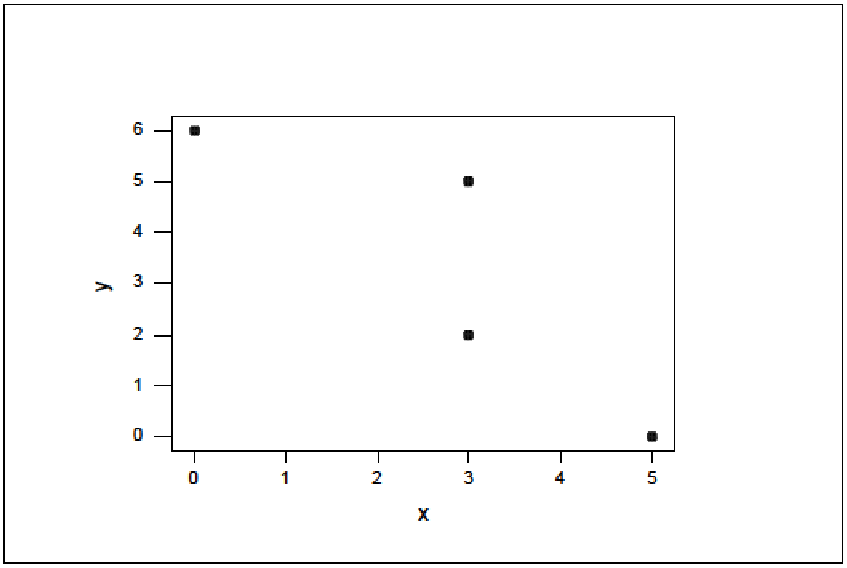
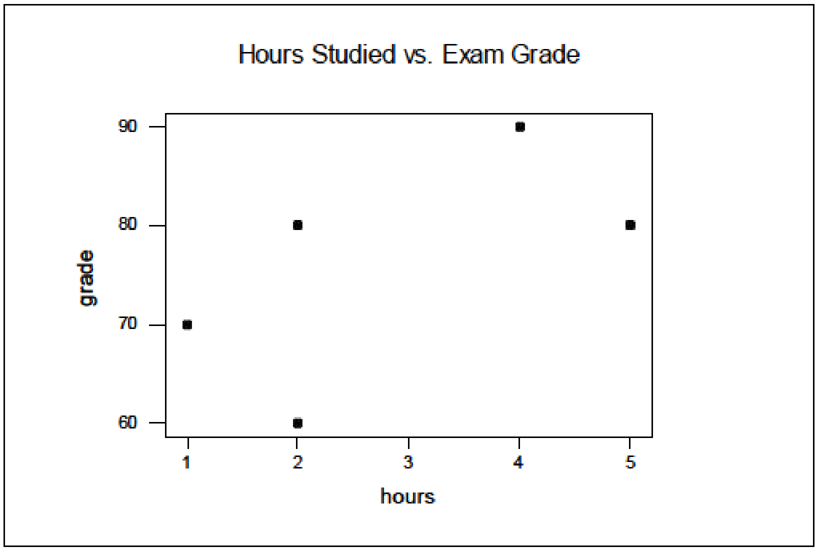
3.5 As the x's are increasing, the y's are decreasing.

3.6
a. age, height,
b. Age = 3 yrs., height = 87 cm.
c. Answers will vary – whether a child's growth is above or below normal, etc.

