Statistics: Two-Variable

Name: _____

Goals:

Target A: I can define and represent two quantitative variables on a scatter plot and describe how the variables are related (direction, shape, strength).

Target B: I can distinguish between data sets that should be modeled by a linear function compared with an exponential function or other relationship.

Target C: I can explain what the parameters of a model mean in the context of the situation (rate of change, initial value).

Target D: I can differentiate between categorical and quantitative data.

Resources:

www.mrnohner.com/twovar

Standard	5A	5B	5C	5D
Do I need to retake?				

Two Variable Stats

L I N E O F B E S T F I T Y L C J A I G R B O S H

O C Z Z J J O M F C V K I H W A Z R I M I O H B N

N O I T C E R I D P A N X K Y A U S E P O T C C X

L W T E K C D Y G E T T B W R P E D S Y P S O O K

S H I R S G Z R M E G Q T W R S M B I G V X R W Y

S O C X I H T U R N H O V E K Y O I N S D P R S K

W Q W A P U A C U O T P M D R P L E N E E U E H U

O G I P R Q E P F E G P S L O P E I R Z L R L F D

Q M K C J P U E E T N V M F P V L E C Q T H A B B

Z H D U T W M L M E E T K S R C I O B B B T T R M

E R Q S I Z M A F C R D I R R L Q J T B C W I I O

D S J X S H Z L F O T Z G U T G C O Y T W O O N R

K F U F X Q X B B J S L N U T F I V F C B P N E K

Y Y L E X D R E H J I M O X A Y Y Q R U N Y A K A

S I V Z N A R A Y J U D G T A N F W C L H H A M C

-Scatterplot
-Line of best fit
-Shape
- Correlation
- Y-intercept
- Strength
- Outlier
-Slope
-Residual
-Direction

Guide notes

Title:	Date:
Goal:	

Notes:

1) Draw a line of best fit

2) Find the equation of the line of best fit

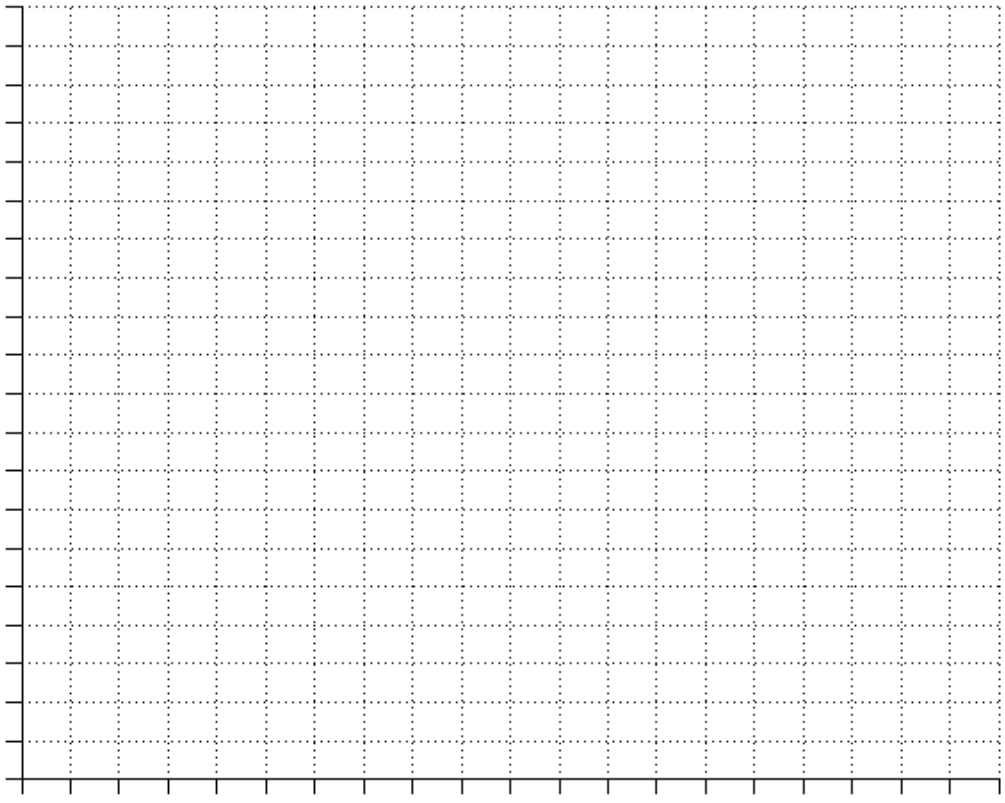
3) What does the slope mean?

4) What does the y-intercept mean?

5) How many grams of fat in a 4000 calorie meal?

6) How many calories in a 100 gram of fat meal?

Grams of fat per serving



Calories

Count by 50s and 2s

Driver Age	Sign Visibility (feet)
25	475
35	430
40	440
40	417
43	390
50	280
75	211
85	190

1) Find the equation of the line of best fit

2) What does the slope mean?

3) What does the y-intercept mean?

4) Use it to predict how far a 100 year old driver could see

5) Use it to predict how long until a driver could only see a sign 200 feet away.

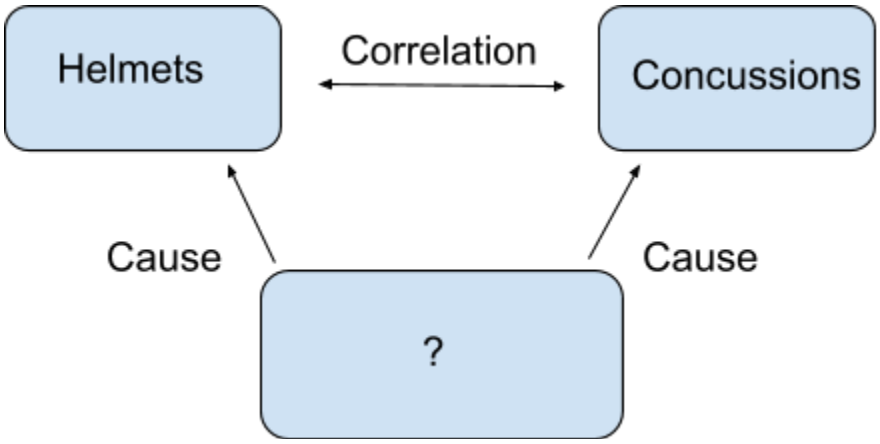
Summary:

Guide notes

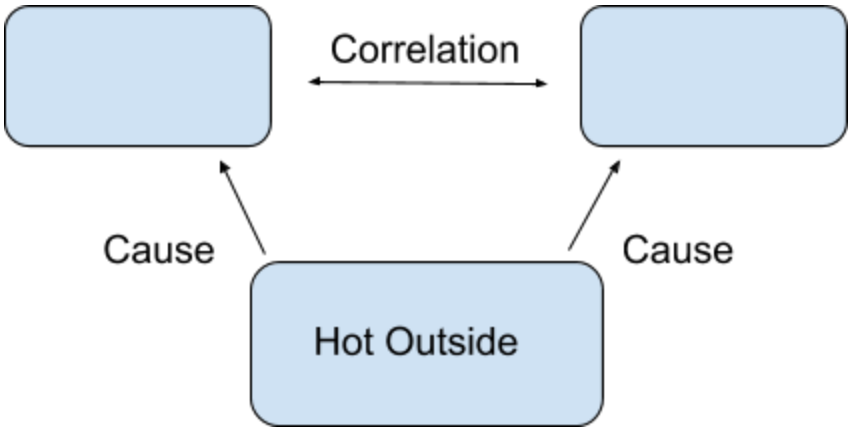
Title:	Date:
Goal:	

Notes:

These are obviously not causes of each other, what underlies each one of these?



Finish this diagram.



