  5   160|00
  9   162|0000
 17   165|00000000
 (4)  167|0000
 16   170|00000000
  8   172|00000
  3   175|00
  1   177|0
```

6-31. Stem-and-leaf display Strength: unit = 1.0 1|2 represents 12

```
  1  533|0
  2  534|1
  4  535|47
  5  536|6
  9  537|5678
 20  538|12345688888
 26  539|016999
 37  540|11166677899
 46  541|123566679
(12) 542|001122235789
 42  543|001111556
 33  544|00012455678
 22  545|223447899
 13  546|23569
  8  547|357
  5  548|11257
```

$\dfrac{i - 0.5}{100} \times 100 = 95 \Rightarrow i = 95.5 \Rightarrow 95^{th}$ percentile is 5481

6-32. Stem-and-leaf of concentration, `N = 60, Leaf Unit = 1.0, 2|9 represents 29`
Note: Minitab has dropped the value to the right of the decimal to make this display.

```
   1     2|9
   2     3|1
   3     3|9
   8     4|22223
  12     4|5689
  20     5|01223444
(13)     5|5666777899999
  27     6|11244
  22     6|556677789
  13     7|022333
   7     7|6777
   3     8|01
   1     8|9
```

The data have a symmetrical bell-shaped distribution, and therefore may be normally distributed.

Sample Mean $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3592.0}{60} = 59.87$

Sample Standard Deviation

$$\sum_{i=1}^{60} x_i = 3592.0 \quad \text{and} \quad \sum_{i=1}^{60} x_i^2 = 224257$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{224257 - \frac{(3592.0)^2}{60}}{60-1} = \frac{9215.93}{59}$$
$$= 156.20$$

and

$$s = \sqrt{156.20} = 12.50$$

Sample Median $\tilde{x} = 59.45$

```
Variable           N     Median
concentration     60      59.45
```

$$\frac{i-0.5}{60} \times 100 = 95 \Rightarrow i = 57.5 \Rightarrow 95^{th} \ \ percentile \ \ is\, 78.75$$

6-33. Stem-and-leaf display Meter unit = 1.0
Note: Minitab has dropped the value to the right of the decimal to make this display.

```
1      20 | 4
2      20 | 8
8      21 | 000244
14     21 | 667789
20     22 | 000123
37     22 | 55666668888999999
(14)   23 | 00011123344444
49     23 | 555677788899
```

```
37      24 | 00011233334444
23      24 | 5555556666788
10      25 | 00011122
2       25 | 56
```

Sample Mean $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = \dfrac{\sum_{i=1}^{100} x_i}{100} = \dfrac{23423.1}{100} = 234.2$ meters

Sample Standard Deviation

$$\sum_{i=1}^{100} x_i = 23423.1 \quad \text{and} \quad \sum_{i=1}^{100} x_i^2 = 5500847$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{5500847 - \frac{(23423.1)^2}{100}}{100-1} = \frac{14430.86}{99}$$

$$= 145.766 \text{ m}^2$$

and

$$s = \sqrt{145.766} = 12.07 \text{ m}$$

Sample Median

```
Variable              N     Median
meters               100     234.0
```

$$\frac{i-0.5}{100} \times 100 = 90 \Rightarrow i = 90.5 \Rightarrow 90^{th} \text{ percentile is } 250.0$$

6-34. Stem-and-leaf of speed (in megahertz) `N = 120`

```
Leaf Unit = 1.0    63|4 represents 634 megahertz

    2   63|47
    7   64|24899
   16   65|223566899
   35   66|0000001233455788899
   48   67|0022455567899
  (17)  68|00001111233333458
   55   69|0000112345555677889
   36   70|011223444556
   24   71|0057889
   17   72|000012234447
    5   73|59
    3   74|68
    1   75|
    1   76|3
```

(120-9)/120 = 92.5% exceed 650 megahertz.

Sample Mean $\bar{x} = \dfrac{\sum_{i=1}^{120} x_i}{120} = \dfrac{82413}{120} = 686.78$ *mhz*

Sample Standard Deviation

$$\sum_{i=1}^{120} x_i = 82413 \quad \text{and} \quad \sum_{i=1}^{120} x_i^2 = 56677591$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{56677591 - \frac{(82413)^2}{120}}{120-1} = \frac{78402.925}{119}$$
$$= 658.85 \quad mhz^2$$
and
$$s = \sqrt{658.85} = 25.67 \quad mhz$$

Sample Median $\tilde{x} = 683.0\, mhz$

```
Variable              N     Median
speed               120     683.00
```

6-35. Stem-and-leaf display Rating: unit = 0.10 1|2 represents 1.2

```
 1    83|0
 2    84|0
 5    85|000
 7    86|00
 9    87|00
13    88|0000
18    89|00000
(8)   90|00000000
14    91|000000
 8    92|0000
 4    93|
 3    94|0
 3    95|000
```

Sample Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{40} x_i}{40} = \frac{3578}{40} = 89.45$$

Sample Standard Deviation

$$\sum_{i=1}^{40} x_i = 3578 \quad \text{and} \quad \sum_{i=1}^{40} x_i^2 = 320384$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{320384 - \frac{(3578)^2}{40}}{40-1} = \frac{331.9}{39}$$
$$= 8.51$$
and
$$s = \sqrt{8.51} = 2.92$$

Sample Median

```
Variable              N       Median
rating               40       90.000
```

14/40 or 35% of the taste testers considered this particular Pinot Noir truly exceptional.

6-36. Stem-and-leaf diagram of $NbOCl_3$ N = 27

Leaf Unit = 100 0|4 represents 40 gram-mole/liter x 10^{-3}

```
  6    0|444444
  7    0|5
 (9)   1|001122233
 11    1|5679
  7    2|
  7    2|5677
  3    3|124
```

Sample mean $\bar{x} = \dfrac{\sum_{i=1}^{27} x_i}{27} = \dfrac{41553}{27} = 1539 \quad \text{gram} - \text{mole/liter} \, \text{x}10^{-3}$

Sample Standard Deviation

$$\sum_{i=1}^{27} x_i = 41553 \qquad \text{and} \qquad \sum_{i=1}^{27} x_i^2 = 87792869$$

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1} = \frac{87792869 - \frac{(41553)^2}{27}}{27-1} = \frac{23842802}{26} = 917030.85$$

and $s = \sqrt{917030.85} = 957.62 \quad \text{gram - mole/liter x}10^{-3}$

Sample Median $\tilde{x} = 1256 \quad \text{gram} - \text{mole/liter} \, \text{x}10^{-3}$

```
Variable              N       Median
NbOCl3               40         1256
```

6-37. Stem-and-leaf display Height: unit = 0.10 1|2 represents 1.2

