Chapter 4

4.1 a $\bar{x} = \frac{\sum x_i}{n} = \frac{7+4+0+2+7+3+1+9+12}{9} = \frac{45}{9} = 5.0$

Ordered data: 0, 1, 2, 3, 4, 7, 7, 9, 12; Median = 4

Mode = 7

4.2 $\bar{x} = \frac{\sum x_i}{n} = \frac{0+5+3+8+6+3+12+8+4+9+16+8+0+6+2}{15} = \frac{90}{15} = 6.0$

Ordered data: 0, 0, 2, 3, 3, 4, 5, 6, 6, 8, 8, 8, 9, 12, 16; Median = 6

Mode = 8

4.3 a $\bar{x} = \frac{\sum x_i}{n} = \frac{15.2+17.3+8.8+21.7+6.6+20.5+15.3+1\grave{}9.9+3.7+32.0+14.4+13.3}{12}$

$= \frac{188.7}{12} = 15.73$

Ordered data: 3.7, 6.6, 8.8, 13.3, 14.4, 15.2, 15.3, 17.7, 19.9, 20.5, 21.7, 32.0; Median = 15.25

Mode = all

b The mean distance is 15.73 km. Half the sample commute more than 15.25 km and half commute less.

4.4 a $\bar{x} = \frac{\sum x_i}{n} = \frac{10+6+5+2+6+4+9+13+10+12+7+4+9+13+15+8+11+12+4+0}{20}$

$= \frac{160}{20} = 8.0$

Ordered data: 0, 2, 4, 4, 4, 5, 6, 6, 7, 8, 9, 9, 10, 10, 11, 12, 12, 13, 13, 15; Median = 8.5

Mode = 4

b The mean number of days to submit grades is 8.2, the median is 8.5, and the mode is 4.

4.5 a $\bar{x} = \frac{\sum x_i}{n} = \frac{36+25+45+60+42+19+52+38+36}{9} = \frac{353}{9} = 39.2$

Ordered data: 19, 25, 36, 36, 38, 42, 45, 52, 60; Median = 38

Mode: 36

b The mean amount of time is 39.2 minutes. Half the group took less than 38 minutes.

4.6 $R_g = \sqrt[3]{(1+R_1)(1+R_2)(1+R_3)} - 1 \;=\; \sqrt[3]{(1+.25)(1-.10)(1+.50)} - 1 = .19$

4.7 $R_g = \sqrt[4]{(1+R_1)(1+R_2)(1+R_3)(1+R_4)} - 1 \;=\; \sqrt[4]{(1+.50)(1+.30)(1-.50)(1-.25)} - 1 = -.075$

4.8 a $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{.10+.22+.06-.05+.20}{5} = \dfrac{.53}{5} = .106$

Ordered data: –.05, .06, .10, .20, .22; Median = .10

b $R_g = \sqrt[5]{(1+R_1)(1+R_2)(1+R_3)(1+R_4)(1+R_5)} - 1 \;=\; \sqrt[5]{(1+.10)(1+.22)(1+.06)(1-.05)(1+.20)} - 1 = .102$

c The geometric mean

4.9 a $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{-.15-.20+.15-.08+.50}{5} = \dfrac{.22}{5} = .044$

Ordered data: –.20, –.15, –.08, .15, .50; Median = –.08

b $R_g = \sqrt[5]{(1+R_1)(1+R_2)(1+R_3)(1+R_4)(1+R_5)} - 1 \;=\; \sqrt[5]{(1-.15)(1-.20)(1+.15)(1-.08)(1+.50)} - 1 = .015$

c The geometric mean

4.10 a Year 1 rate of return = $\dfrac{1200-1000}{1000} = .20$

Year 2 rate of return = $\dfrac{1200-1200}{1200} = 0$

Year 3 rate of return = $\dfrac{1500-1200}{1200} = .25$

Year 4 rate of return = $\dfrac{2000-1500}{1500} = .33$

b $\bar{x} = \dfrac{\sum x_i}{n} = \dfrac{.20+0+.25+.33}{4} = \dfrac{.78}{4} = .195$

Ordered data: 0, .20, .25, .33; Median = .225

c $R_g = \sqrt[4]{(1+R_1)(1+R_2)(1+R_3)(1+R_4)} - 1 \;=\; \sqrt[4]{(1+.20)(1+0)(1+.25)(1+.33)} - 1 = .188$

d The geometric mean is because $1000(1.188)^4 = 2000$

4.11 a Year 1 rate of return = $\dfrac{10-12}{12} = -.167$

Year 2 rate of return = $\dfrac{14-10}{10} = .40$

Year 3 rate of return = $\frac{15-14}{14}$ = .071

Year 4 rate of return = $\frac{22-15}{15}$ = .467

Year 5 rate of return = $\frac{30-22}{22}$ = .364

Year 6 rate of return = $\frac{25-30}{30}$ = –.167

b $\bar{x} = \frac{\sum x_i}{n} = \frac{-.167+.40+.071+.467+.364-.167}{6} = \frac{.968}{6}$ = .161

Ordered data: –.167, –.167, .071, .364, .40, .467; Median = .218

c $R_g = \sqrt[6]{(1+R_1)(1+R_2)(1+R_3)(1+R_4)(1+R_5)(1+R_6)} - 1$

$= \sqrt[6]{(1-.167)(1+.40)(1+.071)(1+.467)(1+.364)(1-.167)} - 1$ = .130

d $12(1.130)^6 = 25$

4.12 a $\bar{x}$ = 24,329; median = 24,461

b The mean starting salary is $24,329. Half the sample earned less than $24,461.

4.13a $\bar{x}$ = 30.53; median = 31

b The mean training time is 30.53. Half the sample trained for less than 31 hours.

4.14a $\bar{x}$ = 32.91; median = 32; mode = 32

b The mean speed is 32.91 mph. Half the sample traveled slower than 32 mph and half traveled faster. The mode is 32.

4.15 $\bar{x}$ = 519.20; median = 523.00

b The mean expenditure is $519.20. Half the sample spent less than $523.00

4.16 a $\bar{x}$ = 11.19; median = 11

b The mean number of days is 11.19 and half the sample took less than 11 days and half took more than 11 days to pay.

4.17a $\bar{x}$ = 128.07; median = 136.00

b The mean expenditure is $128.07 and half the sample spent less than $136.00

14.18a $\bar{x}$ = 29.48; median = 30.00

b $\bar{x}$ = 40.18; median = 41.00

b The mean commuting time in New York is larger than that in Los Angeles.

4.19 $\bar{x} = \frac{\sum x_i}{n} = \frac{2+8+9+4+1+7+5+4}{8} = \frac{40}{8} = 5$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(2-5)^2 + (8-5)^2 + ... + (4-5)^2}{8-1} = \frac{56}{7} = 8$

4.20 $\bar{x} = \frac{\sum x_i}{n} = \frac{8+9+3+6+5+6+5+6}{8} = \frac{48}{8} = 6$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(8-6)^2 + (9-6)^2 + ... + (6-5)^2}{8-1} = \frac{24}{7} = 3.43$

4.21 $\bar{x} = \frac{\sum x_i}{n} = \frac{9+15+11+31+23+13+15+17+21}{9} = \frac{155}{9} = 17.22$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(9-17.22)^2 + (15-17.22)^2 + ... + (21-17.22)^2}{9-1} = \frac{371.56}{8} = 46.45$

$s = \sqrt{s^2} = \sqrt{46.45} = 6.82$

4.22 $\bar{x} = \frac{\sum x_i}{n} = \frac{7+(-5)+(-3)+8+4+(-4)+1+(-5)+9+3}{10} = \frac{15}{10} = 1.5$

$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(7-1.5)^2 + ((-5)-1.5)^2 + ... + (3-1.5)^2}{10-1} = \frac{272.5}{9} = 30.28$

$s = \sqrt{s^2} = \sqrt{30.28} = 5.50$

4.23 The data in (b) appear to be most similar to one another.

4.24 a: $s^2 = 51.5$

b: $s^2 = 6.5$

c: $s^2 = 174.5$

4.25 Variance cannot be negative because it is the sum of *squared* differences.

4.26 5, 5, 5, 5, 5

4.27 a about 68%

b about 95%

c About 99.7%

4.28 a From the empirical rule we know that approximately 68% of the observations fall between 46 and 54. Thus 16% are less than 46 (the other 16% are above 54).

b Approximately 95% of the observations are between 42 and 58. Thus, only 2.5% are above 58 and all the rest, 97.5% are below 58.

c See (a) above; 16% are above 54.

4.29 a at least 75%

b at least 88.9%

4.30 a Nothing

b At least 75% lie between 60 and 180.

c At least 88.9% lie between 30 and 210.

4.31 $s^2 = 40.73 \text{ mph}^2$, and s = 6.38 mph; at least 75% of the speeds lie within 12.76 mph of the mean; at least 88.9% of the speeds lie within 19.14 mph of the mean

4.32 Range = 25.85, $s^2 = 29.46$, and s = 5.43; there is considerable variation between prices; at least 75% of the prices lie within 10.86 of the mean; at least 88.9% of the prices lie within 16.29 of the mean.

4.33 a

Punter	Variance	Standard deviation
1	40.22	6.34
2	14.81	3.85
3	3.63	1.91

b Punter 3 is the most consistent.

4.34 $\bar{x} = 175.73, s^2 = 3{,}851.82$; s = 62.1; At least 75% of the withdrawals lie within \$124.20 of the mean; at least 88.9% of the withdrawals lie within \$186.30 of the mean..

4.35 $s^2 = .0858 \text{ cm}^2$, and s = .2929cm; at least 75% of the lengths lie within .5858 of the mean; at least 88.9% of the rods will lie within .8787 cm of the mean.

4.36a s = 15.01

b In approximately 68% of the days the number of arrivals falls within 30.02 of the mean; on approximately 95% of the hours the number of arrivals falls within 45.03 of the mean

4.37 First quartile: $L_{25} = (13+1)\dfrac{25}{100}$ = (14)(.25) = 3.5; the first quartile is 13.05.

Second quartile: $L_{50} = (13+1)\dfrac{50}{100}$ = (14)(.5) = 7; the second quartile is 14.7.

Third quartile: $L_{75} = (13+1)\dfrac{75}{100}$ = (14)(.75) = 10.5; the third quartile is 15.6.

4.38 Third decile: $L_{30} = (15+1)\dfrac{30}{100}$ = (16)(.30) = 4.8; the third decile is 5 + .8(7 – 5) = 6.6.

Sixth decile: $L_{60} = (15+1)\dfrac{60}{100}$ = (16)(.60) = 9.6; the sixth decile is 17 + .6(18 – 17) = 17.6.

4.39 First quartile: $L_{25} = (15+1)\dfrac{25}{100}$ = (16)(.25) = 4; the fourth number is 3.

Second quartile: $L_{50} = (15+1)\dfrac{50}{100}$ = (16)(.5) = 8; the eighth number is 5.

Third quartile: $L_{75} = (15+1)\dfrac{75}{100}$ = (16)(.75) = 12; the twelfth number is 7.

4.40 30th percentile: $L_{30} = (10+1)\dfrac{30}{100}$ = (11)(.30) = 3.3; the 30th percentile is 22.3.

80th percentile: $L_{80} = (10+1)\dfrac{80}{100}$ = (11)(.80) = 8.8; the 80th percentile 30.8.

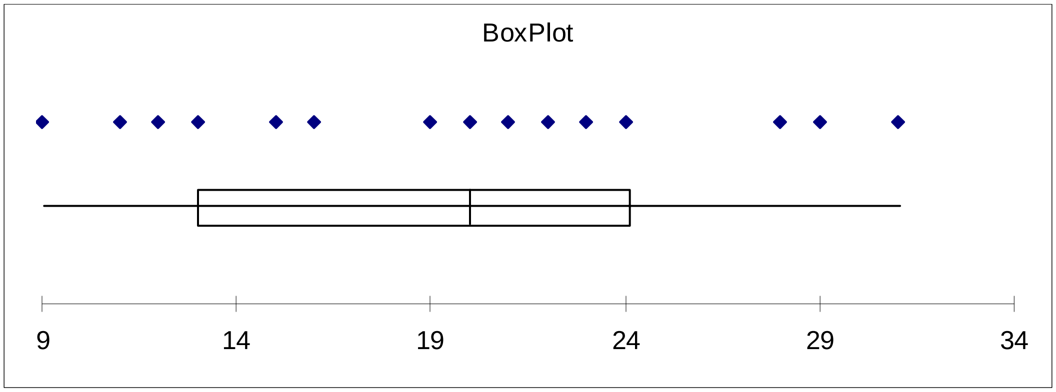
4.41 20th percentile: $L_{20} = (10+1)\dfrac{20}{100}$ = (11)(.20) = 2.2; the 20th percentile is 43 + .2(51–43) = 44.6.

