Componentes_Principales

Arturo

2023-09-26

```
# Instalacion y carga de paquetes
if (!require(stats) || !require(factoextra) || !require(ggplot2) || !require(FactoMineR)) {
  install.packages("stats")
  install.packages("factoextra")
  install.packages('ggplot2')
  install.packages("FactoMineR")
}
```

```
## Loading required package: factoextra
```

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

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

```
## Loading required package: FactoMineR
```

```
library(stats)
library(factoextra)
library(ggplot2)
library(FactoMineR)
```

Nombre: Arturo Garza Campuzano

Matrícula: A00828096

Componentes principales

PARTE I

Funciones para imprimir resultados

```
imprimir_resultados_matrices <- function(resultados_S, resultados_Z, nombre_resultados) {
  cat(nombre_resultados,"de matriz de varianza-covarianza:\n")
  print(resultados_S)
  cat(nombre_resultados, "de matriz de correlaciones:\n")
  print(resultados_Z)
}

imprimir_resultados <- function(resultado, nombre_resultado) {
  cat(nombre_resultado,":\n")
  print(resultado)
}
```

1. Calcule las matrices de varianza-covarianza S con cov(X) y la matriz de correlaciones con cor(X).

```r
# Lectura del archivo
X = read.csv("paises_mundo.csv")
# Matriz de varianza-covarianza
S = cov(X)
# Matriz de correlaciones
R = cor(X)
```

2. Calcule los valores y vectores propios de cada matriz.

```r
# Valores y vectores propios de cada matriz
vvp_S = eigen(S)
vvp_R = eigen(R)

imprimir_resultados_matrices(vvp_S, vvp_R, "Valores y vectores propios")
```

```
## Valores y vectores propios de matriz de varianza-covarianza:
## eigen() decomposition
## $values
##  [1] 6.163576e+10 6.581612e+09 4.636256e+06 3.107232e+05 1.216015e+04
##  [6] 5.137767e+02 3.627885e+02 4.542082e+01 5.800868e+00 1.438020e+00
## [11] 4.768083e-01
##
## $vectors
##                [,1]          [,2]          [,3]          [,4]          [,5]
##  [1,] -1.658168e-06  4.706785e-07  0.0001263736 -1.928408e-05 -0.0055373971
##  [2,] -4.048139e-05 -1.774254e-05  0.0082253821 -2.493257e-03 -0.0944030204
##  [3,]  5.739096e-06 -1.084543e-05  0.0001318149  5.538307e-03  0.0314036410
##  [4,]  8.880376e-01  4.597632e-01  0.0026022071 -3.893588e-04 -0.0003327409
##  [5,]  4.597636e-01 -8.880405e-01  0.0005694896  1.096305e-03  0.0002207819
##  [6,]  3.504341e-04  4.016179e-04 -0.0619424889  7.641174e-03  0.9921404486
##  [7,]  2.625508e-04 -1.122118e-03 -0.0401453227 -9.991411e-01  0.0057795144
##  [8,]  4.089564e-06  7.790843e-06  0.0012719918  6.435797e-03  0.0419331615
##  [9,] -1.073825e-06  2.350808e-07  0.0001916177  4.043796e-05 -0.0018090751
## [10,]  2.547156e-03  7.126782e-04 -0.9972315499  3.973568e-02 -0.0625729475
## [11,]  4.643724e-06 -1.315731e-06 -0.0020679047 -5.626049e-05 -0.0042367120
##                [,6]          [,7]          [,8]          [,9]         [,10]
##  [1,]  1.243456e-02  5.359089e-03  8.390810e-02 -6.778358e-02 -1.158091e-01
##  [2,]  9.917515e-01  2.258019e-02  7.891128e-02 -1.637836e-02  4.264872e-04
##  [3,]  8.552991e-02 -1.136481e-01 -9.856498e-01 -1.468464e-02  8.241465e-03
##  [4,] -8.621005e-06 -7.566477e-06 -1.217248e-05 -3.971469e-07  4.274451e-07
##  [5,]  1.955408e-05  1.544658e-05  2.558998e-05  1.059471e-06 -1.353881e-06
##  [6,]  9.109622e-02  4.748682e-02  3.416811e-02 -5.379549e-03 -3.409423e-03
##  [7,] -1.087229e-03 -6.863294e-03 -4.698731e-03  7.965261e-05  3.621425e-05
##  [8,]  1.721948e-02 -9.920538e-01  1.169638e-01  1.416566e-03  5.891758e-03
##  [9,]  1.758667e-03 -7.455427e-03 -1.811443e-02  1.283039e-01 -9.859317e-01
## [10,]  2.639673e-03 -3.764707e-03 -1.267052e-03  2.262931e-03  2.672618e-04
## [11,] -1.877994e-02 -1.709137e-03  5.204823e-03 -9.891529e-01 -1.200519e-01
##               [,11]
##  [1,]  9.872887e-01
##  [2,] -2.092491e-02
##  [3,]  8.344324e-02
##  [4,]  2.723996e-07
##  [5,] -2.086857e-07
##  [6,]  4.944398e-04
##  [7,]  4.780416e-04
```

```
##  [8,] -3.748976e-03
##  [9,] -1.052934e-01
## [10,]  5.906241e-05
## [11,] -8.221371e-02
##
## Valores y vectores propios de matriz de correlaciones:
## eigen() decomposition
## $values
##  [1] 4.02987902 1.92999195 1.37041115 0.86451597 0.79414057 0.72919997
##  [7] 0.57130511 0.32680096 0.16806846 0.14632819 0.06935866
##
## $vectors
##               [,1]        [,2]        [,3]        [,4]        [,5]        [,6]
##  [1,] -0.314119414  0.34835747 -0.07352541 -0.44028717  0.32972147 -0.18392437
##  [2,] -0.392395442 -0.04136238 -0.17759254 -0.13398483 -0.08340489 -0.08656390
##  [3,]  0.116546319 -0.58283641  0.16686305  0.05865031 -0.18654100  0.16835650
##  [4,]  0.295393771 -0.17690839 -0.53343025 -0.26248209  0.14110658  0.04653378
##  [5,]  0.258964724 -0.17356372 -0.61438847 -0.17389644  0.07521971  0.02821905
##  [6,]  0.446082934 -0.02719077  0.15177250  0.04959796  0.05416498  0.02442175
##  [7,]  0.092410503  0.32060987 -0.37024258  0.73603097 -0.02671021 -0.30940890
##  [8,]  0.005692925 -0.45742697  0.16480339  0.04024882  0.41531702 -0.75356463
##  [9,] -0.243652293 -0.15408201 -0.02961449  0.33650345  0.73261463  0.50894232
## [10,]  0.415029554  0.23286257  0.20608749 -0.06730166  0.23100421  0.05806466
## [11,]  0.374531032  0.29168698  0.20631751 -0.14843513  0.24028756 -0.02809233
##                [,7]        [,8]        [,9]       [,10]       [,11]
##  [1,]  0.1628974320 -0.09481963  0.52181220  0.34674573  0.10062784
##  [2,]  0.6398040762 -0.32307802 -0.29031618 -0.38959240 -0.17487096
##  [3,]  0.5310867107  0.05209889  0.23599758  0.42854658  0.16786800
##  [4,] -0.1490207046 -0.44913216 -0.36995675  0.34911534  0.15247432
##  [5,]  0.1082745817  0.50343911  0.30681318 -0.33770404 -0.12366382
##  [6,] -0.0008501608 -0.56975094  0.44733110 -0.20997673 -0.44992596
##  [7,]  0.2357666690 -0.05962470  0.08358225  0.20561803  0.07067780
##  [8,] -0.0806036686  0.04275404 -0.07438520 -0.08671232  0.01493710
##  [9,]  0.0112333588 -0.01607505 -0.01868615 -0.03209758 -0.07259619
## [10,]  0.2711228006 -0.05023582 -0.04339752 -0.36147417  0.67912543
## [11,]  0.3352822144  0.30978009 -0.37666244  0.28779437 -0.46737561
```

