SOUTH-WESTERN
CENGAGE Learning

Chapter 4

4.1 a $\bar{x} = \frac{\sum x_i}{n} = \frac{52+25+15+0+104+44+60+30+33+81+40+5}{12} = \frac{489}{12} = 40.75$

Ordered data: 0, 5, 15, 25, 30, 33, 40, 44, 52, 60, 81, 104; Median = (33 + 40)/2 = 36.5

Mode = all

4.2 $\bar{x} = \frac{\sum x_i}{n} = \frac{5+7+0+3+15+6+5+9+3+8+10+5+2+0+12}{15} = \frac{90}{15} = 6.0$

Ordered data: 0, 0, 2, 3, 3, 5, 5, 5, 6, 7, 8, 9, 10, 12, 15; Median = 5

Mode = 5

4.3 a $\bar{x} = \frac{\sum x_i}{n} = \frac{5.5+7.2+1.6+22.0+8.7+2.8+5.3+3.4+12.5+18.6+8.3+6.6}{12} = \frac{102.5}{12} = 8.54$

Ordered data: 1.6, 2.8, 3.4, 5.3, 5.5, 6.6, 7.2, 8.3, 8.7, 12.5, 18.6, 22.0; Median = 6.9

Mode = all

b The mean number of miles jogged is 8.54. Half the sample jogged more than 6.9 miles and half jogged less.

4.4 a $\bar{x} = \frac{\sum x_i}{n} = \frac{33+29+45+60+42+19+52+38+36}{9} = \frac{354}{9} = 39.3$

Ordered data: 19, 29, 33, 36, 38, 42, 45, 52, 60; Median = 38

Mode: all

b The mean amount of time is 39.3 minutes. Half the group took less than 38 minutes.

4.5 a $\bar{x} = \frac{\sum x_i}{n} = \frac{14+8+3+2+6+4+9+13+10+12+7+4+9+13+15+8+11+12+4+0}{20}$

$= \frac{164}{20} = 8.2$

Ordered data: 0, 2, 3, 4, 4, 4, 6, 7, 8, 8, 9, 9, 10, 11, 12, 12, 13, 13, 14, 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.

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

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

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 is best.

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

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 is best.

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

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

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

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

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

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 best because $1000(1.188)^4 = 2000$.

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

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

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 The geometric mean is best because $12(1.130)^6 = 25$.

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.

4.13 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.

4.14a $\bar{x}$ = 128.07; median = 136.00

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

4.15a $\bar{x}$ = 29.48; median = 30.00

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

c The mean and median of commuting time in New York is larger than that in Los Angeles.

4.16a $\bar{x}$ = 30.53; median = 31

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

4.17a $\bar{x}$ = 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.

4.18a $\bar{x}$ = 519.20; median = 523.00

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

4.19 $\bar{x} = \frac{\sum x_i}{n} = \frac{9+3+7+4+1+7+5+4}{8} = \frac{40}{8} = 5$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(9-5)^2 + (3-5)^2 + ... + (4-5)^2}{8-1} = \frac{46}{7} = 6.57$$

4.20 $\bar{x} = \frac{\sum x_i}{n} = \frac{4+5+3+6+5+6+5+6}{8} = \frac{40}{8} = 5$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(4-5)^2 + (5-5)^2 + ... + (6-5)^2}{8-1} = \frac{8}{7} = 1.14$$

4.21 $\bar{x} = \frac{\sum x_i}{n} = \frac{12+6+22+31+23+13+15+17+21}{9} = \frac{160}{9} = 17.78$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(12-17.78)^2 + (6-17.78)^2 + ... + (21-17.78)^2}{9-1} = \frac{433.56}{8} = 54.19$$

$$s = \sqrt{s^2} = \sqrt{54.19} = 7.36$$

4.22 $\bar{x} = \frac{\sum x_i}{n} = \frac{0+(-5)+(-3)+6+4+(-4)+1+(-5)+0+3}{10} = \frac{-3}{10} = -.30$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{[(0-(-.3))^2 + ((-5)-(-.3))^2 + ... + (3-(-.3))^2}{10-1} = \frac{136.1}{9} = 15.12$$

$$s = \sqrt{s^2} = \sqrt{15.12} = 3.89$$

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 6, 6, 6, 6, 6

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 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.

4.32 $s^2 = 40.73$ 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

4.33 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 $s^2 = .0858$ 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.

4.35 $\bar{x} = 175.73$ and 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..

4.36a s = 15.01

b In approximately 68% of the days the number of arrivals falls within 15.01 of the mean; in approximately 95% of the hours the number of arrivals falls within 30.02 of the mean; in approximately 99.7% of the hours the number of arrivals falls within 45.03 of the mean

4.37 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.

4.38 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.

4.39 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.

4.40 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.

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

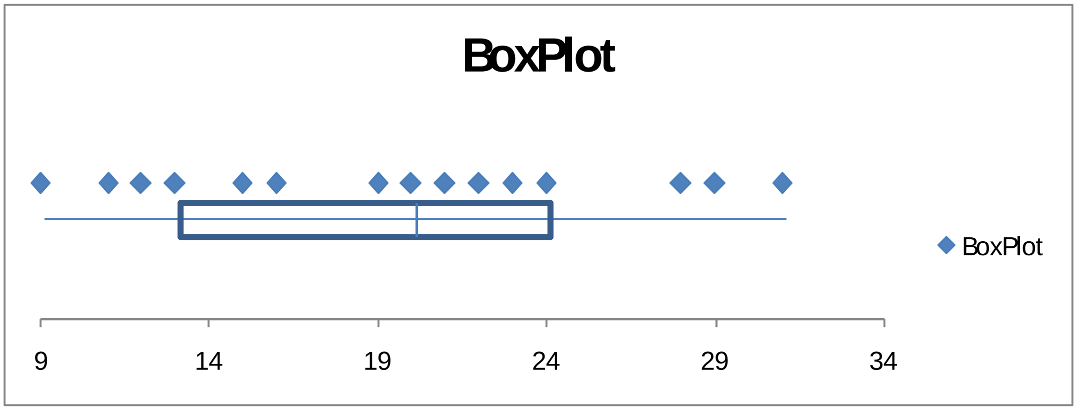
Sixth decile: $L_{60} = (15+1)\dfrac{60}{100}$ = (16)(.60) = 9.6; the sixth decile is 17 + .6(18 – 17) = 17.6.

4.42 Interquartile range = 15.6 –13.05 = 2.55

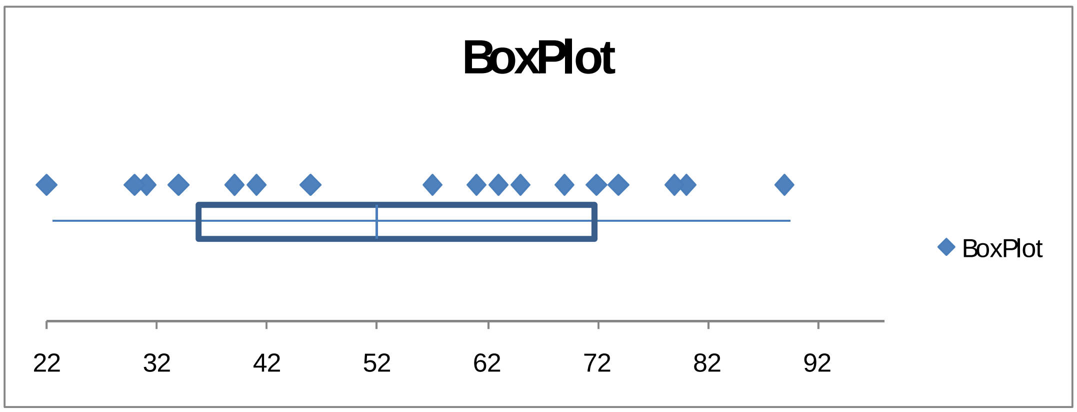
4.43 Interquartile range = 7 – 3 = 4

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

4.45



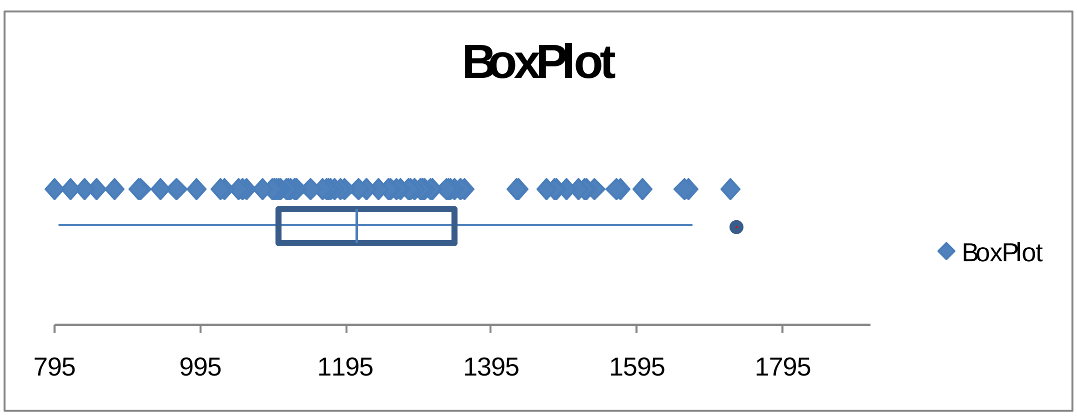
4.46



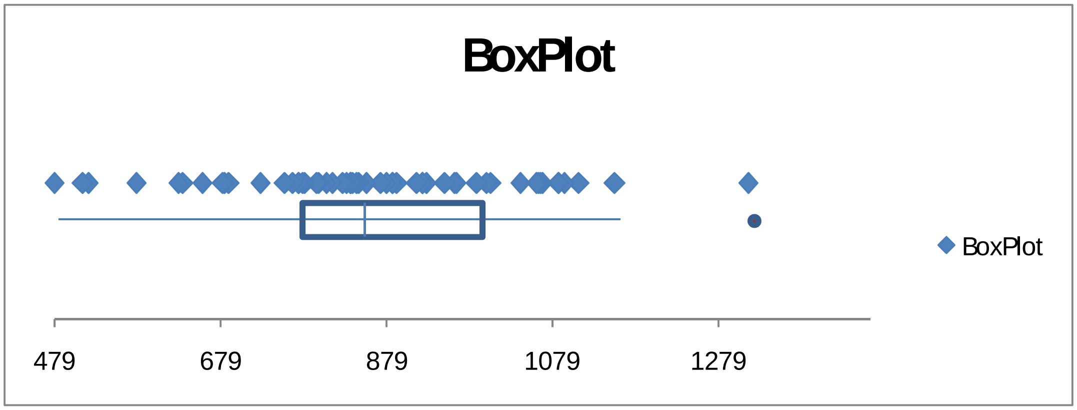
4.47a First quartile = 2, second quartile = 4, and third quartile = 8.

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

4.48 Dogs



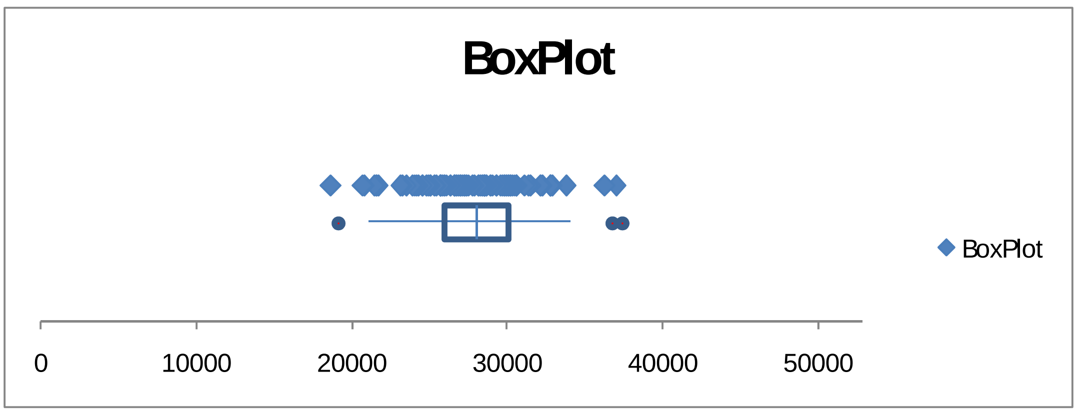
Cats



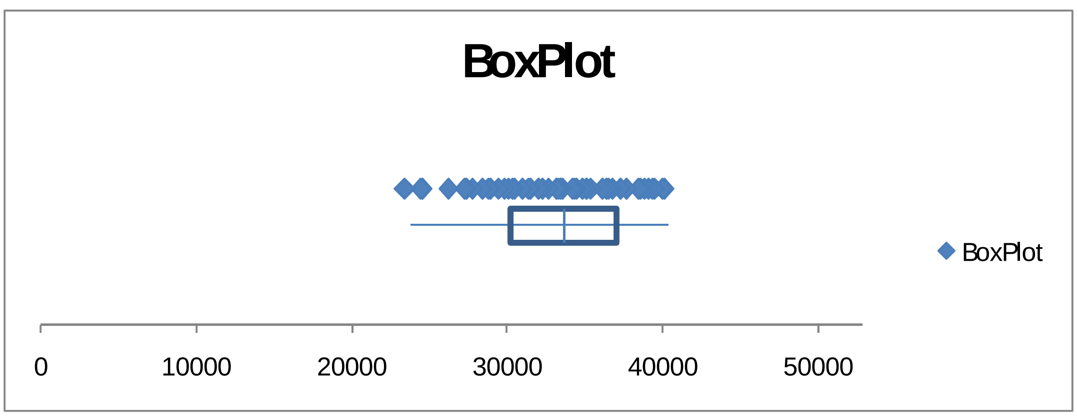
Dogs cost more money than cats. Both sets of expenses are positively skewed.

