Shivan Ajay Iyer be22b048

Analysis and Interpretation of Chlorophyll Data

Assignment 1

BE22B048_DAIB_AS1_Deciduos.ipynb

Introduction

This report examines the distribution and statistical characteristics of chlorophyll a and b concentrations from deciduous and evergreen forests. The analysis is based on the given dataset comprising 100 plant samples, with 50 samples collected from deciduous forests and 50 from evergreen forests.

The following information aids in interpreting the results from the code and plots, facilitating a deeper understanding of the data distribution

Chlorophyll a is the primary photosynthetic pigment that directly participates in the light reactions of photosynthesis. In contrast, chlorophyll b serves as an accessory pigment that extends the light absorption spectrum and transfers energy to chlorophyll a.

Evergreen Forests	Deciduous Forests
Trees of Evergreen forests shed their leaves at different times of the year.	In the case of deciduous forests, they have a particular time for their leaves to shed.
Evergreen forests are found in the areas which receive heavy rainfall and have warmer climates.	Are found in a variety of places.
Important evergreen trees are mahogany, ebony and rosewood.	Important deciduous forest trees are sal, teak, peepal, neem and Sheesham.

The dataset had no units for observed values, so I assumed them to be in the standard units of micrograms of chlorophyll per gram of biological matter, i.e. µg/g.

Below are the method of analysis, assumptions and conclusions drawn from the code corresponding to each question in the assignment.

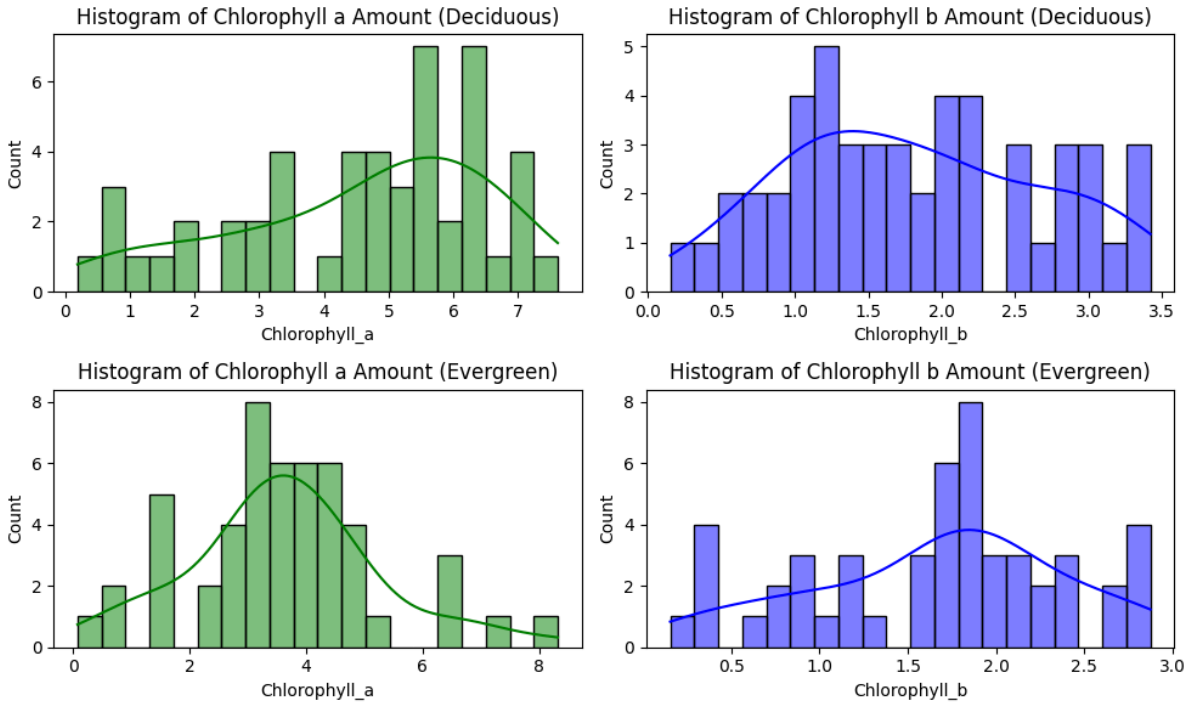
Answer One and Two

Method and Assumptions

Data Filtering: The dataset was first filtered to select the 50 samples corresponding to Deciduous forests.