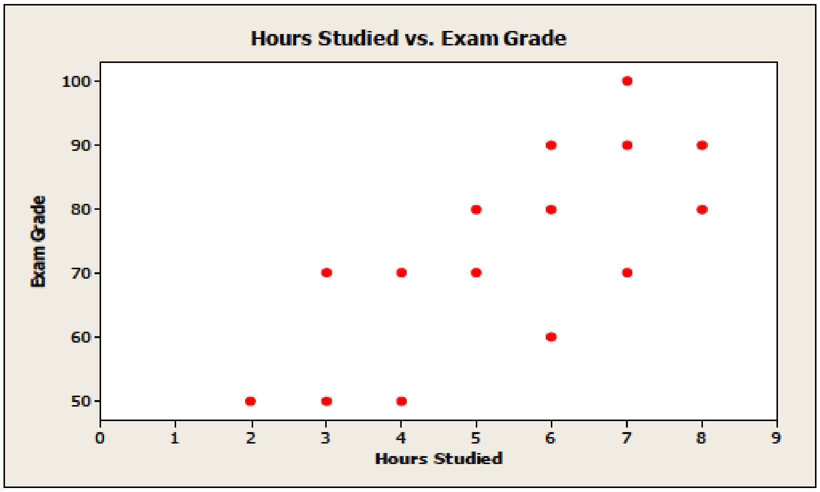


3.7 a.

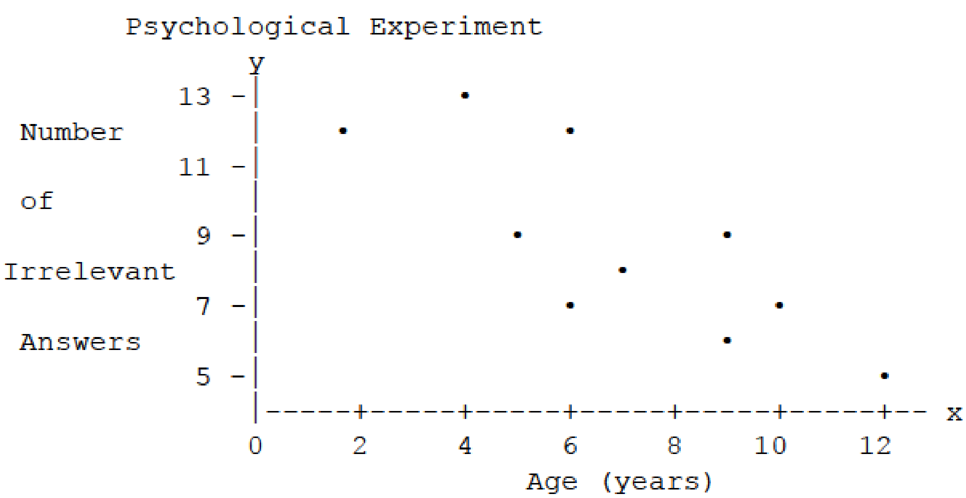


b. As hours studied increased, there seems to be a trend for the exam grades to also increase.

3.8



3.9



3.12 Coefficient values near zero indicate that there is very little or no linear correlation.

3.13 Impossible. The correlation coefficient must be a numerical value between -1 and +1. There must be a calculation or typographical error.

3.14 The scatter diagram suggests a non-linear relationship between the two variables. The correlation coefficient measures the strength of a linear relationship, therefore a value near zero indicates no linear relationship.

Calculating r - the linear correlation coefficient

Preliminary Calculations:

1. Set up a table with the column headings: x, y, x^2, xy and y^2.
2. Insert the bivariate data into corresponding x and y columns. Perform the various algebraic functions to fill in the remaining columns.
3. Sum all columns, that is, find $\sum x$, $\sum y$, $\sum x^2$, $\sum xy$, $\sum y^2$.
4. Double check calculations and summations.
5. Calculate: SS(x) - the sum of squares of x
   SS(y) - the sum of squares of y
   SS(xy) - the sum of squares of xy
   where:
   $SS(x) = \sum x^2 - ((\sum x)^2/n)$
   $SS(y) = \sum y^2 - ((\sum y)^2/n)$
   $SS(xy) = \sum xy - ((\sum x \cdot \sum y)/n)$

Final Calculation:

6. Calculate r:
   $$r = \frac{SS(xy)}{\sqrt{SS(x)SS(y)}}$$ (round to the nearest hundredth)
7. Retain the summations and the sums of squares, as they will be needed for later calculations.

NOTE: Remember $SS(x) \neq \sum x^2$, $SS(y) \neq \sum y^2$ and $SS(xy) \neq \sum xy$.

The computer and/or calculator command to calculate the correlation coefficient can be found on the Chapter 2 Tech card.

