Test 1 – Fall 2025

Instructions

1. Please create your own **answer sheet** (using Word, notebook, etc.) in the format shown below.
2. Save your answer sheet as a **PDF** and submit it **before the end of the test**.
3. For **numerical calculation problems**, **no partial credit** will be given if the final answer is incorrect.
4. Clearly label your answers and plots as instructed.

Answer Sheet Format

Name: _____

Part I: Short Answers

No partial credit will be given for incorrect answers. (Tentatively 70%)

- **Question 1:** _____
- **Question 2:** _____
- **...**
- **Question 7:** _____
- **Question 8:** If the question asks for a plot, please **insert or display the plot here.**
- **...**

Part II: Code and Output

(30%)

Please include your **code and output** for the corresponding questions.
Consistency between your code and answers may be checked for some selected questions, as well as completeness and correctness.

Test Questions

The dataset includes observations on **red and white wine varieties**, focusing on their **chemical properties** and **quality rankings** by tasters.

- All chemical properties are **continuous variables**.

- **Quality** is ranked on a scale from **4 to 7**:
  - 4: “VI” level
  - 5: “V”
  - 6: “VI”
  - 7: “VII”

Each question is worth **5 points**, unless otherwise stated.

1. Construct a plot (e.g., a **correlation plot**) to visually represent the **correlation matrix** of all variables.

2. Calculate the **standard deviation** for the first quantitative variable for each **quality group**. Report the total of the standard deviations (adding all these standard deviations).

3. Compute the **multivariate kurtosis** of all quantitative variables (excluding “Quality”).

4. Let X denote the data matrix containing only quantitative variables. Calculate `X X^T`. What is the **dimension** of the resulting matrix? Report the **number of rows**.

5. For observations with **quality = “VI”**, plot `free.sulfur.dioxide` vs. `total.sulfur.dioxide` and include a **98% prediction region ellipse**. Report the **number of observations outside** this region.

6. Include the **plot (confidence ellipse)** from Question 5.

7. Perform a test to determine whether the **population mean vector** of `free.sulfur.dioxide` and `total.sulfur.dioxide` for **quality ‘V’** wines differs from (15, 50) at a **significance level of 0.05**.

   - Report: `H0`, `H1`, test statistic, p-value, and conclusion. *(10 points)*

8. Conduct an appropriate **multivariate analysis** to test whether the **mean vector of all variables** depends on **quality**.

   - Report: `H0`, `H1`, test statistic, p-value, and conclusion. *(10 points)*

9. Construct a **multivariate QQ plot** for the **first four quantitative variables**. Assess whether the data approximately follow a **multivariate normal distribution**.

10. Test whether the **population covariance matrices** are **equal** across different **quality groups** (equal covariance assumption). Report the p-value.

11. Compute the group means (means of all quantitative variables) for each quality group. Append these group-mean vectors to the dataset as additional cases, then create Chernoff-face plots (face plots) for the extended dataset. Briefly describe notable differences among the quality levels (how the faces differ and what that suggests about the group profiles).

12. Suppose

$$X = (X_1, X_2, X_3, X_4) \sim N_4(\mu, \Sigma), \quad \mu = \begin{bmatrix} 1 \\ 5 \\ 3 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 7 & 1 & 1 & 0 \\ 1 & 9 & 0 & 1 \\ 1 & 0 & 12 & 1 \\ 0 & 1 & 1 & 9 \end{bmatrix}$$

Find the **variance of**

$$Y = 0.2X_1 + 0.3X_2 + 0.2X_3 + 0.2X_4$$

and report the variance.