Visualisation: Two separate histograms were generated for chlorophyll a and b using the Seaborn library.

A consistent number of bins (20) was chosen to allow for a clear comparison between the two measurements

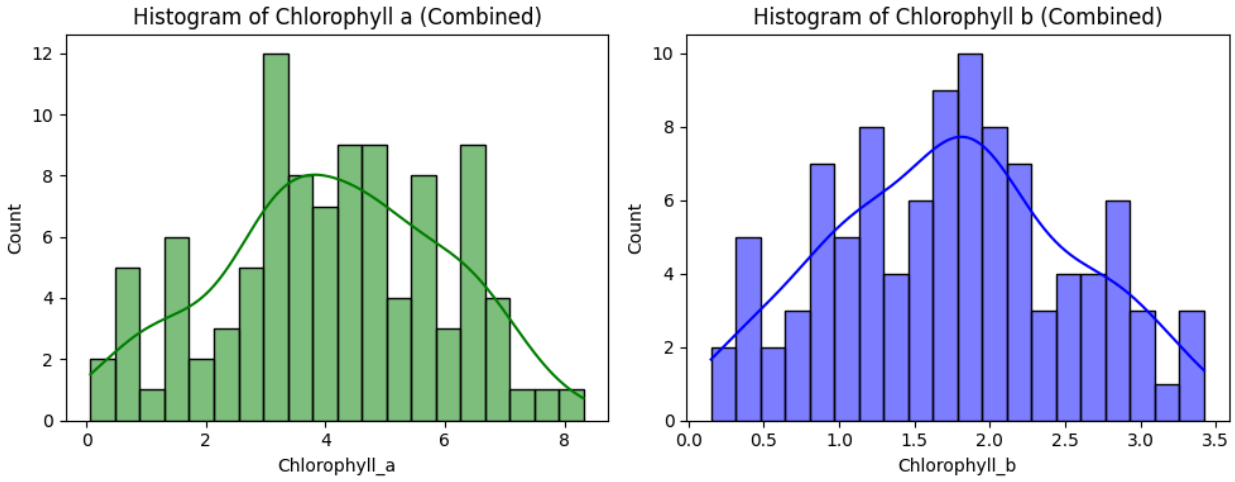


The histograms for deciduous forests indicate distinct distribution patterns for chlorophyll a and b. Chlorophyll a has a broader distribution, predominantly lying between 4-7 μg/g, with notable outliers below 2 μg/g. The distribution appears bimodal, with peaks around 1-2 μg/g and 5-7 μg/g, suggesting potential subgroups that may reflect different species compositions or physiological states.

In contrast, chlorophyll b shows a more concentrated distribution between 1-3 μg/g, with a central tendency around 2 μg/g and a slight positive skew. This narrower range indicates more conservative regulation of this accessory pigment in deciduous plants.

For evergreen forests, chlorophyll-a values primarily fall between 3-5 μg/g, demonstrating a more symmetrical distribution and suggesting a uniform adaptation strategy across species. However, outliers above 6 μg/g and below 2 μg/g indicate some diversity in chlorophyll-a content. Chlorophyll-b in evergreen forests is tightly distributed between 1-2 μg/g, with a slight positive skew and fewer higher values, reflecting adaptation to consistent photosynthetic requirements throughout the year.

