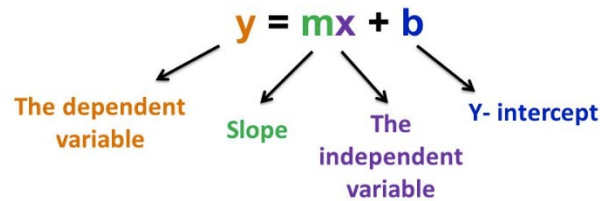


## Lesson #9.5: Using Statistics with Spreadsheets

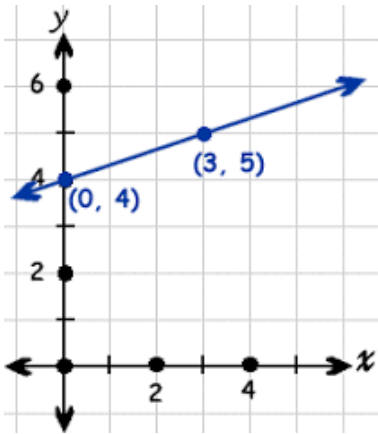
**Learning Goal:** We are learning to use spreadsheets to create scatter plots.

Since you are already on a computer, you will learn this lesson by researching and trying things on your own. Today's lesson is mainly a self-directed and self-paced lesson.

Let's begin with a quick review:



**Recap: 1.** What is the equation for this line?

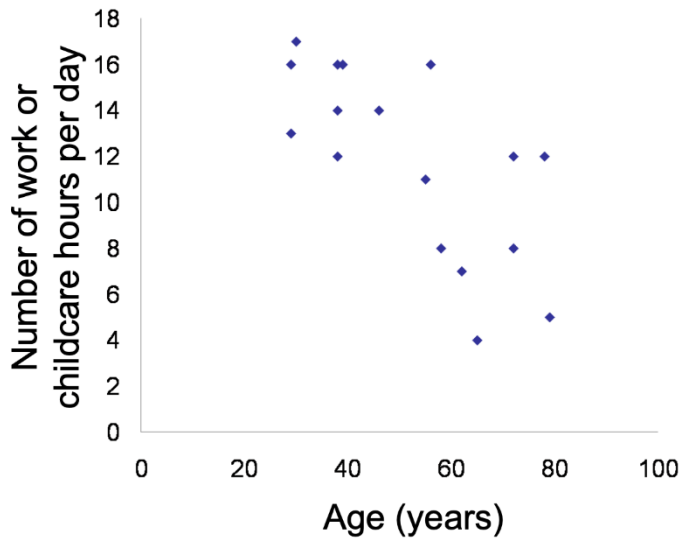


**2.** Let  $y$  = score on a test,  $x$  = number of hours that you study for the test

Given the relationship:  $y = 7x + 44$ , choose the correct option for the questions that follow:

- a) What does the y-intercept represent?
  - i. The score for an average student
  - ii. The average number of hours a student studies
  - iii. The score for a student who doesn't study at all
- b) What does the slope represent?
  - i. The expected number of hours an average student studies
  - ii. The expected increase in the score per hour of studying
  - iii. The expected score for a student who studies for one hour

**Line of Best Fit:** Draw the best estimated line of best fit through the points in the given scatter plot



We have already observed in a previous lesson that making a scatter plot and calculating the line of best fit manually is extremely time consuming and tedious. Hence, the main thing that we will do today is learn how to make a scatter plot and line of best fit using Excel.

But before we do that, we need to understand whether our line of best fit is truly a good fit.

Look up the following and answer the research questions (Q2 has been answered for you. Read through it carefully.):

1. Using your own words, explain what correlation is.

2. What is the correlation coefficient?

The linear **correlation coefficient** (sometimes called Pearson's Correlation Coefficient), commonly denoted **r**, is a measure of the strength of the *linear* relationship between two variables. The value of **r** has the following properties:

- **r** is always a value between **-1 and +1**
- The **further an r value is from zero**, the **stronger the relationship** between the two variables.
- The **sign of r** indicates the nature of the relationship: **A positive r indicates a positive relationship, and a negative r indicates a negative relationship.**

Generally speaking, you may think of the values of **r** in the following manner:


- If **|r| is between 0.85 and 1**, there is a **strong correlation**.
- If **|r| is between 0.5 and 0.85**, there is a **moderate correlation**.
- If **|r| is between 0.1 and 0.5**, there is a **weak correlation**.
- If **|r| is less than 0.1**, there is **no apparent correlation**.

3. Draw the three sketches of a scatter plot that is strong positive, random (no correlation), and perfect negative.

Alright, time to learn about scatter plots in Excel! Start a new document and put this information into Excel:

B	C	D
Month	Advertising	Items sold
Jan	\$45	15
Feb	\$55	25
Mar	\$47	17
Apr	\$75	34
May	\$90	41
Jun	\$100	47
Jul	\$100	50
Aug	\$95	46
Sep	\$88	37
Oct	\$50	22
Nov	\$45	20
Dec	\$58	30

1. Highlight the Advertising cell all the way to the bottom of last row (the 30).

2. Click insert – then the scatter plot button.  Choose the first option.

3. You should be in “Chart Design”. Click on “Add Chart Element” (first button in the row” and select “axis titles” – “Primary Horizontal”. Double click the new text box on the x-axis and change it to “Advertising Dollars Spent”. Do the same steps, but select “Primary Vertical”. Edit the text box to be “Items Sold”. Edit the Chart Title to be “Advertising Spend vs Items Sold”.

4. Click on a data point (on the scatter plot), right click, the select “Add Trendline”. This is the line of best fit. A Format Trendline will appear. Scroll down and click on “Display Equation on Chart”. Click on the equation and move it to on open spot to the right of the title.

5. Click on the x-axis numbers. On the right, you will get new options. Click the bar graph button, then “Axis Options”. Change the minimum bound to 40.

You are done the scatter plot!! You can edit the colours, grid style, and size if you wish.

6. The last thing is to now see whether Advertising does cause you to sell more items. To calculate the correlation of this data, go two spaces under the “advertising” and type “Correlation”. In the cell next to that, type:

=correl(advertising column, items sold column)

This should give you the correlation coefficient. Under the number, write what that number means (from your research earlier!).

Save and submit to Edsby.