```
 Female Students|| Male Students
      0|154   1  |
     00|157   3  |
     00|160   5  |
   0000|162   9  |
00000000|165  17 |   2   165|00
   0000|167  (4) |   3   167|0
00000000|170  16 |   7   170|0000
  00000|172   8  |  17   172|0000000000
     00|175   3  | (15)  175|000000000000000
      0|177   1  |  18   177|0000000
                 |  11   180|00000
                 |  6     182|00
                 |  4    185|00
                 |  2    187|0
```

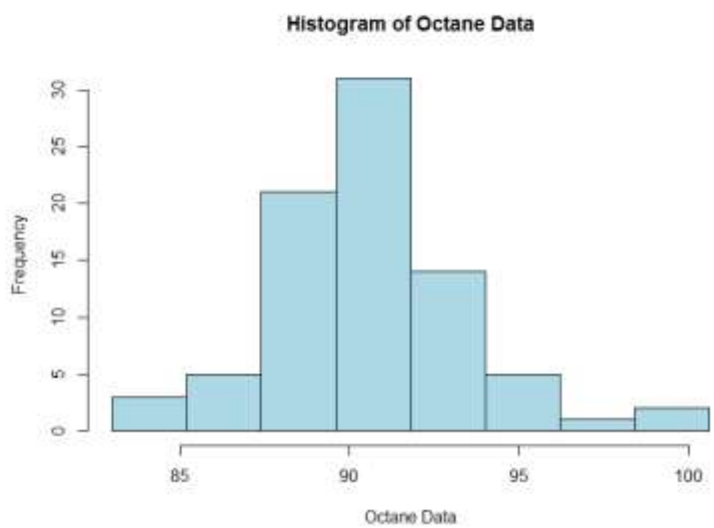
```
1      190|0
```

The male engineering students are taller than the female engineering students. Also there is a slightly wider range in the heights of the male students.

Section 6-3

6-38. Solution uses the n = 82 observations from the data set.
Frequency Tabulation for motor fuel octane data from Exercise 6-22

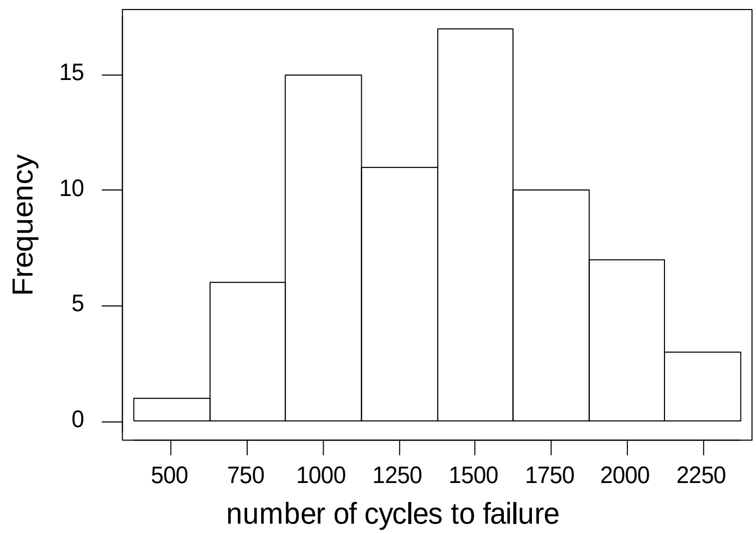
Class	[83,85.2)	[85.2,87.4)	[87.4,89.6)	[89.6,91.8)	[91.8,94)	[94,96.2)	[96.2,98.4)	[98.4,101)
Frequency	3	4	21	30	16	5	1	2
Relative Frequency	0.0366	0.0488	0.2561	0.3659	0.1951	0.0610	0.0122	0.0244
Cumulative Frequency	3	7	28	58	74	79	80	82
Cumulative Relative Frequency	0.0366	0.0854	0.3415	0.7073	0.9024	0.9634	0.9756	1.0000



6-39.

Frequency Tabulation for Exercise 6-23.Cycles

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		.000		0	.0000	0	.0000
1	.000	266.667	133.333	0	.0000	0	.0000
2	266.667	533.333	400.000	1	.0143	1	.0143
3	533.333	800.000	666.667	4	.0571	5	.0714
4	800.000	1066.667	933.333	11	.1571	16	.2286
5	1066.667	1333.333	1200.000	17	.2429	33	.4714
6	1333.333	1600.000	1466.667	15	.2143	48	.6857
7	1600.000	1866.667	1733.333	12	.1714	60	.8571
8	1866.667	2133.333	2000.000	8	.1143	68	.9714
9	2133.333	2400.000	2266.667	2	.0286	70	1.0000
above	2400.000			0	.0000	70	1.0000



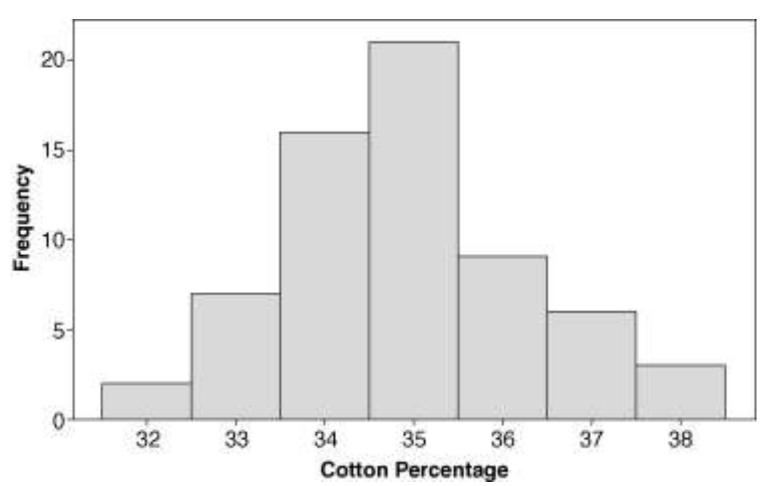
Mean = 1403.66 Standard Deviation = 402.385 Median = 1436.5

6-40.

Frequency Tabulation for Exercise 6-24.Cotton content

	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	31.5	32.5	32	2	2	0.031250	0.03125
2	32.5	33.5	33	7	9	0.109375	0.14063
3	33.5	34.5	34	16	25	0.250000	0.39063
4	34.5	35.5	35	21	46	0.328125	0.71875
5	35.5	36.5	36	9	55	0.140625	0.85938
6	36.5	37.5	37	6	61	0.093750	0.95313
7	37.5	38.5	38	3	64	0.046875	1.00000

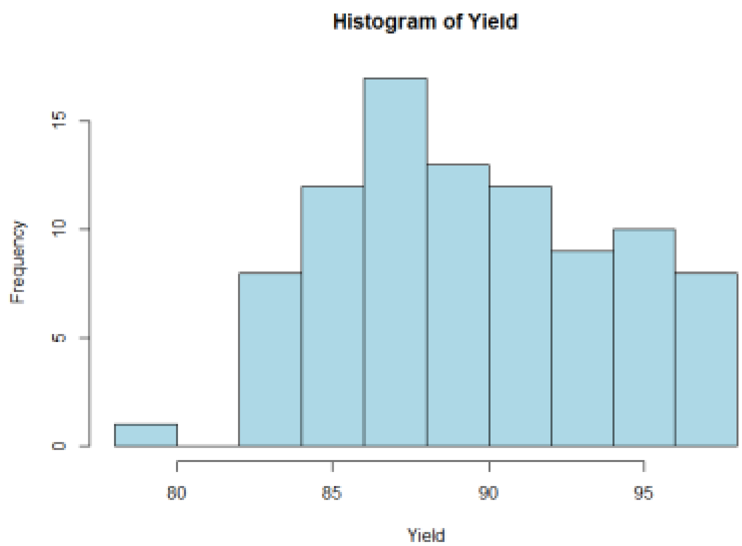
Mean = 34.797 Standard Deviation = 1.367 Median = 34.700



6-41.

Frequency Tabulation for Exercise 6-25.Yield

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
(78,80]	1	0.0111	1	0.0111
(80,82]	0	0.0000	1	0.0111
(82,84]	8	0.0889	9	0.1000
(84,86]	12	0.1333	21	0.2333
(86,88]	17	0.1889	38	0.4222
(88,90]	13	0.1444	51	0.5667
(90,92]	12	0.1333	63	0.7000
(92,94]	9	0.1000	72	0.8000
(94,96]	10	0.1111	82	0.9111
(96,98]	8	0.0889	90	1.0000



6-42. Solutions uses the n = 82 observations from the data set.
Frequency Tabulation for Exercise 6-22.Octane Data (10 bins)

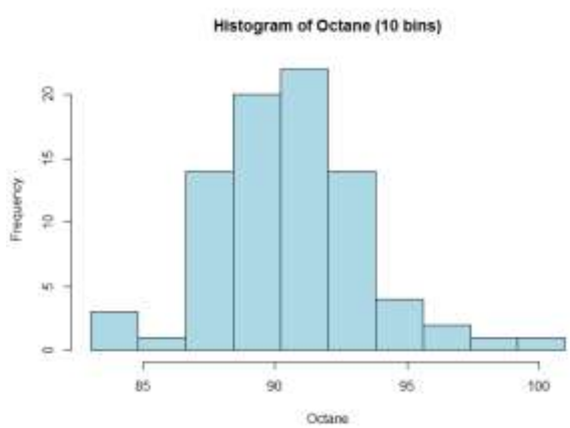
Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
(83, 84. 8]	3	0. 0366	3	0. 0366
(84. 8, 86. 6]	1	0. 0122	4	0. 0488
(86. 6, 88. 4]	14	0. 1707	18	0. 2195
(88. 4, 90. 2]	20	0. 2439	38	0. 4634
(90. 2, 92]	22	0. 2683	60	0. 7317
(92, 93. 8]	14	0. 1707	74	0. 9024
(93. 8, 95. 6]	4	0. 0488	78	0. 9512
(95. 6, 97. 4]	2	0. 0244	80	0. 9756

(97.4, 99.2]	1	0.0122	81	0.9878
(99.2, 101]	1	0.0122	82	1.0000

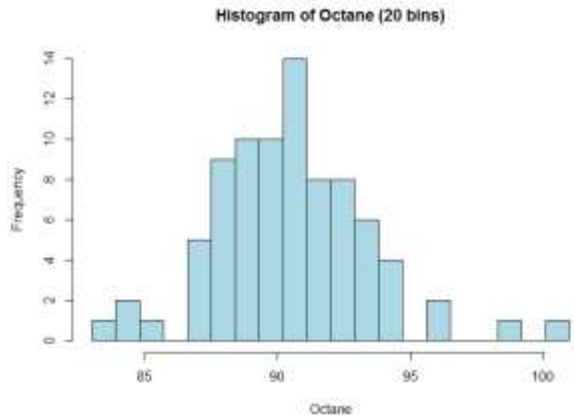
Frequency Tabulation for Exercise 6-22.Octane Data (20 bins)

Class	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
(83, 83.9]	1	0.0122	1	0.0122
(83.9, 84.8]	2	0.0244	3	0.0366
(84.8, 85.7]	1	0.0122	4	0.0488
(85.7, 86.6]	0	0.0000	4	0.0488
(86.6, 87.5]	5	0.0610	9	0.1098
(87.5, 88.4]	9	0.1098	18	0.2195
(88.4, 89.3]	10	0.1220	28	0.3415
(89.3, 90.2]	10	0.1220	38	0.4634
(90.2, 91.1]	14	0.1707	52	0.6341
(91.1, 92]	8	0.0976	60	0.7317
(92, 92.9]	8	0.0976	68	0.8293
(92.9, 93.8]	6	0.0732	74	0.9024
(93.8, 94.7]	4	0.0488	78	0.9512
(94.7, 95.6]	0	0.0000	78	0.9512
(95.6, 96.5]	2	0.0244	80	0.9756
(96.5, 97.4]	0	0.0000	80	0.9756
(97.4, 98.3]	0	0.0000	80	0.9756
(98.3, 99.2]	1	0.0122	81	0.9878
(99.2, 100]	0	0.0000	81	0.9878
(100, 101]	1	0.0122	82	1.0000

Histogram 10 bins:



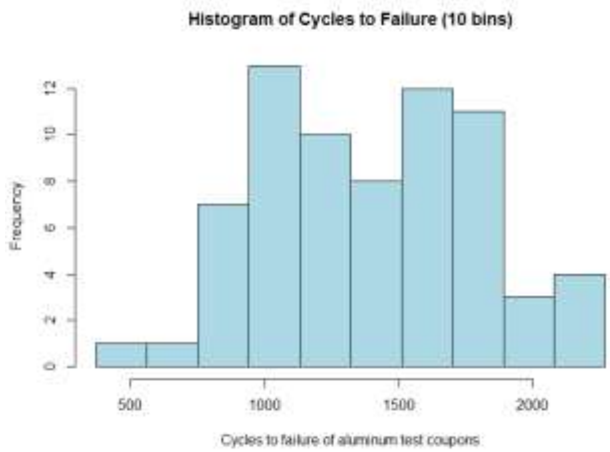
Histogram 20 Bins:



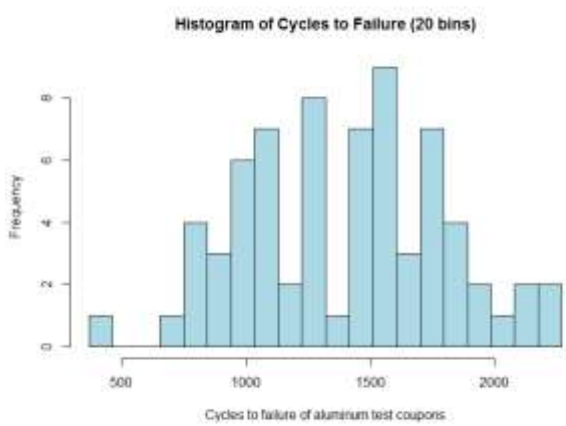
Yes, both of them give the similar information.

6-43.

Histogram 10 bins:



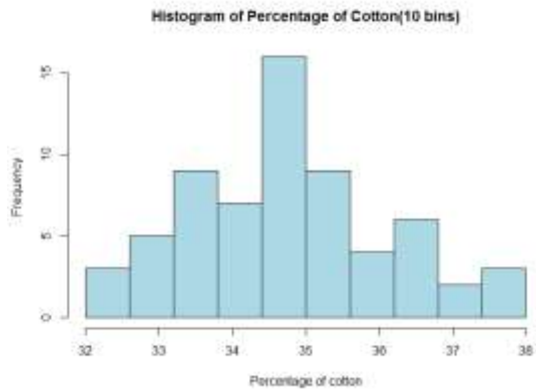
Histogram 20 bins:



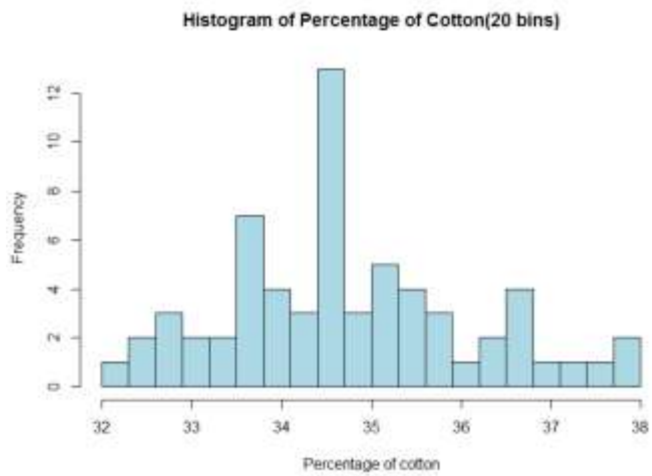
Yes, both of them give the same similar information

6-44.

Histogram 10 bins:



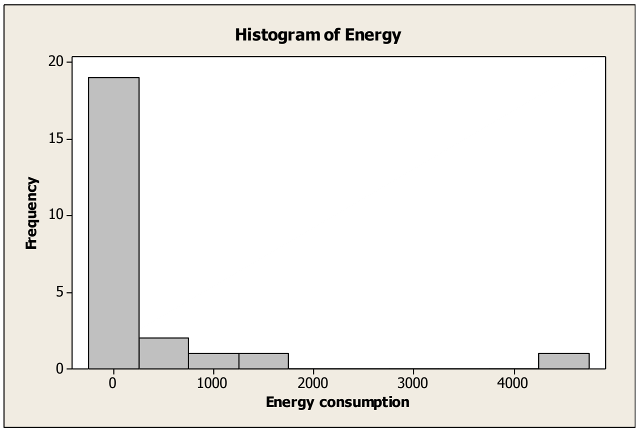
Histogram 20 Bins:



Yes, both of them give similar information.

6-45.

Histogram



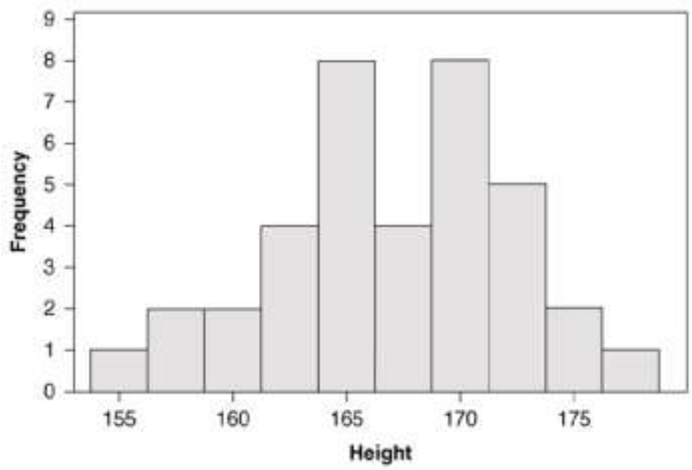
The data are skewed.

6-46.

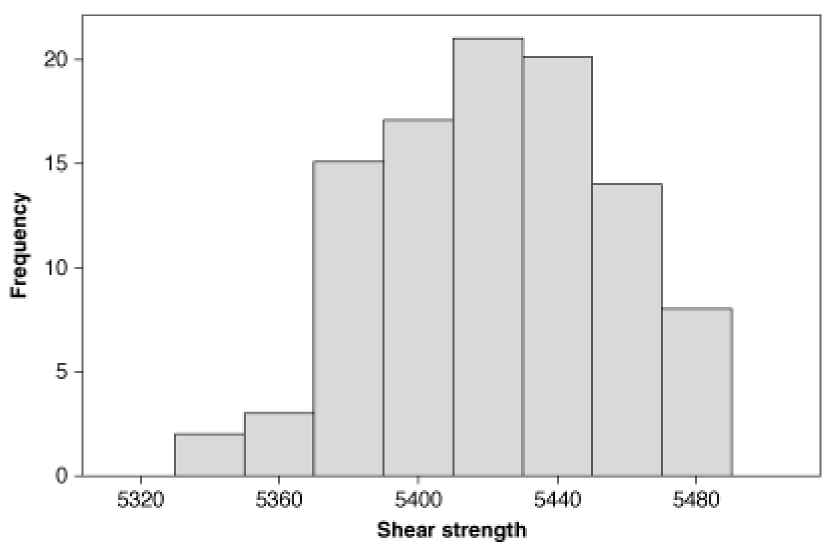
Frequency Tabulation for Problem 6-30. Height Data

	Lower Limit	Upper Limit	Midpoint	Frequency	Cumulative Frequency	Relative Frequency	Cum. Rel. Frequency
1	153.75	156.25	155.0	1	1	0.027027	0.02703
2	156.25	158.75	157.5	2	3	0.054054	0.08108
3	158.75	161.25	160.0	2	5	0.054054	0.13514
4	161.25	163.75	162.5	4	9	0.108108	0.24324
5	163.75	166.25	165.0	8	17	0.216216	0.45946
6	166.25	168.75	167.5	4	21	0.108108	0.56757
7	168.75	171.25	170.0	8	29	0.216216	0.78378
8	171.25	173.75	172.5	5	34	0.135135	0.91892
9	173.75	176.25	175.0	2	36	0.054054	0.97297
10	176.25	178.75	177.5	1	37	0.027027	1.00000

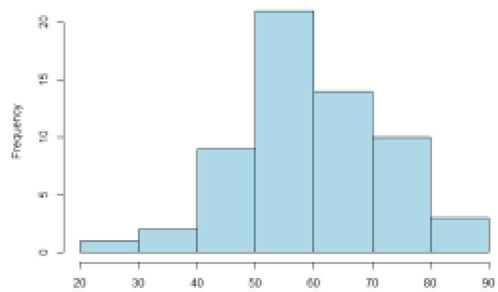
Mean = 166.78   Standard Deviation = 5.329   Median = 167.5



6-47. The histogram for the spot weld shear strength data shows that the data appear to be normally distributed (the same shape that appears in the stem-leaf-diagram).

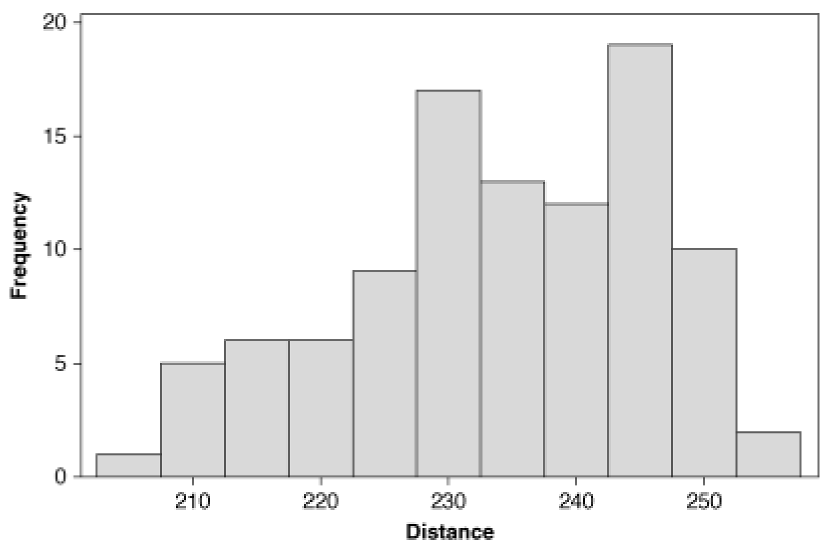


6-48.

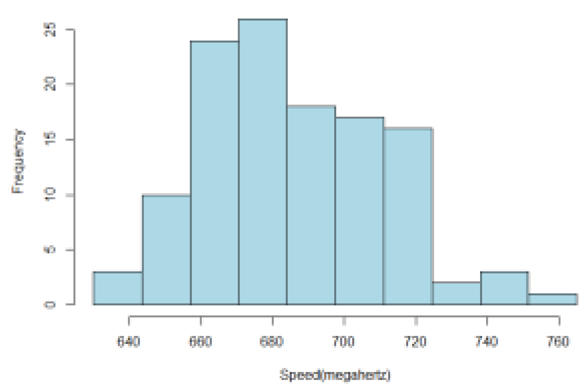


Yes, the histogram shows the same shape as the stem-and-leaf display.

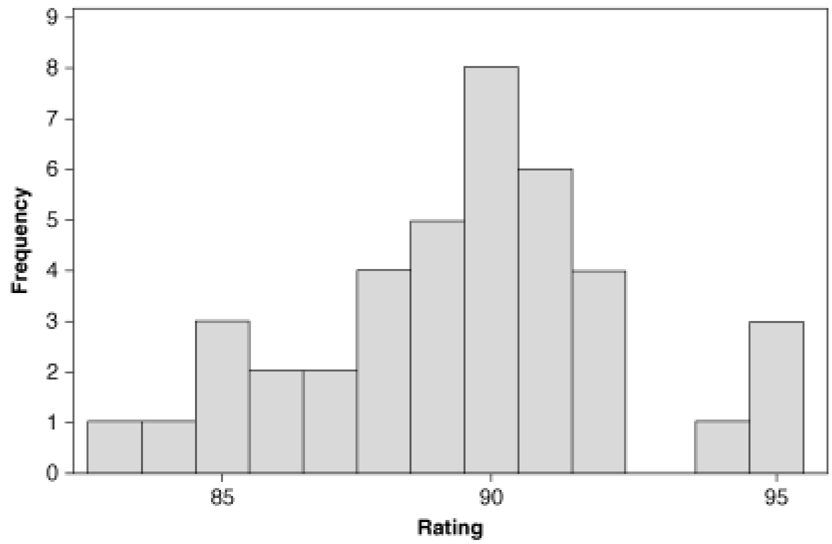
6-49. Yes, the histogram of the distance data shows the same shape as the stem-and-leaf display in exercise 6-33.



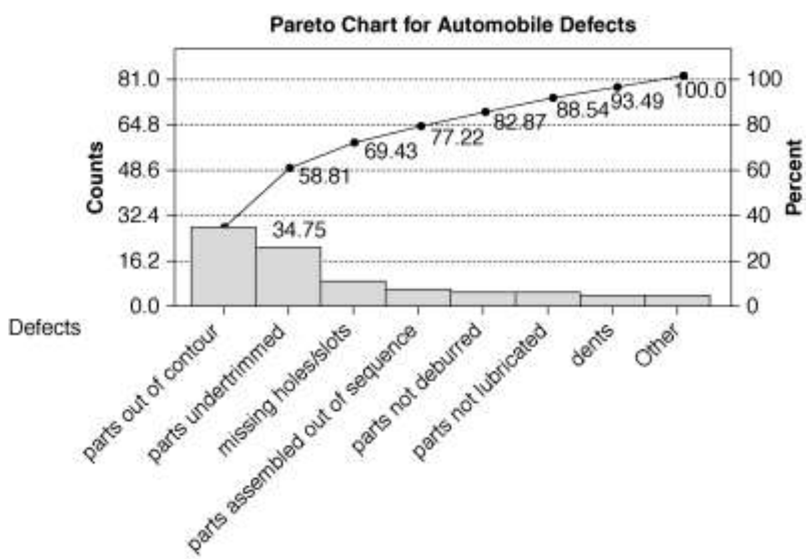
6-50. Histogram for the speed data. Yes, the histogram of the speed data shows the same shape as the stem-and-leaf display.



6-51. Yes, the histogram of the wine rating data shows the same shape as the stem-and-leaf display in exercise 6-35.



6-52.

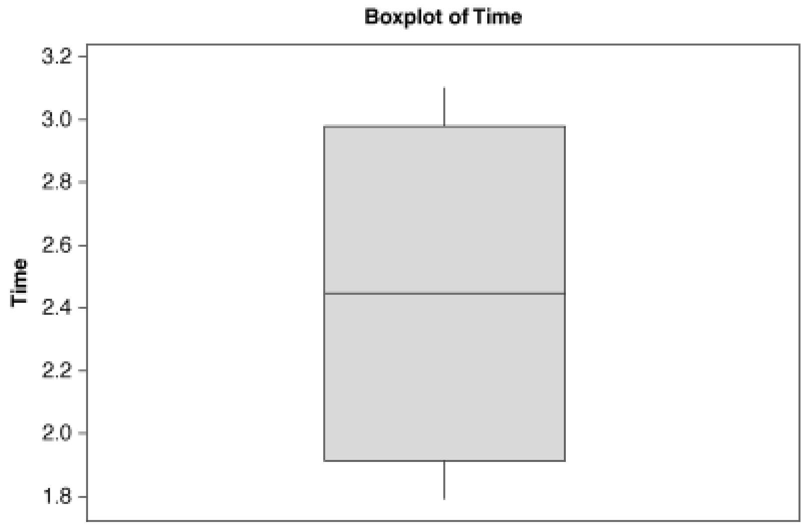


Roughly 60% of defects are described by parts out of contour and parts under trimmed.

6-53.

