

Tugas8

Kenneth Manuel (160419041), Jehuda Rivaldo (160419133)

4/24/2021

```
library("readxl")
library("factoextra")
library("dplyr")
library("MASS")
library("nnet")
library(purrr)
library(caret)
library(psych)
```

```
dat <- read_excel('irisdata.xlsx')
data <- dat[2:6]
data1 <- as.data.frame(map_if(data, is.character, as.factor))
str(data1)
```

```
## 'data.frame':   150 obs. of  5 variables:
##  $ Sepal.Length: num   5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width: num   3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ 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 ...
##  $ Species      : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

```
head(data1)
```

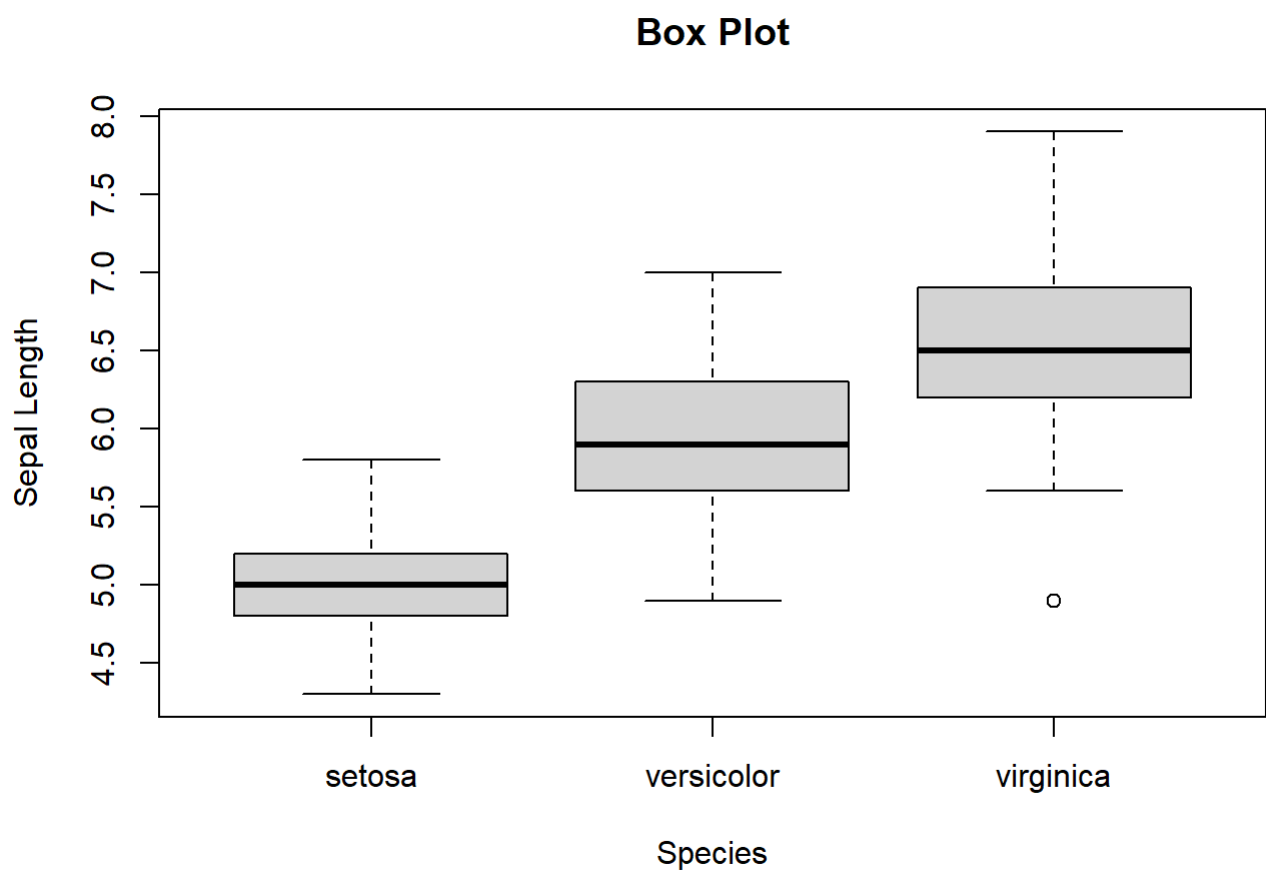
	Sepal.Length<dbl>	Sepal.Width<dbl>	Petal.Length<dbl>	Petal.Width<dbl>	Species<fct>
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

6 rows

```
summary(data1)
```

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

```
boxplot(Sepal.Length ~ Species, data = iris, main = "Box Plot", xlab = "Species", ylab = "Sepal Length")
```



