mid-test

Mehrab Atighi

12/8/2021

# mid-test-Mehrab Atighi - question 2  
library(MASS)  
noshabe=read.table("C:/Users/Frostless/Desktop/noshabe.txt",header=T)  
head(noshabe)

میخواهیم یک شماتیک کلی از داده ها داشته باشیم:

## y1 y2 x1 x2 x3 x4 x5 x6 x7 x8  
## 1 1.0 0.8 4.05 1.68 0.85 3.0 3.97 5.00 16.90 122  
## 2 0.1 0.2 3.81 1.39 0.30 0.6 3.62 4.52 15.80 62  
## 3 0.5 0.0 4.20 1.63 0.92 -2.3 3.48 4.46 15.80 139  
## 4 0.7 0.7 4.35 1.43 0.97 -1.6 3.45 3.98 15.40 150  
## 5 -0.1 -1.1 4.35 1.53 0.87 -2.0 3.67 4.22 15.40 138  
## 6 0.4 0.5 4.05 1.84 0.95 -2.5 3.61 5.00 16.78 123

r=cor(noshabe)  
eigen(r)

برای مشاهده مقادیر و بردار های ویژه ماتریس همبستگی داده هایمان داریم:هر ستون یک بردار ویژه هست و 4 مقدار ویژه بزرگتر را هم قرمز کرده ایم.

## eigen() decomposition  
## $values  
## [1] 3.17448729 2.56493022 1.43212322 1.27657967 0.54223281 0.47275791  
## [7] 0.25092430 0.11845314 0.10353588 0.06397556  
##   
## $vectors  
## [,1] [,2] [,3] [,4] [,5] [,6]  
## [1,] -0.11997230 0.18567314 0.691067226 0.10148541 -0.14966451 -0.54463822  
## [2,] -0.05501841 0.31830313 0.543460141 0.26385314 0.27978122 0.59052627  
## [3,] -0.46364614 -0.05771886 0.069761205 -0.37819443 -0.05615696 -0.22652342  
## [4,] -0.28650440 0.16525793 -0.176461716 0.48673258 -0.75943071 0.10630907  
## [5,] -0.52341375 0.14388701 -0.037523917 -0.01456060 0.11699667 0.07188040  
## [6,] 0.08872045 -0.42401572 0.072786799 0.55418521 0.22264763 -0.32990142  
## [7,] 0.31387769 0.44934153 -0.006184096 -0.14152559 -0.04855183 -0.32241953  
## [8,] 0.22619200 0.53500587 -0.139460474 -0.10368650 -0.04029456 -0.03811413  
## [9,] -0.09255340 0.36412931 -0.383212843 0.44113184 0.41207503 -0.25308608  
## [10,] -0.49542474 0.11110761 -0.134262013 -0.08620706 0.28703214 -0.09258569  
## [,7] [,8] [,9] [,10]  
## [1,] 0.07478701 -0.21801220 0.29951438 -0.03738168  
## [2,] 0.03288863 0.23307443 -0.19919997 -0.09275719  
## [3,] 0.45699056 0.43343742 -0.42583103 -0.03600333  
## [4,] -0.06926643 0.13970428 -0.05290404 -0.08606464  
## [5,] -0.19508330 -0.46846902 -0.23428436 0.61028652  
## [6,] -0.15712636 0.38663840 -0.17942157 0.36761831  
## [7,] -0.48476000 0.04359167 -0.56480044 -0.13153551  
## [8,] 0.19159642 0.39767489 0.33861452 0.56552201  
## [9,] 0.40873750 -0.19700917 -0.06176513 -0.27173894  
## [10,] -0.52742114 0.35064381 0.40239792 -0.25048692

pca = princomp(noshabe , scores = TRUE , cor = TRUE)  
summary(pca)

از آنجایی که جمع تجمعی واریانس های هریک از مولفه های ما با رنگ قرمز نشان داده شده و میزان واریانس هریک از آنها نسبت به کل واریانس با آبی میتوانیم ببینیم که برای 4 مولفه اول ما حدود 84 درصد از واریانس کل را پوشش میدهیم و میتوانیم ما بقی را حذف بکنیم (یاداوری که 4 مقدار ویژه مقادیر بزرگی داشتند). و میزان انجراف استاندارد آنها هم با مشکی نشون داده شده است.

## Importance of components:  
## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5  
## Standard deviation 1.7817091 1.6015400 1.1967135 1.129858 0.73636459  
## Proportion of Variance 0.3174487 0.2564930 0.1432123 0.127658 0.05422328  
## Cumulative Proportion 0.3174487 0.5739418 0.7171541 0.844812 0.89903532  
## Comp.6 Comp.7 Comp.8 Comp.9 Comp.10  
## Standard deviation 0.68757393 0.50092345 0.34417022 0.32176992 0.252933895  
## Proportion of Variance 0.04727579 0.02509243 0.01184531 0.01035359 0.006397556  
## Cumulative Proportion 0.94631111 0.97140354 0.98324886 0.99360244 1.000000000

Pca$Loadings

Loadings:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10

y1 0.120 0.186 0.691 0.101 0.150 0.545 0.218 0.300

y2 0.318 0.543 0.264 -0.280 -0.591 -0.233 -0.199

x1 0.464 -0.378 0.227 0.457 -0.433 -0.426

x2 0.287 0.165 -0.176 0.487 0.759 -0.106 -0.140

x3 0.523 0.144 -0.117 -0.195 0.468 -0.234 -0.610

x4 -0.424 0.554 -0.223 0.330 -0.157 -0.387 -0.179 -0.368

x5 -0.314 0.449 -0.142 0.322 -0.485 -0.565 0.132

x6 -0.226 0.535 -0.139 -0.104 0.192 -0.398 0.339 -0.566

x7 0.364 -0.383 0.441 -0.412 0.253 0.409 0.197 0.272

x8 0.495 0.111 -0.134 -0.287 -0.527 -0.351 0.402 0.250

در بالا میتوانیم مقدایر ضرایب هر یک از متغیرها را در تابع تحلیل مولفه اصلی مشاهده بکنیم برای مثال آنهایی که با سبز نشان داده شده اند همگی ضرایب درون تابع مولفه اصلی اول هستند. و آبی و مشکی و قرمز هم برای مولفه دوم و سوم و چهارم.

در پایین نیز به رسم نمودارهای سنگ ریزع و نمودار 2 مولفه اصلی اول میپردازیم:

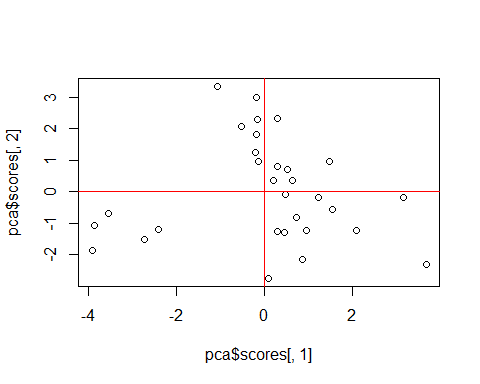
بر اساس نمودار هم ما 4 مولفه اول را انتخاب میکنیم نمودار سنگ ریزه در آخر هستش.زیر شیب یهویی از آن به بعد کم خواهد شد.

plot(pca$scores[,1] , pca$scores[,2])  
abline(h=0 , v= 0 , col="red")  
library(factoextra)

## Loading required package: ggplot2

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

معادله هردو سوال رو بصورت عکس ارسال خواهم کرد با هم.



fviz\_eig(pca)

