MA20277 COURSEWORK 1 COVER SHEET

Contents

General Instructions

Set: 15:30 on Friday, 27 October 2023

Due: 17:00 on Tuesday, 7 November 2023

Estimated time required: 12-15 hours for students familiar with the concepts covered in Chapters 1–3 of the lecture notes.

Submission: Submission is via the MA20277 Moodle page only. You have to submit a PDF file which contains all your answers (including plots) and R code, along with the R Markdown file you used for creating your PDF file. You should use the provided R Markdown file for the analysis and knit it to a PDF or Word. If you knit to Word, you should export the document to PDF for submission. Please do not remove the question text from the R Markdown file and use it to separate your answers in order to facilitate the marking process.

Conditions: This is an individual assignment, and you should not discuss it with anyone other than the lecturer or your tutor. It should be completed during your computer lab in the week 30 October - 3 November and in your own time.

Value: This assessment carries 50% of your total mark for MA20277.

Marking: There are a possible 50 marks available and you will be marked based on the quality of your answers and presentation. There is not necessarily a single correct analysis for the tasks below, so you will not be marked on the basis of how close you get to some particular model answer. Each question is worth 10 marks and marks will be awarded according to the following principles:

7-10/10 (First): An analysis that fully addresses the research question, within the limits of the provided data, and that could be presented with little or no revision. The analysis is described in sufficient detail, and conclusions are reasonably mature and well-supported by data graphics and/or R output, without getting lost in details.

5-6/10 (2.2-2.1): An analysis that could be presented after a round of revision. The analysis is reasonably mature and mostly addresses the research question, but contains some minor flaws in terms of presentation and/or quality of the conclusions.

4-5/10 (Third - 2.2): An analysis that demonstrates some understanding of the data science methods introduced and makes a reasonable attempt at addressing the research question, but contains major flaws in the analysis, presentation and/or conclusions.

0-3/10 (Fail): An analysis that does not address the research question, demonstrates little or no understanding of the data science methods introduced, and/or has an incomprehensible/very badly organised presentation.

You should only use the R packages already listed in the R Markdown file. Should you load any additional packages, your mark will be capped at 40% for the question parts they are used for.

Length: There is no minimum or maximum length for this assignment; in marking emphasis will be placed upon clear argument and precision. It is anticipated that a submission addressing all questions will be 10-15 pages long (depending on size of figures). Submissions that are substantially longer are likely to be deducted marks for poor presentation.

Support and advice: You will be able to ask questions during the lecture on Monday 30 October and during tutorials. Questions can also be posted on Moodle or the MA20277 Padlet board. Please formulate your questions such that you don't give away your approach. Questions that are deemed to break this rule will be removed from the Padlet board and will not be answered.

Feedback: Feedback for the overall cohort will be provided within two weeks following the submission deadline. Individual feedback and provisional marks will be released within three weeks following the submission deadline.

Late submission of coursework: If there are valid circumstances preventing you from meeting the deadline, your Director of Studies may grant you an extension to the specified submission date, if it is requested before the deadline. Forms to request an extension are available on SAMIS.

- If you submit a piece of work after the submission date, and no extension has been granted, the maximum mark possible will be the pass mark.
- If you submit work more than five working days after the submission date, you will normally receive a mark of 0 (zero), unless you have been granted an extension.

Academic integrity statement: Academic misconduct is defined by the University as "the use of unfair means in any examination or assessment procedure". This includes (but is not limited to) cheating, collusion, plagiarism, fabrication, or falsification. The University's Quality Assurance Code of Practice, QA53Examination and Assessment Offences, sets out the consequences of committing an offence and the penalties that might be applied.