3. Calcule la proporción de varianza explicada por cada componente.

```
pve_S = vvp_S$values / sum(diag(S))
imprimir_resultados(pve_S, "Proporción de varianza explicada por cada componente de matriz de varianza-
```

```
## Proporción de varianza explicada por cada componente de matriz de varianza-covarianza :
##  [1] 9.034543e-01 9.647298e-02 6.795804e-05 4.554567e-06 1.782429e-07
##  [6] 7.530917e-09 5.317738e-09 6.657763e-10 8.502887e-11 2.107843e-11
## [11] 6.989035e-12
```

4. Acumule los resultados anteriores.

```
ar_S = cumsum(pve_S)
imprimir_resultados(ar_S, "Acumulado de los resultados anteriores de matriz de varianza-covarianza")
```

```
## Acumulado de los resultados anteriores de matriz de varianza-covarianza :
##  [1] 0.9034543 0.9999273 0.9999953 0.9999998 1.0000000 1.0000000 1.0000000
##  [8] 1.0000000 1.0000000 1.0000000 1.0000000
```

5. Según los resultados anteriores, ¿qué componentes son los más importantes?, ¿qué variables

son las que más contribuyen al primer y segundo componentes principales?, ¿por qué lo dice?, ¿influyen las unidades de las variables?

Se observa que los componentes más importantes son el primer y segundo componente. Por otro lado, las variables que más contribuyen a estos mismos componentes principales son: PNB95 y ProdElec. Estas variables cuentan con constantes predominantes en los vecotres correpondientes a los componentes principales. Contemplando los datos proporcionados se observa que los rangos numéricos de las variables son significativamente diferentes entre sí; las variables que más contribyen cuentan con valores muy grandes. Por lo tanto, se puede afirmar que hay una influencia de las unidades de las variables.

