Principal Components Analysis in R for thired exercise

Mehrab Atighi - Zahra Ghasabi - Zahra Barzegar - Morteza Hoseingholipur -Mina Daliry

10/19/2021

Contents

- Whats the issue?

- How to use PCA in R

- How to visualizating PCA Data in R

We want to display two way to dimantion reduction our data in R. When we want to do Dimantion reduction we should make a data as type of Data Frame, after that we should use princomp function that is in base of R.

For Example we have ramus data(exercise3). we import Dataset from Excel csv file with read.csv("addres",...) into R. and using head(data name ,number to show) to see the 4 observation of our data. and using dim to see the dimantion of our data.

```
data<-read.csv("F:/lessons/Multi countios Variate2/data/ramos-data.csv")
data=data[,-1]
#View(data)
head(data , 4)
```

```
##   year.8 year.8.5 year.9 year.9.5
## 1   47.8     48.8   49.0     49.7
## 2   46.4     47.3   47.7     48.4
## 3   46.3     46.8   47.8     48.5
## 4   45.1     45.3   46.1     47.2
```

```
dim(data)
```

```
## [1] 20  4
```

```
eigen(cor(data))
```

```
## eigen() decomposition
## $values
## [1] 3.69524710 0.25534580 0.03256133 0.01684576
##
## $vectors
##            [,1]       [,2]       [,3]       [,4]
## [1,] -0.4937453  0.5843508  0.5627874 -0.3130814
## [2,] -0.5066122  0.3816454 -0.5250456  0.5674662
## [3,] -0.5088771 -0.3386797 -0.4550180 -0.6475329
## [4,] -0.4905138 -0.6310127  0.4478348  0.4008283
```

Principal Components Analysis in R

Now we are going to use Principal Components Analysis with auto way in R:

we should use princomp function.

we should write the data as type as data.frame into this function.

we have a cor argument in this function that we can equal it with TRUE when we want to do Principal Components method with Correlation matrix, or FALSE when we want to do Principal Components method with Covariance matrix.

the other argument in this function is scores that we can equal it with TRUE and it will show the score of each component.

```
pc.r<-princomp( data,cor = TRUE ,scores = TRUE )
pc.c<-princomp(data , cor = FALSE ,scores = TRUE)
```

Principal Components Analysis in R

```
summary(pc.r)

## Importance of components:
##                          Comp.1    Comp.2      Comp.3    Comp.4
## Standard deviation    1.9223026 0.50531752 0.180447589 0.129
## Proportion of Variance 0.9238118 0.06383645 0.008140333 0.004
## Cumulative Proportion 0.9238118 0.98764823 0.995788559 1.000

head(pc.r$loadings,2)

##            Comp.1    Comp.2     Comp.3     Comp.4
## year.8   0.4937453 0.5843508  0.5627874  0.3130814
## year.8.5 0.5066122 0.3816454 -0.5250456 -0.5674662

head(pc.r$scores, 2)

##          Comp.1    Comp.2      Comp.3      Comp.4
## [1,]  -0.9747545 0.2908385 -0.03664369 -0.05337138
## [2,]  -2.0616581 0.2063934 -0.02811289 -0.02050069
```

Principal Components Analysis in R

```
summary(pc.c)

## Importance of components:
##                           Comp.1     Comp.2      Comp.3      Comp.3
## Standard deviation      4.8787246 1.28693742 0.455691682 0.328
## Proportion of Variance 0.9234897 0.06425905 0.008056783 0.004
## Cumulative Proportion  0.9234897 0.98774873 0.995805512 1.000
```

```
head(pc.c$loading,2)

##             Comp.1    Comp.2     Comp.3     Comp.4
## year.8   0.4744413 0.5917697  0.5478640  0.3529293
## year.8.5 0.4922786 0.4055451 -0.4725736 -0.6081687
```

```
head(pc.c$scores, 2)

##          Comp.1    Comp.2      Comp.3      Comp.4
## [1,] -2.522507 0.7302355 -0.07435367 -0.14057794
## [2,] -5.266811 0.5032162 -0.06231710 -0.05625445
```

