Name: _____ Date: _____

QUANTIFYING PREDICTABILITY
COMMON CORE ALGEBRA I

In the last few lessons we have worked with generating **lines and curves of best** fit for **bivariate** data sets. In every circumstance, though, the data did not fall on a straight line or on a perfect curve. We have never answered the question of how well specifically a **linear model** does in predicting the **correlation** between the two variables.

***Exercise* #1:** Your teacher will explain how to ensure that your calculator has its "r-value" on. Since this varies by graphing calculator, write down the procedure below if necessary.

***Exercise* #2:** In the following exercises four data sets with equal *x*-values are given to illustrate different types of **positive correlations**. For each, enter the data, observe the scatter plot, and record the *r*-value, known as the **correlation coefficient**, for a **linear fit** to the nearest *thousandth*.

(a)

x	2	5	8	11	15	18
y	4	13	22	29	43	52

(b)

x	2	5	8	11	15	18
y	16	14	22	41	37	51

(c)

x	2	5	8	11	15	18
y	18	8	41	28	62	44

(d)

x	2	5	8	11	15	18
y	44	51	30	55	45	47

(d) How does the **correlation coefficient** quantify the fit of a positive correlation?

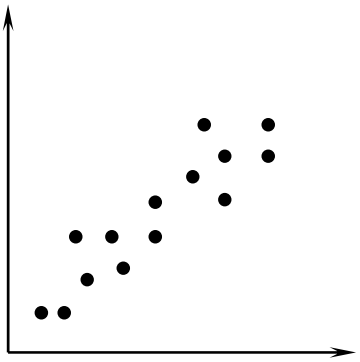
***Exercise* #3:** The following data set is that of two variables that have a **negative correlation**. Enter the data, produce the scatter plot, and record the *r*-value. How is the negative correlation reflected in the *r*-value?

x	2	5	8	11	15	18
y	52	47	28	32	25	10

***Exercise* #4:** Given the scatter plot shown below, which of the *r*-values would most likely represent the correlation between the two variables? Explain your choice.

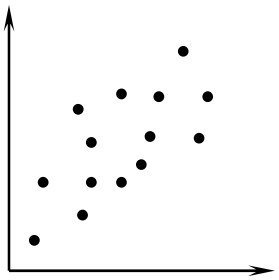
(1) $r = 0.88$ (3) $r = 1$

(2) $r = 0.28$ (4) $r = -0.94$

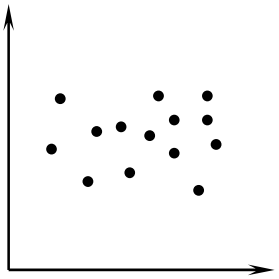


***Exercise* #5:** Which of the following scatter plots would have a correlation coefficient closest to -1?

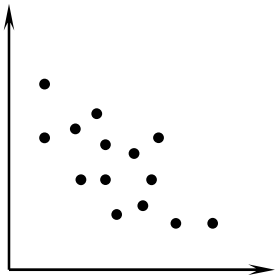
(1)



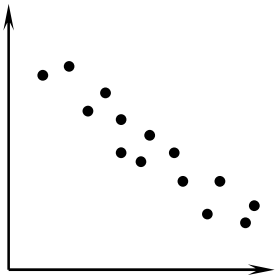
(3)



(2)



(4)



***Exercise* #6:** There are two primary types of crude oil sold in the world, West Texas Intermediate (WTI) and Brent Crude. Each is priced differently on a daily basis and each has a correlation with the average price per gallon for unleaded gasoline. The two linear regression models, along with their *r*-values, are shown below. Give a prediction for the price per gallon of unleaded gasoline, *y*, on a day when the price for WTI is \$103 and the price for Brent is \$109, *x*. Which model did you choose and why?

