multivariate2 project

Mehrab Atighi

10/29/2021

# Contents

* Pca Method On Data
* Factor Analysis on Data

# PCA Method On Data

Now we are importing our data from Excel and csv format to R.

rm(list=ls())  
Data<-read.csv("F:/lessons/Multi countios Variate2/project/edited-data.csv")  
#View(Data)  
new.data.y<-data.frame(y1=Data$How.much.time.did.you.spend.exercising.last.week...Total.duration.of.activity.per.week.,  
 y2=Data$How.much.time.did.you.spend.last.week.on.social.and.recreational.activities..travel..excursions..etc.....Total.duration.of.,  
 y3=Data$How.much.time.did.you.spend.on.art.activities.last.week...Total.duration.of.activity.per.week.,  
 y4=Data$How.much.time.did.you.spend.on.audio.visual.activities..TV..radio..cinema..etc...last.week...Total.duration.of.activity.per.we,  
 y5=Data$How.much.time.did.you.spend.on.virtual.social.activities.last.week...Total.duration.of.activity.per.week.,  
 y6=Data$How.much.time.did.you.spend.studying.outside.of.school.last.week...Total.duration.of.activity.per.week.)  
#View(new.data.y)  
head(new.data.y,5)

## y1 y2 y3 y4 y5 y6  
## 1 2 10 0 7 10 2  
## 2 5 0 6 6 1 4  
## 3 3 5 2 10 20 0  
## 4 2 7 0 6 20 0  
## 5 13 0 5 15 20 10

# PCA Method On Data

Now We want to see the dimantiom of our data and get Correlation and Variance Covarince matrix of our variables.

dim(new.data.y)

## [1] 95 6

cor(new.data.y)

## y1 y2 y3 y4 y5 y6  
## y1 1.0000000 0.229320331 0.22077799 0.20190028 0.139483494 0.468624239  
## y2 0.2293203 1.000000000 0.03552107 -0.13350996 -0.004623956 -0.003322887  
## y3 0.2207780 0.035521065 1.00000000 -0.09152112 0.138033595 0.372941913  
## y4 0.2019003 -0.133509959 -0.09152112 1.00000000 0.373966882 0.184092013  
## y5 0.1394835 -0.004623956 0.13803359 0.37396688 1.000000000 0.122167726  
## y6 0.4686242 -0.003322887 0.37294191 0.18409201 0.122167726 1.000000000

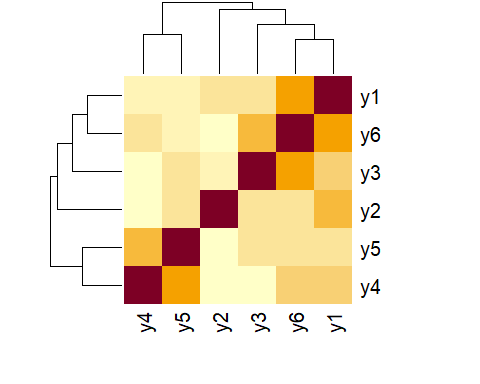
cov(new.data.y)

## y1 y2 y3 y4 y5 y6  
## y1 29.113868 6.31297695 5.8580041 7.913185 5.3514978 12.32806831  
## y2 6.312977 26.03055201 0.8911927 -4.947887 -0.1677482 -0.08265662  
## y3 5.858004 0.89119267 24.1817582 -3.269111 4.8264889 8.94139869  
## y4 7.913185 -4.94788712 -3.2691111 52.762925 19.3152419 6.51958131  
## y5 5.351498 -0.16774823 4.8264889 19.315242 50.5597773 4.23525336  
## y6 12.328068 -0.08265662 8.9413987 6.519581 4.2352534 23.77062803

# PCA Method On Data

Now we want to see the Correlation between variables in heatmap

heatmap(cor(new.data.y))



# PCA Method On Data

Now its time to see the eigen values of correlation matrix. and use principal components method on our dataset with two matrix(Correlation and Variance Covarianve matrix.)