```
#partition data (mengumpulkan data iris untuk diuji)
set.seed(100)
ind <- createDataPartition(iris$Species,p=0.80,list = F)
data_train <- iris[ind,]
test <- iris[-ind,]

dim(data_train)
```

```
## [1] 120 5
```

```
dim(test)
```

```
## [1] 30 5
```

```
#data set train memiliki 120 dan data uji 30
```

```
#PCA (membuat bentuk dari analisis PCA)
data_pc <- prcomp(data_train[,5], center = T, scale = T)
data_pc
```

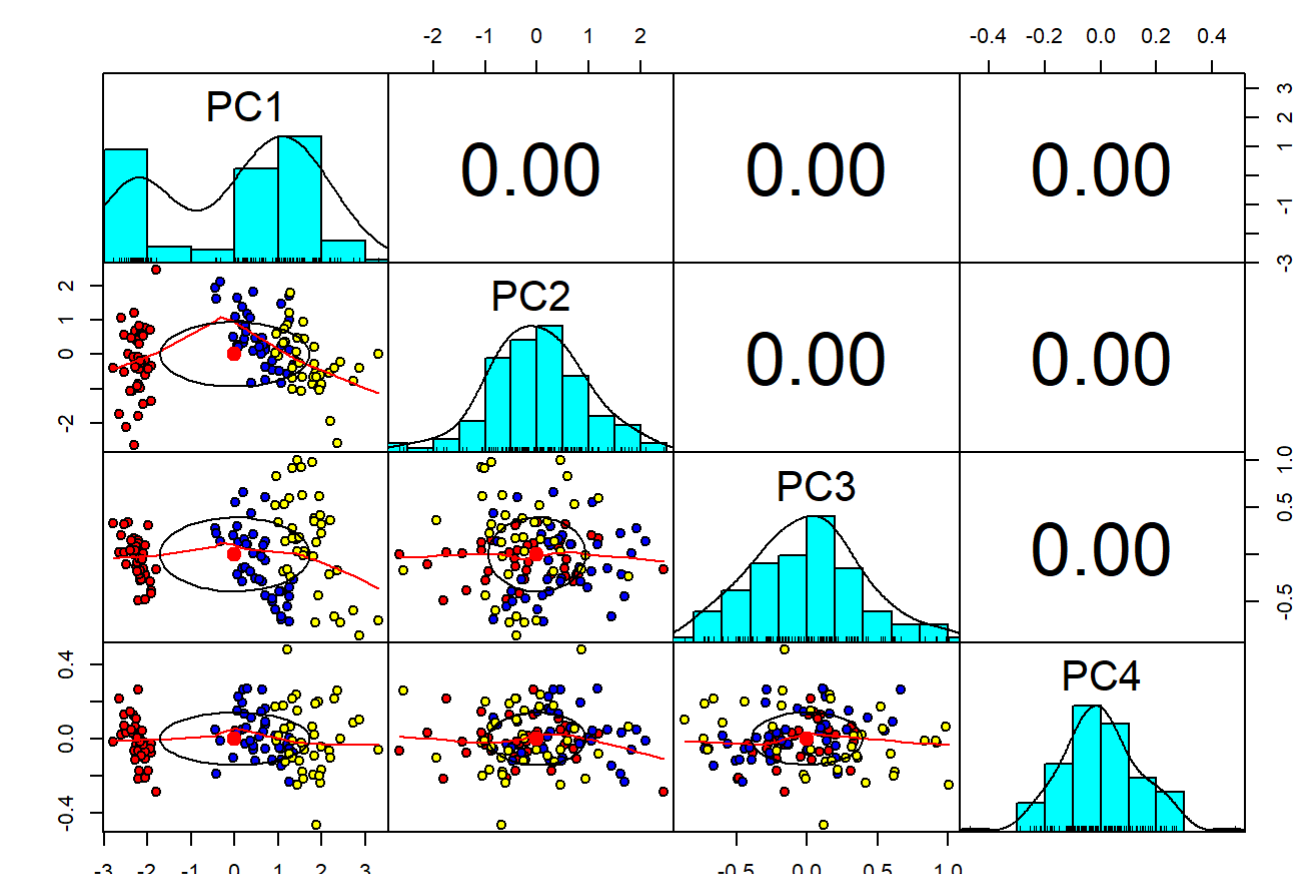
```
## Standard deviations (1, .., p=4):
## [1] 1.7161062 0.9378761 0.3941760 0.1413965
##
## Rotation (n x k) = (4 x 4):
## PC1 PC2 PC3 PC4
## Sepal.Length 0.5201525 -0.37173825 -0.7181697 -0.2747443
## Sepal.Width -0.2864515 -0.92354821 0.2239707 0.1218250
## Petal.Length 0.5777030 -0.04107397 0.1319176 0.8044687
## Petal.Width 0.5600413 -0.08474854 0.6454976 -0.5123517
```

```
summary(data_pc)
```

```
## Importance of components:
## PC1 PC2 PC3 PC4
## Standard deviation 1.7161 0.9379 0.39418 0.1414
## Proportion of Variance 0.7363 0.2199 0.03884 0.0050
## Cumulative Proportion 0.7363 0.9562 0.99500 1.0000
```

```
#dari summary diatas, dapat dilihat PC1 menjelaskan sekitar 74% dan PC2 sekitar 96% dari varian yang dijelaskan
```

```
#Membuat scatter plot
data2 <- pairs.panels(data_pc$x,gap=0,bg=c("red","blue","yellow")[data_train$Species],
pch = 21)
```



```
#Dari scatter plot diatas, kita dapat melihat korelasi 0 diantara principal component
```