Correlation of Two Variables to Explore Causal Relationship

Now that you have explored the large datasets available and found the summary statistics for your variable(s) of interest, it is time to see if we can gain a better understanding of what influences the variable(s). One of the most common ways to look for patterns in data is to choose another variable that you think might be impacting your variable of interest. For example, if I notice that there is an increase in ice cream sales during summer, I might think that as temperature increases people buy more ice cream. So, I can explore that relationship. In this case, ice cream is the **dependent variable** (the variable we think might depend on another variable - in this case temperature). Temperature then is the **independent variable** (the variable we think might affect the dependent variable). Once we choose that relationship, we can look at the relationship and ask some questions about any patterns we think we might see.



In this graph, we can see that as the independent variable increases, the dependent variable increases. The positive value of the trendline slope (0.5) confirms this, and the R-squared valuetells us that we have a reasonably strong relationship (the trendline fits the data pretty well).

Now, we will explore your data to look for an independent variable that may be impacting your variable of interest (dependent variable). You will accomplish the following:

1. Decide on an independent variable that might be influencing your dependent variable
2. Find that data or similar in the online datasets
3. Generate summary statistics about the relationship between the variables
4. Interpret your results

Step 1: Think about your dependent variable (the variable of interest). What might influence this variable? Come up with 3 or 4 possible independent variables that might influence your dependent variable.

Step 2: Explore the dataset for variables that you chose above. You may need to choose something close if you cannot find the exact variable of interest. Now that you know your options, choose the **independent variable** you most want to explore. Write your independent variable here and explain why you chose it. Why do you think this independent variable will affect your dependent variable?

Step 3: Use the spreadsheet provided to add in the data for your independent variable and your dependent variable. Be sure to include country names or other identifiers to keep track of the data that goes with each country. You should have three columns: label for each datapoint, independent variable, and dependent variable. These three columns will make it possible to make a correlation between your dependent and independent variable. Paste the first 5 rows of your data below. Check with your instructor if you are not familiar with how to copy and paste spreadsheet data.

Step 4: Follow the instructions in the worksheet to generate a scatterplot for your data.

Step 5: Answer the following questions based on your results.

1. Just looking at the two columns of data, do you think the two variables are related? Why or why not? Does this match your prediction?



1. Once you have completed the practice tab in the Excel workbook, paste your practice graph here. Check with your instructor if you are not familiar with how to copy and paste graphs. What is your interpretation of the data?
2. Once you have made the plot of your data, post it below.
3. Looking at the plot, do you think the two variables are related? Why or why not? Does this match your prediction?
4. Add in the trendline, the equation of the trendline, and the R2. Record them here.
	1. Regression equation (form dependent data = mx + b):
	2. R2:
5. Based on these results, does there appear to be a relationship between your variables? Is it a positive or negative relationship? Does this match your prediction?
6. Do you **know** that your dependent variable is causing the change in your dependent variable? What else could be causing both variables to change? What could be other reasons the variables are related besides one causes the other? If the variables are not related, reflect on why you think you did not find a relationship and what your next steps might be to look for a cause.
7. Discuss a possible follow-up study.

If you found a relationship between the variables:

* 1. What kind of experiment could you run to decide if the relationship is causal?
	2. What kind of data would you need for your experiment?

 If you found no relationship between the variables:

1. What kind of experiment could you run to determine if your independent variable has any effect on the dependent variable?
2. What kind of data would you need for your experiment?

This material is based upon work supported by the National Science Foundation under Grant No.1919613. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.