Answer Three



When examining the combined distributions of chlorophyll a and chlorophyll b across different forest types, distinct patterns emerge. Chlorophyll a exhibits a broad, multimodal distribution, ranging from approximately 0.5 to 8 μg/g. This variation suggests the presence of different subpopulations linked to specific forest types, indicating the necessity of separate analyses to understand their unique photosynthetic adaptations.

In contrast, chlorophyll b demonstrates a more unimodal distribution, with most values concentrated between 1-3 μg/g. The narrower range of chlorophyll b across various forest types implies that this accessory pigment experiences more consistent selective pressure, likely due to its essential role in the photosynthetic process.

Answer 4

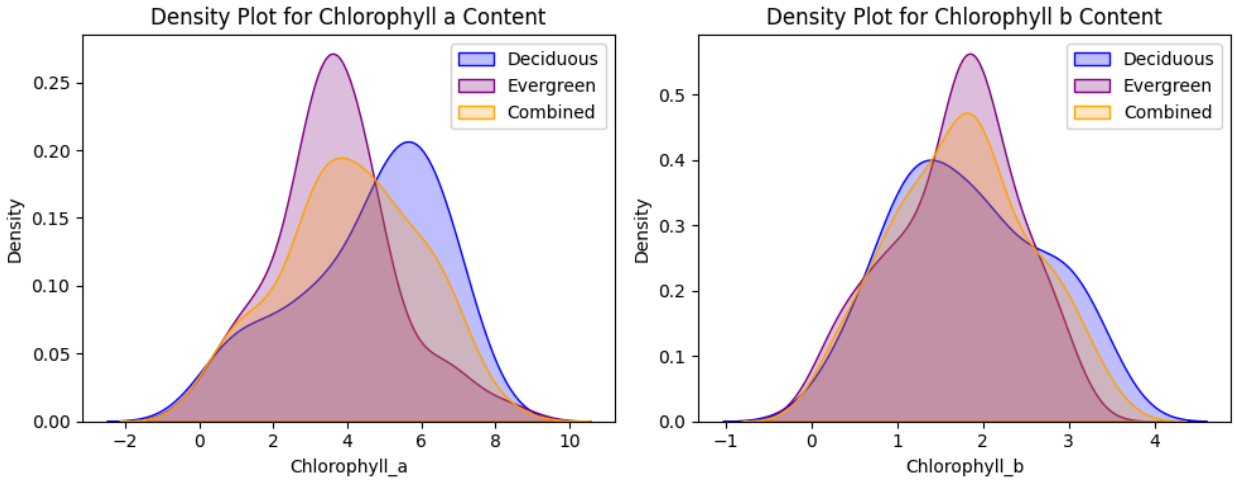
Method and Assumptions

Data Preparation: Density plots were computed from the histogram data for each of the three conditions (Deciduous, Evergreen, Combined).

Visualization: All three density plots were overlaid on a single plot.

Assumptions:

- The kernel density estimation (KDE) used assumes a smooth underlying distribution.
- The same bandwidth was used across groups for comparability.
- The samples are representative of the population distribution



Notable differences in peak locations and spread are observed, providing clues to forest-specific environmental effects and highlight the importance of analysing the data separately.

Answer 5: Summary Statistics

```
Summary Statistics for Deciduous Forests:
Chlorophyll_a: Mean = 4.564, Median = 4.975, Mode = 0.195, Std = 1.964
Chlorophyll_b: Mean = 1.818, Median = 1.764, Mode = 0.150, Std = 0.864


Summary Statistics for Evergreen Forests:
Chlorophyll_a: Mean = 3.626, Median = 3.531, Mode = 0.071, Std = 1.643
Chlorophyll_b: Mean = 1.666, Median = 1.781, Mode = 0.155, Std = 0.719


Summary Statistics for Combined Data:
Chlorophyll_a: Mean = 4.095, Median = 4.099, Mode = 0.071, Std = 1.862
Chlorophyll_b: Mean = 1.742, Median = 1.769, Mode = 0.150, Std = 0.794
```

The statistics demonstrate that deciduous forests typically exhibit higher chlorophyll a and b levels than evergreen forests. The increased standard deviation in chlorophyll

measurements across both forest types reflects it high variability, potentially reflecting its direct response to environmental conditions. Additionally, the chlorophyll a to b ratio is consistently elevated in deciduous forests, suggesting distinct light-harvesting strategies between the two forest types.

