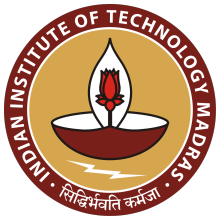
BT3041: Analysis and Interpretation of Biological Data

Assignment 1

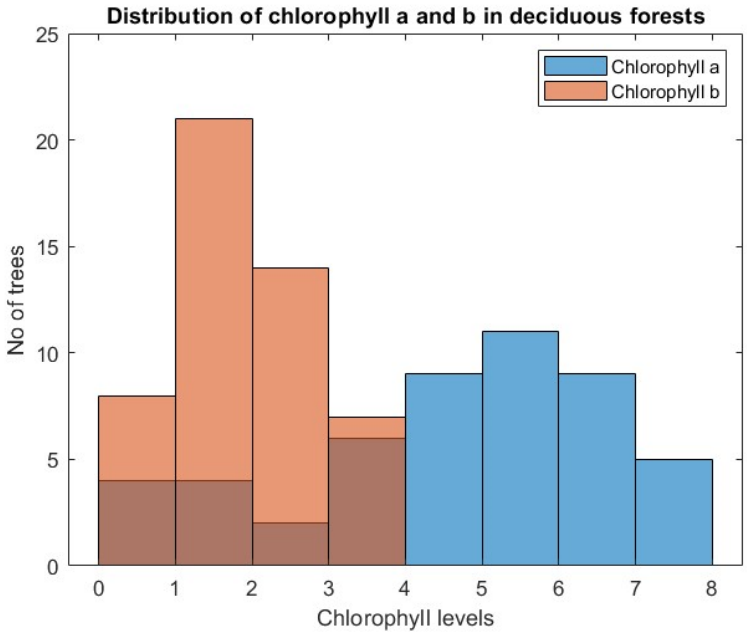
Sumedh Sanjay Kangne (BE21B040)
Department of Biotechnology
Indian Institute of Technology, Madras

A BT3041: Analysis and Interpretation of Biological Data Assignment



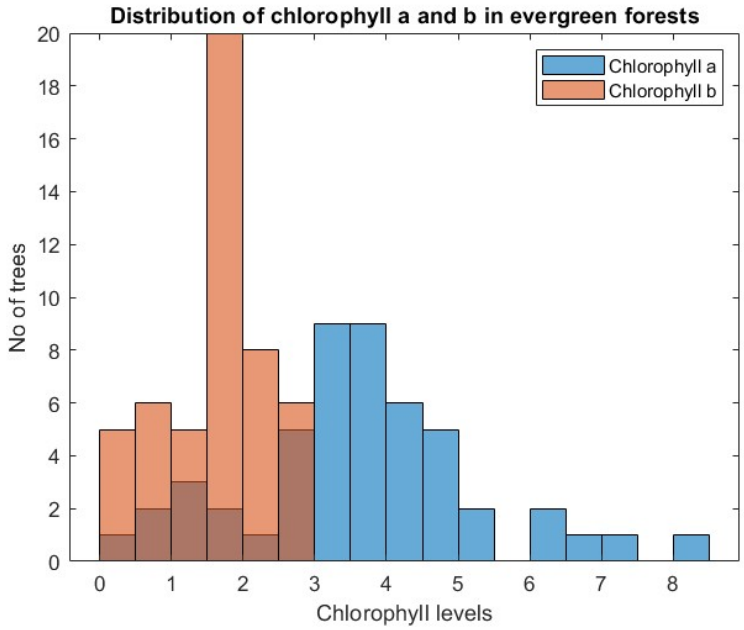
March 9, 2024

1. Visualize the distribution of chlorophyll a and chlorophyll b values using histogram or density plots in Deciduous forests.



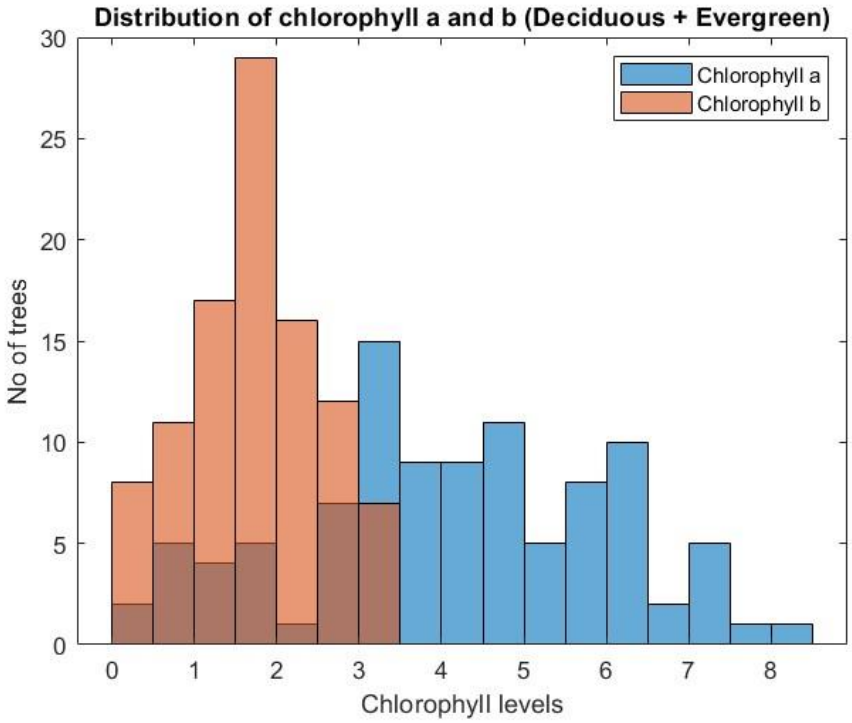
Chlorophyll a shows a left-skewed uniform distribution whereas **chlorophyll b** has a pretty centered uniform distribution

