

Assignment 1: Case 2

Jyotsna Dalal, Pallav Walia, Shashank Singhal

04/02/2020

DATA

The data gives us information about three different categories of flowers and the properties of their petals. Each species of flower has 50 samples accounting to a total of 150 different observations.

```
## 'data.frame':    150 obs. of  3 variables:
## $ petal_length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ petal_width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ class       : Factor w/ 3 levels "Iris-setosa",...: 1 1 1 1 1 1 1 1 1 1 ...
```

DATA INTRODUCTION

This dataset consists of mainly three types of variables viz the class(species) of flowers, the length and the width of their petals. The variables petal length and petal width have numeric values whereas class is a factor. For instance, one of the classes of flower1, Iris-setosa has a petal length of 1.4 and petal width of 0.2.

```
head(flower1)
```

```
##   petal_length petal_width      class
## 1         1.4         0.2 Iris-setosa
## 2         1.4         0.2 Iris-setosa
## 3         1.3         0.2 Iris-setosa
## 4         1.5         0.2 Iris-setosa
## 5         1.4         0.2 Iris-setosa
## 6         1.7         0.4 Iris-setosa
```

DATA SUMMARY

The summary of the dataset gives us information about the central tendency indicating the probability distribution. We can observe from the summary statistics that the mean of petal length is 3.759 and that of petal width is 1.199 which means the average of all values of petal length is more than the average of all values of petal width. Additionally, the range of petal length is from 1 to 6.9 whereas the range of petal width is from 0.1 to 2.5. Furthermore, we can see that petal length and petal width are positively correlated. The petal width increases with an increase in petal length.

```
##   petal_length  petal_width      class
## Min.   :1.000   Min.   :0.100   Iris-setosa   :50
## 1st Qu.:1.600   1st Qu.:0.300   Iris-versicolor:50
## Median :4.350   Median :1.300   Iris-virginica :50
## Mean   :3.759   Mean   :1.199
## 3rd Qu.:5.100   3rd Qu.:1.800
## Max.   :6.900   Max.   :2.500
```

Correlation

```
## [1] 0.9627571
```

In the below table we have shown the maximum and minimum length and width of each flower. According to the sample data Iris-setosa has the minimum length and width whereas the Iris-Virginica has the maximum length and width.

```
unique(nfpl1)
```

```
## # A tibble: 3 x 5
## # Groups:   class [3]
##   class      max_len max_width min_len min_width
##   <fct>      <dbl>    <dbl>   <dbl>    <dbl>
## 1 Iris-setosa      1.9      0.6     1      0.1
## 2 Iris-versicolor  5.1      1.8     3      1
## 3 Iris-virginica   6.9      2.5    4.5     1.4
```

DATA EXPLORATION

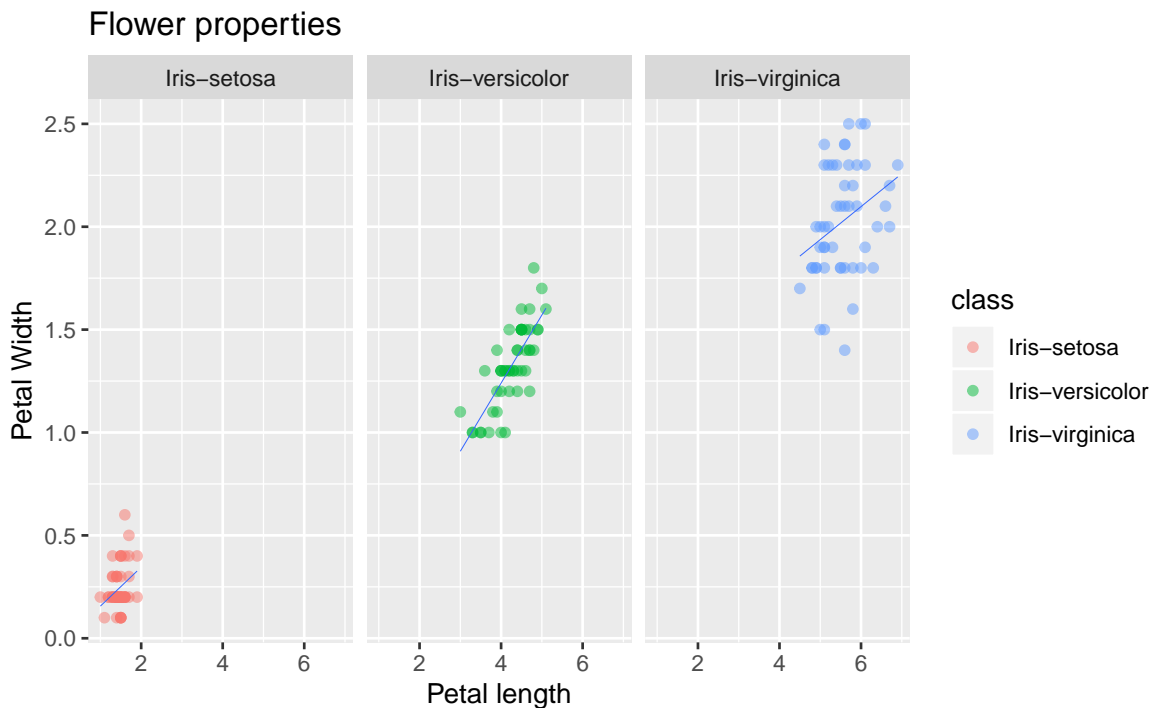
From the graph, we can infer that the median of Iris Virginica is more than the median of Iris versicolor and Iris setosa. Even though the median is higher in Iris virginica there are some petal widths of Iris versicolor that are more than the petal width of Iris Virginica.