eigen(cor(new.data.y))

## eigen() decomposition  
## $values  
## [1] 1.8996804 1.3135788 1.0222821 0.8376152 0.4786501 0.4481934  
##   
## $vectors  
## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] -0.5407557 0.1940342 0.2546516 -0.3614315 -0.46856970 0.5048726  
## [2,] -0.1005109 0.4553460 0.7620038 0.2320411 0.33420840 -0.1907078  
## [3,] -0.3891328 0.3495745 -0.4731513 0.4910754 0.35464057 0.3682067  
## [4,] -0.3288327 -0.6310244 0.1853864 -0.2225726 0.60762439 0.2014030  
## [5,] -0.3555008 -0.4601811 0.1605149 0.6511339 -0.41360272 -0.2025957  
## [6,] -0.5581620 0.1511621 -0.2655136 -0.3175772 0.05198702 -0.7011062

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

# PCA Method On Data

To see the Standard deviation and proprtion of Varince for each new components we use the summary function.

summary(pc.r)

## Importance of components:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5  
## Standard deviation 1.3782889 1.1461146 1.0110797 0.9152132 0.69184541  
## Proportion of Variance 0.3166134 0.2189298 0.1703804 0.1396025 0.07977501  
## Cumulative Proportion 0.3166134 0.5355432 0.7059235 0.8455261 0.92530110  
## Comp.6  
## Standard deviation 0.6694725  
## Proportion of Variance 0.0746989  
## Cumulative Proportion 1.0000000

According to above outputs we should select 4 components until good cumulative proportion of variance. we can see that the first Components just have 31% of variance and the second 21% and 4 components have 84% cumulative proprtion of variance.

# Pca Method On Data

Now we want to see the Cofficient of each y in each components:

pc.r$loadings

##   
## Loadings:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## y1 0.541 0.194 0.255 0.361 0.469 0.505  
## y2 0.101 0.455 0.762 -0.232 -0.334 -0.191  
## y3 0.389 0.350 -0.473 -0.491 -0.355 0.368  
## y4 0.329 -0.631 0.185 0.223 -0.608 0.201  
## y5 0.356 -0.460 0.161 -0.651 0.414 -0.203  
## y6 0.558 0.151 -0.266 0.318 -0.701  
##   
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## SS loadings 1.000 1.000 1.000 1.000 1.000 1.000  
## Proportion Var 0.167 0.167 0.167 0.167 0.167 0.167  
## Cumulative Var 0.167 0.333 0.500 0.667 0.833 1.000

# Pca Method On Data

Now we want to see the values of each observation in new dimantions, so we have values for each Componetes(all of them). Attention that here we will see just 10 observation.

head(pc.r$scores, 10)

## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6  
## [1,] -0.8312907 0.78800045 1.12424599 -0.2291562 -0.3911689 -0.58407635  
## [2,] -0.5174568 1.16325204 -1.15464474 0.7564151 -0.3693710 0.46632498  
## [3,] -0.2614935 -0.45647905 0.64091222 -1.0931362 0.2343108 -0.06526045  
## [4,] -0.6637798 -0.10678415 0.98458600 -1.1743495 0.4966841 -0.49652451  
## [5,] 2.2640996 -0.45415577 -0.34473645 0.3165415 0.6914200 -0.01717232  
## [6,] -0.4421232 -0.23444266 0.79595248 -1.2860104 1.1466105 -0.37491776  
## [7,] -0.3442112 0.05856331 0.06441728 -0.7962395 -0.2681955 0.36999988  
## [8,] -1.9204466 0.78481683 -0.70643484 0.5211554 0.2423603 -0.17157760  
## [9,] 0.4892685 -1.97192890 0.59807666 -0.3625204 0.3922635 0.54114852  
## [10,] 2.0631809 -1.16656444 0.80883491 1.3869356 0.8672390 1.17471957

