MATH1324 Introduction to Statistics Assignment 3

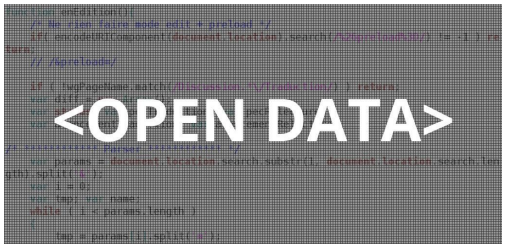
Final Project (Last Updated 10.05.2019)

Overview

The Course Evaluation Surveys (CES) have opened. Please take the time to complete these important surveys. You can access this survey through myRMIT. (by clicking on Students Surveys link). Positive feedback provides evidence of effective course delivery and constructive feedback can help improve course delivery for future students.

Thanks for collaboration.

Laleh



The final assignment is simple. I want you to think of an interesting statistical question, find some open data, and use your knowledge gained during the course to answer your question. This is your opportunity to demonstrate all that you have learnt during this course. You will be awarded (with marks) the clearer you demonstrate your skills. This isn't about showing off. You won't be rewarded for applying advanced statistical analysis outside the course. That's not what this course is about. This course is about asking and answering interesting questions about the world using the fundamentals of statistics. Keep it simple! There is only one major constraint for this course. You must use Open Data because the aim of this assignment is turn your project into a slideshow presentation that you can include in an ePortfolio.

This assignment is worth 20% and must be uploaded to the **Assignment 3 Turnitin link** submission by **2/06/2019**.

Open Data

Assignment 3 is open-ended, but there will be one key requirement. The data to be used must be open and ideally have a Creative Commons Licence. This will ensure you can share your work with the anyone provided you make proper attribution. If you're not sure if data is Open, contact the provider, read the documentation or post on the discussion board and I will investigate. Some open data sources are provided below, but I encourage you to find others:

- https://www.kaggle.com
- UCI Machine Learning Repository
- data.gov

- world bank
- amazon web services
- google data sets
- youtube video data sets
- analytics vidhya
- quandl
- driven data
- http://www.abs.gov.au/
- https://www.data.vic.gov.au/
- http://www.bom.gov.au/

You can also collect your own data. If you choose to collect your own data, then explain how you collected your data and explain the sampling method. There should be enough detail here so that someone else could replicate your data collection.

Groups

Students are permitted to work individually or in groups of up to 3 for Assignment 3. **Each group must fill out the following form before 26/05/2019 to register their group details.** Submit the details of your group here.

Group Registration Form

All group members must submit a copy of the report! Group members that are not registered or do not submit a presentation will not be acknowledged. One group member's submission will be marked and given feedback. It will be the responsibility of the marked group member to share the group's feedback with the other group members. The other group members will receive a mark only.

Submission Instructions

The report must be uploaded as a **PDF** with your code showing.

or

The assignment 3 report can be completed using the R Markdown template provided here:

R Markdown Template - Assignment 3

Note that this is a slideshow template. The template includes basic instructions. You can read more here.

You can also use other format to provide slideshow.

Or

The slideshow presentation will be in a reproducible R Markdown format with written sections, R code and output. **Presentations are limited (maximum) to 20 slides.** The presentation must

be composed of the following sections. You can add more if your wish, but you must include these sections as a minimum.

