CHAPTER 3
STATISTICAL DESCRIPTION OF DATA

SECTION EXERCISES

3.1 c/a/m $\bar{x}$ = 613.48/30 = $20.449. Median = $20.495, the average of the two values in the middle when the data are arranged in order of size.

3.2 c/a/m $\bar{x}$ = 894/20 = 44.70 goals per season. Median = 44.50, average of the two values in the middle when the data are arranged in order of size.

3.3 c/a/m $\bar{x}$ = 1141/20 = 57.05 visitors. Median = 57.50, average of the two values in the middle when the data are arranged in order of size. The mode is 63. There were three different days with 63 visitors.

3.4 c/a/m $\bar{x}$ = 198/10 = 19.8. Median = 18.50, average of the two values in the middle when the data are arranged in order of size. The mode is 30. There were two cartoons that had 30 incidents.

3.5 c/a/m $\bar{x}$ = 138.59/20 = 6.93. Median = 7.095, average of the two values in the middle when the data are arranged in order of size.

3.6 c/a/m $\bar{x}$ = 972.21/15 = $64.81. Median = $65.50, the middle value when the data are arranged in order of size.

3.7 c/a/m $\bar{x}$ = 838/22 = 38.1 yrs. Median = 36.5 yrs., average of the two values in the middle when the data are arranged in order of size.

3.8 c/a/d Expressing dollars and employees in thousands, the weighted mean expenditure per employee is

$$\bar{x} = \frac{[50,845(4738) + 43,690(4637) + 47,098(4540) + 56,121(4397) + 49,369(4026)]}{(4738 + 4637 + 4540 + 4397 + 4026)} = \$49,370.70$$

3.9 c/a/m $\bar{x}$ = 90(0.35) + 78(0.45) + 83(0.20) = 83.2

3.10 d/p/d

a. Motorcyclists usually ride 1 to a motorcycle, so this would be the most frequent value.
b. Mean will be greater, because there is sometimes more than one rider, but always at least one.
c. Mean will be greater, because the distribution is skewed to the right.

3.11 d/p/d

a. Mean will be higher since salaries are usually skewed to the right. Management will emphasize the mean to make the present wage situation look brighter.
b. The union representative will wish to make the present wage situation look worse and therefore will emphasize the median.

3.12 c/c/m The Minitab printout is shown below.

```
Descriptive Statistics: Acad, LawJud

Variable   N  N*    Mean  SE Mean   StDev  Minimum      Q1  Median      Q3
Acad      50   0  3.6080   0.0851  0.6020   2.7000  3.1750  3.4000  4.1250
LawJud    50   0  3.8200   0.0740  0.5233   2.9000  3.4000  3.7000  4.3250

Variable  Maximum
Acad       4.8000
LawJud     4.8000
```

In each set of ratings, the mean exceeds the median. Each distribution is positively skewed.

3.13 c/c/m The Minitab and Excel printouts are shown below.

```
Descriptive Statistics: PSI

Variable             N       Mean     Median     TrMean      StDev    SE Mean
PSI                100     398.86     396.75     398.64      19.58       1.96

Variable       Minimum    Maximum         Q1         Q3
PSI             351.70     454.50     385.18     414.53
```

	E	F
1	*PSI*	
2		
3	Mean	398.86
4	Standard Error	1.96
5	Median	396.75
6	Mode	403.90
7	Standard Deviation	19.58
8	Sample Variance	383.50
9	Kurtosis	-0.24
10	Skewness	0.23
11	Range	102.80
12	Minimum	351.70
13	Maximum	454.50
14	Sum	39885.60
15	Count	100

The mean exceeds the median. The distribution is positively skewed.

3.14 c/c/m

```
Descriptive Statistics: absent by gender

Variable   gender             N       Mean     Median     TrMean      StDev
absent     1                 50      8.040      8.000      8.000      3.326
           2                 50     10.520     10.500     10.568      2.589

Variable   gender       SE Mean    Minimum    Maximum         Q1         Q3
absent     1              0.470      1.000     17.000      5.750     10.000
           2              0.366      3.000     16.000      9.000     12.000
```

The mean number of absences for female employees is less than that for males. The median for the female employees is also lower. For each gender, the mean exceeds the median and the distribution is positively skewed.

3.15 c/c/m

Descriptive Statistics: age by gender

Variable	gender	N	Mean	Median	TrMean	StDev
age	1	50	40.62	39.00	40.70	11.03
	2	50	41.08	41.50	41.05	9.95

Variable	gender	SE Mean	Minimum	Maximum	Q1	Q3
age	1	1.56	19.00	60.00	32.75	50.25
	2	1.41	19.00	63.00	35.00	46.50

The mean age for female employees is less than that for males. The median for the female employees is also lower. For females, the mean age exceeds the median and the distribution is positively skewed.

3.16 d/p/d An ad claim such as "Get up to 70% more miles per gallon by using product x." Most cars tested may have obtained little or no increase in mpg.

3.17 c/a/m Range = 75 - 36 = 39 visitors. MAD = 207.00/20 = 10.35 visitors.
s^2 = 2922.95/19 = 153.84, and s = $\sqrt{153.84}$ = 12.40 visitors.

3.18 c/a/m Range = 30 - 11 = 19 incidents. MAD = 62.0/10 = 6.2 incidents.
s^2 = 475.60/9 = 52.84, and s = $\sqrt{52.84}$ = 7.27 incidents.

3.19 c/a/m

a. μ = 203.8/7 = 29.11 million visitors. Median = 22.4 million visitors. Range = 61.7 - 17.2 = 44.5 million visitors.
   Midrange = (61.7 + 17.2)/2 = 39.45 million visitors.

b. MAD = 90.943/7 = 12.99 million visitors.

c. σ^2 = 1668.769/7 = 238.396, σ = 15.44 million visitors.

3.20 c/a/m

a. $\bar{x}$ = 229/11 = 20.82 cents. Median = 18 cents. Range = 55 - 2 = 53 cents. Midrange = (2 + 55)/2 = 28.5 cents.

b. MAD = 133.454/11 = 12.13 cents.

c. s^2 = 2553.6/10 = 255.36, and s = $\sqrt{255.36}$ = 15.98 cents

3.21 c/a/m

a. $\bar{x}$ = 272/10 = 27.2 mpg. Median = (27 + 29)/2 = 28 mpg. Range = 40 - 10 = 30 mpg.
   Midrange = (10 + 40)/2 = 25 mpg.

b. MAD = 56/10 = 5.6 mpg.

c. s^2 = 583.6/9 = 64.84, and s = $\sqrt{64.84}$ = 8.052 mpg.

3.22 c/a/m

a. $\bar{x}$ = 511.20/8 = 63.9 percent. Median = 67.20 percent, average of the two values in the middle when the data are arranged in order of size.
   Range = 79.1 - 43.9 = 35.2 percent. Midrange = (79.1 + 43.9)/2 = 61.5 percent.

b. MAD = 73.0/8 = 9.125 percent.

c. s^2 = 965.58/7 = 137.94, and s = $\sqrt{137.94}$ = 11.74 percent.