# Pca Method On Data

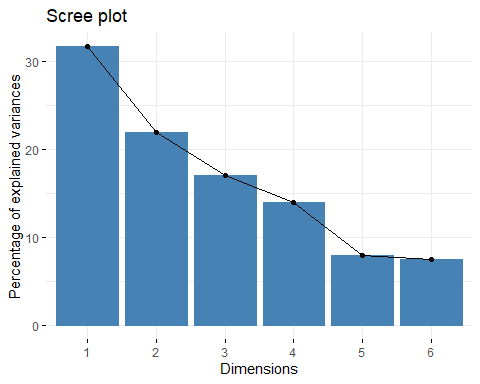
Now we want to visualazing Pca components.

library(factoextra)

## Loading required package: ggplot2

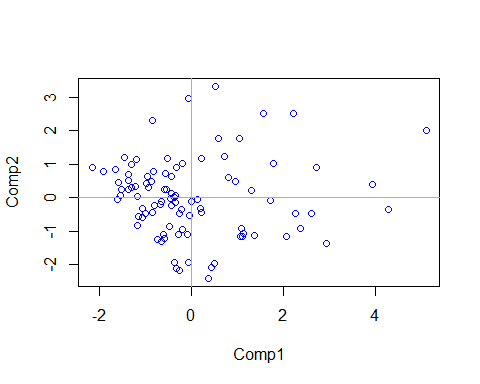
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

fviz\_eig(pc.r)



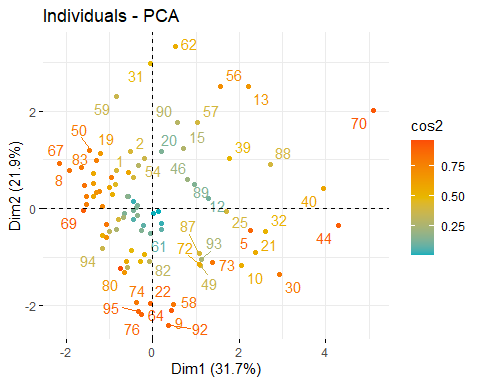
# Pca Method On Data

plot(pc.r$scores[,1],pc.r$scores[,2]  
 ,xlab = "Comp1" , ylab="Comp2" ,col="Blue")  
abline(h=0 , col="orange")  
abline(v=0 , col="orange")



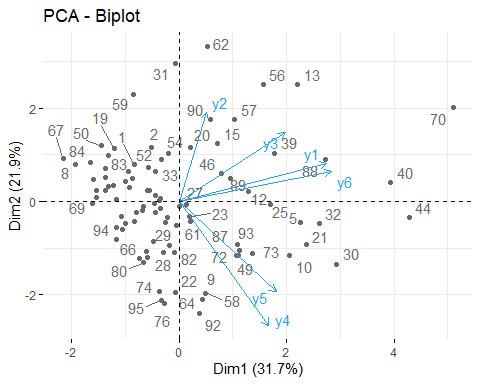
# Pca Method On Data

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



# Pca Method On Data

fviz\_pca\_biplot(pc.r, repel = TRUE,  
 col.var = "#2E9FDF", # Variables color  
 col.ind = "#696969" # Individuals color  
 )



# Factor Analysis on Data

Now we want to do Factor Analysis on our Data. at the first we start with factor analysis method with correlation matrix that calculated with pearson method.

#with Correlation Matrix  
fa1<-factanal(new.data.y , 3 ,scores = "regression" ,  
 rotation = "none", cor="pearson")  
fa2<-factanal(new.data.y , 3 ,scores = "Bartlett",  
 cor="pearson")  
fa3<-factanal(new.data.y , 3 ,scores = "regression" ,  
 rotation = "varimax", cor="pearson")

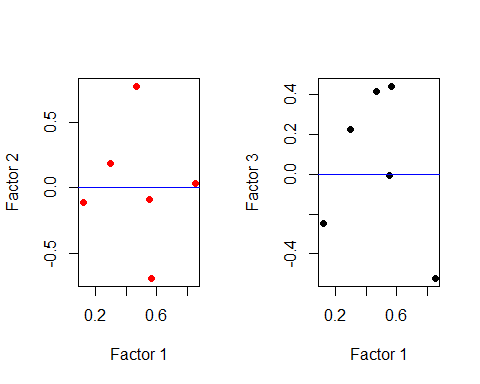
# Factor Analysis on Data

in last slide we made 3 factor analysis with diffrent rotation type and type of scores . in next sildes we want to ploting these factor analysis points with new axes.

