IM2 Problem Set 4.1 - Bivariate Data and Scatter Plots

BIG PICTURE of this UNIT:	• How do we analyze and then make conclusions from a data set when we collect information on two variables? • How do I describe and analyze bivariate data? • What functions can I use to model bivariate data sets? • How do I decide on the validity/reliability of my data? Of my analysis? Of my conclusions? Of my decision?

PART 1 - More Investigations with Data - Univariate vs Bivariate

In Unit 3, you reviewed how to present and analyze ***univariate*** data. So, to review, please compile the following data sets, wherein you will measure your height, your foot size and your hand span. Enter all three data points on the same line please.

a. Data Set #1: Student Heights: Enter your data on the following spreadsheet

b. Data Set #2: Student Foot Size: Enter your data on the same spreadsheet

c. Data Set #3: Student Hand Span: Enter your data on the same spreadsheet

https://docs.google.com/spreadsheets/d/1zj-KgbQ8FR2CN-RQv5uymfHD5F3ynN3ou6Rz4T9yM44/edit#gid=0

So, in Unit 3, you were asked to visualize (histogram, BW plot) a data set and then analyze the data set (mean, median, quartiles, probabilities).

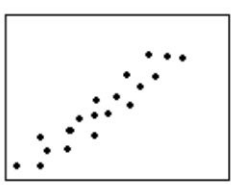
But now that we have compiled this data about you, we can ask the questions about associations or relationships, for example, is your height related to your foot size (or vice versa) or is your hand span related to your foot size. So now, we are considering bivariate data ⇒ as we now have more information about you and we can now compare 2 data sets at the same time (hence the idea of bivariate)

PART 2 - Visualizing and Describing Bivariate Data

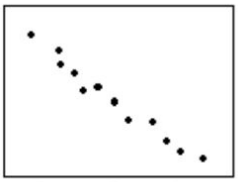
To visualize bivariate data, we will prepare scatter plots and then use the plots to start making descriptions & analysis about the data and the relationships that may/may not appear from the data sets.

A scatter plot is helpful in understanding the form, direction, and strength of the relationship between two variables. ***Correlation*** is the strength and direction of the linear relationship between the two variables.

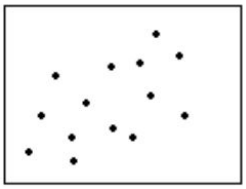
To describe the scatter plots, we will use the following terms/ideas:



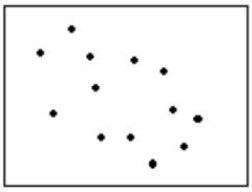
Strong Positive



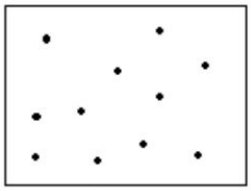
Strong Negative



Weak Positive

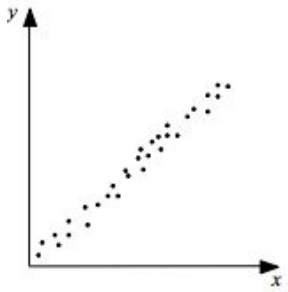


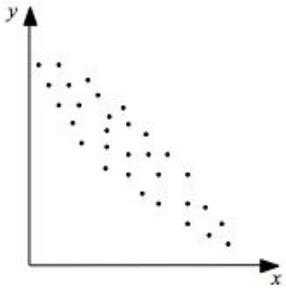
Weak Negative

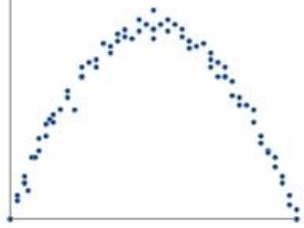


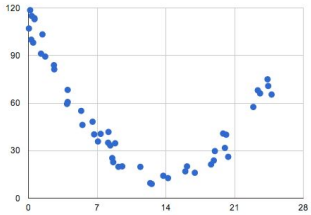
None

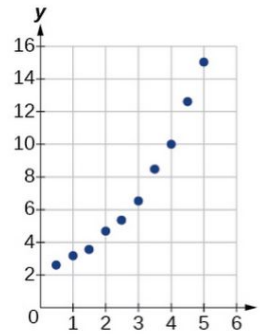
We will also distinguish between linear and non-linear relationships (and these will be the relationships we explore for this semester - linear, exponential and quadratic)