```
Descriptive Statistics
Variable      N       Mean      Median      TrMean      StDev    SE Mean
time          8      2.426       2.440       2.426      0.5201     0.184

Variable       Minimum     Maximum          Q1          Q3
time             1.8       3.150         1.9275      2.9725
```

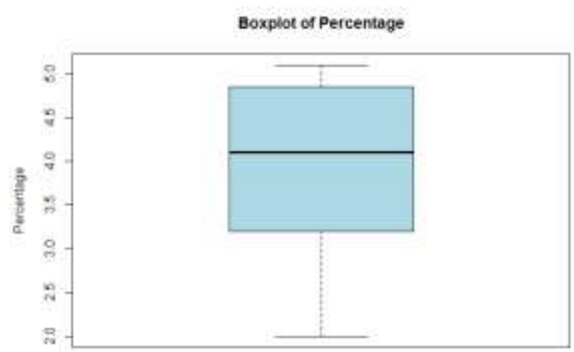
a) Sample Mean: 2.426
   Sample Standard Deviation: 0.5201
b) Box Plot – There are no outliers in the data.



6-54. Descriptive Statistics

min	max	range	sum	median	mean	SE.mean	CI.mean. 0.95	var	std.dev	coef.var
2	5.1	3.1	80	4.1	4	0.2074	0.4340	0.86	0.9274	0.2318

a) Sample Mean = 4, Sample Variance = 0.86, Sample Standard Deviation = 0.9274
b)

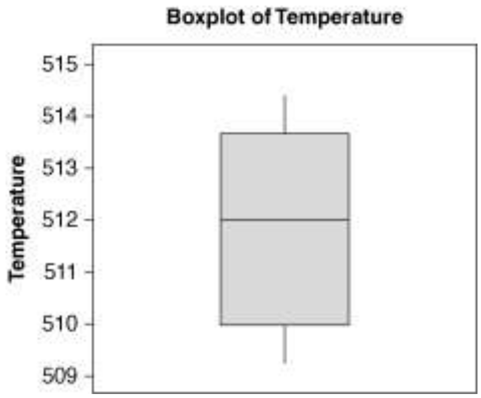


6-55. Descriptive Statistics

```
Variable         N      Mean    Median  Tr Mean    StDev  SE Mean
Temperat         9    511.91    512     511.91     4.33     1.44
Variable       Min      Max      Q1       Q3
Temperat       509    515      509.95   513.65
```

a) Sample Mean: 511.91
   Sample Variance: 9.53
   Sample Standard Deviation: 4.33
b) Median: 512: Any increase in the largest temperature measurement will not affect the median.

c)



6-56. Descriptive statistics

Variable	N	Mean	Median	TrMean	StDev	SE Mean
drag coefficients	9	83.11	82.00	83.11	7.11	2.37

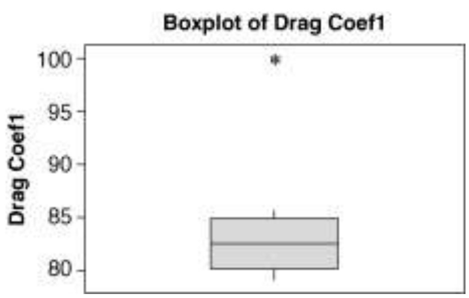
Variable	Minimum	Maximum	Q1	Q3
drag coefficients	74.00	100.00	79.50	84.50

a) Median: $\tilde{x} = 82.00$

Upper quartile: Q_1 =79.50

Lower Quartile: Q_3=84.50

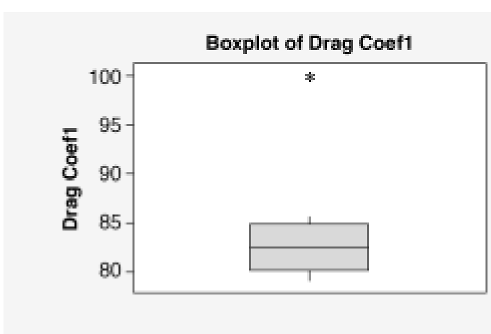
b)



c)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
drag coefficients	8	84.25	82.50	86.52	6.67	2.36

Variable	Minimum	Maximum	Q1	Q3
drag coefficients				



Removing the largest observation (100) decreases the mean and the median. Removing this "outlier" also greatly reduces the variability as seen by the smaller standard deviation and the smaller difference between the upper and lower quartiles.

6-57. Descriptive Statistics of O-ring joint temperature data

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	range	sum	SE.mean	CI.mean.0.95	var	std.dev	coef.var
31	59.5	67.5	65.86	75	84	53	2371	2.0265	4.1140	147.8373	12.1588	0.1846

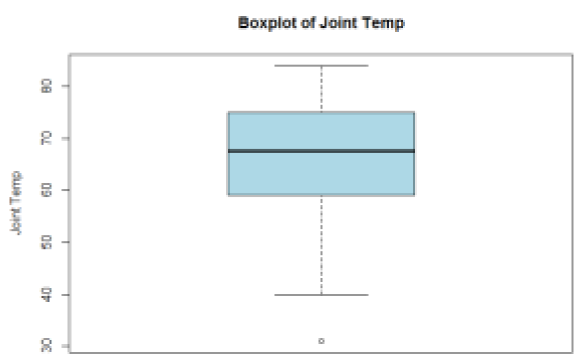
a) Median = 67.50, Lower Quartile: Q_1 = 59.50, Upper Quartile: Q_3 = 75.00

b) Data with lowest point removed

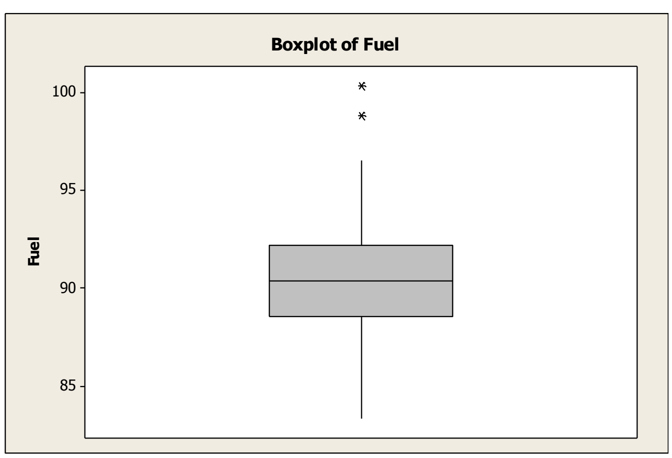
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	range	sum	SE.mean	CI.mean.0.95	var	std.dev	coef.var
40	60.5	68	66.86	75	84	44	2340	1.8160	3.6905	115.4202	10.7434	0.1607

The mean and median have increased and the standard deviation and difference between the upper and lower quartile have decreased.

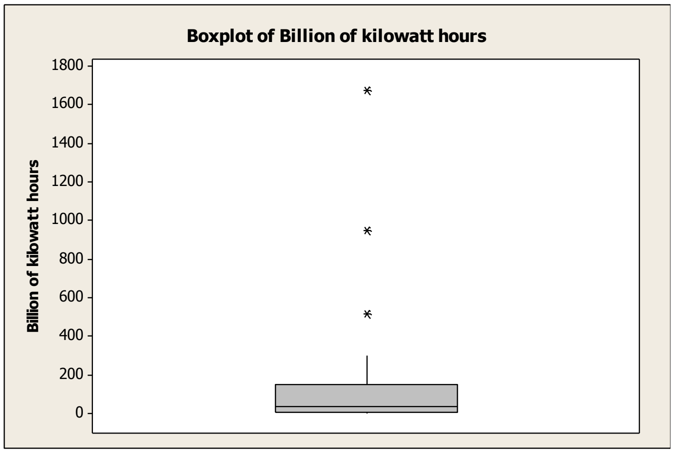
c)Box Plot: The box plot indicates that there is an outlier in the data.



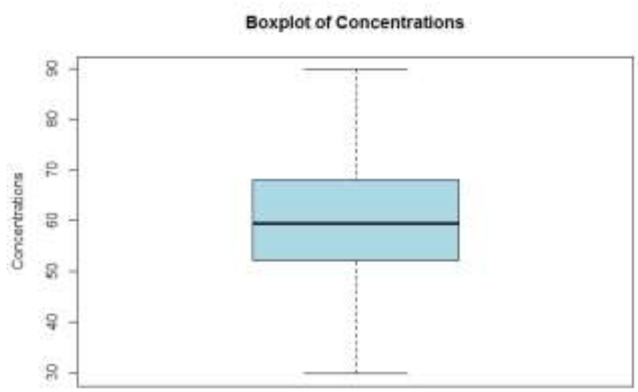
6-58. This plot conveys the same basic information as the stem and leaf plot but in a different format. The outliers that were separated from the main portion of the stem and leaf plot are shown here separated from the whiskers.



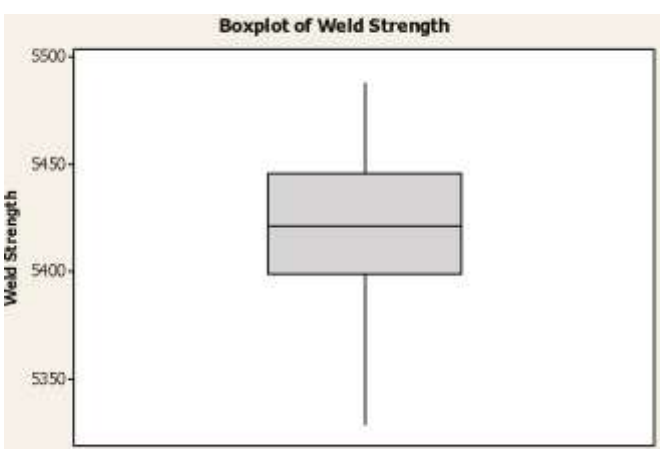
6-59. The box plot shows the same basic information as the stem and leaf plot but in a different format. The outliers that were separated from the main portion of the stem and leaf plot are shown here separated from the whiskers.



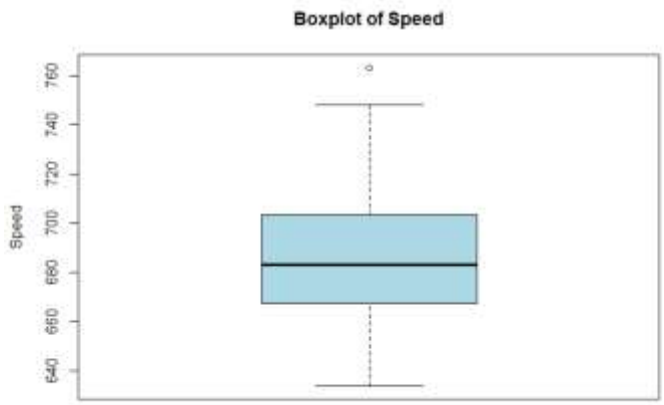
6-60. The box plot and the stem-leaf-diagram show that the data are very symmetrical about the mean. It also shows that there are no outliers in the data.



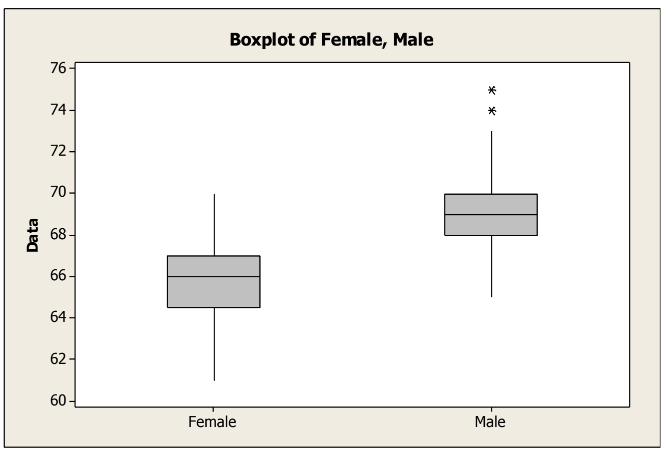
6-61. This plot, as the stem and leaf one, indicates that the data fall mostly in one region and that the measurements toward the ends of the range are more rare.



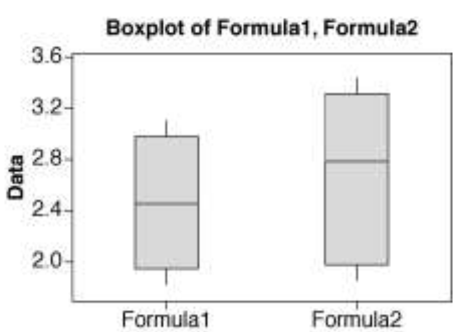
6-62. The box plot shows that the data are symmetrical about the mean. It also shows that there is an outlier in the data. These are the same interpretations seen in the stem-leaf-diagram.



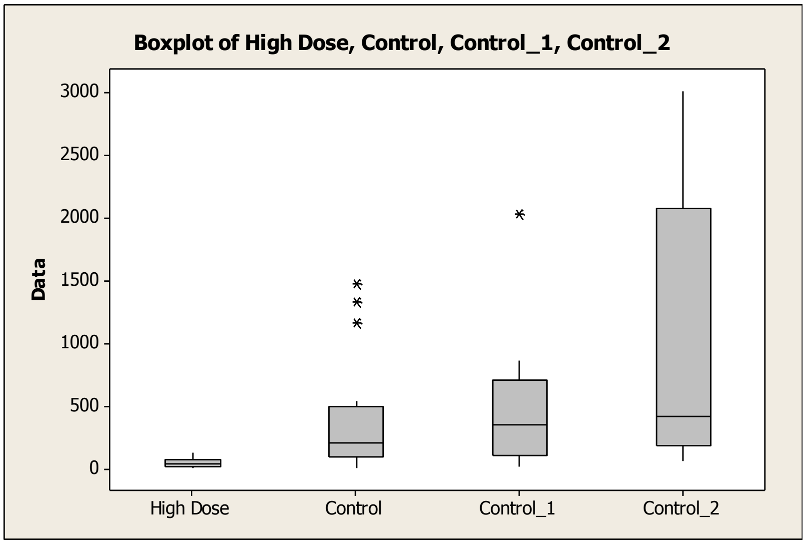
6-63. We can see that the two distributions seem to be centered at different values.



6-64. The box plot indicates that there is a difference between the two formulations. Formulation 2 has a higher mean cold start ignition time and a larger variability in the values of the start times. The first formulation has a lower mean cold start ignition time and is more consistent. Care should be taken, though since these box plots for formula 1 and formula 2 are made using 8 and 10 data points respectively. More data should be collected on each formulation to get a better determination.



6-65. All distributions are centered at about the same value, but have different variances.



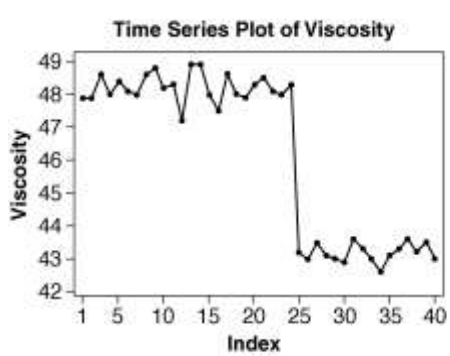
<u>Section 6-5</u>

6-66. Stem-leaf-plot of viscosity N = 40
Leaf Unit = 0.10

