Name _____ Date _____

Like a Glove
Least Squares Regression

Vocabulary

Write a definition for each term.

1. least squares regression line

2. interpolation

3. extrapolation

Problem Set

Determine the least squares regression line for each set of points. Round your answer to the nearest hundredth.

1. (3, 4), (7, 6) and (−2, −4)

$n = 3$

$\Sigma x = 3 + 7 + (-2)$
$= 8$

$\Sigma y = 4 + 6 + (-4)$
$= 6$

$\Sigma x^2 = 3^2 + 7^2 + (-2)^2$
$= 9 + 49 + 4$
$= 62$

$\Sigma xy = (3 \cdot 4) + (7 \cdot 6) + (-2 \cdot -4)$
$= 12 + 42 + 8$
$= 62$

$(\Sigma x)^2 = 8^2$
$= 64$

$$a = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{n\Sigma x^2 - (\Sigma x)^2}$$
$$= \frac{(3)(62) - (8)(6)}{(3)(62) - (64)}$$
$$= \frac{186 - 48}{186 - 64} = \frac{138}{122}$$
$$a \approx 1.13$$

$$b = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n\Sigma x^2 - (\Sigma x)^2}$$
$$= \frac{(6)(62) - (8)(62)}{(3)(62) - (64)}$$
$$= \frac{372 - 496}{186 - 64} = \frac{-124}{122}$$
$$b \approx -1.02$$

The least squares regression line for the points is $y = 1.13x - 1.02$.

2. (−7, 1), (3, 8) and (9, 7)

Name _____ Date _____

3. (−3, 6), (−2, −1) and (6, −4)

4. (−8, 7), (−5, 3), (3, 6) and (9, 0)

5. (−7, −1), (−5, −9), (3, 3) and (6, 9)

6. (−8, 6), (−8, −2), (−6, −9) and (−5, −4)

Name _____ Date _____

While in high school, Clayton started his own T-shirt printing business. The table shows the number of T-shirts Clayton has sold each year since starting his business in 2006.

Year	2006	2007	2008	2009	2010	2011	2012
Number of T-shirts	50	75	175	125	250	350	375

The linear regression equation representing the data shown in the table is $y = 57.14x + 28.57$, where x represents the number of years since 2006 and y represents the number of T-shirts sold. Use the regression equation to predict the number of T-shirts Clayton sold during each given year. Then compare the prediction to the actual number of T-shirts or determine if the prediction is reasonable based on the problem situation.

7. 2008

For 2008, $x = 2$.

$y = 57.14x + 28.57$

$y = 57.14(2) + 28.57$

$y = 114.28 + 28.57$

$y = 142.85$

The total number of T-shirts sold in 2008 should be about 143. The actual number of T-shirts sold was 175, so the predicted value is fairly close to the actual value.

8. 2010

9. 2012

10. 2014

11. 2020

12. 2000

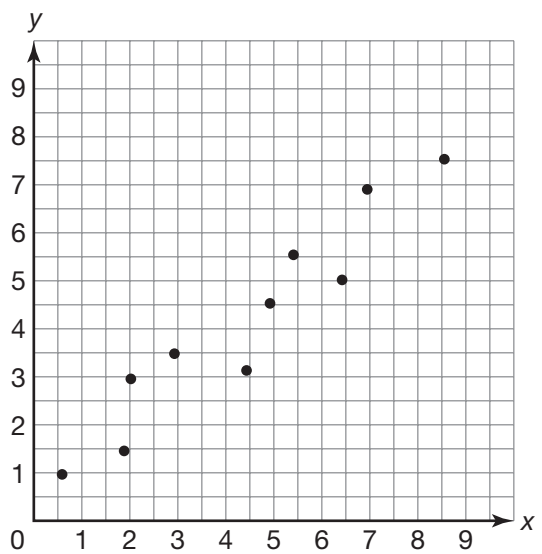
Name _____ Date _____

Gotta Keep It Correlatin'
Correlation

Problem Set

Determine whether the points in each scatter plot have a positive correlation, a negative correlation, or no correlation. Then determine which *r*-value is most accurate.

1.



A $r = 0.8$

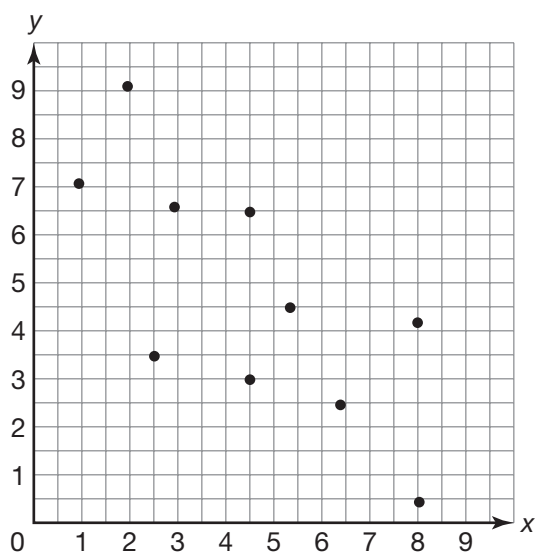
B $r = -0.8$

C $r = 0.08$

D $r = -0.08$

These data have a positive correlation. Because of this the *r*-value must be positive. Also, the data are fairly close to forming a straight line, so $r = 0.8$ (A) would be the most accurate.

