

MDM4U – Statistics Culminating Activity

Overview

You will be completing a statistical analysis of data in order to answer a significant question. There is an introduction, a one-variable analysis, and a two-variable analysis with conclusions. Your culminating activity is a bit more formal than the one and two variable analyses you have done so far.

Introduction

- Background Information
 - o Describe your two variables, what they mean, what they measure
 - o Are there any reasons to expect that there might be a relationship between your two variables? Why might it make sense for there to be a link?
- Question
 - o Is there a relationship between...
- Hypothesis
 - o What do you expect to see when you do your analysis?
 - o As the independent variable increases, what do you expect to see in your dependent variable?
 - o Predict the strength of the correlation (if you think one will be found)
- Data
 - o Raw data table, formatted in your word document (not a screenshot from your spreadsheet!) like the one below, with a proper table title at the top
 - o **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 APA format. You may want to revisit your Portfolio entry from Unit 5 Activity 2 for the data that you found from Statistics Canada. You can also revisit the tutorial video in that activity.

Table 1: Description of the data that is summarized in the table. Data from Author (year).

Year, Country, Player, etc	Name of Variable 1 (with units)	Name of Variable 2 (with units)
1	3	5
2	4	6

One Variable Analysis

You will complete a one variable analysis for **each** of your variables. Include the following in your analysis:

- Mean, median, mode, standard deviation
- Histogram and Box and whisker plot
- Description of the type of data distribution (normal, negatively skewed, positively skewed)
- Description of quartiles, IQR, and any potential outliers
- Calculate and interpret the z-score for one data point of your choice
- Calculate and interpret the 90th percentile

Make sure all graphs have proper figure titles below the graph, like the graph shown below. You should also have descriptive x-axis labels and y-axis labels.

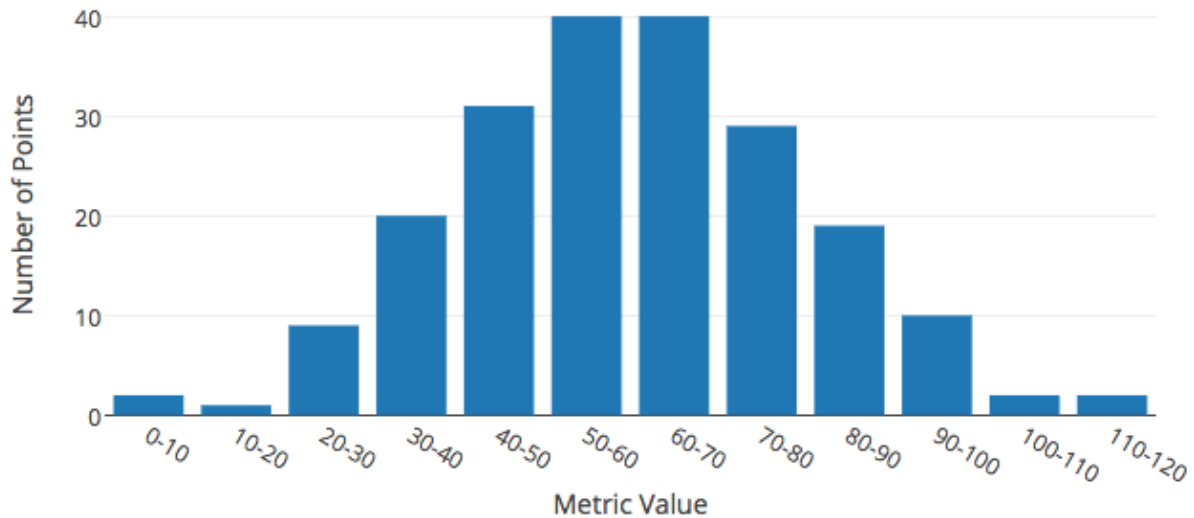


Figure 1: Description of the data that is included in the figure, including the type of graph you have used. Data from Author (year).

Refer to each graph in the main body of your assignment (As seen in Figure 1, ...). Graphs should be page width and included in your word document in the order they are mentioned. Your assignment should follow the order listed above, with a paragraph to summarize or describe each bullet point. Be thorough and include as much detail as possible.

Two Variable Analysis

You will compare your variables using a two-variable analysis, then wrap up your project with a conclusion to summarize your findings and suggest next steps for study.

Two Variable Analysis:

- Linear Regression
 - o Create a scatter plot of your data and add a line of best fit
 - o Determine the correlation coefficient “ r ” and comment on its value.
 - o Explain what the equation of the line of best fit means in the context of the question and what the two variables are (what does the equation tell you? What does the slope mean? What does the y-intercept mean?)
 - o Comment on whether the line appears to be a good fit for the data

- Outliers
 - Comment on the presence of outliers (if any) and any effect they may be having on the linear regression (calculated and identified in your one variable analysis)
 - Create a scatter plot with the outliers removed and add a line of best fit (include a residual plot here)
 - Determine the correlation coefficient “r” and comment on its value.
 - Explain what the equation of the line of best fit means in the context of the question and what the two variables are (What does the equation tell you? What does the slope mean? What does the y-intercept mean?)
 - Comment on whether the line appears to be a better fit for the data.

- Cause and Effect
 - Explain what type (common, accidental, etc.) of causal relationship is likely to be present between the two variables. Provide reasons to support your answer

Conclusion:

- Answer the question you wrote in your introduction
- Was your hypothesis correct? Why or why not?
- Are there any extraneous variables that you did not consider in your analysis that could be affecting your results? What kind of effect could they have?
- What would be an interesting question for further study?

Include APA style references for your data sources (and any other websites that helped you complete your culminating project) at the end of the document.

