4.5 NOTES - Analyzing Lines of Best Fit

LESSON 4.5

Analyzing Lines of Best Fit

LEARNING GOALS:

1) Interpolate & extrapolate data.

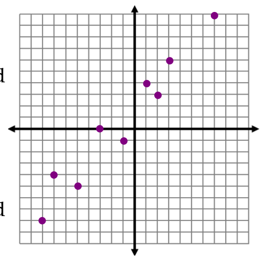
2) Calculate residuals to determine if an equation is a good model for a set of data.

Common Core State Standards

HSF-LE.B.5, HSS-ID.B.6a, HSS-ID.B.6b, HSS-ID.B.6c, HSS-ID.C.7, HSS-ID.C.8, HSS-ID.C.9

LESSON 4.5 - Analyzing Lines of Best Fit

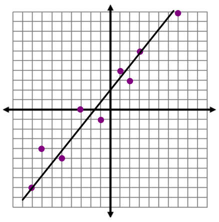
- Approximating a value between two known values is called **INTERPOLATION**
- Predicting a value outside of a range of known values is called **EXTRAPOLATION**
- For example, if you were to guess the value of y at $x = 5$ for this set of data, you would be *interpolating*.
- If you were to guess the value of y at $x = 20$ or $x = -913$ for this set of data, you would be *extrapolating*.



- The further away from the known values the point is, the less confidence you can have about the accuracy of the extrapolation.

INTERPOLATING & EXTRAPOLATING

- Given a set of a data and an equation for the line of best fit for the data, simply plug the value you are trying to interpolate or extrapolate into the equation and simplify.
- For example, let the data at the right be modeled by the equation $y = \frac{5}{4}x + 2$.



- Approximate the value of y when $x = 5$.

$y = \frac{5}{4}(5) + 2$

$y = 6.25 + 2$

$y = 8.25$

- Predict the value of y when $x = 20$.

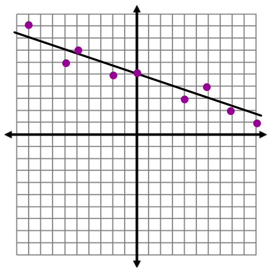
$y = \frac{5}{4}(20) + 2$

$y = 25 + 2$

$y = 27$

INTERPOLATING & EXTRAPOLATING

- Let the data at right be modeled by the equation $y = -\frac{1}{3}x + 5$.



- Approximate the value of y when $x = 2$.

$y = -\frac{1}{3}(2) + 5$

$y = -0.67 + 5$

$y = 4.33$

- Predict the value of y when $x = -138$.

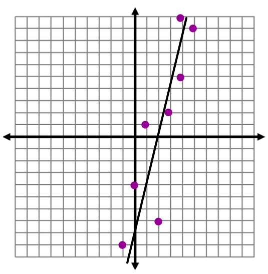
$y = -\frac{1}{3}(-138) + 5$

$y = 46 + 5$

$y = 51$

INTERPOLATING & EXTRAPOLATING

- Let the data at right be modeled by the equation $y = 4x - 8$.



- Approximate the value of x when $y = 7$.

$7 = 4x - 8$

$15 = 4x$

$x = 3.75$

- Predict the value of x when $y = 55$.

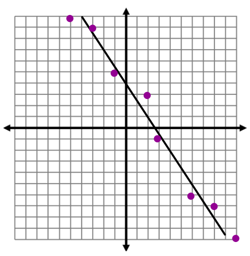
$55 = 4x - 8$

$63 = 4x$

$x = 15.75$

INTERPOLATING & EXTRAPOLATING

- Let the data at right be modeled by the equation $y = -\frac{3}{2}x + 4$.



- Approximate the value of x when $y = -4$.

$-4 = -\frac{3}{2}x + 4$

$-8 = -\frac{3}{2}x$

$x = 5.33$

- Predict the value of y when $x = -24$.