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

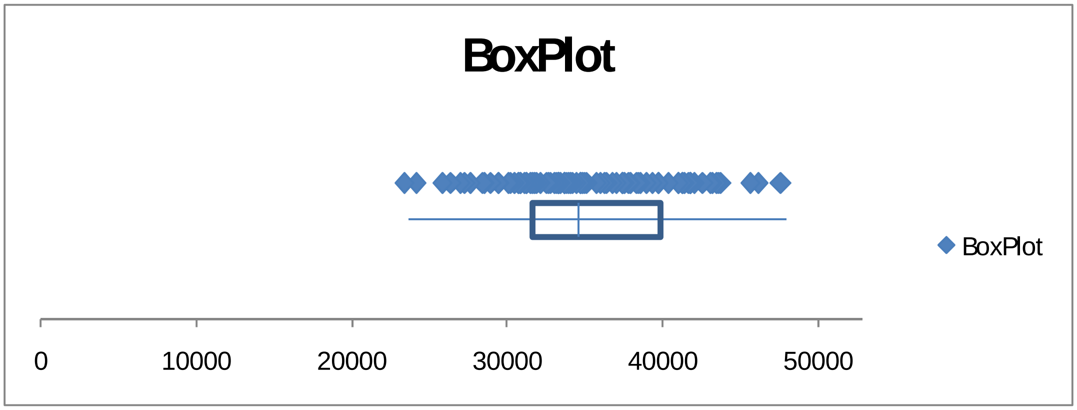
4.50 BA



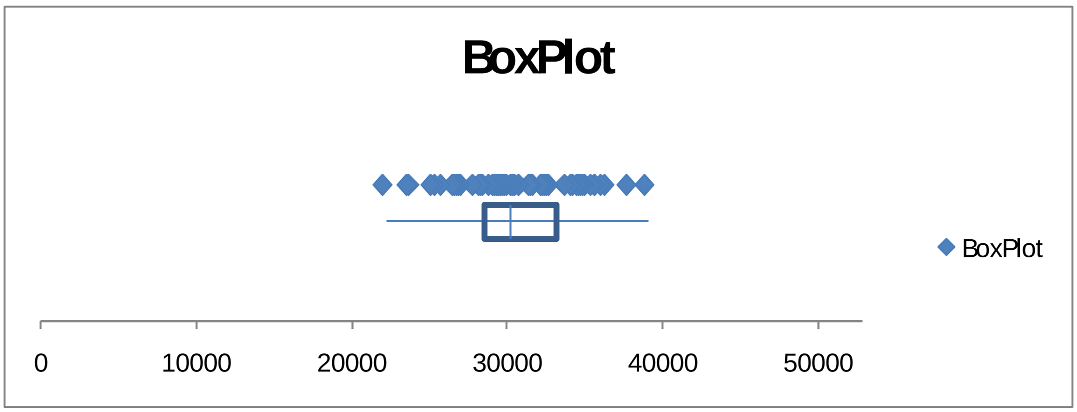
BSc



BBA

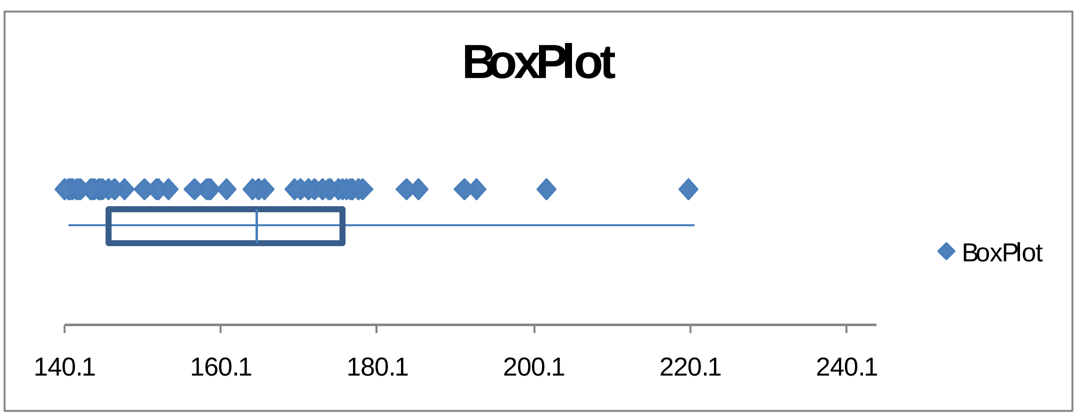


Other



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

4.51 a



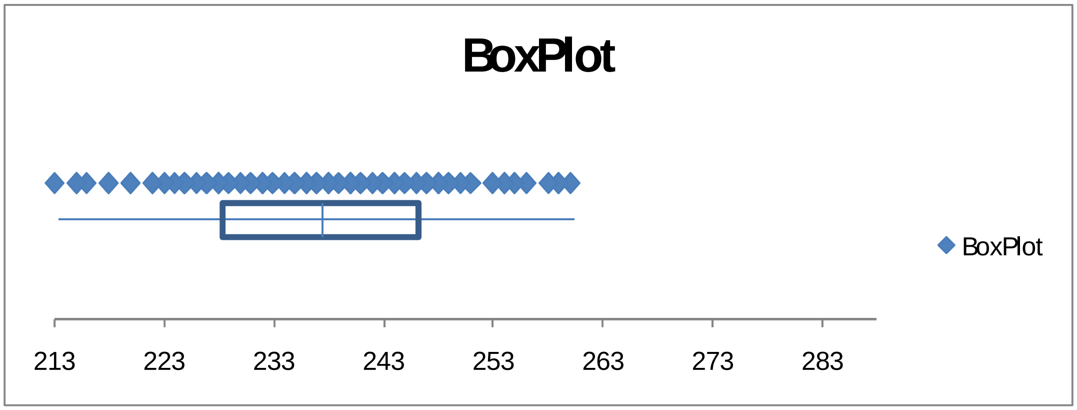
b The quartiles are 145.11, 164.17, and 175.18

c There are no outliers.

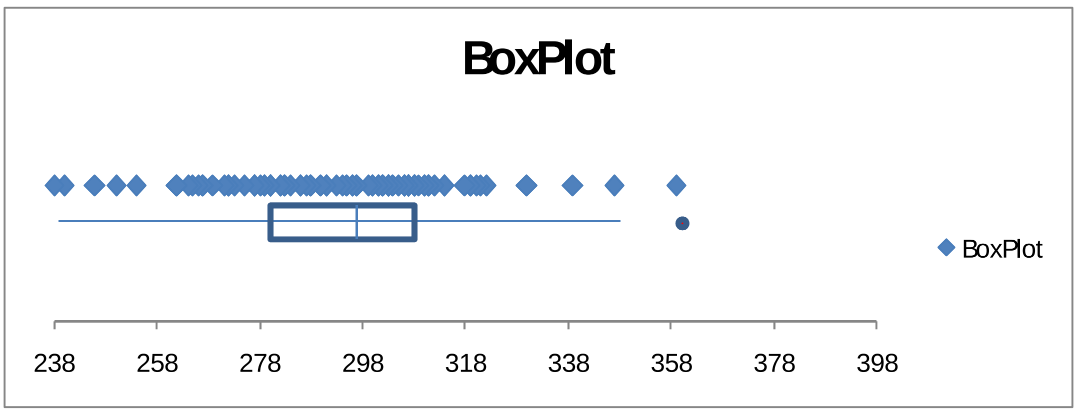
d The data are positively skewed. One-quarter of the times are below 145.11 and one-quarter are above 175.18.

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 a. $r = \frac{s_{xy}}{s_x s_y} = \frac{-150}{(16)(12)} = -.7813$

There is a moderately strong negative linear relationship.

b. $R = r^2 = (-.7813)^2 = .6104$

61.04% of the variation in y is explained by the variation in x.

4.57a.

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	20	14	400	196	280
	40	16	1600	256	640
	60	18	3600	324	1080
	50	17	2500	289	850
	50	18	2500	324	900
	55	18	3025	324	990
	60	18	3600	324	1080
	70	20	4900	400	1400
Total	405	139	22,125	2,437	7,220

$$\sum_{i=1}^{n} x_i = 405 \quad \sum_{i=1}^{n} y_i = 139 \quad \sum_{i=1}^{n} x_i^2 = 22{,}125 \quad \sum_{i=1}^{n} y_i^2 = 2{,}437 \quad \sum_{i=1}^{n} x_i y_i = 7{,}220$$

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{8-1}\left[7{,}220 - \frac{(405)(139)}{8}\right] = 26.16$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{8-1}\left[22{,}125 - \frac{(405)^2}{8}\right] = 231.7$$

$$s_x = \sqrt{s_x^2} = \sqrt{231.7} = 15.22$$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] = \frac{1}{8-1}\left[2{,}437 - \frac{(139)^2}{8}\right] = 3.13$$

$$s_y = \sqrt{s_y^2} = \sqrt{3.13} = 1.77$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{26.16}{\sqrt{(15.22)(1.77)}} = ..9711$$

$$R^2 = r^2 = .9711^2 = .9430$$

The covariance is 26.16, the coefficient of correlation is .9711 and the coefficient of determination is .9430.

94.30% of the variation in expenses is explained by the variation in total sales.

b. $$b_1 = \frac{s_{xy}}{s_x^2} = \frac{26.16}{231.7} = .113$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{405}{8} = 50.63$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{139}{8} = 17.38$$

$$b_0 = \bar{y} - b_1\bar{x} = 17.38 - (.113)(50.63) = 11.66$$

The least squares line is

$$\hat{y} = 11.66 + .113x$$

The estimated variable cost is .113 and the estimated fixed cost is 11.66.

4.58

	x_i	y_i	x_i^2	y_i^2	x_iy_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 $R^2 = r^2 = .8811^2 = .7763$

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

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

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

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

The least squares line is

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

e. 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.

4.59

	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
	599	9.6	358,801	92.16	5750.4
	689	8.8	474,721	77.44	6063.2
	584	7.4	341,056	54.76	4321.6
	631	10.0	398,161	100.00	6310.0
	594	7.8	352,836	60.84	4632.2
	643	9.2	413,449	84.64	5915.6
	656	9.6	430,336	92.16	6297.6
	594	8.4	352,836	70.56	4989.6
	710	11.2	504,100	125.44	7952.0
	611	7.6	373,321	57.76	4643.6
	593	8.8	351,649	77.44	5218.4
	683	8.0	466,489	64.00	5464.0
Total	7,587	106.4	4,817,755	957.2	67,559.2

$$\sum_{i=1}^{n} x_i = 7,587 \quad \sum_{i=1}^{n} y_i = 106.4 \quad \sum_{i=1}^{n} x_i^2 = 4,817,755 \quad \sum_{i=1}^{n} y_i^2 = 957.2 \quad \sum_{i=1}^{n} x_i y_i = 67,559.2$$

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{12-1}\left[67,559.2 - \frac{(7,587)(106.4)}{12}\right] = 26.16$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{12-1}\left[4,817,755 - \frac{(7,587)^2}{12}\right] = 1,897.7$$

$$s_x = \sqrt{s_x^2} = \sqrt{1{,}897.7} = 43.56$$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right] = \frac{1}{12-1}\left[957.2 - \frac{(106.4)^2}{12}\right] = 1.25$$

$$s_Y = \sqrt{s_Y^2} = \sqrt{1.25} = 1.12$$

$$r = \frac{s_{xy}}{s_x s_y} = \frac{26.16}{\sqrt{(43.56)(1.12)}} = .5362$$

$R^2 = r^2 = .5362^2 = .2875$

The covariance is 26.16, the coefficient of correlation is .5362, and the coefficient of determination is .2875. The coefficient of determination tells us that 28.75% of the variation in MBA GPAs is explained by the variation in GMAT scores.

4.60

	A	B	C
1		*Unemployment Rate*	*Employment Rate*
2	Unemployment Rate	1	
3	Employment Rate	-0.6332	1

$R^2 = r^2 = (-.6332)^2 = .4009$; 40.09% of the variation in the employment rate is explained by the variation in the unemployment rate.

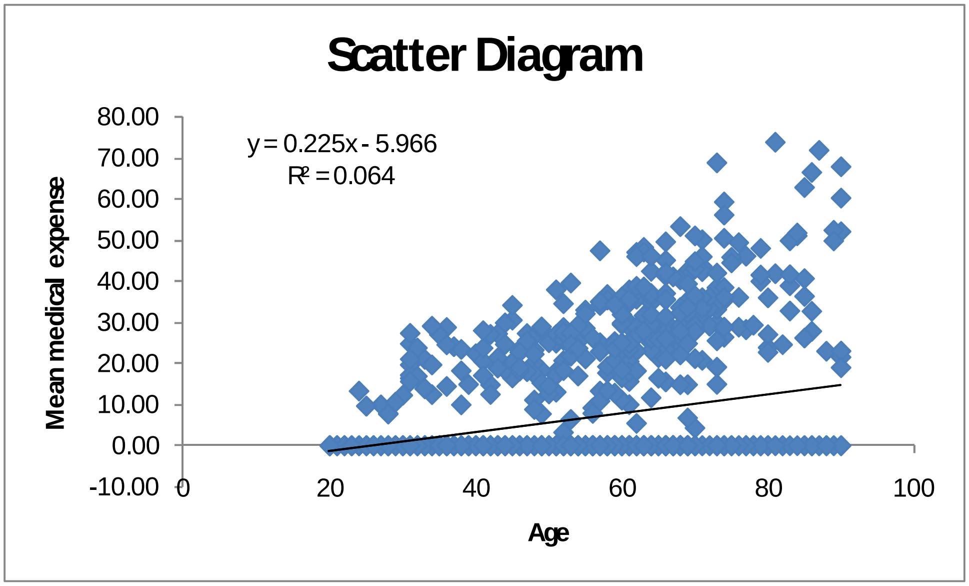
4.61 a

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

$R^2 = r^2 = (.2543)^2 = .0647$.

b There is a weak linear relationship between age and medical expenses. Only 6.47% of the variation in average medical bills is explained by the variation in age.