3.15 a.

x	y	x^2	xy	y^2
42	303	1764	12726	91809
7	212	49	1484	44944
75	401	5625	30075	160801
78	500	6084	39000	250000
126	536	15876	67536	287296
22	200	484	4400	40000
23	278	529	6394	77284
373	2430	30411	161615	952134

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 30411 - (373^2/7) = \underline{10535.429}$

$SS(y) = \sum y^2 - ((\sum y)^2/n) = 952134 - (2430^2/7) = \underline{108576.857}$

$SS(xy) = \sum xy - ((\sum x \cdot \sum y)/n) = 161615 - (373 \cdot 2430/7) = \underline{32130.714}$

b. $r = SS(xy)/\sqrt{SS(x) \cdot SS(y)} = 32130.174/\sqrt{(10535.429)(108576.857)} = 0.94998 = \underline{0.95}$

3.18 a. Estimate r to be near 2/3 or 0.7.

b.

Data	x	y	x^2	xy	y^2
1	2	5	4	10	25
2	3	5	9	15	25
3	3	7	9	21	49
4	4	5	16	20	25
5	4	7	16	28	49
6	5	7	25	35	49
7	5	8	25	40	64
8	6	6	36	36	36
9	6	9	36	54	81
10	6	8	36	48	64
11	7	7	49	49	49
12	7	9	49	63	81
13	7	10	49	70	100
14	8	8	64	64	64
15	8	9	64	72	81
$\sum$	81	110	487	625	842

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 487 - (81^2/15) = 49.6$

$SS(y) = \sum y^2 - ((\sum y)^2/n) = 842 - (110^2/15) = 35.333$

$SS(xy) = \sum xy - ((\sum x \cdot \sum y)/n) = 625 - (81 \cdot 110/15) = 31.0$

$r = SS(xy)/\sqrt{SS(x) \cdot SS(y)} = 31.0/\sqrt{49.6 \cdot 35.333} = \underline{0.741} = \underline{0.74}$

3.19 a. Estimate r to be near -0.75

b.

Data	x	y	x^2	xy	y^2
1	2	12	4	24	144
2	4	13	16	52	169
3	5	9	25	45	81
4	6	7	36	42	49
5	6	12	36	72	144
6	7	8	49	56	64
7	9	6	81	54	36
8	9	9	81	81	81
9	10	7	100	70	49
10	12	5	144	60	25
$\sum$	70	88	572	556	842

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 572 - (70^2/10) = 82.0$

$SS(y) = \sum y^2 - ((\sum y)^2/n) = 842 - (88^2/10) = 67.6$

$SS(xy) = \sum xy - ((\sum x \cdot \sum y)/n) = 556 - (70 \cdot 88/10) = -60.0$

$r = SS(xy)/\sqrt{SS(x) \cdot SS(y)} = -60.0/\sqrt{82.0 \cdot 67.6} = \underline{-0.806} = \underline{-0.81}$