Summary:

Guide notes

Title:	Date:
Goal:	
Notes:	
Summary:	

Guide notes

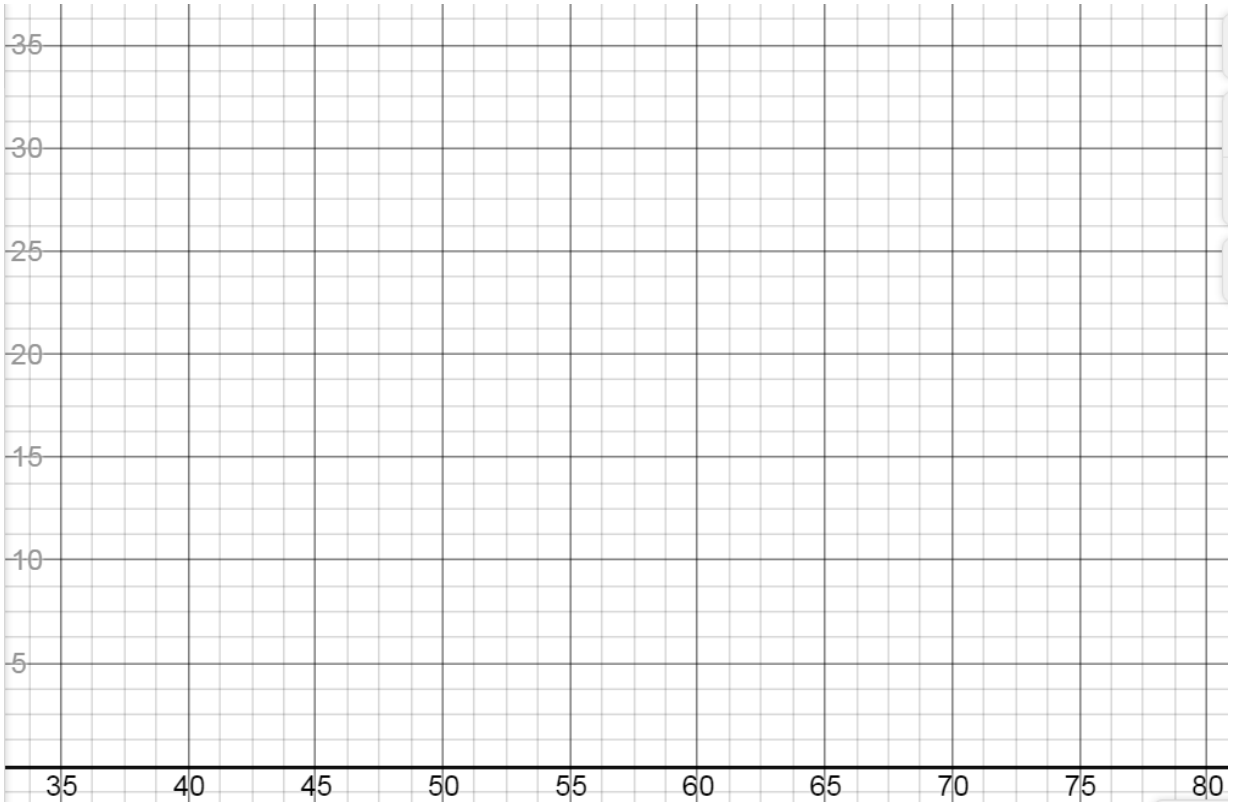
Title:	Date:
Goal:	
Notes:	
Summary:	

Guide notes

Title:	Date:
Goal:	

Notes:

Length of Ulna (centimeters)



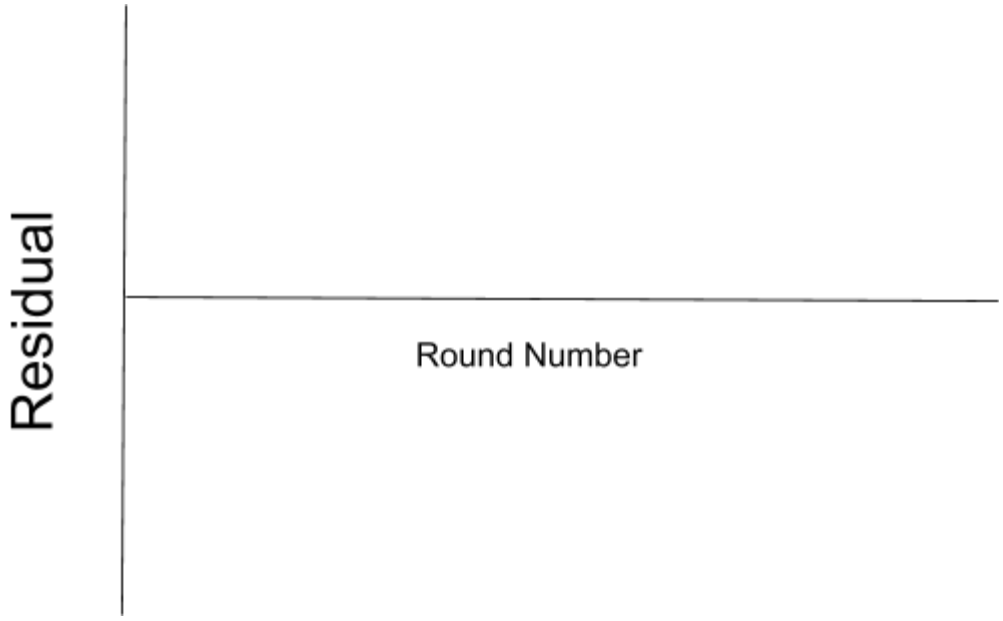
Your height (Inches)

How tall is the humanoid with the 26.4 cm bone ulna?

What if we add UPPER and LOWER Bounds?!

The murderer is going to get you all, it is only a matter of time, if you roll a 1 you are dead. Every other number and you live until the next round.

Everyone grab a die and stand up and come to the front of the room. If you roll a 1 you are dead (please sit down)!



Summary:

Guide notes

Title:	Date:
Goal:	
Notes:	
Summary:	

Guide notes

Title:	Date:
Goal:	
Notes:	
Summary:	

Guide notes

Title:	Date:
Goal:	
Notes:	
Summary:	

Guide notes

Title:	Date:
Goal:	
Notes:	
Summary:	

Finding the line of best fit from a scatter plot and a table.

The past population for Smallville over a 25 year period is shown below.

Year	1985	1990	1995	2000	2005	2010
Population	248	241	219	216	199	189

1) Using your calculator, find the line of best fit (round your slope to 5 decimal places).

2) Now, use your equation to find the population of Smallville in 2020.

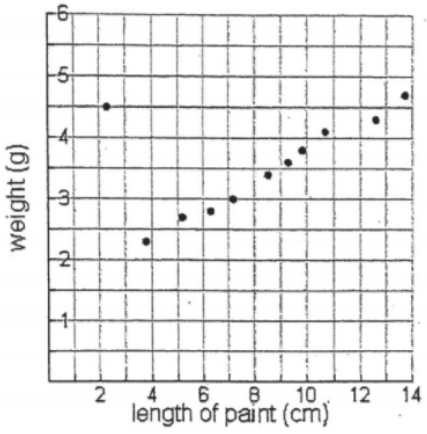
---

3) What is the line of best fit for this data?

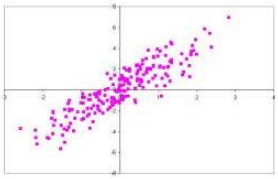
4) What is the correlation between the weight and the length of the pencil.

5) Write a possible explanation for the outlier in the data.

Sam collected data by sharpening her pencil and comparing the length of the painted part of the pencil to its weight. Her data is shown on the graph below:



6) Interpret the y-intercept in the context of this problem.