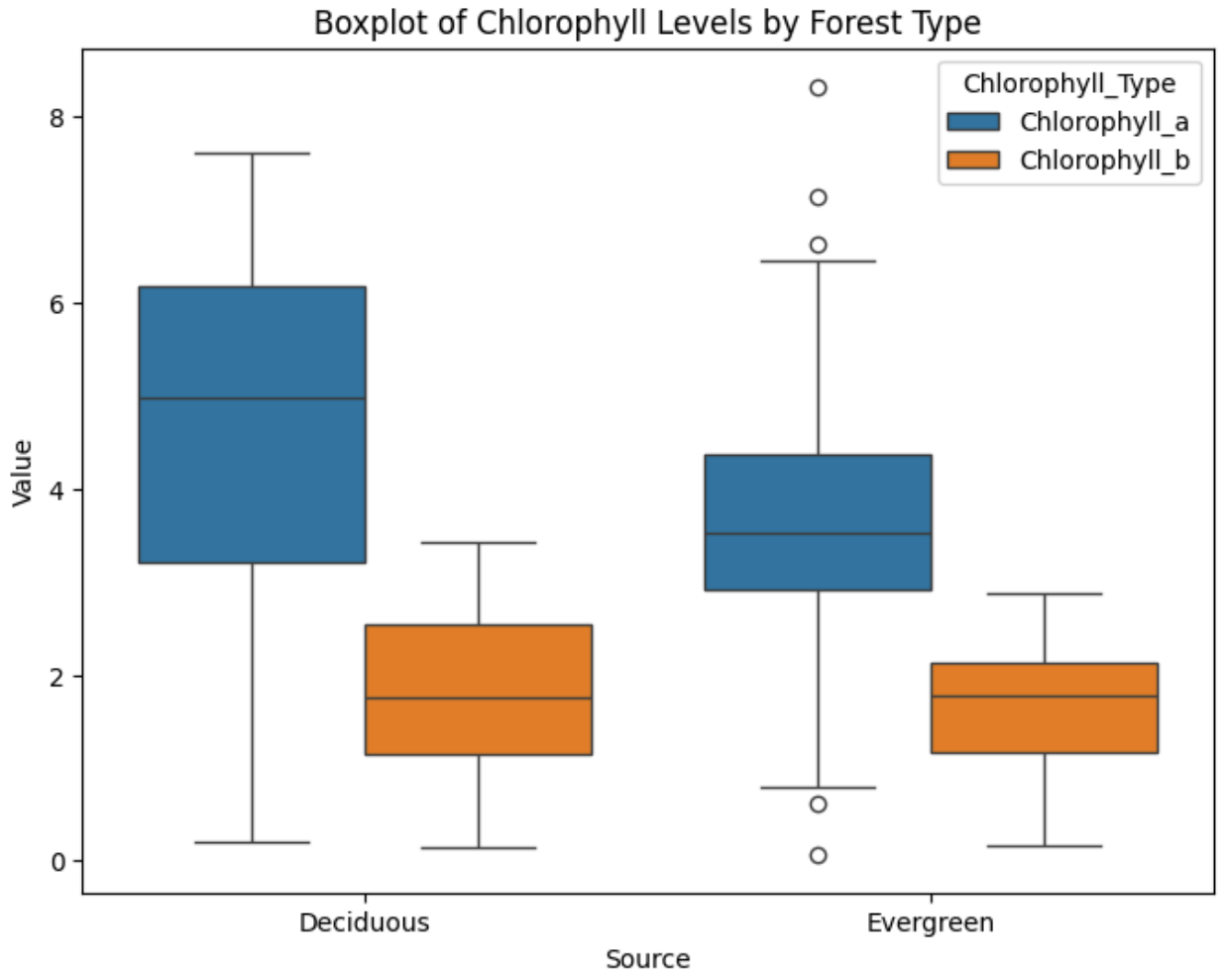
Answer 6: Boxplot

Method and Code Overview

Visualization: A combined boxplot was generated to compare chlorophyll a and b distributions across both forest types.

Assumptions:

- The boxplot assumes that the quartile summary (25th, median, 75th percentiles) is a meaningful representation of the data distribution.
- Outliers are identified using the standard 1.5*IQR rule.



The boxplot comparison of chlorophyll a and b between forest types provides a visual representation of the differences mentioned above. Deciduous forests display not only higher median values for chlorophyll a but also a wider interquartile range, indicating greater diversity in photosynthetic capacity among deciduous species. This variation likely reflects adaptations to the seasonal changes in light availability that deciduous forests experience.

Evergreen forests show more consistent chlorophyll levels with narrower interquartile ranges, particularly for chlorophyll b. This consistency aligns with the year-round photosynthetic activity of evergreen species, which requires more stable pigment compositions to maintain photosynthesis through different seasons.

Answer 7

Testing Strategy

Separate tests were conducted for Deciduous, Evergreen, and combined datasets.

The null hypothesis in each test was that the variances of chlorophyll a and b were equal.

Assumptions:

- The F tests assume independent samples, random sampling and that the sample is normally distributed.

Conclusions from Analysis

Statistical tests comparing the variance of chlorophyll a and b measurements indicate significant differences across all three scenarios: deciduous, evergreen, and combined. The results of the F-test consistently demonstrate that chlorophyll a exhibits greater variance than chlorophyll b in each forest type. This observed difference supports the hypothesis that chlorophyll a, as the primary photosynthetic pigment, displays more plasticity in response to environmental conditions. In contrast, chlorophyll b maintains more stable levels to fulfil its role as an accessory pigment.