3.20

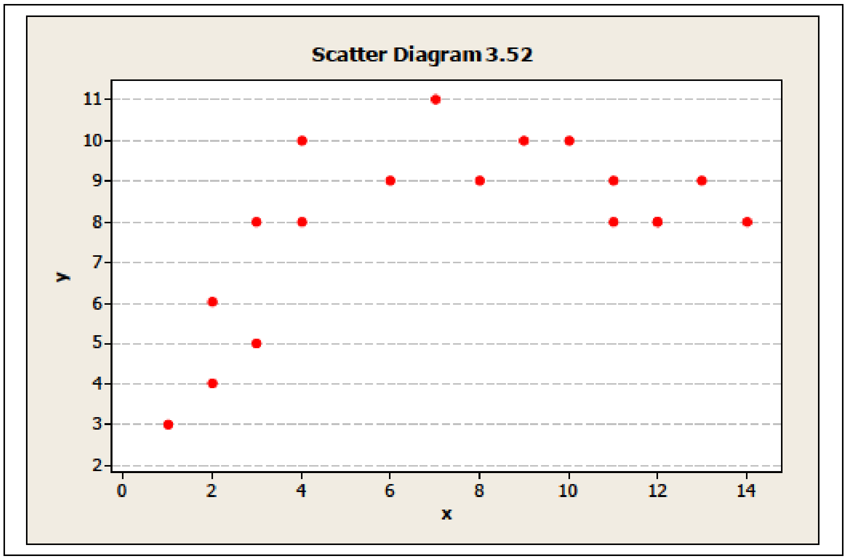
a. 0.95

b. 1.00

c. answers will vary

d. yes, CO_2 will double

3.23



No. The pattern shown is definitely not that of a straight line, thus the results from using any of the linear techniques would be meaningless.

CALCULATING $\hat{y} = b_0 + b_1 x$ - THE EQUATION OF THE LINE OF BEST FIT

1. Retrieve preliminary calculations from previous *r* calculations.

2. Calculate b_1 where $b_1 = \frac{SS(xy)}{SS(x)}$

3. Calculate b_0 where $b_0 = \frac{1}{n}(\Sigma y - b_1 \Sigma x)$

$\hat{y}$ = predicted value of y (based on the regression line)

NOTE: See Review Lessons in Appendix A for additional information about the concepts of slope and intercept of a straight line.

DRAWING THE LINE OF BEST FIT ON THE SCATTER DIAGRAM

1. Pick two *x*-values that are within the interval of the data *x*-values. (one value near either end of the domain)
2. Substitute these values into the calculated $\hat{y} = b_0 + b_1 x$ equation and find the corresponding $\hat{y}$ values.
3. Plot these points on the scatter diagram in such a manner that they are distinguishable from the actual data points.
4. Draw a straight line connecting these two points. This line is a graph of the line of best fit.
5. Plot a third point, the ordered pair $(\overline{x}, \overline{y})$ as an additional check. It should be a point on the line of best fit.

OR:

Computer and/or calculator commands to find the equation of the line of best fit and also draw it on a scatter diagram can be found on your Chapter 3 Tech card.

3.24 a. Summations from extensions table: $n = 5$, $\sum x = 14$, $\sum y = 380$, $\sum x^2 = 50$, $\sum xy = 1110$, $\sum y^2 =$ 29,400

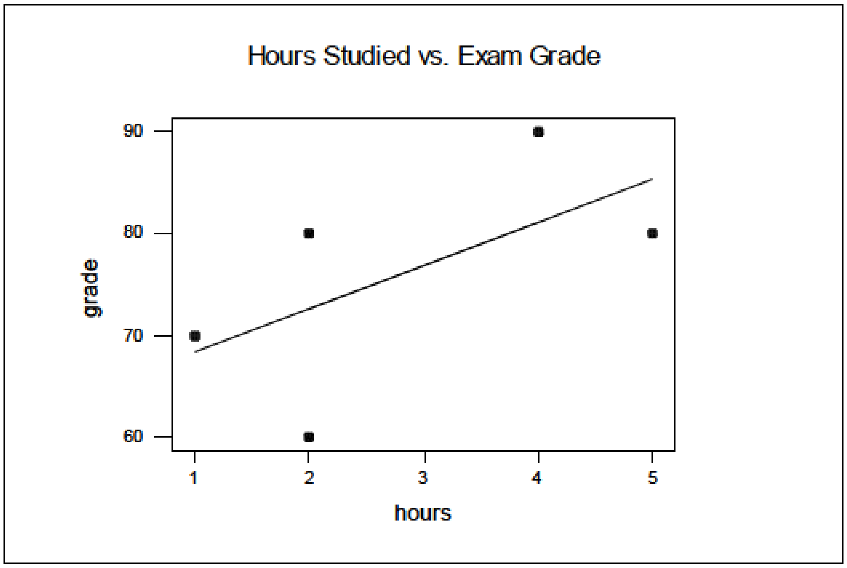
$SS(x) = \sum x^2 - ((\sum x)^2/n) = 50 - (14^2/5) = 10.8$
$SS(xy) = \sum xy - ((\sum x \cdot \sum y)/n) = 1110 - (14 \cdot 380/5) = 46$

