Name: _____ Name: _____

Quiz #2 (Topics 3.1 and 3.2)

1. (5 points each) Suppose you were to collect data for the following pair of variables. You want to make a scatterplot. Which variable would you use as the explanatory variable (predictor) and which as the response variable? Give a brief explanation of why you chose them the way that you did. Additionally, discuss the likely direction of the association and explain why you think the direction would be that way.

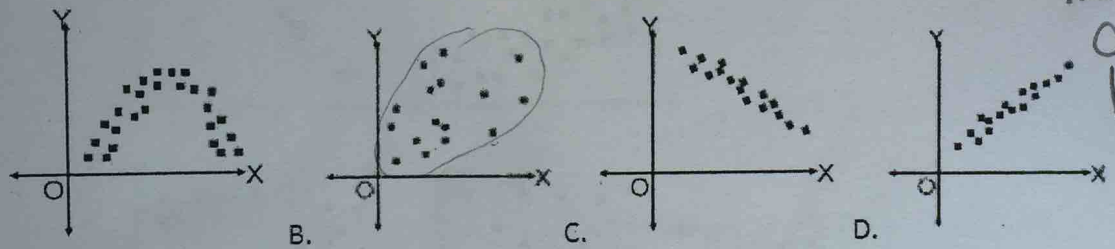
   a. Distance driven; Cost of a cab ride

   The cost of a cab ride depends on the distance driven, so the explanatory variable is the distance driven and the cost is the response variable. Positive Association - the further you go, the more it will cost.

   b. Apparent brightness; Distance from a streetlamp

   Apparent brightness depends on the distance you are from the streetlamp. The explanatory variable is distance and response is apparent brightness

2. (10 points total) Given the following Scatterplots, answer the questions below: Negative Association The further you are away, the less bright it will be.



A. B. C. D.

   a. Which scatterplot(s) show little to no association? B

   b. Which scatterplot(s) show a negative association? C

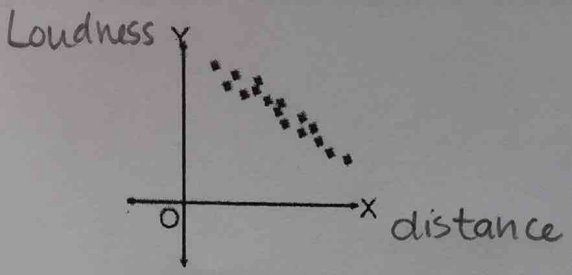
   c. Which scatterplot(s) show a positive association? D (maybe B too)

   d. Which scatterplot(s) show a strong association? A, C, D

   e. Which scatterplot(s) show a curved association? A

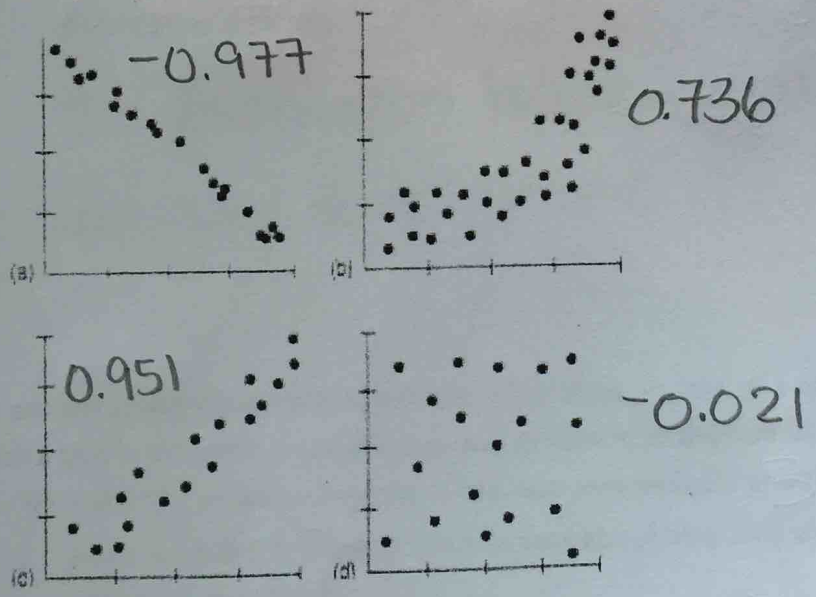
   f. Which scatterplot(s) show a linear association? C, D

