

# UNIVERSITY OF THE FREE STATE

## DEPARTMENT OF MATHEMATICAL STATISTICS AND ACTUARIAL SCIENCE

### STSM 2634

**SEMESTER TEST 1**

**28 March 2024**

**Time: 2h00**

**Marks: 100**

**FOLLOW THESE INSTRUCTIONS METICULOUSLY, OTHERWISE MARKS WILL BE SUBTRACTED:**

- Use R-markdown to answer the questions. Save the answer file in **MS-Word format** (generated by the R-markdown) **using your student number as the author's name and 'STSM2634 Test 1' as the file name**. You need to submit this MS-Word document.
- Any other form of submission will **NOT** be accepted. If your R-markdown codes fail to knit, you will lose 50% of the marks you obtained.
- Use the following packages: **tidyverse**, **MASS**, **ggrepel**, **GGally**, **dplyr**. All plots **should be** created using the **ggplot2** package.
- Create a folder as "Test 1" on the Desktop and save all your work in that folder. Do not use any other drive.
- You are allowed to use the internet for R-helps, **but social media, email, and any use of AI tools are strictly prohibited during the test**.
- The code and corresponding output must be included in your answers failing to which 50% marks will be deducted from the marks you obtained.
- Please make sure that all questions are numbered clearly and correctly. No paperwork is allowed.
- **You are advised to create regular backups during this test**, that is to save your R-code and the output regularly. If you did not create a backup and your documents are lost at the end of the test, nothing can be done.
- It is your responsibility that your answers are submitted correctly.
- Additional instructions will be given at the test venue as needed and will be considered binding.

Q1. Use the dataset 'Occupancy\_data.csv' from the 'Extra Resources' section in BlackBoard. Then answer the following questions. Use the 'classic' theme as background for all the plots.

- (i) Create a boxplot to compare the distribution of CO2 levels between occupied and unoccupied rooms.
- (ii) Plot a histogram of the Humidity values. Use 20 bins and fill the bars in blue with white borders.
- (iii) Create a line chart of Temperature over time.
- (iv) Multiply 100 with all 'Temperature' values and save these values with the same variable name as 'Temperature'. Plot both Temperature and CO2 over time on the same chart using different colors.
- (v) Create a boxplot showing the distribution of Light levels for different occupancy statuses.

[10× 5 = 50]

Q2.

Use the 'mpg' dataset.

- (i) Now create a histogram of the 'displ' variable which is the engine displacement in liters. Add labels for the X and Y axis as 'Displacement' and 'Frequency', respectively. Also add a title for the histogram plot as 'Histogram of the engine displacement in liters'. The histogram bars should be filled with 'blue' color and the borders should be in 'black'. Use a 'classic' theme for the background.
- (ii) Create a subset of the 'mpg' data for the following manufacturers: "corolla", "altima", "jetta", "maxima", "pathfinder 4wd", "grand prix". Now use this subset of the 'mpg' data, then create boxplots for the 'displ' variable (Engine displacement in liters) for different manufacturers. All the boxplots should be in the same plot. Use different colors for different manufacturers. Provide a title of the boxplot as 'Engine displacement for different manufacturers'. Provide X and Y axis labels as 'Manufacturer' and 'Engine displacement in liters', respectively. Use the minimal theme for the plot.

[10× 2 = 20]

Q3. Use the 'airquality' dataset. Create a dataframe with the variables 'Ozone', 'Solar.R', 'Temp' from the dataset. Now create a correlation plot for these variables using the GGally package. Interpret the correlation plot.

[8+2 = 10]

Q4. Write a code using a 'for' loop to calculate the sum of all even numbers between (and inclusive) 1 to 100.

[10]

Q5. Write a code using a 'while' loop that finds the smallest integer such that its square is greater than 1000. The code should print the integer and its square.

[10]

--- END ---