Name: _____ Date: _____

BIVARIATE DATA ANALYSIS
COMMON CORE ALGEBRA I

Oftentimes, statistical studies are done where data is collected on **two variables** instead of one in order to establish whether there is a **relationship** between the **two variables**. This is called a **bivariate data analysis**.

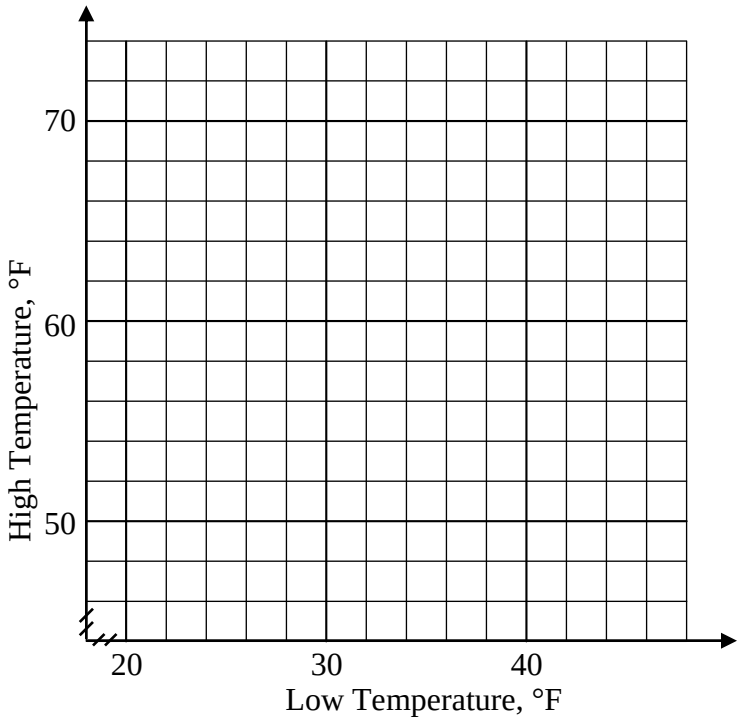
***Exercise* #1:** A survey was taken of 10 low and high temperatures, in Fahrenheit, in the month of April to try to establish a relationship between a day's low temperature and high temperatures.

Low Temperature, x	26	28	30	32	34	35	37	38	41	45
High Temperature, y	49	50	57	54	60	58	64	66	63	72

(a) Construct a scatter plot of this bivariate data set on the grid below.

(b) Draw a line of best fit through this data set.

(c) Calculate the slope of this line by picking off two points (not necessarily data points).



(d) Use your line of best fit to estimate the high temperature for a day in April given that the low temperature was 42 degrees. Illustrate your answer on your graph.

(e) Would you characterize the relationship between the low and high temperature as a **positive correlation** or a **negative correlation?** Explain.

eMATHINSTRUCTION, RED HOOK, NY 12571, © 2013

Two variables can have a **strong relationship** with one another, as seen on a scatterplot, but might not have a **causal relationship**. A causal relationship exists when the **change in one variable** actually **causes** the **change** in the other (or is one of the primary causes).

***Exercise* #2:** In each of the following scenarios, two variables are given that if plotted would have a strong correlation (a scatterplot where the data falls nearly in a line). Determine if there exists a **causal** relationship between the two variables. If so, which variable causes the other?

(a) The high temperature in New York City and the number of bottles of water sold.

(b) A person's height and a person's shoe size.

(c) A person's weight loss and the number of hours a person spends in the gym per week.

(d) The years of education a person achieves and the salary that person starts at upon entering the work force.

Variables can have extremely strong **correlations** but no causal relationship. This is often the case if there is a third variable that causes both (known interestingly enough as a **lurking variable**).

***Exercise* #3:** The table below shows the number of firefighters required to fight a given fire versus the dollar damage done to the house by the fire.

Number of firefighters	2	3	5	8	9	12	16
Damage done by fire (in dollars)	2,932	9,750	15,575	23,190	22,900	35,400	52,900

(a) Are the data positively or negatively correlated? How can you tell?

