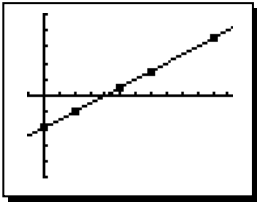


THE CORRELATION COEFFICIENT

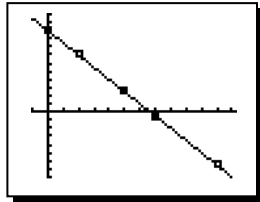
ALGEBRA 2 WITH TRIGONOMETRY

Data collected by scientists and others can oftentimes be fit with a line or other curve, but that fit is not always perfect. It is important for people who use linear models based on data to know how well the linear model predicts what is happening in the data set. The measurement of this fit is known as the **linear correlation coefficient** or r -value.

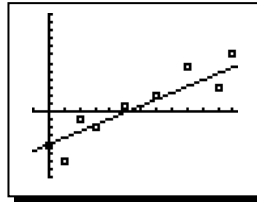
Exercise #1: Below are five scatter plots along with their lines of best fit and r -values. Based on these graphs answer the following questions.



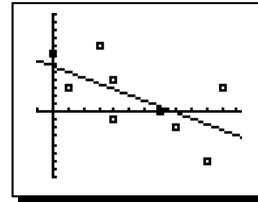
$$r = 1$$



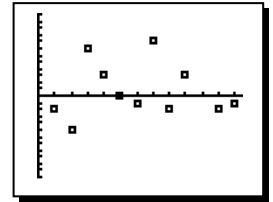
$$r = -1$$



$$r = 0.89$$



$$r = -0.67$$



$$r = 0$$

- (a) Why do some graphs have positive r -values and others negative r -values? (In other words, what does the sign of the r -value tell us about the correlation?)
- (b) What does an r -value whose absolute value is one tell us about the nature of the fit? (The first two graphs)
- (c) Does the fit get better or worse as the r -value gets closer to zero?
- (d) Why is there no line of best fit drawn on the graph where $r = 0$?

THE CORRELATION COEFFICIENT

Perfect Negative Correlation

$$-1 \leq r \leq 1$$

Perfect Positive Correlation

Exercise #2: Which of the following r -values would indicate a linear model with the best predictive value?

(1) $r = 0.85$

(3) $r = 0.27$

(2) $r = -0.62$

(4) $r = -0.94$



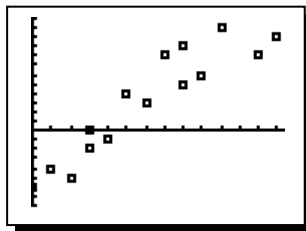
Exercise #3: Which of the following *could* be the value of the linear correlation coefficient for the following scatter plot?

(1) $r = 1$

(3) $r = 0.78$

(2) $r = 0$

(4) $r = -0.84$



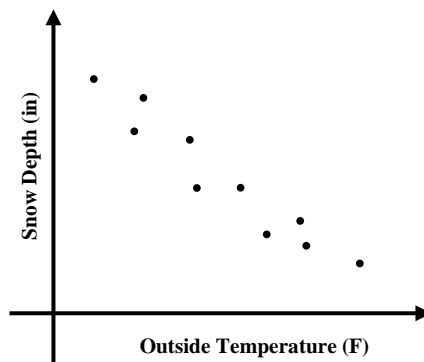
Exercise #4: The points on the following scatter plot represent the high temperature outside and the depth of snow in Spring when it is melting. From this data, which of the following would be reasonable to assume?

(1) The linear correlation coefficient is between 0 and 0.5.

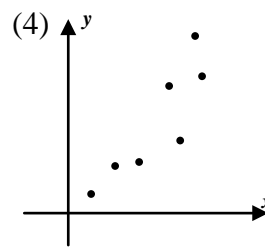
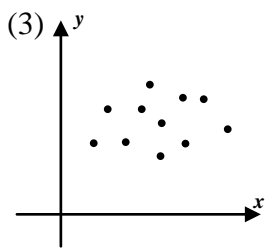
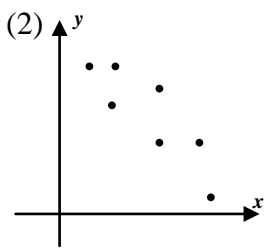
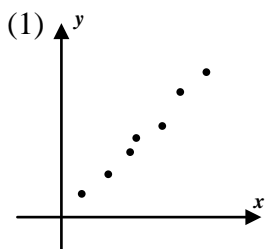
(2) The linear correlation coefficient is less than zero.

(3) The linear correlation coefficient is zero.

(4) The linear correlation coefficient is greater than 0.5.



Exercise #5: Each of the following graphs shows a scatter plot of data. Which of the graphs would have a linear correlation coefficient closest to zero?



Exercise #6: A scientist found that the least squares regression line relating two variables was given by the equation $y = 8.72 - 3.25x$. From this equation which of the following can we *definitely* conclude?

