Correlation vs Causation Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Go to PhET simulations

Math menu

Least squares regression

Custom/ check grid lower right hand corner

1. a. Using the orange circles drag points onto the grid

(3,5), (6,6) (9,7) (12, 8)

b. predict **my** line of fit by using the sliders a and b on the right

y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

c. check answer by using the computers **best line** of fit on the left side

y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

How did your equations comapre? Use slope and y-intercept in your answer.

d. click on the correlation coefficient r = \_\_\_\_\_\_\_\_\_\_

1. Clear number 1 and find five points that would give a correlation close to -1

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

What is the best fit line for these points? y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Clear all lines and instead of custom find the height versus shoe size data.
2. using MY LINE feature use sliders to create your line of best fit

y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ predict an r value \_\_\_\_\_\_\_\_\_\_\_\_\_

aq

1. find the actual line of best fit y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

and r value \_\_\_\_\_\_\_\_\_\_\_\_ **clear and unclick before moving on**

1. Repeat number 3 with minimum wage versus time

a. predictions y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ r = \_\_\_\_\_\_\_\_\_\_\_\_\_

b. actual y = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ r = \_\_\_\_\_\_\_\_\_\_\_\_\_

1. By looking at the stat plots for each set of data, **predict** the r value for each and then give the actual correlation coefficient. Also hypothesize on a 3rd variable being a possible causation for each correlation.
2. spending and salaries

prediction actual

causation

1. orbital speed and distance

prediction actual

causation

1. life expectancy and TVs

prediction actual

causation

1. Come up with two variables with a possible strong positive correlation and a 3rd variable for the causation. Do no use the violence – ice cream example.

7. Come up with two variables having a strong negative correlation and a 3rd variable for the causation.

Statistics extension

Click on the gasoline price vs time data and find the best fit line. Click on the residual . By looking at the result what do you think a residual is ?