6. Hacer los mismos pasos anteriores, pero con la matriz de correlaciones.

```
# Proporcion de varianza explicada por cada componente
pve_R = vvp_R$values / sum(diag(R))
imprimir_resultados(pve_R, "Proporción de varianza explicada por cada componente de matriz de correlaci
```

```
## Proporción de varianza explicada por cada componente de matriz de correlación :
##  [1] 0.366352638 0.175453813 0.124582832 0.078592361 0.072194597 0.066290906
##  [7] 0.051936828 0.029709178 0.015278951 0.013302563 0.006305332
```

```
# Acumulado de los resultados anteriores
ar_R = cumsum(pve_R)
imprimir_resultados(ar_R, "Acumulado de los resultados anteriores de matriz de correlaciones")
```

```
## Acumulado de los resultados anteriores de matriz de correlaciones :
##  [1] 0.3663526 0.5418065 0.6663893 0.7449816 0.8171762 0.8834671 0.9354040
##  [8] 0.9651132 0.9803921 0.9936947 1.0000000
```

7. Compare los resultados de los incisos 5 y 6. ¿Qué concluye?.

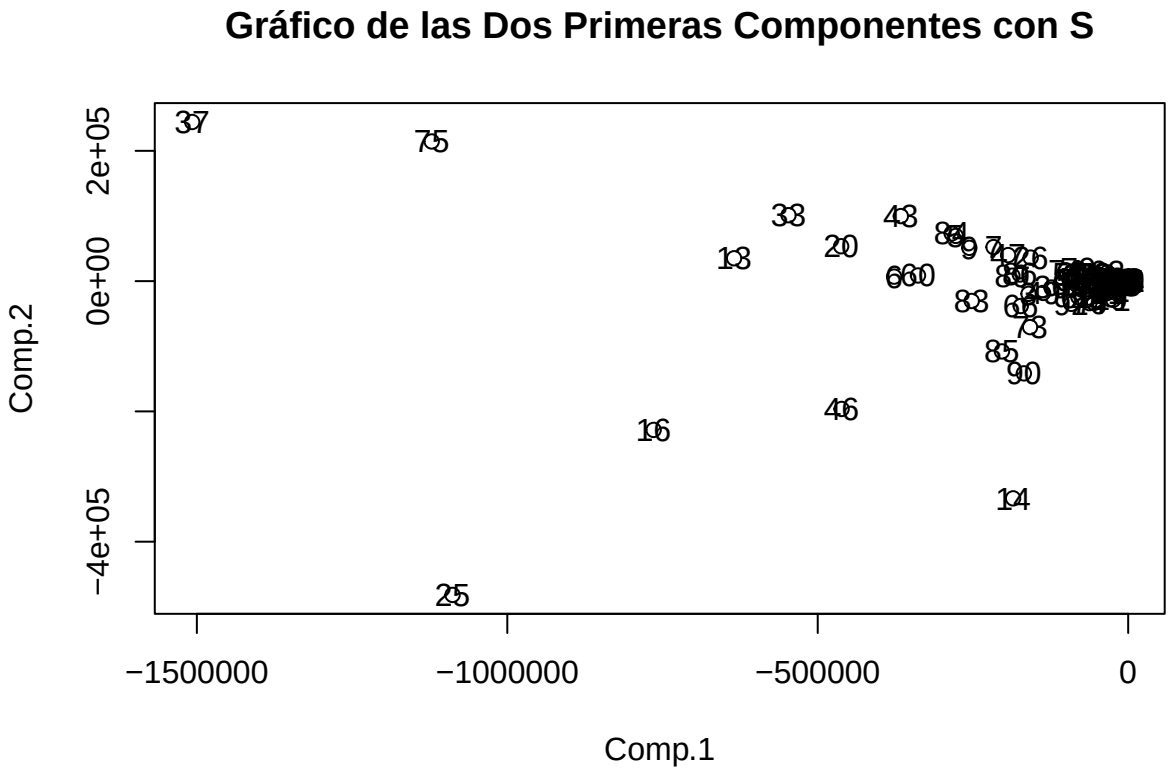
Por un lado, por medio de la matriz de varianza-covarianza se determinó que los componentes más importantes son el primer y segundo componente; considerando que las variables que más contribuyen a estos mismos componentes son PNB95 y ProdElec. Mientras que, por medio de la matriz de correlación se observa que los componentes más importantes son del primer al quinto; considerando que las variables que más contribuyen a estos mismo componentes varían de vector en vector. Por lo tanto, los resultados que corresponden a la matriz de correlación son los más apropiados para tomar la decisión de la seleccion de variables dado que la cantidad de componentes más importantes y las variables que más contribuyen son más realistas.