40th percentile: $L_{40} = (10+1)\dfrac{40}{100}$ = (11)(.40) = 4.4; the 40th percentile is 52 +.4(60–52) = 55.2.

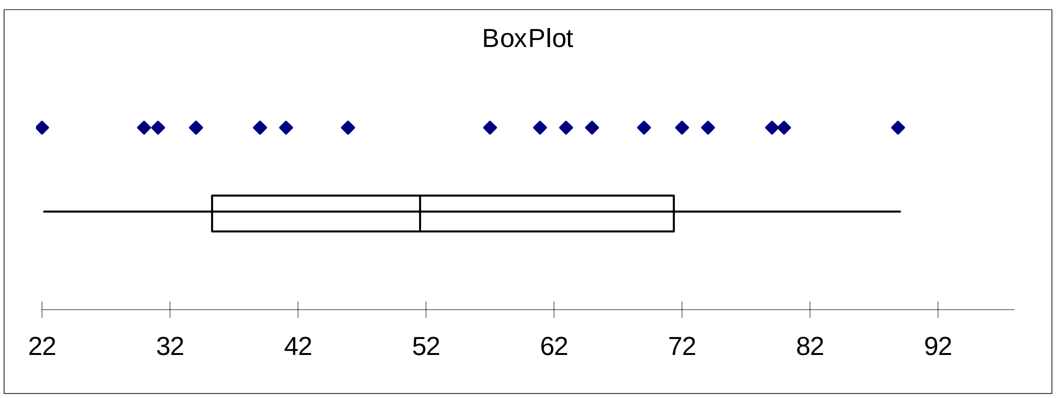
4.42 Interquartile range = 15.6 –13.05 = 2.55

4.43 Interquartile range = 7 – 3 = 4

4.44 First quartile = 5.75, third quartile = 15; interquartile range = 15 – 5.75 = 9.25



4.45

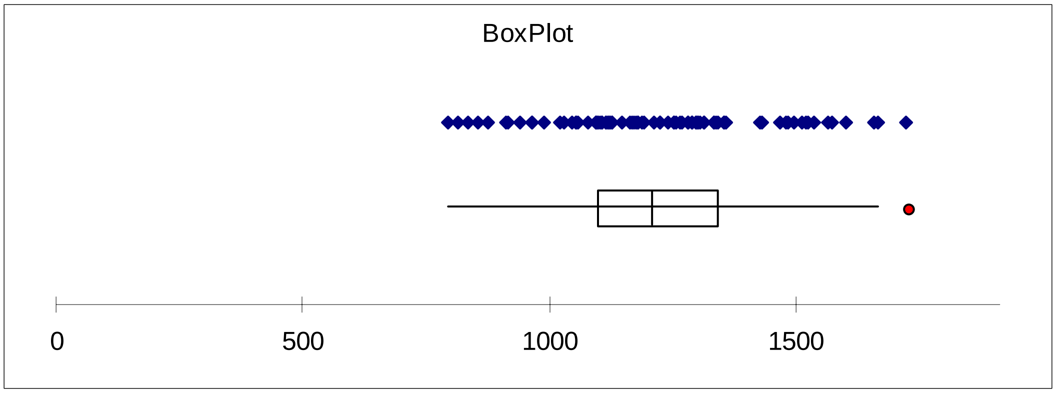


4.46

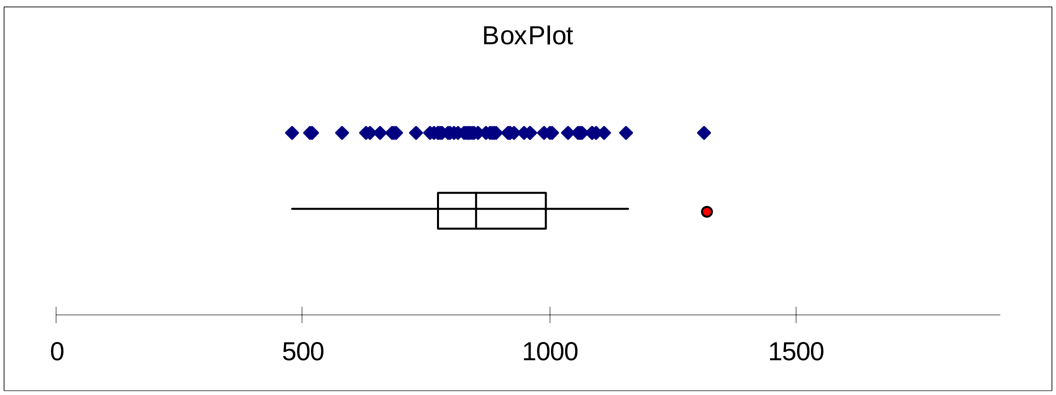
4.47a First quartile = 2, second quartile = 4, and third quartile = 8.

b Most executives spend little time reading resumes. Keep it short.

4.48 Dogs



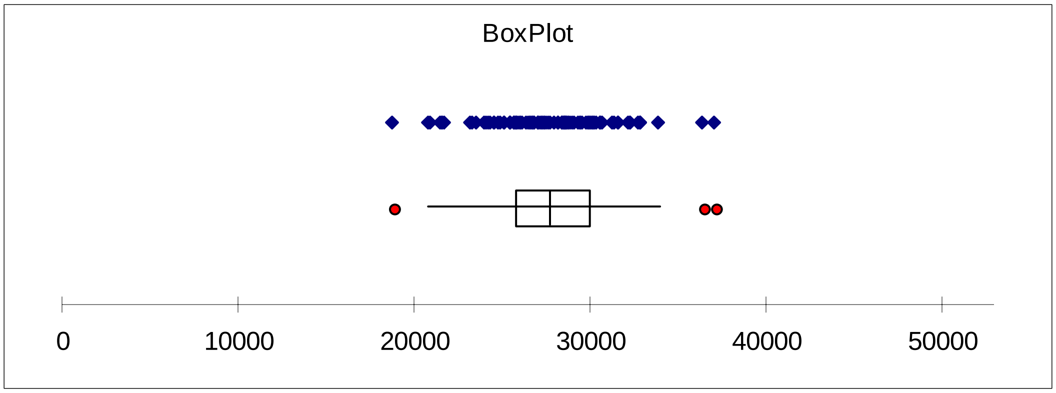
Cats

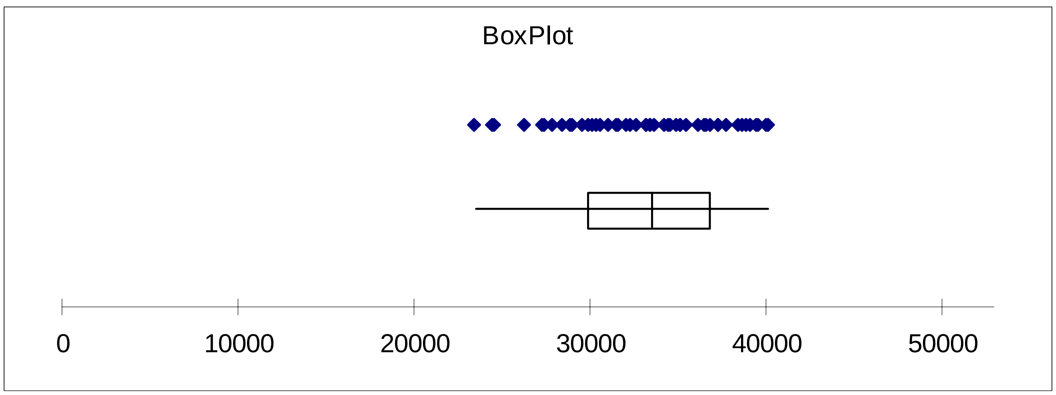


Dogs cost more money than cats. Both sets of expenses are positively skewed.

4.49 First quartile = 50, second quartile = 125, and third quartile = 260. The amounts are positively skewed.

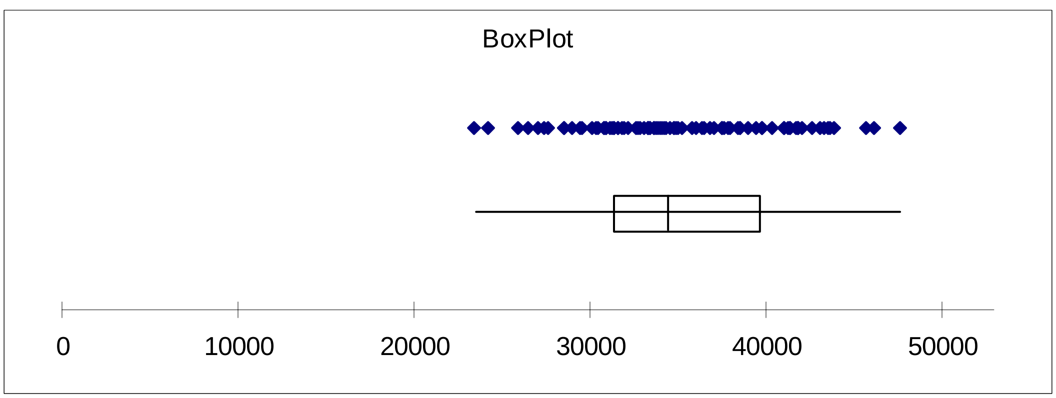
4.50 BA



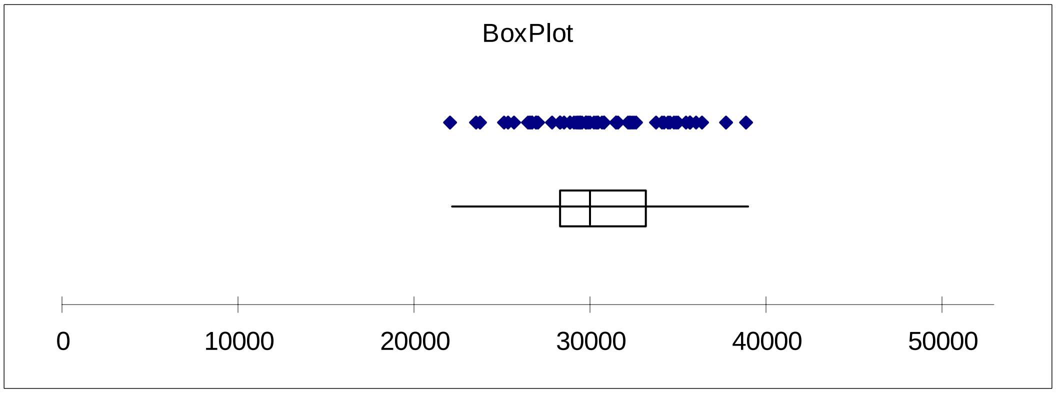


BSc

BBA

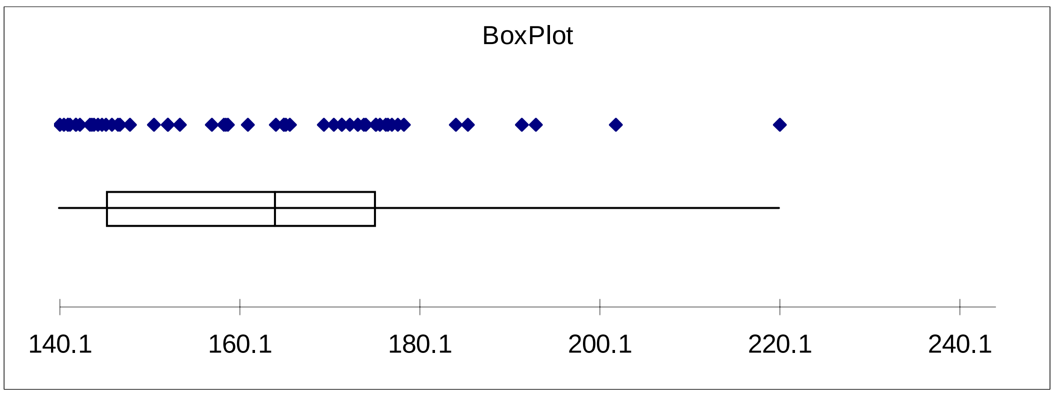


Other



The starting salaries of BA and other are the lowest and least variable. Starting salaries for BBA and BSc are higher.