to see the Percentage of Explained Variances of each variable we should install a new packages that we call factoextra and library it. the output is a scree plot that we can see codes in next silde.
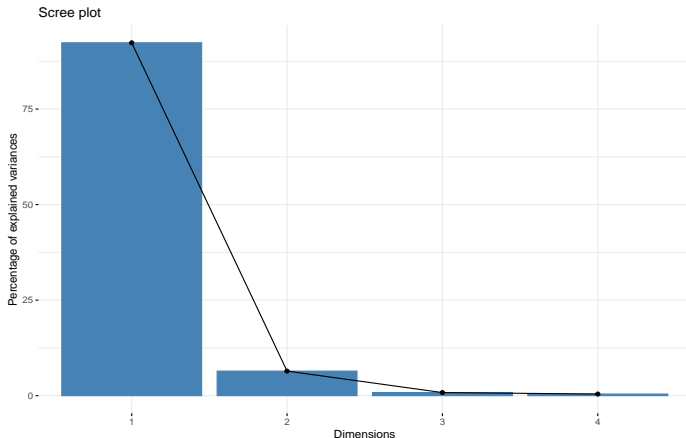
How to visualizing PCA Data in R

```r
#install.packages(factoextra)
library(factoextra)
```

```
## Loading required package: ggplot2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```
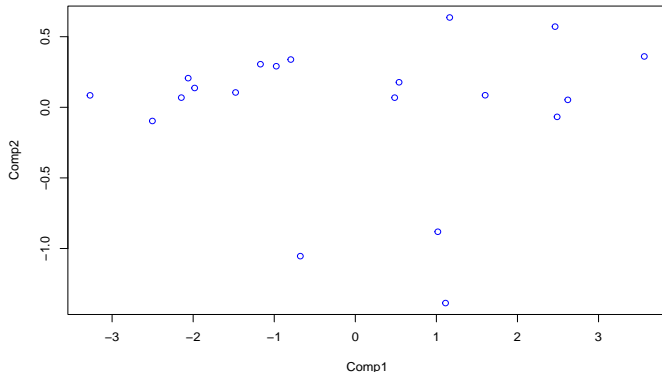
```r
fviz_eig(pc.c)#plot(pc.c)
```



```r
#fviz_eig(pc.r)#plot(pc.r)
```

How to visualizing PCA Data in R

So Now we select 2 components of our data here and ploting them.
we can use 2 function here, the first function is plotting components 1 , 2
with plot() function and the second function is biplot.

```r
plot(pc.r$scores[,1],pc.r$scores[,2]
     ,xlab = "Comp1" , ylab="Comp2" ,col="Blue")
```

How to visualizating PCA Data in R
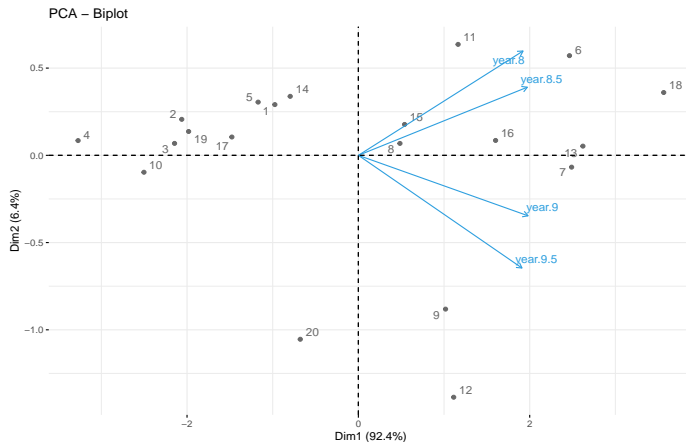
Now we want to see the Graph of individuals. Individuals with a similar profile are grouped together.

```
#biplot(pc.r)
fviz_pca_ind(pc.r,
             col.ind = "cos2", # Color by the quality of representation
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)     #Avoid text overlapping)
```
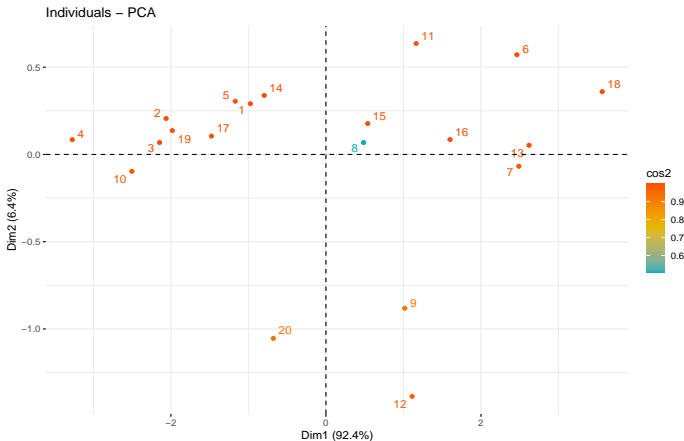
How to visualizating PCA Data in R

```
fviz_pca_biplot(pc.r, repel = TRUE,
                col.var = "#2E9FDF", # Variables color
                col.ind = "#696969"  # Individuals color
                )
```


PCA – Biplot

End.