PARTE II

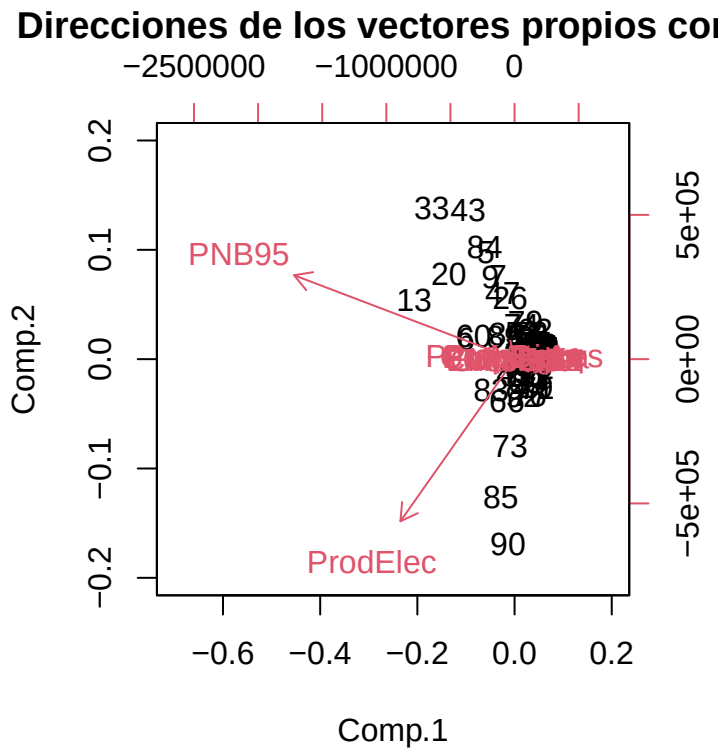
Obtenga las gráficas de respectivas con S y con R de los dos primeros componentes e interprete los resultados en término de agrupación de variables.

```
datos = X
xlim <- c(-0.2, 0.2)
ylim <- c(-0.2, 0.2)

# PCA con S
cpS=princomp(datos,cor=FALSE)
cpaS=as.matrix(datos)%*%cpS$loadings
plot(cpaS[,1:2],type="p", main = "Gráfico de las Dos Primeras Componentes con S")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```

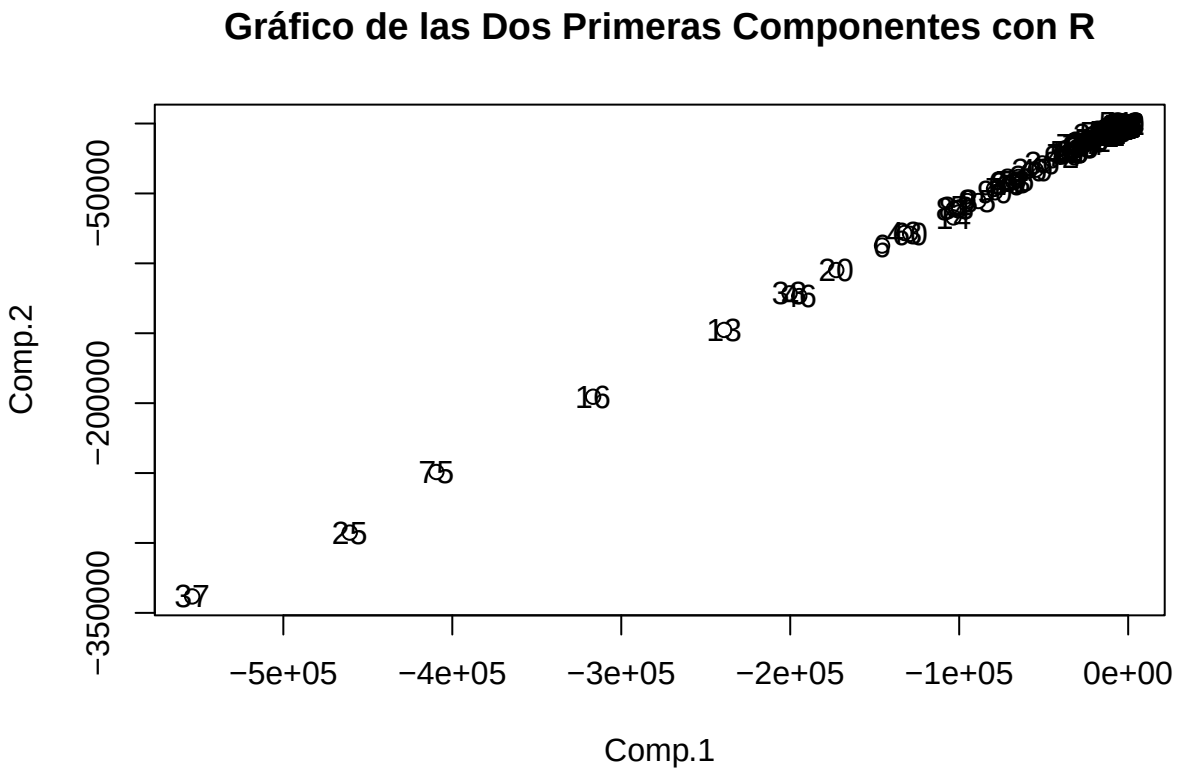