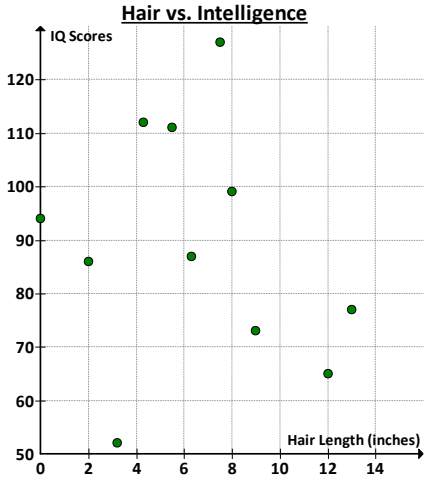
CORRELATION OF SCATTERPLOTS

Name: _____

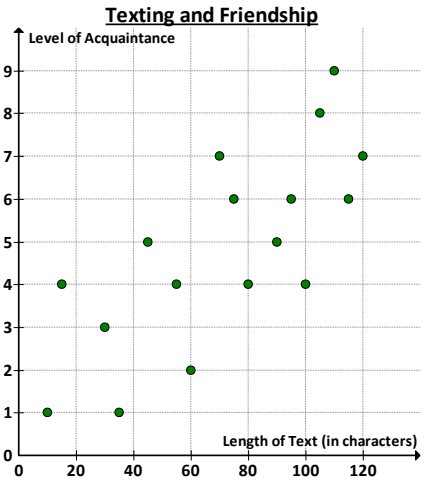
Date: _____ **Period:** _____

In math, we use the word ***correlation*** to describe the relationship that exists between the independent and dependant variables. Correlation addresses both the general trend of the points as you read the graph from left to right and how closely those points fit to a line. *For the following graphs and tables, choose the most appropriate correlation description. Justify your response.*

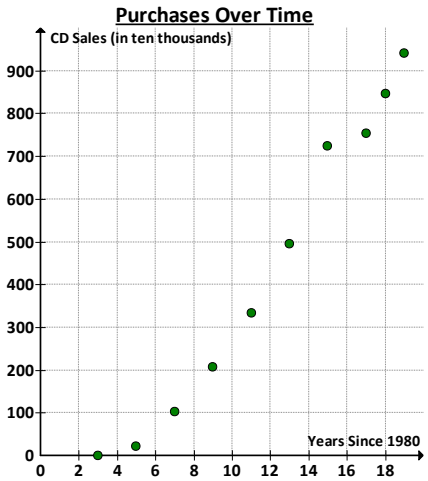
a. Strong Positive b. Strong Negative c. Weak Positive d. Weak Negative e. No Correlation



Correlation Type:



Correlation Type:



Correlation Type:

What is the correlation of the following? Justify your response.

1. Income vs. Years of Education

2. IQ vs. Weight

3. Video Game Usage vs. GPA

4. Age of a Car vs. Value of Car

5. Hours of Exercise vs. Body Fat

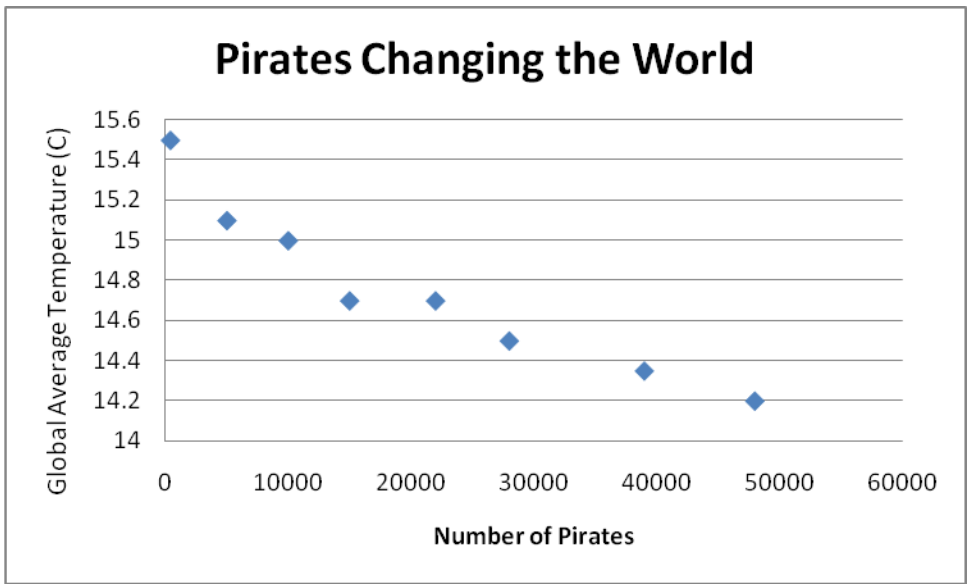
6. Studying vs. Test Errors

Now, create your own examples of:

7. Strong Positive Correlation: _____
8. Weak Negative Correlation: _____
9. No Correlation: _____

Let's look at various situations and think about "causality." That is, let's think about whether variables are associated with each other or whether one variable actually causes a change in another variable.

10. Over time, global temperature has changed. Also, the approximate number of pirates (they don't like to show up for the annual census) has changed. Eight data points are plotted below to show the relationship between these two variables over the last few hundred years.



- Describe the correlation type for this data.

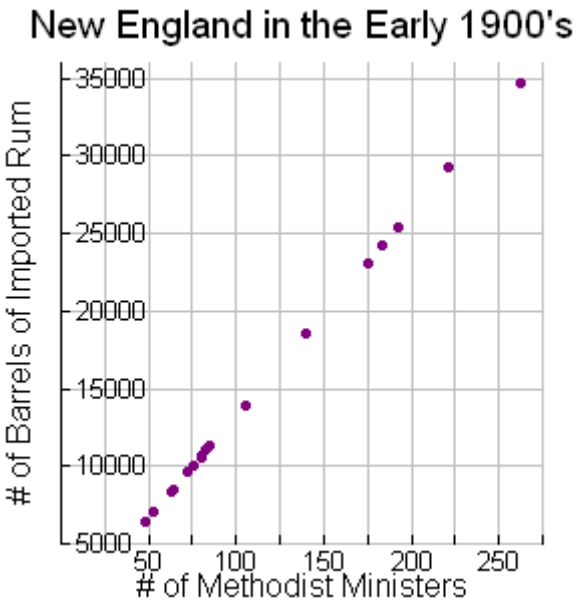
- Complete this sentence to describe the relationship between the variables.
  "As the _____ increases, the _____."

- Jack Sparrow concludes that increasing the number of pirates will combat global warming. Is this conclusion justified? Explain any flaws in his reasoning.

- Explain why you think that the global temperature is higher when there are fewer pirates.

11. A very peculiar relationship was found between two interesting variables.
Consider the following dataset (actual data– I did not make this up ☺) that shows the increase of Methodist ministers and the increase in the amount of Cuban rum imported from 1860 to 1940.

Year	Number of Methodist Ministers in New England	Number of Barrels of Cuban Rum Imported to Boston
1860	63	8376
1865	48	6406
1870	53	7005
1875	64	8486
1880	72	9595
1885	80	10,643
1890	85	11,265
1895	76	10,071
1900	80	10,547
1905	83	11,008
1910	105	13,885
1915	140	18,559
1920	175	23,024
1925	183	24,185
1930	192	25,434
1935	221	29,238
1940	262	34,705



a. Describe the type of correlation between these variables.

b. Complete this sentence to describe the relationship between the variables.
"As the _____ increases, the _____."

c. Is the increase in the number of Methodist ministers responsible for the increase in imported rum? Explain your thinking.

d. What other factor(s) might have caused an increase in both variables?

Name:_____ Residuals

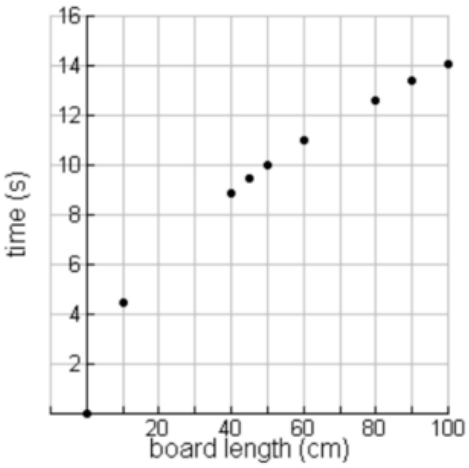
Battle Creek Cereal is trying a variety of packaging for their Toasted Oats cereal. They wish to predict the net weight of cereal based on the amount of cardboard used for the package. Below is a list of six current packages.

