ESE 326: Final Project

Eric Stewart & Gabriel Minton

November 2024

1 Introduction

The maiin objective of this project involoved R's built-in dataset, Iris. Through graphical exploration and mathematical analysis, the researchers determine whether there are clear rules as to which of the features determine the species of a given specimen.

2 Methods

2.1 Exploratory Analysis

2.2 Confidence Interval Estimate

2.3 Hypothesis Test

$\sigma_1 == \sigma_2$ T-test using Sp:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p{}^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim T_{n_1 + n_2 - 2} \tag{1}$$

$$S_p{}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \tag{2}$$

$\sigma_1 \neq \sigma_2$ T-test using γ:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim T_\gamma \tag{3}$$

$$\gamma = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} \tag{4}$$
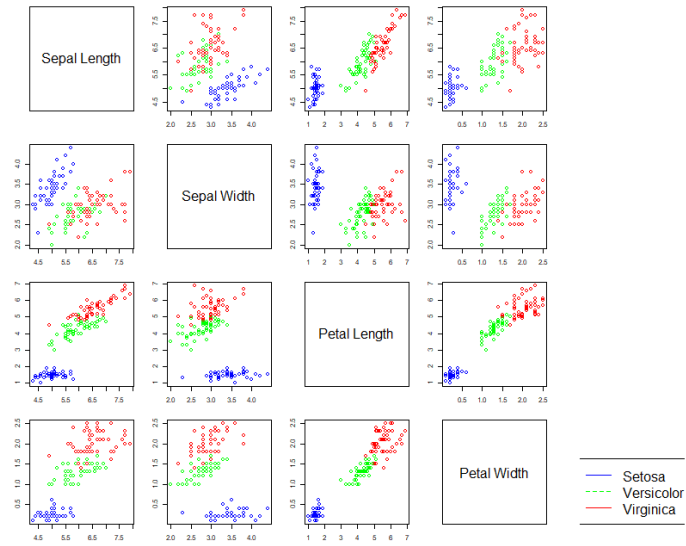
3 Results and Observations

Figure 1: A visualization of the Iris dataset showing scatterplots of each pair of the features.

(a) Petal Lengths

(b) Petal Widths

(c) Sepal Lengths

(d) Sepal Widths

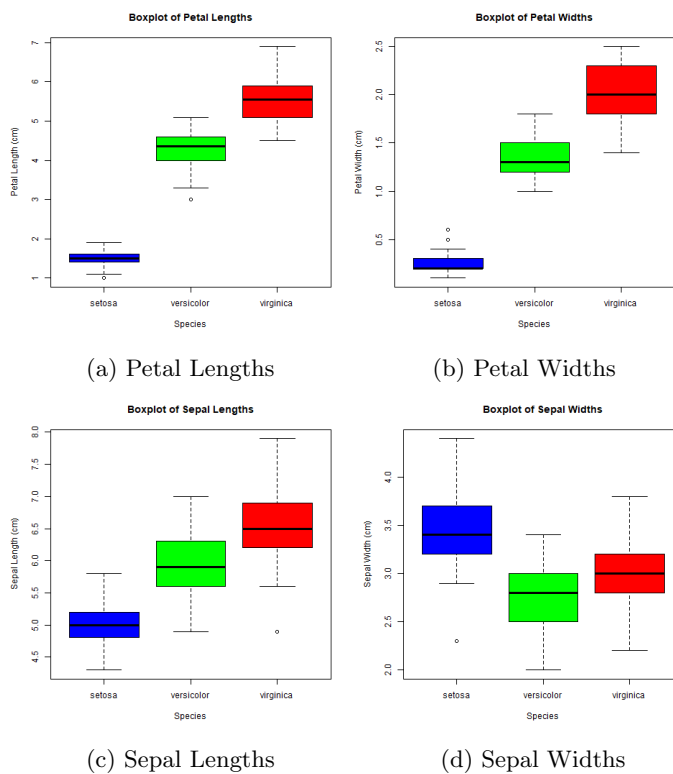Figure 2: Boxplots for each of the four features.

Figure 2a is petal length

4   Conclusions

5   Appendix

5.1   R-scripts

5.2   Extra Figures and Tables