# Factor Analysis on Data

ploting for fa1:

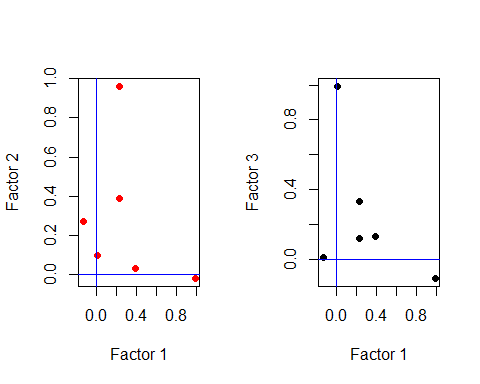
#windows(10,10)  
par(mfrow=c(1,2))  
plot(loadings(fa1)[,1],loadings(fa1)[,2],pch=16,xlab="Factor 1",  
 ylab="Factor 2",col="red")  
abline(h=0 , col="blue")  
abline(v=0 , col="blue")  
plot(loadings(fa1)[,1],loadings(fa1)[,3],pch=16,xlab="Factor 1",  
 ylab="Factor 3",col="black")  
abline(h=0 , col="blue")  
abline(v=0 , col="blue")



# Factor Analysis on Data

ploting for fa2:

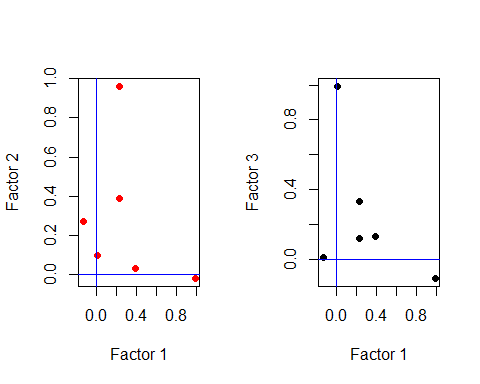
#windows(10,10)  
par(mfrow=c(1,2))  
plot(loadings(fa2)[,1],loadings(fa2)[,2],pch=16,xlab="Factor 1",  
 ylab="Factor 2",col="red")  
abline(h=0 , col="blue")  
abline(v=0 , col="blue")  
plot(loadings(fa2)[,1],loadings(fa2)[,3],pch=16,xlab="Factor 1",  
 ylab="Factor 3",col="black")  
abline(h=0 , col="blue")  
abline(v=0 , col="blue")



# Factor Analysis on Data

ploting for fa3:

#windows(10,10)  
par(mfrow=c(1,2))  
plot(loadings(fa3)[,1],loadings(fa3)[,2],pch=16,xlab="Factor 1",  
 ylab="Factor 2",col="red")  
abline(h=0 , col="blue")  
abline(v=0 , col="blue")  
plot(loadings(fa3)[,1],loadings(fa3)[,3],pch=16,xlab="Factor 1",  
 ylab="Factor 3",col="black")  
abline(h=0 , col="blue")  
abline(v=0 , col="blue")



# Factor Analysis on Data

We can you and have Factor Analysis method without Correlation matrix and using covariance matrix in this algorithm.