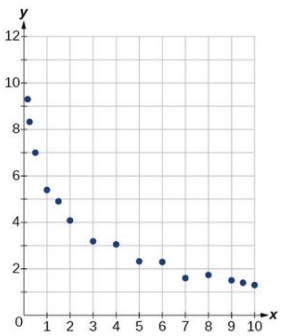


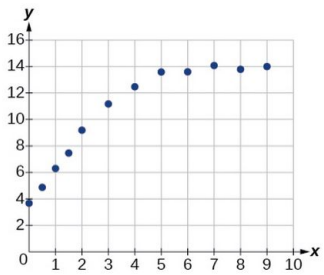


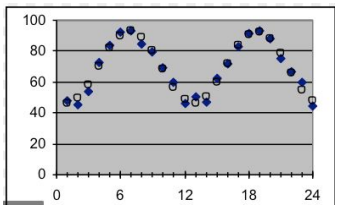








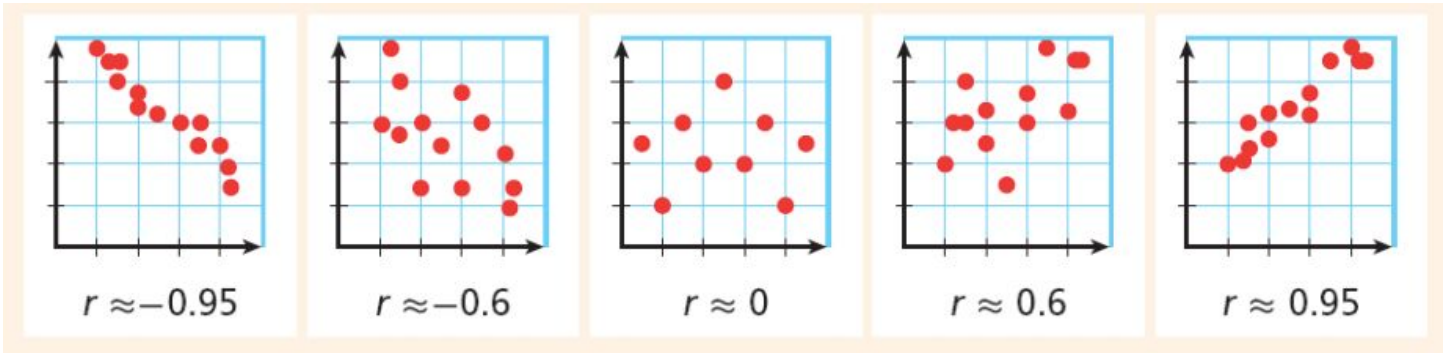


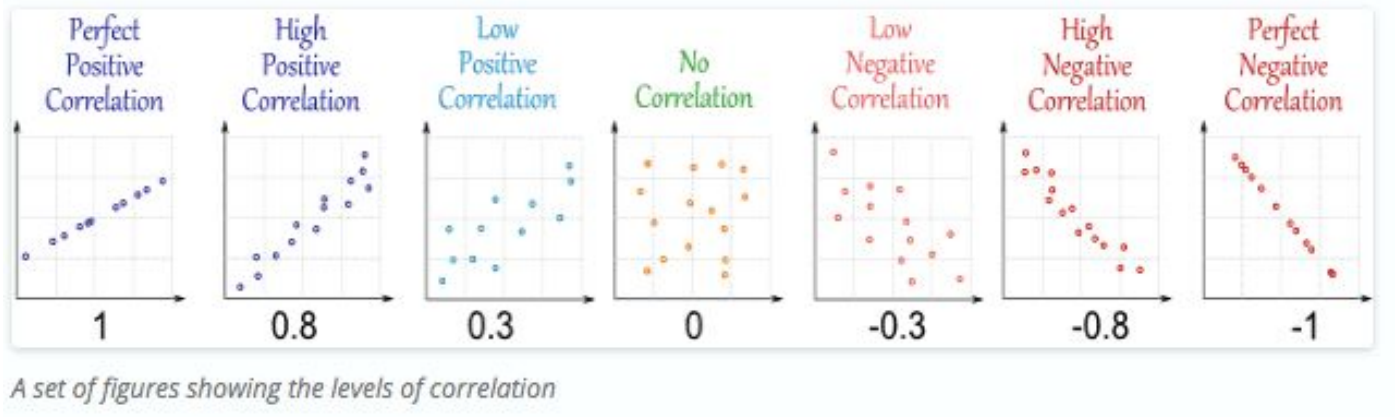


For relationships that we feel are linear, we will describe the correlation with a number ⇒ the ***correlation coefficient***, ***r*** , is a measure of how well the data set is fit by a model. In other words, how well it fits the line of best fit.

- ***r*** is a value in the range $-1 \leq r \leq 1$