(b) Does the number of firefighters cause the damage done to the house? If not, what hidden variable is causing both variables to change?

eMATHINSTRUCTION, RED HOOK, NY 12571, © 2013

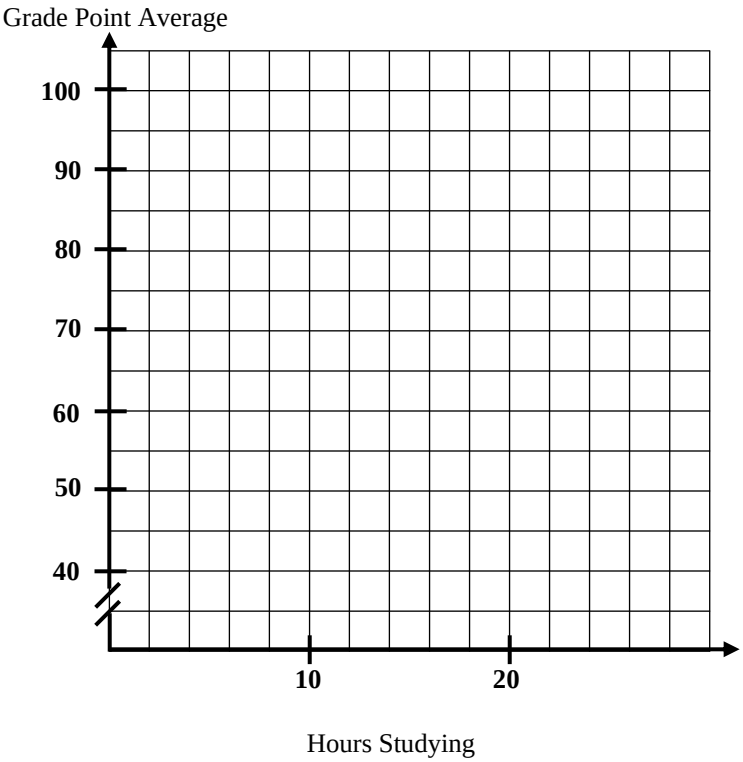
Name: _____ Date: _____

BIVARIATE DATA ANALYSIS
COMMON CORE ALGEBRA I HOMEWORK

1. A survey was done at Ketcham High School to determine the effect of time spent on studying and grade point average. The table below shows the results for 10 students randomly selected.

Study time (Hours per week)	2	4	5	7	10	12	14	17	19	20
GPA (out of 100)	64	71	69	74	81	86	84	94	91	96

(a) Create a scatter plot for this data set on the grid provided. Draw a best fit line through the scatterplot drawn.



(b) Determine the equation for the best fit line for this data set. Use two points from the line you drew in (a) to determine the slope and estimate the *y*-intercept graphically. Round your slope to the nearest *tenth* and the *y*-intercept to the nearest whole number.

(c) Use your answer from part (b) to determine the expected GPA from studying for 8 hours per week. Round your answer to the nearest whole number.

(d) Is there a causal relationship between these two variables? If so, which variable causes the other?

eMATHINSTRUCTION, RED HOOK, NY 12571, © 2013

2. A survey was done to determine if there was any connection between the price that people pay for their most expensive car and the current value of their house. The results, for eight participants, are given below.

Car Cost, x (in dollars)	11,500	14,750	19,500	26,750	32,900	43,000	45,750	54,500
House Value, y (in dollars)	160,000	195,000	255,000	400,000	440,000	525,000	475,000	725,000

A computer was used to determine the line of best fit. Its equation was:

$$y = 12x + 33,766$$

(a) Use the line of best fit to predict the house value of a person whose most expensive car costs $19,500.

(b) Was the prediction in (a) an overestimate or underestimate of the actual house value? Explain.

(c) Is there a positive or negative correlation between these two variables? Explain.

(d) Is there a causal relationship between these two variables? If you answer yes, then determine which variable causes the other. If you answer no, then explain a third variable that could be causing both.

3. It has been noted that on any given day, there is a strong correlation between the number of ice-cream cones sold and the number of people who go swimming. Is there a causal relationship between the eating ice-cream and going swimming? If not, what could be causing this strong correlation?