3.23 c/a/m With the data arranged in order of size:

First quartile is in ranked position (11 + 1)/4 = 3; Q_1 = first quartile = 7
Second quartile is in ranked position 2(11 + 1)/4 = 6; Q_2 = second quartile = 18
Third quartile is in ranked position 3(11 + 1)/4 = 9; Q_3 = third quartile = 30
Interquartile range = 30 - 7 = 23; quartile deviation = 23/2 = 11.5

3.24 c/a/m With the data arranged in order of size:
First quartile is in ranked position (10 + 1)/4 = 2.75; Q_1 = 21(0.25) + 23(0.75) = 22.5
Second quartile is in ranked position 2(10 + 1)/4 = 5.5; Q_2 = 27(0.5) + 29(0.5) = 28
Third quartile is in ranked position 3(10 + 1)/4 = 8.25; Q_3 = 32(0.75) + 33(0.25) = 32.25
Interquartile range = 32.25 - 22.5 = 9.75; quartile deviation = 9.75/2 = 4.875

3.25 c/c/m a. and c. The Excel and Minitab descriptive statistics are shown below.

	C	D
1	*Seconds*	
2		
3	Mean	23.3498
4	Standard Error	0.7764
5	Median	22.8600
6	Mode	22.7400
7	Standard Deviation	5.4897
8	Sample Variance	30.1372
9	Kurtosis	0.6938
10	Skewness	0.6721
11	Range	26.13
12	Minimum	13.40
13	Maximum	39.53
14	Sum	1167.49
15	Count	50

```
Descriptive Statistics: Seconds

Variable             N       Mean     Median     TrMean      StDev    SE Mean
Seconds             50     23.350     22.860     23.089      5.490      0.776

Variable       Minimum    Maximum         Q1         Q3
Seconds         13.400     39.530     19.095     26.718
```

The midrange is (13.40 + 39.53)/2 = 26.465

b. The mean absolute deviation must be calculated separately. It is 215.809/50 = 4.316 seconds.

3.26 c/c/m a. and c. The Excel and Minitab descriptive statistics are shown below.

	E	F
1	*absent*	
2		
3	Mean	9.2800
4	Standard Error	0.3216
5	Median	9
6	Mode	8
7	Standard Deviation	3.2164
8	Sample Variance	10.3451
9	Kurtosis	-0.0461
10	Skewness	-0.2065
11	Range	16
12	Minimum	1
13	Maximum	17
14	Sum	928
15	Count	100

```
Descriptive Statistics: absent

Variable            N       Mean     Median     TrMean      StDev    SE Mean
absent            100      9.280      9.000      9.300      3.216      0.322

Variable      Minimum    Maximum         Q1         Q3
absent          1.000     17.000      8.000     12.000
```

The midrange is (1 + 17)/2 = 9

b. The mean absolute deviation must be calculated separately. It is 255.12/100 = 2.55 absences.

3.27 c/c/m a. and c. The Excel and Minitab descriptive statistics are shown below.

	C	D
1	*meters*	
2		
3	Mean	90.7713
4	Standard Error	0.9347
5	Median	91.4
6	Mode	85.6
7	Standard Deviation	8.3606
8	Sample Variance	69.9000
9	Kurtosis	-0.1468
10	Skewness	0.0717
11	Range	40.4
12	Minimum	71.8
13	Maximum	112.2
14	Sum	7261.7
15	Count	80

```
Descriptive Statistics: meters

Variable            N       Mean     Median     TrMean      StDev    SE Mean
meters             80     90.771     91.400     90.742      8.361      0.935

Variable      Minimum    Maximum         Q1         Q3
meters         71.800    112.200     85.025     96.275
```

The midrange is (71.8 + 112.2)/2 = 92.0

b. The mean absolute deviation must be calculated separately. It is 536.3575/80 = 6.70 meters.

3.28 c/a/m

a. The median is approximately 37.5 defects per day. The first quartile is approximately 37 defects per day. The third quartile is approximately 39 defects per day.
b. The asterisks at the right are outliers, indicating two days on which unusually large numbers of defects were produced. The production supervisor should try to determine if anything out of the ordinary was happening at the plant on those days.
c. The distribution is positively skewed.

3.29 c/a/e
a. At least $(1-(1/2.5^2))*100 = 84\%$
b. At least $(1-(1/3^2))*100 = 88.89\%$
c. At least $(1-(1/5^2))*100 = 96\%$

3.30 c/a/m Standardized data values: -1.18, -0.99, -0.87, -0.55, -0.36, -0.18, -0.05, 0.26, 0.57, 1.20, and 2.14; 90.9% of them are within 1.5 standard deviation units of the mean. Chebyshev's Theorem states that at least $(1 - (1/1.5^2))*100 = 55.6\%$ should fall within that interval, and these results support the theorem.

3.31 c/a/m Standardized data values: -2.14, -0.77, -0.52, -0.02, -0.02, 0.22, 0.35, 0.60, 0.72, and 1.59; 90% of them are within 2.0 standard deviations of the mean. Chebyshev's Theorem states that at least $(1 - (1/2^2))*100 = 75\%$ should fall within that interval, and these results support the theorem.

3.32 c/a/m Using the empirical rule:
a. 95%. This is the percentage of values that are within ±2 standard deviations of the mean.
b. 16%, or 50% - 34%. Recall that 68% of the values are within ±1 standard deviation of the mean.
c. 2.5%, or 50% - 47.5%; 95% of the values are within ±2 standard deviations of the mean.
d. 81.5%, obtained by 34% (the area between the mean and 11,500) plus 47.5% (the area from the mean to 13,000).

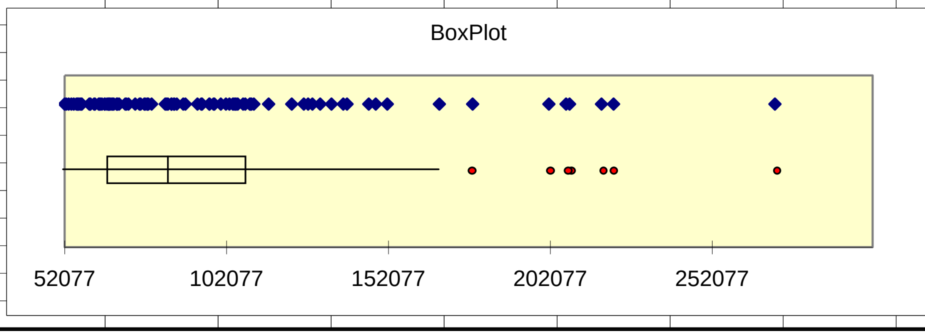
3.33 c/a/m Using the empirical rule:
a. 68%. This is the percentage of values that are within ±1 standard deviation of the mean.
b. 2.5%, or 50% - 47.5%; 95% of the values are within ±2 standard deviations of the mean.
c. 84%, or 50% (the area to the left of the mean) plus 34% (the area from the mean to 580).
d. 13.5%, obtained by 47.5% (the area between the mean and 680) minus 34% (the area between the mean and 580).

3.34 c/a/m Coefficient of variation = $s/\bar{x}$ = (140/1235)*100 = 11.34% for data set A. Coefficient of variation = $s/\bar{x}$ = (1.87/15.7)*100 = 11.91% for data set B. Set B has greater relative dispersion.

3.35 c/a/m Coefficient of variation = $s/\bar{x}$ = (87/315)*100 = 27.62% for Barnsboro. Coefficient of variation = $s/\bar{x}$ = (1800/8350)*100 = 21.56% for Wellington. Barnsboro has greater relative dispersion.

