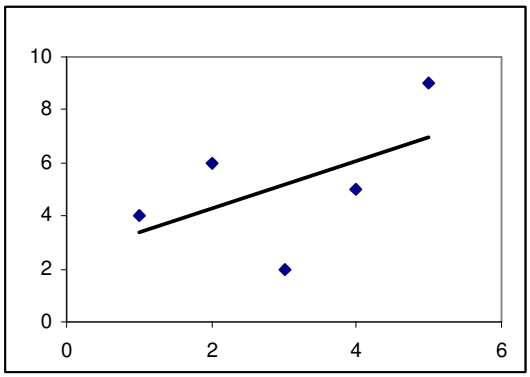
Linear Regression & Correlation Coefficient Worksheet

Name _____

Hr _____

1. Recall that the least squares line minimizes the squares of the residuals.

   The plot to the right shows 5 data points and the least squares line. Sketch and shade the squares of the residuals.



2. The table below shows the percentage of females in the U.S. labor force at various times throughout history.

Years after 1900	50	60	70	80	90	100
Percentage	29.6	33.4	38.1	42.5	45.3	52.0

   a. Enter the data into your calculator and give the equation of the least squares line. *Round slope & y-intercept to 2 decimal places.*

   b. What is the meaning of the slope? Of the y-intercept?

   c. Use your equation to predict the percentage of women working in 2010. (Be careful about what value you enter for x.)

   d. What is the residual for the point (70, 38.1)?

3. Recently, Adam put larger wheels on his skateboard and noticed it would coast farther. He decided to test this relationship and gathered the following data.

Wheel diameter (in.)	1	2	3.5	5	5.5	7	8.5	9.5	10
Rolling Distance (in.)	17	23	32	30	36	52	57	55	70

   a. Find the equation of the least squares line that models these data. *Round to 2 decimals.*

   b. Describe the real-world meaning of the slope and the y-intercept of your line.

   c. If Jimmy's skateboard has 8 inch wheels, approximately how far will it coast?

   d. Use your model (line of best fit) to determine the size wheel of a skateboard that rolls 50 inches.

4. For each research finding, decide whether there is evidence of causation, correlation or both. If it is only a correlation, name a possible lurking variable that may be the cause of the results.

a. As the sales of television sets has increased, so has the number of overweight adults. Does television cause weight gain?

b. A study in an elementary school found that children with larger shoe sizes were better readers than those with smaller shoe sizes. Do big feet make children read better?

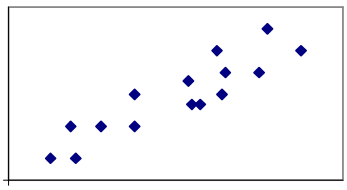
c. The more firefighters sent to a fire, the longer it takes to put out the fire. Does sending more firefighters cause a fire to burn longer?

5. *Multiple Choice.* What does it mean to say that data has a strong negative correlation?

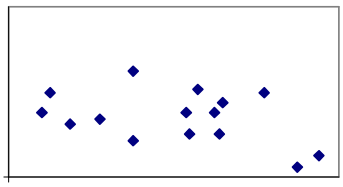
a. The variables in the data have no relationship at all.
b. A linear model is appropriate and the slope of that line is negative.
c. There is a negative causation relationship in the data.
d. One variable in the data set always has a negative value.

6. The correlation coefficients for the six scatter plots shown below are -0.85, -0.40, 0, 0.50, 0.90 and 0.99. Match each scatter plot with the correct correlation coefficient.

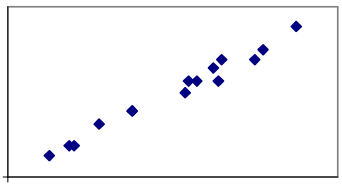
a.



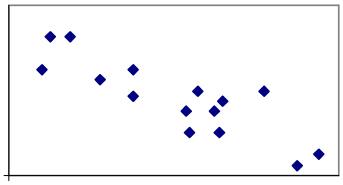
b.



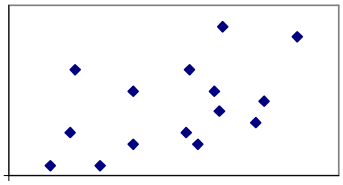
c.



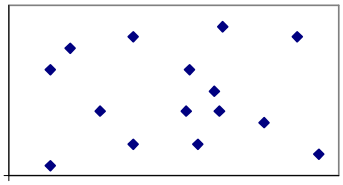
d.



e.



f.



7. Sketch the graph of a scatter plot that has a correlation coefficient of exactly 1, but the slope of the line of best fit is greater than 1.

8. Using your graphing calculator, find the correlation coefficients for the data in problem #2 and 3 from the front side of this worksheet.

#2: _____

#3: _____

With these correlation coefficients, how confident do you feel about your predictions in 2c and 3c,d? Explain.