Componentes principales

Ileana Parra

10/10/2022

#PARTE A

##1.Matriz de datos centrados en sus medias

```
x1=c(2.5, 0.5, 2.2, 1.9, 3.1, 2.3, 2, 1, 1.5, 1.1)
x2=c( 2.4, 0.7, 2.9, 2.2, 3.0, 2.7, 1.6, 1.1, 1.6, 0.9)
m=matrix(c(x1,x2),ncol=2)
m1=c(rep(mean(x1),10))
m2=c(rep(mean(x2),10))
mu=matrix(c(m1,m2),ncol=2)
M=m-mu
M
```

```
##        [,1]  [,2]
##  [1,]  0.69  0.49
##  [2,] -1.31 -1.21
##  [3,]  0.39  0.99
##  [4,]  0.09  0.29
##  [5,]  1.29  1.09
##  [6,]  0.49  0.79
##  [7,]  0.19 -0.31
##  [8,] -0.81 -0.81
##  [9,] -0.31 -0.31
## [10,] -0.71 -1.01
```

##2.Matriz de varianza-covarianza de la matriz de datos centrados

```
cov(M)
```

```
##           [,1]      [,2]
## [1,] 0.6165556 0.6154444
## [2,] 0.6154444 0.7165556
```

##3. Valores propios y vectores propios de la matriz de varianza-covarianza de la matriz de datos centrados.

```
lambda <- eigen(cov(M))
lambda$values
```

```
## [1] 1.2840277 0.0490834
```

```
lambda$vectors
```

```
##            [,1]       [,2]
## [1,] 0.6778734 -0.7351787
## [2,] 0.7351787  0.6778734
```

##4.Obtenga las matrices transpuestas de los vectores propios

```
t(lambda$vectors)
```

```
##             [,1]      [,2]
## [1,]  0.6778734 0.7351787
## [2,] -0.7351787 0.6778734
```

y la traspuesta de la matriz de datos centrados.

```
t(M)
```

```
##      [,1]  [,2] [,3] [,4] [,5] [,6]  [,7]  [,8]  [,9] [,10]
## [1,] 0.69 -1.31 0.39 0.09 1.29 0.49  0.19 -0.81 -0.31 -0.71
## [2,] 0.49 -1.21 0.99 0.29 1.09 0.79 -0.31 -0.81 -0.31 -1.01
```

##5.Multiplique la matriz transpuesta de los vectores propios con la transpuesta de la matriz de datos centrados.

```
t(t(lambda$vectors)%*%t(M))
```

```
##              [,1]        [,2]
##  [1,]  0.82797019 -0.17511531
##  [2,] -1.77758033  0.14285723
##  [3,]  0.99219749  0.38437499
##  [4,]  0.27421042  0.13041721
##  [5,]  1.67580142 -0.20949846
##  [6,]  0.91294910  0.17528244
##  [7,] -0.09910944 -0.34982470
##  [8,] -1.14457216  0.04641726
##  [9,] -0.43804614  0.01776463
## [10,] -1.22382056 -0.16267529
```

##Conclusión *El componente principal 1 tiene mayores coeficientes para la mayoría de las variables a comparación del componente principal 2.