3.36 c/c/m
a. Box-and-whisker plot and listing of key descriptors. The distribution is positively skewed.

	A	B	C	D	E	F	G	H	I
1	**Box Plot**								
2									
3	*Income*								
4	Smallest = 52077								
5	Q1 = 65615								
6	Median = 83786								
7	Q3 = 107754.75								
8	Largest = 271535								
9	IQR = 42139.75								
10	Outliers: 271535, 221684, 217977, 208105, 206967, 201723, 178107,								



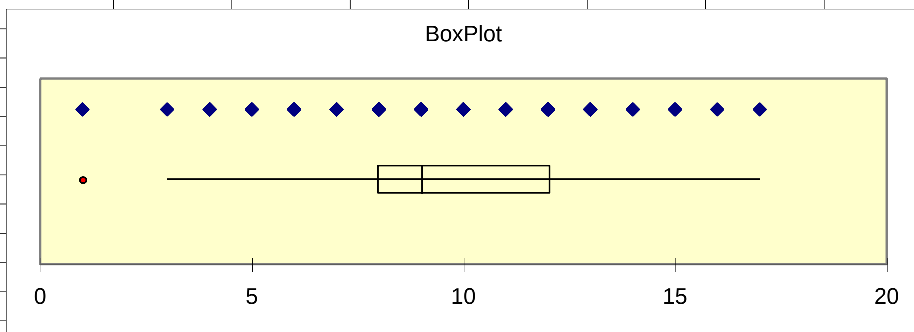
b. A portion of the data and standardized data, and descriptive statistics for the 100 standardized values.

	A	B	C	D	E
1	**Income**	**Std_Inc**		*Std_Inc*	
2	109289	0.3354			
3	68419	-0.6239		Mean	0.0000
4	178107	1.9505		Standard Error	0.1000
5	84018	-0.2578		Median	-0.2632
6	201723	2.5048		Mode	#N/A
7	77965	-0.3998		Standard Deviation	1.0000
8	78928	-0.3772		Sample Variance	1.0000
9	56128	-0.9123		Kurtosis	3.6763
10	94504	-0.0116		Skewness	1.8345
11	76700	-0.4295		Range	5.1507
12	56330	-0.9076		Minimum	-1.0074
13	96830	0.0429		Maximum	4.1433
14	98100	0.0728		Sum	0.0000
15	115090	0.4715		Count	100

3.37 c/c/m

a. Box-and-whisker plot and listing of key descriptors. The distribution is positively skewed.

	A	B	C	D	E	F	G	H
1	**Box Plot**							
2								
3	*absent*							
4	Smallest = 1							
5	Q1 = 8							
6	Median = 9							
7	Q3 = 12							
8	Largest = 17							
9	IQR = 4							
10	Outliers: 1, 1,							
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								



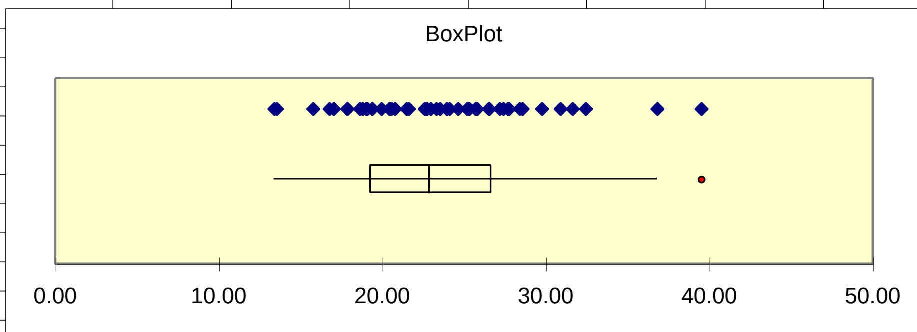
b. A portion of the data and standardized data, and descriptive statistics for the 100 standardized values.

	D	E	F	G	H
1	absent	StdAbsent		*StdAbsent*	
2	8	-0.3980			
3	10	0.2239		Mean	0.0000
4	13	1.1566		Standard Error	0.1000
5	8	-0.3980		Median	-0.0871
6	13	1.1566		Mode	-0.3980
7	10	0.2239		Standard Deviation	1.0000
8	11	0.5348		Sample Variance	1.0000
9	7	-0.7089		Kurtosis	-0.0461
10	1	-2.5743		Skewness	-0.2065
11	11	0.5348		Range	4.9745
12	4	-1.6416		Minimum	-2.5743
13	8	-0.3980		Maximum	2.4002
14	13	1.1566		Sum	0.0000
15	8	-0.3980		Count	100
16	11	0.5348			

3.38 c/c/m

a. Box-and-whisker plot and listing of key descriptors. The distribution is positively skewed.

	A	B	C	D	E	F	G	H
1	**Box Plot**							
2								
3	*Seconds*							
4	Smallest = 13.4							
5	Q1 = 19.095							
6	Median = 22.86							
7	Q3 = 26.7175							
8	Largest = 39.53							
9	IQR = 7.6225							
10	Outliers: 39.53,							



b. A portion of the data and standardized data, and descriptive statistics for the 50 standardized values.

	A	B	C	D	E
1	**Seconds**	**StdSecs**		*StdSecs*	
2	19.11	-0.7723			
3	13.56	-1.7833		Mean	0.0000
4	22.98	-0.0674		Standard Error	0.1414
5	32.46	1.6595		Median	-0.0892
6	19.05	-0.7832		Mode	-0.1111
7	27.19	0.6995		Standard Deviation	1.0000
8	19.39	-0.7213		Sample Variance	1.0000
9	23.96	0.1112		Kurtosis	0.6938
10	27.70	0.7924		Skewness	0.6721
11	19.02	-0.7887		Range	4.7598
12	22.60	-0.1366		Minimum	-1.8124
13	20.44	-0.5300		Maximum	2.9474
14	28.59	0.9545		Sum	0.0000
15	24.13	0.1421		Count	50

3.39 c/a/d

a. Frequency distribution with classes having widths of 1:

class	m_i	f_i	f_im_i	$f_im_i^2$
4 - under 5	4.5	2	9.0	40.50
5 - under 6	5.5	4	22.0	121.00
6 - under 7	6.5	3	19.5	126.75
7 - under 8	7.5	6	45.0	337.50
8 - under 9	8.5	3	25.5	216.75
9 - under 10	9.5	2	19.0	180.50
		sum = 20	sum = 140.0	sum = 1023.0

The estimates are $\bar{x} = 140.0/20 = 7.00$ and $s^2 = \frac{1023.0 - (20)(7.0)^2}{19} = 2.263,\ s = 1.504$

b. The mean and standard deviation for the actual data were 6.930 and 1.399, respectively.

c. Frequency distribution with classes having widths of 0.5:

class	m_i	f_i	f_im_i	$f_im_i^2$
4 - under 4.5	4.25	0	0.00	0.0000
4.5 - under 5.0	4.75	2	9.50	45.1250
5.0 - under 5.5	5.25	3	15.75	82.6875
5.5 - under 6.0	5.75	1	5.75	33.0625
6.0 - under 6.5	6.25	1	6.25	39.0625
6.5 - under 7.0	6.75	2	13.50	91.1250
7.0 - under 7.5	7.25	5	36.25	262.8125
7.5 - under 8.0	7.75	1	7.75	60.0625
8.0 - under 8.5	8.25	1	8.25	68.0625
8.5 - under 9.0	8.75	2	17.50	153.1250
9.0 - under 9.5	9.25	2	18.50	171.1250
9.5 - under 10.0	9.75	0	0.00	0.0000
		sum = 20	sum =139.0	sum = 1006.25

The estimates are now $\bar{x} = 139.0/20 = 6.95$ and $s^2 = \frac{1006.25 - (20)(6.95)^2}{19} = 2.116,\ s = 1.455$

The approximations have improved.

d. If each data value were the midpoint of its own class, the approximate values would be identical to the exact values.