2.



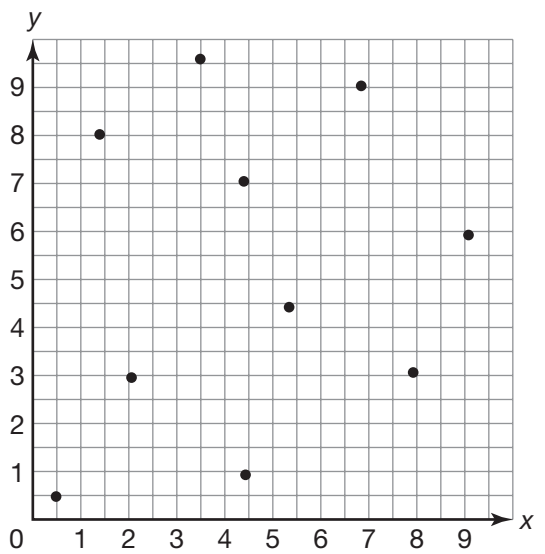
A $r = 0.9$

B $r = -0.6$

C $r = 0.02$

D $r = -0.006$

3.



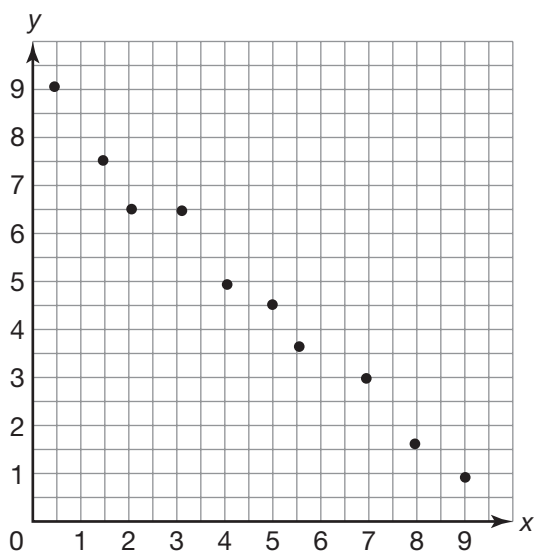
A $r = 0.01$

B $r = 0.8$

C $r = -0.5$

D $r = 0.5$

4.



A $r = -0.009$

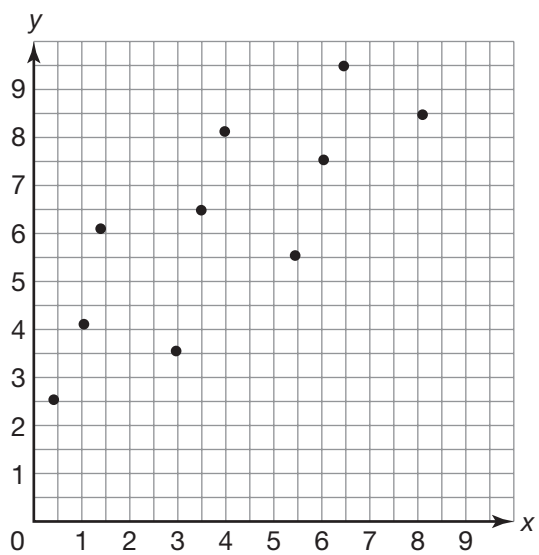
B $r = 0.8$

C $r = -0.9$

D $r = 0.2$

Name _____ Date _____

5.



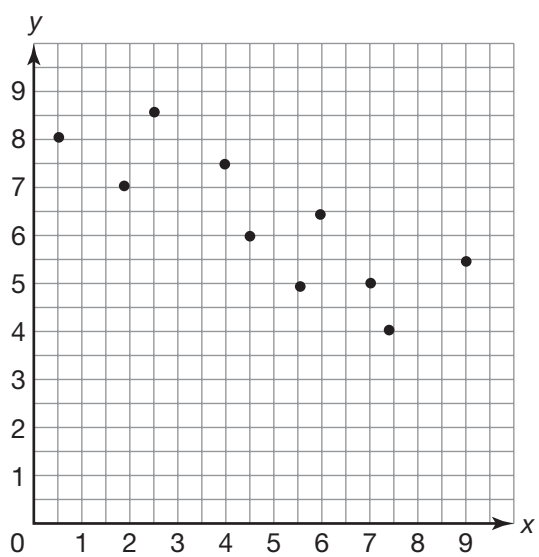
A $r = -0.003$

B $r = -0.6$

C $r = 0.004$

D $r = 0.7$

6.



A $r = 0.01$

B $r = -0.8$

C $r = -0.01$

D $r = 0.9$

Determine the correlation coefficient of each data set. Round your answer to the nearest ten thousandth.