- If ***r*** = 1, then the data set forms a straight line with a positive slope
- If ***r*** = -1, then the data set forms a straight line with a negative slope
- If ***r*** = 0, then the data set has no correlation

For example:



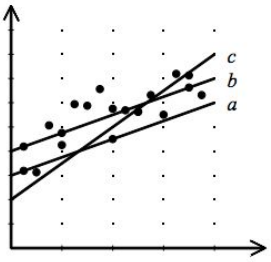


A set of figures showing the levels of correlation

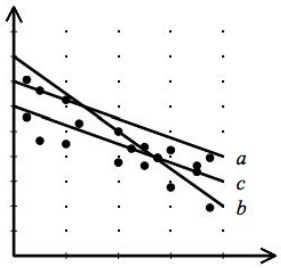
PART 3 - Working with Scatter Plots - Lines (Curves) of Best Fit

1. Which is the best trend line for the 2 scatter plots shown below?

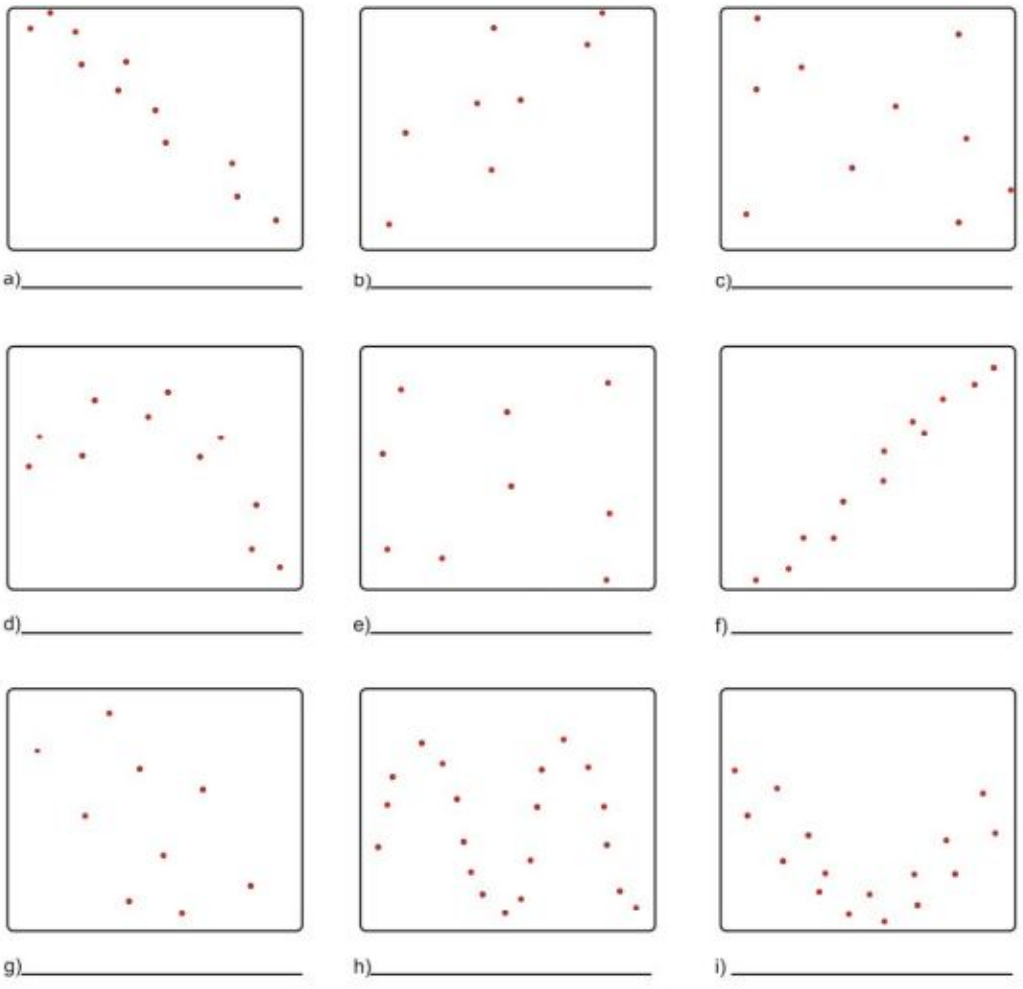
Which is the best trend line for the data shown?



