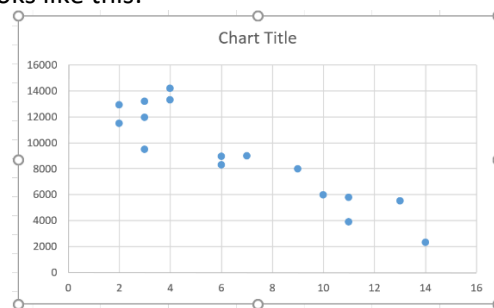


Excel: Correlation and Regression Line

A recent search of Craigslist found several used Honda Civics for sale by owner in the SF Bay Area. Listed below are the ages of the cars and the price that the owner is asking for. This data has already been entered into an Excel document called "Used Civics.xlsx" in the **Excel documents** folder in **Files** on Canvas.

Age (yr)	Asking Price (\$)
2	12900
2	11500
3	11950
3	13200
3	9500
4	13300
4	14200
6	8300
6	8950
7	9000
9	8000
10	5988
11	3900
11	5800
13	5500
14	2300

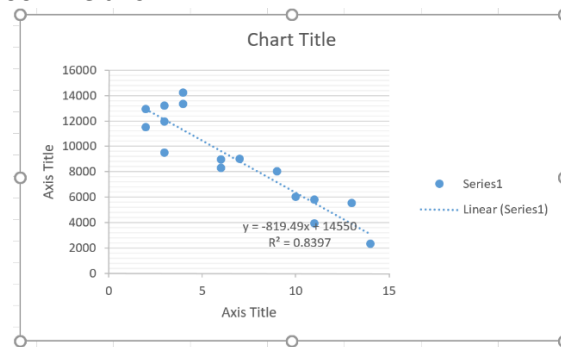
1. Since we should always start with a graph of a relationship, we will make a scatterplot for these data. It makes sense that the age of the car will affect its price, so age will be our explanatory variable and asking price will be the response variable. Highlight (by clicking and dragging) all the data in cells A2:B17. (Do not include the variable titles.) Go to the **Insert** menu and find the option that looks like a scatterplot (without lines) in the **Charts** group. Click on it. In the dropdown menu that appears, choose the top left option. You should now have a graph that looks like this:



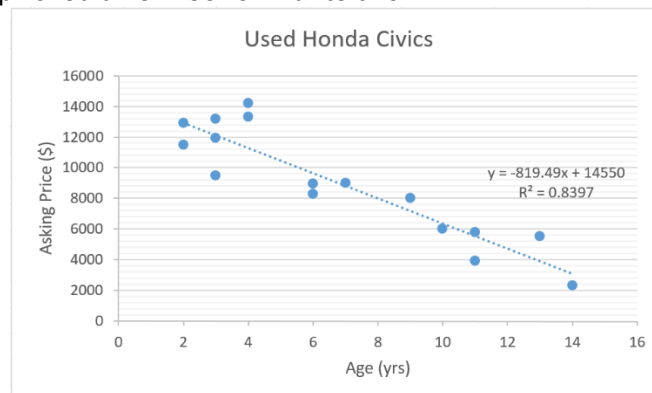
2. This scatterplot has no axes labels, so anyone looking at this graph would have no idea what relationship it displays! In the **Design** menu of **Chart Tools**, click on **Quick Layout** near the far left. You will get a dropdown window with options that looks like this:



Choose Layout 9, which I have highlighted above. This choice is perfect because it will not only give us the option to add axes labels and a chart title, but it will also graph the regression line and give us the equation of it. Your graph should now look like this:



3. Edit the chart title so that it says “Used Honda Civics” at the top of the graph, change the x-axis title to “Age (yrs),” and change the y-axis title to “Asking Price (\$).” There is no need to rescale the axes, but if you ever need to, you just double-click on any of the numbers along the x-axis so that a sliding window appears on the right side of your document titled Format Axis. There you may change the minimum and/or maximum for each axis.
4. We do not need the legend on the right side of the graph with “Series1” and “Linear (Series1).” Get rid of it by clicking on it so that there is a box with bubbles around it, and then hit **Delete** on your keyboard. Also, move the regression line equation and R^2 (by clicking on it and moving the box) so that it is more easily seen. Your completed graph should now look similar to this:



5. Now let’s find the value of the correlation. (The number given on the scatterplot is r^2 , not r .) In an empty cell of your spreadsheet, enter **=CORREL(A2:A17,B2:B17)**. (Instead of manually entering the x- and y-ranges, you can also click and drag over the data.) You could have also found the square root of the r^2 given by Excel, but that would result in a positive value of r , and this relationship obviously has a negative association.
6. Describe (preferably in a text box, or handwritten on your paper later) the association between age and price of a used Civic. Include information about the direction, form, and strength of the association. In general, what happens to the asking price as the age of the car increases?
7. The slope of the regression line is -819.49. What does that number mean in regards to the age of the car and the asking price? (Both the negative sign and the number itself.)
8. The value of r^2 is 0.8397. Explain what that means about the differences in asking price for used Honda Civics on Craigslist.
9. List at least three other (lurking) variables that could have an effect on the asking price of a vehicle.

10. We can use our regression line equation to predict the asking price of a 5-year-old Honda Civic by plugging the number 5 in for x. But instead of using our calculator, there is an Excel command for that:

=FORECAST(x value, y-range, x-range)

In an empty cell in your spreadsheet, enter **=FORECAST(5,B2:B17,A2:A17)**. (Notice that the y-range must come first!) Indicate on your page that this is the predicted asking price for a 5-year-old Civic.

11. Use the forecast command to predict the asking price of a 10-year-old Civic. There is a 10-year-old Civic on Craigslist with an asking price of \$5988. What is the residual for that car? Does buying this car seem like a good deal?
12. You have the chance to buy one of two used Civics. They are the same age, are in equally good condition, and you like them both the same. If you want to get a good deal, should you buy the one with a positive residual or the one with a negative residual? Explain.
13. Would this regression model be useful in establishing a fair price for a 20-year-old Civic? Explain.

Your printed page should have the scatterplot with regression line, the value of the correlation, and answers to all the questions in #6-13.