**Culminating Final Project**

You will need to do:

* Video recording a 3-mins presentation to discuss your findings, along with a brief PowerPoint.
* Formal report to fully display your research
* Peer feedback

The goal of this culminating statistics project is to identify the factors that might have some influence on a chosen topic/variable. You will be collecting and analysing data for three different variables using a secondary source. There will be **one dependent variable** and **two independent variables** for which data will be needed.

For example, what are the factors that affect climate change? If climate change was your dependent variable, what could you pick for your independent variable? You will choose 2 factors that will create models that you can compare, to see if you can find a high correlation. Your final report and summary presentation will include the following sections.

**Section 1: Dependent variable**: Choose the main variable that you want to measure versus other variables. This will go on the vertical axis of your scatter plot. Provide **rationale for why** you chose this variable and brainstorm different variables that may impact it. Research the variable on the internet and collect data for **at least 10 years** (the more the better) or over **at least 10 different places** in the same year.

Note: Statistics Canada is our most reliable source for data and one of the best sources to get a wide variety of data. Be sure to cite your source in proper MLA or APA format.

**Section 2: Independent variables**: Choose the independent variables that you want to measure against the dependent variable. You will find 2 different variables for which you can create a line of best fit with to test correlation. This will be your biggest challenge of the assignment. You need to find variables such that they overlap with the data that you chose for your dependent variable.

For example: If you found the years 1990 to 2010 for your dependent variable, you will need to find those same years for each of your independent variables. If you had data across the provinces and territories for your independent variable for 2016, you will need data across the provinces and territories for 2016 for each of your independent variables.

**Section 3: One variable measure with graphs and distributions**: Describe each of the 3 of your variables (two independent and one dependent) by calculating and making reference to the **measures of central tendency**, **spread**, and creating a **histogram and box-whisker-plots**. Describe the distribution of your data as well. For each of the variables, describe the meaning of each of the measures as you state what they are.

**Section 4: Two variable measure with linear models and correlations**: Create 2 **scatter plots**, one for each of your independent variables. Create 2 scatterplots, one for each of your independent variables. Create the **linear model** to describe the **relationship**. Describe the **strength of the correlation** and the **meaning of the slopes and y-intercepts**.

**Section 5: Identify and remove relevant points**: identify one point on each scatter plot that has the greatest impact on the model by using residual plots. Describe the impact of the point by comparing the new linear model to the original.

**Section 6: Classification of correlation type**: For each of the models, classify the correlation type as **cause and effect**, reverse cause and effect, accidental or common cause. Explain your choice and brainstorm possible common cause factors where applicable.

**Section 7: Conclusions from linear models:** Make conclusions from the two variable analysis taking into consideration the removal of points that are far from the line of best fit. Discuss the overall impact, if any, that the independent variables have on the dependent variable by referring to other sections of the report.

**Section 8: Response to peer feedback:** Read two reports from 2 classmates, include the feedback that you received from your classmate as well as the follow up questions. Identify the steps you took to act on the feedback and your responses to the follow up questions.

**Section 9: Areas for improvement and recommendations for further exploration**: Identify the areas of improvement in your study. If you had more time and resources, what are you still curious about that you may want to pursue further. Identify the possible **biases** in your investigation and overall how confident you are with your conclusions.

**Rubric: Use the following list to check if you’ve covered all requirements before submit it.**

|  |  |  |
| --- | --- | --- |
| Requirements | Yes | No |
| 1. The collected data covers either at least 10 years or over at least 10 different places in the same year. |  |  |
| 1. All recourses were recorded and cited in proper MLA or APA format. |  |  |
| 1. Provided rationale for why you chose this variables. |  |  |
| 1. Created and calculated measures of central tendency, spread, histogram, and box-whisker plots for dependent variable. |  |  |
| 1. Described the meaning of each of the measures for dependent variable. |  |  |
| 1. Created and calculated, measures of central tendency, spread, histogram, and box-whisker plots for the first independent variable. |  |  |
| 1. Described the meaning of each of the measures for the first independent variable. |  |  |
| 1. Created and calculated measures of central tendency, spread, histogram, and box-whisker plots for the second independent variable. |  |  |
| 1. Described the meaning of each of the measures for the second independent variable. |  |  |
| 1. Described the type of distribution of dependent and two independent variables. |  |  |
| 1. Created 2 scatterplots, one for each of your independent variables. |  |  |
| 1. Created the linear model to show the equation of linear regression and R2 value for each of your independent variables. |  |  |
| 1. Described the relationships and strength of the correlation for each of our independent variables. |  |  |
| 1. Described the meaning of the slopes and y-intercepts of linear regression for each of your independent variables. |  |  |
| 1. Identified possible outlier(s) on each scatter plot for each of your independent variables by using residual plots. |  |  |
| 1. Described the impact of the outlier(s) for each of your independent variables by comparing the new linear model (the one you regenerated after remove the outlier) to the original. |  |  |
| 1. For each of the models, classified the cause and effect. Explained any possible common cause factors where applicable. |  |  |
| 1. Properly showed conclusion of what you find. |  |  |
| 1. Identified the areas of improvement and recommendations. |  |  |
| 1. Posted 1 question for each of 2 classmates before the deadline. |  |  |
| 1. Answered or attempted to answer the questions you got from classmates before the deadline. |  |  |