

Iris data

Using R studio

Marwan Mahmoud Ibrahim sec [17]

Shereen Mohamed Ibrahim sec[9]

Aya Rabea Said [5]

Year 2

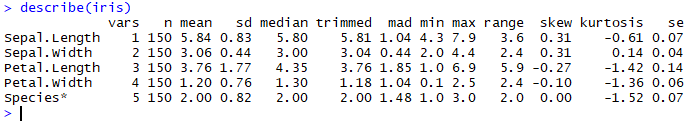
**By Doctor Maryam Nabil**

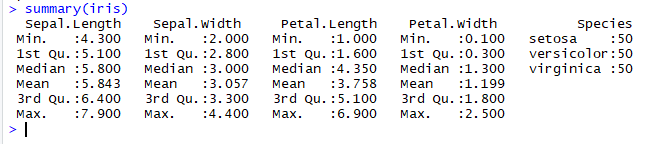
**Why we choose this project and how will we analyze it using the topics we'd covered during the lectures?**

We have chosen this project due to it’s easy, non-complicated idea, besides we want to deal practically with statistical analysis through this project.

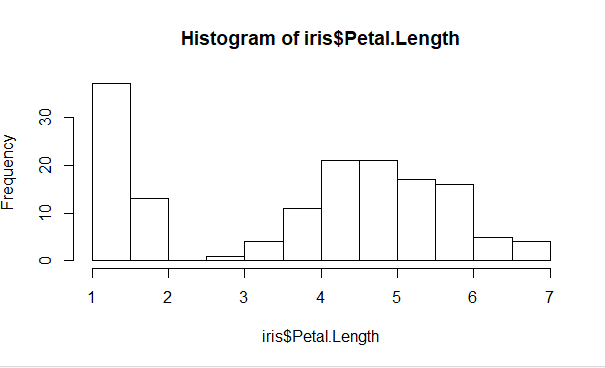
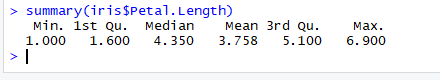
we planned to analyze this data by using "R Studio" software using the ideas we had discussed in the lecture, we analyzed the whole idea in 3 points as follows:

1. Descriptive Analysis:   
   As you taught us; what we do with data which we called "Descriptive Statistics", Furthermore, we have shown how data measurements vary across observations by showing measures of central tendency and variability also by visual "Histogram" graph (as it is quantitative interval data set). Finally, those were the outputs:





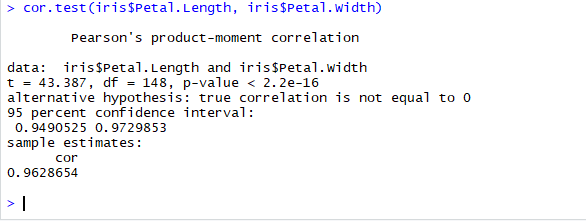
Now we will study one case of the table (Petal.length)

1. 
2. Correlations:

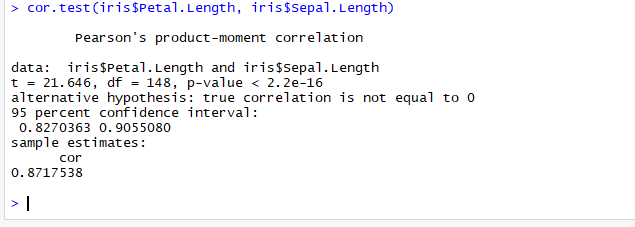
In this point, we wanted to show how the variables are related to each other, we identified relationships between the measurements by using the most common measure of correlation "Pearson's correlation" which is also called "Pearson's r".

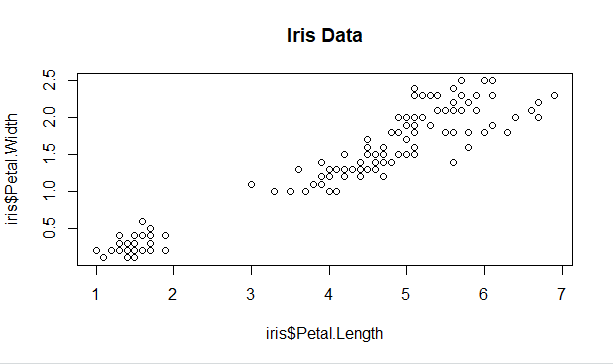
As a conclusion from the "r" value which was 0.9 as we will see in the following picture, we noticed that there is a high positive relationship between the variables. However we can predict it theoretically from the previous box plot.

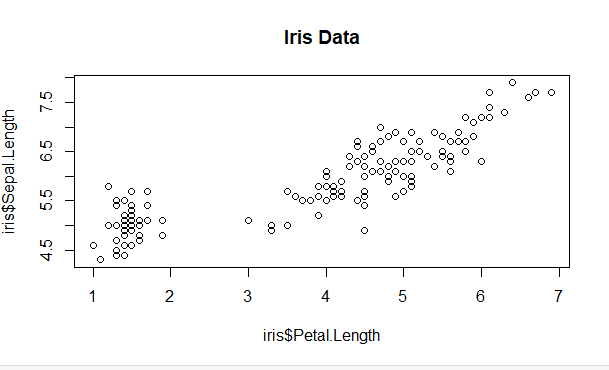
So we now make the correlation for the petal in a general way



And now we make the correlation for the length in general way :



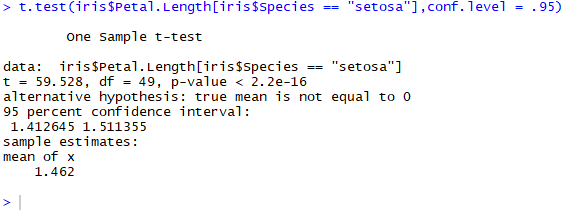




1. Inferential Statistics:

This was our last point to be analyzed, we tested hypotheses about our data

We will study small sample from it (setosa) and will continue our statistics on petal. Length.



Conclusion

This report introduced how to deal with "quantitative -interval data set" statistically by using R language, also it answered our mentioned question. We presented information about how to calculate measures of central tendency & dispersion. Also test hypotheses for a certain confidence interval by means of algorithms we had at the lectures. This would be changed a little bit for another test cases with different variables and confidence interval percent.