```
  2   42 69
 12   43 0000112233
 16   43 5566
 16   44
 16   44
 16   45
 16   45
 16   46
 16   46
 17   47 2
 (4)  47 5999
 19   48 000001123334
  7   48 5666899
```

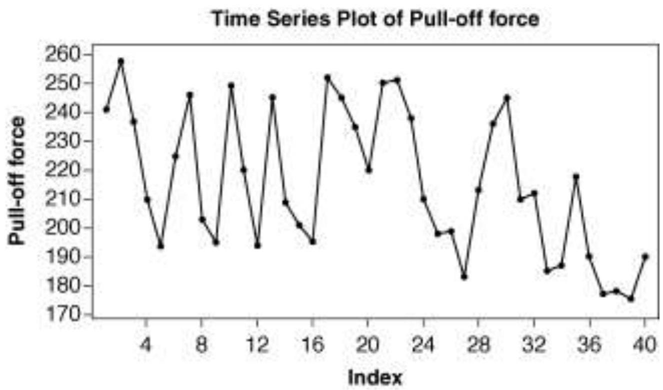
The stem-leaf-plot shows that there are two “different” sets of data. One set of data is centered about 43 and the second set is centered about 48. The time series plot shows that the data starts out at the higher level and then drops down to the lower viscosity level at point 24. Each plot gives us a different set of information.

If the specifications on the product viscosity are 48.0±2, then there is a problem with the process performance after data point 24. An investigation needs to take place to find out why the location of the process has dropped from around 48.0 to 43.0. The most recent product is not within specification limits.



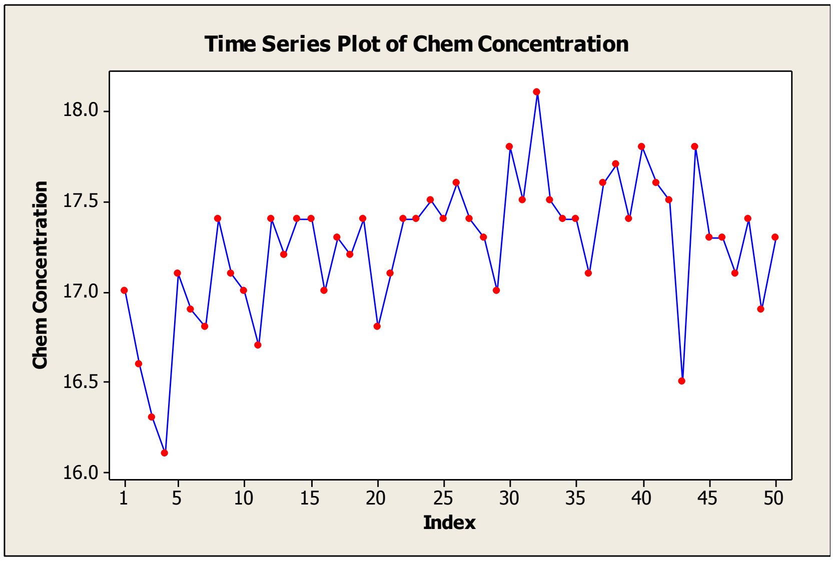
6-67. Stem-and-leaf display for Force: unit = 1   1|2 represents 12

```
 3    17|578
 6    18|357
14    19|00445589
17    20|139
(6)   21|000238
17    22|005
14    23|5678
10    24|155569
 4    25|0128
```



In the time series plot there appears to be a downward trend beginning after time 30. The stem and leaf plot does not reveal this.

6-68.
```
Stem-and-leaf of Chem Concentration  N  = 50
Leaf Unit = 0.10

 1   16  1
 2   16  3
 3   16  5
 5   16  67
 9   16  8899
 18  17  000011111
 25  17  2233333
 25  17  4444444444445555
 8   17  6667
 4   17  888
 1   18  1
```

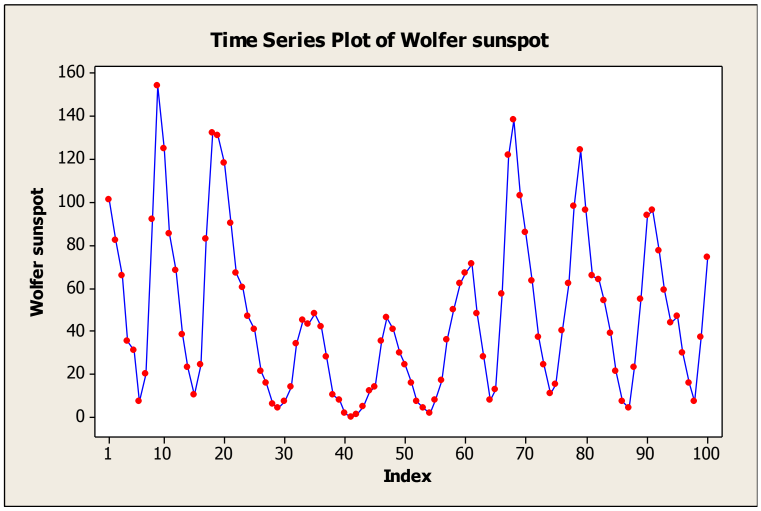


In the time series plot there appears to be trends with higher and lower concentration (probably autocorrelated data). The stem-and-leaf plot does not reveal this.

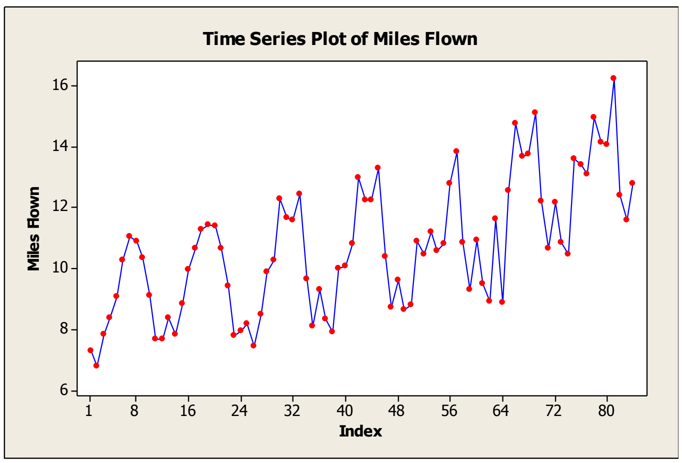
6-69.

```
Stem-and-leaf of Wolfer sunspot  N  = 100
Leaf Unit = 1.0

 17   0    01224445677777888
 29   1    001234456667
 39   2    0113344488
 50   3    00145567789
 50   4    011234567788
 38   5    04579
 33   6    0223466778
 23   7    147
 20   8    2356
 16   9    024668
 10   10   13
 8    11   8
 7    12   245
 4    13   128
 1    14
 1    15   4
```

The data appears to decrease between 1790 and 1835, the stem and leaf plot indicates skewed data.



6-70. Time Series Plot



Each year the miles flown peaks during the summer hours. The number of miles flown increased over the years 1964 to 1970.

Stem-and-leaf of Miles Flown N = 84

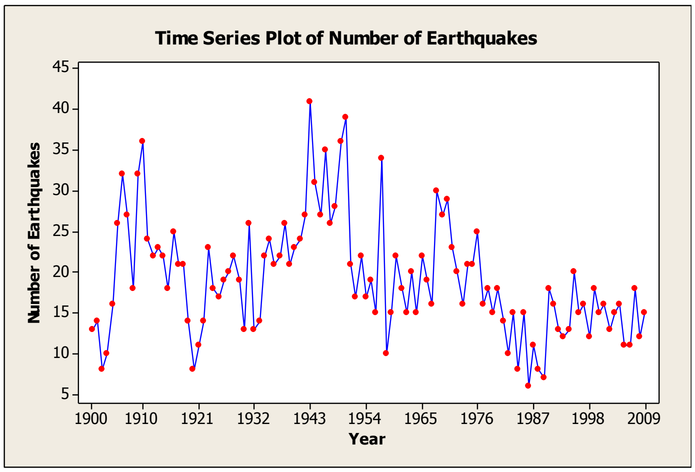
```
Leaf Unit = 0.10

 1     6    7
 10    7    246678889
 22    8    013334677889
 33    9    01223466899
(18)   10   022334456667888889
 33    11   012345566
 24    12   11222345779
 13    13   1245678
 6     14   0179
 2     15   1
 1     16   2
```

When grouped together, the yearly cycles in the data are not seen. The data in the stem-leaf-diagram appear to be nearly normally distributed.

6-71.
```
Stem-and-leaf of Number of Earthquakes  N  = 110
Leaf Unit = 1.0
```

```
 2     0  67
 6     0  8888
 13    1  0001111
 22    1  222333333
 38    1  4444455555555555
 49    1  66666666777
(13)   1  8888888889999
 48    2  00001111111
 37    2  222222223333
 25    2  44455
 20    2  66667777
 12    2  89
 10    3  01
 8     3  22
 6     3  45
 4     3  66
 2     3  9
 1     4  1
```

Time Series Plot



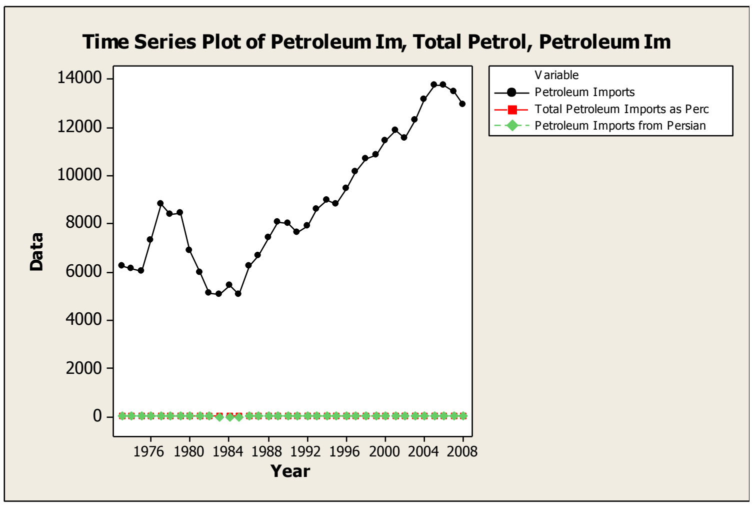
6-72.

