Homework 1

Giannopoulou Ilektra Eleni - 23202690

2024-06-07

"Analysis of the Iris Dataset"

Introduction

The iris dataset provides the measurements in centimeters of the variables sepal length and width, and petal length and width, for 50 flowers from each of three species of iris. The species are **Iris setosa**, **Iris versicolor**, and **Iris virginica**.

Data Overview

The dataset consists of 150 observations with five variables: Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species. Each variable represents a measurement in centimeters, except for Species, which is a factor variable with three levels.

Descriptive Statistics:

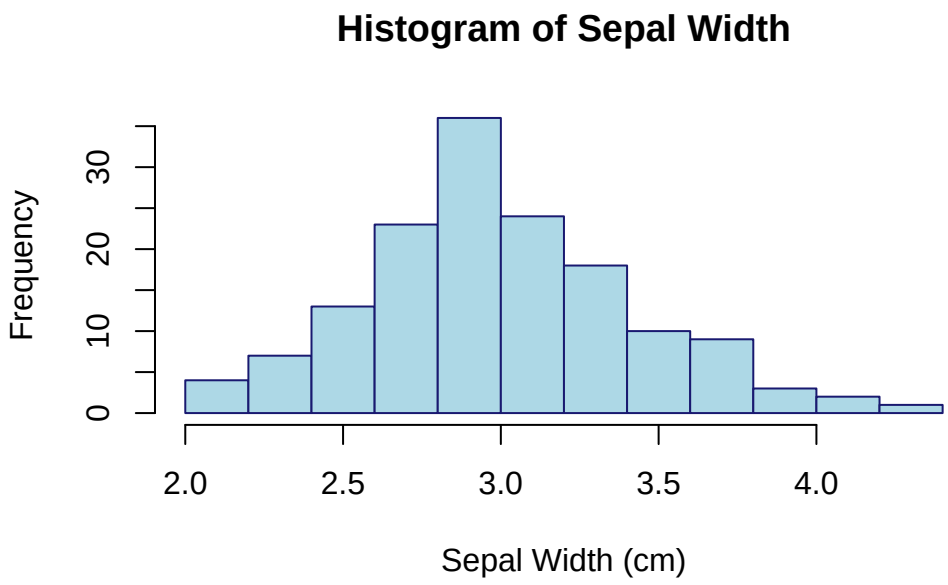
Table 1: Summary Statistics of Iris Dataset

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300
Median :5.800	Median :3.000	Median :4.350	Median :1.300
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199
3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800
Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500

Histogram of Sepal Width

Approach Explanation:

To analyze the iris dataset, first of all we need to have loaded the data using the **data()** function. Next, we will create a histogram of the *Sepal Width* variable to visualize its distribution. To accomplish this task, we will use the **hist()** function with custom colors and border parameters to improve the visual appeal of the plot.



This histogram illustrates the distribution of *Sepal Width* across the iris dataset, showing a **normal distribution** with a central peak around the mean. The most frequent sepal width values range between about 2.5 and 3.5 cm confirming the result of the summary statistic above.