BT3041 – Analysis and Interpretation of Biological Data

Assignment 1

You are provided with a dataset comprising measurements of chlorophyll levels (chlorophyll a and chlorophyll b) from 100 plant samples collected from two types of forests. Fifty samples were taken from Deciduous forests and the remaining fifty from Evergreen forests. Perform the below mentioned analysis on the given data:

1. Visualize the distribution of chlorophyll a and chlorophyll b values using histogram or density plots in Deciduous forests. (2 marks)
2. Visualize the distribution of chlorophyll a and chlorophyll b values using histogram or density plots in Evergreen forests. (2 marks)
3. Visualize the distribution of chlorophyll a and chlorophyll b values using histogram or density plots without separating the measurements from each forests. (2 marks)
4. Plot 1, 2 and 3 in same plot together and explore how it changes. (2 marks)
5. Calculate summary statistics (mean, median, mode and standard deviation) of chlorophyll a and chlorophyll b measurements from Deciduous forests separately, Evergreen forests separately and both the forests together. (12 marks)
6. In a same boxplot, compare how the distribution of chlorophyll a and chlorophyll b values compare in Deciduous forests and Evergreen forests. (6 marks)
7. Are the variances between chlorophyll a and chlorophyll b measurements differ significantly? Perform appropriate statistical tests to support your claim. Compare variances of chlorophyll content from Deciduous forests separately, Evergreen forests separately and both the forests together. (12 marks)
8. Finally, test whether the mean of chlorophyll a is greater than mean of chlorophyll b using appropriate statistical test in all three combinations: Deciduous forests separately, Evergreen forests separately and both the forests together. Clearly state your null hypothesis, chosen significance criteria and the result of hypothesis testing. (12 marks)

You can download the data from here.

Please provide your analyses results in a report form, specifically answering each of the above 7 questions with relevant plots, etc. Also, do state any assumptions made clearly in the report. Attach the Google drive link to your software codes (MATLAB/Python) used for performing calculations with the report.