```
biplot(cpS, main="Direcciones de los vectores propios con S", xlim = c(-0.7, 0.2),ylim = ylim)
```

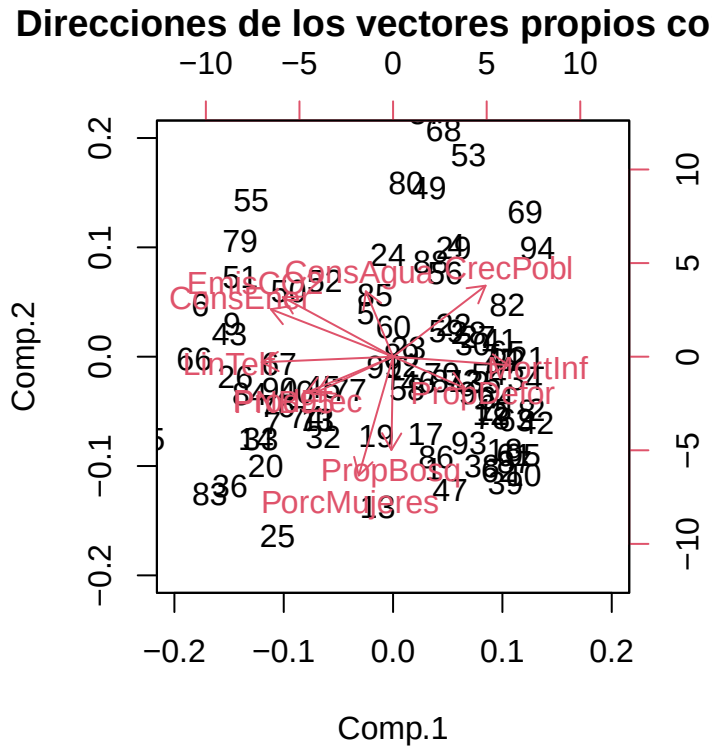



```
# PCA con R
cpR=princomp(datos,cor=TRUE)
cpaR=as.matrix(datos)%*%cpR$loadings
plot(cpaR[,1:2],type="p", main = "Gráfico de las Dos Primeras Componentes con R")
```