```
Stem-and-leaf of Petroleum Imports   N  = 36
Leaf Unit = 100

 5    5   00149
 11   6   012269
 15   7   3468
(8)   8   00346889
 13   9   4
 12   10  178
 9    11  458
 6    12  29
 4    13  1477
```

```
Stem-and-leaf of Total Petroleum Imports as Perc  N  = 36
Leaf Unit = 1.0

 4    3  2334
 9    3  66778
 14   4  00124
(7)   4  5566779
 15   5  0014
 11   5  5688
 7    6  013
```

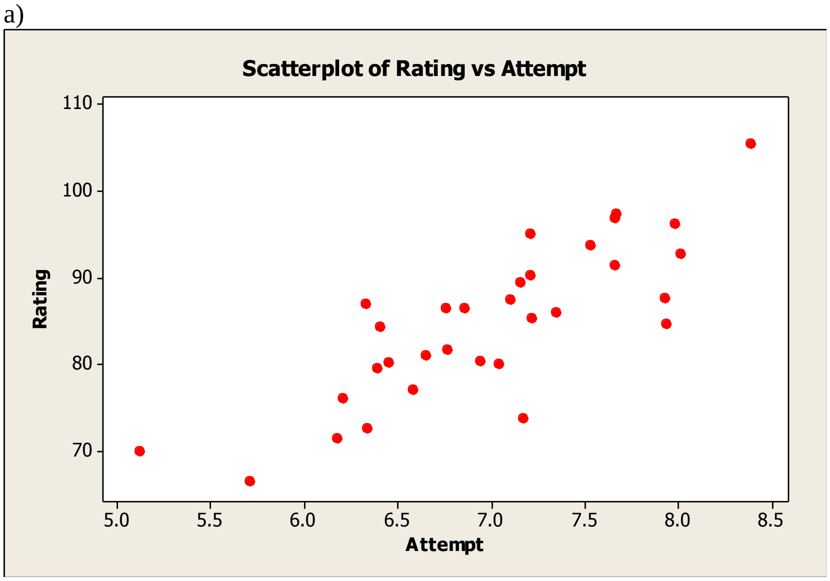
```
 4    6   5566

Stem-and-leaf of Petroleum Imports from Persian  N  = 36
Leaf Unit = 1.0

 1    0   6
 3    0   89
 3    1
 5    1   33
 6    1   4
 14   1   66667777
(6)   1   889999
 16   2   000011
 10   2   2233
 6    2   4445
 2    2   67
```

Time Series plot:



Section 6-6

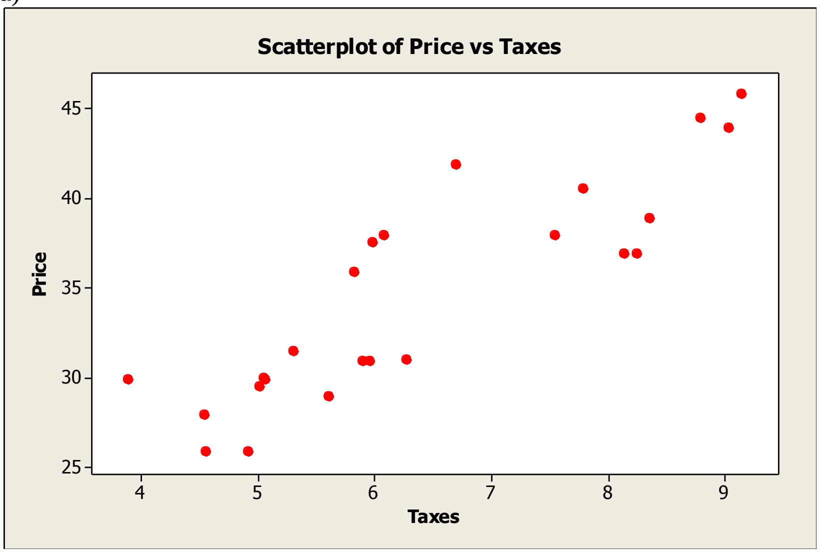
6-73. a)



As the *yards per attempt* increase, the *rating* tends to increase.

b) The correlation coefficient from computer software is 0.820

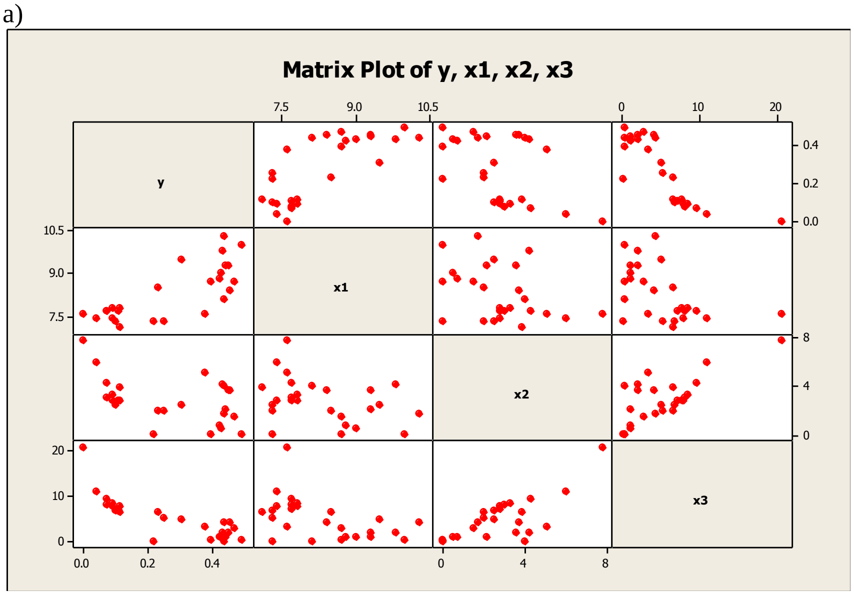
6-74. a)



As the *taxes* increase, the *price* tends to increase.

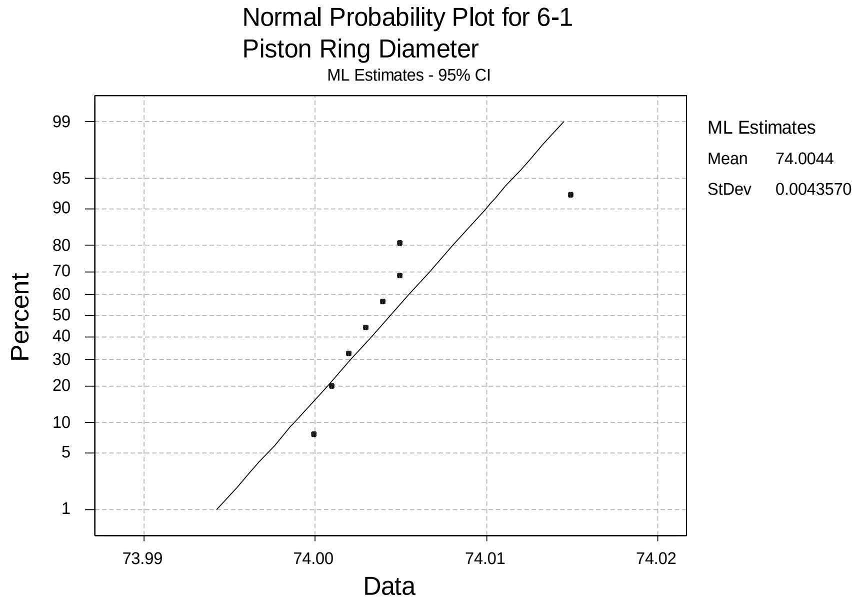
b) From computer software the correlation coefficient is 0.876

6-76. a)

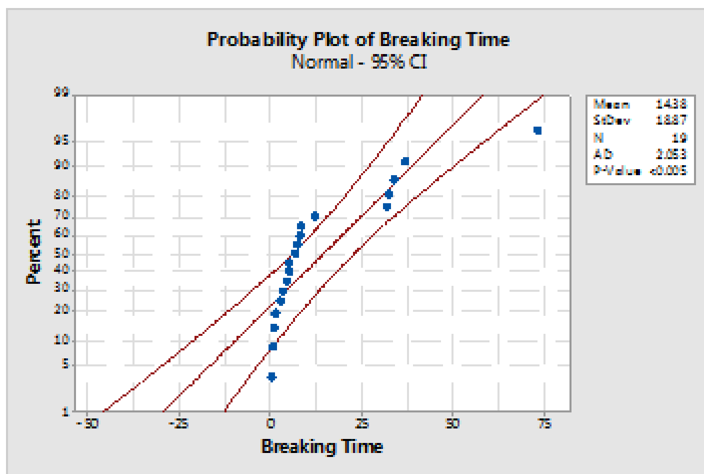


b) Values for y tend to increase as values for x_1 increase. However, values for y tend to decrease as values for x_2 or x_3 decrease.

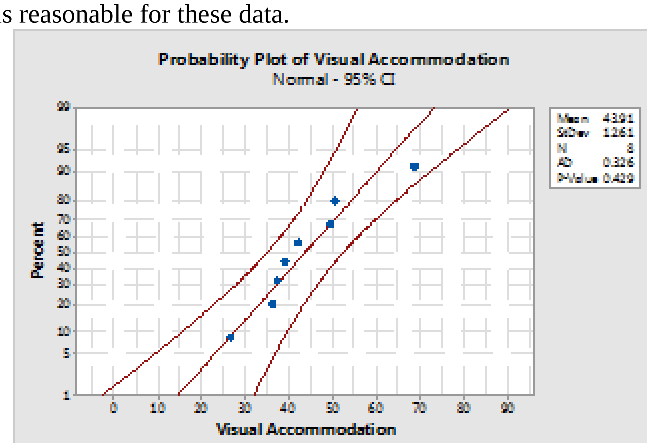
6-77. The pattern of the data indicates that the sample may not come from a normally distributed population or that the largest observation is an outlier. Note the slight bending downward of the sample data at both ends of the graph.



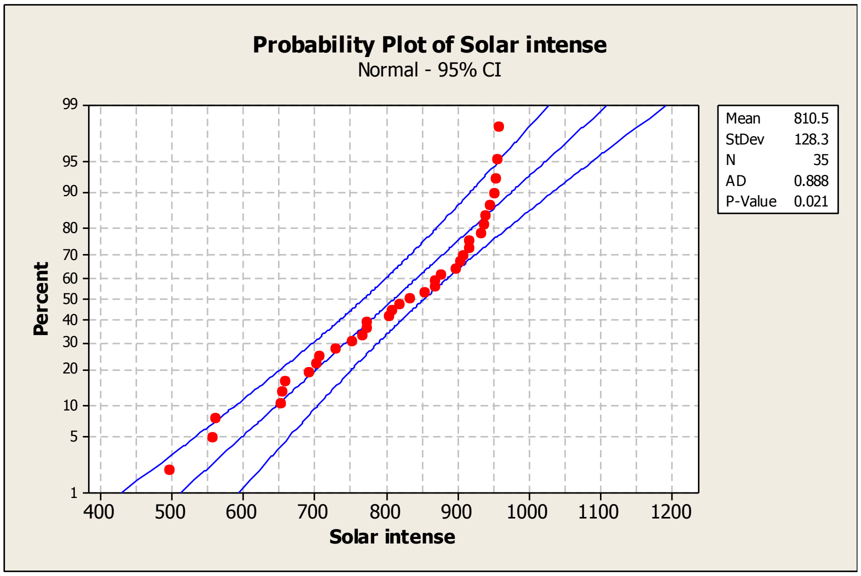
6-78. It appears that the data do not come from a normal distribution. Very few of the data points fall near the line.



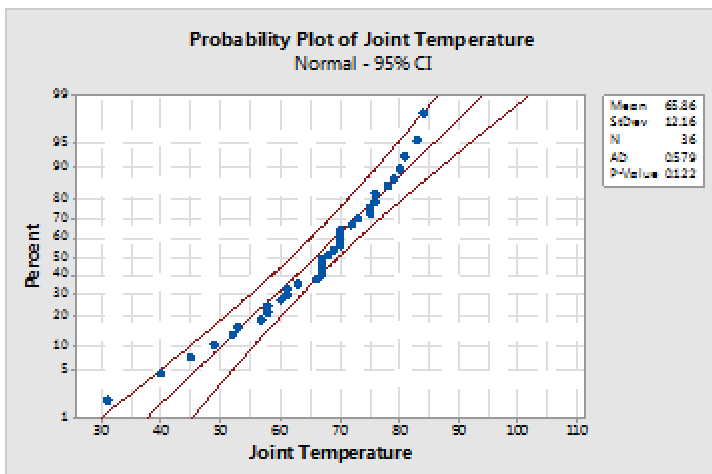
6-79. A normal distribution is reasonable for these data.



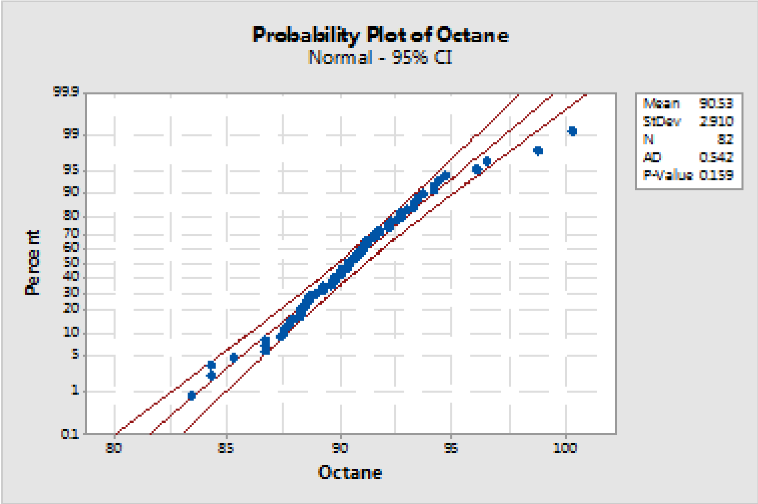
6-80. The normal probability plot shown below does not seem reasonable for normality.



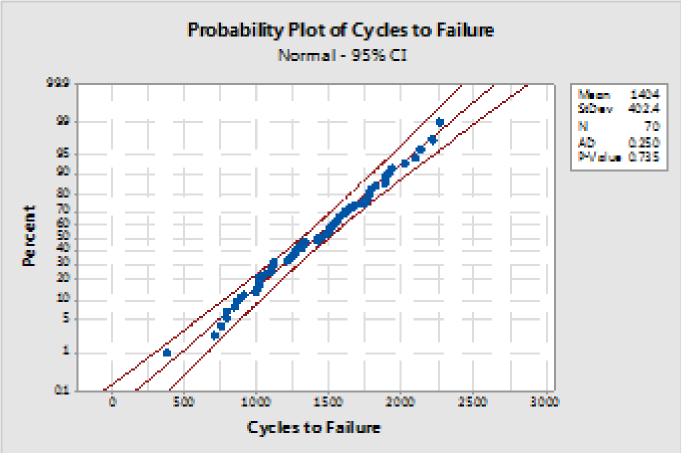
6-81. The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.



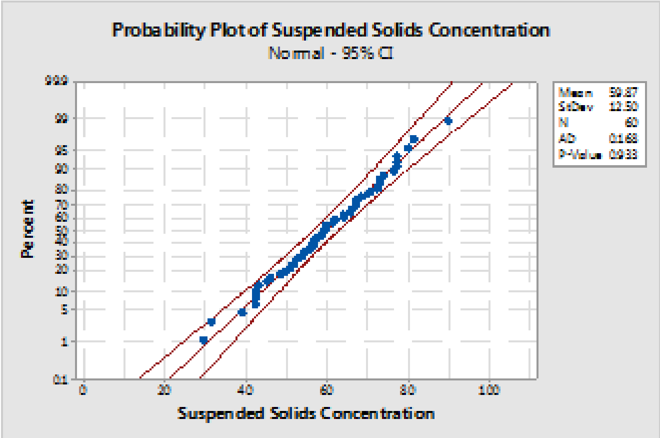
6-82. The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.



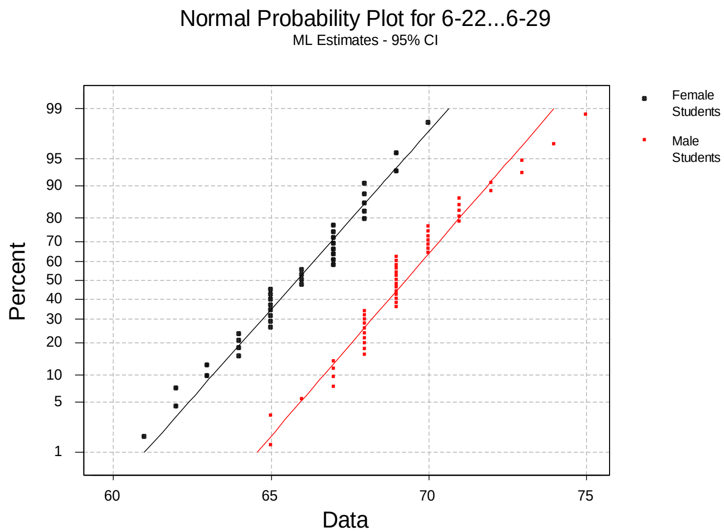
6-83. The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.



6-84. The data appear to be approximately normally distributed. However, there are some departures from the line at the ends of the distribution.



6-85.

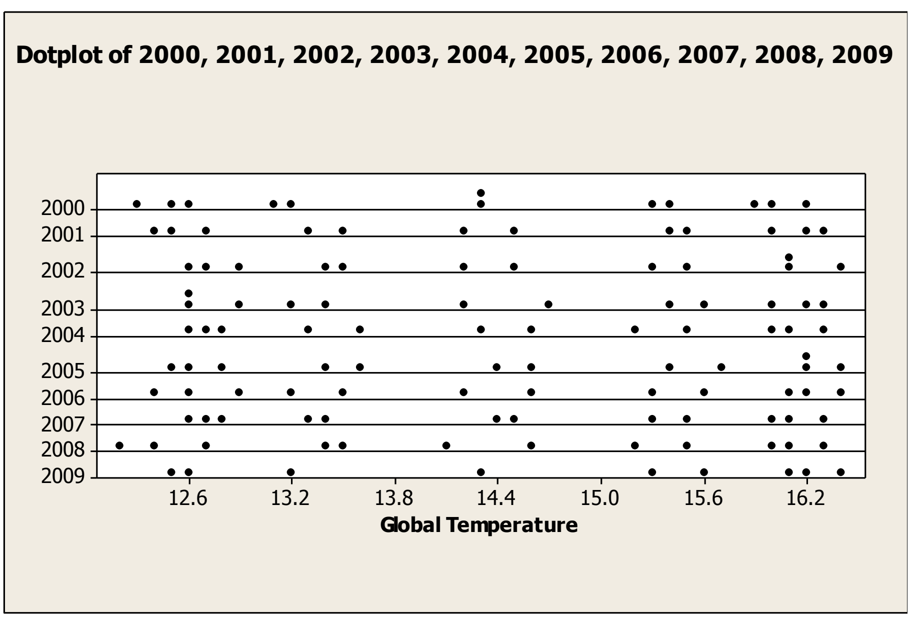


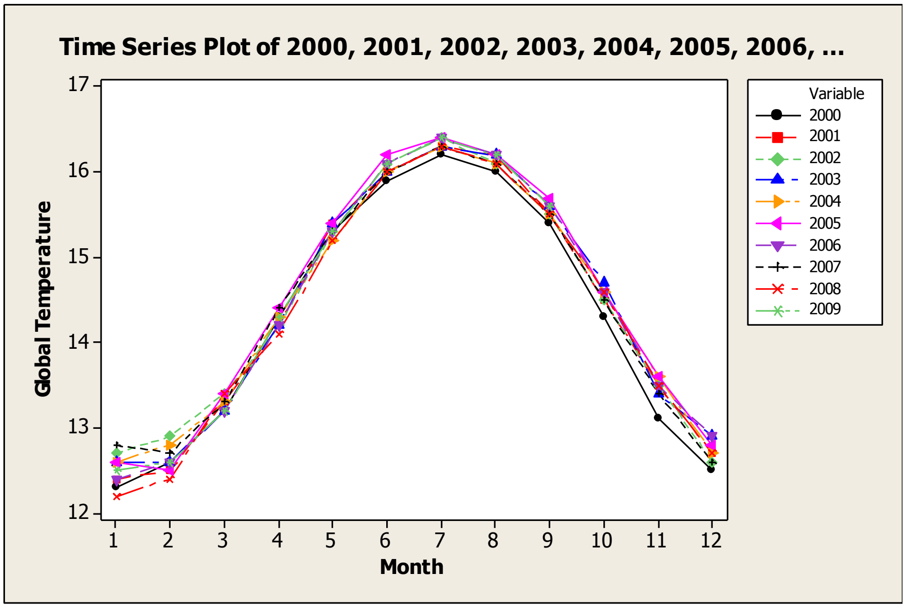
Both populations seem to be normally distributed. Moreover, the lines seem to be roughly parallel indicating that the populations may have the same variance and differ only in the value of their mean.

6-86. Yes, it is possible to obtain an estimate of the mean from the 50th percentile value of the normal probability plot. The 50th percentile point is the median, and for a normal distribution the median equals the mean. An estimate of the standard deviation can be obtained from the 84th percentile minus the 50th percentile . From the z-table, the 84th percentile of a normal distribution is one standard deviation above the mean.

Supplemental Exercises

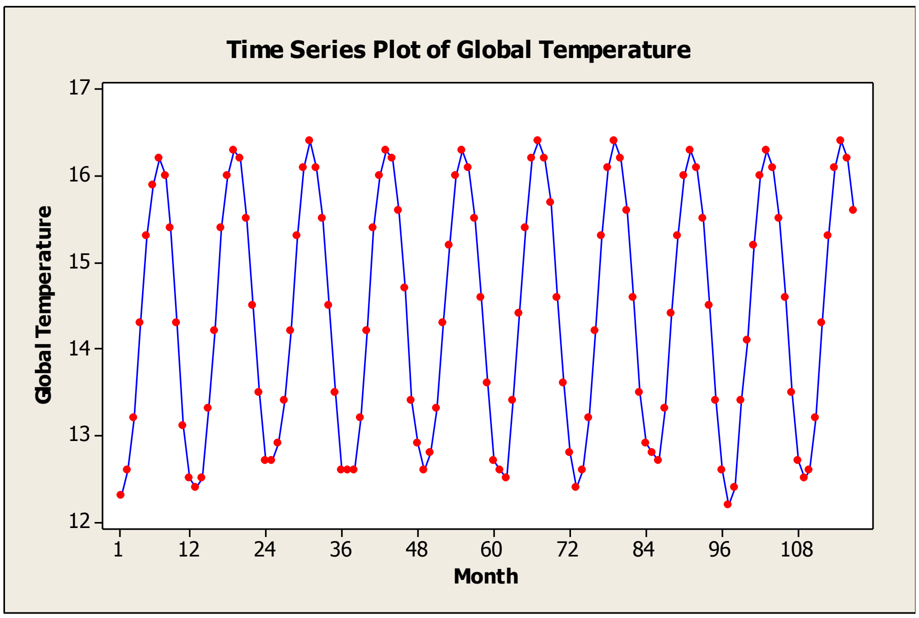
6-87. Based on the digidot plot and time series plots of these data, in each year the temperature has a similar distribution. In each year, the temperature increases until the mid year and then it starts to decrease.