Please not forget to submit **Course Evaluation Surveys (CES).**

1. **Presentation title and group/individual details [Plain text]**: You can add the title of your presentation and student(s) details by updating the "title" and "author" entries at the top of the R Markdown Template.
2. **Introduction [Plain text]:** A good introduction provides a brief background to the problem, defines important terms, and leads to a strong rationale.
3. **Problem Statement [Plain text]:** State the overall problem/question driving the investigation. Summarise how you will use statistics to solve the problem or answer your question.
4. **Data [Plain text]:** If you collected your own data, explain how you collected your data. There should be enough detail here so that someone else could replicate your data collection. Explain the sampling method if known. Ensure you reference the data source if you have used Open Data. List and explain the important variables. Explain everything that you do to preprocess the data.
5. **Descriptive Statistics and Visualisation [Plain text & R code & Output]:** Summarise the important variables in your investigation. Use visualisation to highlight interesting features of the data and tell the overall story. Explain how you dealt with data issues (if any), e.g. missing data and outliers.
6. **Hypothesis Testing [Plain text & R code & Output]:** Apply an appropriate hypothesis test for your investigation. Ensure you state the hypotheses and check any assumptions. Report the appropriate values and interpret the results.
7. **Discussion [Plain text]:** Discuss the major findings of your investigation. Discuss any strengths and limitations. Propose directions for future investigations. This is a good place to re-state your findings as a final conclusion. What is the one take home message the reader should leave with?
8. **References [Plain text]:** Provide a list of any references you use in the presentation.

The presentation pdf must be uploaded to the **Assignment 3 Turnitin link** submission by **2/06/2019**.

Extensions will only be granted in accordance with the RMIT University Extension and Special Consideration Policy. No exceptions. Assignments submitted late will be penalised (see Course Information for further details).

Collaboration

You are permitted to discuss and collaborate on the assignment with your classmates and other groups. However, the write-up of the presentation must be an individual/group effort.

Assignments will be submitted through Turnitin, so if you've copied from a fellow classmate/group, it will be detected. It is your responsibility to ensure you do not copy or do not allow another classmate/groups to copy your work. If plagiarism is detected, both the copier and the student/group copied from will be responsible. It is good practice to never share assignment files with other students/groups. You should ensure you understand your responsibilities by reading the RMIT University website on academic integrity. Ignorance is no excuse.

Assignment 3 Marking Rubric

Criteria	Not acceptable (0)	Needs Improvement (1)	Good (2)	Outstanding (3)
Introduction (15%)	No background to investigation and/or rationale. The aim of the investigation was not clear.	A simple background and rationale to the investigation were provided, but it lacked detail/clarity.	Background and rationale driving the investigation were provided, but some minor details were missing/not clear.	Background builds a strong and interesting rationale driving the investigation. Important concepts are detailed. The reader wants to know the answer.
Method/ Data (15%)	The data/data collection method was described/poorly. The quality and nature of the data were unclear.	The data/data collection method description needed improvement. Important insight into the nature of the data was missing.	The data/data collection method was described in detail, but some minor aspects were not entirely clear/correct.	The data/data collection method was clear, concise and detailed. The presenters clearly had an in-depth knowledge of the data.
Analysis (40%)	Major aspects of the analysis were inappropriate and/or completed incorrectly. Results of the statistical analysis are poorly presented and/or missing. Results were misinterpreted.	The analysis and presentation of results were mostly suitable, but some important elements (e.g. statistical tests, descriptive statistics or visualisations) were poorly described/ and or interpreted.	Most of the statistical tests and summaries were reported and interpreted correctly. There was some room for improvement in the analyses selected, reporting of results and/or interpretations.	The analysis is justified, assumptions checked, and results reported clearly. Results are presented and summarised meticulously. The interpretation of the results were clearly and accurately communicated.
Discussion (20%)	There was no discussion or conclusion or the discussion and conclusion do not match the results. There is no attempt to interpret the results back to the context of the original investigation. The presenter did not appear to understand the implications of the investigation's findings.	A discussion and conclusion are presented, however, there is a lack of insight into the results of the investigation. There is either insufficient discussion or the discussion raises many irrelevant points.	The results of the investigation are discussed in detail and an appropriate conclusion was reached. Minor details could be added, or some minor irrelevant points were included.	The results of the investigation are critically analysed and a conclusion was reached which related back to the initial investigation question. The discussion is succinct and conveys the take home message from the investigation.
Presentation (10%)	The presentation appears messy, is difficult to read, has poor spelling/grammar, and poor referencing. The slide are crowded/empty.	The presentation has a number of issues with appearance, readability, spelling and/or referencing.	Overall, the presentation appears professional and well laid out. Only a few minor issues with crowding, appearance, readability, spelling/grammar or	The presentation shows considerable effort and attention to detail. Its is succinct and difficult to fault.

			referencing.	