```
text(cpaR[,1],cpaR[,2],1:nrow(cpaR))
```




```
biplot(cpR, main="Direcciones de los vectores propios con R", xlim = xlim, ylim = ylim)
```



Análisis de Componentes Principales con la Matriz de Varianza-Covarianza

- Gráfico de dispersión de las Dos Primeras Componentes: las observaciones están concentradas en el origen, esto sugiere que las variables originales tienen una alta correlación entre sí. También esto puede indicar que algunas de las variables originales pueden ser redundantes o proporcionar información similar.
- Direcciones de los vectores propios: los vectores más largos y visibles correponden a las variables PNB95 y ProdElec, lo cual indica que son las variables que más contribuyen a las dos primeras componentes principales. Su dirección indica que las variables están correlacionadas negativamente con la driección positiva de la primera componente principal.

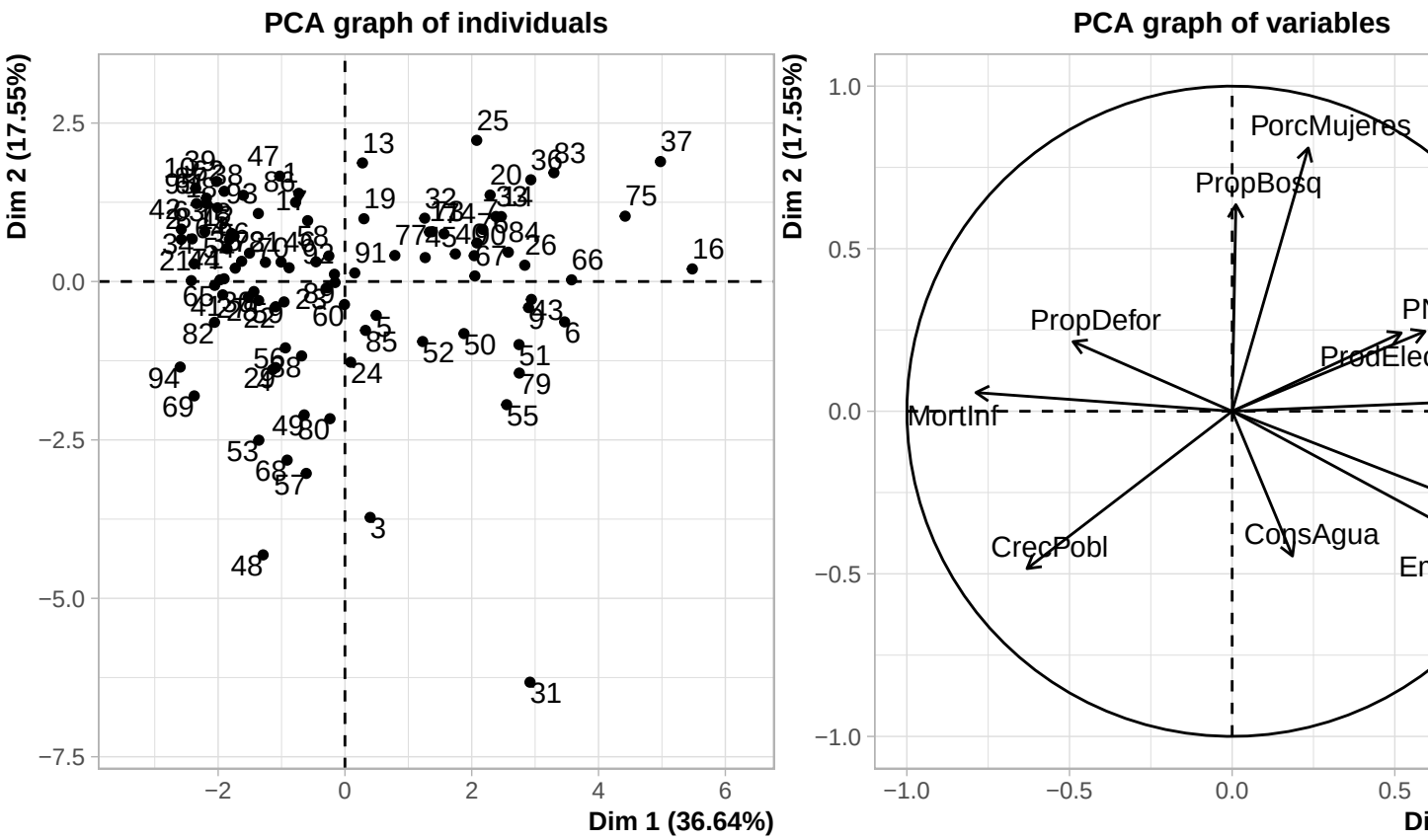
Análisis de Componentes Princiaples con la Matriz de Correlación