4.51 a



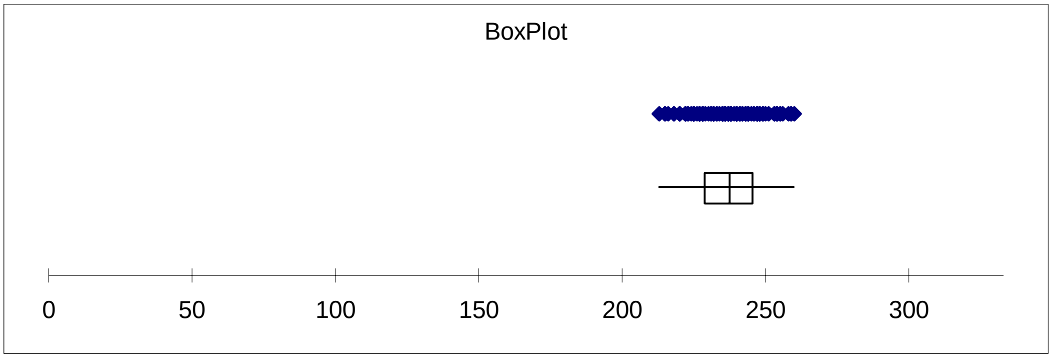
b The quartiles are 145.11, 164.17, and 175.18

c There are no outliers.

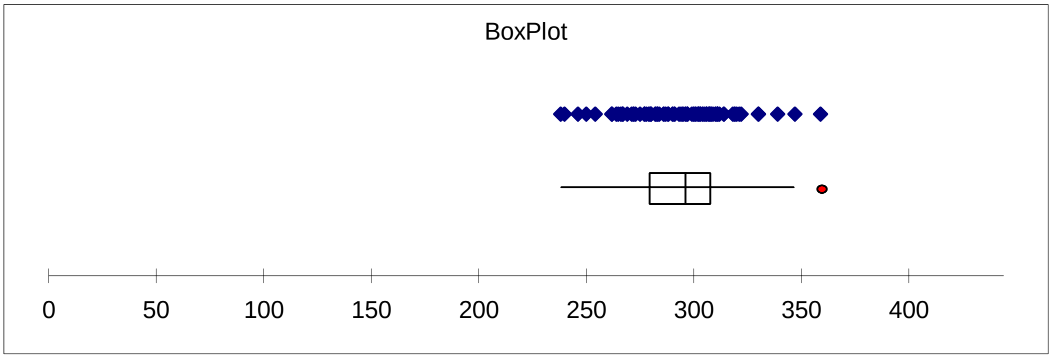
d The data are positively skewed. One-quarter of the times are below 145.11 and one-quarter are above 175.18.

4.52a

Private course:



Public course:



b The amount of time taken to complete rounds on the public course are larger and more variable than those played on private courses.

4.53 a The quartiles are 26, 28.5, and 32

b the times are positively skewed.

4.54 The quartiles are 697.19, 804.90, and 909.38. One-quarter of mortgage payments are less than $607.19 and one quarter exceed $909.38.

4.55 There is a negative linear relationship. The strength is unknown.

4.56 $r = \frac{s_{xy}}{s_x s_y} = \frac{-14}{(3)(7)} = -.667$

There is a moderately strong negative linear relationship.

4.57

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
27.0	12.02	729.00	144.48	324.54
28.5	12.04	812.25	144.96	343.14

	30.8	12.32	948.64	151.78	379.46
	31.3	12.27	979.69	150.55	384.05
	31.9	12.49	1,017.61	156.00	398.43
	34.5	12.70	1,190.25	161.29	438.15
	34.0	12.80	1,156.00	163.84	435.20
	34.7	13.00	1,204.09	169.00	451.10
	37.0	13.00	1,369.00	169.00	481.00
	41.1	13.17	1,689.21	173.45	541.29
	41.0	13.19	1681.00	173.98	540.79
	38.8	13.22	1,505.49	174.77	512.94
	39.3	13.27	1,544.49	176.09	521.51
Total	449.9	165.49	15,826.67	2,109.19	5,751.59

$$\sum_{i=1}^{n} x_i = 449.9 \quad \sum_{i=1}^{n} y_i = 165.49 \quad \sum_{i=1}^{n} x_i^2 = 15,826.67 \quad \sum_{i=1}^{n} y_i^2 = 2,109.19 \quad \sum_{i=1}^{n} x_i y_i = 5,751.59$$

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{13-1}\left[5,751.59 - \frac{(449.9)(165.49)}{13}\right] = 2.030$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{13-1}\left[15,826.67 - \frac{(449.9)^2}{13}\right] = 21.39$$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] = \frac{1}{13-1}\left[2,109.19 - \frac{(165.49)^2}{13}\right] = .2085$$

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{2.030}{21.39} = .0949$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{449.9}{13} = 34.61$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{165.49}{13} = 12.73$$

$$b_0 = \bar{y} - b_1\bar{x} = 12.73 - (.0949)(34.61) = 9.446$$

The least squares line is

$$\hat{y} = 9.446 + .0949x$$

The appropriate compensation is 9.49 cents per degree API.

4.58

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	60	260	3,600	67,600	15,600
	45	266	2,025	70,756	11,970
	55	257	3,025	66,049	14,135
	35	199	1,225	39,601	6,965
	55	271	3,025	73,441	14,905
	60	322	3,600	103,684	19,320
	35	165	1,225	27,225	5,775
	60	388	3,600	150,544	23,280
	50	233	2,500	54,289	11,650
	45	243	2,025	59,049	10,935
	40	242	1,600	58,564	9,680
	45	233	2,025	54,289	10,485
Total	585	3,079	29,475	825,091	154,700

$$\sum_{i=1}^{n} x_i = 585 \qquad \sum_{i=1}^{n} y_i = 3{,}079 \qquad \sum_{i=1}^{n} x_i^2 = 29{,}475 \qquad \sum_{i=1}^{n} y_i^2 = 825{,}091 \qquad \sum_{i=1}^{n} x_i y_i = 154{,}700$$

a $$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{12-1}\left[154{,}700 - \frac{(585)(3{,}079)}{12}\right] = 418.1$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{12-1}\left[29{,}475 - \frac{(585)^2}{12}\right] = 86.93$$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] = \frac{1}{12-1}\left[825{,}091 - \frac{(3{,}079)^2}{12}\right] = 3{,}188$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{418.1}{\sqrt{(86.93)(3{,}188)}} = .7942$$

b $$b_1 = \frac{s_{xy}}{s_x^2} = \frac{418.1}{86.93} = 4.809$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{585}{12} = 48.75$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{3{,}079}{12} = 256.6$$

$$b_0 = \bar{y} - b_1\bar{x} = 256.6 - (4.809)(48.75) = 22.16$$

The least squares line is

$\hat{y} = 22.16 + 4.809x$

For each additional minute of exercise the metabolic rate increases on average by 4.809.

4.59

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	40	77	1,600	5,929	3,080
	42	63	1,764	3,969	2,646
	37	79	1,369	6,241	2,923
	47	86	2,209	7,396	4,041
	25	51	625	2,601	1,276
	44	78	1,936	6,084	3,432
	41	83	1,681	6,889	3,403
	48	90	2,304	8,100	4,320
	35	65	1,225	4,225	2,275
	28	47	784	2,209	1,316
Total	387	719	15,497	53,643	28,712

$\sum_{i=1}^{n} x_i = 387$ $\quad \sum_{i=1}^{n} y_i = 719$ $\quad \sum_{i=1}^{n} x_i^2 = 15{,}497$ $\quad \sum_{i=1}^{n} y_i^2 = 53{,}643$ $\quad \sum_{i=1}^{n} x_i y_i = 28{,}712$

a $s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{10-1}\left[28{,}712 - \frac{(387)(719)}{10}\right] = 98.52$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{10-1}\left[15{,}497 - \frac{(387)^2}{10}\right] = 57.79$$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] = \frac{1}{10-1}\left[53{,}643 - \frac{(719)^2}{10}\right] = 216.32$$

b $r = \frac{s_{xy}}{s_x s_y} = \frac{98.52}{\sqrt{(57.79)(216.32)}} = .8811$

c $b_1 = \frac{s_{xy}}{s_x^2} = \frac{98.52}{57.79} = 1.705$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{387}{10} = 38.7$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{719}{10} = 71.9$$

$$b_0 = \bar{y} - b_1\bar{x} = 71.9 - (1.705)(38.7) = 5.917$$

The least squares line is

$$\hat{y} = 5.917 + 1.705x$$

There is a strong positive linear relationship between marks and study time. For each additional hour of study time marks increased on average by 1.705.

4.60 a

	A	B	C
1		*Age*	*Expense*
2	Age	228.1	
3	Expense	51.49	179.7

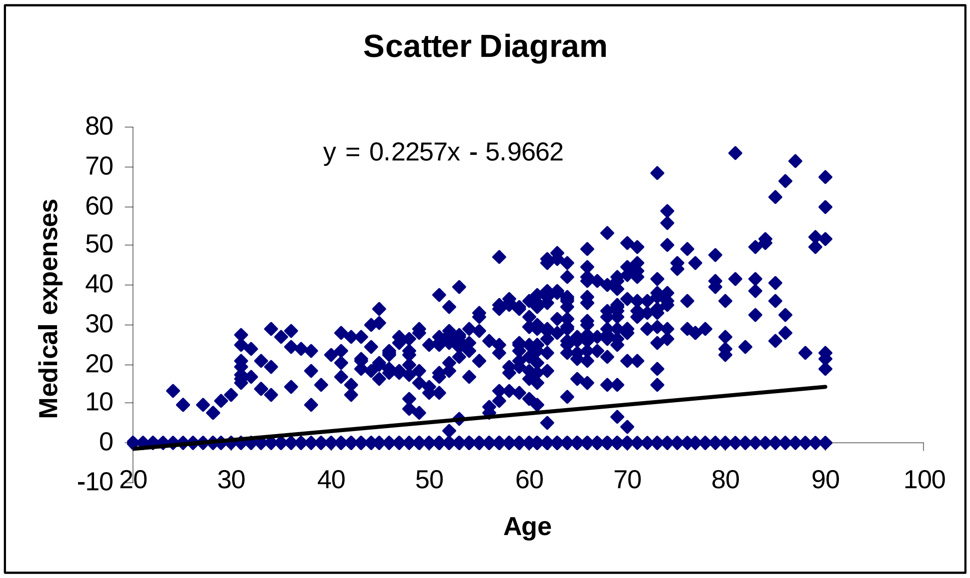
Covariance: $s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{(1,348)(51.49)}{1,347} = 51.53$

	A	B	C
1		*Age*	*Expense*
2	Age	1	
3	Expense	0.2543	1

Coefficient of correlation = .2543

b There is a weak linear relationship between age and medical expenses.

c

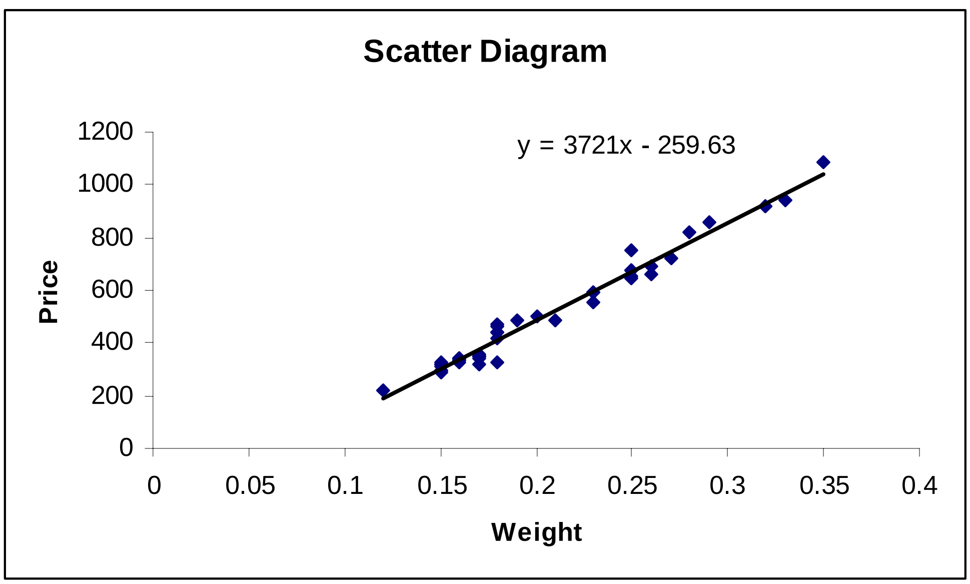


The least squares line is $\hat{y} = -5.966 + .2257x$

d For each additional year of age mean medical expenses increase on average by $.2257 or 23 cents.

e Charge 25 cents per day per year of age.