Packaging Cardboard (in^2)	Net Weight of Cereal (g)
47	28
69	85
100	283
111	425
125	566
138	850

1) What is the line of best fit that is modeled by this table?

2) A new “green” package will use 76 square inches of cardboard. Use your line of best fit to predict how much cereal it can hold.

3) To calculate the residual you need to use the equation “*residual = actual - predicted”*. We just found what are predicted value is, what would the residual be if we discovered that the 76 square inch “green box” actually held 198 grams of cereal?



4) Ms. Lang's class conducted an experiment by rolling a marble down different lengths of boards and timing how long it took. The results are shown to the left. Describe the correlation.

5) Home prices in Smallville can be modeled by the equation ***y = 150 + 62x*** where x is the size of the house in square feet and y is the cost of the house. Fancyville can be modeled by the equation ***y = 250 + 198x*** where x once again models the size of the house and y is the cost of the house. Ashley found a 2000 square foot house selling for $125,000.

a) Calculate the residual if you were to predict the house was located in the Smallville neighborhood.

b) Calculate the residual if you were to predict the house was located in Fancyville.

c) Using your residuals as proof, which neighborhood is the house more likely to be located in?

Two Variable Stats in Context Name:_____

1. Is there a relationship between the fat grams and the total calories in fast food? The chart below shows the total fat and calories for fast food items.

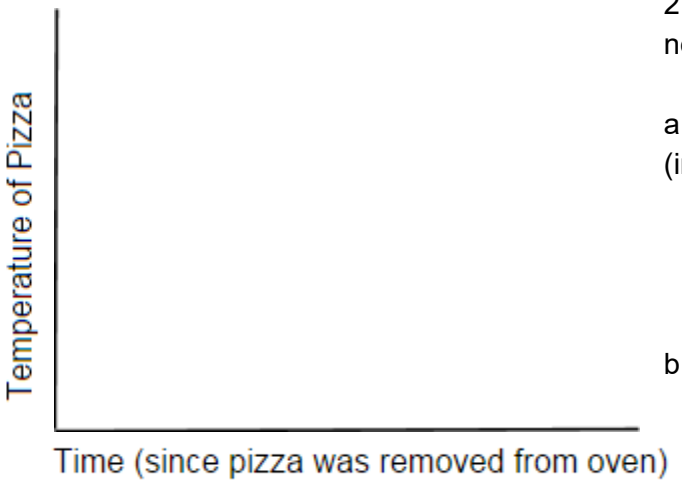
Total Fat (g)	Total Calories
9	260
13	320
21	420
30	530
31	560
31	550
34	590
25	500
28	560
20	440

a. Calculate the line of best fit from the data.

b. The table to the left is the ***actua***l data. Use your line of best fit to ***predict*** how many calories would be in each of the fast food items

Total Fat (g)	Predicted Calories
9	
13	
21	
30	
31	
31	
34	
25	
28	
20	

c. Find the residual for the 13 g **and** 28 g fat meal.

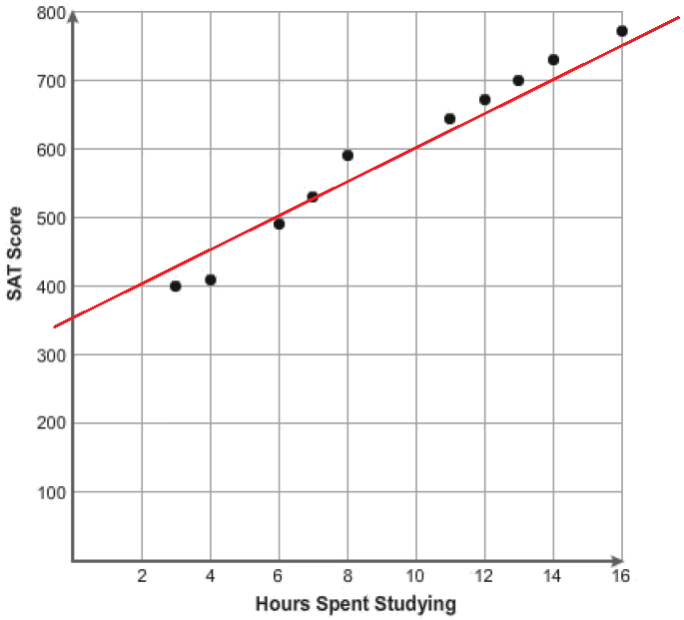


2. Create a fake scatter plot that represents weak negative correlation. Draw the line of best fit.

a. What does the y-intercept in your graph mean? (include units)

b. What does the slope mean in your graph? (include units)

3. The graph to the right was gathered from polling 10 students on their study habits and their SAT scores.



a. What is the correlation?

b. Find the y-intercept. What does it mean in this context?

c. Find the slope. What does it mean in this context?

d. Frederick predicted that if you study 14 hours you would get a 720 on the SAT. Did he predict correctly? Why or why not?

e. Estimate the residual for the student that studied 8 hours.

4. Ashley is trying to figure out the relationship between these two variables.

Portland Temperature	Amount of families that go to OMSI
40	730
74	482
58	594
80	447
86	375
50	640
68	524
93	349

a. Calculate the line of best fit.

b. What is the y-intercept? What does it mean in this context?

c. What is the slope? What does it mean in this context?

d. Use your line of best fit to predict the temperature if 510 families show up.

e. Use your line of best fit to predict how many people will show up if the temperature is 100^0.

5. Whole Foods sells over a thousand bananas a day. Instead of counting them all when they show up on each truck, they weigh them and use an equation, y = .58x + 50 to predict the number of bananas in the box. The number of bananas is x and the weight is y.

a. Why do you think the y-intercept is (0,50)?

b. Why do you think the slope is .58?

c. If a box weighs 100 pounds then how many bananas are in it?

d. If the shipper promised you 5000 bananas, then how many pounds are you expecting the box to be?

Review: Find the equation of the line parallel to y= 13x-9 and passess through the point (-7,-10)

Name:_____ **Residual Plots & Upper and Lower Bound**

1) Susan got the equation y = 22.7x + 19 when she found the line of best fit from the data of her classes height in feet (x) and weight in pounds (y). What does the slope mean in this situation?

---

2) Your house is infested with mice. Everyday you record how many mice you see and set traps to try to decrease the number of mice.

Days since you started trapping	Number of mice spotted
1	17
2	16
3	12
5	6

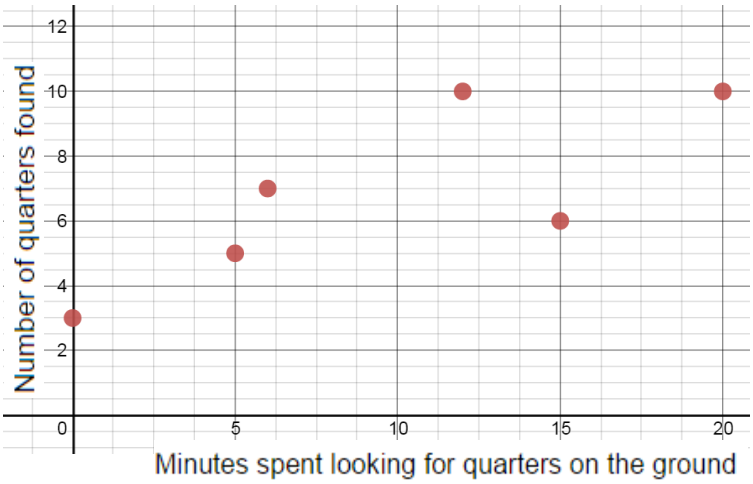
a) What is the line of best fit?

b) What does the slope mean in the context of this situation?

c) What does the y intercept mean in the context of this situation?