- Gráfico de dispersión de las Dos Primeras Componentes: la mayoría de las observaciones están agrupadas en un punto específico, esto sugiere una fuerte correlación negativa entre las variables originales en las direcciones de las primeras dos componentes principales. Este patrón podría indicar que algunas de las variables originales están altamente correlacionadas negativamente entre sí.
- Direcciones de los vectores propios: todos los vectores cuentan con la misma proporción y apuntan hacia afuera de la gráfica, esto es una característica comun en un biplot cuando se utiliza una matriz de correlación. Esto sugiere que todas las variables tienen la misma importancia relativa en términos de varianza.

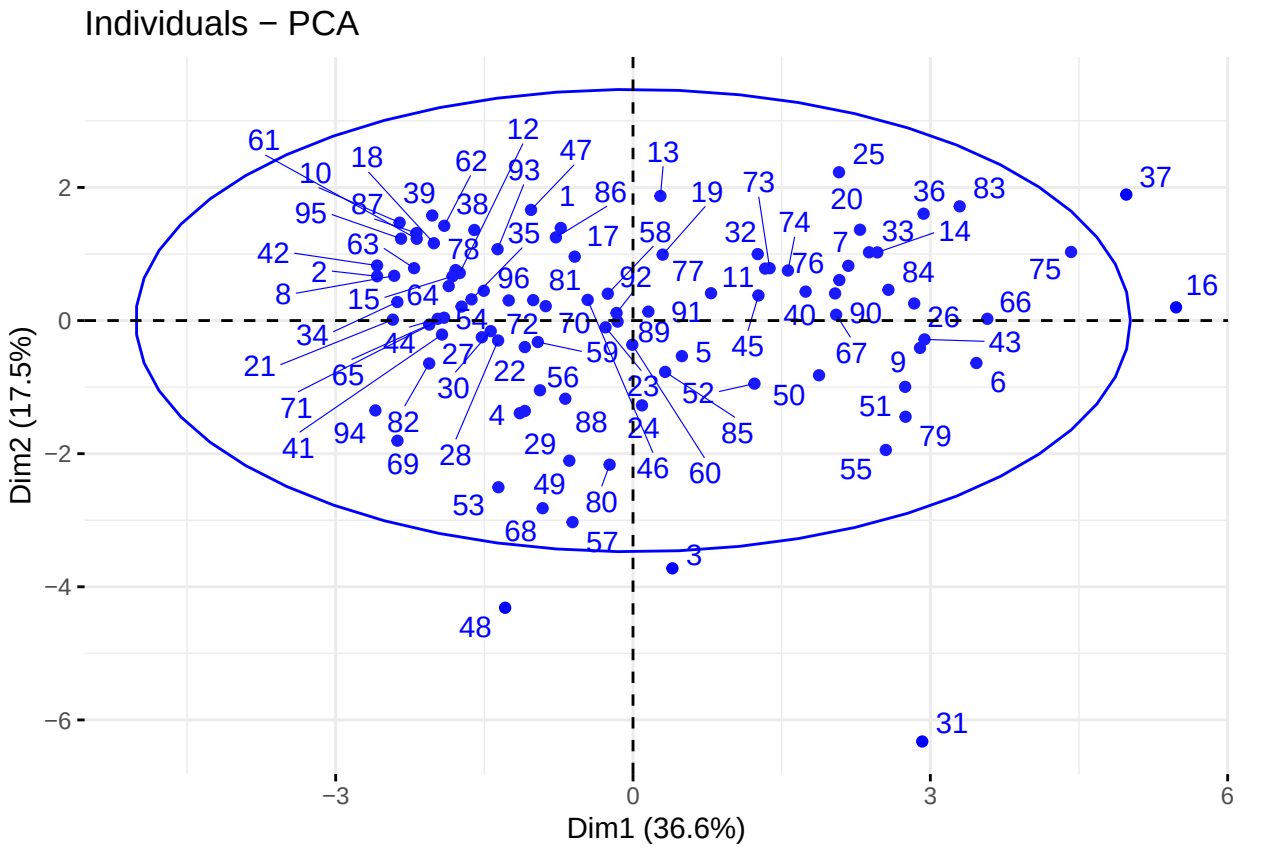
PARTE III

Explore los gráficos relativos al problema Componentes Princiales y dé una interpretación de cada gráfico.

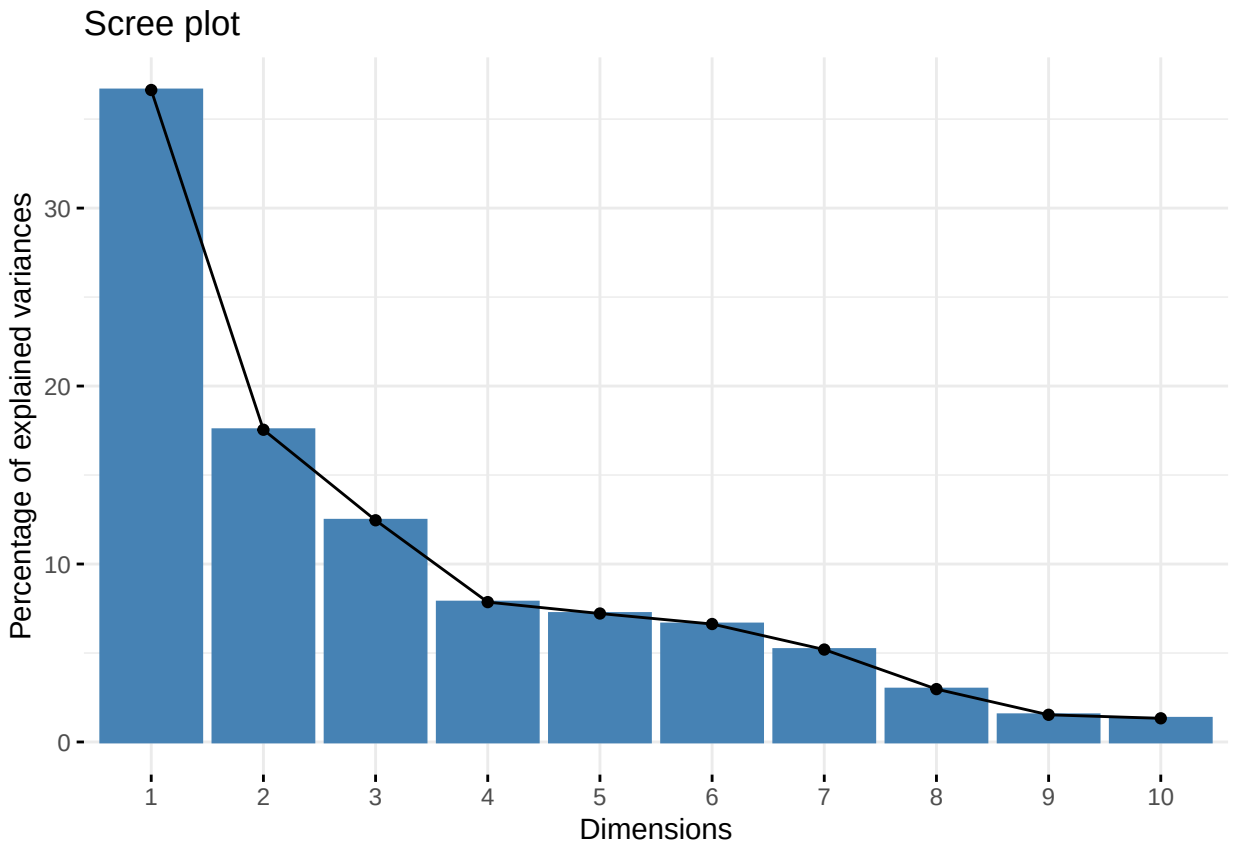
```
datos=X
cp3 = PCA(datos)
```