DATA ANALYSIS

In order to compare the flowers based on their petal length and petal width, we make three scatter plots with petal length on the x axis and petal width on the y axis based on the class of the flower viz Iris-setosa, Iris-versicolor and Iris-Virginica. From the graph, we can deduce that Iris-setosa falls under the smallest category of flowers when compared with Iris-versicolor and Iris-Virginica. This is mainly due to their small petal size and petal width. The Iris-versicolor is bigger in size than Iris-setosa.

Iris-virginica can have the biggest size when compared to Iris-setosa and Iris-versicolor, although there are some Iris-versicolor which have more petal length and petal width than Iris-virginica. This type of graph also helps us to classify different species of flowers based on their petal length and petal width.



The table below gives us a brief information about the three classes of flower and the mean and standard deviation of their petal length and petal width. It can be seen that Iris-setosa has the smallest mean length of 1.464 and mean width of 0.244 as compared to Iris-versicolor and Iris-virginica. The standard deviation of petal length and petal width of Iris Virginica is the highest with 0.55 and 0.27 which means that it has the highest spread of data over a given range of data. On the other hand, Iris-versicolor has a standard deviation of petal length and petal width of 0.46 and 0.19, which means the data is more spread out than Iris-setosa.

```
## # A tibble: 3 x 5
## # Groups:   class [3]
##   class      mean_length mean_width sd_length sd_width
##   <fct>          <dbl>      <dbl>    <dbl>    <dbl>
## 1 Iris-setosa      1.46        0.244    0.174    0.107
## 2 Iris-versicolor  4.26        1.33     0.470    0.198
## 3 Iris-virginica   5.55        2.03     0.552    0.275
```

From the results of the following table, we can be 95percent confident that the petal width of Iris-setosa lies between 0.231 and 0.2745 and the petal length is between 1.415 and 1.513 as compared to the petal width of Iris-Virginica which lies between 1.9840 and 2.1040 and the petal length between 5.395 and 5.709. We notice that there is no overlapping between the three classes of flower. Hence, we can say that Iris-Virginica has the largest petal and Iris-setosa has the smallest petal.

```
s=groupwiseMean(petal_length~class,data=flower1,conf=.95,digits = 4)
s
```

##	class	n	Mean	Conf.level	Trad.lower	Trad.upper
## 1	Iris-setosa	50	1.464	0.95	1.415	1.513
## 2	Iris-versicolor	50	4.260	0.95	4.126	4.394
## 3	Iris-virginica	50	5.552	0.95	5.395	5.709

```
l=grouppwiseMean(petal_width~class,data=flower1,conf=.95,digits = 4)
```

##	class	n	Mean	Conf.level	Trad.lower	Trad.upper
## 1	Iris-setosa	50	0.244	0.95	0.2135	0.2745
## 2	Iris-versicolor	50	1.326	0.95	1.2700	1.3820
## 3	Iris-virginica	50	2.026	0.95	1.9480	2.1040

CONCLUSION

Based on the data analysis, it is fair to conclude that different classes of flowers have different sizes. The three categories of flowers namely, Iris-setosa, Iris-versicolor and Iris Virginica can be classified based on their petal length and petal width. It can be concluded that the mean petal length and petal width of Iris-virginica is the largest and that of Iris-setosa is smallest. Furthermore, there is more variation in data of Iris-virginica than in Iris-versicolor and Iris-setosa. To sum up, we can say that most of Iris Virginica tend to be bigger in size and Iris setosa tend to be the smallest in size when compared between the three classes of flowers.