4.61



There is very strong linear relationship between price and weight. For each additional carat the price of the diamond increases on average by $3,721 Singapore.

4.62 a

	A	B	C
1		*Rate*	*Houses*
2	Rate	10.21	
3	Houses	-13.96	230.5

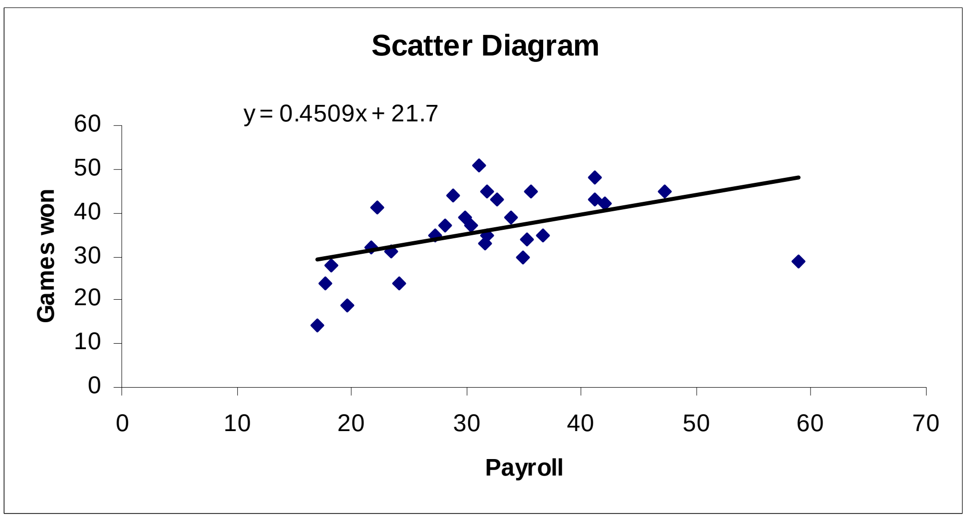
Covariance: $s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{492(-13.96)}{491} = -13.99$

	A	B	C
1		*Rate*	*Houses*
2	Rate	1	
3	Houses	-0.2878	1

Coefficient of correlation = –.2878

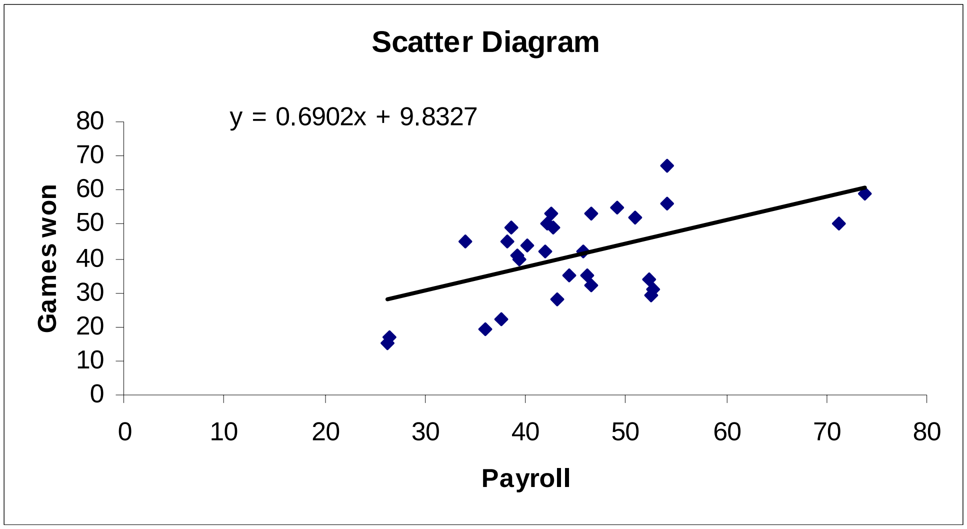
There is a weak negative linear relationship between the number of houses built and the prime bank rate

4.63



$\hat{y}$ = 21.7 + .4509x; Cost to win one more game = 1million/..4509 = \$2,217,787

4.64

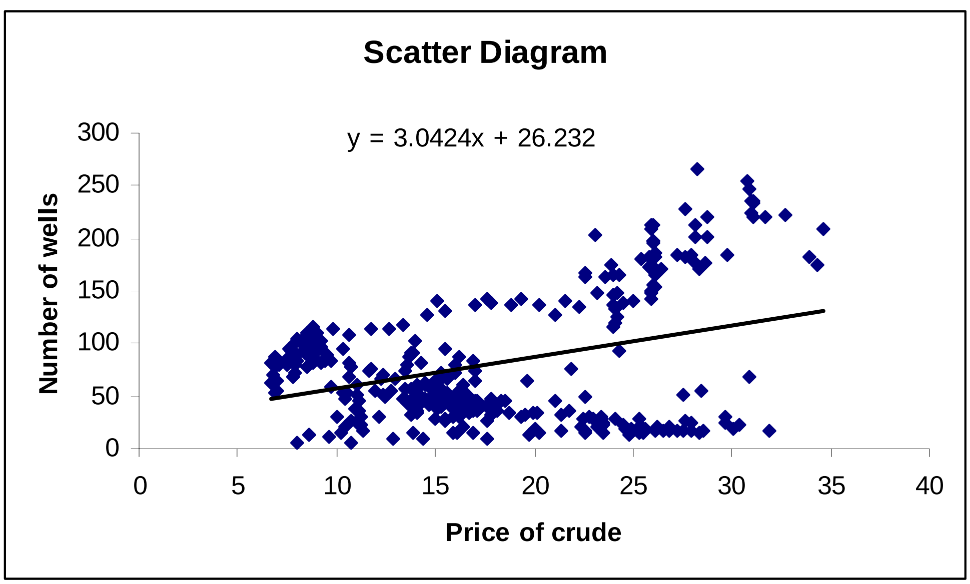


$\hat{y} = 9.8327 + .6902x$; Cost to win one more game = 1million/.6902 = $1,448,855

4.65

	A	B	C
1		*Crude*	*Wells*
2	Crude	1	
3	Wells	0.3715	1

The coefficient of correlation is .3715. There is a weak positive linear relationship between the price of a barrel of crude oil and the number of exploratory wells drilled.



For each additional dollar increase in the price of crude oil the number of exploratory wells drilled increases on average by 3.0424.

4.66

	A	B	C
1		*Index*	*Rate*
2	Index	1	
3	Rate	0.0669	1

There is a very weak positive linear relationship between the unemployment rate and the index. We would have expected a negative linear relationship.

4.67a

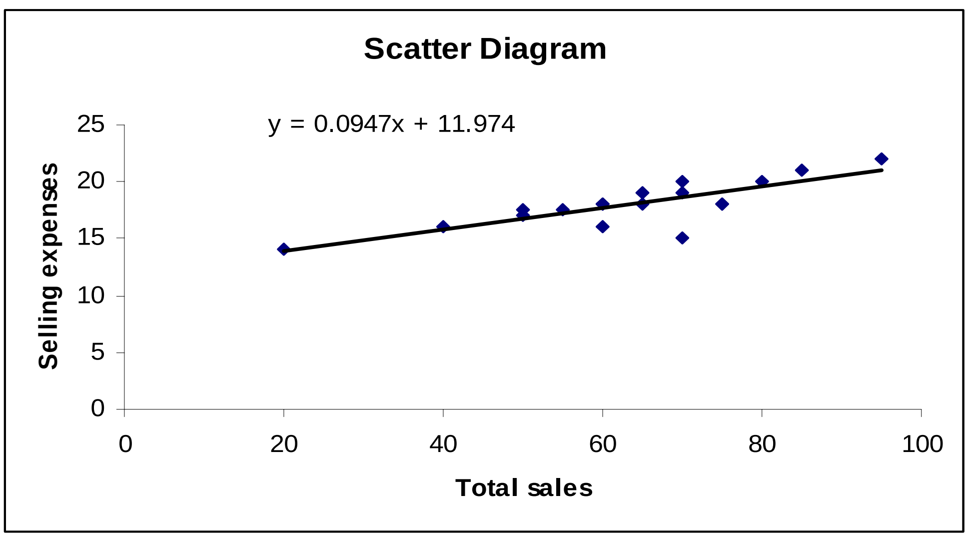
	Selling Expenses	*Total Sales*
Selling Expenses	3.86	
Total Sales	26.53	280.02

Covariance: $s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{18(26.53)}{17} = 28.09$

	Selling Expenses	*Total Sales*
Selling Expenses	1	
Total Sales	0.8068	1

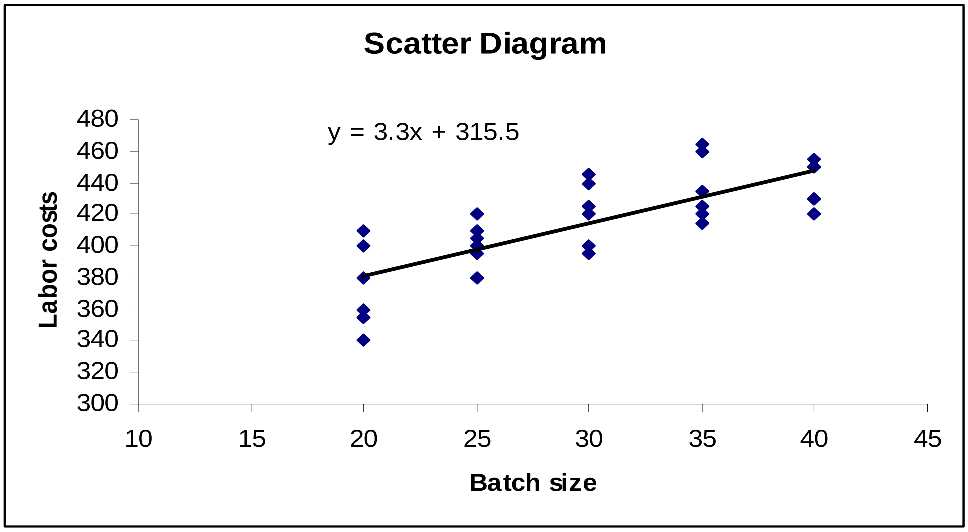
Coefficient of correlation: r = .8068

b



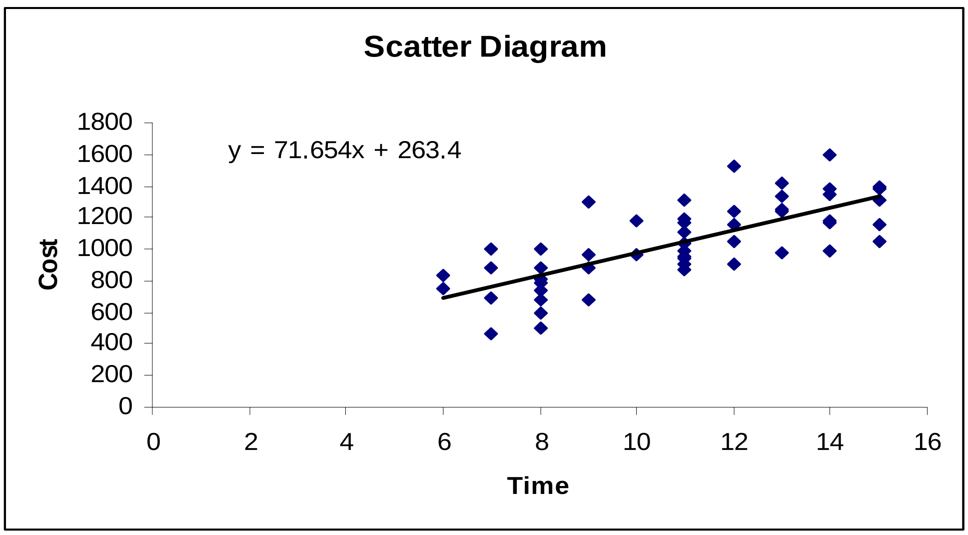
$\hat{y}$ = 11.974 + .0947x; Fixed costs = $11,974, variable costs = $0.0947