$b_1 = SS(xy)/SS(x) = 46/10.8 = 4.259$

$b_0 = [\sum y - b_1 \cdot \sum x]/n = [380 - (4.259 \cdot 14)]/5 = 64.0748$

$\hat{y}$ = <u>64.1 + 4.26x</u>

b. At $x = 1$, $\hat{y} = 64.1 + 4.26(1) = 68.36$; thus (1,68.4)
At $x = 3$, $\hat{y} = 64.1 + 4.26(3) = 76.9$; thus (3,76.9)
Points (1,68.4) and (3,76.9) are used to locate the line.



c. Yes, as the hours studied increased, the exam grades appear to increase, also.

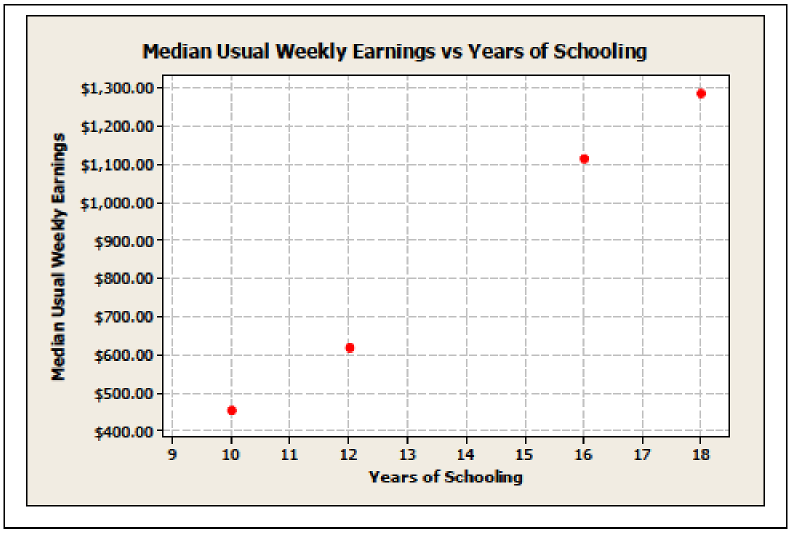
3.25 a. The average value is $34,030 for a used SUV that is virtually brand new

b. SUV value decreases $3040 for each additional year in age

3.27 $\hat{y} = 7.31 - 0.01x$ when $x = 50$ is
$\hat{y} = 7.31 - 0.01(50) = 6.81$

The predicted value is 6.81(10,000) or $<u>68,100</u>

3.28 a.



b. Yes, as years of schooling increase, so do median weekly earnings.