7. (3, 2), (5, 7) and (10, 9)

$\bar{x} = \frac{3 + 5 + 10}{3}$

$= 6$

$\bar{y} = \frac{2 + 7 + 9}{3}$

$= 6$

$(x_i - \bar{x})$

$3 - 6 = -3$

$5 - 6 = -1$

$10 - 6 = 4$

$(y_i - \bar{y})$

$2 - 6 = -4$

$7 - 6 = 1$

$9 - 6 = 3$

$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$-3 \cdot -4 = 12$
$-1 \cdot 1 = -1$
$4 \cdot 3 = 12$
$\Big\} 12 + (-1) + (12) = 23$

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

$(-3)^2 = 9$
$(-1)^2 = 1$
$(4)^2 = 16$
$\Big\} 9 + 1 + 16 = 26$

$\sum_{i=1}^{n}(y_i - \bar{y})^2$

$(-4)^2 = 16$
$(1)^2 = 1$
$(3)^2 = 9$
$\Big\} 16 + 1 + 9 = 26$

$\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \sqrt{26} \cdot \sqrt{26}$

$= 26$

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

$= \frac{23}{26}$

≈ 0.8846

The correlation coefficient of this data set is 0.8846.

Name _____ Date _____

8. (2, 10), (3, 3) and (10, 5)

9. (2, 2), (5, 3) and (7, 6)

Name _____ Date _____

10. (5, 6), (7, 4) and (8, 2)

11. (2, 8), (3, 5) and (6, 6)

Name _____ Date _____

12. (4, 8), (6, 11) and (8, 15)

Determine the linear regression equation and correlation coefficient for each data set. State if the linear regression equation is appropriate for the data set. Round your answer to the nearest ten thousandth.

13.

Year	2007	2008	2009	2010	2011	2012
Profit (dollars)	50,000	75,000	150,000	125,000	195,000	225,000

x = years since 2007

$y = 34{,}571.4286x + 50{,}238.0952$

$r = 0.9571$

Because the r-value is close to 1, the linear regression equation is appropriate for the data set.

14.

Year	2007	2008	2009	2010	2011	2012
Profit (dollars)	100,000	85,000	91,000	82,000	79,500	74,000

15.

Time (seconds)	0	1	2	3	4	5
Height (feet)	5	21	34	31	18	3

Name _____ Date _____

16.

Time (seconds)	0	1	2	3	4	5
Height (feet)	63	56	42	36	28	12

17.

Year	2007	2008	2009	2010	2011	2012
Units Sold	1480	14,105	8925	18,750	5250	2650

18.

Year	2007	2008	2009	2010	2011	2012
Units Sold	5245	7840	7075	9130	10,620	12,635

9

Name _____ Date _____

The Residual Effect
Creating Residual Plots

Vocabulary

Write a definition for each term.

1. residual

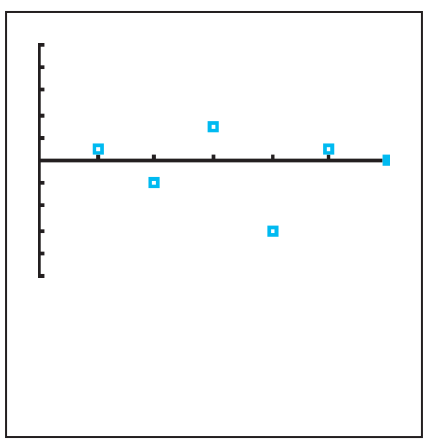
2. residual plot

Problem Set

Complete each table. Round your answers to the nearest tenth. Construct a residual plot.

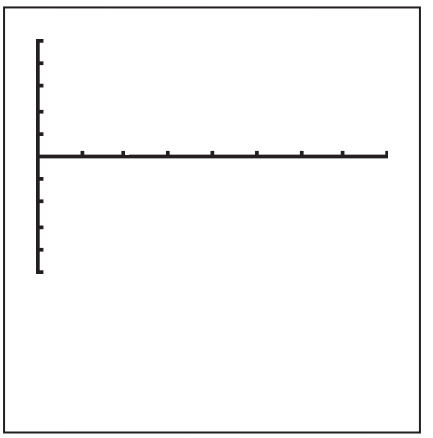
1. Linear regression equation: $y = 0.5x$

x	*y*	Predicted Value	Residual Value
5	3	2.5	0.5
10	4	5	−1
15	9	7.5	1.5
20	7	10	−3
25	13	12.5	0.5
30	15	15	0