c

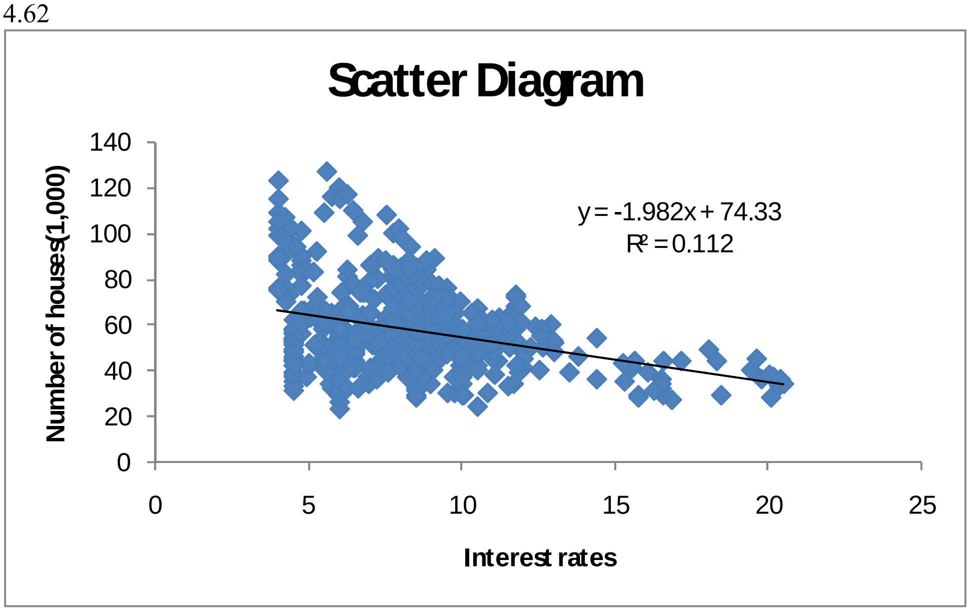


The least squares line is $\hat{y} = -5.966 + .2257x$

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

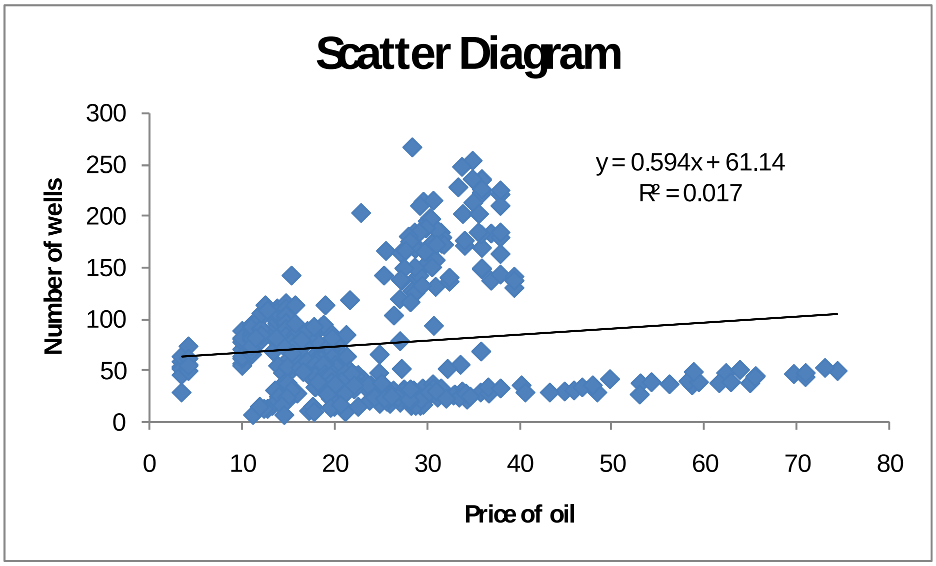
e Charge 25 cents per day per year of age.

4.62



The coefficient of determination is .112. Only 11.2% of the variation in the number of houses sold is explained by the variation in interest rates.

4.63



Only 1.7% of the variation in the number of wells drilled is explained by the variation in the price of oil.

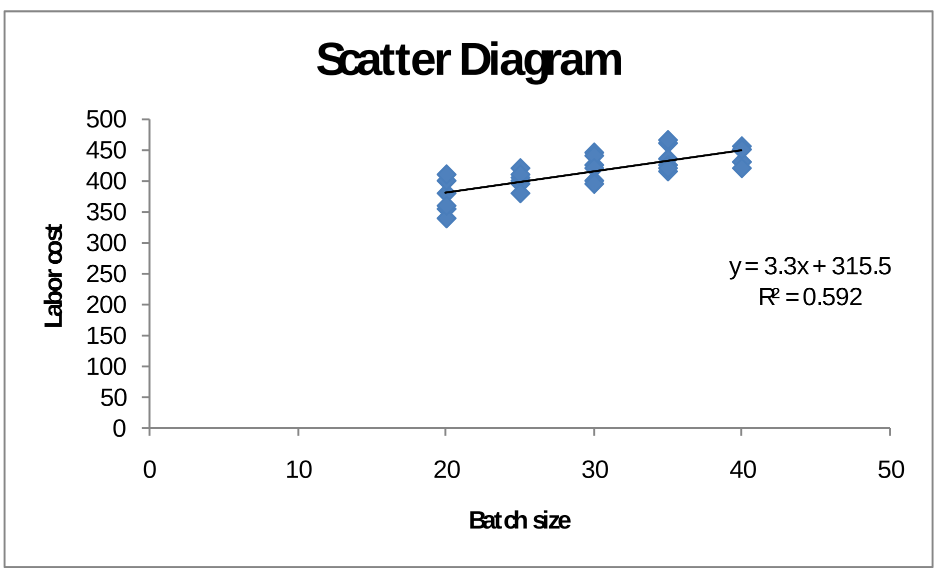
4.64

	A	B	C
1		*Unemployment rate*	*Help wanted index*
2	Unemployment rate	1	
3	Help wanted index	0.0830	1

There is a positive relationship between the two variables.

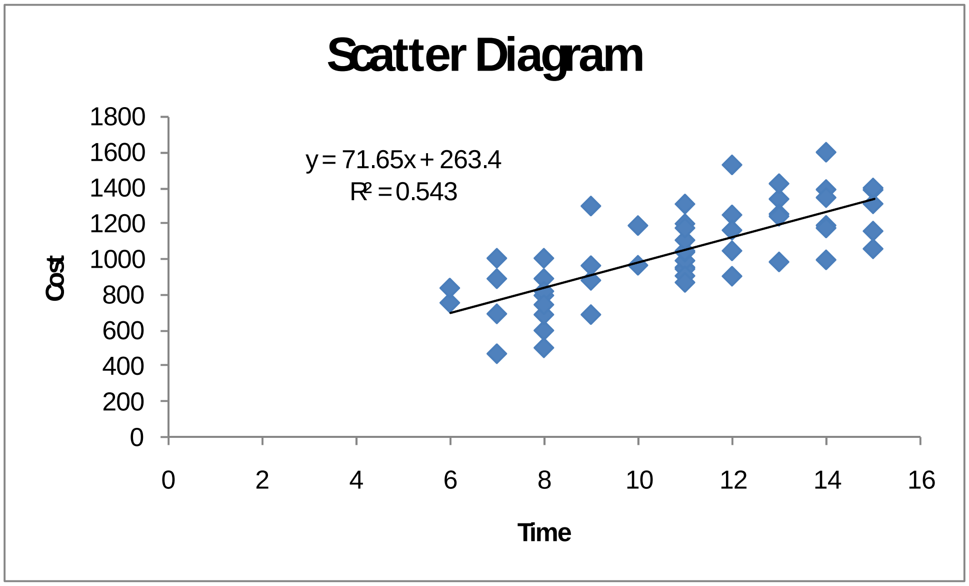
$R^2 = r^2 = (.0830)^2 = .0069$. The relationship is very week.

4.65



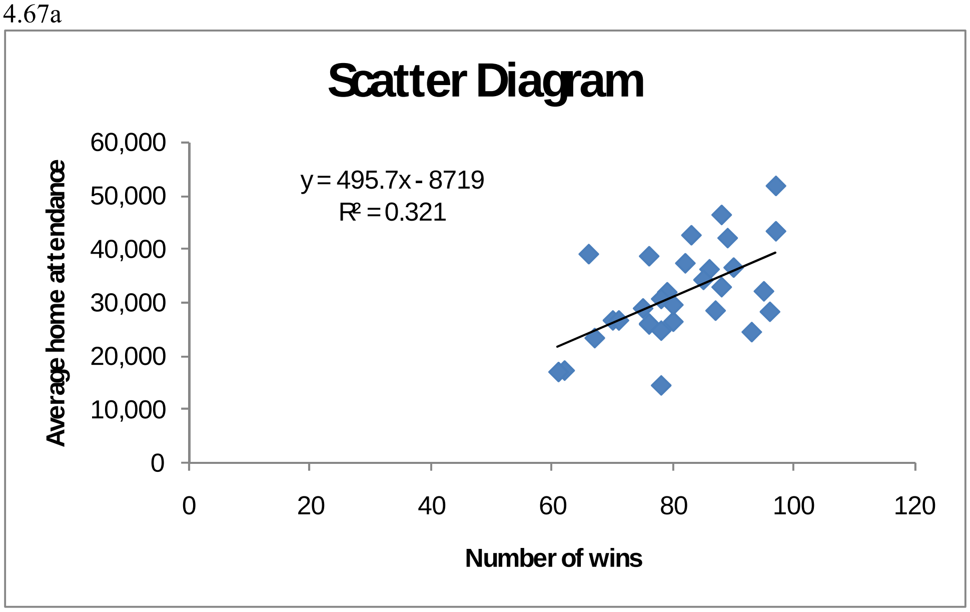
$\hat{y} = 315.5 + 3.3x$; Fixed costs = \$315.50, variable costs = \$3.30

4.66



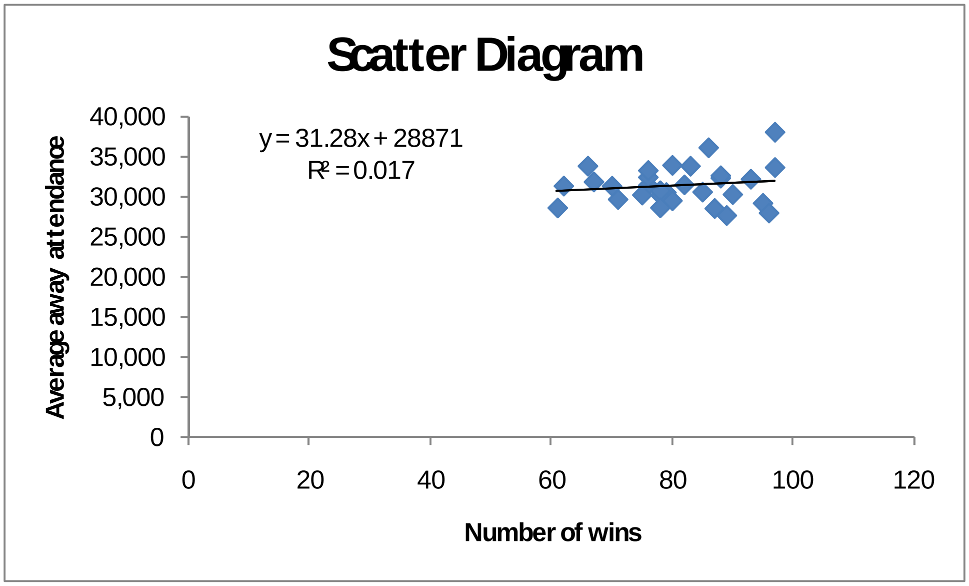
$\hat{y} = 263.4 + 71.65x$; Estimated fixed costs = \$263.40, estimated variable costs = \$71.65

4.67a



b The slope coefficient is 495.7; home attendance increases on average by 495.7 for each win. There is a moderately weak positive linear relationship between the two variables.

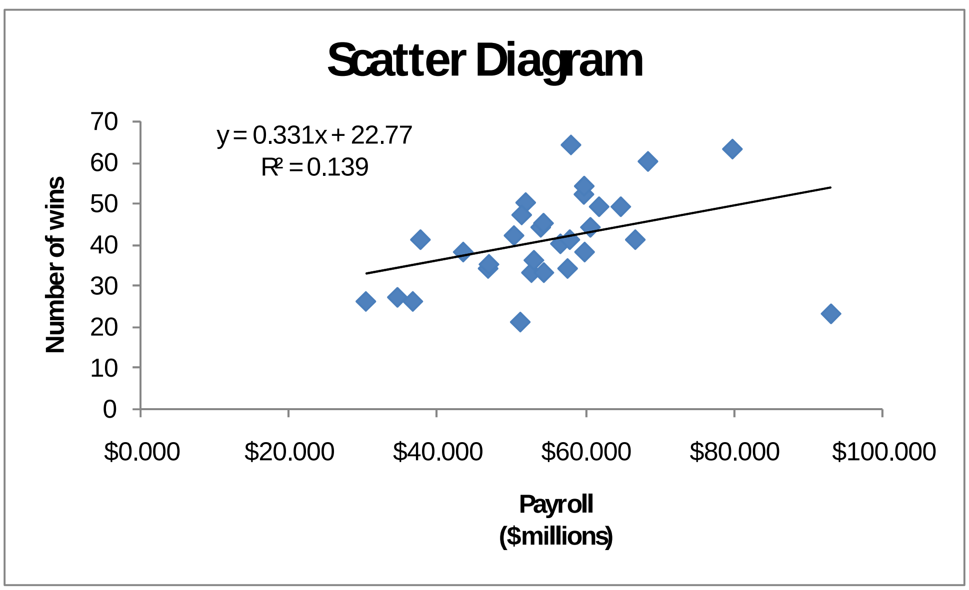
4.68a



R^2 = .017; there is a very week relationship between the two variables.

b The slope coefficient is 31.28; away attendance increases on average by 31.28 for each win. However, the relationship is very weak.

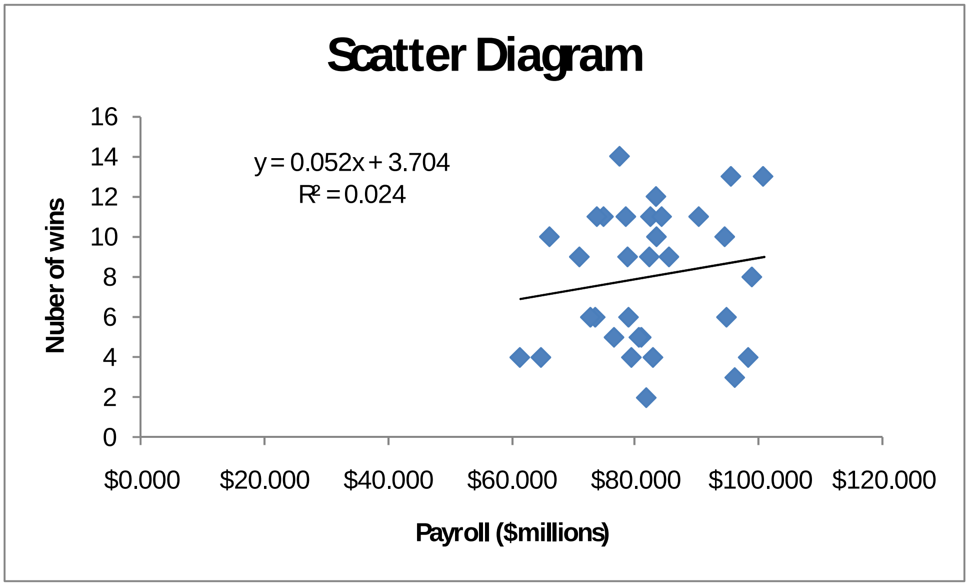
4.69



a. The slope coefficient is .331; for each million dollars in payroll the number of wins increases on average by .331. Thus, to cost of winning one addition game is 1/.331 million = $3.021 million.

b. The coefficient of determination tells us that only 13.9% of the variation in the number of wins is explained by the variation in payroll,

4.70



a. The slope coefficient is .052; for each million dollars in payroll the number of wins increases on average by .052. Thus, to cost of winning one addition game is 1/.052 million = $19.231 million.

b. The coefficient of determination = .024, which reveals that the linear relationship is very weak.