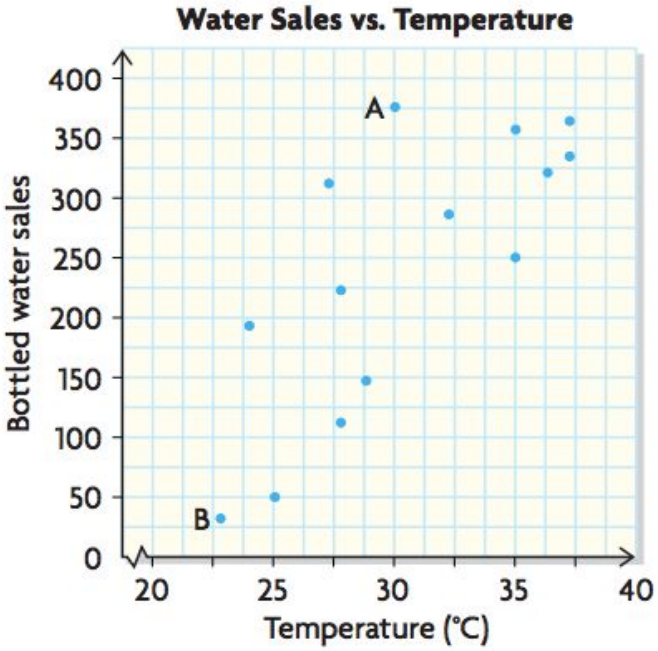
Which is the best trend line for the data shown?



2. Given the following scatterplots, draw the line (or curve) that you feel best fits the relationship shown in the graph of the data set. Write two or three key words to describe each relation on the line below the scatter plot. (rises upward to the right, falls downward to the right, no relationship, strong, weak, linear, non-linear)



3. The scatter plot shows the sales of bottled water at GRECOS during different days over the fall semester.



   a. What information does point A represent? What does point B represent?
   b. What does the scatter plot show about the relationship between water sales and temperature?

4. The following data table shows how many sit-ups Sam did in PE class.

Time (min)	0.5	1	1.5	2	2.5	3	3.5	4
Sit-ups completed	17	33	48	62	72	80	86	91

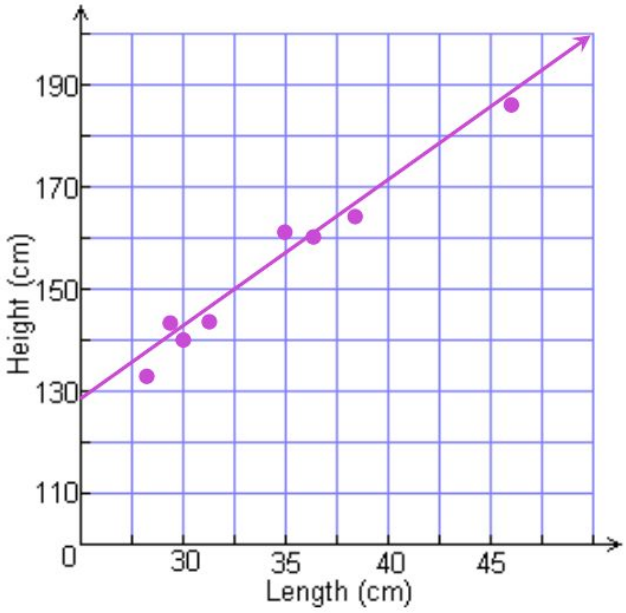
   Prepare a scatter plot and use it to describe the relationship between completed sit-ups and time.