4.68



$\hat{y} = 315.5 + 3.3x$; Fixed costs = \$315.50, variable costs = \$3.30

4.69



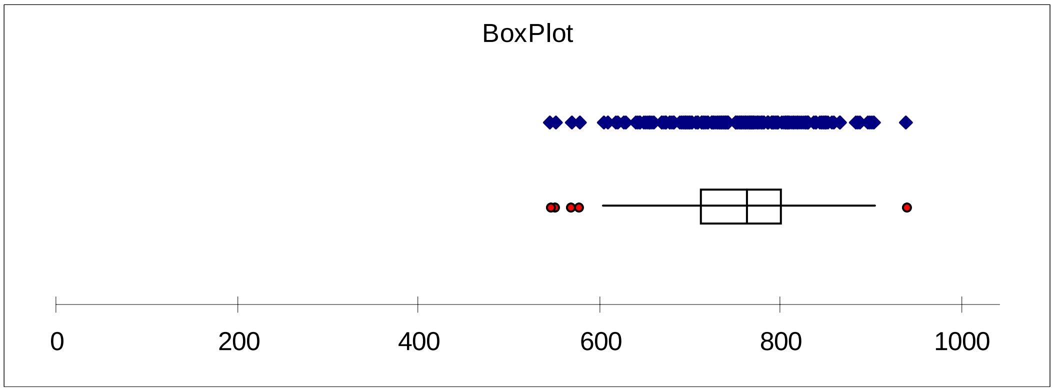
$\hat{y} = 263.4 + 71.65x$; Fixed costs = \$263.40, variable costs = \$71.65

4.70 a

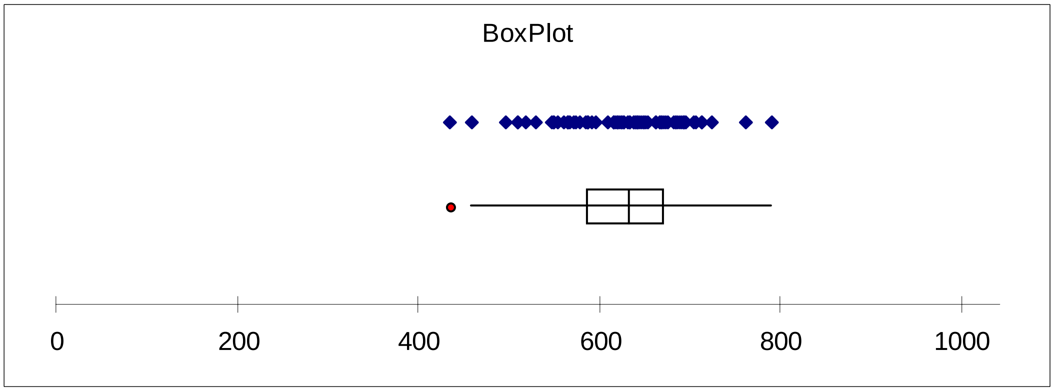
	A	B	C	D	E
1	*Repaid*			*Defaulted*	
2					
3	Mean	751.5		Mean	545.7
4	Standard Error	3.75		Standard Error	6.07
5	Median	753		Median	549.5
6	Mode	753		Mode	552
7	Standard Deviation	49.16		Standard Deviation	54.25
8	Sample Variance	2416		Sample Variance	2943
9	Kurtosis	0.177		Kurtosis	-0.673
10	Skewness	-0.170		Skewness	-0.094
11	Range	253		Range	237
12	Minimum	625		Minimum	419
13	Maximum	878		Maximum	656
14	Sum	129265		Sum	43658
15	Count	172		Count	80

b We can see that among those who repaid the mean score is larger than that of those who did not and the standard deviation is smaller. This information is similar but more precise than that obtained in Exercise 2.54.

4.71 Repaid loan:



Defaulted on loan:



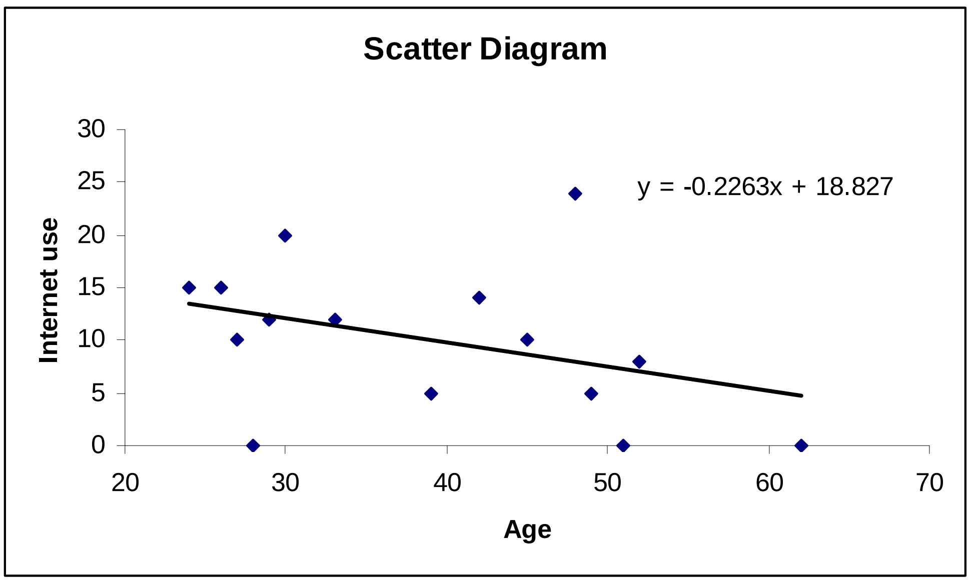
The box plots make it a little easier to see the overlap between the two sets of data (indicating that the scorecard is not very good).

4.72

	A	B	C
1		*Calculus*	*Statistics*
2	Calculus	1	
3	Statistics	0.6784	1

The coefficient of correlation provides a more precise indication of the strength of the linear relationship. However, we cannot define exactly what this value tells us.

4.73



The slope coefficient tells us that for each additional year of age Internet use decreases on average by .2263 hour. The scatter diagram alone is not that precise.

4.74

	A	B	C
1		*Price:Gasoline*	*Price: Crude*
2	Price:Gasoline	0.05782	
3	Price: Crude	1.374	46.51

Covariance: $s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{336(1.374)}{335} = 1.378$

	A	B	C
1		*Price:Gasoline*	*Price: Crude*
2	Price:Gasoline	1	
3	Price: Crude	0.8376	1

The coefficient of correlation .8376 provides a somewhat more precise measure of the strength of the linear relationship.

4.75

	A	B	C
1		*Temperature*	*Tickets*
2	Temperature	1	
3	Tickets	0.6570	1

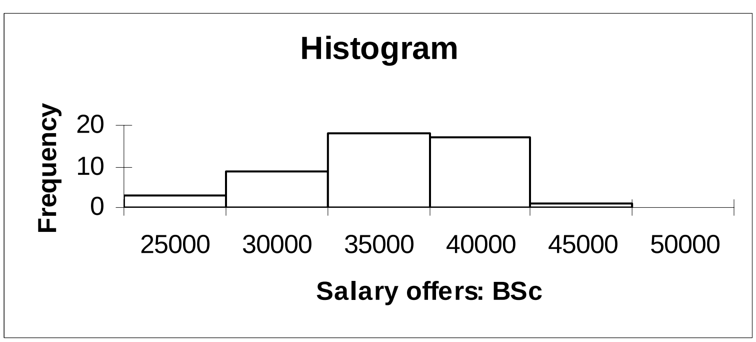
The coefficient of correlation .6570 provides a more precise measure of the strength of the linear relationship. However, because we cannot exactly interpret the coefficient of correlation the information acquired is limited.

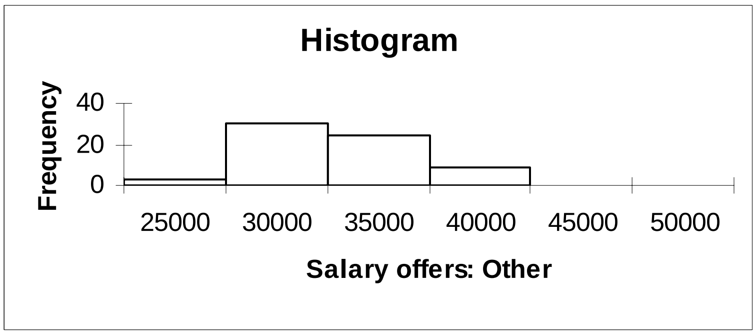
4.76

	A	B	C
1		*Height*	*Income*
2	Height	1	
3	Income	0.2248	1

The coefficient of correlation .2248 is more precise than the scatter diagram. However, the information is limited.

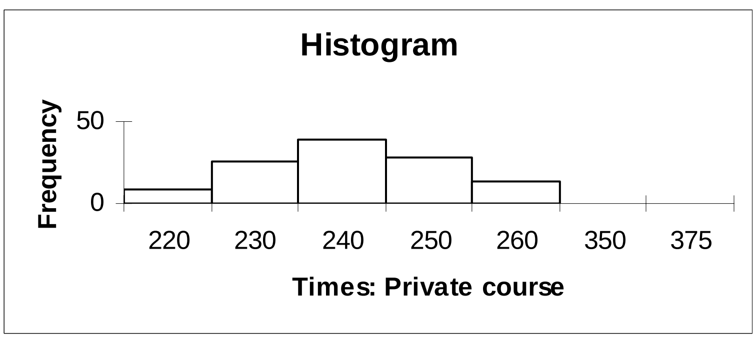
4.77

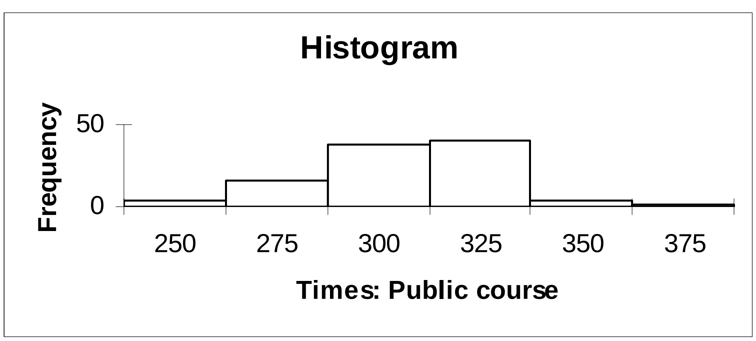




Using the same class limits the histograms provide more detail than do the box plots.

4.78





The information obtained here is more detailed than the information provided by the box plots.

4.79

	A	B
1	*Bone Loss*	
2		
3	Mean	35.01
4	Standard Error	0.69
5	Median	36
6	Mode	38
7	Standard Deviation	7.68
8	Sample Variance	59.04
9	Kurtosis	0.08
10	Skewness	-0.19
11	Range	38
12	Minimum	15
13	Maximum	53
14	Sum	4376
15	Count	125

a $\bar{x} = 35.0$, median = 36

b s = 7.68

c Half of the bone density losses lie below 36. At least 75% of the numbers lie between 19.64 and 50.36, at least 88.9% of the numbers lie between 11.96 and 58.04.