3. (5 points) For the following scatterplot, make up two variables that could be described by the graph. Explain why you think so (and do NOT use the same variables as from #1 – if you do, you will receive 0 points).



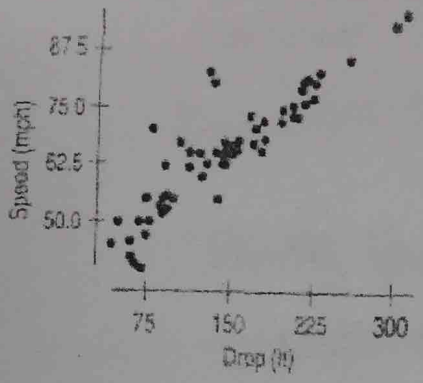
How loud a fire truck is depends on how far away from it you are. As the distance increases, the loudness decreases

4. (10 points) Match the four scatterplots below with their correct correlations: -0.977, -0.021, 0.736 and 0.951



the closer the correlation, r, is to 1 or -1, the more linear the scatterplot will appear

5. (5 points) Roller coasters get all of their speed by dropping down a steep initial incline, so it makes sense that the height of the drop must be related to the speed of the roller coaster. Here is a scatterplot of Top Speed and Height of Drop for 75 roller coasters around the world (sorry it's crooked!).



a. Does the scatterplot indicate that it is appropriate to calculate correlation? Explain.

Yes. The scatterplot looks linear

b. In fact, the correlation between Speed and Drop is 0.91. Describe the association (strength, direction and form).

The association between speed & drop is strong, positive and linear.

6. (10 points) A regression analysis of the initial drop (in feet) of a roller coaster ride and the DURATION (notice this is different from the previous problem) of the ride (in seconds) yields the model: $\hat{y} = 91 + 0.242x$, where Duration is the response variable and the Drop is the explanatory variable.

a. Explain what the slope of the line says about how long a roller coaster may last and the height of the roller coaster.

$\hat{Duration} = 91 + 0.242(Drop)$

$slope = \frac{0.242 \text{ seconds}}{1 \text{ foot}}$

An increase in initial drop of 1 foot increases the predicted duration by 0.242 seconds.

Problem 6 continues here. Here is the prompt again: A regression analysis of the initial drop (in feet) of a roller coaster ride and the DURATION (notice this is different from the previous problem) of the ride (in seconds) yields the model: $\hat{y} = 91 + 0.242x$, where Duration is the response variable and the Drop is the explanatory variable.

b. A new roller coaster advertises an initial drop of 200 feet. How long would you predict the ride to last?

$\hat{\text{Duration}} = 91 + 0.242(200) = 139.4 \text{ seconds}$

Using the regression line, we predict that a roller coaster with an initial drop of 200 feet will last 139.4 seconds.

c. If the coaster with the initial drop of 200 feet actually lasts 2 minutes, calculate the residual and explain what it means using the context of this problem.

residual = actual − estimated
= 120sec − 139.4sec
= −19.4sec

The regression line overestimates the duration of the roller coaster by 19.4 seconds.

d. Identify the y-intercept of the least squares regression line and write what it means using the context of the problem. Does the y-intercept actually make logical sense? (Yes or No?) Why or why not? (0, 91)

A ride with a 0ft initial drop will last 91 seconds. This doesn't make sense – this would be a flat ride with no bumps... doesn't sound like a "roller coaster"!