3.40 c/a/d

m_i	f_i	f_im_i	$f_im_i^2$
10	7	70	700

20	9	180	3,600
30	12	360	10,800
40	14	560	22,400
50	13	650	32,500
60	9	540	32,400
70	8	560	39,200
80	11	880	70,400
90	10	900	81,000
100	7	700	70,000
	sum = 100	sum = 5400	sum = 363,000

Approximate values are $\bar{x} = 5400/100 = 54$ and $s^2 = \dfrac{363,000 - (100)(54)^2}{99} = 721.21,\ s = 28.86$

3.41 c/a/d

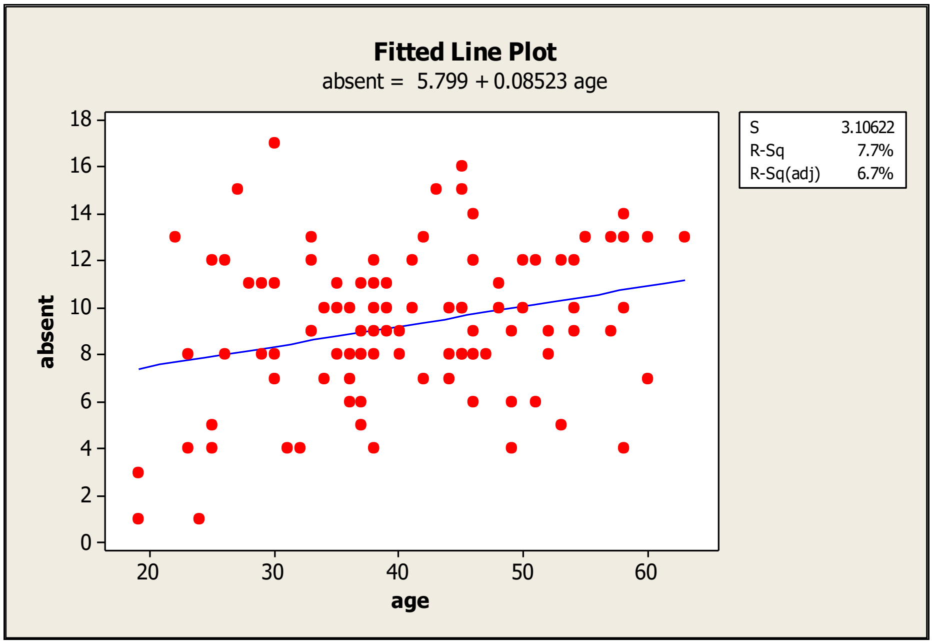
m_i	f_i	$f_i m_i$	$f_i m_i^2$
5	25	125	625
15	17	255	3,825
25	15	375	9,375
35	9	315	11,025
45	10	450	20,250
55	4	220	12,100
	sum = 80	sum = 1740	sum = 57,200

Approximate values: $\bar{x} = 1740/80 = 21.75$ and $s^2 = \dfrac{57,200 - (80)(21.75)^2}{79} = 245.00,\ s = 15.65$

3.42 d/p/e The coefficient of determination is the proportion of the variation in y that is explained by the best-fit linear equation. It is a measure of the strength of the relationship between the variables.

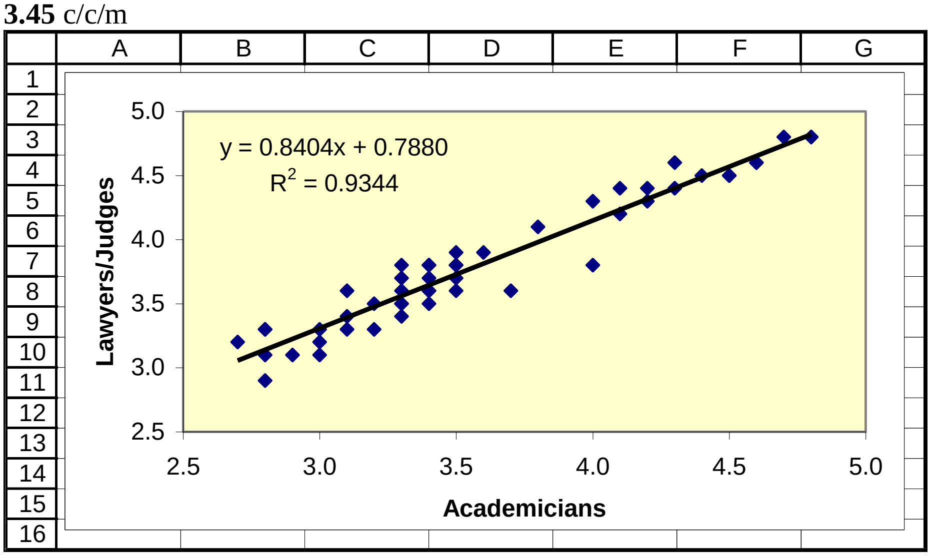
3.43 c/a/e Because the variables are inversely related, r will be negative. Thus, r will be the negative square root of 0.64, or r = -0.8.

3.44 c/c/m



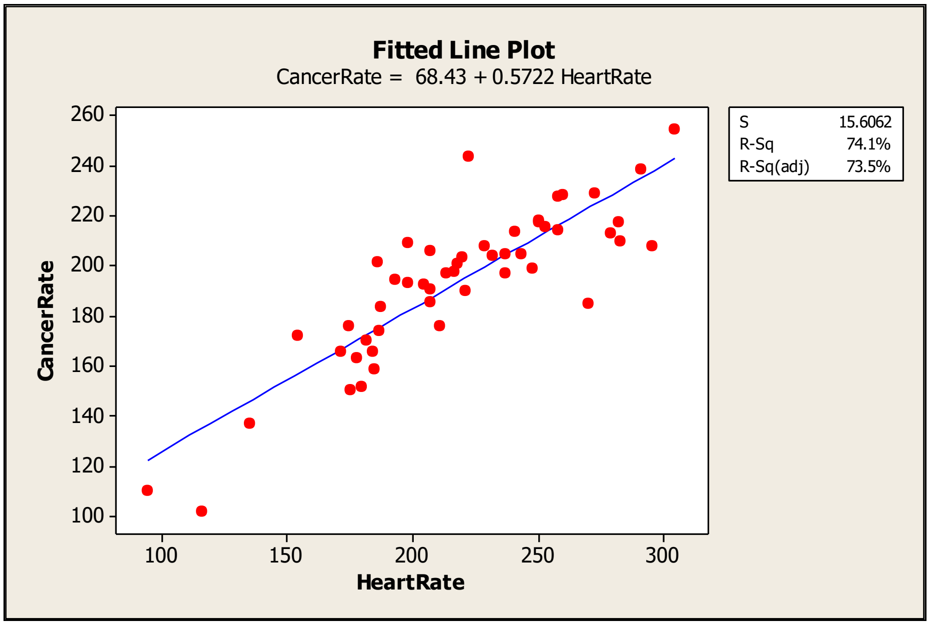
The equation explains 7.7% of the variation in the number of absences. The coefficient of correlation is the positive (since the slope is positive) square root of 0.077, or r = 0.277.