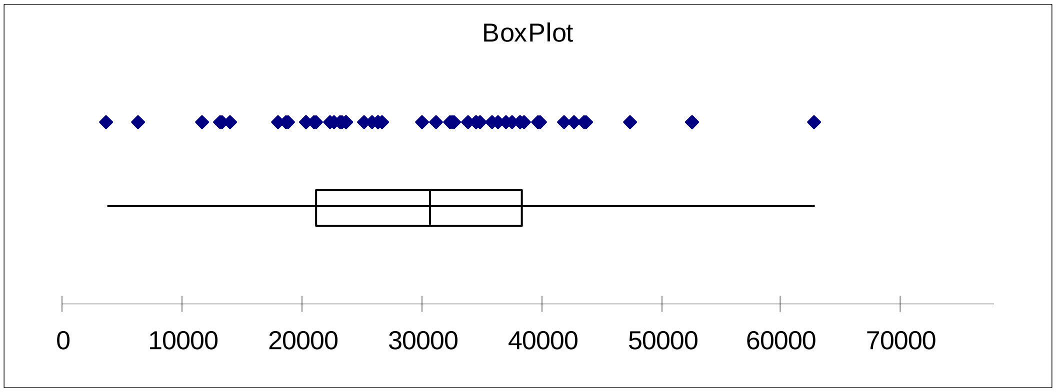
4.80

	A	B
1	*Coffees*	
2		
3	Mean	29,913
4	Standard Error	1,722
5	Median	30,660
6	Mode	#N/A
7	Standard Deviation	12,174
8	Sample Variance	148,213,791
9	Kurtosis	0.12
10	Skewness	0.22
11	Range	59,082
12	Minimum	3,647
13	Maximum	62,729
14	Sum	1,495,639
15	Count	50

a $\bar{x}$ = 29,913, median = 30,660

b s^2 = 148,213,791; s = 12,174

c



d The number of coffees sold varies considerably.

4.81

	A	B	C
1		*Bone Loss*	*Age*
2	Bone Loss	1	
3	Age	0.5742	1

r = .5742; there is a moderately strong linear relations ship between age and bone density loss.

4.82a

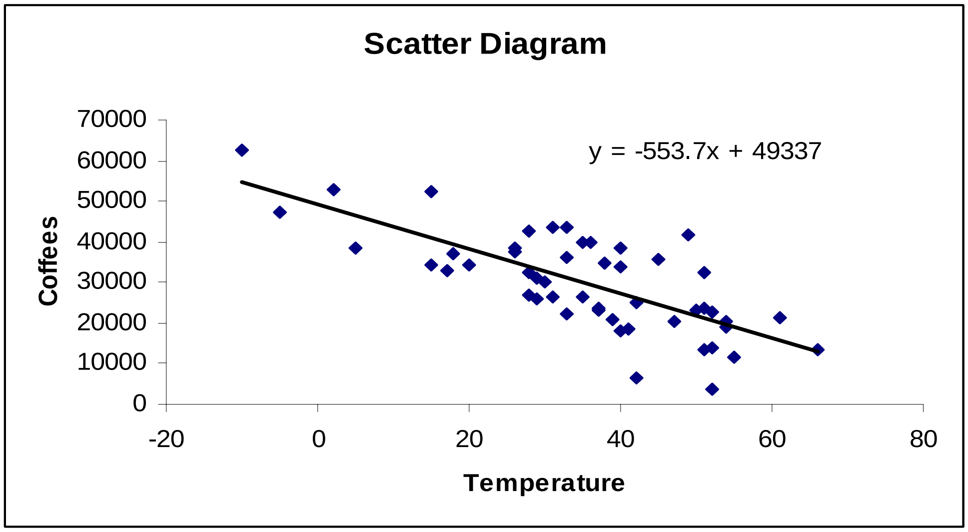
	A	B	C
1		*Coffees*	*Temperature*
2	Coffees	145,249,515	
3	Temperature	-144,003	260

$$s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{50(-144{,}003)}{49} = -146{,}942$$

	A	B	C
1		*Coffees*	*Temperature*
2	Coffees	1	
3	Temperature	-0.7409	1

r = –.7409

b $\hat{y} = 49{,}337 - 553.7x$



c There is a moderately strong negative linear relationship. For each additional degree of temperature the number of coffees sold decreases on average by 554 cups.

d In this exercise we determined that the number of cups of coffee sold is related to temperature, which may explain the variability in coffee sales. .

4.83a mean, median, and standard deviation

b

	A	B
1	*Total Score*	
2		
3	Mean	93.90
4	Standard Error	0.77
5	Median	94
6	Mode	94
7	Standard Deviation	7.72
8	Sample Variance	59.55
9	Kurtosis	0.20
10	Skewness	0.24
11	Range	39
12	Minimum	76
13	Maximum	115
14	Sum	9390
15	Count	100

$\bar{x}$ = 93.90, s = 7.72

c We hope Chris is better at statistics than he is golf.

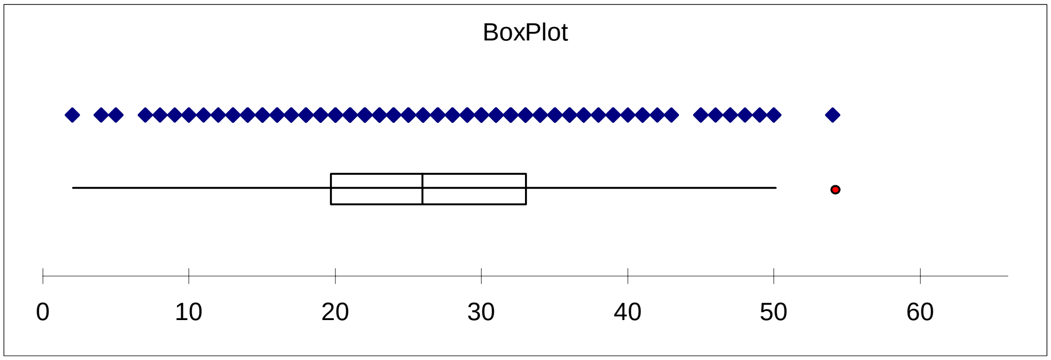
4.84

	A	B
1	*Internet*	
2		
3	Mean	26.32
4	Standard Error	0.60
5	Median	26
6	Mode	21
7	Standard Deviation	9.41
8	Sample Variance	88.57
9	Kurtosis	-0.071
10	Skewness	0.15
11	Range	52
12	Minimum	2
13	Maximum	54
14	Sum	6579
15	Count	250

a $\bar{x}$ = 26.32 and median = 26

b s^2 = 88.57, s = 9.41

c



d The times are positively skewed. Half the times are above 26 hours.

4.85

	A	B	C
1		*Total Score*	*Putts*
2	Total Score	1	
3	Putts	0.8956	1

There is a strong positive linear relationship between total score and number of putts.

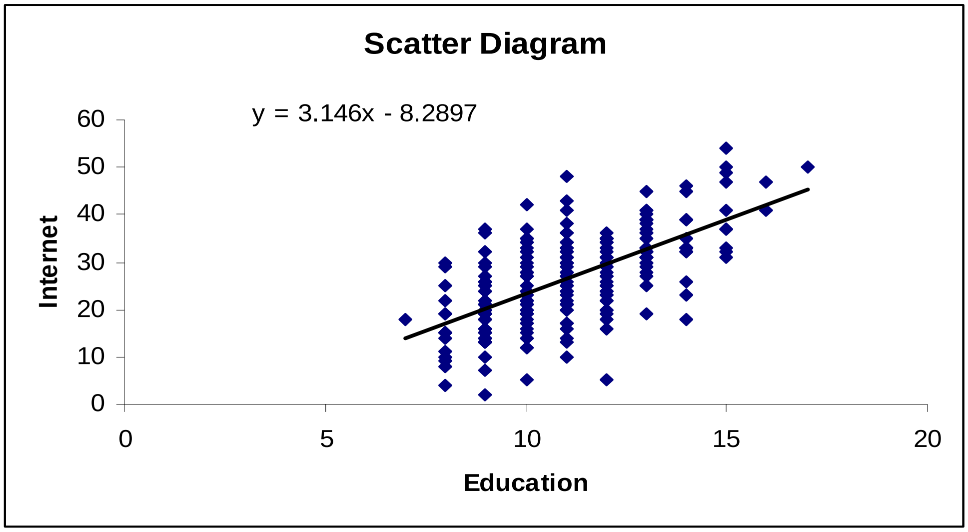
4.86a

	A	B	C
1		*Internet*	*Education*
2	Internet	88.22	
3	Education	11.55	3.67

$$s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{250(11.55)}{249} = 11.60$$

	A	B	C
1		*Internet*	*Education*
2	Internet	1	
3	Education	0.6418	1

b



c There is a moderately strong positive linear relationship between Internet use and education. For each additional year of education Internet use increases on average by 3.15 hours per month.

d This exercise helps explain the variation in Internet use.

4.87

	A	B
1	*Corn*	
2		
3	Mean	150.77
4	Standard Error	1.61
5	Median	150.50
6	Mode	154
7	Standard Deviation	19.76
8	Sample Variance	390.38
9	Kurtosis	-0.13
10	Skewness	0.08
11	Range	107
12	Minimum	101
13	Maximum	208
14	Sum	22,616
15	Count	150

$\bar{x}$ = 150.77, median = 150.50, and s = 19.76. The average crop yield is 150.77 and there is a great deal of variation from one plot to another.

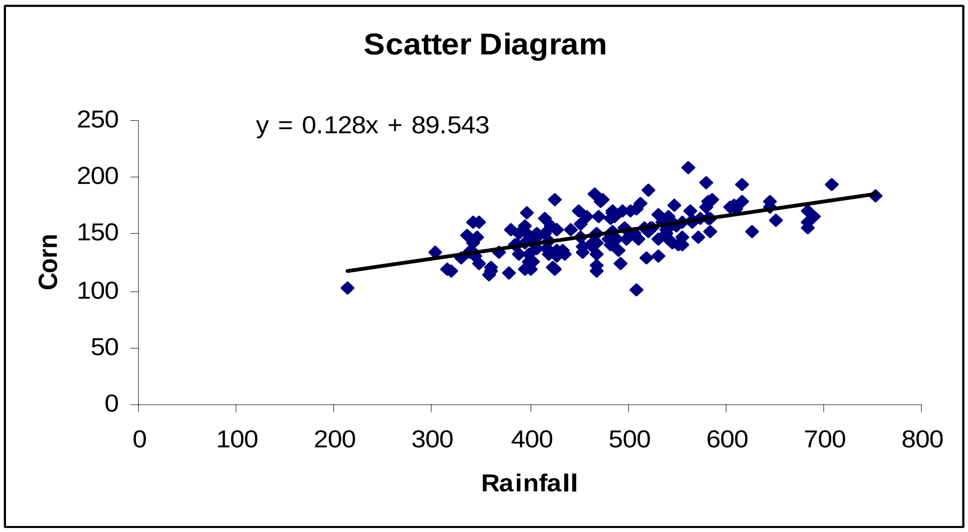
4.88a

	Corn	*Rainfall*
Corn	387.78	
Rainfall	1,118.57	8,738.66

$$s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{150(1{,}118.57)}{149} = 1{,}126.07$$

	A	B	C
1		*Corn*	*Rainfall*
2	Corn	1	
3	Rainfall	0.6076	1

b



$\hat{y}$ = 89.54 + .128 Rainfall

c There is a moderately strong positive linear relationship between yield and rainfall. For each additional inch of rainfall yield increases on average by .13 bushels.

d In this exercise we determined that yield is related to rainfall, which helps explain the variability in corn yield.

4.89a

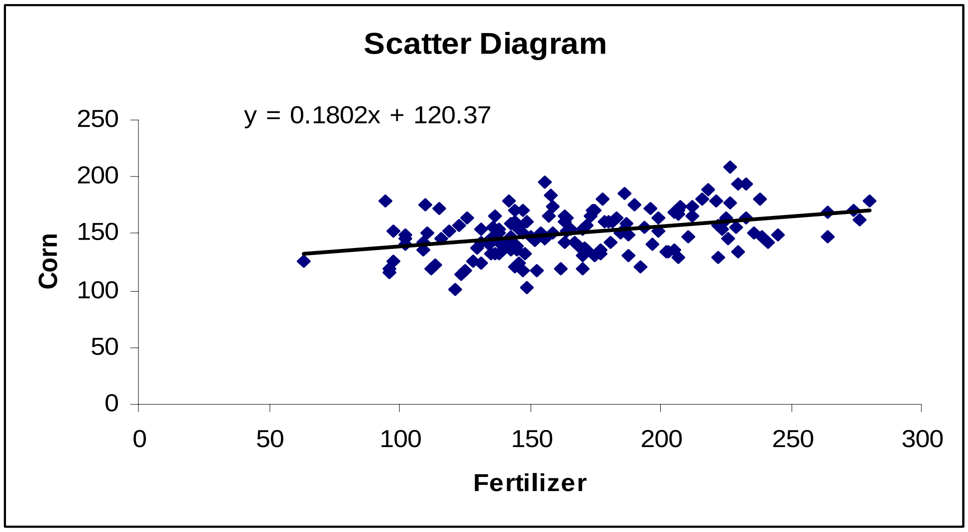
	A	B	C
1		*Corn*	*Fertilizer*
2	Corn	387.78	
3	Fertilizer	333.39	1849.79

$$s_{xy} = \frac{n(\text{Excel covariance})}{n-1} = \frac{150(333.39)}{149} = 335.62$$

	A	B	C
1		*Corn*	*Fertilizer*
2	Corn	1	
3	Fertilizer	0.3936	1

r = .3936

b



$\hat{y}$ = 120.37 + .180 Fertilizer

c There is a relatively weak positive linear relationship.

d Some of the variation in crop yields is explained by the variation in fertilizer.

