

MATH 9102 - Probability and Statistical Inference

Assignment 4

Due 6pm, Sunday 5th May, 2024

15 marks

Submission guidelines:

- You will need to upload only **single** R markdown (.Rmd) file.
- File name of your Rmd file must be **YourName_StudentID_CA4**
- Do not upload given datasets (or use zip).
- Make use of R built in datasets (if mentioned in the question). If you have considered external dataset instead of R built in, upload the dataset without zipping it.
- Use the following statement if installing any package:
`if(!require(packageName))install.packages("packageName")`

General Instructions:

- Read the questions carefully and answer all parts to secure full marks.
- Do not ask for direct solutions. This is part of your assessment.
- Assignment will be penalized if you miss any of the submission guidelines.
- Please complete assignment individually and avoid plagiarism as it will lead to penalties and negatively affect your overall grade.
- Including comments/markup in your code to explain what you did and provide answers to all questions.

1. Linear regression with dummy variable [5 marks]

Consider the dataset "weatherhistory.csv".

- a) Describe dependent variable pressure and independent variable temperature.
- b) Explore the relationship between pressure and temperature.
- c) Build a linear model considering temperature and pressure.
- d) Identify a dummy variable and build extended model considering dummy variable.
- e) Report your findings.

2. **Multiple Linear Regression [6 marks]**

Consider the dataset "weatherhistory.csv".

- a) Explore the relationship between pressure and windspeed.
- b) Build a linear model considering (windspeed, humidity, temperature) and pressure.
- c) Assess how model meets key assumptions of linear regression.
- d) Investigate a differential effect by adding dummy variable.
- e) Investigate an interaction effect for windspeed and dummyvariable.
- f) Report your findings.

3. **Logistic regression [4 marks]**

Consider the dataset "heartfailure.csv"

- a) Build a model considering diabetes as predictor.
- b) Calculate and analyze odds ratio of the model.
- c) Extend the model by considering variable age. (Convert the age into categorical data, if age < 55 , category 1; age is between 55 and 68 category 2 ; otherwise category 3)
- d) Report your finding.