---

3) Use the scatter plot to the right to answer the following questions.



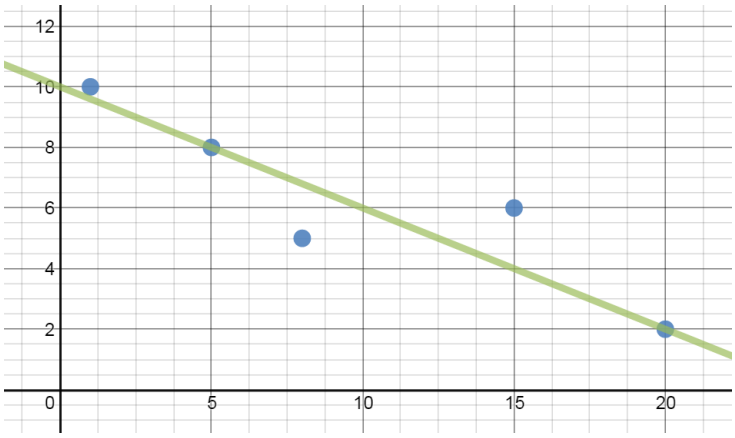
a) Draw the line of best fit.

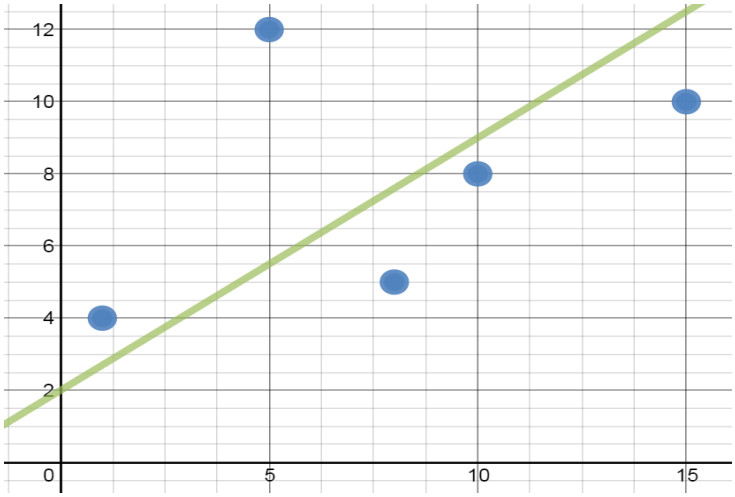
b) Draw both the upper and lower bounds.

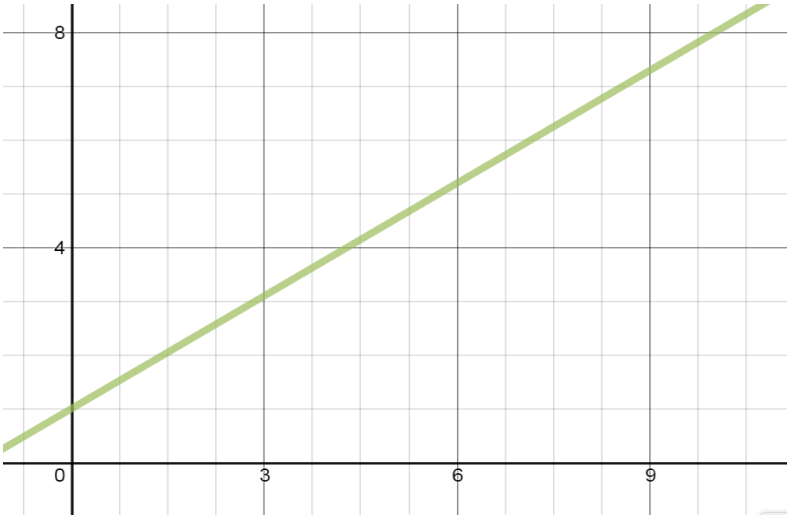
c) Predict the amount of quarters someone would have if they spent 10 minutes using your line of best fit.

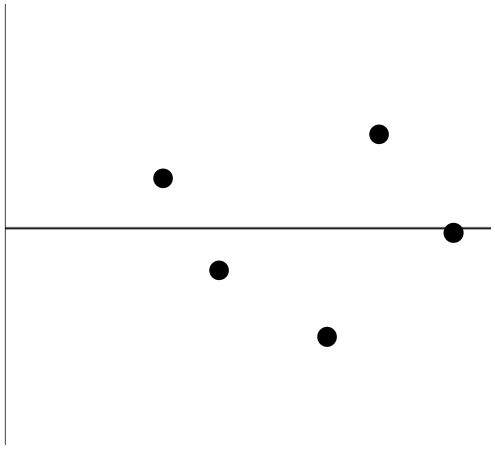
d) Predict the amount of quarters someone would have if they spent 10 minutes using your upper and lower bound

Make the residual plots for the following scatterplots.





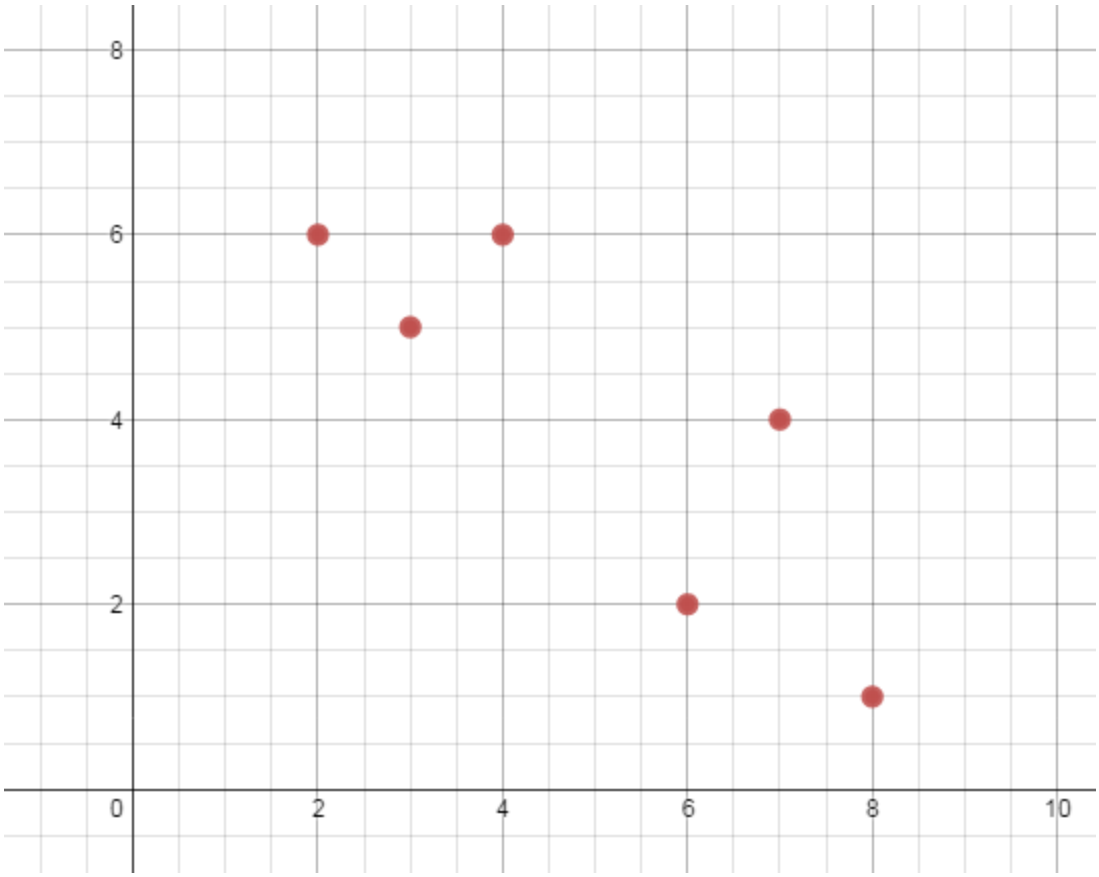




Place the dots that match this residual plot on the scatterplot to the left

1. Describe the directions, shape and strength of the scatterplot on the below

2. Draw the line of best fit, what is the equation for the line you drew?



Cecilia is trying to track the amount of cars that pass by her a house that she wants to purchase. She observes a little bit each day and sees how many cars pass by. After 7 days she made the following table.