2. Visualize the distribution of chlorophyll a and chlorophyll b values using histogram or density plots in Evergreen forests.



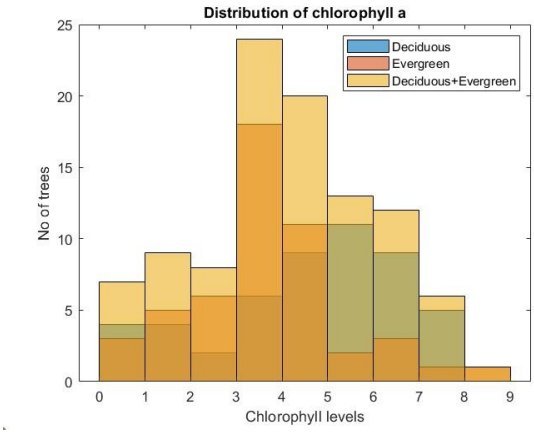
Chlorophyll a shows a right-skewed uniform distribution whereas **chlorophyll b** has a pretty centered uniform distribution.

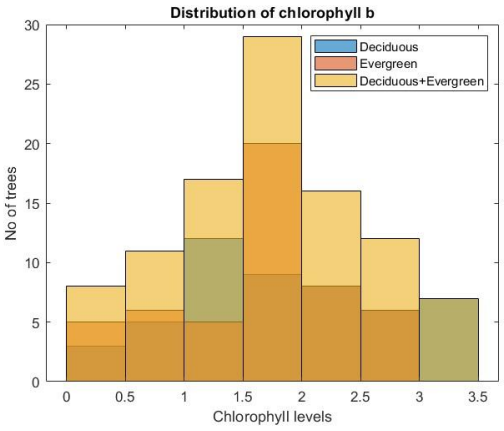
3. Visualize the distribution of chlorophyll a and chlorophyll b values using histogram or density plots without separating the measurements from each forests.



Both of the chlorophyll samples show a centered uniform distribution.

4. Plot 1, 2, and 3 in the same plot together and explore how it changes.





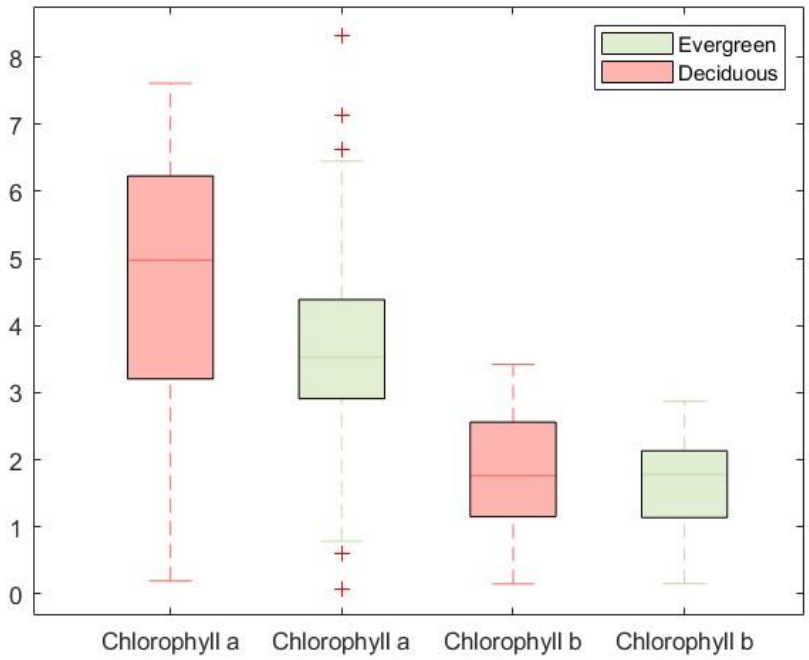
Analysing the histograms we can see that the individually plotted deciduous and evergreen samples for **Chlorophyll a** are uniform distributions skewed to the left and right respectively. When they are taken as one combined sample the distribution is a roughly centred uniform distribution.