4.71	b_1	R^2
Coca- Cola	.471	.136
Genentech	.811	.068
General Electric	.824	.306
General Motors	1.17	.195
McDonalds	1.077	.337
Motorola	1.330	.270

4.72		
Barrick Gold	.574	.086
Bell Canada Enterprises	.414	.103
Bank of Montreal	.447	.151
MDS Laboratories	.882	.173
Petro-Can	.535	.093
Research in Motion	2.698	.302

4.73

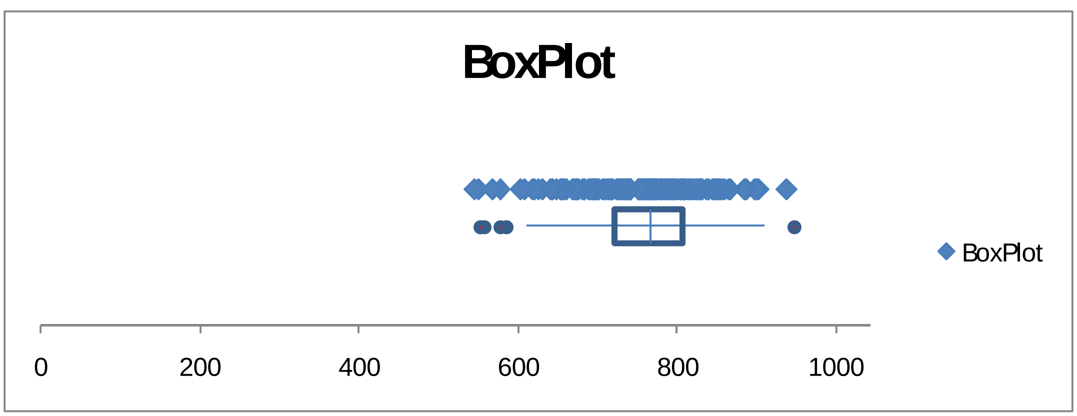
Amgen	.378	.132
Ballard Power Systems	1.684	.428
Cisco Systems	1.446	.636
Intel	1.411	.654
Microsoft	.810	.452

4.74 a

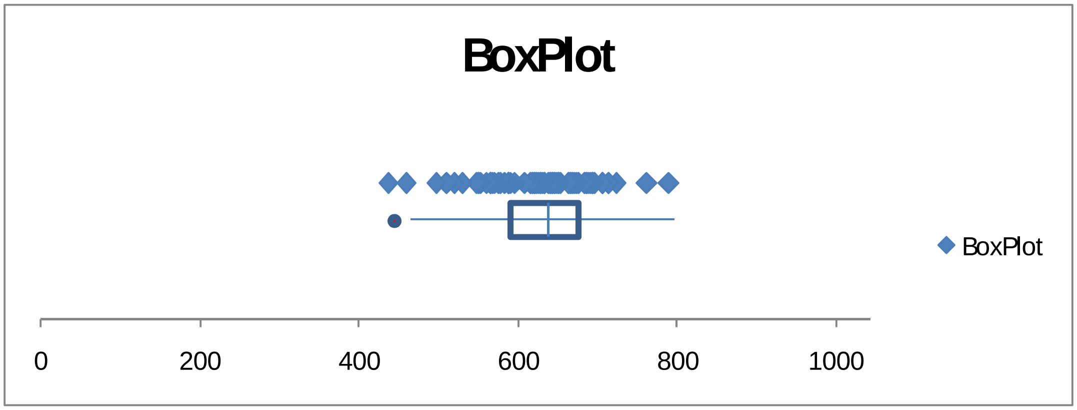
	A	B	C	D	E
1	*Repaid*			*Defaulted*	
2					
3	Mean	752.45		Mean	545.73
4	Standard Error	3.53		Standard Error	6.07
5	Median	757		Median	549.5
6	Mode	759		Mode	552
7	Standard Deviation	52.35		Standard Deviation	54.25
8	Sample Variance	2740.61		Sample Variance	2943.49
9	Kurtosis	-0.23		Kurtosis	-0.67
10	Skewness	-0.11		Skewness	-0.09
11	Range	252		Range	237
12	Minimum	616		Minimum	419
13	Maximum	868		Maximum	656
14	Sum	165539		Sum	43658
15	Count	220		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.

4.75 Repaid loan:



Defaulted on loan:



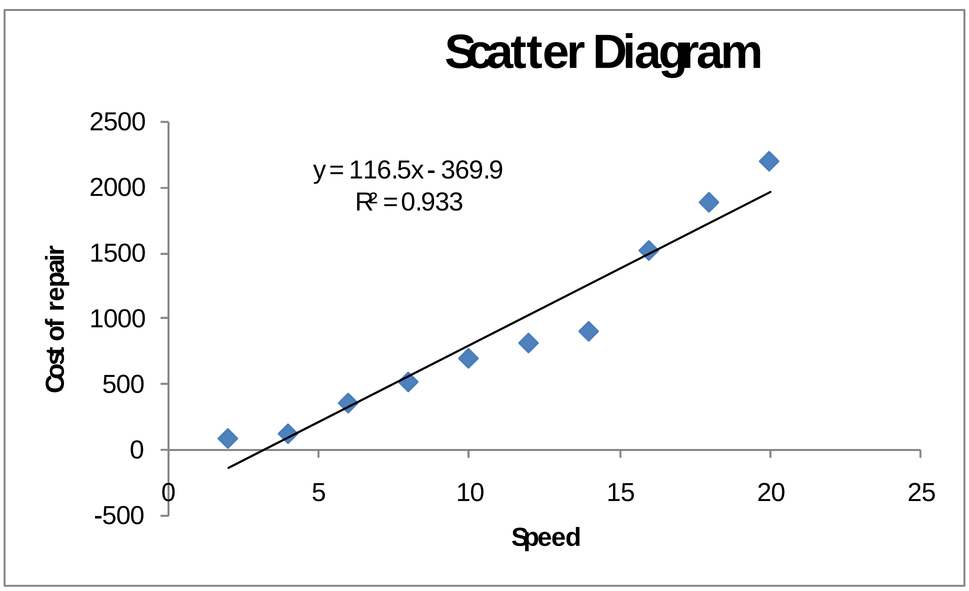
The box plots make it a little easier to see the overlap between the two sets of data (indicating that the scorecard is not very good).

4.76

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

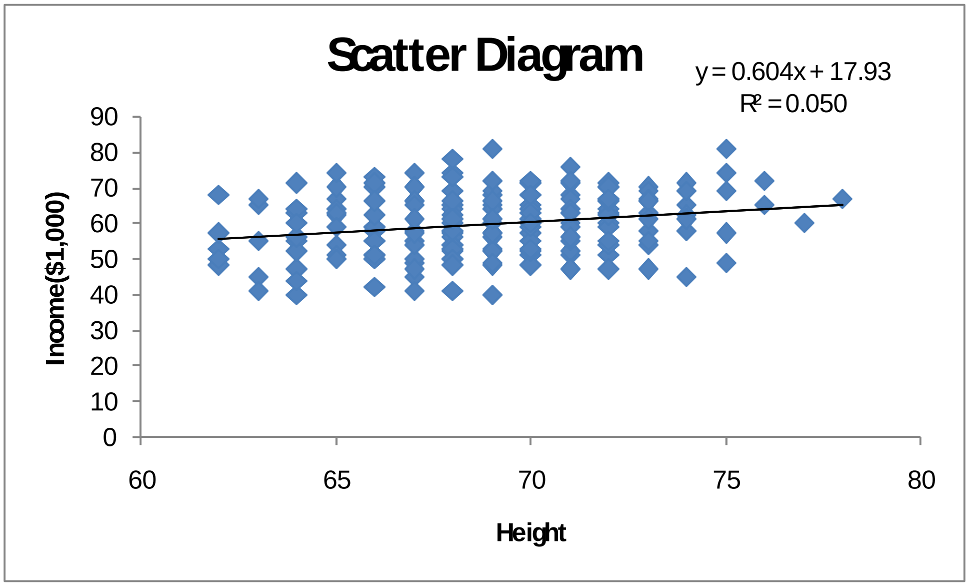
$R^2 = .6784^2 = .4603$; 46.03% of the variation in statistics marks is explained by the variation in calculus marks. The coefficient of determination provides a more precise indication of the strength of the linear relationship.

4.77



The least squares line is $\hat{y} = 369.9 + 116.5x$. On average for each addition mph the cost of repair increases by $116.50.

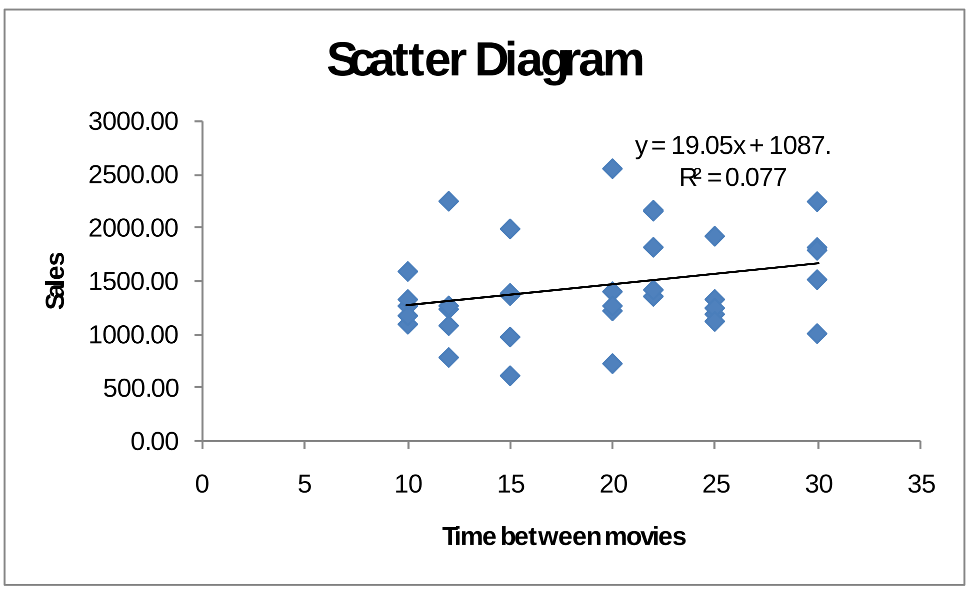
4.78



a $\hat{y} = 17.93 + .6045x$

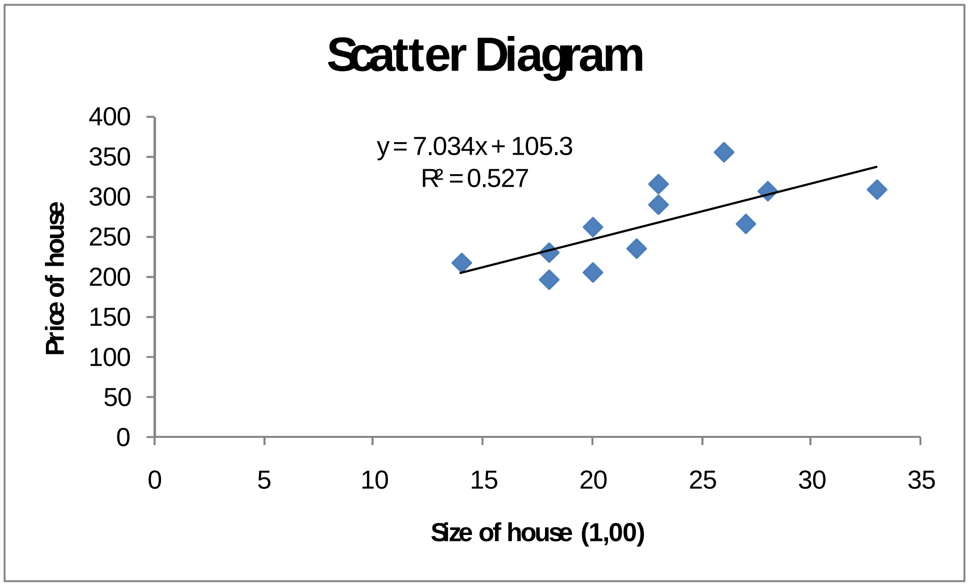
b The coefficient of determination is .050, which indicates that only 5% of the variation in incomes is explained by the variation in heights.

4.79



The coefficient of determination is .077, which indicates that only 7.7% of the variation in sales is explained by the time between movies.

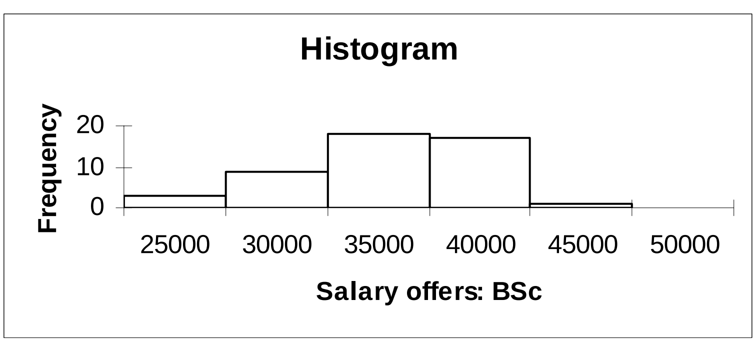
4.80a

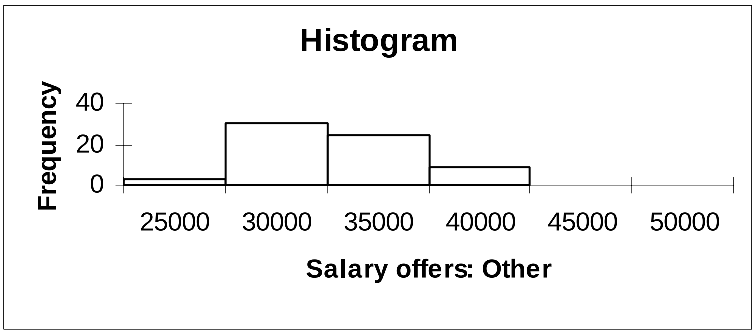


b. The slope coefficient is 7.034; For each additional 100 square feet the price increases on average by $7.034 thousand. More simply for each additional square foot the price increases on average by$70.34.

c. From the least squares line we can more precisely measure the relationship between the two variables.

