

Attached is a dataset containing information about 156 countries throughout the world. Some of the data were gathered from the World Bank Data Archives. Current Coronavirus data were obtained from Worldometer's Corona Virus Statistics Tracker. Using the data set provided, you will create a statistical model to help explain the transmission and/or death rate of the Covid 19 virus.

First, pick a dependent (Y) variable. You may choose either Total Cases/1M or Total Deaths/1M population. Use the ratio and not the totals since you have countries of varying populations. Next, pick at least 4 independent variables that you think are significant predictors of your outcome variable. Identify your predictors and state your hypothesized relationship (positive for factors that increase the dependent and negative for those that decrease it). Note that the Meta-Data tabs have information to define each variable as well as information about how the data were collected. You will have to turn in your Excel sheet showing all of your work. To make the project easier, create a separate tab for your final dataset. This should only include your dependent variable and your independent variables.

Section 1:

Define your dependent and independent variables and state your hypotheses (at least 4). Do not worry about stating the null as we normally don't do that with multiple regression. Also be advised that I do not expect you to create an entire literature review for this project. Simply state the ones you are using and hypothesize the direction that you are predicting.

(i.e.,

H1: There is a significant and positive relationship between Variable 1 and Dependent Variable such that as Variable 1 increases (decreases), Dependent Variable increases (decreases).

Or

H2: There is a significant and negative relationship between Variable 2 and Dependent Variable such that as Variable 2 increases (decreases), Dependent Variable decreases (increases).

You may use these outlines or create your own, but it must be logically sound.)

Section 2:

Provide descriptive statistics. (I would use the Data Analysis Toolpak, clean up the table, and paste it in here.) Notice that some of the countries have missing data for some factors. If you use these variables. Be sure to identify and state your treatment for missing data and/or extreme outliers. The regression will not run in Excel if the dataset has missing values. At this point, with a large dataset, I would simply delete any country that had a missing value or extreme

outlier. Just make sure you report what you did and why and make sure your final sample size is at least 100.

Once you've done that, be sure to create a correlation table for all of your variables. This can be easily done using the Data Analysis Toolpak. Paste the table into this section to show that you do not have any collinearity issues that would alter your regression model. If any of your independent variables have a correlation higher than .6, I would delete one of them and add another one in its place.

Section 3:

Run the regression model in Excel and briefly explain your results. Just paste the tables into this section and state what it means.

(i.e., The model is significant ($F=?$), and explains ?? percent of the dependent variable. The regression results show a positive and significant relationship between $X1$ and Y ($b=3.245$; $p<.041$) as well as $X4$ and Y ($b=4.676$; $p<.010$), thereby providing strong support for $H1$ and $H4$. Hypothesis 2 was not supported since factor $X2$ was significantly related to Y , but in the opposite direction as we hypothesized ($b=-2.33$; $p<.022$). The relationship between $X4$ and Y was not significant ($b=0.993$; $p<.460$), thereby offering no support for hypothesis 3.)

After that, I usually provide a little table to make it easier to read. Feel free to use this one. Just make sure to use the variable names, not X and Y . Also, made sure to provide your own numbers and valences, not the ones I made up.

Hypothesis	Hypothesized Relationship	Valence	Beta	p-value	Result
H1	$X1 \rightarrow Y$	+	3.245	.041	Supported
H2	$X2 \rightarrow Y$	+	-2.335	.022	Not Supported
H3	$X3 \rightarrow Y$	-	.993	.460	Not Supported
H4	$X4 \rightarrow Y$	+	4.676	.010	Supported

Section 4:

Write a brief summary of what these results mean and how they could be used to improve public policy and create better governance structures, both domestically and internationally.

Once you've completed this, upload both the Excel file and your Word document into the Blackboard dropbox provided in Week 7. Note the you must use Excel and Word and you are only allowed to use the Data Analysis Toolpak to perform your calculations.