In deciduous forests, the variance in chlorophyll a is 3.39, significantly higher than that of chlorophyll b at 0.71, yielding an F-ratio of 4.77 ($p < 0.001$). In evergreen forests, chlorophyll a demonstrates a variance of 2.69, again surpassing chlorophyll b's variance of 0.45, resulting in an F-ratio of 5.98 ($p < 0.001$). These significant differences affirm the distinct regulatory mechanisms governing the two pigments within each forest type.

Answer 8

Method and Assumptions

Hypothesis Testing: For each forest type (Deciduous, Evergreen, and combined), a one-tailed, separate variance t-test was used since the hypothesis is directional—we want to test specifically if the mean of Chlorophyll a is greater than that of Chlorophyll b. The inbuilt function in the scipy.stats library *stats.ttest_ind* was used.

Null Hypothesis (H_0): Mean (Chlorophyll a) = Mean (Chlorophyll b)

Alternative Hypothesis (H_1): Mean (Chlorophyll a) > Mean (Chlorophyll b)

Assumptions:

- Data are approximately normally distributed or the t-test is robust enough for the sample size.
- The tests assume independence between the two chlorophyll measurements within each group.
- Significance Criteria: A significance level (α) of 0.05 was used.

Conclusions from Analysis

The analysis shows that chlorophyll-a concentrations are significantly higher than chlorophyll-b in both forest types ($p < 0.001$ for all comparisons).

The mean difference between chlorophyll a and b is more pronounced in deciduous forests (2.750 µg/g) than in evergreen forests (1.96 µg/g). This larger gap in deciduous species may relate to their need for more efficient light harvesting during their limited growing season, requiring a higher proportion of the primary photosynthetic pigment.