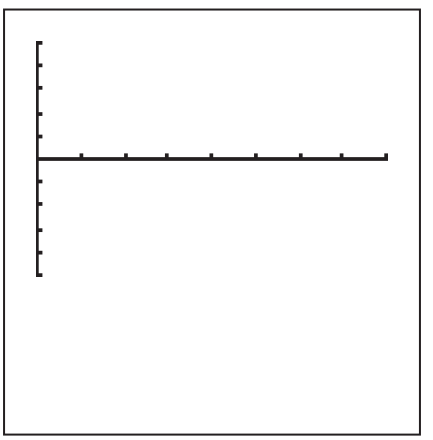
2. Linear regression equation: $y = -0.4x + 16.3$

x	y	Predicted Value	Residual Value
2	5		
4	15		
6	26		
8	23		
10	11		
12	3		



3. Linear regression equation: $y = 3x - 2.1$

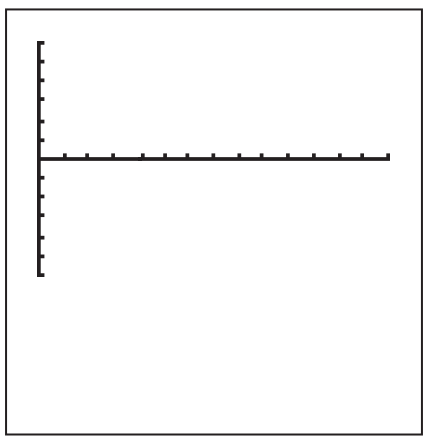
x	y	Predicted Value	Residual Value
1	1.5		
3	6.5		
5	12.5		
7	19.5		
9	24.5		
11	31.5		



Name _____ Date _____

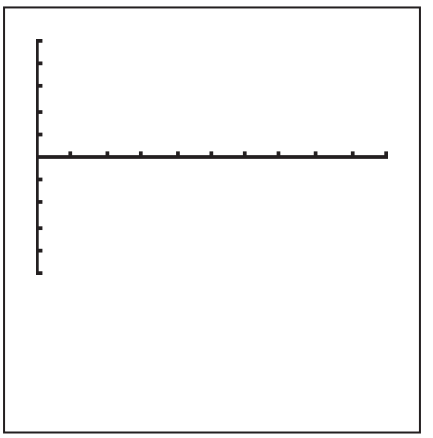
4. Linear regression equation: $y = -9.6x + 641.7$

x	*y*	Predicted Value	Residual Value
10	600		
20	450		
30	300		
40	200		
50	150		
60	125		



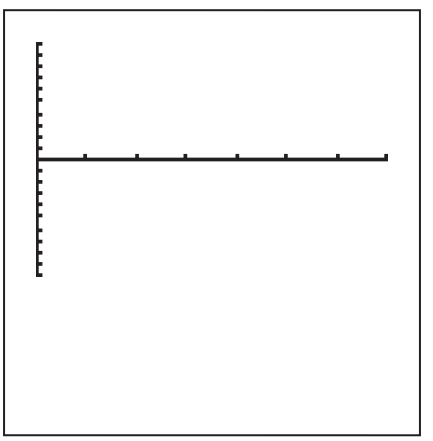
5. Linear regression equation: $y = 4.9x + 16.4$

x	*y*	Predicted Value	Residual Value
100	505		
90	460		
80	415		
70	360		
60	305		
50	265		



6. Linear regression equation: $y = -x + 19.7$

x	*y*	Predicted Value	Residual Value
2	17		
4	16		
6	15		
8	12		
10	9		
12	8		



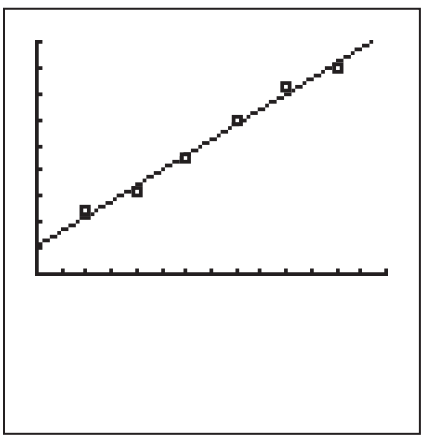
Name _____ Date _____

Consider the scatter plot, its line of best fit, and the corresponding residual plot of each data set. State if a linear model is appropriate for the data.

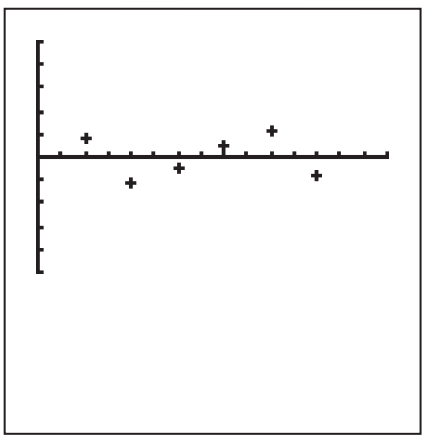
7. Linear regression equation: $y = 2.96x + 5.30$, $r = 0.9964$

x	2	4	6	8	10	12
y	12	16	22.5	29.5	36	40

Scatter Plot & Line of Best Fit



Residual Plot

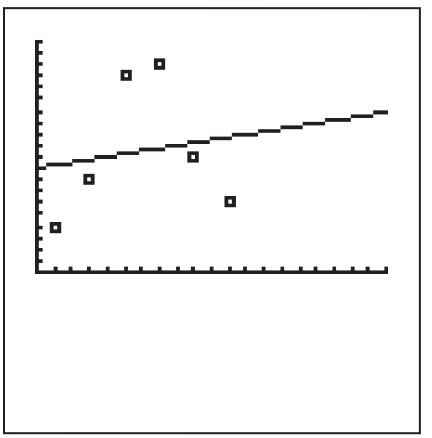


Based on the shape of the scatter plot and the correlation coefficient, a linear model appears to be appropriate for the data. Based on the residual plot, a linear model appears to be appropriate for the data.