Similarly, for **Chlorophyll b** all of them appear to be centered uniform distributions.

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.

```
Mean of chlorophyll a levels in deciduous forest samples:
    4.5644

Mean of chlorophyll b levels in deciduous forest samples:
    1.8178

Mean of chlorophyll a levels in evergreen forest samples:
    3.6257

Mean of chlorophyll a levels in evergreen forest samples:
    1.6663

Mean of chlorophyll a levels in the entire sample:
    4.0950

Mean of chlorophyll b levels in the entire sample:
    1.7421

Median of chlorophyll a levels in deciduous forest samples:
    4.9747

Median of chlorophyll b levels in deciduous forest samples:
    1.7636

Median of chlorophyll a levels in evergreen forest samples:
    3.5315

Median of chlorophyll a levels in evergreen forest samples:
    1.7813

Median of chlorophyll a levels in the entire sample:
    4.0987

Median of chlorophyll b levels in the entire sample:
    1.7694

Mode of chlorophyll a levels in deciduous forest samples:
    0.1947

Mode of chlorophyll b levels in deciduous forest samples:
    0.1499

Mode of chlorophyll a levels in evergreen forest samples:
    0.0714

Mode of chlorophyll a levels in evergreen forest samples:
    0.1550

Mode of chlorophyll a levels in the entire sample:
    0.0714

Mode of chlorophyll b levels in the entire sample:
    0.1499

Standard deviation of chlorophyll a levels in deciduous forest samples:
    1.9447

Standard deviation of chlorophyll b levels in deciduous forest samples:
    0.8550

Standard deviation of chlorophyll a levels in evergreen forest samples:
    1.6261

Standard deviation of chlorophyll a levels in evergreen forest samples:
    0.7116

Standard deviation of chlorophyll a levels in the entire sample:
    1.8529

Standard deviation of chlorophyll b levels in the entire sample:
    0.7902
```

6. In the same boxplot, compare how the distribution of chlorophyll a and chlorophyll b values compare in Deciduous forests and Evergreen forests



Chlorophyll a:

Deciduous forest:	25^{th} %ile = 3.20454	median=4.9747	a75^{th} %ile=6.722761
Evergreen forest:	25^{th} %ile = 2.91088	median=3.5315	a75^{th} %ile=4.38797

Chlorophyll b:

Deciduous forest:	25^{th} %ile = 1.15059	median=1.7636	a75^{th} %ile=2.56148
Evergreen forest:	25^{th} %ile = 1.13786	median=1.7813	a75^{th} %ile=2.13301

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.

```
Null Hypothesis = Variance of both the chlorophyll samples (a and b) for deciduous trees is equal
Alternate Hypothesis = Variance of both the chlorophyll samples (a and b) for deciduous trees is not equal
h = 1
p = 5.227318e-08

Null Hypothesis = Variance of both the chlorophyll samples (a and b) for evergreen trees is equal
Alternate Hypothesis = Variance of both the chlorophyll samples (a and b) for evergreen trees is not equal
h = 1
p = 4.444923e-08

Null Hypothesis = Variance of both the chlorophyll samples (a and b) is equal
Alternate Hypothesis = Variance of both the chlorophyll samples (a and b) is not equal
h = 1
p = 1.083480e-15
```

Using the f-test analysis to check whether the variances of the given samples are equal or not.
Significance criteria, α = 0.05 (Default)

As the function returns a **h value** of 1 for all three cases, we can safely reject the null hypothesis for all of them. Therefore, neither of the samples have an equal population variance

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.

```
Null Hypothesis = Mean of chlorophyll a is greater than mean of chlorophyll b for deciduous forest samples
Alternate Hypothesis = Mean of chlorophyll a is not greater than mean of chlorophyll b for deciduous forest samp
h = 1
p = 1.503502e-13

Null Hypothesis = Mean of chlorophyll a is greater than mean of chlorophyll b for evergreen forest samples
Alternate Hypothesis = Mean of chlorophyll a is not greater than mean of chlorophyll b for evergreen forest samp
h = 1
p = 3.701411e-11

Null Hypothesis = Mean of chlorophyll a is greater than mean of chlorophyll b
Alternate Hypothesis = Mean of chlorophyll a is not greater than mean of chlorophyll b
h = 1
p = 2.530134e-22
```

Using the separate t-test analysis to check whether the population mean of ‘chlorophyll a’ in the given sample is greater than ‘chlorophyll b’.
Significance criteria, α = 0.05 (Default)

As the function returns a **h value** of 1 for all three cases, we can safely reject the null hypothesis for all of them. Therefore, the population mean of ‘chlorophyll a’ values is not greater than ‘chlorophyll b’ in any of the given samples.

Code

The link for the code and all the images used can be found here:
https://drive.google.com/drive/folders/19UIi92eD5XKcX0EjL5y5fkqHave_vGqC?usp=sharing

The End