4.90a, b, and c

	A	B
1	*Vocabulary*	
2		
3	Mean	226.49
4	Standard Error	3.11
5	Median	223
6	Mode	215
7	Standard Deviation	43.99
8	Sample Variance	1934.8
9	Kurtosis	-0.27
10	Skewness	0.089
11	Range	240
12	Minimum	114
13	Maximum	354
14	Sum	45,297
15	Count	200
16	Largest(50)	259
17	Smallest(50)	193

a The mean vocabulary is 226.49 words. The median is 223, which tells us that half the children had vocabularies greater than 223 words and half had less.

b The variance is 1934.8 and the standard deviation is 43.99. Both indicate a great deal of variation between children.

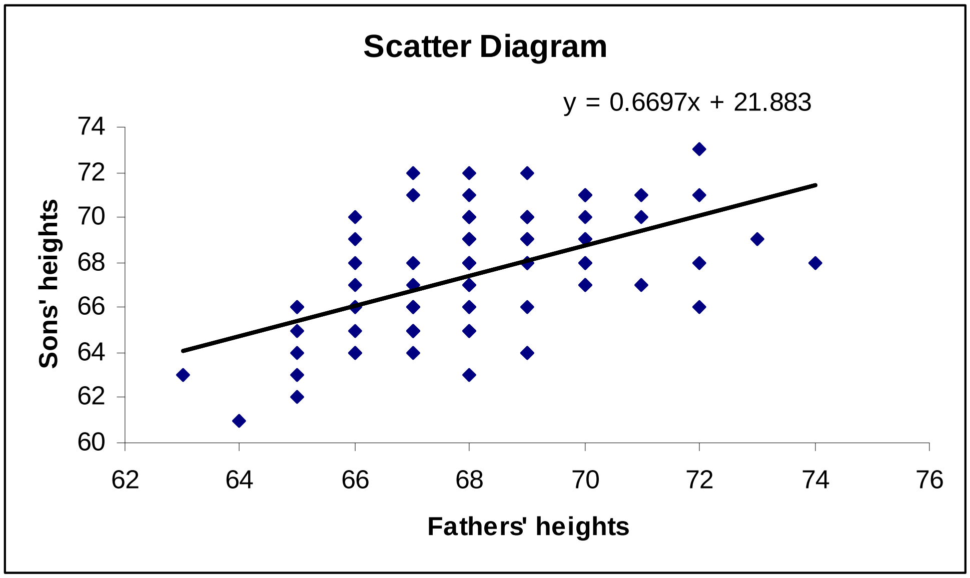
c The quartiles are 193, 223, and 259.

4.91 a

	A	B	C
1		*Father*	*Son*
2	Father	1	
3	Son	0.5347	1

There is a moderately strong positive linear relationship between the heights of fathers and sons.

b



For each additional inch of a father's height the height of his son increases on average by .6697 inch.

4.92a

	A	B	C
1		*Temperature*	*Winning times*
2	Temperature	1	
3	Winning times	0.5984	1

b There is a moderately strong positive linear relationship between the winning times of women and temperatures.

c They appear to provide the same type of information.

4.93a

	A	B	C
1		*Temperature*	*Winning times*
2	Temperature	1	
3	Winning times	0.7242	1

b There is a moderately strong positive linear relationship between the winning times of men and temperatures.

c They appear to provide the same type of information.

4.94a

	A	B
1	*Debts*	
2		
3	Mean	12,067
4	Standard Error	179.9
5	Median	12,047
6	Mode	11,621
7	Standard Deviation	2,632
8	Sample Variance	6,929,745
9	Kurtosis	-0.41325
10	Skewness	-0.2096
11	Range	12,499
12	Minimum	4,626
13	Maximum	17,125
14	Sum	2,582,254
15	Count	214

b The mean debt is $12,067. Half the sample incurred debts below $12,047 and half incurred debts above. The mode is $11,621.

Case 4.1

Ages:

	Means	Medians	Standard deviations
BMW	45.3	45	4.4
Cadillac	61.0	61	3.7
Lexus	50.4	50	6.1
Lincoln	59.7	60	4.7
Mercedes	52.3	52	7.7

Incomes:

	Means	Medians	Standard deviations
BMW	140,544	139,908	33,864
Cadillac	107,832	106,997	15,398
Lexus	154,404	155,846	30,525
Lincoln	111,199	110,488	21,173
Mercedes	184,215	186,070	47,554

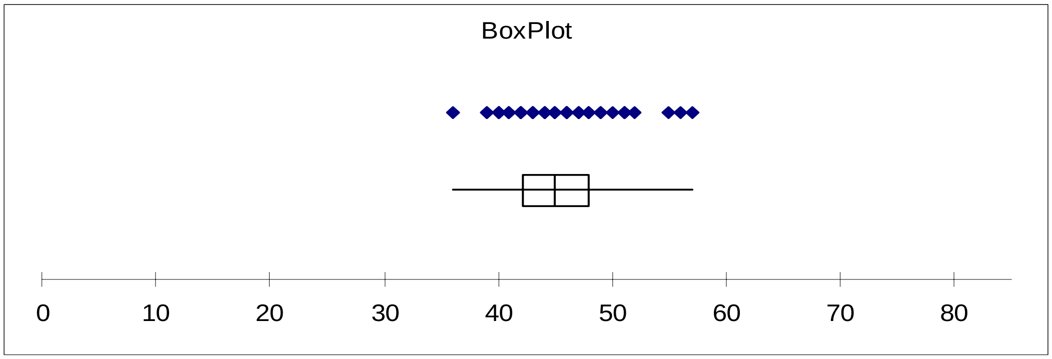
Education

	Means	Medians	Standard deviations
BMW	15.8	16	1.9
Cadillac	12.8	13	1.6
Lexus	15.8	16	2.5
Lincoln	13.1	13	1.6
Mercedes	17.3	17	1.8

