UNIVERSITY OF THE FREE STATE
DEPARTMENT OF MATHEMATICAL STATISTICS AND ACTUARIAL SCIENCE
STSM 2634

Tutorial 3

Full marks: 100 (All 7 tutorials together)

Date: 22 April, 2025

Deadline: 24 April, 2025

FOLLOW THESE INSTRUCTIONS METICULOUSLY, OTHERWISE MARKS WILL BE SUBTRACTED:

- Name the answer file as **'Tutorial3_student number'** as the file name. The **code and the output must be included in your answers**.
- **Submit the MS-Word file generated by R-markdown**. Any other form of submission will not be accepted.
- You have freedom to write the code in your own way.
- **No need to print unnecessary long data/output.**
- You are allowed to use the class notes, or any other help from the internet.
- All computations must be done with the help of suitable R functions. Manual or calculator-based answers will not be accepted.
- 0 marks for no submission.

Q1. Consider the ‘airquality’ dataset. This dataset contains daily readings of the air quality measures from May 1, 1973 to September 30, 1973 in New York. Perform a statistical analysis of the data and write a 5-page (at most) report. Your report should contain a brief description of the dataset, basic statistical analysis and exploratory analysis with the help of statistical plots.

[10]

Q2. Use the ‘economics2’ data from the ‘Extra Resource’ section in BlackBoard. A data analytics professional used AI tools to analyse the data and claimed that “Based on the dataset, countries with higher GDP per capita tend to have higher unemployment rates. This is likely due to more advanced labor markets and automation in wealthy economies.”

Do you agree with this claim? Think about a possible critical analysis of the above claim with relevant (i) graphical/visualization tools, (ii) correlation and (iii) regression analysis of the data. (all three distinct methods mentioned above must be used).

[10]

--- End ---