**Project 2018**

**Programming and Scripting**

**Fisher’s Iris dataset**

**Student Name: Somanathan Subramaniyan**

**Student Id: G00364742**

**Contents**

* **Problem Statement**
* **R.A Fisher and Irish Data Set**
* **Data Set - Analysis and Investigation**
* **Conclusion**
* **References**

**Problem statement**

This project requires researching the Fisher’s Iris data set, and then writing documentation and code in the Python programming language based on that research. The below information and programming tasks to be performed as part of the project.

* Background information and summary of the Iris Data set
* Download the data set and write Python code to investigate it.
* Summarise the data set and document the investigations.
* Document the supporting tables and graphics.
* Document the references used

**R.A Fisher and Irish Data Set**

**Sir Ronald Aylmer Fisher** (R.A Fisher) was a British statistician and geneticist. His work in statistics created the foundations for modern statistical science and considered as most important person in 20th century statistics.

The key contributions of R.A Fisher are listed below

* One of the key founders of population genetics
* Fisher’s principle
* Fisherian Runaway or runaway selection
* Sexy son hypothesis
* Analysis of Variance (AVOVA)

**Irish Data Set:**

The Iris flower is a multivariate data set introduced by the R.A Fisher in his 1936 paper and it is sometimes called Anderson's Iris data set because Edgar Anderson collected the data to quantify the morphologic variation of Iris flowers of three related species.

Two of the three species were collected in the Gaspé Peninsula and picked on the same day and measured at the same time by the same person with the same apparatus.

The dataset contains

* 150 records in total for 3 species of iris flower
* 50 samples from each of Iris flower namely Iris setosa, Iris virginica and Iris versicolor.
  + Four features were measured from each sample
    - Petal length
    - Petal width
    - Sepal length
    - Sepal width
  + The length and the width of the sepals and petals, in centimetres.

Based on the combination of these four features, R.A Fisher developed a linear discriminant model to distinguish the species from each other.

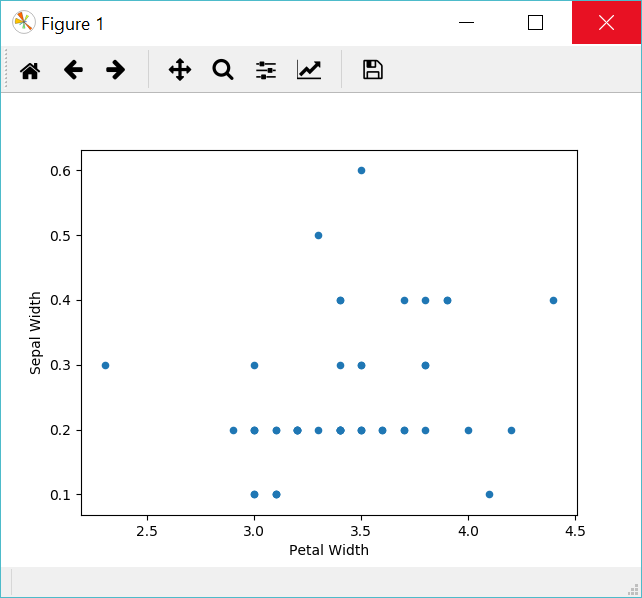
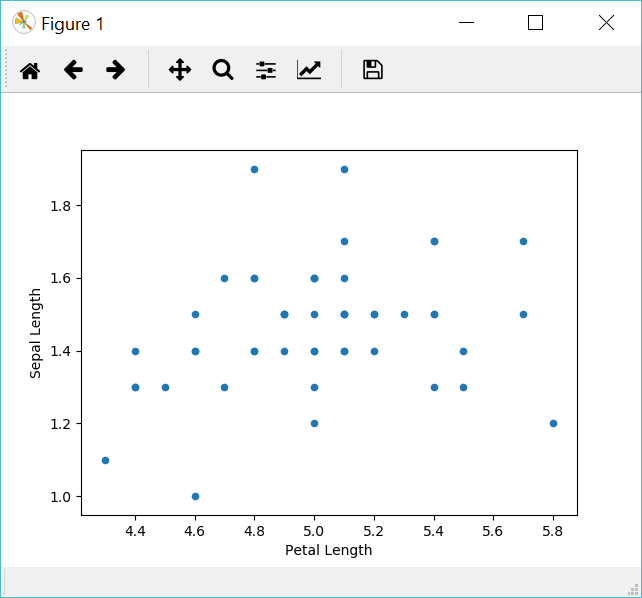
**Data Set - Analysis and Investigation**

**Iris Setosa**

Python file name: Iris-setosa.py

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Category | Petal Length | Petal Width | Sepal Length | Sepal Width |
| count | 50 | 50 | 50 | 50 |
| mean | 5.006 | 3.418 | 1.464 | 0.244 |
| std | 0.35249 | 0.381024 | 0.173511 | 0.10721 |
| min | 4.3 | 2.3 | 1 | 0.1 |
| 25% | 4.8 | 3.125 | 1.4 | 0.2 |
| 50% | 5 | 3.4 | 1.5 | 0.2 |
| 75% | 5.2 | 3.675 | 1.575 | 0.3 |
| max | 5.8 | 4.4 | 1.9 | 0.6 |

Scatter Plot:



Box Plot:

