

DUE: October 22nd, 2021 at 11:59 PM

QMM1001 Case Study 1 [20%]

DATA

Have you ever wondered how you spend your day? For this case study, you will collect data on your daily habits to create a personalized data set. This data set must contain nine (9) variables that meet the following requirements. A template (.csv) to use for your personalized data was provided in Moodle.

1. Date
2. Hours spent in Zoom class
3. Hours spent studying
4. Hours spent sleeping
5. Number of times leaving the house
6. Did you watch the news today?
7. A nominal level categorical variable of your choice
8. An ordinal level categorical variable of your choice
9. A quantitative variable of your choice

OUTLINE

The goal of the case study is to use the personalized data collected to answer the questions:

1. What do you do in a day?
2. Do you spend your days how you expect?

You will answer these questions by analyzing your personalized data set in R. You will write up a summary of your methods and the interpretations of your results in a short report using Word (a template is provided). Your report must include three sections:

1. Introduction
2. Data Analysis
3. Conclusion

Please submit your personalized data set (.csv), an R script that includes your full code, and your report in Word before October 22nd at 11:59 PM. **Late submissions will not be accepted** without notifying the professor and discussing any extenuating circumstances before the assessment due date. **You may not work with another student and/or collaborate on your answers.** To do so would be considered cheating and would be a violation of Cambrian's Academic Integrity policy. You may refer to your notes but you must provide your own unique solutions to the case study questions.

REQUIREMENTS

In the **Introduction** section you must include the following:

- State the questions that you are attempting to answer:
 - What do you do in a day?
 - Do you spend your days how you expect?

DUE: October 22nd, 2021 at 11:59 PM

- List all 9 variables in your personalized data set including how they are recorded (for example Yes/No, scale of 1 – 5, time in hrs, etc.).
- For the 3 personalized variables of your choice explain why you selected them.

In the **Data Analysis** section, you will analyze your personalized data set and interpret your results. Throughout this section, it is extremely important that you relate your findings to your daily activities to answer the question: do you spend your days how you expect? Throughout the data analysis section you must include 5 or more examples of your daily activities that help to explain the results of your analysis for full marks. Please note that the date will be treated as an identifier variable and will NOT be included in the analysis of variables below. You must complete the following steps.

For EACH categorical variable:

1. Create a frequency table and interpret it.
2. Create an appropriate visual display and interpret it.
3. Create a contingency table for the two categorical variables of your choice (variable 7 and variable 8) and interpret your findings in relation to what you do each day.

For EACH quantitative variable:

1. Create a properly labelled histogram and comment on the shape of the distribution (mode, symmetry, outliers).
2. Calculate the mean, median, standard deviation and IQR for each variable (these can be displayed in a tabular format).
3. Calculate if there are any outliers (using standardized values for symmetric data and boxplots and the 1.5 IQR rule for asymmetric data) and comment on why these may have occurred.
4. Determine the correlation coefficients between all pairs of quantitative variables.

For the TWO quantitative variables with the highest correlation:

1. What is the direction and strength of the relationship?
2. Check the conditions required for regression – are they met?
3. Determine the equation of the regression line. State the equation of the line and interpret the slope and y-intercept values.
4. Create a scatter chart that includes the regression line. The chart must be properly labelled.
5. Use the line to make a prediction for the dependent variable by choosing an appropriate value for the independent variable (this can be any value you want to make a prediction for!).

In the **Conclusion** section summarize your findings about what you do in an average day by answering the questions:

- What do you do in a day?
- Do you spend your days how you expect?

Also comment on what you learned about your time management skills by completing this case study and if you can make any changes/improvements for the remainder of the semester.

DUE: October 22nd, 2021 at 11:59 PM

FORMAT

The report should be formal. Here is a checklist for writing your report:

- ✓ Size 12 font, no required length
- ✓ Three sections: introduction, data analysis, and conclusion
- ✓ You can write in the first person since you are describing your daily activities.
- ✓ DO include all visualizations and tables created in your report.
- ✓ Use a consistent and professional format for visualizations and tables.
- ✓ Include examples of your daily activities to support your analysis.
- ✗ Do NOT include any code within your report.
- ✗ Do NOT include the steps from the case study on pages 1 and 2. You need to describe the steps of your analysis and the methods used for outside readers to understand.

Please consult the marking guide on the final page for full details on how you will be assessed.

TIMELINES

To help you stay on track for this case study assignment, steps to complete are included below. There will be multiple check points built into your applied activities before the final case study is due.

Step 1: Case Study 1 Assigned

September 10

Choose variables 7 – 9 that will be unique to you and start collecting data. You must update your personalized data set **EVERY DAY** until October 15.

Step 2: Case Study Check Points

September 17 – October 15

Case study check points will be built into your applied activities throughout the semester to help you prepare for the case study. There will be case study question(s) on the following assessments:

Module 1 Applied Activity – Due September 17th, 2021 @ 11:59PM

Module 2 Applied Activity – Due September 24th, 2021 @ 11:59PM

Module 3 Applied Activity – Due October 8th, 2021 @ 11:59PM

Module 4 Applied Activity – Due October 15th, 2021 @ 11:59PM

Step 3: Case Study Work Periods

October 20 and 22

Class on October 20 and October 22 will be used as work periods for the case study. Outside of class time, the case study should take between 4 - 6 hours to complete.