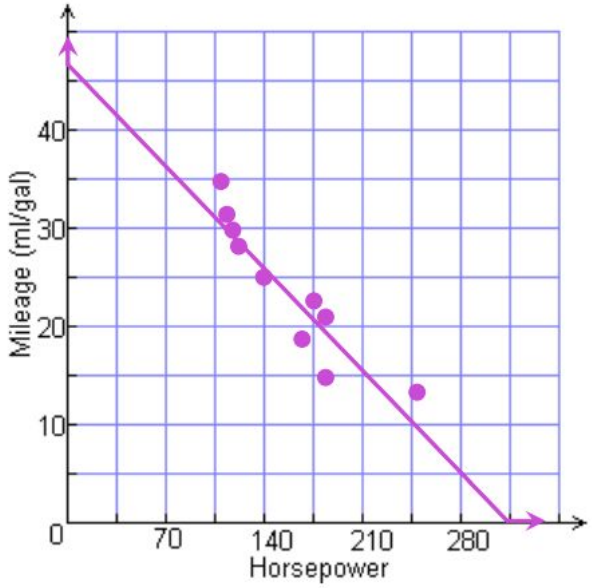
5. Students at CAC are now trying out for Coach Monte's varsity baseball team. The speeds of their pitches are measured and compared against their ages as shown on the table below.

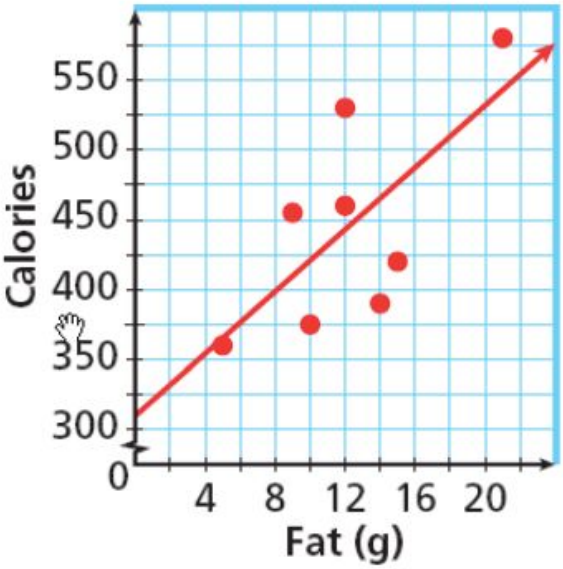
Age (years)	14.1	14.6	14.9	15.3	15.5	15.6	15.7	15.8	15.9	16.3	16.4
Speed (km/h)	79.3	50.2	66.1	103.3	62.3	40.4	91.6	75.8	55.9	52.7	62.4

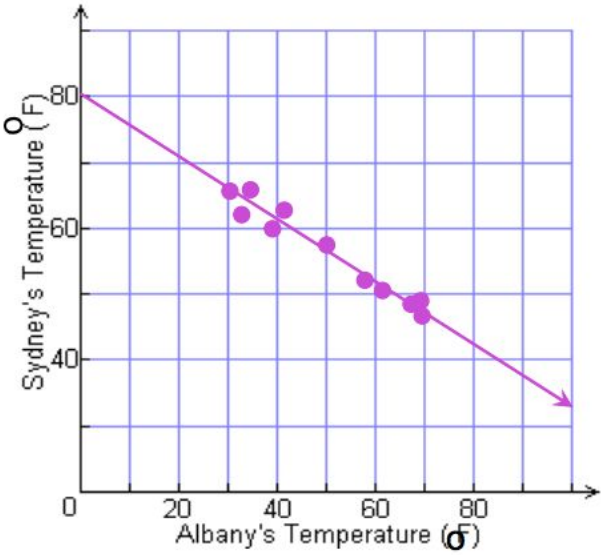
   Prepare a scatter plot and use it to describe the relationship between age and throwing speed.

6. Determine the equation of the line of best fit for the given data sets.









HOMEWORK: From Nelson 9, Chap 6.1, on pages 327-329, complete Q3,4,7,8