Minutes Cecilia Watched	Cars that passed by
15	40
10	25
5	14
2	12
1	3
6	14
9	24

3. Find the line of best fit from the following table.

4. What does the slope mean in your equation?

5. Using your line of best fit, how long would it take for 100 cars to pass by?

6. Using your line of best fit, how many cars would pass by in 22 minutes?

Teddy gathered some data about car price and fuel efficiency (miles it gets per gallon).

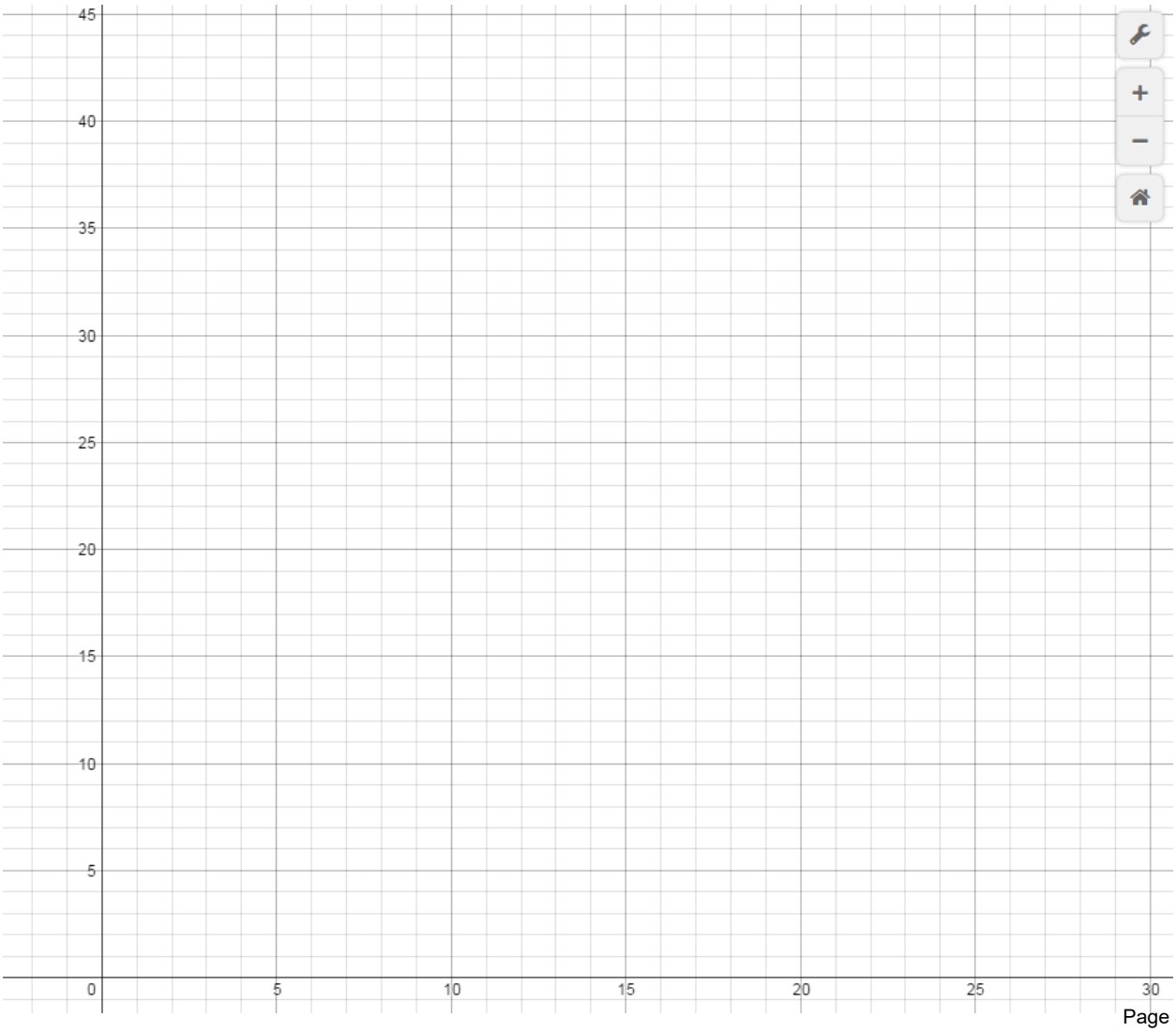
Cost of Car (in thousands)	Fuel Efficiency (miles per gallon)
5	18
10	25
13	27
20	30
23	40
25	42

7. Create a scatter plot given her data.

8. Draw and label the line of best fit, upper and lower bound line.

9. Using the upper and lower bound estimate how many miles per gallon a car would get that costs $8,000.

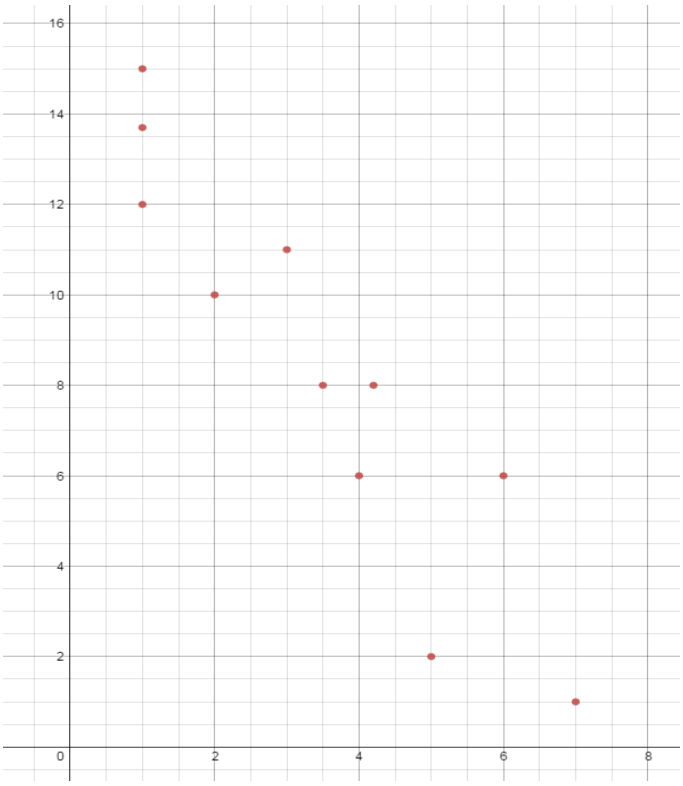
10. What does your y-intercept mean in this situation?



11) Sarah heard that there is a strong correlation between the amount of sunscreen sold and the amount of drownings that occur. To her it is clear that there must be something in sunscreen that makes her sink. She proposes a new law to combat the effects of this evil product for the good of the community. Is she justified in her thinking? Why or why not? Include any lurking variables in your answer.

12) Create a scatter plot with negative weak correlation. Come up with two variables that fit that correlation.

1. Describe the direction, shape and strength of the scatterplot on the right.



2. Draw the line of best fit, what is the equation for the line you drew?

Days	Amount Saved
10	584
16	832
25	1270
30	1397
37	1699
50	2260
55	2574

Simon was tracking his bank account throughout the summer and came up with the following data.

3. Find the line of best fit from the following table.

4. Using your line of best fit, what would you expect him to save by the 73rd day?

5. Using your line of best fit, when would you expect his savings to reach 10,000 dollars?

Sarah runs a shipping company and has gathered information based on the length of boxes and their overall weight.