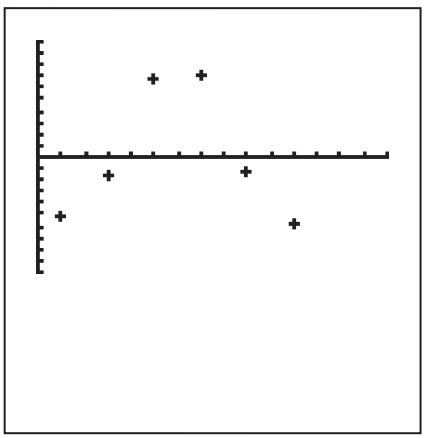
8. Linear regression equation: $y = 0.24x + 9.04$, $r = 0.1570$

x	1	3	5	7	9	11
y	4	8	17	18	10	6

Scatter Plot & Line of Best Fit



Residual Plot

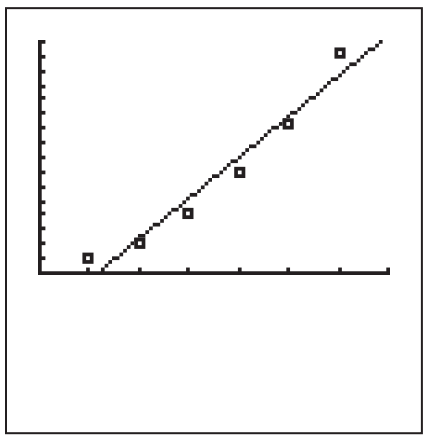


Name _____ Date _____

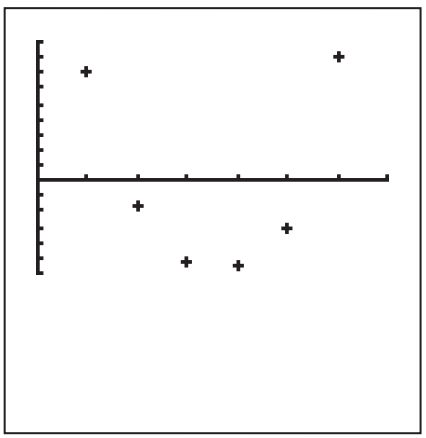
9. Linear regression equation: $y = 14.08x - 163.13$, $r = 0.9746$

x	10	20	30	40	50	60
y	49	103	207	346	511	762

Scatter Plot & Line of Best Fit



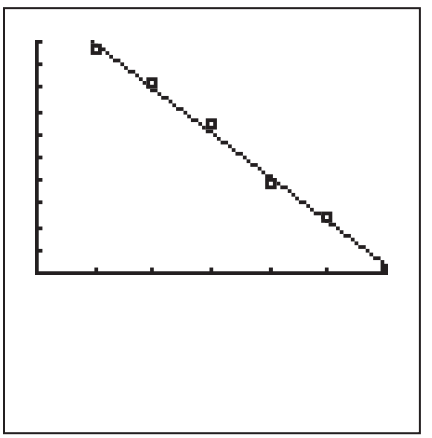
Residual Plot



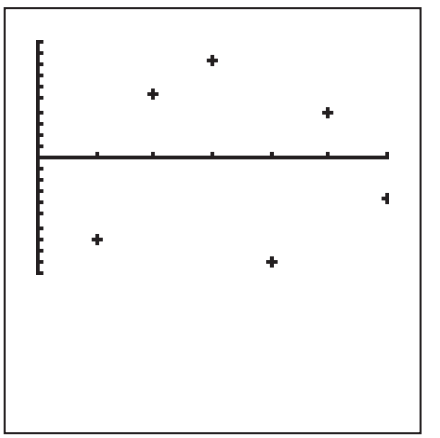
10. Linear regression equation: $y = -1.91x + 59$, $r = -0.9968$

x	5	10	15	20	25	30
y	48	41	32	19	12	1

Scatter Plot & Line of Best Fit



Residual Plot

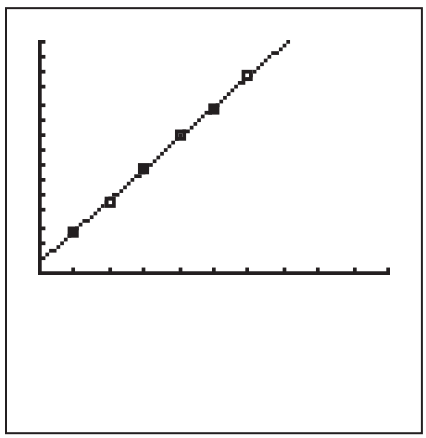


Name _____ Date _____

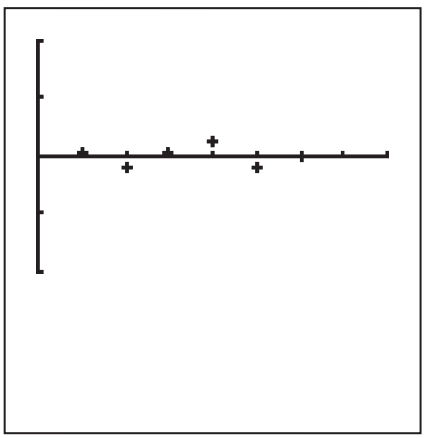
11. Linear regression equation: $y = 4.01x + 1.43$, $r = 0.9997$