3.45 c/c/m



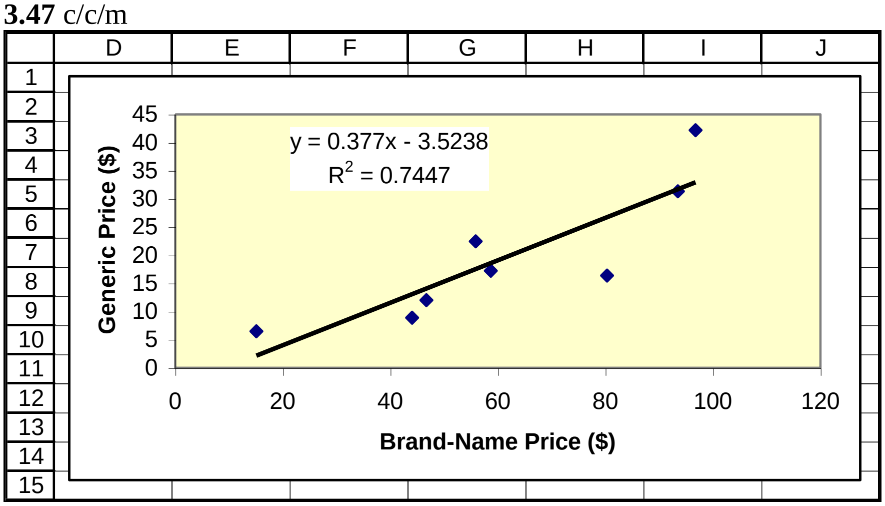
Ratings from the academicians explain 93.44% of the variation in the ratings of the lawyers/judges. The coefficient of correlation is the positive (since the slope is positive) square root of 0.9344, or r = 0.967.

3.46 c/c/m



The equation explains 74.1% of the variation in the cancer rates. The coefficient of correlation is the positive (since the slope is positive) square root of 0.741, or r = 0.861.

3.47 c/c/m



The equation explains 74.47% of the variation in the generic prices. The coefficient of correlation is the positive (since the slope is positive) square root of 0.7447, or r = 0.863.

CHAPTER EXERCISES

3.48 c/a/m $\bar{x}$ = (1.25 + 2.36 + 2.50 + 2.15 + 4.55 + 1.10 + 0.95)/7 = $2.12. Yes, the service to the first seven customers was profitable.

3.49 c/a/d $\bar{x}$ = (5(50) + 2(30) + 4(60) + 10(20))/(50 + 30 + 60 + 20) = $4.69

3.50 c/a/m

$\bar{x} = \sum x / n = 4225.7 / 20 = 211.285$; Median = (206.4 + 210.6)/2 = 208.5; There is no mode.

3.51 c/a/m $\bar{x} = \sum x / n = 942 / 8 = 117.75$ Median = (113 + 122)/2 = 117.5; There is no mode.

3.52 c/a/m

a. $\bar{x} = \sum x / n = 552.26 / 18 = 30.68$ mph. Median = (30 + 30)/2 = 30 mph. b. Mode = 30 mph.

3.53 d/p/m The distribution is not symmetrical. It is positively skewed.

3.54 c/p/d

a. The mean exceeds the median and, based on the rough character-graph boxplot shown below, the distribution appears to be very slightly positively skewed.

```
                 ---------
     ----------I    +   I------------
                 ---------
-----+---------+---------+---------+-----
     50       100       150        200
```

b. Approximately 2.5%, obtained by 50% (the area to the left of the mean) minus 47.5% (the area between 64 cups and the mean). According to the empirical rule, approximately 95% of the data values will lie within 2 standard deviations of the mean; 64 cups is about two standard deviations less than the mean.

3.55 d/p/m

a. Since all values should be increased by 0.1, the sample mean will increase by 0.1 to 3.1 lbs. Since the relative variation is unchanged, the sample standard deviation will still be 0.5 lbs.

b. Using the empirical rule, this would be 4.1 lbs., obtained by 3.1 + 2(0.5). Approximately 95% of the data values will lie within 2 standard deviations of the mean.

3.56 c/a/m

a. $\bar{x} = \sum x / n = 1116.0 / 5 = 223.2$ stoppages. Median = 235 stoppages (3rd value when data are arranged in order of size).
Range = 424 – 44 = 380 stoppages
Midrange = (44 + 424)/2 = 234.0 stoppages

b. $\text{MAD} = \sum |x_i - \bar{x}| / n = 514.8 / 5 = 102.96$ stoppages

c. $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1} = 80{,}694.8 / 4 = 20{,}173.7$, $s = \sqrt{s^2} = 142.0$ stoppages

3.57 c/a/m

a. $\bar{x} = \sum x / n = 33.59 / 16 = 2.10$ tons, Median = (2.08 + 2.15)/2 = 2.115 tons.
Range = 2.31 - 1.85 = 0.46 tons Midrange = (1.85 + 2.31)/2 = 2.08 tons.

b. $\text{MAD} = \sum | x_i - \bar{x} | / n = 2.19 / 16 = 0.137$ tons

c. $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1} = 0.3669 / 15 = 0.02446, \; s = \sqrt{s^2} = 0.156$ tons

3.58 c/a/m The median is approximately 99 gallons. The first quartile is approximately 92 gallons. The third quartile is approximately 104 gallons. The range is approximately 120 - 80 = 40 gallons. The distribution appears to be slightly negatively skewed.

3.59 c/a/m The median is approximately 120 watts. The first quartile is approximately 116 watts. The third quartile is approximately 124 watts. The range is approximately 130 - 110 = 20 watts. The distribution appears to be symmetrical.

3.60 c/a/m

a. $\bar{x} = \sum x / n = 1.84 / 25 = 0.0736\%$ $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1} = 0.112176 / 24, \; s = 0.0684\%$

b. Chebyshev's Theorem states that at least $(1 - (1/1.5^2))*100 = 55.6\%$ should fall within 1.5 standard deviation units. For this data, all except the largest three values, or 88% of the data set, fall within 1.5 standard deviation units.

c. Coefficient of variation = $(s/\bar{x})*100\%$ = (0.0684/0.0736)*100% = 92.9%

3.61 c/a/m Exercise 3.57: coefficient of variation = $(s/\bar{x})*100$ = (0.156/2.10)*100 = 7.43 %
Exercise 3.60: coefficient of variation = $(s/\bar{x})*100$ = (0.0684/0.0736)*100% = 92.9%
There is greater variation for the data in exercise 3.60.