(1) $r = 0$

(2) $0 < r < 0.5$

(3) $r > 0$

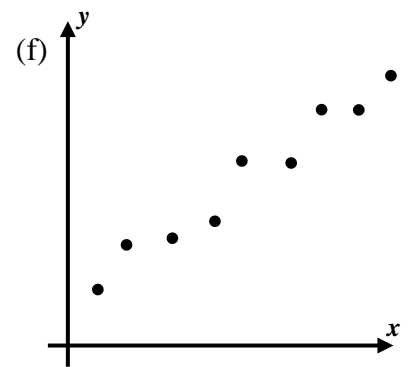
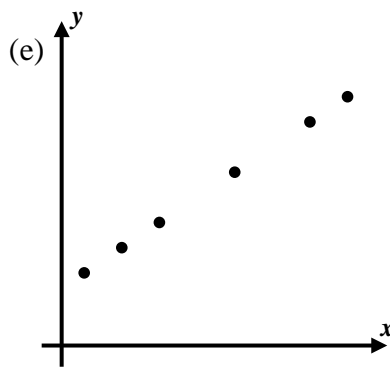
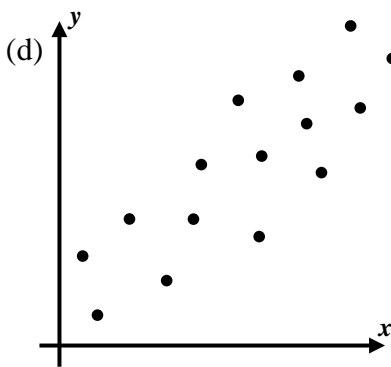
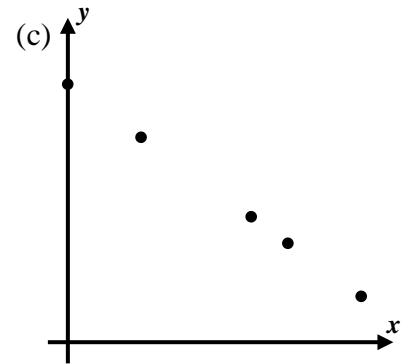
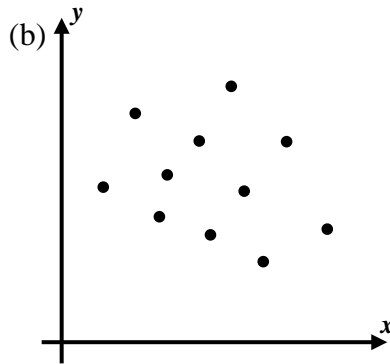
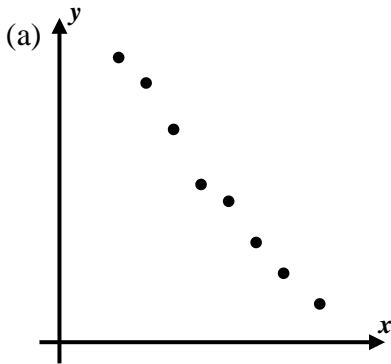
(4) $r < 0$



THE CORRELATION COEFFICIENT
ALGEBRA 2 WITH TRIGONOMETRY - HOMEWORK

SKILLS

1. Shown below are six scatter plots. Below the scatter plots are six r -values. Using a straightedge, draw what you believe to be the line of best fit on each graph. Then, place the most appropriate r -value in the blank below the corresponding scatter plot.



Possible r -values: -1 , 1 , 0 , 0.85 , -0.96 , 0.50

2. Which of the following linear correlation coefficients indicates a linear regression equation with the most predictive accuracy?

(1) $r = 0.72$

(3) $r = -0.85$

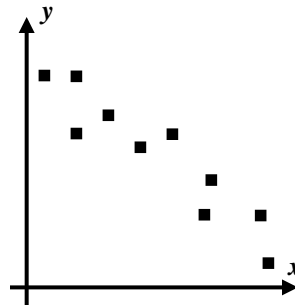
(2) $r = 0.92$

(4) $r = -0.12$



3. Which of the following could be the correlation coefficient for the scatter plot shown below?

- (1) $r = -0.88$ (3) $r = -1$
 (2) $r = 0.93$ (4) $r = 1$



4. A physiologist is trying to relate the number of weekly hours spent at the gym, h , to a person's weight loss while, w . The best fit line was found to have the equation $w = 2.72h - 4.25$. From this information, which of the following *must* be true about the correlation coefficient?

- (1) $r = 1$ (3) $r > 0$
 (2) $r < 0$ (4) $r = 0.5$

APPLICATIONS

5. The price that a driver pays at the pump is dependent on many different variables, one of which is the price per barrel of oil. The table below gives a price for a barrel of oil and the corresponding average price for a gallon of regular unleaded gasoline.

Oil Price (per barrel)	45	75	84	96	125	140
Gasoline Price (per gallon)	2.37	2.54	2.84	3.25	3.54	4.08

- (a) Find the best fit linear equation, in $y = ax + b$ form, that relates the oil price, x , to the gasoline price, y . Round all coefficients to the nearest *hundredth*.
- (b) The correlation coefficient for this data set is $r = 0.97$ (rounded to the nearest hundredth). Based on this, would you characterize the relationship between these two variables as strongly positive, weakly positive, strongly negative, or weakly negative? Explain.
- (c) Based on your equation from part (a), predict the price for a gallon of gas when the oil price hits 200 dollars per barrel.
- (d) Is the calculation you made in part (c) an example of extrapolation or interpolation? Explain.