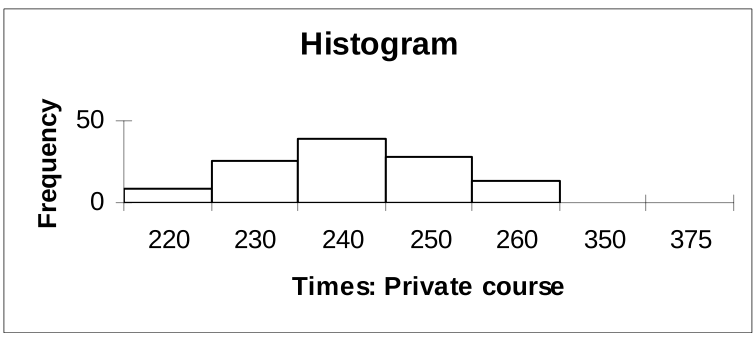
4.81

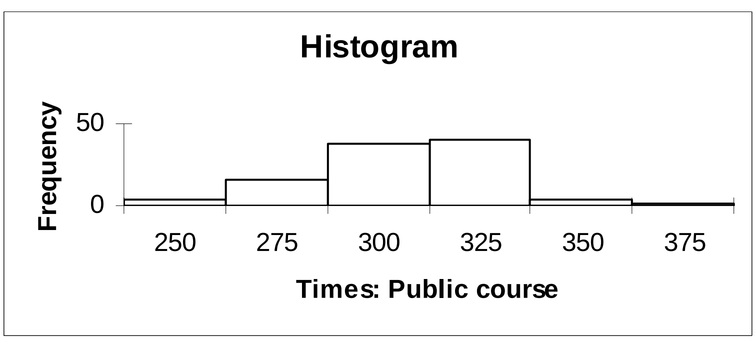




Using the same class limits the histograms provide more detail than do the box plots.

4.82





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

4.83a

	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.

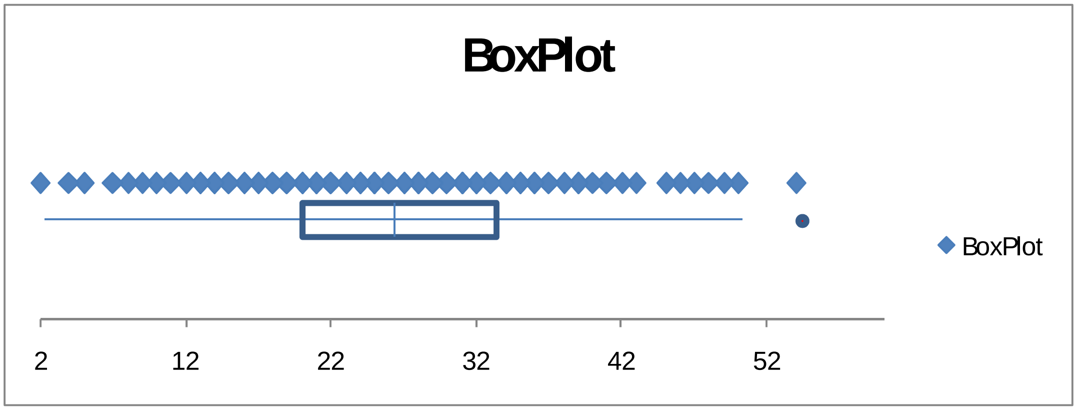
4.84

	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
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.

4.86a mean, median, and standard deviation

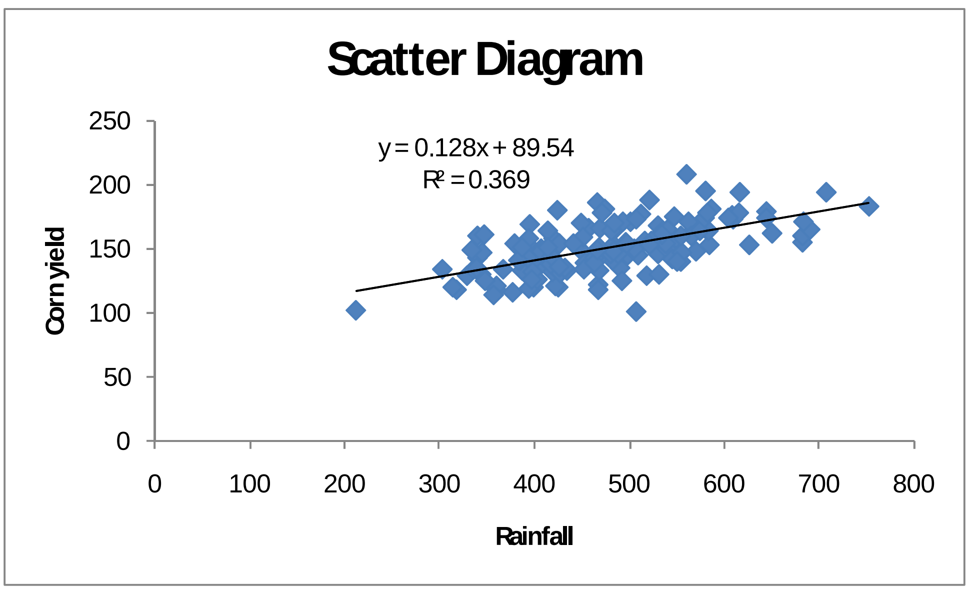
b

	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.

4.87



$R^2 = .369$ and the least squares line is $\hat{y} = 89.54 + .128$ Rainfall

c 36.9% of the variation in yield is explained by the variation in rainfall. For each additional inch of rainfall yield increases on average by .128 bushels.

d We can measure the strength of the linear relationship accurately and the slope coefficient gives information about how rainfall and crop yield are related.

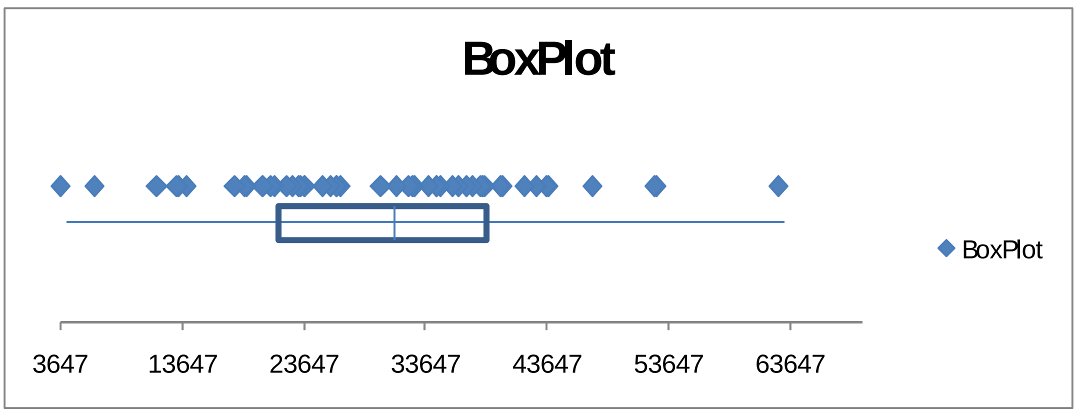
4.88

	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

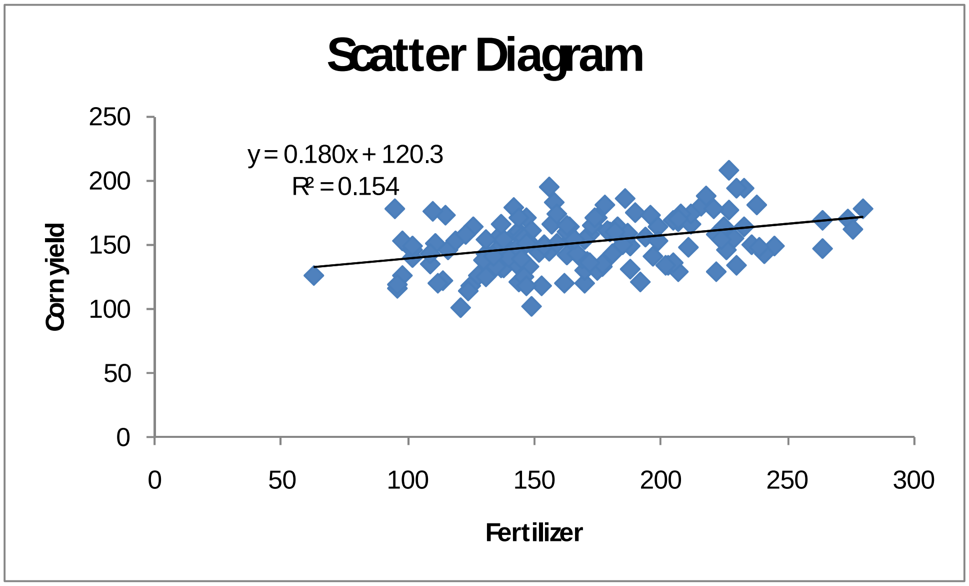
b s^2 = 148,213,791; s = 12,174

c



d The number of coffees sold varies considerably.

4.89



R^2 = .154 and the least squares line $\hat{y}$ = 120.37 + .180 Fertilizer

c 15.4% of the variation in yield is explained by the variation in the amount of fertilizer. For each additional unit of fertilizer yield increases on average by.180 bushels.

d We can measure the strength of the linear relationship accurately and the slope coefficient gives information about how the amount of fertilizer and crop yield are related.

4.90

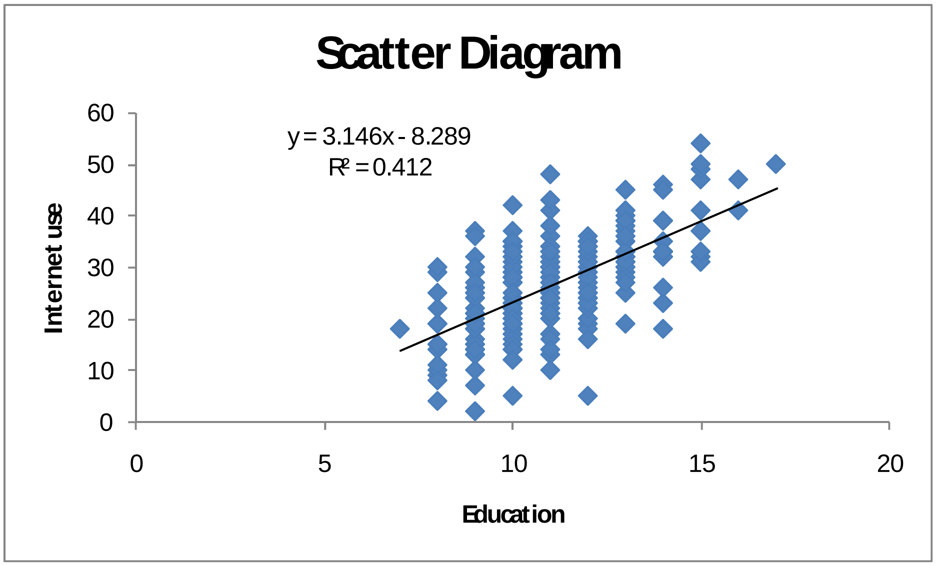
	A	B
1	*Bone Loss*	
2		
3	Mean	35.01
4	Standard Error	0.69
5	Median	36
6	Mode	38
7	Standard Deviation	7.68
8	Sample Variance	59.04
9	Kurtosis	0.08
10	Skewness	-0.19
11	Range	38
12	Minimum	15
13	Maximum	53
14	Sum	4376
15	Count	125

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.

4.91 a & b

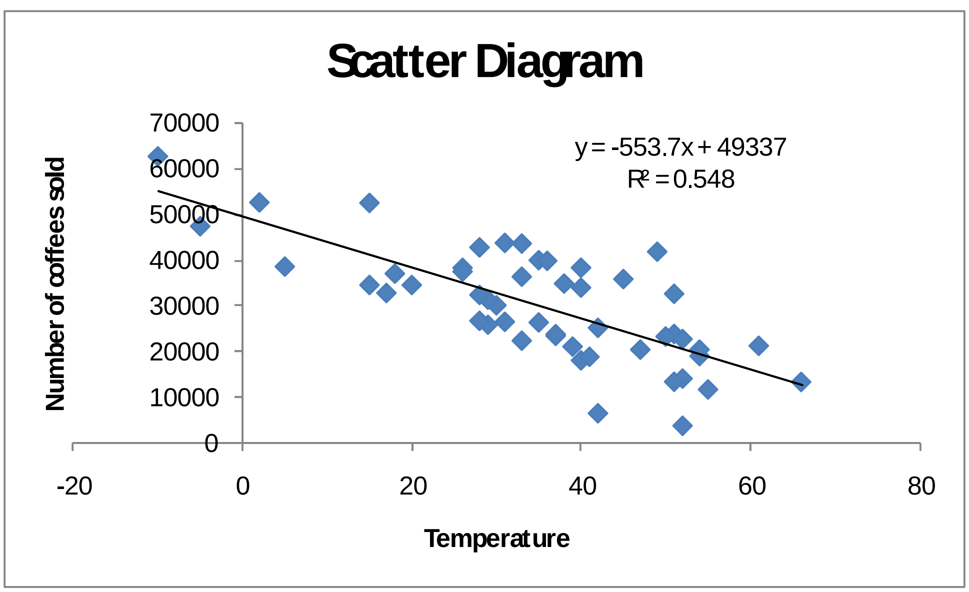


$R^2 = .412$ and the least squares line is $\hat{y} = -8.289 + 3.146x$

c 41.2% of the variation in Internet use is explained by the variation in education. For each additional year of education Internet use increases on average by 3.146 hours.

d We can measure the strength of the linear relationship accurately and the slope coefficient gives information about how education and Internet use are related.