```
Stem-and-leaf of Global Temperature  N  = 117
Leaf Unit = 0.10


  5    12  23444
 29    12  555566666666677777888999
 42    13  1222233344444
 48    13  555566
(11)   14  12222333344
 58    14  55566667
 50    15  22333334444
 39    15  5555566679
 29    16  00000011111112222222333334444
```

Time-series plot by month over 10 years



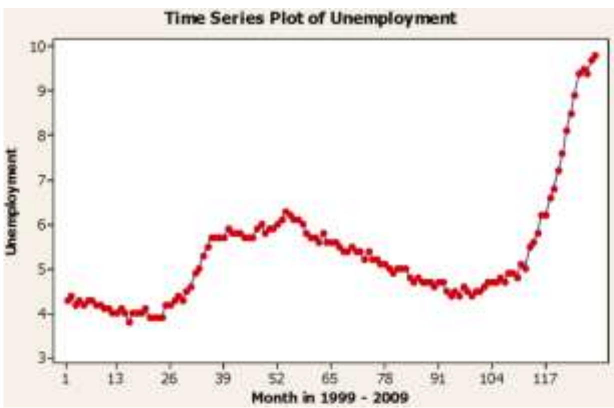
6-88. a) Sample Mean = 65.083
The sample mean value is close enough to the target value to accept the solution as conforming. There is a slight difference due to inherent variability.

b) s^2 = 1.86869 s = 1.367

c) A major source of variability might be variability in the reagent material. Furthermore, if the same setup is used for all measurements it is not expected to affect the variability. However, if each measurement uses a different setup, then setup differences could also be a major source of variability.

A low variance is desirable because it indicates consistency from measurement to measurement. This implies the measurement error has low variability.

6-89. The unemployment rate is steady from 200-2002, then it starts increasing till 2004 and then decreases steadily from 2004 to 2008 and then increases again dramatically in 2009. It reaches its peak up till now.



6-90. a) $\sum_{i=1}^{6} x_i^2 = 10{,}433$ $\left(\sum_{i=1}^{6} x_i\right)^2 = 62{,}001$ $n = 6$

$$s^2 = \frac{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{n}}{n-1} = \frac{10{,}433 - \frac{62{,}001}{6}}{6-1} = 19.9\Omega^2$$

$$s = \sqrt{19.9\Omega^2} = 4.46\Omega$$

b) $\sum_{i=1}^{6} x_i^2 = 353$ $\left(\sum_{i=1}^{6} x_i\right)^2 = 1521$ $n = 6$

$$s^2 = \frac{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{n}}{n-1} = \frac{353 - \frac{1{,}521}{6}}{6-1} = 19.9\Omega^2$$

$$s = \sqrt{19.9\Omega^2} = 4.46\Omega$$

Shifting the data from the sample by a constant amount has no effect on the sample variance or standard deviation.

c) $\sum_{i=1}^{6} x_i^2 = 1043300 \qquad \left(\sum_{i=1}^{6} x_i\right)^2 = 6200100 \qquad n = 6$

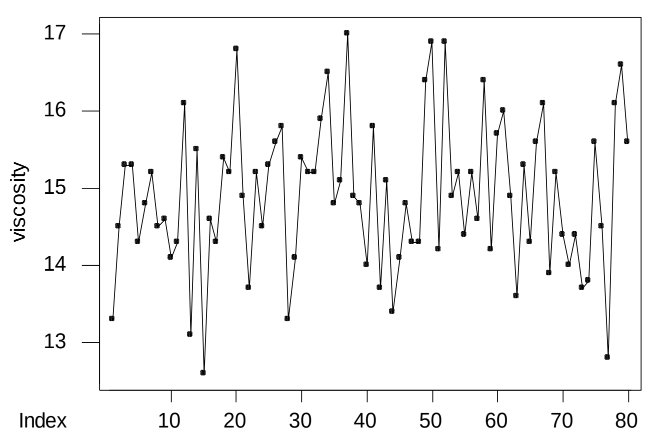
$$s^2 = \frac{\sum_{i=1}^{6} x_i^2 - \frac{\left(\sum_{i=1}^{6} x_i\right)^2}{n}}{n-1} = \frac{1043300 - \frac{6200100}{6}}{6-1} = 1990\Omega^2$$

$$s = \sqrt{1990\Omega^2} = 44.61\Omega$$

Yes, the rescaling is by a factor of 10. Therefore, s^2 and s would be rescaled by multiplying s^2 by 10^2 (resulting in $1990\Omega^2$) and s by 10 (44.6Ω).

6-91. a) Sample 1 Range = 6
Sample 2 Range = 6
Yes, the two appear to exhibit the same variability
b) Sample 1 s = 2.317
Sample 2 s = 2.098
No, sample 1 has a larger standard deviation.
c) The sample range is a relatively crude measure of the sample variability as compared to the sample standard deviation since the standard deviation uses the information from every data point in the sample whereas the range uses the information contained in only two data points - the minimum and maximum.

6-92. a) It appears that the data may shift up and then down over the 80 points.



b) It appears that the mean of the second set of 40 data points may be slightly higher than the first set of 40.

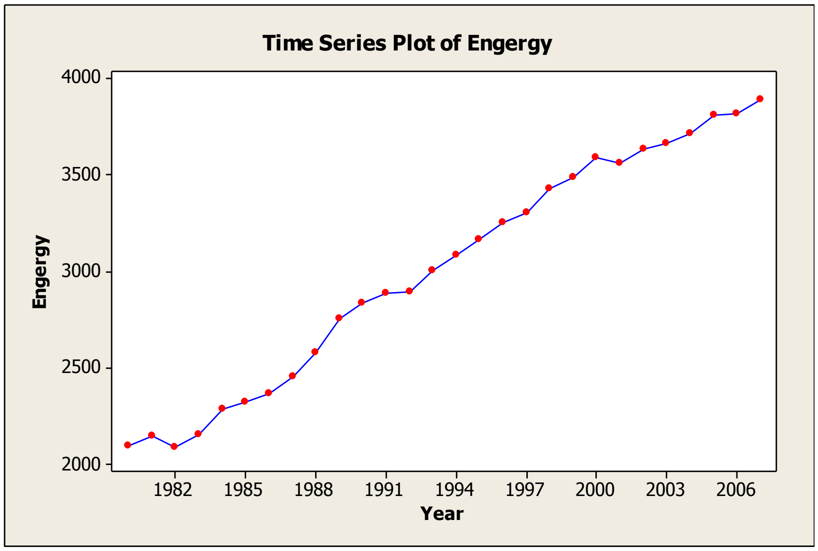
c) Descriptive Statistics: viscosity 1, viscosity 2

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Viscosity1	40	14.875	14.900	14.875	0.948	0.150
Viscosity2	40	14.923	14.850	14.914	1.023	0.162

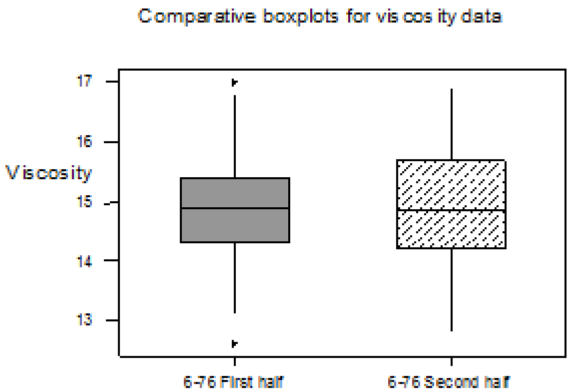
There is a slight difference in the mean levels and the standard deviations.

6-93. From the stem-and-leaf diagram, the distribution looks like the uniform distribution. From the time series plot, there is an increasing trend in energy consumption.