3.62 c/a/m

m_i	f_i	$f_i m_i$	$f_i m_i^2$

50	27	1350	67,500
150	11	1650	247,500
250	4	1000	250,000
350	1	350	122,500
450	2	900	405,000
550	1	550	302,500
650	0	0	0
750	1	750	562,500
850	1	850	722,500
950	0	0	0
1050	1	1050	1,102,500
1150	1	1150	1,322,500
	sum = 50	sum = 9600	sum = 5,105,000

Approximate values:

$$\mu = 9600/50 = 192, \text{ and } \sigma^2 = \frac{5{,}105{,}000 - (50)(192)^2}{50} = 65{,}236 \text{ and } \sigma = 255.4$$

3.63 c/a/m Median = (24 + 25)/2 = 24.5 pages. First Quartile = 22(0.75) + 22(0.25) = 22 pages. Third Quartile = 29(0.25) + 35(0.75) = 33.5 pages.

```
Variable              N       Mean     Median     TrMean      StDev    SE Mean
pages                20      25.65      24.50      25.72       8.01       1.79

Variable        Minimum    Maximum         Q1         Q3
pages             11.00      39.00      22.00      33.50
```

3.64 c/a/m

Class	m_i	f_i	$f_i m_i$	$f_i m_i^2$
10 - under 20	15	4	60	900
20 - under 30	25	11	275	6,875
30 - under 40	35	5	175	6,125
		sum = 20	sum = 510	sum = 13,900

Approximate values: $\bar{x}$ = 510/20 = 25.5 $\quad s^2 = \frac{13{,}900 - (20)(25.5)^2}{19} = 47.105, \; s = 6.863$

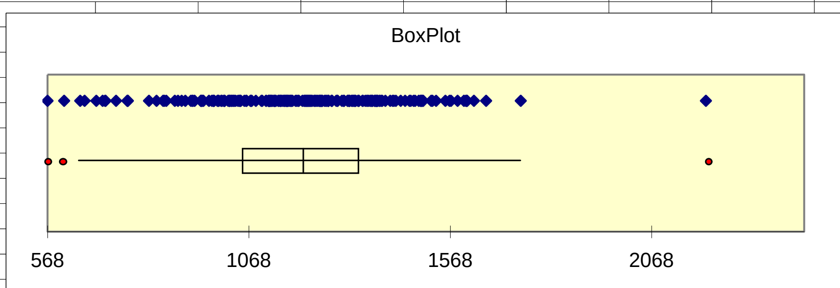
3.65 c/c/m

a. Descriptive statistics.

	A	B
1	*Electricity*	
2		
3	Mean	1196.000
4	Standard Error	13.953
5	Median	1203.000
6	Mode	1317.000
7	Standard Deviation	220.624
8	Sample Variance	48674.916
9	Kurtosis	1.495
10	Skewness	0.113
11	Range	1635
12	Minimum	568
13	Maximum	2203
14	Sum	299000
15	Count	250

b. Boxplot with interpretation statistics.

	A	B	C	D	E	F	G	H	I
1	**Box Plot**								
2									
3	*Electricity*								
4	Smallest = 568								
5	Q1 = 1047.75								
6	Median = 1203								
7	Q3 = 1334.5								
8	Largest = 2203								
9	IQR = 286.75								
10	Outliers: 2203, 609, 568,								



c. As shown in part (b), there are two outlier households ($568 and $609) at the low end and one ($2203) at the high end of electricity expenditures. Energy-conservation officials may wish to examine these households for habits or characteristics that should either be emulated or avoided.

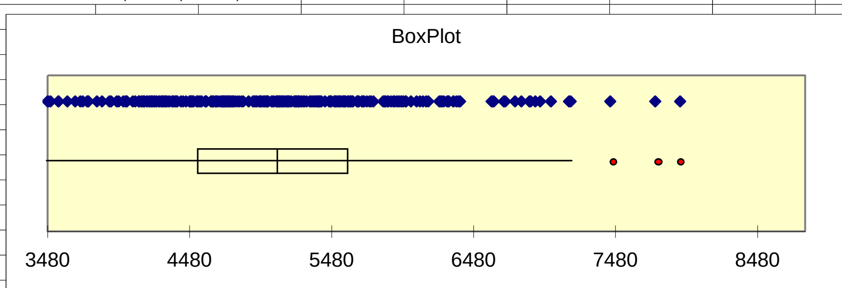
3.66 c/c/m

a. Descriptive statistics.

	A	B
1	*$cost*	
2		
3	Mean	5111.000
4	Standard Error	46.551
5	Median	5101.000
6	Mode	4482.000
7	Standard Deviation	806.288
8	Sample Variance	650100.602
9	Kurtosis	0.390
10	Skewness	0.528
11	Range	4455
12	Minimum	3480
13	Maximum	7935
14	Sum	1533300
15	Count	300

b. Boxplot with interpretation statistics.

	A	B	C	D	E	F	G	H	I
1	**Box Plot**								
2									
3	*$cost*								
4	Smallest = 3480								
5	Q1 = 4532.25								
6	Median = 5101								
7	Q3 = 5589.75								
8	Largest = 7935								
9	IQR = 1057.5								
10	Outliers: 7935, 7759, 7444,								



c. As shown in part (b), there are three outlier couples ($7935, $7759, and $7444) at the high end of honeymoon expenditures. Cruise lines, resort areas, and various governmental tourism-promotion agencies could be interested in finding out more about the age, media habits, and other characteristics of these people so as to be able to reach and persuade others like them to spend their honeymoons or vacations at their venues.

3.67 c/c/m

a. Descriptive statistics.

	A	B
1	*ACT*	
2		
3	Mean	20.465
4	Standard Error	0.257
5	Median	20.800
6	Mode	25.000
7	Standard Deviation	5.131
8	Sample Variance	26.329
9	Kurtosis	-0.388
10	Skewness	-0.272
11	Range	26.1
12	Minimum	6.1
13	Maximum	32.2
14	Sum	8185.8
15	Count	400

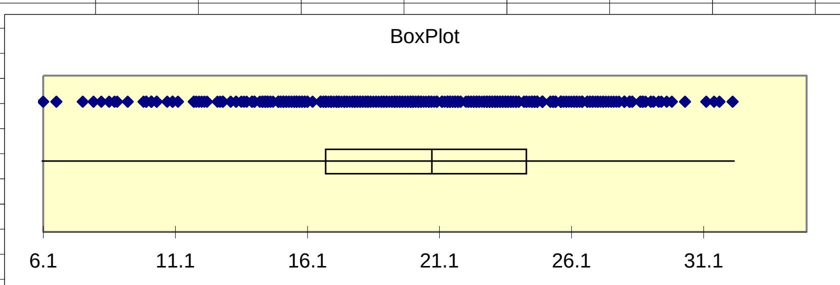
Descriptive Statistics: ACT

```
Variable     N  N*    Mean  SE Mean  StDev  Minimum      Q1  Median      Q3
ACT        400   0  20.465    0.257  5.131    6.100  16.800  20.800  24.375

Variable  Maximum
ACT        32.200
```

