# Act3 Componentes principales C_IACD_Estadística

A01750164 | Paul Martín García Morfín

2022-10-11

## Componentes principales

### Parte A

De los siguientes datos:

- x1: 2.5, 0.5, 2.2, 1.9, 3.1, 2.3, 2, 1, 1.5, 1.1
- x2: 2.4, 0.7, 2.9, 2.2, 3.0, 2.7, 1.6, 1.1, 1.6, 0.9

#### 1. Obtenga una 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 = data.frame(x1, x2)
M = as.matrix(M)
H = diag(nrow(M))-(1/nrow(M)) * matrix(1, nrow(M), nrow(M))
MCen = H %*% M
MCen
```

```
##          x1    x2
##  [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. Obtenga la matriz de varianza-covarianza de la matriz de datos centrados

```
MCov = cov(MCen)
MCov
```

```
##           x1        x2
## x1 0.6165556 0.6154444
## x2 0.6154444 0.7165556
```

**3. Obtenga los valores propios y vectores propios de la matriz de varianza-covarianza de la matriz de datos centrados.**

```
L = eigen(MCov)
L

## eigen() decomposition
## $values
## [1] 1.2840277 0.0490834
##
## $vectors
##            [,1]       [,2]
## [1,] 0.6778734 -0.7351787
## [2,] 0.7351787  0.6778734
```

**4. Obtenga las matrices transpuestas de los vectores propios y la traspuesta de la matriz de datos centrados.**

```
t_Vprop = t(L$vectors)
t_MCen = t(MCen)
cat("Matriz transpuesta de los vectores propios:\n")

## Matriz transpuesta de los vectores propios:

t_Vprop

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

cat("\nMatriz transpuesta de la matriz de datos centrados:\n")

##
## Matriz transpuesta de la matriz de datos centrados:

t_MCen

##    [,1]  [,2] [,3] [,4] [,5] [,6]  [,7]  [,8]  [,9] [,10]
## x1 0.69 -1.31 0.39 0.09 1.29 0.49  0.19 -0.81 -0.31 -0.71
## x2 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.**

```
CP = t_Vprop%*%t_MCen
rownames(CP) = c("CP1", "CP2")
t(CP)

##                CP1         CP2
##  [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
```

### 6. Interprete los resultados.

Al final obtenemos los datos por sustituidos en la combinación lineal de vectores propios, o sea, el valor de los datos rotados (los datos centrados, en este caso sin escalar, multiplicados por la matriz cuyas columnas contienen los vectores propios)

## Parte B

### Aplique a los mismos datos las fórmulas de R para Componentes principales e interpreta resultados.

```
cpa = prcomp(M, scale=TRUE)
names(cpa)

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

Se realiza un análisis de componentes principales en la matriz de datos M (contiene los vectores dados al inicio) y se indica que las variables deben estar escaladas antes de realizar el análisis.

```
cat("Desviaciones estándar:\n")

## Desviaciones estándar:

cpa$sdev

## [1] 1.3877785 0.2721594

cat("\nMedias.\n")

##
## Medias.

cat("Center y scale dan las medias y desv estándar previa
estandarización:\n")

## Center y scale dan las medias y desv estándar previa estandarización:

cpa$center

##   x1   x2
## 1.81 1.91

cat("\n")

cpa$scale

##        x1        x2
## 0.7852105 0.8464960
```

```r
cat("\nLos coeficientes de la combinación lineal normalizada de componete:\n")
```

```
##
## Los coeficientes de la combinación lineal normalizada de componete:
```

```r
cpa$rotation
```

```
##           PC1        PC2
## x1 -0.7071068  0.7071068
## x2 -0.7071068 -0.7071068
```

```r
cat("\nLos datos por sustituidos en la combinación lineal de vectores propios:\n")
```

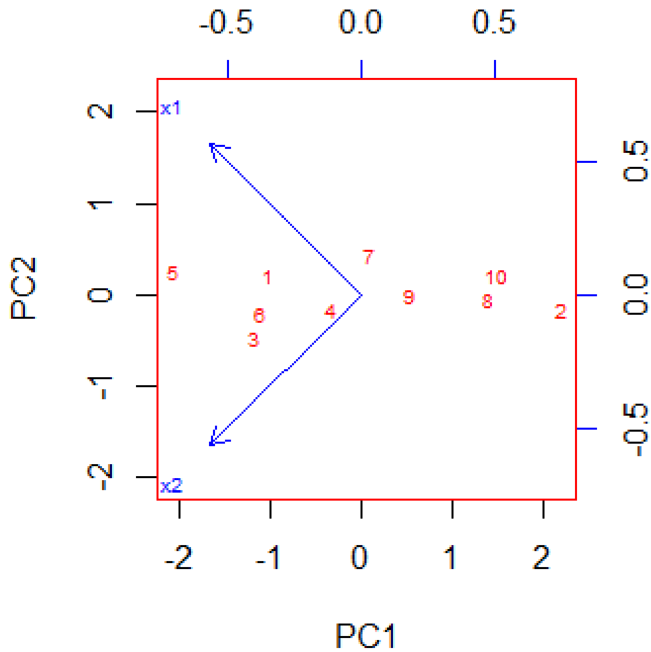
```
##
## Los datos por sustituidos en la combinación lineal de vectores propios:
```

```r
cpa$x
```

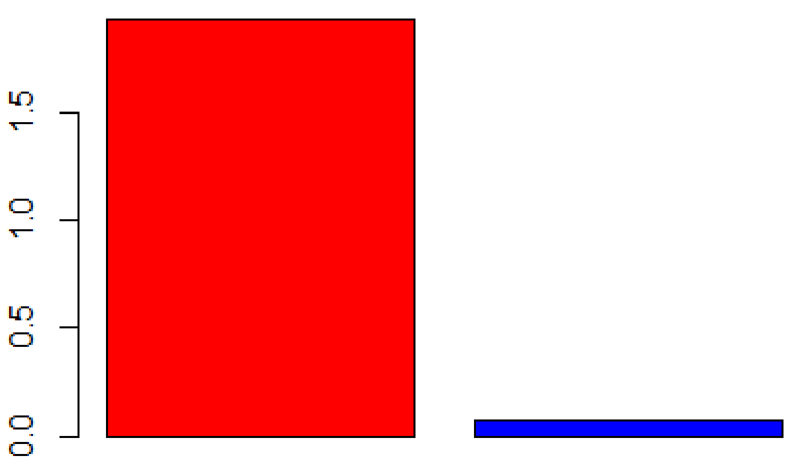
```
##               PC1         PC2
##  [1,] -1.03068029  0.21205314
##  [2,]  2.19045016 -0.16894230
##  [3,] -1.17818776 -0.47577321
##  [4,] -0.32329464 -0.16119898
##  [5,] -2.07219947  0.25117173
##  [6,] -1.10117414 -0.21865330
##  [7,]  0.08785251  0.43005447
##  [8,]  1.40605089 -0.05281009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

Obtenemos las desviaciones estándar de los componentes principales (es decir, las raíces cuadradas de los valores propios de la matriz de covarianza), el centrado y la escala utilizados, una matriz cuyas columnas contienen los vectores propios y, finalmente, el valor de los datos rotados (los datos centrados y escalados multiplicados por la matriz de rotación)

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




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



En el biplot se representan tanto las observaciones como las variables de la matriz de datos multivariados en el mismo gráfico.

```
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
```

Hacer un análisis de componentes principales es una técnica útil para el análisis exploratorio de datos, que permite visualizar mejor la variación presente en un conjunto de datos con muchas variables.