4.92 a & b

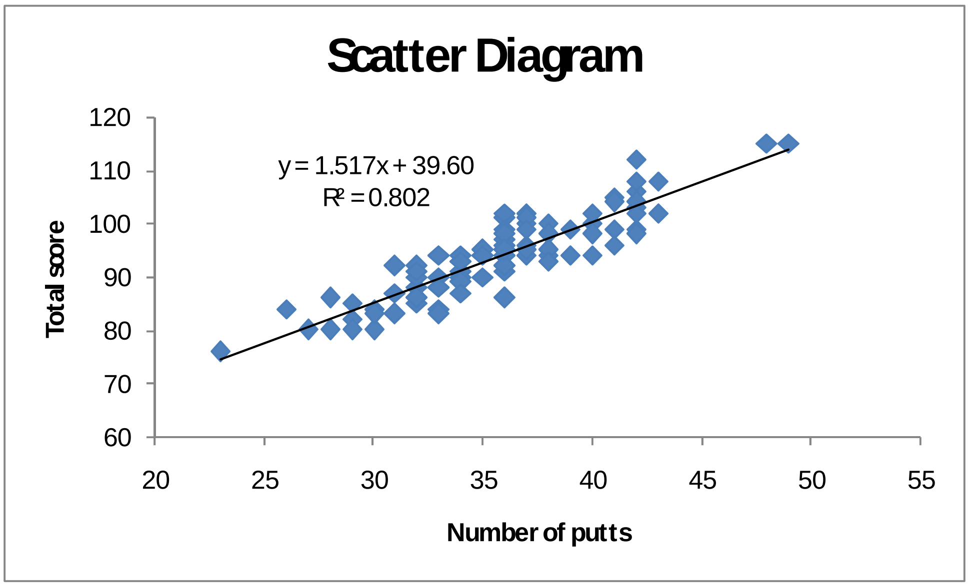


$R^2 = .548$ and the least squares line is $\hat{y} = 49{,}337 - 553.7x$

c 54.8% of the variation in the number of coffees sold is explained by the variation in temperature. For each additional degree of temperature the number of coffees sold decreases on average by 554 cups. Alternatively for each 1-degree drop in temperature the number of coffees increases on average, by 553.7 cups.

d We can measure the strength of the linear relationship accurately and the slope coefficient gives information about how temperature and the number of coffees sold are related.

4.93



80.2% of the variation in scores is explained by the variation in the number of putts.

4.94

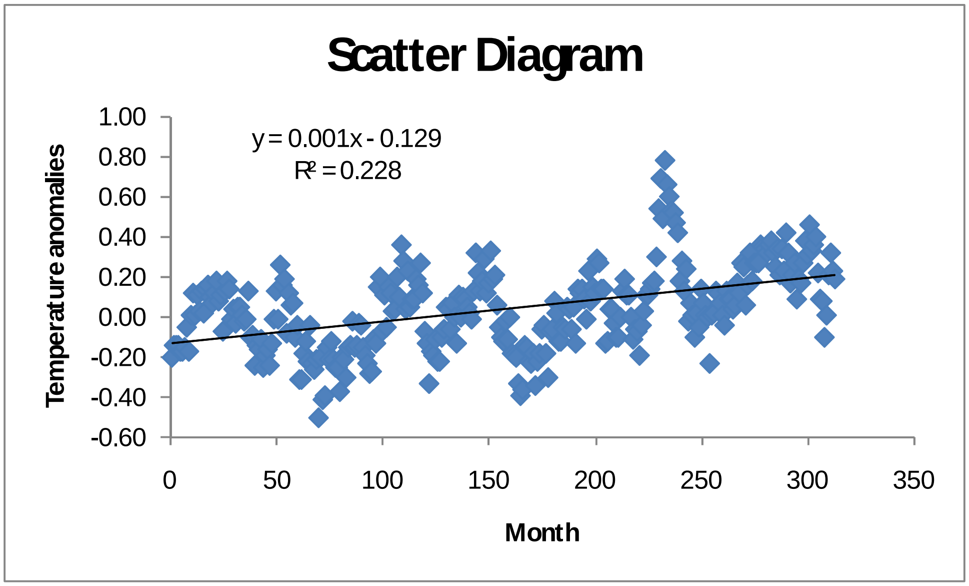
	A	B	C
1		*Bone Loss*	*Age*
2	Bone Loss	1	
3	Age	0.5742	1

R^2 r^2 $.5742^2 = .3297$; 32.97% of the variation in bone loss is explained by the variation in age.

Case 4.1

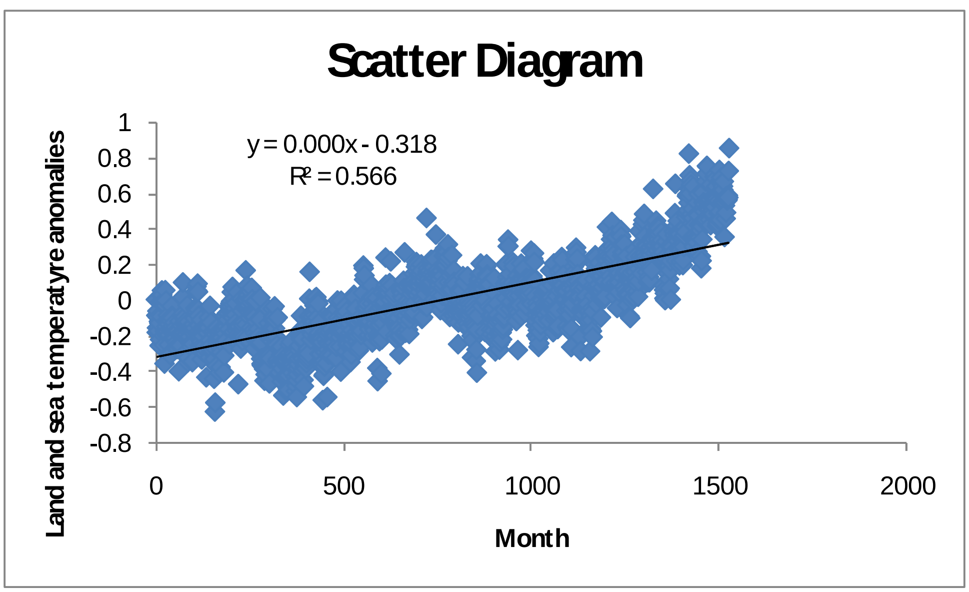
a Scatter diagrams with time as the independent variable and temperature anomalies as the dependent variable

1. Satellite-Measured Temperature Anomalies



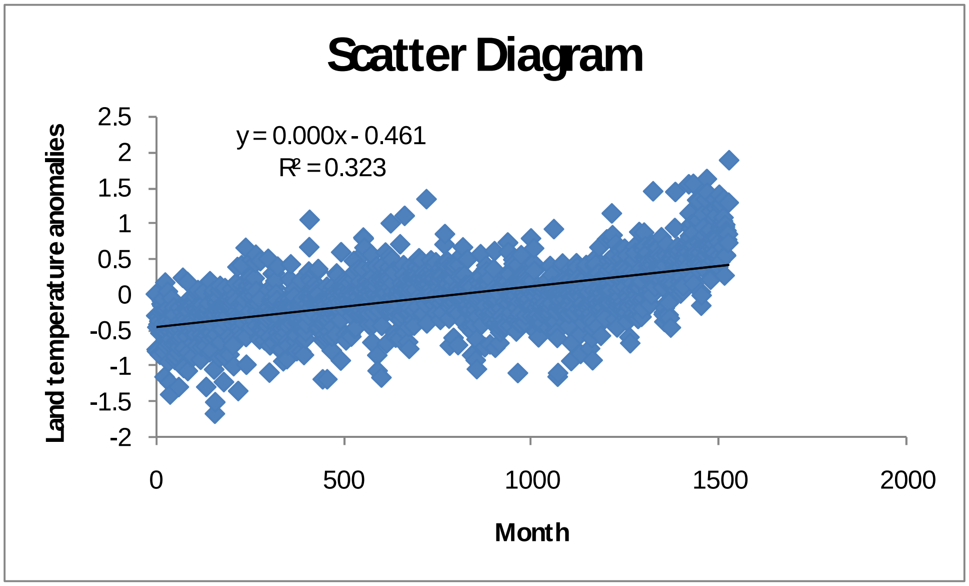
Monthly average increase is .001.

2. National Climate Data Center Land and Sea Temperature Anomalies



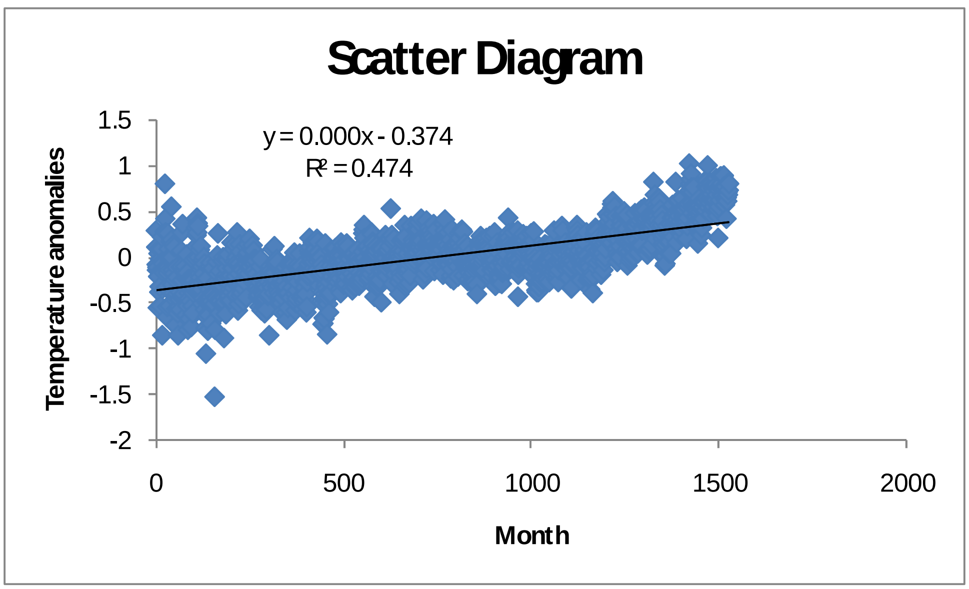
Monthly average increase is less than .001.

3. National Climate Data Center Land Temperature Anomalies



Monthly average increase is less than .001.

4. Goddard Institute for Space Studies Temperature Anomalies

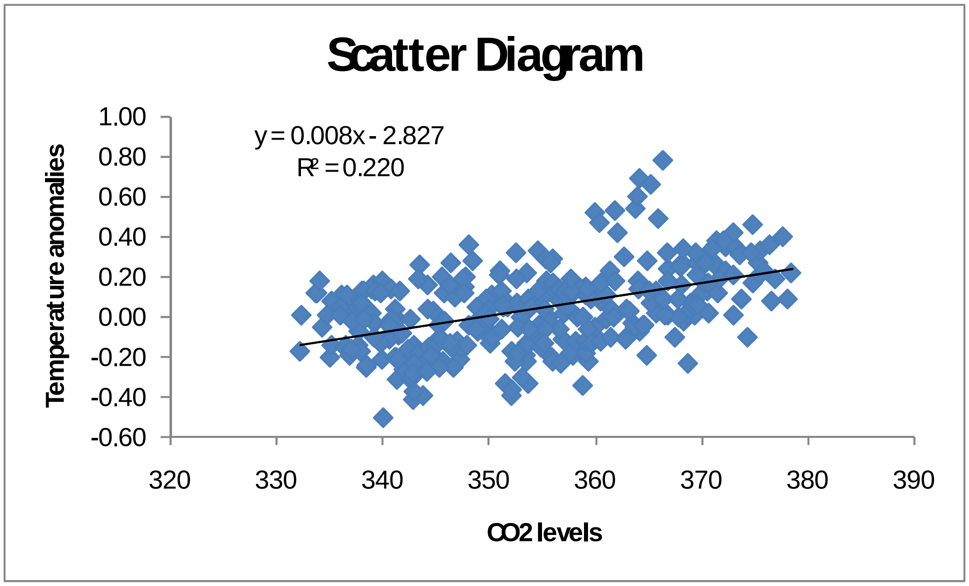


Monthly average increase is less than .001.

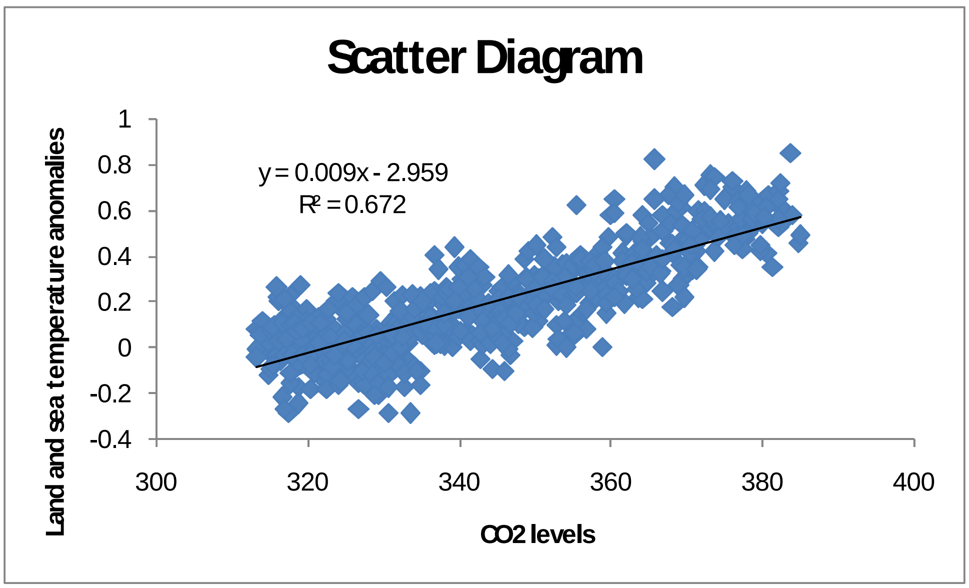
The four line charts (scatter diagrams) indicate that the temperature anomalies have increased. The largest increase is a rise from −.3 to + .3 over 127 years (scatter diagrams 1 and 2).