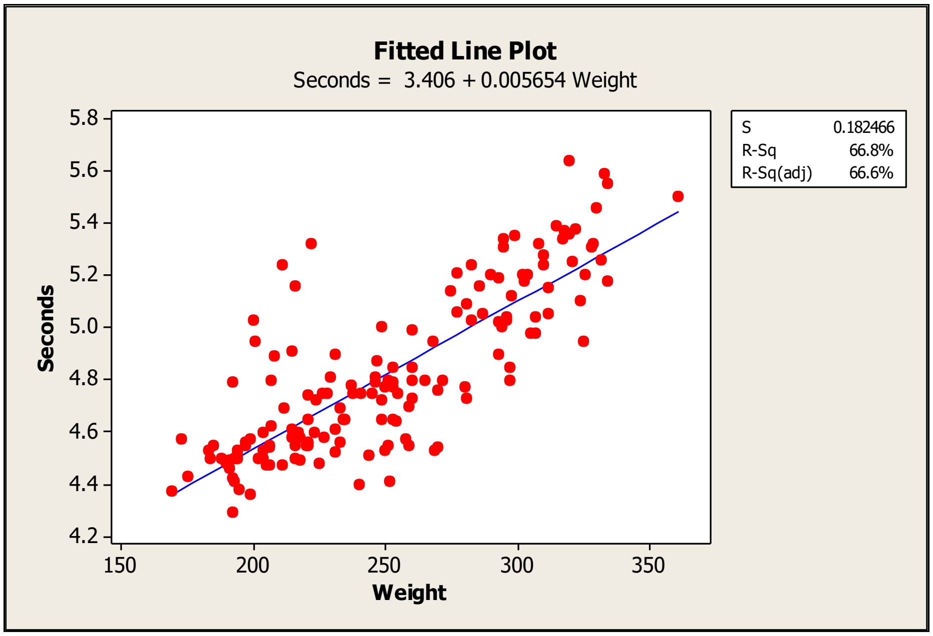
b. Boxplot with interpretation statistics.

	A	B	C	D	E	F	G	H	I
1	**Box Plot**								
2									
3	*ACT*								
4	Smallest = 6.1								
5	Q1 = 16.8								
6	Median = 20.8								
7	Q3 = 24.375								
8	Largest = 32.2								
9	IQR = 7.575								
10	Outliers:								



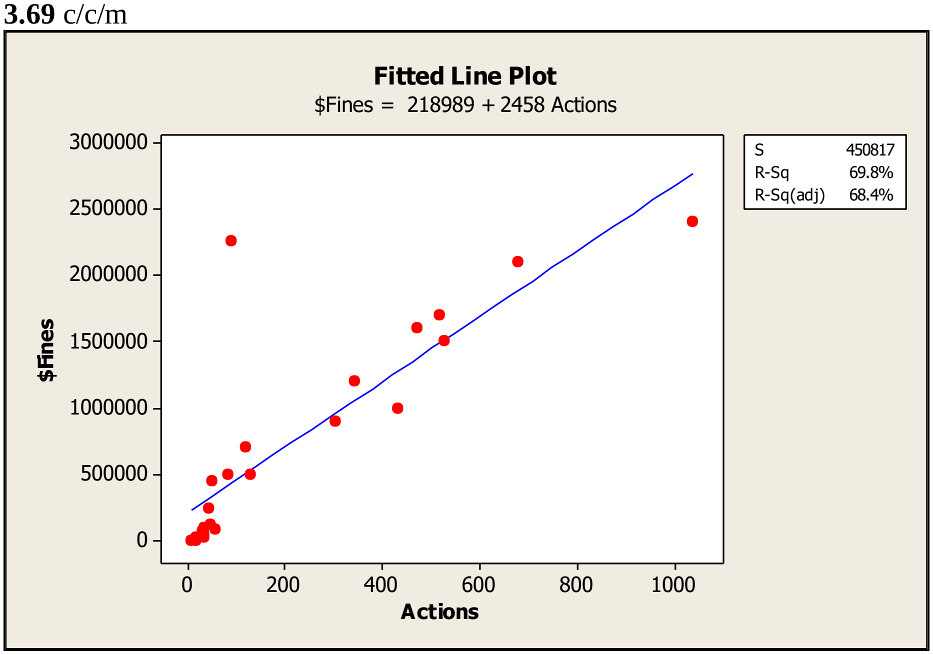
c. A test-taker would have to score 25 (24.375, rounded up) on the math portion to be higher than 75% of the sample members. He or she would have to score 17 (16.8, rounded up) to be higher than 25% of the sample members. These correspond to the third and first quartiles, respectively.

3.68 c/c/m



With the linear estimation equation, player weight explains 66.8% of the variation in 40-yard times. Since the slope is positive, the coefficient of correlation is the positive square root of 0.668, or r = 0.82.

3.69 c/c/m



Through the linear estimation equation, the number of actions explains 69.8% of the variation in fine amounts. Because the slope is positive, the coefficient of correlation is the positive square root of 0.698, or r = 0.84.

INTEGRATED CASES

THORNDIKE SPORTS EQUIPMENT

1. Measures of central tendency and dispersion for the new golf balls, using Minitab:

Descriptive Statistics: NewBall

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
NewBall	25	251.53	3.57	17.86	223.70	235.45	252.80	264.70	294.10

The mean is 251.53 and the median is 252.80. Both are good measurements to reflect central tendency. The standard deviation is 17.86, measuring the dispersion of the data.

2. Measures of central tendency and dispersion for the conventional golf balls:

Descriptive Statistics: ConBall

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
ConBall	25	238.04	3.86	19.29	201.00	222.45	240.30	254.15	267.90

The mean is 238.04 and the median is 240.30. The standard deviation is 19.29.

3. The mean and median distances traveled by the new ball are considerably larger than the corresponding values for the old ball. This indicates that the new ball is “more lively” than the old ball, and on average travels further. Another indication of a greater distance for the new ball can be seen in the ranges. The range of the new ball is from 223.70 to 294.10; whereas, the range of the old ball is from 201.00 to 267.90. The standard deviations of the samples are relatively similar, with a larger dispersion among the distances of the old ball than the new one.

SPRINGDALE SHOPPING SURVEY

This exercise is based on SHOPPING, the Springdale shopping survey database. There are 30 variables and 150 cases (respondents) in this database. Using Minitab:

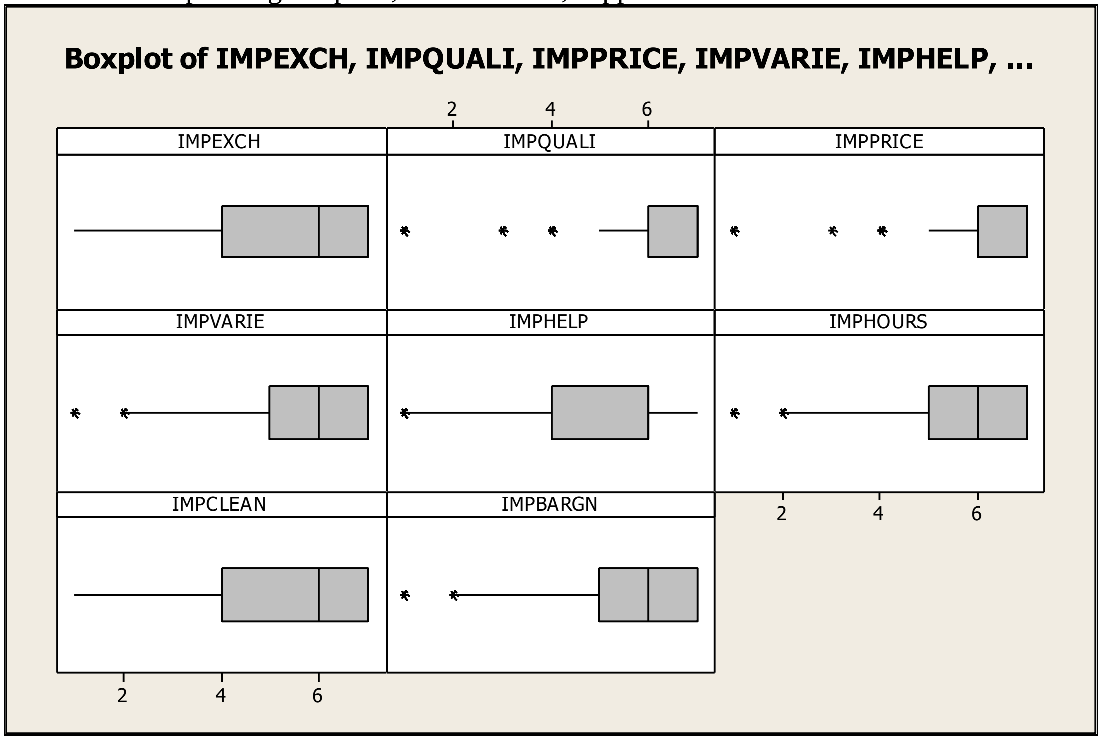
1a. Descriptive statistics, including mean and median.