```
Stem-and-leaf of Energy  N  = 28
Leaf Unit = 100

 4    2   0011
 7    2   233
 9    2   45
 10   2   7
 13   2   888
(3)   3   001
 12   3   23
 10   3   4455
 6    3   667
 3    3   888
```

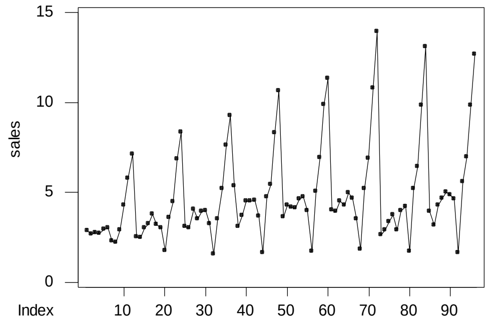


6-94.



Both sets of data appear to have the same mean although the first half of the data seems to be concentrated a little more tightly. Two data points appear as outliers in the first half of the data.

6-95.



There appears to be a cyclic variation in the data with the high value of the cycle generally increasing. The high values are during the winter holiday months.

b) We might draw another cycle, with the peak similar to the last year's data (1969) at about 12.7 thousand bottles.

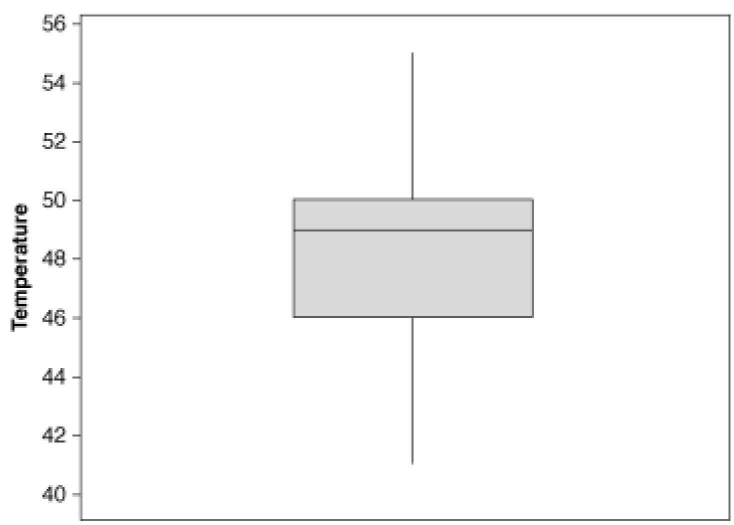
6-96. Descriptive Statistics

```
Variable         N      Mean    Median  Tr Mean     StDev  SE Mean
temperat        24    48.167    49.000   49.917     3.171    0.647
Variable       Min       Max        Q1       Q3
temperat    40.000    55.000    46.000   50.000
```

a) Sample Mean: 48.167
Sample Median: 49

b) Sample Variance: 10.058
Sample Standard Deviation: 3.171

c)



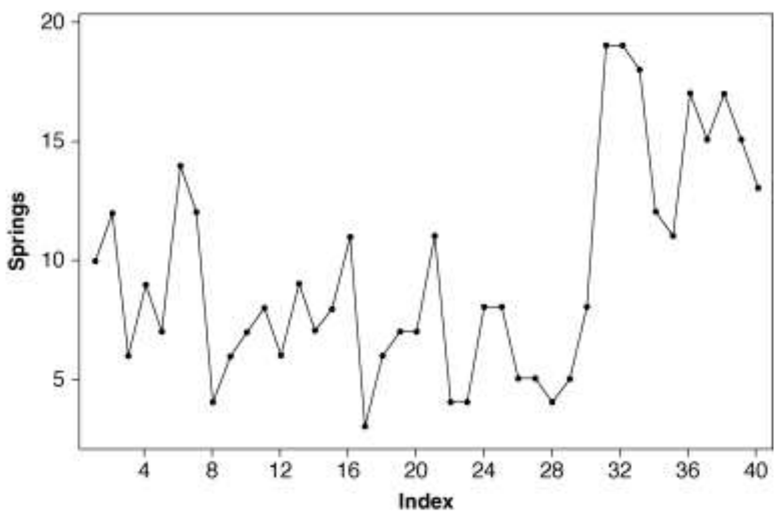
The data appear to be slightly skewed.

6-97. a) Stem-and-leaf display for Problem 2-35: unit = 1   1|2 represents 12

```
  1    0  3
  8    0  4444555
 17    0  666677777
 (7)   0  8888899
 16    1  0111
 12    1  2223
  8    1  455
  5    1  77
  3    1  899
```

b) Sample Average = 9.425
Sample Standard Deviation = 4.5285

c)



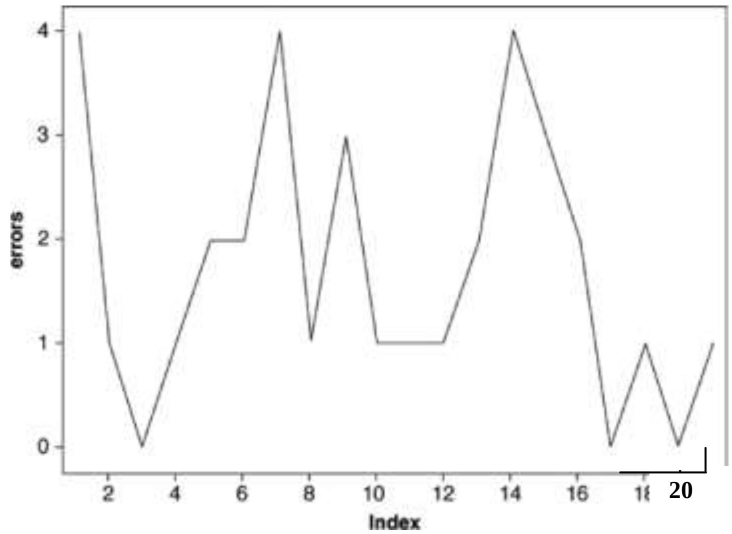
The time series plot indicates there was an increase in the average number of nonconforming springs during the 40 days. In particular, the increase occured during the last 10 days.

6-98. a) Stem-and-leaf of errors   N = 20
Leaf Unit = 0.10

```
  3    0 000
 (8)   1 00000000
  9    2 0000
  5    3 00
  3    4 000
```

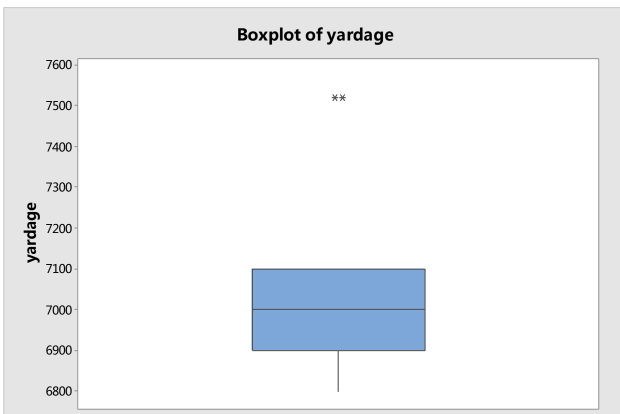
b) Sample Average = 1.700
Sample Standard Deviation = 1.302

c)

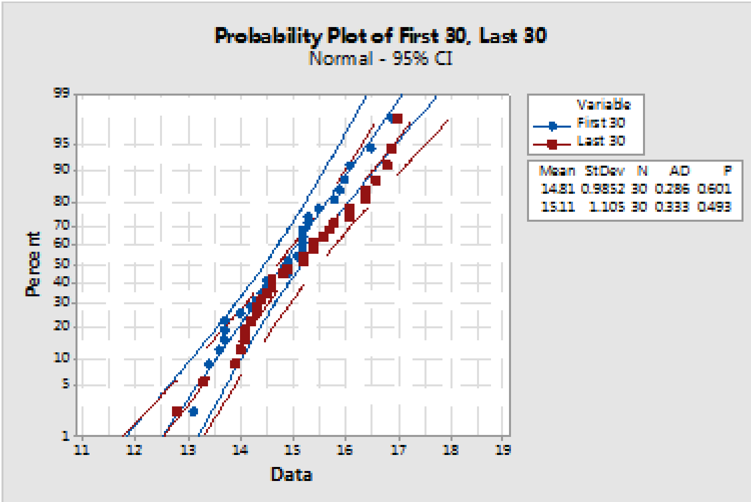


The time series plot indicates a slight decrease in the number of errors for strings 17 - 20.

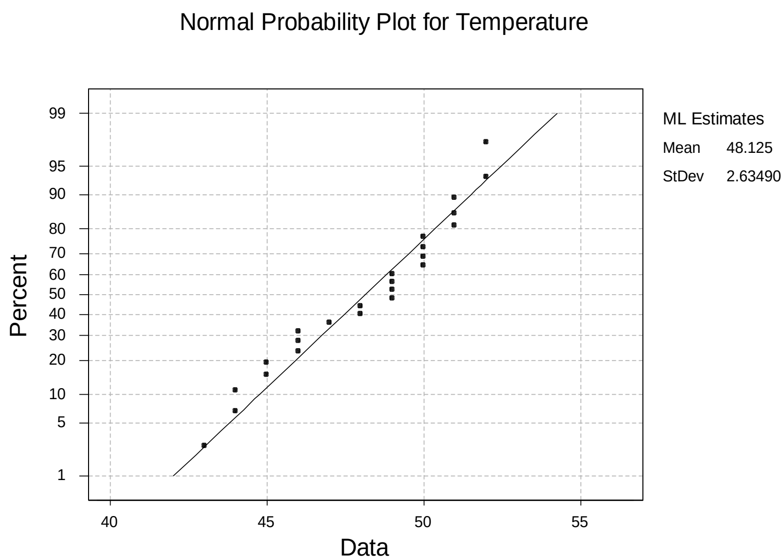
6-99. The golf course yardage data appear to be skewed. Also, there is an outlying data point above 7500 yards.



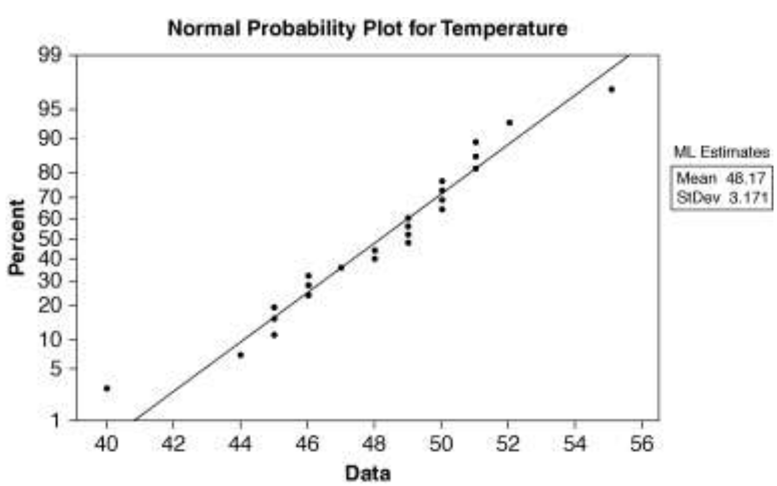
6-100. Both sets of data appear to be normally distributed and with roughly the same mean value. The difference in slopes for the two lines indicates that a change in variance might have occurred. This could have been the result of a change in processing conditions, the quality of the raw material or some other factor.



6-101.

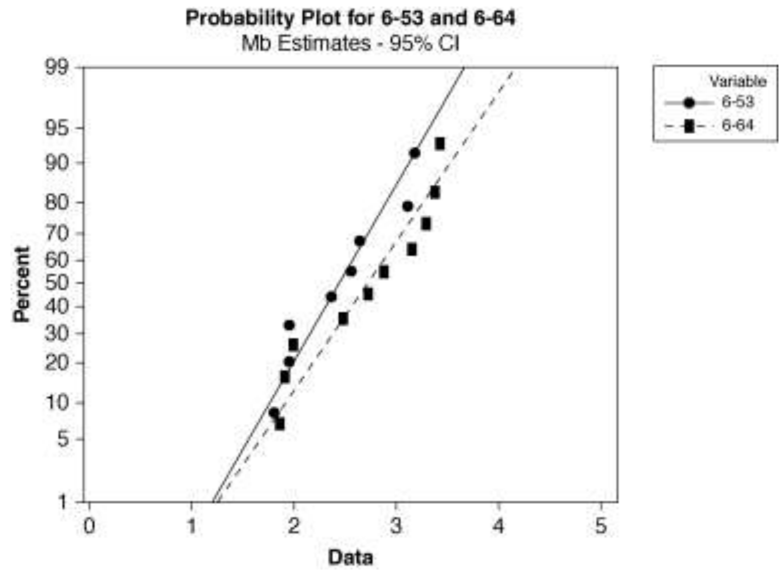


A normal distribution is reasonable for these data. There are some repeated values in the data that cause some points to fall off the line.



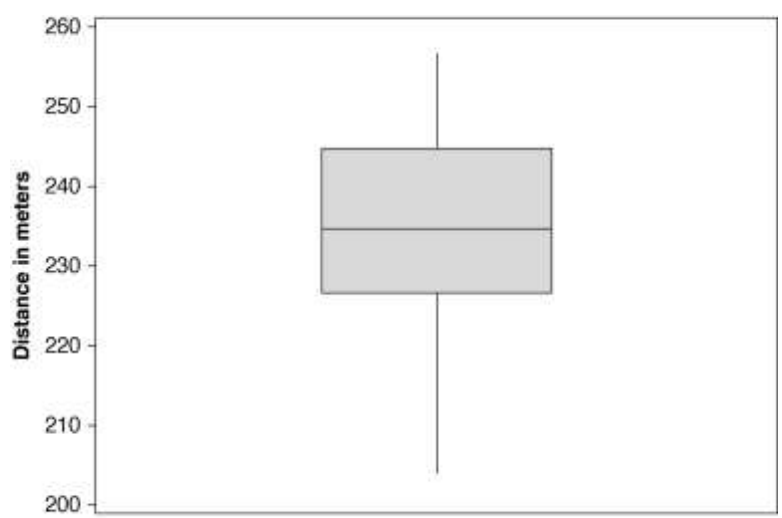
There appears to be no evidence that the data are not normally distributed. There are some repeat points in the data that cause some points to fall off the line.

6-102.



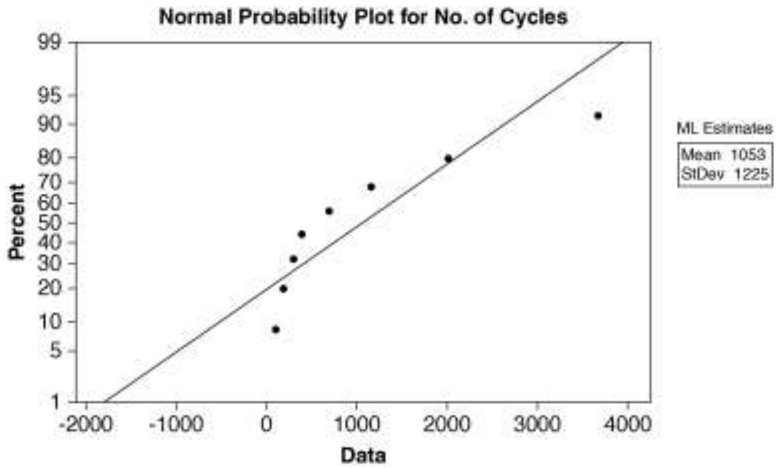
Although we do not have sufficient data points to really see a pattern, there seem to be no significant deviations from normality for either sample. The large difference in slopes indicates that the variances of the populations are very different.

6-103.



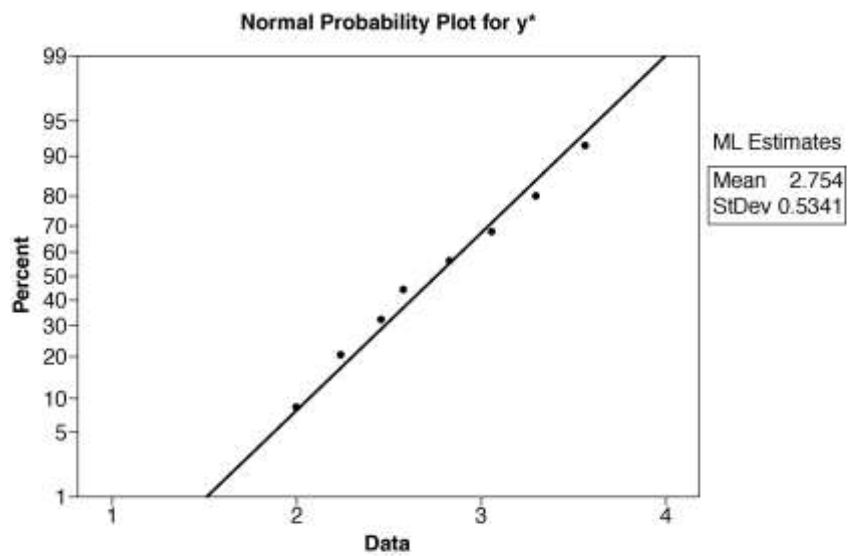
The plot indicates that most balls will fall somewhere in the 225-245 range. In general, the population is grouped more toward the high end of the region. This same type of information could have been obtained from the stem and leaf graph of problem 6-33.

6-104. a)



The data do not appear to be normally distributed. There is a curve in the line.

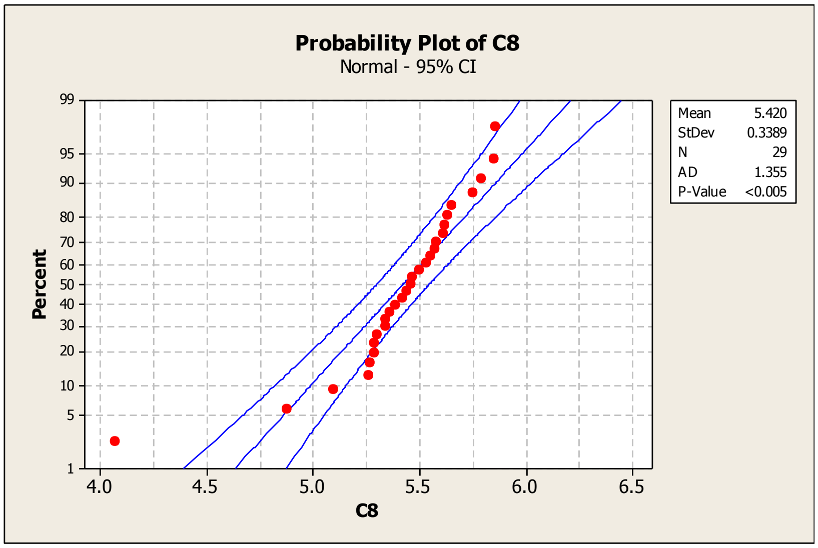
b)



After the transformation y*=log(y), the normal probability plot shows no evidence that the data are not normally distributed.

6-105. a) Descriptive Statistics