$y = -\frac{3}{2}(-24) + 4$

$y = 36 + 4$

$y = 40$

4.5 NOTES - Analyzing Lines of Best Fit

RESIDUALS

- We can use **RESIDUALS** to determine if a line of best fit models a set of data well or not.
- A **RESIDUAL** is the difference of the *y*-value of a data point and the corresponding *y*-value found using the line of best fit.
- Residuals can be positive, negative, or zero.
- A scatter plot of the residuals of a set of data can show how well a model (equation) fits the data set.
- If the model is a good fit, then the absolute values of the residuals will be relatively small.
- Also, the residuals will be more or less evenly dispersed around the horizontal axis.
- If the model is NOT a good fit, then the absolute values of the residuals will be relatively large.
- Also, the residuals will form a nonlinear shape suggesting that the model is not a good fit.
- If the residuals are wildly scattered around the graph, the original data may have no correlation at all.

CALCULATING RESIDUALS

When given a set of data and an equation that models the data, follow these steps to calculate the residuals:

STEP 1- Plug the *x*-values of the data points into the model. This will yield the *y*-values of model equation.

STEP 2- Subtract the *y*-values from the original data set from the *y*-values you found in Step 1.

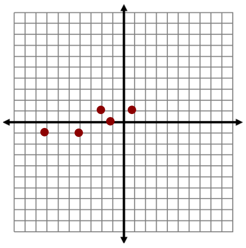
STEP 3- Graph the points formed by the *x*-values of the original data set and the residuals as the *y*-values.

STEP 4- Analyze the scatter plot formed in Step 3 to determine if the equation is a good model for the data.

Calculate the residuals of the given set of data, then graph to determine if the given equation is a good model for the data.

$$y = 2x + 7$$

x	y	y_1	Residual	Point
-7	-6	-7	-1	(-7, -1)
-4	-2	-1	1	(-4, 1)
-2	2	3	1	(-2, 1)
-1	5	5	0	(-1, 0)
1	8	9	1	(1, 1)

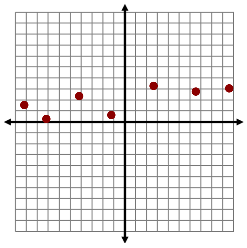


The residuals are small values, closely grouped around the x-axis in a more or less horizontal line. Therefore, the given line of best fit is a good model for the set of data.

Calculate the residuals of the given set of data, then graph to determine if the given equation is a good model for the data.

$$y = \frac{2}{5}x + 3$$

x	y	y_1	Residual	Point
-9	-2	-0.6	1.4	(-9, 1.4)
-7	0	0.2	0.2	(-7, 0.2)
-4	-1	1.4	2.4	(-4, 2.4)
-1	2	2.6	0.6	(-1, 0.6)
3	1	4.2	3.2	(3, 3.2)
7	3	5.8	2.8	(7, 2.8)
10	4	7	3	(10, 3)

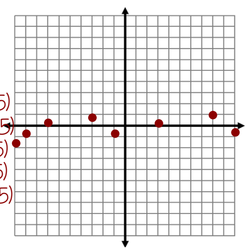


Some of the residuals are rather large and the points are not evenly dispersed around the x-axis, therefore the given line of best fit is not a very good model for the set of data.

Calculate the residuals of the given set of data, then graph to determine if the given equation is a good model for the data.

$$y = -\frac{1}{4}x - 5$$

x	y	y_1	Residual	Point
-10	-1	-2.5	-1.5	(-10, -1.5)
-9	-2	-2.75	-0.75	(-9, -0.75)
-7	-3	-3.25	0.25	(-7, 0.25)
-3	-5	-4.25	0.75	(-3, 0.75)
-1	-4	-4.75	-0.75	(-1, -0.75)
3	-6	-5.75	0.25	(3, 0.25)
8	-8	-7	1	(8, 1)
10	-7	-7.5	-0.5	(10, -0.5)

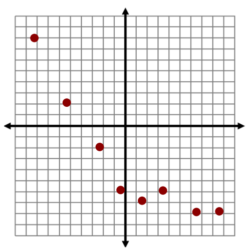


The residuals are small values, closely grouped around the x-axis in a more or less horizontal line. Therefore, the given line of best fit is a good model for the set of data.