Descriptive Statistics: IMPEXCH, IMPQUALI, IMPPRICE, IMPVARIE, IMPHELP, ...

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
IMPEXCH	150	5.260	0.159	1.947	1.000	4.000	6.000	7.000
IMPQUALI	150	6.293	0.111	1.359	1.000	6.000	7.000	7.000
IMPPRICE	150	6.4267	0.0962	1.1778	1.0000	6.0000	7.0000	7.0000
IMPVARIE	150	5.653	0.113	1.381	1.000	5.000	6.000	7.000
IMPHELP	150	5.160	0.139	1.699	1.000	4.000	6.000	6.000
IMPHOURS	150	5.387	0.129	1.579	1.000	5.000	6.000	7.000
IMPCLEAN	150	5.320	0.132	1.619	1.000	4.000	6.000	7.000
IMPBARGN	150	5.667	0.114	1.398	1.000	5.000	6.000	7.000

Variable	Maximum
IMPEXCH	7.000
IMPQUALI	7.000
IMPPRICE	7.0000
IMPVARIE	7.000
IMPHELP	7.000
IMPHOURS	7.000
IMPCLEAN	7.000
IMPBARGN	7.000

1b. In part (a), for all 8 variables, the median exceeds the mean, indicating negative skewness. The corresponding boxplots, shown below, support this conclusion.



2. Quality and price seem to be the most important attributes in respondents' choice of a shopping area. Helpful staff, clean store, and convenient hours are the least important attributes.

3. Descriptive statistics for variables 29 and 30.

Descriptive Statistics: RESPHOUS, RESPAGE

```
Variable    N   Mean  SE Mean  StDev  Minimum     Q1  Median     Q3  Maximum
RESPHOUS  150  3.120    0.143  1.757    1.000  1.750   3.000  4.000    8.000
RESPAGE   150  32.05     1.22  14.96    17.00  21.00   26.00  38.25    74.00
```

		Mean	StDev	Coef. of Variation (StDev/Mean)*100
C29	RESPHOUS	3.120	1.757	56.3
C30	RESPAGE	32.05	14.96	46.7

Based on the coefficients of variation shown above, C29 (RESHOUS) exhibits greater variation than C30 (RESPAGE).

4. Coefficient of correlation between variables 29 and 30.

Correlations: RESPHOUS, RESPAGE

```
Pearson correlation of RESPHOUS and RESPAGE = -0.099
P-Value = 0.228
```

With r = -0.099, $(-0.099)^2$*100 is just 0.98%. Slightly less than 1% of the variation in the number of persons in the respondent's household is explained by the respondent's age.

BUSINESS CASE

BALDWIN COMPUTER SALES (A)

1. The mean score on the screening test is higher for those who did not default, shown in the Minitab printout below as 63.439 versus 56.65.

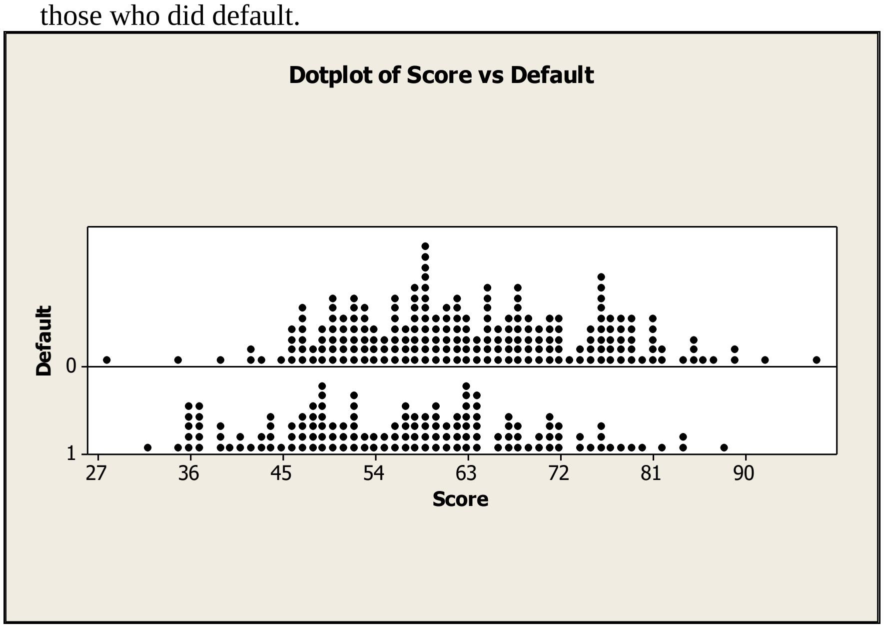
Descriptive Statistics: Score

Variable	Default	N	Mean	SE Mean	StDev	Minimum	Q1	Median
Score	0	205	63.439	0.835	11.954	28.000	54.000	62.000
	1	137	56.65	1.06	12.45	32.00	48.00	57.00

Variable	Default	Q3	Maximum
Score	0	72.000	97.000
	1	64.00	88.00

2. The third quartile for those who did not default was 72.00 -- for this group, 75% scored 72.00 or lower on the screening test. If a score of 72.00 had been had been established as a cutoff for receiving a computer loan, 25% of those who repaid would have been denied a loan in the first place. Granting a loan solely on the basis of a screening test score of 72.00 or above would seem to be rather unfair to those students who end up repaying the loan, as 25% of them would not have received the loan they ended up repaying.

3. The Minitab dotplots below visually compare the screening test scores of students who did not default on their computer loan to the scores of those who defaulted. The distribution of screening test scores for those who did not default is most definitely shifted to the right of the distribution of scores for those who did default.



4. Based on the preceding results, the screening test does appear to be potentially useful as one of the factors in helping Baldwin predict whether a given applicant will end up defaulting on his or her computer loan. However, Baldwin might benefit from considering other factors as well -- note that four students with screening test scores well over 72.00 (ranging from the low 80s to the high 80s) ended up defaulting on their computer loans. Also, one of the students who did not default had the lowest screening score of all, shown in the dotplots above as slightly above 27.