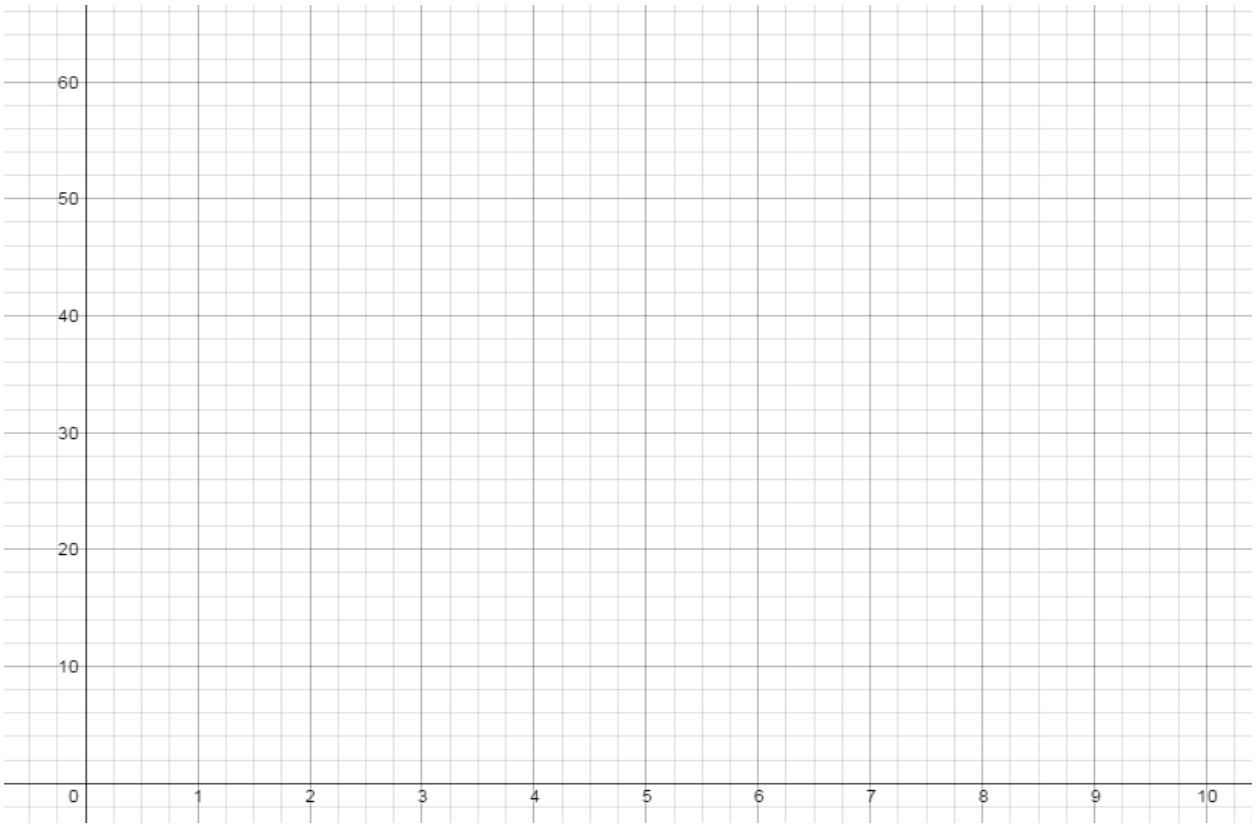
6. Create a scatter plot given her data.

Length (Feet)	Weight (Pounds)
4	26
2	13
5	26
4	22
1	6
8	37
9	47
6	35
5	22
7	33
2	8
10	52

7. Draw the line of best fit, upper and lower bound line.

8. Using the upper and lower bound estimate how much a package would weigh if its length was 3 feet.

9. What does your slope mean in this situation?



10) Circle all for the questions that would produce categorical data and put a box around all the questions that produce quantitative data.

a) How many hours do you sleep a night?

b) What is the thread count of your sheets?

c) What color are your sheets?

d) How many mistakes do you make on your tests after getting less than 5 hours of sleep?

e) How many times do you wake up a night?

f) Did you eat breakfast before your test?

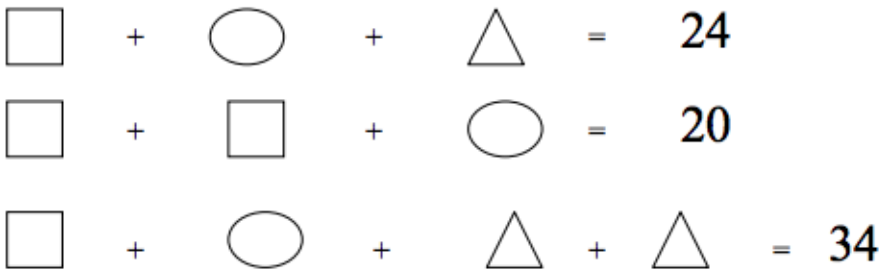
Challenge Problems

1. The average of four numbers is 10. You add a fifth number and the average becomes 12. What number did you add?

2. The average of four numbers is 10. You change one of the numbers to 30 and the average is now 12. What was the number before you changed it?

3. At the first meeting of the House of Eccentricities in the government of the Gnomes, each member shook hands with each other member. There are 25 members of the House. How many handshakes took place?

4. In the 3 equations below, each of the shapes represents a number. Each shape represents the same number everywhere it occurs. What is the numerical value of the square?



5.

Player Name	1st Place	2nd Place	3rd Place	4th Place
Player A				
Player B				
Player C				
Player D				

A group of 4 soccer players, each from a different team, stopped at the same restaurant to eat after a tournament. They began to discuss how their teams did at the tournament. Determine what order their teams placed in the 4-team tournament (1st through 4th place) based on the discussion below. (Note: The players from the 1st and 2nd place team did not want to brag about their victories, so the statements they made are NOT truthful!)

Player A: "We did not come in last place."
Player B: "We did not come in 1st place."
Player C: "We came in 3rd place."
Player D: "My team did not come in 3rd place."

7. ABRAHAM WALD'S MEMO

Abraham is tasked with reviewing damaged planes coming back from sorties over Germany in the Second World War. He has to review the damage of the planes to see which areas must be protected even more.

Abraham finds that the fuselage and fuel system of returned planes are much more likely to be damaged by bullets or flak than the engines. What should he recommend to his superiors?

Record your answer here then google the answer:

Two Variable Stats

Assignment 1: Finding the line of best fit from a scatter plot and a table.

1) Y= -2.42x + 5058.3 2) 170 3)y=.124x + 2.53 4)The longer the pencil, the heavier it will be, or the shorter the pencil, the less it will weigh 5) possibly has removable eraser, error in calculation… 6) The non-painted part of the pencil weighs 2.53 including outlier or 1.37 without the outlier

Assignment 2: Correlation of Scatter Plots

Graphs- hair v Intelligence- no correlation, texting v Friendship- weak positive, purchase over time- strong positive

1) Strong positive 2) No correlation 3)weak negative 4)strong negative 5)Strong negative 6)strong negative 7)8)9) answers will vary 10)strong negative, "number of pirates/ global average temperature decreases", yes-according to the data although neither of them have anything to do with the other., Time-there were more pirates back in time when global temperature was cooler. 11)a. Strong positive b."Methodist Ministers/ number of imported bottles of rum c. No, correlation doesn't equal causation. D. The population of New England growing between the 1860s and 1940 cause both to rise.