Calculate the residuals of the given set of data, then graph to determine if the given equation is a good model for the data.

$$y = -x - 1$$

x	y	y_1	Residual	Point
-8	-1	7	8	(-8, 8)
-5	2	4	2	(-5, 2)
-2	3	1	-2	(-2, -2)
0	5	-1	-6	(0, -6)
2	4	-3	-7	(2, -7)
4	1	-5	-6	(4, -6)
7	0	-8	-8	(7, -8)
10	-3	-11	-8	(10, -8)



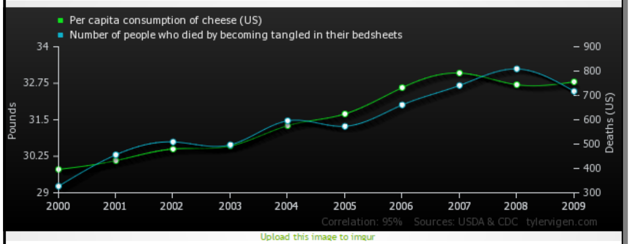
Most of the residuals are rather large and the points are not evenly dispersed around the x-axis, and they are not even close to forming a horizontal line, therefore the given line of best fit is definitely not a very good model for the set of data. In fact, it is possible the original data had no correlation at all.

4.5 NOTES - Analyzing Lines of Best Fit

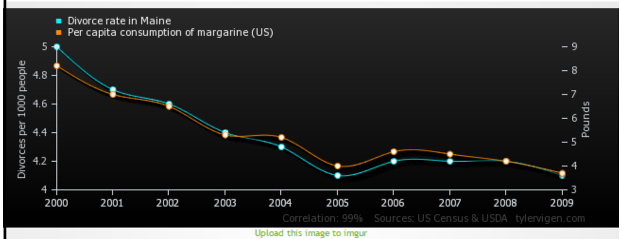
CORRELATION -VS- CAUSATION

- **CORRELATION** is a relationship between two sets of data.
- **CAUSATION** is when a change in one variable causes a change in another variable.
- Causation produces a strong correlation between two variables. However, the opposite is NOT true.
- In other words, correlation doesn't imply causation.

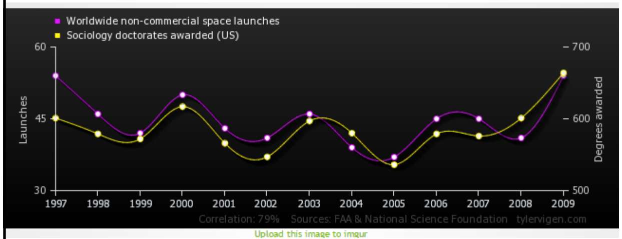
CORRELATION -VS- CAUSATION



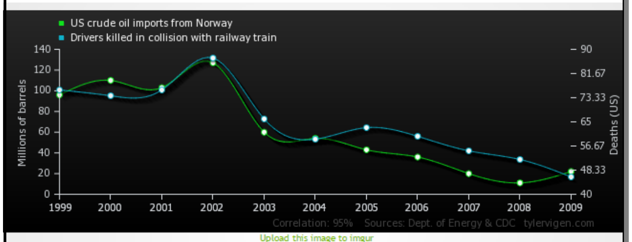
CORRELATION -VS- CAUSATION



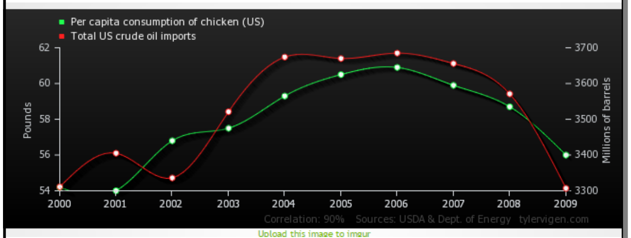
CORRELATION -VS- CAUSATION



CORRELATION -VS- CAUSATION



CORRELATION -VS- CAUSATION



4.5 NOTES - Analyzing Lines of Best Fit

Decide if a correlation is likely in the situation. If so, tell whether there is a causal relationship.