Scatter diagrams with carbon dioxide levels as the independent variable and temperature anomalies as the dependent variable

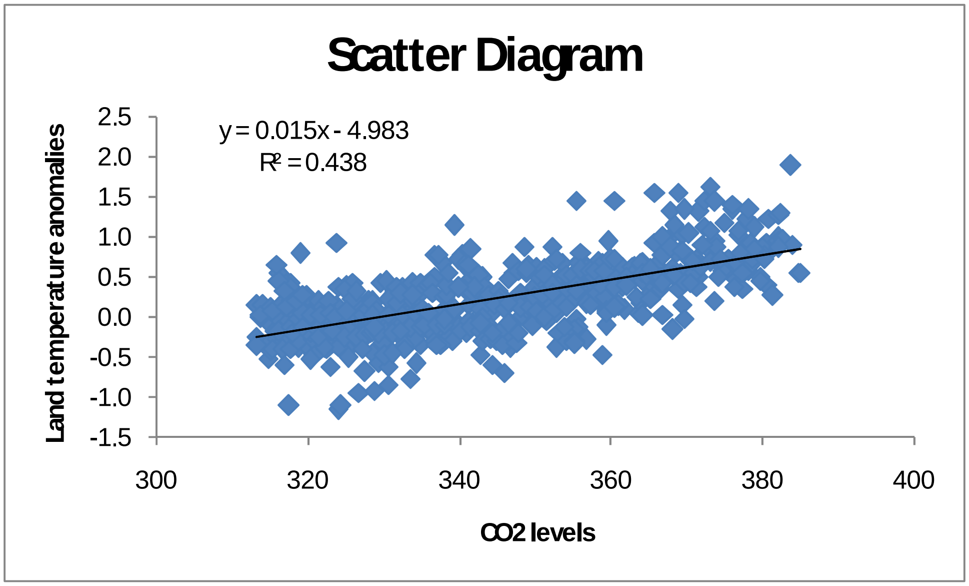
1. Satellite-Measured Temperature Anomalies and CO2 Levels



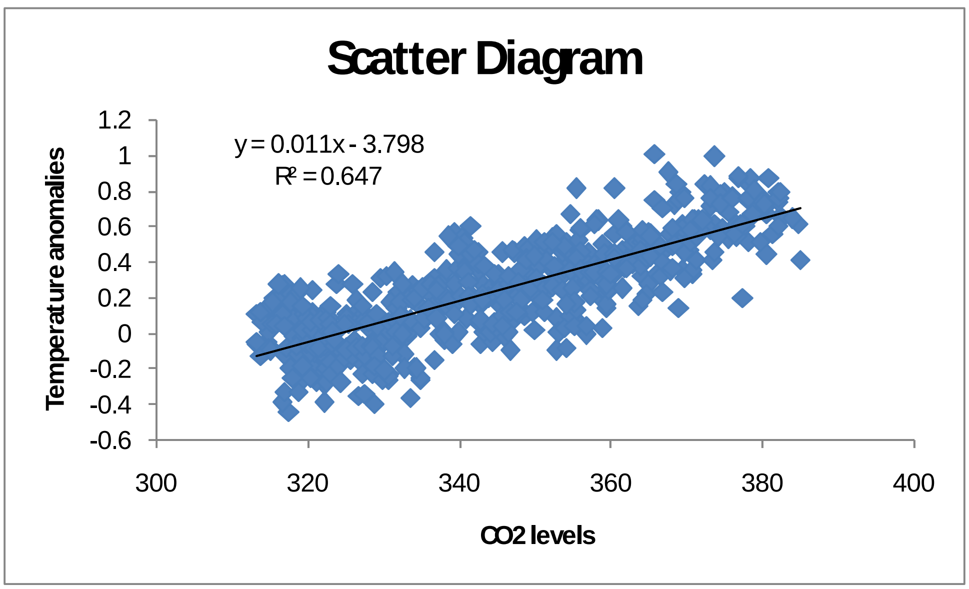
2. National Climate Data Center Land and Sea Temperature Anomalies and CO2 Levels



3. National Climate Data Center Land Temperature Anomalies and CO2 Levels



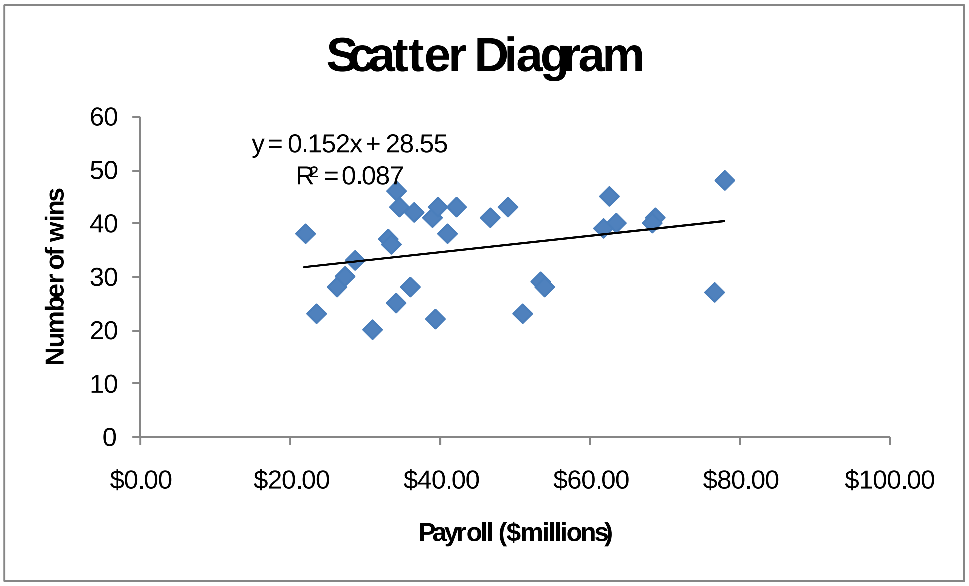
4. Goddard Institute for Space Studies Temperature Anomalies and CO2 Levels



The coefficients of determination range from .220 to .672. There appears to be some linear relationship but it is far from conclusive that CO_2 is the cause of temperature increases.

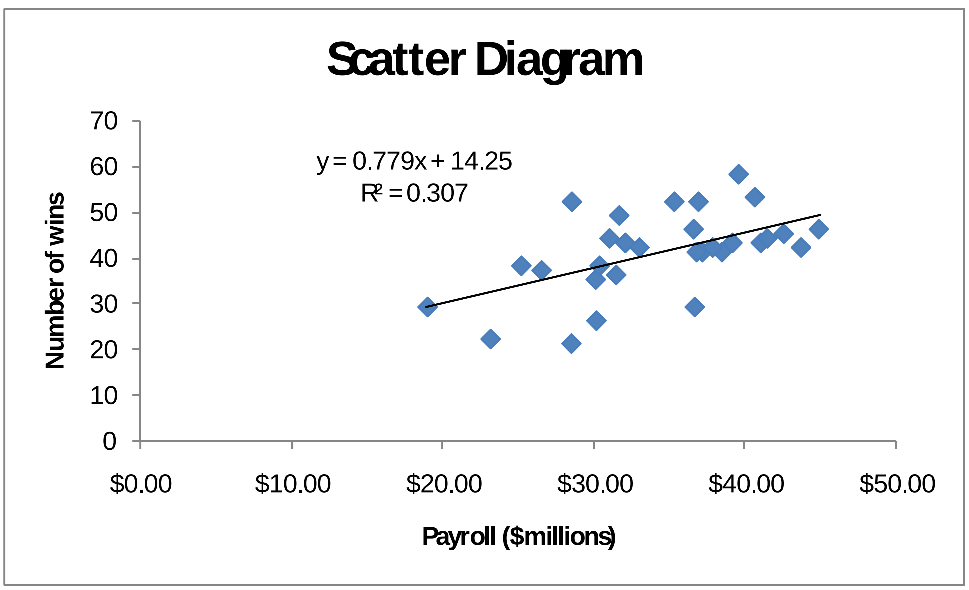
Case 4.2

2003-04 Season



The cost of winning one additional game is 1million/.152 = \$6.58 million. However, the coefficient of determination is only .087, which tells us that there are many other variables that determine how well a team will do.

2005-06 Season



The cost of winning one additional game is 1million/.779 = \$1.28 million. The coefficient of determination is .307. The small coefficient of determination in the year before the strike seems to indicate that team owners were spending large amounts of money and getting little in return. The results are markedly different in the year after the strike. There is a much stronger linear relationship between payroll and the number of wins and the cost of winning one additional game is considerably smaller.