Brent Crude: $y = 0.028x + 0.71$, $r = 0.973$ WTI Crude: $y = 0.031x + 0.67$, $r = 0.924$

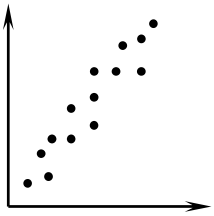
Name: _____ Date: _____

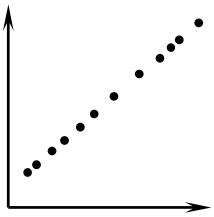
QUANTIFYING PREDICTABILITY

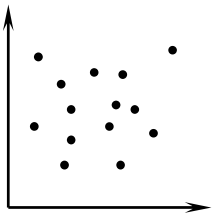
COMMON CORE ALGEBRA I HOMEWORK

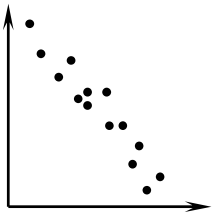
1. Below there are six scatter plots, six correlation coefficients, and six terms. Match the appropriate *r*-value with the scatter plot it most likely corresponds to. Then match the term you think is most appropriate to the *r*-value as well (not to the graph).

(a)  $r = 1.0$ Weak Negative

(b)  $r = 0.35$ Perfect Positive

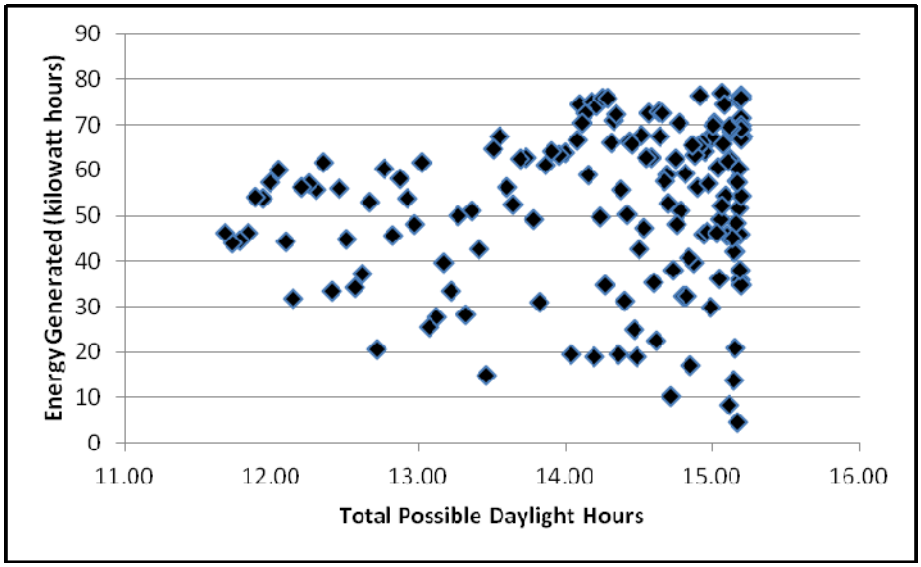
(c)  $r = -0.82$ Strong Positive

(d)  $r = 0$ Weak Positive

(e)  $r = -0.56$ Moderate Negative

(f)  $r = 0.93$ No Correlation

2. A solar power company is trying to correlate the total possible hours of daylight (simply the time from sunrise to sunset) on a given day to the production from solar panels on a residential unit. They created a scatter plot for one such unit over the span of five months. The scatter plot is shown below.



The equation line of best fit for this bivariate data set was:

$y = 2.26x + 20.01$.

(a) How many kilowatt hours would the model predict on a day that has 14 hours of possible daylight?

(b) To the nearest tenth of an hour, how many hours of possible daylight would be needed to produce 50 kilowatt hours of energy?

(c) The correlation coefficient for this regression was $r = 0.134$. Would you characterize this as strongly positive, moderately positive, or a weakly positive correlation? Explain.

(d) Based on (c), do you have confidence in the model to accurately predict the energy production based on the total possible daylight hours? Explain.

(e) What environmental factors might contribute to the "noise" in the data? Noise are factors that prevent the correlation from being perfect.