(Separate page)

References

Author, A. (year). *Title of Webpage*. Retrieved from
<https://www.somewebsite.com/somethingsomething>

Name of Organization (year). *Title of Webpage: Maybe it has a subtitle*. Retrieved from
<https://www.somewebsite.com/somethingsomething>

** Look at the OWL APA website to help you with formatting for number of authors, types of sources...

** Reference list entries must be in alphabetical order

Name: _____

VARIABLE 1

	Level 1 (50-59%)	Level 2 (60-69%)	Level 3 (70-79%)	Level 4 (80-100%)
Introduction (C) - <i>What are the variables</i> - <i>Background information</i>	Describes the variables with limited effectiveness	Describes the variables with some effectiveness	Describes the variables with considerable effectiveness	Describes the variables with a high degree of effectiveness
Question (T)	Develops questions with limited effectiveness	Develops questions with some effectiveness	Develops questions with considerable effectiveness	Develops questions with a high degree of effectiveness
Hypothesis (T)	Develops a hypothesis with limited effectiveness	Develops a hypothesis with some effectiveness	Develops a hypothesis with considerable effectiveness	Develops a hypothesis with a high degree of effectiveness
Data (C) - <i>Data collection methods</i> - <i>Data table</i>	Describes the data being used and how it was collected with limited effectiveness	Describes the data being used and how it was collected with some effectiveness	Describes the data being used and how it was collected with considerable effectiveness	Describes the data being used and how it was collected with a high degree of effectiveness
	Organizes data in a table with limited effectiveness	Organizes data in a table with some effectiveness	Organizes data in a table with considerable effectiveness	Organizes data in a table with a high degree of effectiveness
Calculations (K) <i>Mean, median, mode, standard deviation, quartiles, IQR, z-score, percentile</i>	Measures of central tendency and spread are calculated with limited accuracy	Measures of central tendency and spread are calculated with some accuracy	Measures of central tendency and spread are calculated with considerable accuracy	Measures of central tendency and spread are calculated with a high degree of accuracy
Interpretations of Calculations (C) <i>Mean, median, mode, standard deviation, quartiles, IQR, z-score, percentile</i>	Explains and interprets calculated values with limited accuracy	Explains and interprets calculated values with some accuracy	Explains and interprets calculated values with considerable accuracy	Explains and interprets calculated values with a high degree of accuracy
Graphs (C) <i>Histogram, box and whisker plot, figure titles</i>	Organizes data into graphs with limited effectiveness	Organizes data into graphs with some effectiveness	Organizes data into graphs with considerable effectiveness	Organizes data into graphs with a high degree of effectiveness
Interpretations of Graphs (C) <i>Histogram, data distribution, box and whisker plot, outliers</i>	Explains and interprets graphs with limited accuracy	Explains and interprets graphs with some accuracy	Explains and interprets graphs with considerable accuracy	Explains and interprets graphs with a high degree of accuracy

VARIABLE 2

	Level 1 (50-59%)	Level 2 (60-69%)	Level 3 (70-79%)	Level 4 (80-100%)
Calculations (K) <i>Mean, median, mode, standard deviation, quartiles, IQR, z-score, percentile</i>	Measures of central tendency and spread are calculated with limited accuracy	Measures of central tendency and spread are calculated with some accuracy	Measures of central tendency and spread are calculated with considerable accuracy	Measures of central tendency and spread are calculated with a high degree of accuracy
Interpretations of Calculations (C) <i>Mean, median, mode, standard deviation, quartiles, IQR, z-score, percentile</i>	Explains and interprets calculated values with limited accuracy	Explains and interprets calculated values with some accuracy	Explains and interprets calculated values with considerable accuracy	Explains and interprets calculated values with a high degree of accuracy
Graphs (C) <i>Histogram, box and whisker plot, figure titles</i>	Organizes data into graphs with limited effectiveness	Organizes data into graphs with some effectiveness	Organizes data into graphs with considerable effectiveness	Organizes data into graphs with a high degree of effectiveness
Interpretations of Graphs (C) <i>Histogram, data distribution, box and whisker plot, outliers</i>	Explains and interprets graphs with limited accuracy	Explains and interprets graphs with some accuracy	Explains and interprets graphs with considerable accuracy	Explains and interprets graphs with a high degree of accuracy
Scatter Plots (C) - Linear - Outliers removed	Organizes data into graphs with limited effectiveness	Organizes data into graphs with some effectiveness	Organizes data into graphs with considerable effectiveness	Organizes data into graphs with a high degree of effectiveness
Regression (A/T) - Equations (interpret slope and y-intercept) - Correlation coefficients - Is each line a good fit for the data? Which is best? How do you know?	Applies information from the regression analyses to real world contexts with limited effectiveness	Applies information from the regression analyses to real world contexts with some effectiveness	Applies information from the regression analyses to real world contexts with considerable effectiveness	Applies information from the regression analyses to real world contexts with a high degree of effectiveness
Cause and Effect (A/T) - Type of relationship and reasons for your choice	Analyzes and explains the type of causal relationship with limited effectiveness	Analyzes and explains the type of causal relationship with some effectiveness	Analyzes and explains the type of causal relationship with considerable effectiveness	Analyzes and explains the type of causal relationship with a high degree of effectiveness
Conclusion (T) - Correct hypothesis? Extraneous variables? - Benefit to community?	Draws conclusions with limited effectiveness	Draws conclusions with some effectiveness	Draws conclusions with considerable effectiveness	Draws conclusions with a high degree of effectiveness
Communication (C) - Spelling - Grammar - Punctuation - Capitalization - Citations	Uses conventions, vocabulary, and terminology with limited effectiveness	Uses conventions, vocabulary, and terminology with some effectiveness	Uses conventions, vocabulary, and terminology with considerable effectiveness	Uses conventions, vocabulary, and terminology with a high degree of effectiveness