Assignment 3: Residuals

1) y= 8.54x - 466.57 2) 182.47 3) 15.53 4) strong positive- the longer the board, the longer the it took 5) a. 850 b. -271, 250 c. Smallville

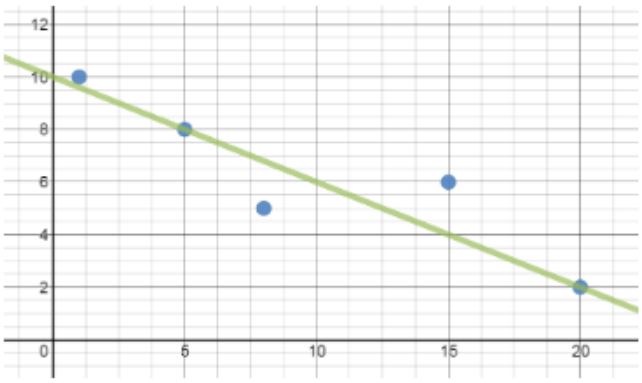
Assignment 4: Two Variable Stats in Context

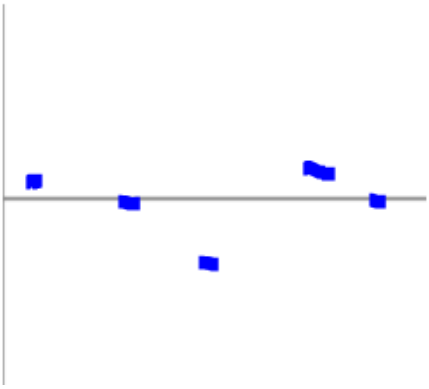
1a) y = 13.2x+ 153.6 1b) Use your line to create a table on your calculator
1c) 13 grams: 320 - 325.2 = -5.2 28 grams: 560-523.2 = 36.8
2a) The starting temp of the pizza when it is removed from the oven.
2b) How many degrees it cools every minute
3a) Strong Positive 3b) (0,350) If you don't study the line would predict you would score a 350. 3c) 25, every hour that you study your score increases by 25 points.
3d) No, he used a single data point to make his prediction. He should use the line of best fit.
3e) 595-550 = 45 4a) y = -7.18x + 1010 4b) (0,1010) If it is zero degrees then expect 1010 people to show up. 4c) 7 less people show up for every degree it gets warmer. 4d) 69.6 degrees 4e) 292 people.
5a) The box they weigh the bananas in must weigh 50 pounds. 5b) Each banana weighs 0.58 pounds. 5c) Roughly 86 bananas. 5d) 2950 pounds
Review: y = 13x + 81

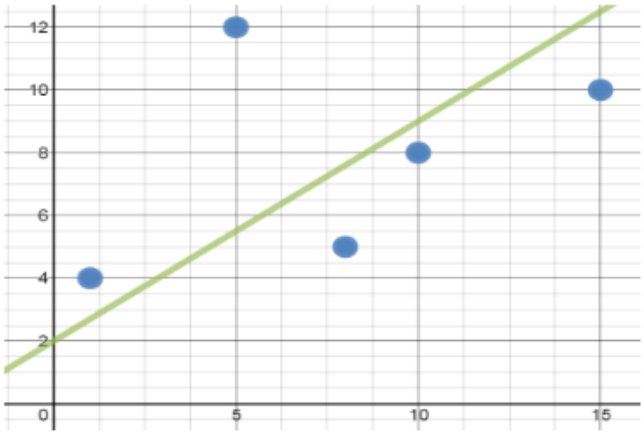
Assignment 5: Residual Plots & Upper and Lower Bound

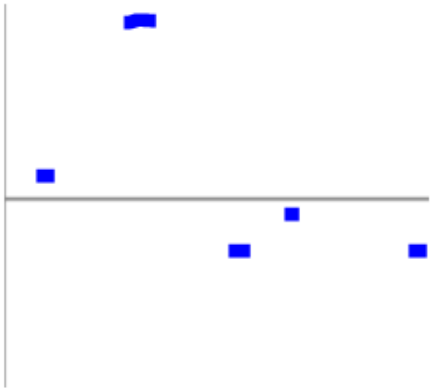
1) Every foot in height adds 22.7 pounds 2a) y = -2.89x + 20.69
2b) You trap 2.88 mice a night 2c) There were between 20 and 21 mice originally

Two Variable Stats









Assignment 6: Two Variable Stats Test Review (PART 1)

1) Shape: Linear Direction: Negative Strength: Weak

2) y = -4/5x+ 8 , but slope and y-intercept may vary.

3) y = 2.393x + 2.448 4) Every minute 2.393 cars go by.

5) 40.77 minutes 6) 55.094 cars

10) A car that is free would get that gas mileage

11) You wear sunscreen and swim more often when it is hot. The lurking variable is that these events are correlated with summer.

Assignment 7: Two Variable Stats Test Review (PART 2)

1) Shape: Linear Direction: Negative Strength: Strong

2) y = -2x+ 15 , but slope and y-intercept may vary. 3) y= 43.12x+142.85

4) $3290.61 5) 228.589 days 9) Every foot that gets added increases that package by that much.

10) a,b,d,e should have a box c,f should be circled

One of the first Directors of the House of Wisdom in Bagdad in the early 9th Century was an outstanding Persian mathematician called **Muhammad Al-Khwarizmi.**

The word “algorithm” is derived from the Latinization of his name, and the word "algebra" is derived from the Latinization of "al-jabr", part of the title of his most famous book, in which he introduced the fundamental algebraic methods and techniques for solving equations.
Perhaps his most important contribution to mathematics was his strong advocacy of the Hindu numerical system, which Al-Khwarizmi recognized as having the power and efficiency needed to revolutionize Islamic and Western mathematics. The Hindu numerals 1 - 9 and 0 - which have since become known as Hindu-Arabic numerals - were soon adopted by the entire Islamic world. Later, with translations of Al-Khwarizmi’s work into Latin by Adelard of Bath and others in the 12th Century, and with the influence of Fibonacci’s “Liber Abaci” they would be adopted throughout Europe as well.

Al-Khwarizmi’s other important contribution was algebra, a word derived from the title of a mathematical text he published in about 830 called “Al-Kitab al-mukhtasar fi hisab al-jabr wa'l-muqabala” (“The Compendious Book on Calculation by Completion and Balancing”). Al-Khwarizmi wanted to go from the specific problems considered by the Indians and Chinese to a more general way of analyzing problems, and in doing so he created an abstract mathematical language which is used across the world today.

His book is considered the foundational text of modern algebra, although he did not employ the kind of algebraic notation used today (he used words to explain the problem, and diagrams to solve it). But the book provided an exhaustive account of solving polynomial equations up to the second degree, and introduced for the first time the fundamental algebraic methods of “reduction” (rewriting an expression in a simpler form), “completion” (moving a negative quantity from one side of the equation to the other side and changing its sign) and “balancing” (subtraction of the same quantity from both sides of an equation, and the cancellation of like terms on opposite sides).

In particular, Al-Khwarizmi developed a formula for systematically solving quadratic equations (equations involving unknown numbers to the power of 2, or x^2) by using the methods of completion and balancing to reduce any equation to one of six standard forms, which were then solvable. He described the standard forms in terms of "squares" (what would today be "x^2"), "roots" (what would today be "x") and "numbers" (regular constants, like 42), and identified the six types as: squares equal roots ($ax^2 = bx$), squares equal number ($ax^2 = c$), roots equal number ($bx = c$), squares and roots equal number ($ax^2 + bx = c$), squares and number equal roots ($ax^2 + c = bx$), and roots and number equal squares ($bx + c = ax^2$). Al-Khwarizmi is usually credited with the development of lattice (or sieve) multiplication method of multiplying large numbers, a method algorithmically equivalent to long multiplication. His lattice method was later introduced into Europe by Fibonacci.

In addition to his work in mathematics, Al-Khwarizmi made important contributions to astronomy, also largely based on methods from India, and he developed the first quadrant (an instrument used to determine time by observations of the Sun or stars), the second most widely used astronomical instrument during the Middle Ages after the astrolabe. He also produced a revised and completed version of Ptolemy's “Geography”, consisting of a list of 2,402 coordinates of cities throughout the known world.

From: https://www.storyofmathematics.com/islamic_alkhwarizmi.html