x	1	2	3	4	5	6
y	5.5	9.25	13.5	17.75	21.25	25.5

Scatter Plot & Line of Best Fit



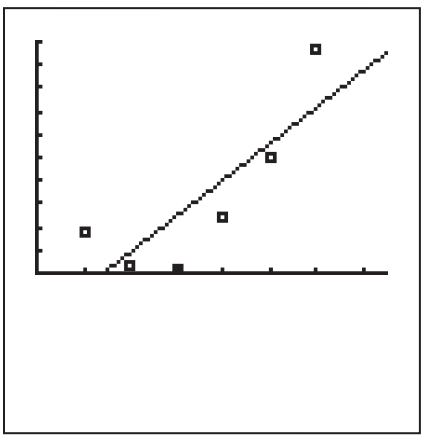
Residual Plot



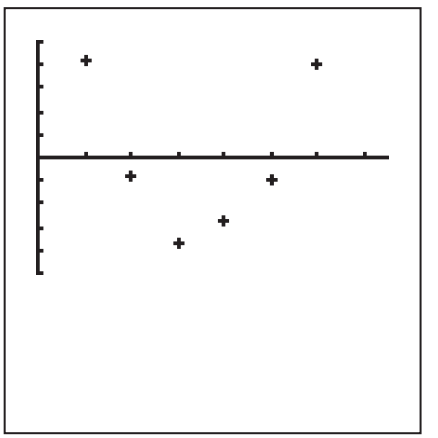
12. Linear regression equation: $y = 3.93x - 11.33$, $r = 0.8241$

x	2	4	6	8	10	12
y	9	2	1	12	25	48

Scatter Plot & Line of Best Fit



Residual Plot



Name _____ Date _____

To Fit or Not To Fit? That Is The Question!
Using Residual Plots

Problem Set

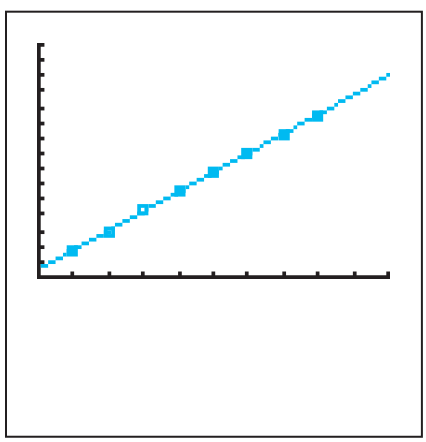
For each data set, determine the linear regression equation. Then, construct a scatter plot and a corresponding residual plot. State if a linear model is appropriate for the data. Round your answers to the nearest hundredth. Round the correlation coefficient to the nearest ten thousandth.

1.

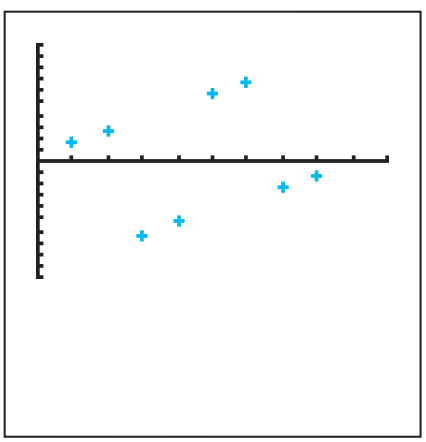
x	10	20	30	40	50	60	70	80
y	351	601	849	1099	1351	1601	1849	2099
Prediction	350.66	600.46	850.26	1100.06	1349.86	1599.66	1849.46	2099.26
Residual	0.34	0.54	−1.26	−1.06	1.14	1.34	−0.46	−0.26

Linear regression equation: $y = 24.98x + 100.86$, $r = 1.0000$

Scatter Plot & Line of Best Fit



Residual Plot



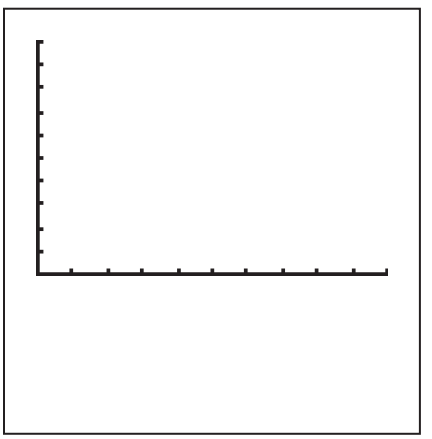
Based on the shape of the scatter plot and the correlation coefficient, a linear model appears to be appropriate for the data. Based on the residual plot, a linear model appears to be appropriate for the data.