A) Time spent exercising and the number of calories burned.

Correlation is likely. It will be a causal relationship, because exercising causes you to burn calories.

B) Number of gas stations in a city and the population of the city.

Correlation is likely. It will NOT be a causal relationship, because having many gas stations doesn't cause the population of a city to increase.

C) Number of hats you own and the size of your head.

Correlation is not likely.

LABELING AXES OF GRAPHS

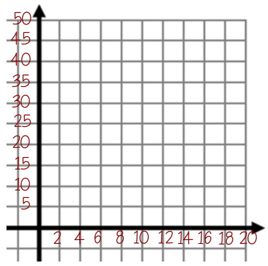
- Oftentimes, the data points we are given lie outside of the -10 to 10 area that we typically graph in. We must chose an increment for our axes to go by to fit the data on the graph.
- The increment you choose for the x-axis does NOT have to be the same as the y-axis.

EXAMPLE: If our x-values range from 1 to 20, what increment should we use for the x-axis?

2

And if our y-values range from 3 to 42, what increment should we use for the y-axis?

5



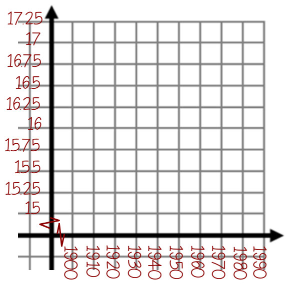
LABELING AXES OF GRAPHS

EXAMPLE: If our x-values range from 1907 to 1984, what increment should we use for the x-axis?

10, with a break between 0 and 1900

And if our y-values range from 15 to 17, what increment should we use for the y-axis?

0.25, with a break between 0 and 15

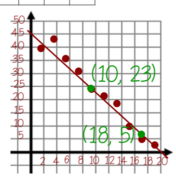


The data shows the number of boxes of chocolate y that a store sold x days before Valentine's Day.

x	2	4	6	8	10	12	14	16	18	20
y	38	42	35	30	23	20	17	8	3	1

A) Graph the data.

B) Sketch in a line of best fit.



C) Find the equation of the line.

$m = \frac{23 - 5}{10 - 18} = \frac{18}{-8} = -2.25$

$y - 23 = -2.25(x - 10)$

$y - 23 = -2.25x + 22.5$

$y = -2.25x + 45.5$

D) What does the point (8, 30) represent?

8 days before Valentine's Day, the store sold 30 boxes of chocolates.

E) How many boxes of chocolate do you think the store sold during the 5th week?

$y = -2.25(5) + 45.5$

$y = 34.25$

They sold about 34 boxes of chocolates.

F) In what week do you think the store sold about 25 boxes of chocolate?

$25 = -2.25x + 45.5$

$x = 9.1$

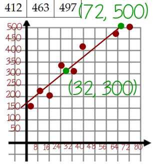
They sold 25 boxes of chocolates during the 9th week.

The data shows the number of strands of Christmas lights families hung on their houses x and the amount of money they paid for their electricity bill in December y

x	5	12	19	27	36	42	66	76
y	150	214	198	333	301	412	463	497

A) Graph the data.

B) Sketch in a line of best fit.



C) Find the equation of the line.

$m = \frac{500 - 300}{72 - 32} = \frac{200}{40} = 5$

$y - 300 = 5(x - 32)$

$y - 300 = 5x - 160$

$y = 5x + 140$

D) What does the point (19, 198) represent?

With 19 strings of Christmas lights, the family will pay $198 for electricity.

E) How much do you think the electric bill would be if a family hung 55 strings of Christmas lights?

$y = 5(55) + 140$

$y = 415$

The family would spend $415 on electricity.

F) How many strings of Christmas lights do you think a family hung up if their electric bill was $800?

$800 = 5x + 140$

$x = 132$

The family would have hung 132 strings of Christmas lights if their electric bill was $800.

HOMEWORK

4.5 Worksheet - Interpreting Data