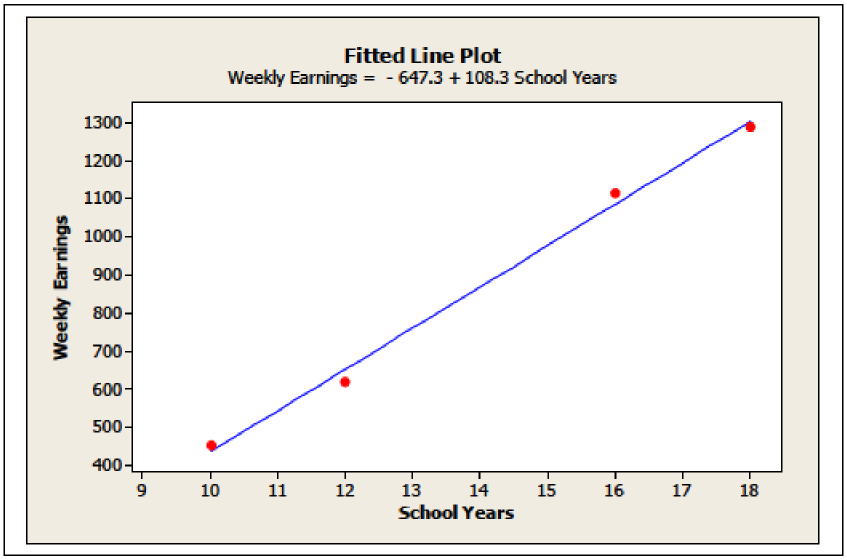
c. 0.997

d. yes

e. $\hat{y}$ = -647.25 + 108.25x

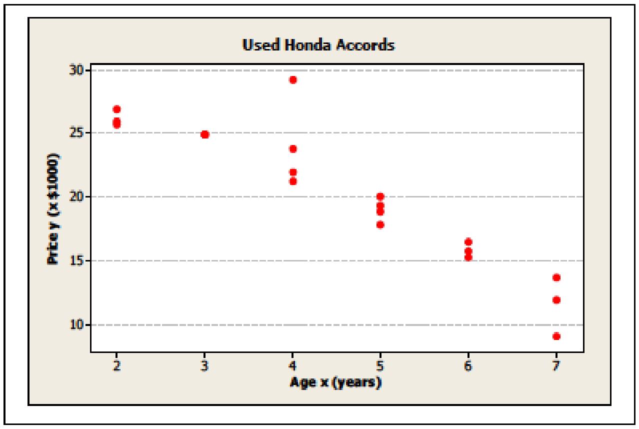
f. For every additional year of schooling, median weekly earnings increase by $108.25.

g.



h. -647.25,x = 0 years of schooling is not in the range of data

3.29 a.



b. Summations from extensions table: $n = 20$, $\sum x = 90$, $\sum y = 407$, $\sum x^2 = 458$, $\sum xy = 1670.3$

$SS(x) = \sum x^2 - ((\sum x)^2/n) = 458 - (90^2/20) = 53$

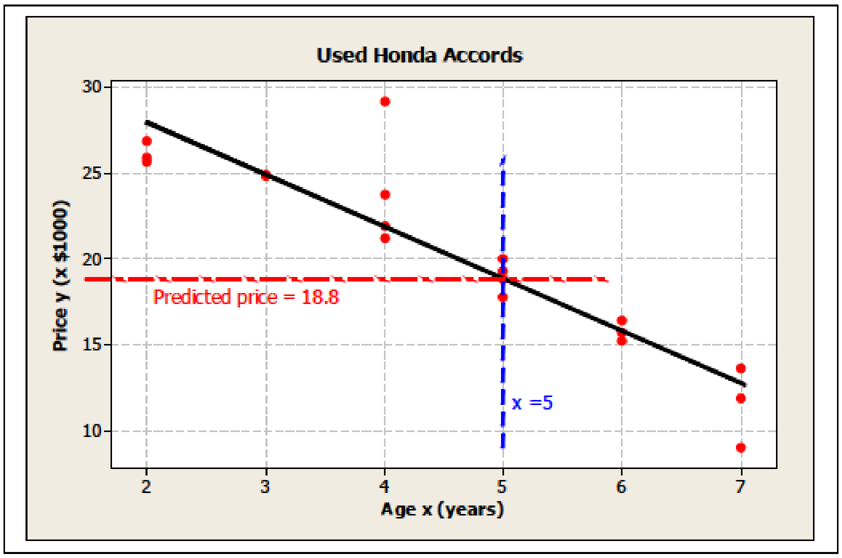
$SS(xy) = \sum xy - ((\sum x \cdot \sum y)/n) = 1670.3 - (90 \cdot 407/20) = -161.2$

$b_1 = SS(xy)/SS(x) = -161.2/53 = -3.04$

$b_0 = [\sum y - b_1 \cdot \sum x]/n = [407 - (-3.04 \cdot 90)]/20 = 34.03$

$\hat{y} = 34.03 - 3.04x$

c.



d. 1) Intersection of vertical line for x = 5 and regression line on above graph gives 18.8($1000 or 18,800) for y.
2) price = 34.03 – 3.04 (5) = 18.83($1000) or $18,830

e. Possible lurking variables include: condition of the car, mileage of car, potential buyers desire to buy car, and many others. Some of these variables would tend to raise the price while others would tend to lower the price.