```
Variable        N  N*    Mean  SE Mean   StDev  Variance
Density        29   0  5.4197   0.0629  0.3389    0.1148


Variable      Minimum       Q1   Median       Q3   Maximum
Density        4.0700   5.2950   5.4600   5.6150    5.8600
```

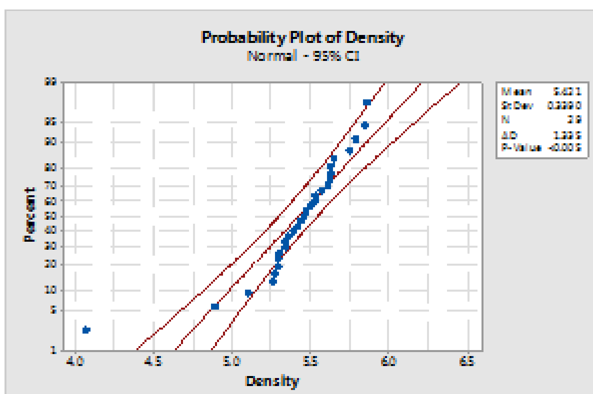


b) There appears to be a low outlier in the data.

c) Due to the very low data point at 4.07, the mean may be lower than it should be. Therefore, the median would be a better estimate of the density of the earth. The median is not affected by a few outliers.

6-106. a) Descriptive Statistics: Density

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Density	29	0	5.4210	0.0630	0.3390	4.0700	5.2950	5.4600	5.6250	5.8600



b) There does appear to be a low outlier in the data.

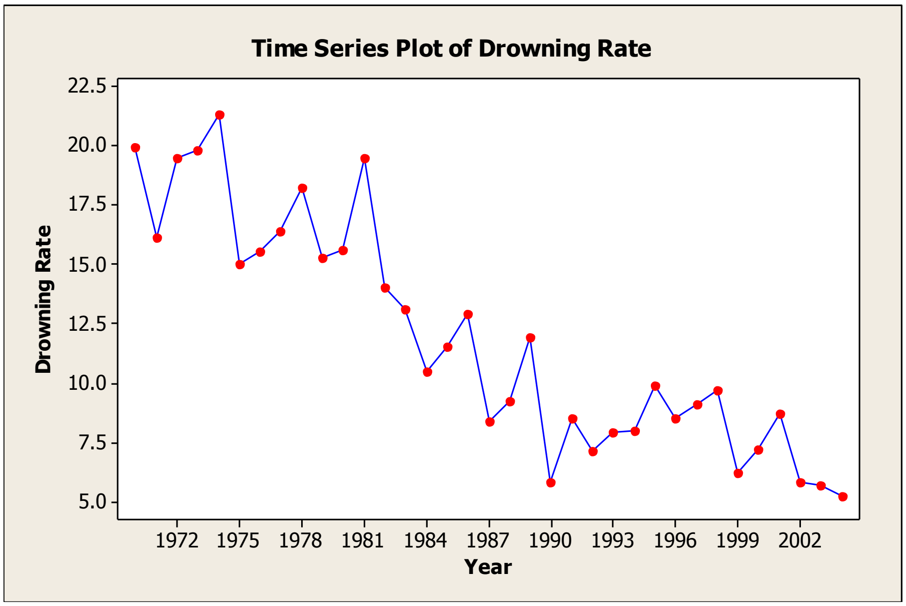
c) Due to the very low data point at 4.07, the mean may be lower than it should be. Therefore, the median would be a better estimate of the density of the earth. The median is not affected by a few outliers.

6-107. a) Stem-and-leaf of Drowning Rate N = 35
Leaf Unit = 0.10

```
 4    5   2788
 5    6   2
 8    7   129
 13   8   04557
 17   9   1279
(1)   10  5
 17   11  59
```

```
15   12   9
14   13   1
13   14   0
12   15   0356
8    16   14
6    17
6    18   2
5    19   5589
1    20
1    21   3
```

Time Series Plots



b)

Descriptive Statistics: Drowning Rate

```
Variable         N  N*    Mean  SE Mean  StDev  Minimum     Q1  Median      Q3
Drowning Rate   35   0  11.911    0.820  4.853    5.200  8.000  10.500  15.600

Variable        Maximum
Drowning Rate    21.300
```

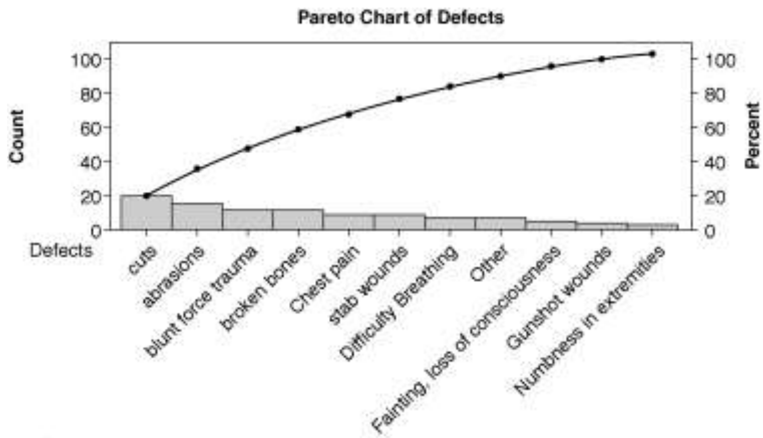
c) Greater awareness of the dangers and drowning prevention programs might have been effective.

d) The summary statistics assume a stable distribution and may not adequately summarize the data because of the trend present.

6-108. a) Sort the categories by the number of instances in each category. Bars are used to indicate the counts and this sorted bar chart is known as a Pareto chart (discussed in Chapter 15).

Cuts	20
Abrasions	16
Blunt force trauma	12
Broken bones	11
Stab wounds	9

Chest pain	9
Other	7
Difficulty breathing	7
Fainting, loss of consciousness	5
Gunshot wounds	4
Numbness in extremities	3



Count	20	16	12	11	9	9	7	7	5	4	3
Percent	19.4	15.5	11.7	10.7	8.7	8.7	6.8	6.8	4.9	3.9	2.9
Cum %	19.4	35.0	46.6	57.3	66.0	74.8	81.6	88.3	93.2	97.1	100.0

b) One would need to follow-up with patients that leave through a survey or phone calls to determine how long they waited before being seen and any other reasons that caused them to leave. This information could then be compiled and prioritized for improvements.

Mind Expanding Exercises

6-109. $\sum_{i=1}^{9} x_i^2 = 63343 \qquad \left(\sum_{i=1}^{9} x_i\right)^2 = 567009 \qquad n = 9$

$$s^2 = \frac{\sum_{i=1}^{9} x_i^2 - \frac{\left(\sum_{i=1}^{9} x_i\right)^2}{n}}{n-1} = \frac{63343 - \frac{567009}{9}}{9-1} = 42.75$$

$$s = \sqrt{42.75} = 6.54$$

Subtract 30 and multiply by 10

$$\sum_{i=1}^{9} x_i^2 = 2626300 \qquad \left(\sum_{i=1}^{9} x_i\right)^2 = 23328900 \qquad n = 9$$

$$s^2 = \frac{\sum_{i=1}^{9} x_i^2 - \frac{\left(\sum_{i=1}^{9} x_i\right)^2}{n}}{n-1} = \frac{2626300 - \frac{23328900}{9}}{9-1} = 4275$$

$$s = \sqrt{4275} = 65.38$$

Yes, the rescaling is by a factor of 10. Therefore, s^2 and s would be rescaled by multiplying s^2 by 10^2 (resulting in 4275) and s by 10 (65.38). Subtracting 30 from each value has no effect on the variance or standard deviation. This is because $V(aX+b) = a^2V(X)$.

6-110. $\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - a)^2$; The sum written in this form shows that the quantity is minimized when a = $\bar{x}$.

6-111. Of the two quantities $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ and $\sum_{i=1}^{n}\left(x_i - \mu\right)^2$, the quantity $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ will be smaller given that $\bar{x} \neq \mu$. This is because $\bar{x}$ is based on the values of the x_i's. The value of μ may be quite different for this sample.

6-112. $y_i = a + bx_i$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad \bar{y} = \frac{\sum_{i=1}^{n}(a + bx_i)}{n} = \frac{na + b\sum_{i=1}^{n} x_i}{n} = a + b\bar{x}$$

$$s_x^{\ 2} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} \quad \text{and} \quad s_x = \sqrt{s_x^{\ 2}}$$

$$s_y^{\ 2} = \frac{\sum_{i=1}^{n}(a + bx_i - a - b\bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n}(bx_i - b\bar{x})^2}{n-1} = \frac{b^2\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = b^2 s_x^{\ 2}$$

Therefore, $s_y = bs_x$

6-113. $\bar{x} = 835.00$ °F $\quad s_x = 10.5$ °F

The results in °C:

$\bar{y} = -32 + 5/9\bar{x} = -32 + 5/9(835.00) = 431.89$ °C

$s_y^{\ 2} = b^2 s_x^{\ 2} = (5/9)^2(10.5)^2 = 34.028$ °C

6-114. Using the results found in a previous exercise with a = $-\frac{\bar{x}}{s}$ and b = 1/*s*, the mean and standard deviation of the z_i are $\bar{z} = 0$ and $s_Z = 1$.

6-115. Yes, in this case, since no upper bound on the last electronic component is available, use a measure of central location that is not dependent on this value. That measure is the median.

Sample Median = $\frac{x_{(4)} + x_{(5)}}{2} = \frac{63+80}{2} = 71.5$ hours

6-116. a) $\bar{x}_{n+1} = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{\sum_{i=1}^{n} x_i + x_{n+1}}{n+1}$

$$\bar{x}_{n+1} = \frac{n\bar{x}_n + x_{n+1}}{n+1}$$

$$\bar{x}_{n+1} = \frac{n}{n+1}\bar{x}_n + \frac{x_{n+1}}{n+1}$$

b) $ns_{n+1}^2 = \sum_{i=1}^{n} x_i^2 + x_{n+1}^2 - \frac{\left(\sum_{i=1}^{n} x_i + x_{n+1}\right)^2}{n+1}$

$$= \sum_{i=1}^{n} x_i^2 + x_{n+1}^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n+1} - \frac{2x_{n+1}\sum_{i=1}^{n} x_i}{n+1} - \frac{x_{n+1}^2}{n+1}$$

$$= \sum_{i=1}^{n} x_i^2 + \frac{n}{n+1}x_{n+1}^2 - \frac{n}{n+1}2x_{n+1}\bar{x}_n - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n+1}$$

$$= \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n+1}\right] + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_{n+1}\bar{x}_n\right]$$

$$= \sum_{i=1}^{n} x_i^2 + \left[\frac{\left(\sum x_i\right)^2}{n} - \frac{\left(\sum x_i\right)^2}{n}\right] - \frac{\left(\sum x_i\right)^2}{n+1} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n} + \frac{(n+1)\left(\sum x_i\right)^2 - n\left(\sum x_i\right)^2}{n(n+1)} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= (n-1)s_n^2 + \frac{\left(\sum x_i\right)^2}{n(n+1)} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= (n-1)s_n^2 + \frac{n\bar{x}^2}{n+1} + \frac{n}{n+1}\left[x_{n+1}^2 - 2x_n\bar{x}_n\right]$$

$$= (n-1)s_n^2 + \frac{n}{n+1}\left(x^2{}_{n+1} - 2x_n\bar{x}_n + \bar{x}_n^2\right)$$

$$= (n-1)s_n^2 + \frac{n}{n+1}\left(x_{n+1}^2 - \bar{x}_n\right)^2$$

c) $\bar{x}_n = 65.811$ inches $\qquad x_{n+1} = 68$

$$s_n^{\ 2} = 4.435 \qquad n = 37 \ s_n = 2.106$$

$$\bar{x}_{n+1} = \frac{37(65.81) + 68}{37 + 1} = 65.87$$

$$s_{n+1} = \sqrt{\frac{(37-1)4.435 + \frac{37}{37+1}(68 - 65.811)^2}{37}} = 2.107$$

6-117. The trimmed mean is pulled toward the median by eliminating outliers.
a) 15% Trimmed Mean = 87.86
b) 12% Trimmed Mean = 90.84
Difference is small
c) No, the differences are small, due to a large data set with no significant outliers.

6-118. If nT/100 is not an integer, calculate the two surrounding integer values and interpolate between the two. For example, if nT/100 = 2/3, one could calculate the mean after trimming 2 and 3 observations from each end and then interpolate between these two means.