Contact Details:

Christian Rohrbeck
Office: 6 West 1.18
E-mail: cr777@bath.ac.uk

Background and Tasks

Background: The country of Utopia has collected large amounts of data over the past years, but the country is short of data scientists who can help with analyzing that data. Two organisations from Utopia thus seek your help with addressing some urgent questions. The organisations have provided you with several data sets they believe are useful for the analysis. A detailed data description for the different files is provided at the end of this document. There is no expectation for you to use any data other than that provided.
Reproducibility of results is crucial for the people in Utopia. Therefore, the organisations provided you with a R Markdown template (which includes the questions) and they insist that you use it. You are not allowed to load any R packages other than the ones already specified in the provided R Markdown file. The organisations would like to receive a PDF file that contains all your answers and necessary details of your analysis, including the R code and outputs your answers are based on.

Tasks: Use the data science techniques covered in Chapters 1–3 of the lecture notes to address the following research questions.

(1) The Utopian charity *Respect for Pets* has collected data on cats and dogs for 1990-2023. Utopia only allows three dog breeds, Beagle, Dachshund and Maltese, and all pets have to be registered. The charity would like to gain some insight regarding the following questions:

   (a) How has the number of dogs and cats changed over time? How has the popularity of the different dog breeds evolved since 1990? **[10 marks]**

   (b) Maltese are known to experience respiratory issues, such as wheezing or asthma. How do environmental and physiological factors affect the risk of a Maltese experiencing these issues? **[10 marks]**

(2) The Utopian Fire Department has gathered data on their activities for 2022. They also managed to provide you with access to some data for the houses in Utopia. The Utopian Fire Department asks you to address the following questions:

(a) How does the frequency of the different causes for fires vary over time? How many casualties were attributed to each cause and are there differences in the frequency with which casualties occur across causes? **[10 marks]**

(b) Are there any differences in the risk of fire for the different types of property? What is the relation between the year a property was built and the risk of fire? **[10 marks]**

(c) The Fire Department wishes to run a campaign which encourages home owners to install smoke and carbon monoxide detectors. What does the data reveal regarding the benefits of installing smoke and carbon monoxide detectors? **[10 marks]**

Data Descriptions

Pets.csv: Number of pets registered at the beginning of the month

Year, Month: Year and month the observation refers to

Beagles: Number of registered beagles (including the ones rescue centres)

Dachshund: Number of registered dachshunds (including the ones in rescue centres)

Maltese: Number of registered Maltese (including the ones in rescue centres)

Cats: Number of registered cats (including the ones in rescue centres)

BeaglesRescue: Number of beagles in rescue centres

DachshundRescue: Number of dachshunds in rescue centres

MalteseRescue: Number of Maltese in rescue centres

CatRescue: Number of cats in rescue centres

Cases.csv: Number of Maltese treated by a vet for respiratory issues

Date: Recorded date

Temperature: Recorded daily maximum temperature in degree Celsius

Number: Reported number of Maltese treated for respiratory issues

Maltese.csv: Health records for some of the Maltese in Utopia

Weight: Recorded weight in pounds at last visit to the vet

Height: Recorded height in inches at last visit to the vet

Age: Reported age in years at last visit to the vet

RespiratoryIssues: Whether the dog suffered from respiratory issues at the last visit to the vet (1='Yes', 0='No')

Fires.csv: List of fires recorded by the Utopian Fire Department for 2022

Date: Day and time and the fire was reported

Cause: Cause the fire was attributed to ('Cooking', 'Electrical Fault', 'Heating' or 'Other')

Casualties: Number of casualties

Damage: Pay-out by the insurance company in US dollars

Smoke: Whether the property had a smoke detector (1='Yes', 0='No')

CO: Whether the property had a carbon monoxide detector (1='Yes', 0='No')

RegisterNumber: Identification number of the property as listed in the register of houses/properties

Housing Register.csv: Register of houses/properties for Utopia at the beginning of 2022

ID: Identification number of the property

Year: Year the property was built

Type: Type of property ('Flat', 'Terraced', 'Semi-detached' or 'Detached')
Bedroom: Number of bedrooms (studio apartments are counted as having 1 bedroom)