The Iris Data Set

Fabian Schroeder

September 13, 2017

Description

This famous dataset (FISHER 1936, Anderson (1936)) gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species or iris. The species are *Iris setosa, versicolor*, and *virginica.*

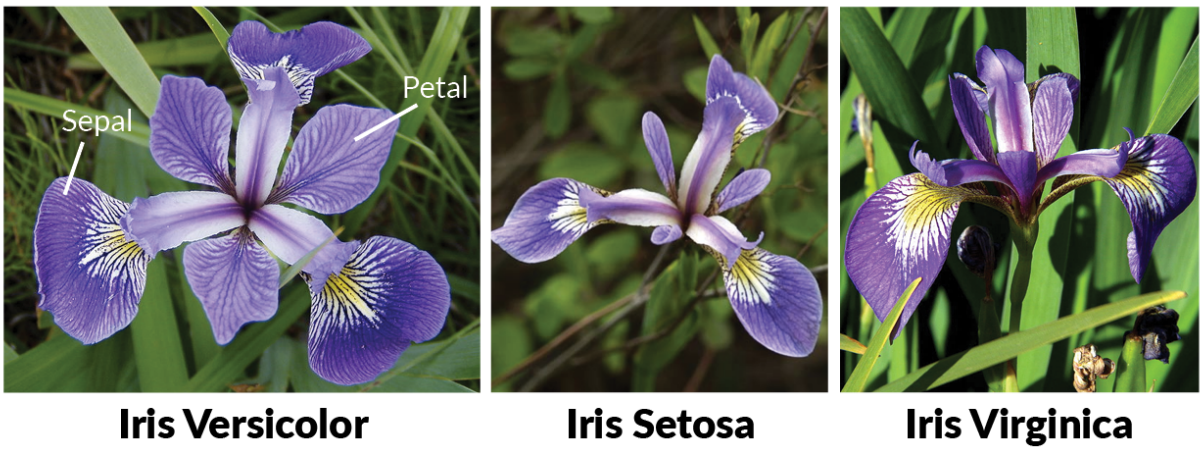


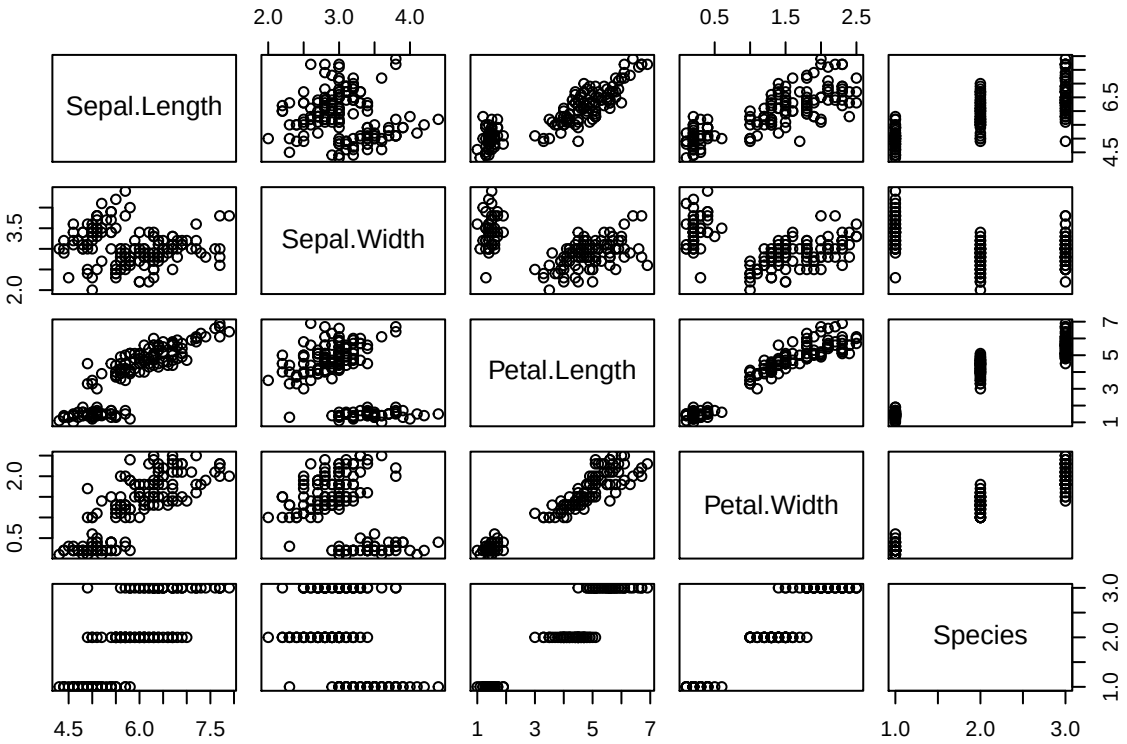
Figure 1: Illustration of the Variables of the iris data set.

Iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100	setosa :50
2	1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300	versicolor:50
3	Median :5.800	Median :3.000	Median :4.350	Median :1.300	virginica :50
4	Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199	
5	3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800	
6	Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500	

Table 1: Summary of the data set

Scatterplot Matrix



Logistic Regression Analysis

Table 2: Regression Results

	Dependent variable:	
	Is.Versicolor	Is.Virginica
	(1)	(2)
Petal.Length	0.148***	0.184***
	(0.043)	(0.031)
Sepal.Length	−0.230**	0.023
	(0.092)	(0.067)
Constant	1.119***	−0.490*
	(0.407)	(0.294)
Observations	150	150
Log Likelihood	−94.823	−45.951
Akaike Inf. Crit.	195.646	97.902
Note:	*p<0.1; **p<0.05; ***p<0.01	

References

Anderson, Edgar. 1936. "The Species Problem in Iris." *Annals of the Missouri Botanical Garden* 23 (3). Missouri Botanical Garden Press: 457–509. http://www.jstor.org/stable/2394164.

FISHER, R. A. 1936. "THE Use of Multiple Measurements in Taxonomic Problems." *Annals of Eugenics* 7 (2). Blackwell Publishing Ltd: 179–88. doi:10.1111/j.1469-1809.1936.tb02137.x.