2.

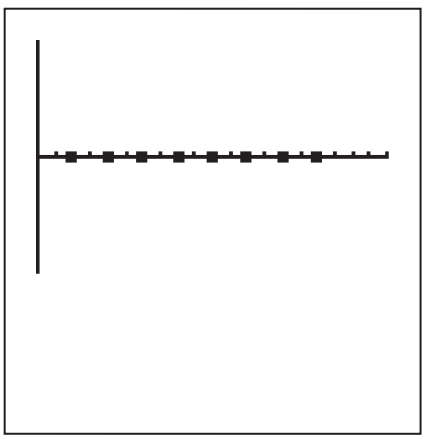
x	2	4	6	8	10	12	14	16
y	8	14	20	26	32	38	44	50
Prediction								
Residual								

Linear regression equation:

Scatter Plot & Line of Best Fit



Residual Plot



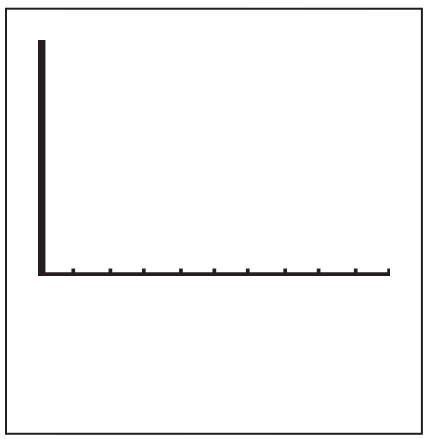
Name _____ Date _____

3.

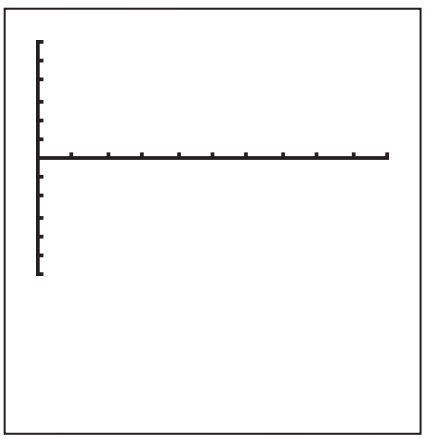
x	1	3	5	7	9	11	13	15
y	2	10	26	50	82	122	170	226
Prediction								
Residual								

Linear regression equation:

Scatter Plot & Line of Best Fit



Residual Plot

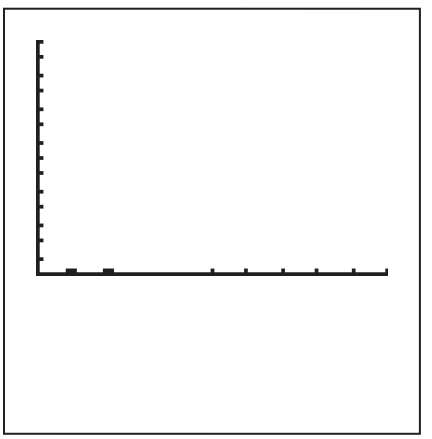


4.

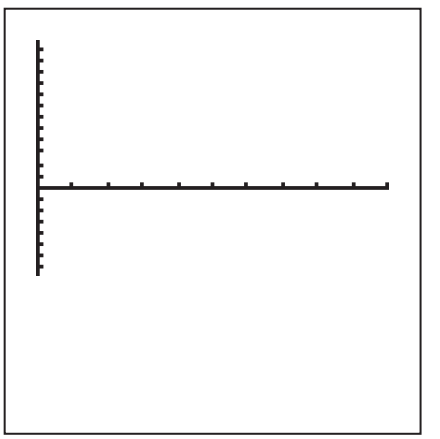
x	2	4	6	8	10	12	14	16
y	2	5	11	25	57	129	291	656
Prediction								
Residual								

Linear regression equation:

Scatter Plot & Line of Best Fit



Residual Plot



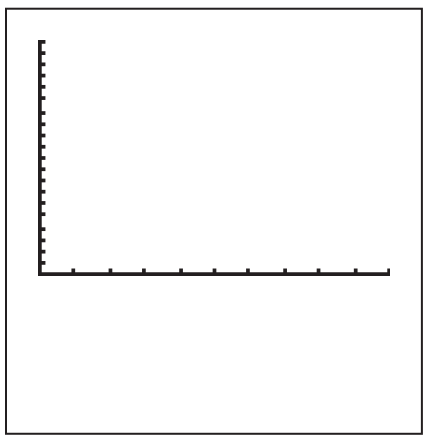
Name _____ Date _____

5.