#install.packages("psych")  
library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(ggplot2)  
fa4 <- fa(new.data.y,nfactors = 6, rotate = "varimax" ,  
 scores = "regression" ,covar = TRUE )

# Factor Analysis on Data

using factor analysis with covariance matrix and varimax method:

fa4

## Factor Analysis using method = minres  
## Call: fa(r = new.data.y, nfactors = 6, rotate = "varimax", scores = "regression",   
## covar = TRUE)  
## Unstandardized loadings (pattern matrix) based upon covariance matrix  
## MR3 MR1 MR2 MR5 MR4 MR6 h2 u2 H2 U2  
## y1 0.38 0.64 3.82 0.63 1.28 0 17 12 0.59 0.41  
## y2 0.01 -0.44 0.52 -0.02 3.62 0 14 12 0.52 0.48  
## y3 0.46 -0.51 1.02 3.23 0.06 0 12 12 0.49 0.51  
## y4 1.80 5.81 1.33 -0.76 -0.87 0 40 13 0.76 0.24  
## y5 5.91 1.50 0.42 0.75 0.06 0 38 13 0.75 0.25  
## y6 0.11 0.62 2.91 1.93 -0.35 0 13 11 0.54 0.46  
##   
## MR3 MR1 MR2 MR5 MR4 MR6  
## SS loadings 38.54 37.30 26.34 15.71 15.62 0.00  
## Proportion Var 0.19 0.18 0.13 0.08 0.08 0.00  
## Cumulative Var 0.19 0.37 0.50 0.57 0.65 0.65  
## Proportion Explained 0.29 0.28 0.20 0.12 0.12 0.00  
## Cumulative Proportion 0.29 0.57 0.77 0.88 1.00 1.00  
##   
## Standardized loadings (pattern matrix)  
## item MR3 MR1 MR2 MR5 MR4 MR6 h2 u2  
## y1 1 0.07 0.12 0.71 0.12 0.24 0 0.59 0.41  
## y2 2 0.00 -0.09 0.10 0.00 0.71 0 0.52 0.48  
## y3 3 0.09 -0.10 0.21 0.66 0.01 0 0.49 0.51  
## y4 4 0.25 0.80 0.18 -0.10 -0.12 0 0.76 0.24  
## y5 5 0.83 0.21 0.06 0.11 0.01 0 0.75 0.25  
## y6 6 0.02 0.13 0.60 0.40 -0.07 0 0.54 0.46  
##   
## MR3 MR1 MR2 MR5 MR4 MR6  
## SS loadings 0.77 0.73 0.95 0.62 0.58 0.00  
## Proportion Var 0.13 0.12 0.16 0.10 0.10 0.00  
## Cumulative Var 0.13 0.25 0.41 0.51 0.61 0.61  
## Cum. factor Var 0.21 0.41 0.67 0.84 1.00 1.00  
##   
## Mean item complexity = 1.4  
## Test of the hypothesis that 6 factors are sufficient.  
##   
## The degrees of freedom for the null model are 15 and the objective function was 180.33 with Chi Square of 16439.72  
## The degrees of freedom for the model are -6 and the objective function was 0   
##   
## The root mean square of the residuals (RMSR) is 0   
## The df corrected root mean square of the residuals is NA   
##   
## The harmonic number of observations is 95 with the empirical chi square 0 with prob < NA   
## The total number of observations was 95 with Likelihood Chi Square = 0 with prob < NA   
##   
## Tucker Lewis Index of factoring reliability = 1.001  
## Fit based upon off diagonal values = 1  
## Measures of factor score adequacy   
## MR3 MR1 MR2 MR5 MR4 MR6  
## Correlation of (regression) scores with factors 0.84 0.81 0.77 0.69 0.73 0  
## Multiple R square of scores with factors 0.70 0.65 0.60 0.48 0.53 0  
## Minimum correlation of possible factor scores 0.40 0.30 0.20 -0.05 0.05 -1

# Factor Analysis on Data

n\_factors <- length(fa4$e.values)  
scree <- data.frame(  
 Factor\_n = as.factor(1:n\_factors),   
 Eigenvalue = fa4$e.values)  
ggplot(scree, aes(x = Factor\_n, y = Eigenvalue, group = 1)) +   
 geom\_point() + geom\_line() +  
 xlab("Number of factors") +  
 ylab("Initial eigenvalue") +  
 labs( title = "Scree Plot",   
 subtitle = "(Based on the unreduced correlation matrix)")

