COAL – Guided Notes Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Correlation and Causation

**S.ID.7** Interpret the slope (rate of change) and the intercept (constant term) of a linear model in the context of the data.

**S.ID.8** Compute (using technology) and interpret the correlation coefficient of a linear fit.

**S.ID.9** Distinguish between correlation and causation.

EQ: What is the difference between correlation and causation?

**Warmup** - In the following situation, interpret the slope and y-intercept in context. Make sure to explain whether or not the y-intercept is relevant.

The Jones family is renting a car. The function y = 2x + 75 for finding the cost of renting the car is based on the # of miles driven.

slope: y-intercept:

**First…what is correlation and causation?**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is the relationship between two events, where a change in one event is related to a change in the second event.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is a relationship between two events, where a change in one event is responsible for a change in the second event.

**CORRELATION DOES NOT MEAN CAUSATION!!!**

Examples of correlation: Examples of causation:

 Height and shoe size Weight and pant size

 SAT scores and college achievement SAT scores and college acceptance

 GPA and hours of television watched

What do you think? Education and years in jail Education and salary

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (\_\_\_\_) – a measure that shows how well data are modeled by a linear equation, In other words, it helps us determine how strongly two events are related.

-1 ≤ r ≤ 1

when r is close to -1, data has a negative correlation (negative slope)

when r = 0, the data have no correlation (not related)

when r is close to 1, data has a positive correlation (positive slope)

 \*\*r is NOT the value of the slope, it simply gives the direction of the slope\*\*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Strong($\left|.8-.99\right|$) | Moderate($\left|.4-.79\right|$) | Weak(below $\left|.4\right|$) | No Correlation  |
| Positive0 ≤ r ≤ 1 |  |  |  | \*\* Perfect Correlation r = 1 |
| Negative-1 ≤ r ≤ 0 |  |  |  |

Guided Example:

An internet-based company launched a new television commercial and recorded the number of new users each week after the commercial aired. We are told r = .99.

|  |  |
| --- | --- |
| Number of weeks | New users in thousands |
| 1 | 6 |
| 2 | 9 |
| 3 | 15 |
| 4 | 19 |
| 5 | 19 |
| 6 | 22 |
| 7 | 32 |
| 8 | 31 |
| 9 | 37 |
| 10 | 45 |
| 11 | 44 |

Step 1: Describe the data based on the scatter plot.

 The scatter plot appears linear in shape. The number of weeks and new users appear to have a strong, positive correlation, meaning as the number of weeks increase, the number of new users increases.

Step 2: Support your conclusion using the correlation coefficient (r).

 The correlation coefficient is .99. Because r is positive and very close to one, r does support our conclusion that the number of weeks and number of new users has a strong, positive correlation. This does NOT tell us that the number of weeks CAUSES the new users to increase, just that they are related.

What if r had not been given to you?

Calculator steps for calculating the line of best fit and r:

|  |  |
| --- | --- |
| TI-30XS Multiview | TI-30XIIS |
| Hold “on” and press “clear” to clear the memoryPress “data” keyEnter x values in L1 (press “enter” after each value)Right arrow over to L2Enter y values in L2 (press “enter” after each value)Press “2nd” and “data” Go to 2: 2-var stats and press enterX-data should be highlighted as L1, y as L2Arrow down to “calc” and press enterArrow down and record a, b, and r. Round a/b to 1 decimal place; r to 3 decimal places | Hold “on” and press “clear” to clear the memoryPress “2nd” and “data” key arrow to 2-Var press enterPress “data” keyX1 = appears; enter your first x-valueArrow down and Y1 = appears; enter your first y-valueArrow down to enter X2, Y2, etc.After entering the last y-value be sure to arrow down Press “2nd” and “data” keyArrow over to 2-var stats and press “enter”Press “statvar” key Arrow to the right and record a, b, and r.Round a/b to 1 decimal place; r to 3 decimal places |

Use the instructions to calculate the line of best fit and the true r from our guided example.

Line of Best Fit: y = ax + b Correlation Coefficient

 a: 4 and b: 1.5, so y = 4x + 1.5 r: .986

Interpret the slope and y-intercept in context: As the weeks increase by 1, the new users increase by about 4000. The starting number of users is 1500. This is relevant because the company has probably has users before the commercial airs.

You try:

The following table records the amount of time a customer waits in line and the number of people in line ahead of that customer. Describe the relationship between waiting time and people ahead of customer in line.

|  |  |
| --- | --- |
| People ahead of customer | Minutes waiting |
| 1 | 10 |
| 2 | 21 |
| 3 | 32 |
| 5 | 35 |
| 8 | 42 |
| 9 | 45 |
| 10 | 61 |

Using the calculator, find the line-of-best fit. y = ax + b

Interpret the slope and y-intercept.

Using the calculator, find the correlation coefficient.

r = \_\_\_\_\_\_\_\_\_\_\_

Using your scatter plot and correlation coefficient, describe the data.

In Summary:

Fill in the blank with increase or decrease.

1. In a positive correlation, as the x values increase, the y values \_\_\_\_\_\_\_\_\_\_\_\_.

2. In a positive correlation, as the x values decrease, the y values \_\_\_\_\_\_\_\_\_\_\_\_.

3. In a negative correlation, as the x values increase, the y values \_\_\_\_\_\_\_\_\_\_\_\_.

4. In a negative correlation, as the x values decrease, the y values \_\_\_\_\_\_\_\_\_\_\_\_.

COAL – Homework Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Correlation and Causation

I. Describe the correlation in the following scatterplots: (perfect, weak, moderate, strong, no correlation). Be sure to include positive / negative.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

II. A researcher wants to see if there is a relationship between a person’s income (in thousands) and the percent of their income they give to charity.



|  |  |
| --- | --- |
| Income (Thousands of $) | Percent given to charity |
| 30 | 5 |
| 20 | 4 |
| 50 | 3 |
| 45 | 3 |
| 60 | 4 |
| 25 | 3 |
| 40 | 3 |
| 65 | 5 |
| 38 | 3 |
| 27 | 3 |

r = \_\_\_\_\_\_\_\_\_

Describe the data using your scatter plot and r.

III. The data below shows the number of years of experience of a car salesman and the number of cars sold each month. Sketch the scatterplot. Find the equation for the line-of-best fit using the calculator. Interpret the slope and y-intercept. Find the correlation coefficient for the data. Describe the data using your scatter plot and r. Show work!

|  |  |
| --- | --- |
| Year, x | Cars Sold, y |
| 3 | 5 |
| 9 | 14 |
| 2 | 12 |
| 5 | 21 |
| 1 | 8 |
| 4 | 15 |
| 11 | 22 |



Line-of-best fit. y = ax + b

Interpret the slope and y-intercept.

r = \_\_\_\_\_\_\_\_\_\_\_

Using your scatter plot and correlation coefficient, describe the data.