#PARTE B

```
cpa <- prcomp(m, scale=TRUE)
names(cpa)
```

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

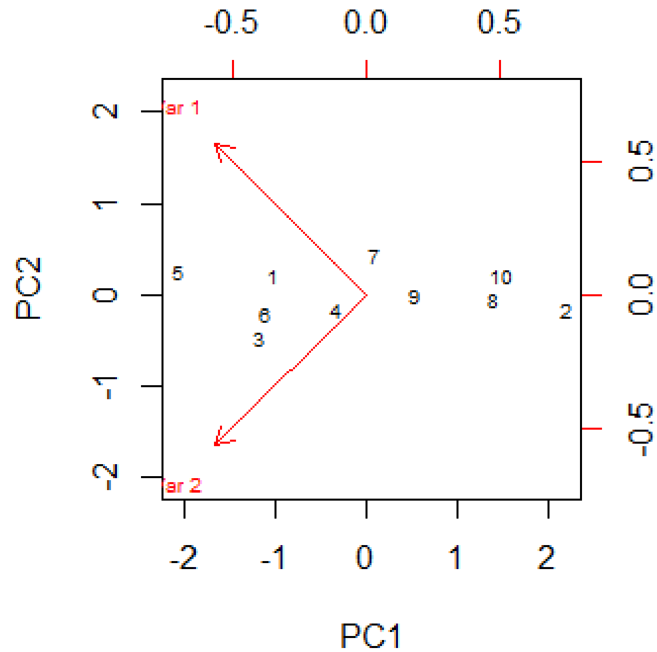
```
print("desviaciones estándar: ")
```

```
## [1] "desviaciones estándar: "
```

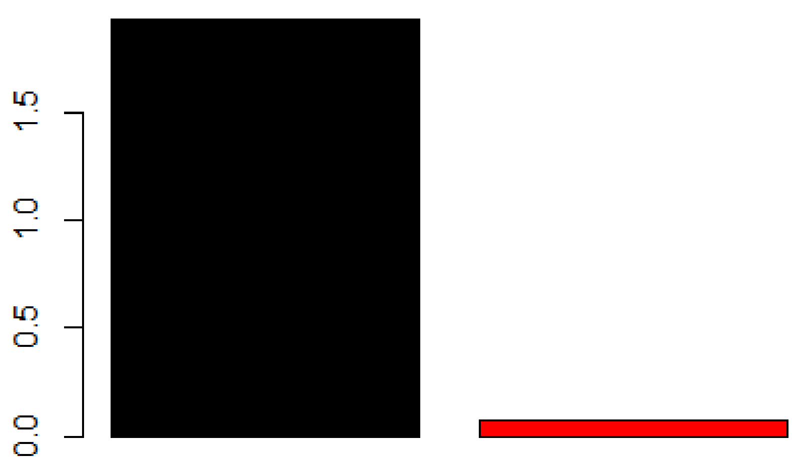
```
cpa$sdev
```

```
## [1] 1.3877785 0.2721594
```

```r
print("medias: ")

## [1] "medias: "

print("center y scale dan las medias y desv estándar previa
estandarización: ")

## [1] "center y scale dan las medias y desv estándar previa
estandarización: "

cpa$center

## [1] 1.81 1.91

cpa$scale

## [1] 0.7852105 0.8464960

print("Los coeficientes de la combinación lineal normalizada de
componete")

## [1] "Los coeficientes de la combinación lineal normalizada de
componete"

cpa$rotation

##              PC1        PC2
## [1,] -0.7071068  0.7071068
## [2,] -0.7071068 -0.7071068

print("Los datos por sustituidos en la combinación lineal de vectores
propios:")

## [1] "Los datos por sustituidos en la combinación lineal de vectores
propios:"

cpa$x

##               PC1         PC2
##  [1,] -1.03068029  0.21205314
##  [2,]  2.19045016 -0.16894230
##  [3,] -1.17818776 -0.47577321
##  [4,] -0.32329464 -0.16119898
##  [5,] -2.07219947  0.25117173
##  [6,] -1.10117414 -0.21865330
##  [7,]  0.08785251  0.43005447
##  [8,]  1.40605089 -0.05281009
##  [9,]  0.53811824 -0.02021127
## [10,]  1.48306451  0.20430982

biplot(x = cpa, scale = 0, cex =0.6, col = c("black", "red"))
```




```r
barplot(cpa$sdev^2,col=c("black","red"))
```




```r
summary(cpa)
```

```
## Importance of components:
##                          PC1     PC2
## Standard deviation     1.388 0.27216
## Proportion of Variance 0.963 0.03704
## Cumulative Proportion  0.963 1.00000
```

Conclusión

*El primer componente principal explica el 96.3% de la varianza, lo que lo hace más importante que el segundo componente.

En ambos métodos, el componente principal 1 es el más importante.