Step 4: Case Study 1 Due [20%]

October 22

Please submit your personalized data set, report in Word, and a separate R script that includes your full code before October 22 at 11:59 PM. You must submit these 3 files.

DUE: October 22nd, 2021 at 11:59 PM

MARKING GUIDE The case study is worth 20% of your final grade and will be evaluated based on the guide below.

	Level 1	Level 2	Level 3	Level 4	Comments
REPORT	Introduction to the report is unclear, explaining few variables investigated, questions, and methods	Introduction to the report is somewhat clear, explaining some variables investigated, questions, and methods	Introduction to the report is mostly clear, explaining most variables investigated, questions, and methods	Introduction to the report is written clearly, summarizing variables investigated, explaining all questions in the case study and methods used	
	Data analysis section uses no relevant examples relating to personal activities	Data analysis section uses few relevant examples relating to personal activities (less than 3)	Data analysis section uses some relevant examples relating to personal activities (between 3 – 5)	Data analysis section uses multiple relevant examples relating to personal activities (more than 5)	
	Conclusion to the report clearly restates none of the questions investigated and few of the main findings	Conclusion to the report is missing two of the two questions investigated or the main findings	Conclusion to the report is missing one of the two questions investigated or the main findings	Conclusion to the report clearly restates two of the questions investigated and the main findings	
	Errors in word selection, usage, sentence structure, spelling, punctuation and capitalization detract from meaning	Many errors in word selection, usage, sentence structure, spelling, punctuation and capitalization	Some errors in word selection, usage, sentence structure, spelling, punctuation and capitalization	Few errors in word selection, usage, sentence structure, spelling, punctuation and capitalization	
	Notation and formatting of tables, graphs, and text is inconsistent, and the report does not look professional	Notation and formatting of few tables, graphs, and text is consistent and the report looks presentable	Notation and formatting of some tables, graphs, and text is consistent and the report looks presentable	Notation and formatting of all tables, graphs, and text is consistent and the report looks professional	
	0 marks	1 mark	2 marks	3 marks	Comments
Categorical	No frequency distributions are included	A correct frequency distribution is included and interpreted for 1/3 categorical variables	A correct frequency distribution is included and interpreted for 2/3 categorical variables	A correct frequency distribution is included and interpreted for all three categorical variables	
	No visual displays are included	A correct visual display and interpretation are included for 1/3 categorical variables	A correct visual display and interpretation are included for 2/3 categorical variables	A correct visual display and interpretation are included for all categorical variables	
	No two-way contingency table	A contingency table and interpretation are included with at least one error	A contingency table and interpretation are included with one error	A correct contingency table and interpretation are included	
Quantitative	No visual display is included	A properly labelled and correctly constructed histogram is included and interpreted for one or two quantitative variables	A properly labelled and correctly constructed histogram is included and interpreted for three or four quantitative variables	A properly labelled and correctly constructed histogram is included and interpreted for all five quantitative variables	
	No measures of centre are calculated	Measure of centre are calculated correctly for one or two quantitative variables	Measure of centre are calculated correctly for three or four quantitative variables	Measure of centre are calculated correctly for all five quantitative variables	
	No measures of spread are calculated	Measures of spread are calculated correctly for one or two quantitative variables	Measures of spread are calculated correctly for three or four quantitative variables	Measures of spread are calculated correctly for all five quantitative variables	
	No check for outliers	Outliers are calculated and explained for one or two quantitative variables	Outliers are calculated and explained for three or four quantitative variables	Outliers are calculated and explained for all five quantitative variables	

DUE: October 22nd, 2021 at 11:59 PM

Correlation & Regression	Correlation coefficients are not calculated	Correlation coefficients are calculated with at least one error	Correlation coefficients are calculated correctly for all pairs of variables	N/A	
	Direction and strength of relationship are not given	One of the direction and strength of the relationship are stated correctly	Both the direction and strength of the relationship are stated correctly	N/A	
	Conditions are not checked	Conditions are checked with more than one error	All conditions are checked with one error	All conditions are checked correctly	
	The equation of the regression line is not determined	The equation of the regression line is calculated with at least one error	The equation of the regression line is calculated with one error and/or is missing from the report	The equation of the regression line is calculated correctly and stated in the report	
	The slope and y-intercept are not interpreted	The slope and y-intercept are both interpreted with errors	One of the slope and y-intercept are interpreted correctly	The slope and y-intercept are interpreted correctly	
	No visual display is included	A visual display and interpretation are included with at least one error	A properly labelled and correctly constructed visual display is included and interpreted	A properly labelled and correctly constructed visual display is included and interpreted	
	No prediction is given	A prediction is included with at least one error	A correct prediction is given	N/A	
	0 marks	1 mark			Comments
Submission	One or more of the following components are missing -Personalized data (.csv) -Report in Word -R script with code	All the following components are submitted: -Personalized data (.csv) -Report in Word -R script with code	N/A		
Comments					
TOTAL					/60