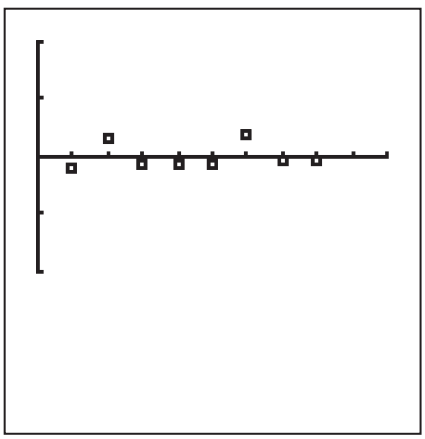
x	1	2	3	4	5	6	7	8
y	37.5	35.5	32.5	30	27.5	25.5	22.5	20
Prediction								
Residual								

Linear regression equation:

Scatter Plot & Line of Best Fit



Residual Plot

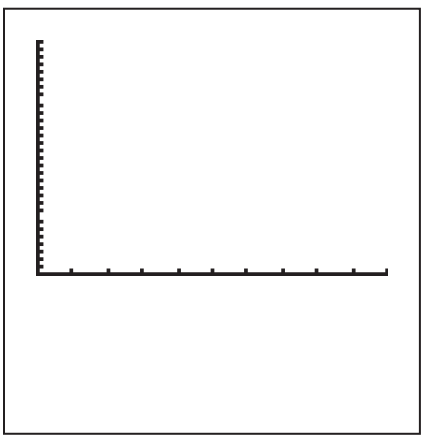


6.

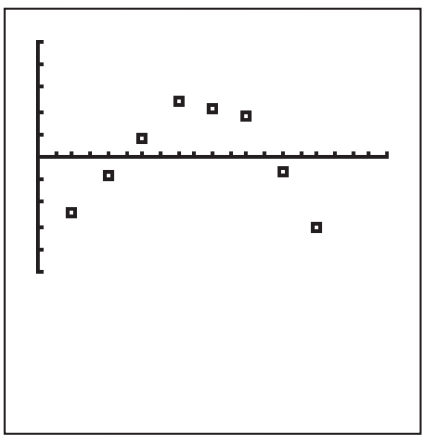
x	2	4	6	8	10	12	14	16
y	50	48	46	44	40	36	30	24
Prediction								
Residual								

Linear regression equation:

Scatter Plot & Line of Best Fit



Residual Plot



Name _____ Date _____

Who Are You? Who? Who?
Causation vs. Correlation

Vocabulary

Choose the word from the box that best completes each sentence.

causation	necessary condition	confounding variable
common response	sufficient condition	

1. A correlation is a _____ for causation, but a correlation is not a _____ for causation.
2. A _____ is when some other reason may cause the same result.
3. _____ is when one event causes a second event.
4. A _____ is when there are other variables that are unknown or unobserved.

Problem Set

Determine whether each correlation implies causation. List reasons why or why not.

1. The amount of ice cream a grocery store sells is negatively correlated to the amount of soup that the grocery store sells.

   The correlation does not imply causation. There may be a correlation between ice cream sales and soup sales. For instance, ice cream sales may increase as soup sales decrease because ice cream sales typically increase in warmer weather and soup sales typically decrease in warmer weather. However, this trend does not mean that an increase in ice cream sales causes the soup sales to decrease.

2. The number of new entry-level jobs in a city is positively correlated to the number of new home sales.

3. There is a positive correlation between the total number of dollars paid toward an education and a person's annual salary.

4. There is a negative correlation between the number of times a person washes their hands during the day and the number of times that person catches a cold.

5. There is a negative correlation between the number of hours a student plays video games per day and the grades a student receives in school.

6. There is a positive correlation between the number of hours a student spends studying and the grades a student receives in school.

Name _____ Date _____

Read each statement. Then answer the questions. Explain your reasoning.

7. A study claims that eating a healthy breakfast improves school performance.

a. Do you think that eating breakfast every morning is a necessary condition for a student to perform well at school?

Yes. It is very difficult for a student to perform well in school without a healthy breakfast.

b. Do you think that eating breakfast every morning is a sufficient condition for a student to perform well at school?

No. Not every student who eats breakfast every morning performs well at school.

8. A teacher said that students who read a book slowly will understand the story.

a. Do you think that reading a book slowly is a necessary condition for understanding the story?

b. Do you think that reading a book slowly is a sufficient condition for a student to understand the story?

9. A reporter claims that when there are a large number of paramedics at a disaster site, there are a large number of fatalities.

a. Do you think that a large number of paramedics at a disaster site is a necessary condition for a large number of fatalities?

b. Do you think that a large number of paramedics at a disaster site is a sufficient condition for a large number of fatalities?

10. An adult claims that if you play with fire, you are going to have bad dreams.

a. Do you think that playing with fire is a necessary condition for a person to have bad dreams?

b. Do you think that playing with fire is a sufficient condition for a person to have bad dreams?

11. A dietician says that if people reduce their caloric intake they will lose weight.

a. Do you think that reducing caloric intake is a necessary condition for a person to lose weight?

b. Do you think that reducing caloric intake is a sufficient condition for a person to lose weight?

12. A cosmetic company claims that if you use sunscreen you will not get skin cancer.

a. Do you think that using sunscreen is a necessary condition for a person to not get skin cancer?

b. Do you think that using sunscreen is a sufficient condition for a person to not get skin cancer?