Case 4.3

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

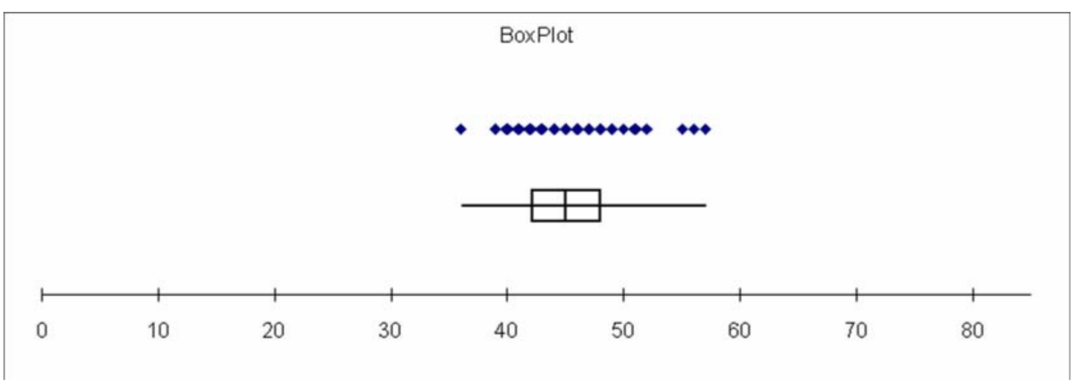
Education

	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

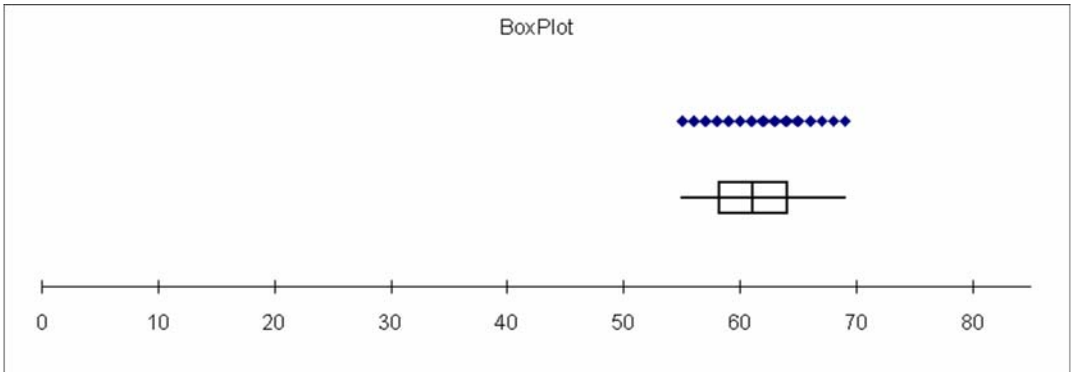
b

Ages

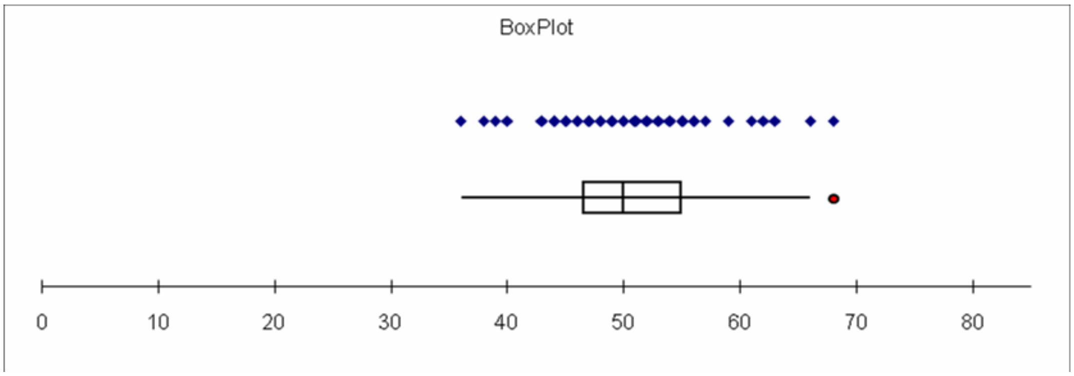
BMW



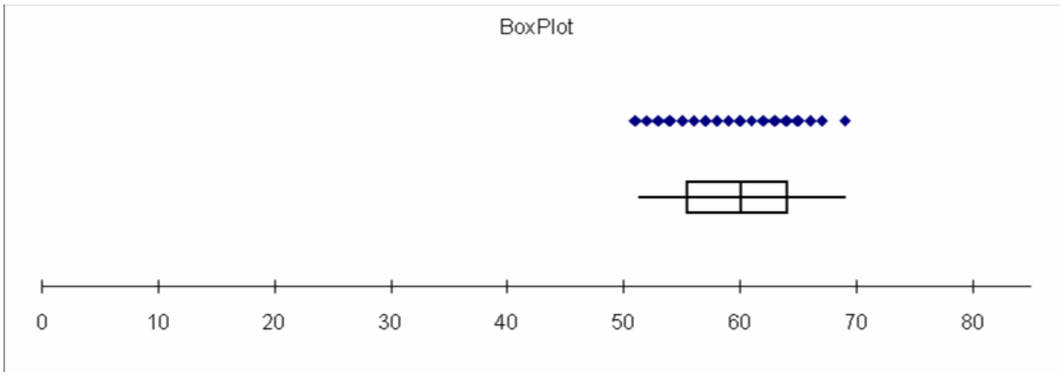
Cadillac



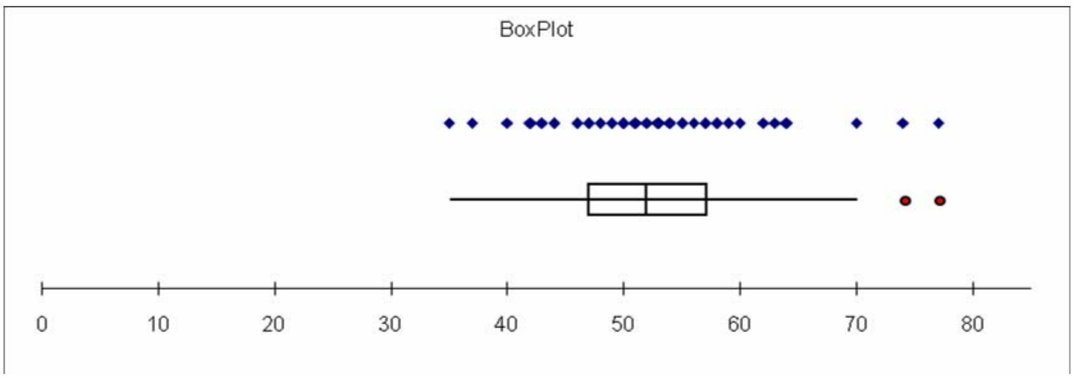
Lexus



Lincoln

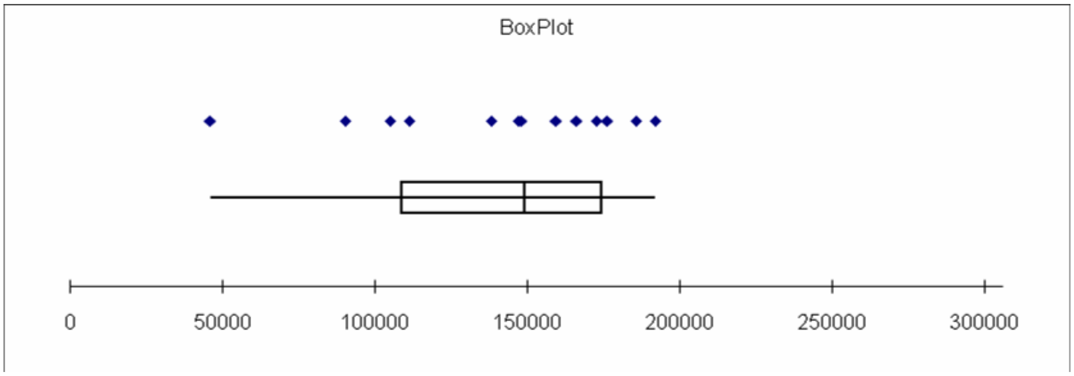


Mercedes

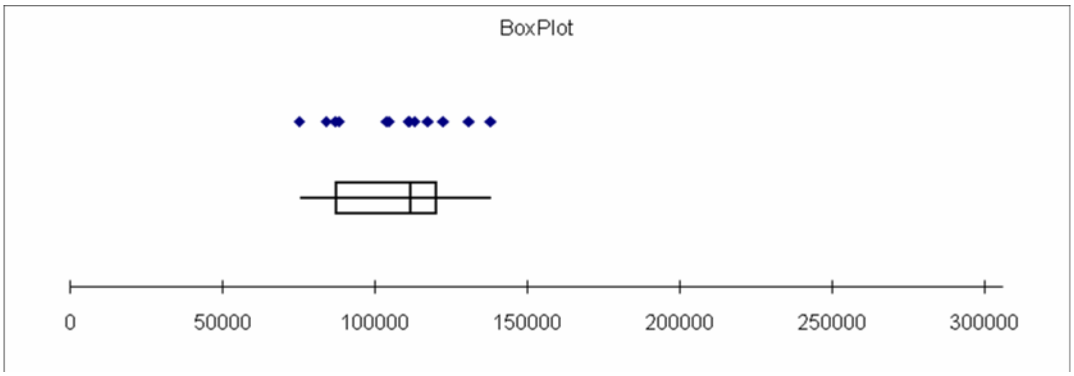


Income

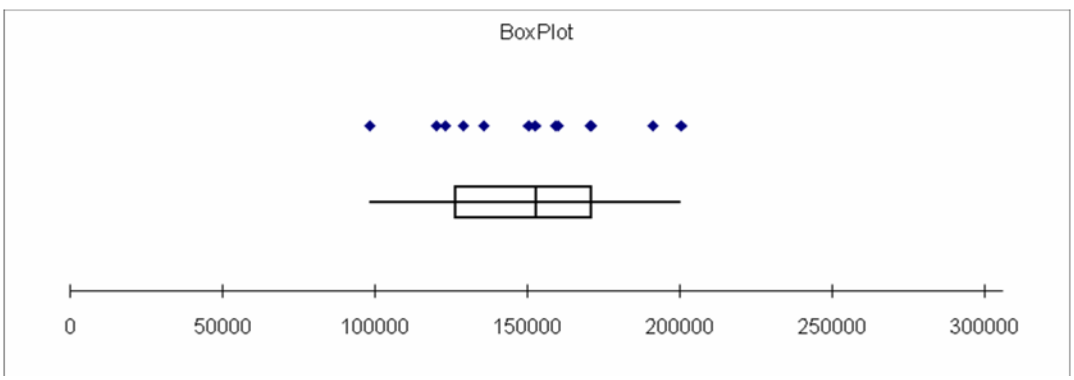
BMW



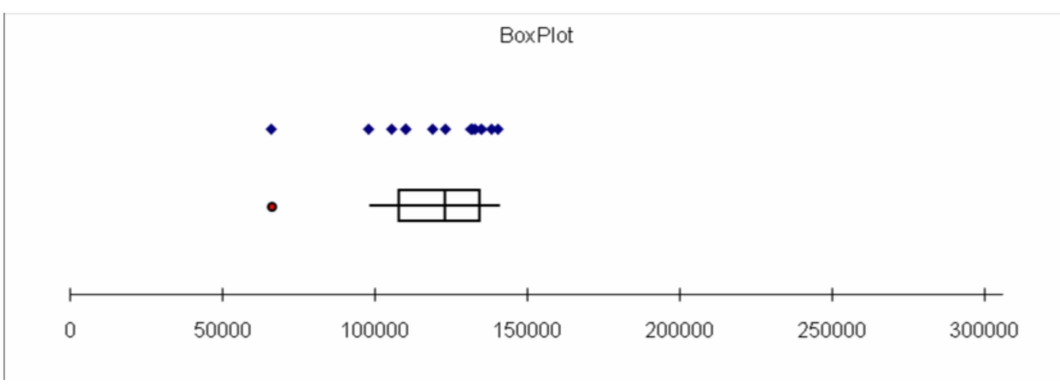
Cadillac



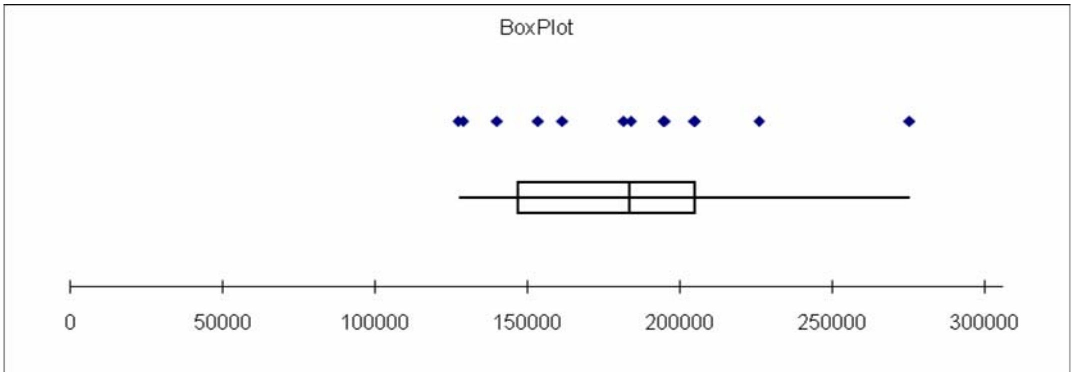
Lexus



Lincoln

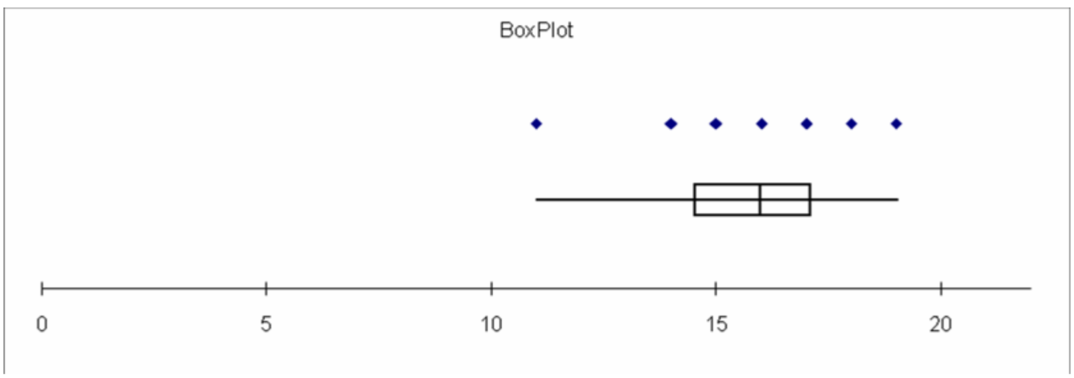


Mercedes

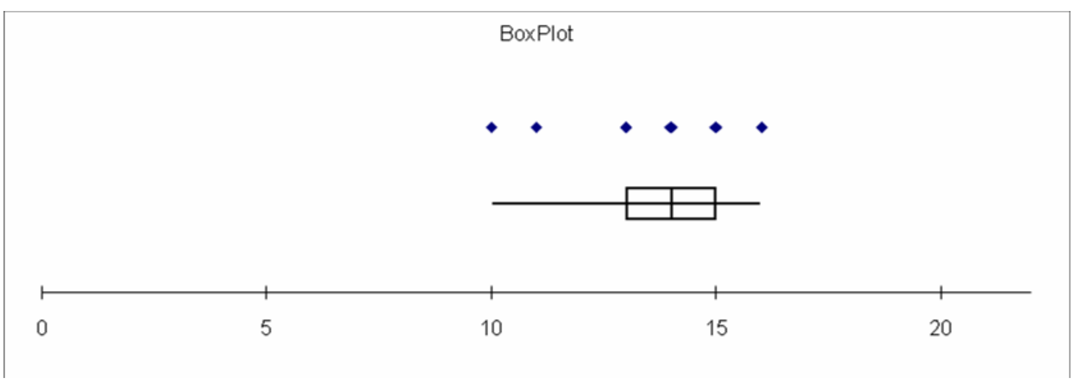


Education

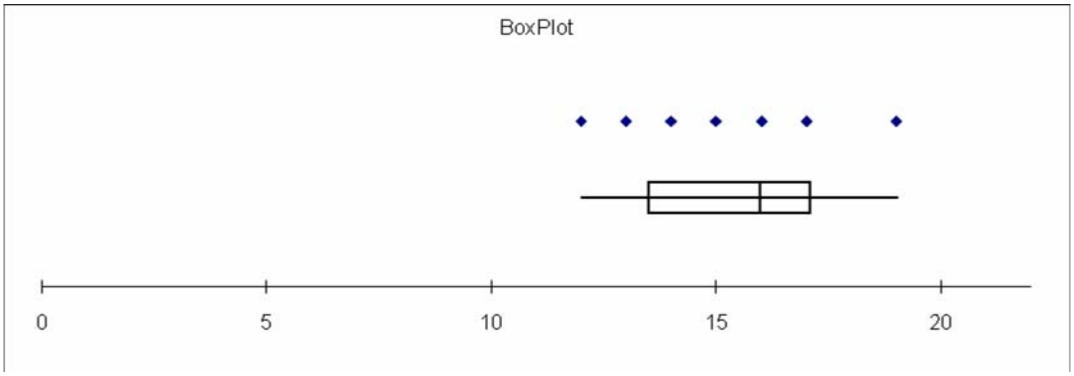
BMW



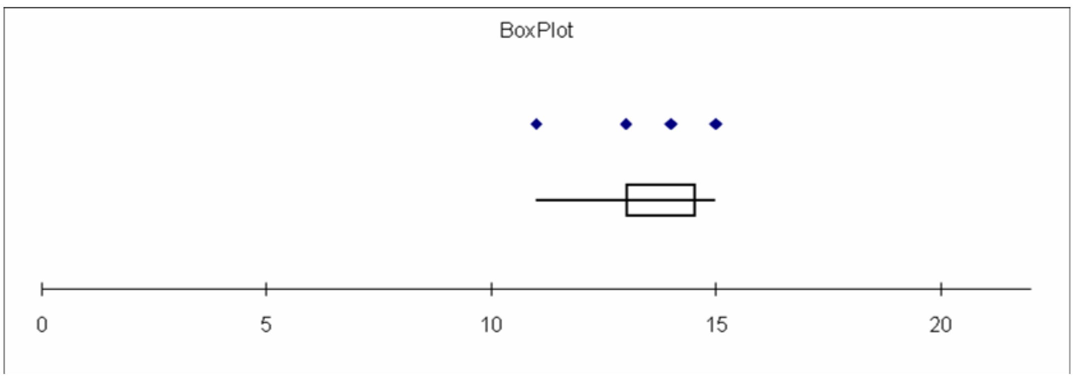
Cadillac



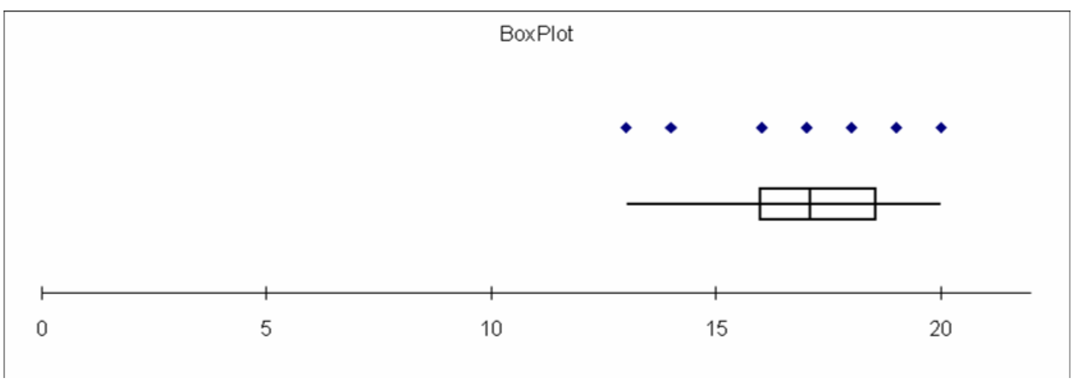
Lexus



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 4.4

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

The coefficient of determination is $(-.1787)^2 = .0319$. 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

The coefficient of determination is $(.3600)^2 = .1296$. 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

The coefficient of determination is $(.0678)^2 = .0046$. 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.