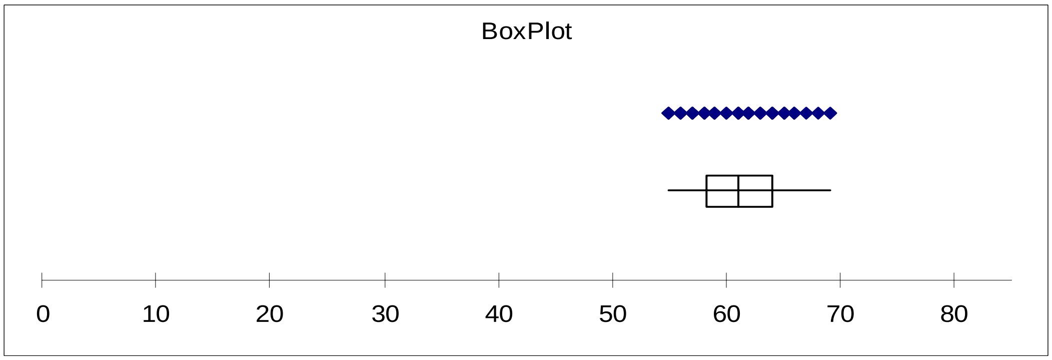
b

Ages

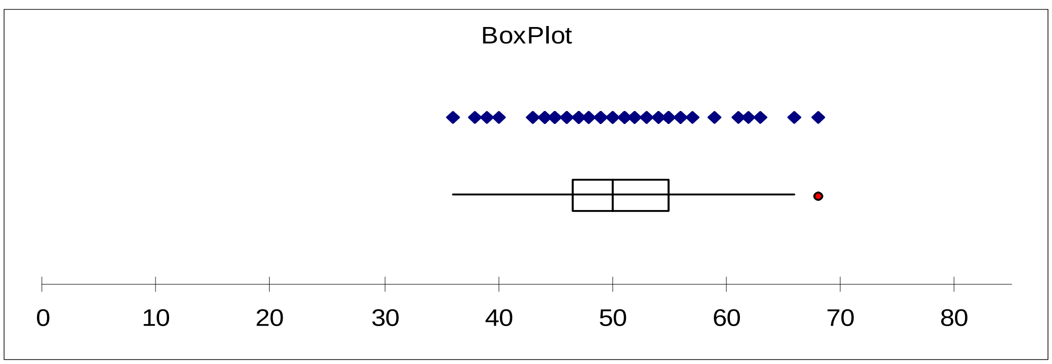
BMW



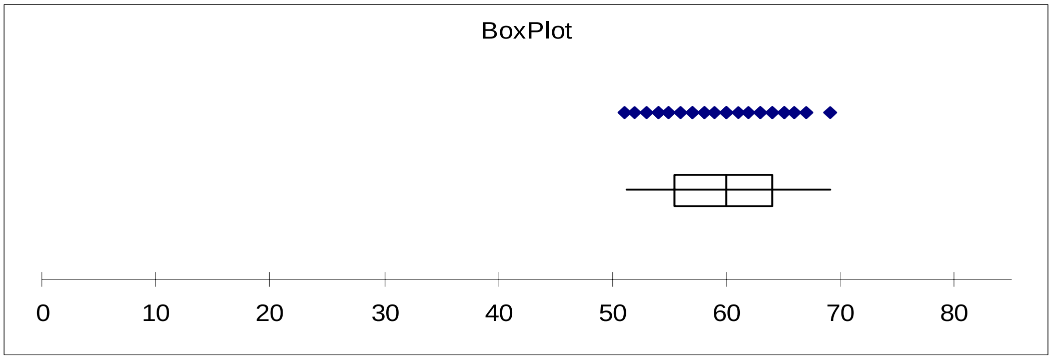
Cadillac



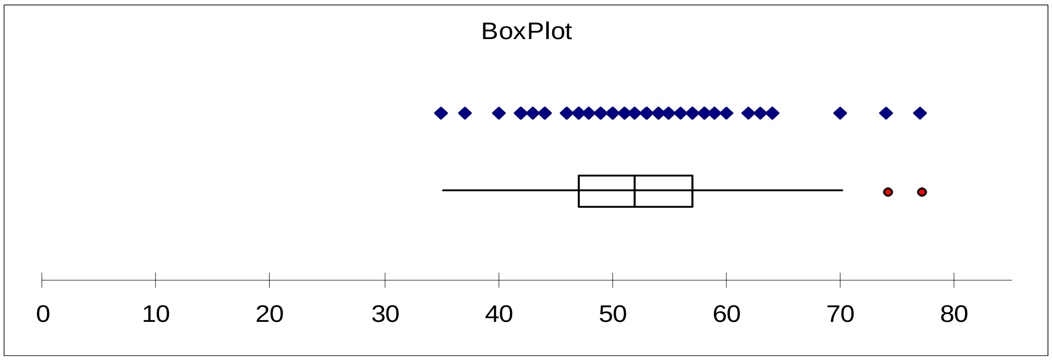
Lexus



Lincoln

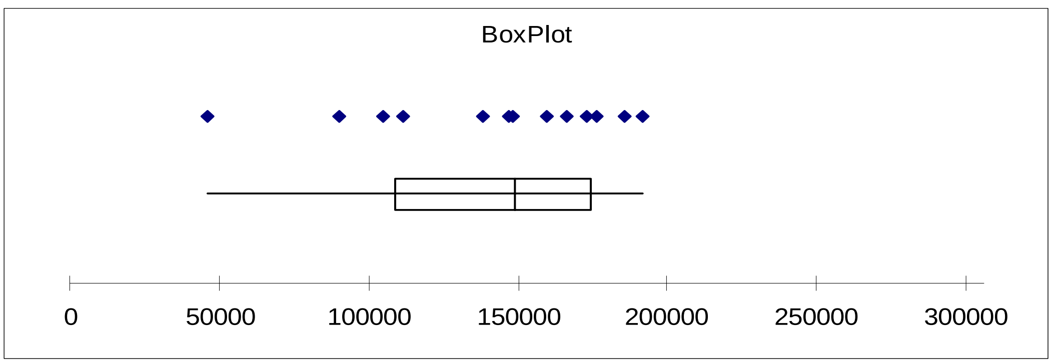


Mercedes

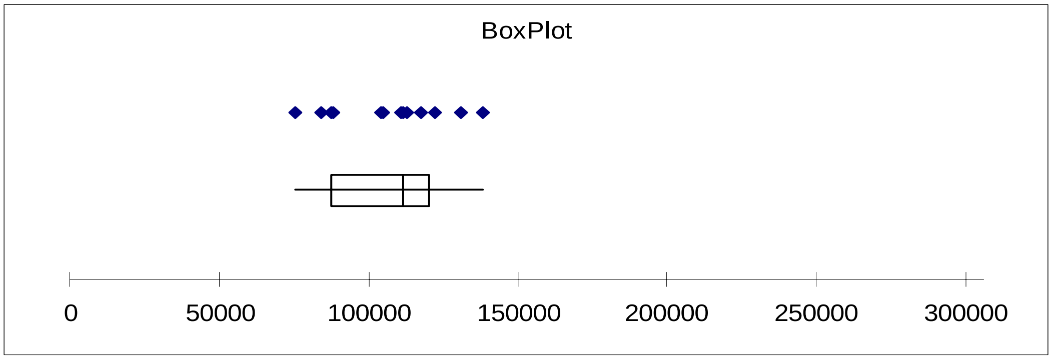


Income

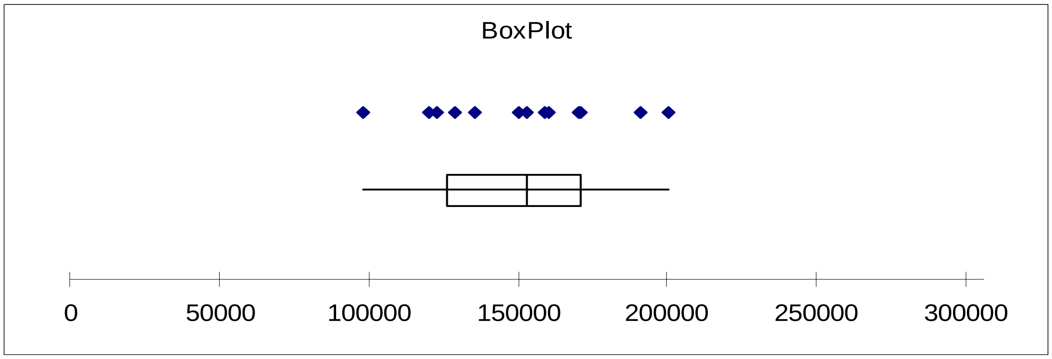
BMW



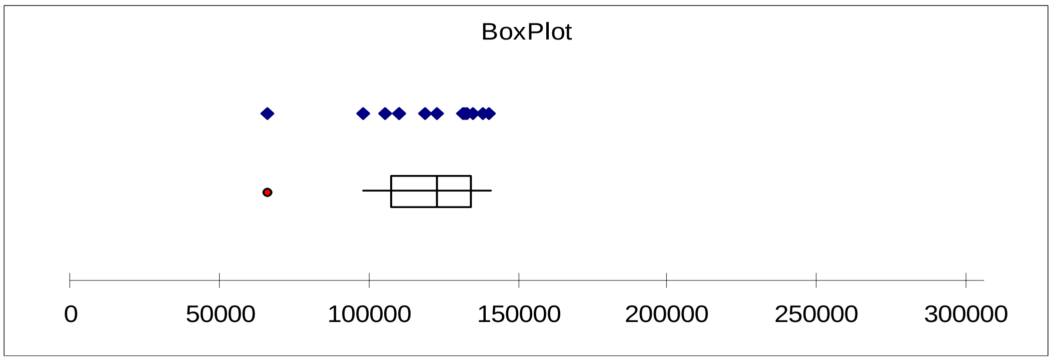
Cadillac



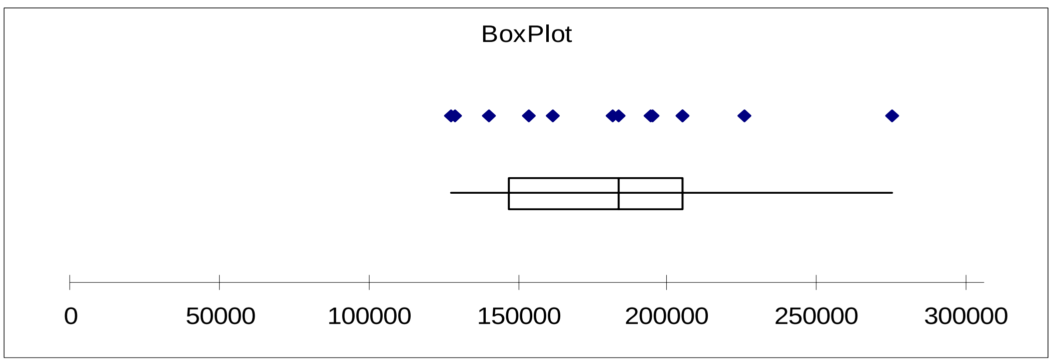
Lexus



Lincoln

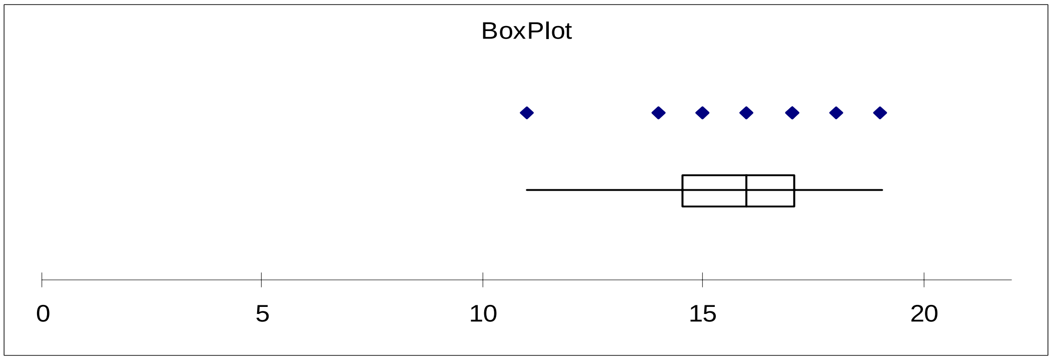


Mercedes

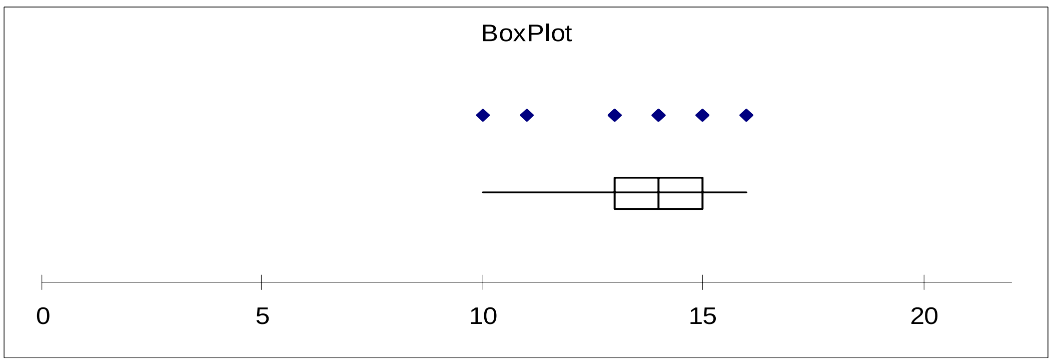


Education

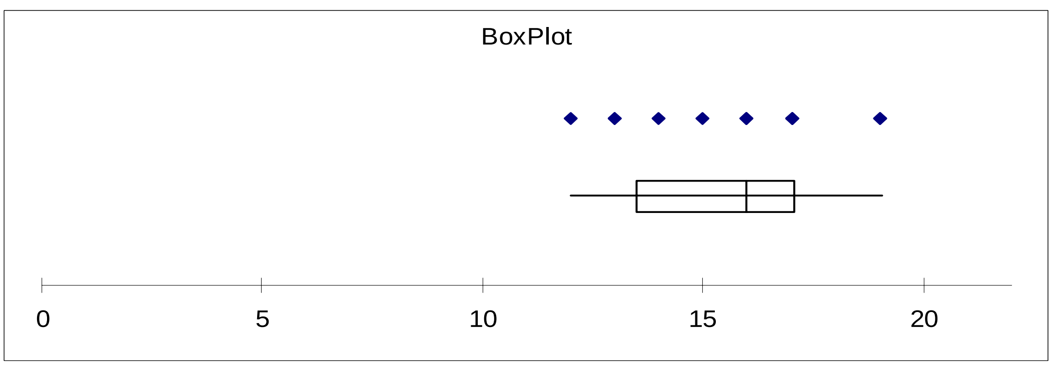
BMW



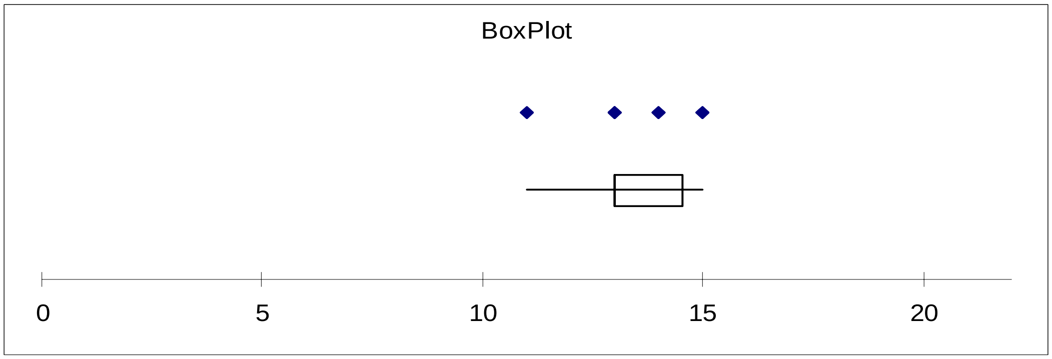
Cadillac



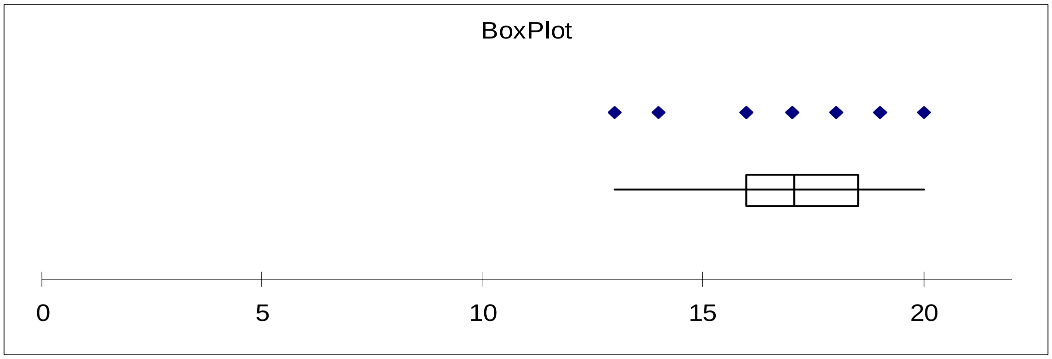
Lexus



Lincoln



Mercedes



The statistics and box plots paint a clear picture. Cadillac owners are older, earn less income, and have less education than the owners of the other luxury cars.

Case 4.2

	A	B	C
1		*Pct Reject*	*Pct Yes*
2	Pct Reject	1	
3	Pct Yes	-0.1787	1

There is a weak negative linear relationship between percentage of rejected ballots and Percentage of "yes" votes.

	A	B	C
1		*Pct Reject*	*Pct Allo*
2	Pct Reject	1	
3	Pct Allo	0.3600	1

There is a moderate positive linear relationship between percentage of rejected ballots and Percentage of Allophones.

	A	B	C
1		*Pct Reject*	*Pct Anglo*
2	Pct Reject	1	
3	Pct Anglo	0.0678	1

There is a very weak positive linear relationship between percentage of rejected ballots and Percentage of Allophones.

The statistics provide some evidence that electoral fraud has taken place.