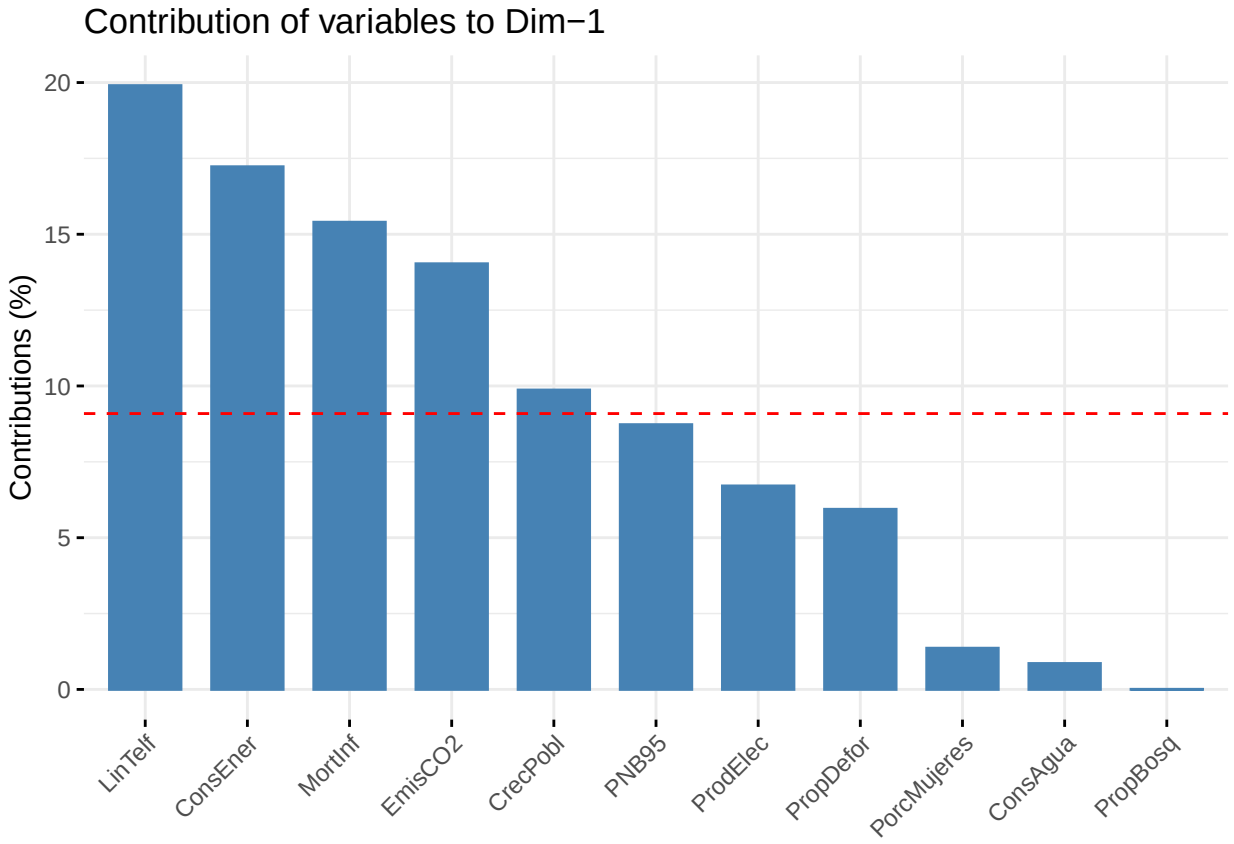
```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```




```
fviz_screeplot(cp3)
```




```
fviz_contrib(cp3, choice = c("var"))
```



- PCA graph of individuals: las observaciones que están concentradas en el segundo cuadrante indican que están relacionadas de alguna manera y comparten patrones similares en las variables originales que se han reducido a través del PCA. La dispersión de observaciones en el tercer y cuarto cuadrante sugiere que estas observaciones son más diversas y heterogéneas en términos de sus características originales.
- PCA graph of variables: las variables que contribuyen más significativamente a la variablilidad total en los datos son ConsEner, EmisCO2, PorcMujeres, MortInf, y CrecPobl.
- Individuals - PCA: la elipse que encierra la mayoría de las observaciones indica una cierta agrupación o similitud entre estas observaciones en términos de sus puntuaciones en las componentes principales. La presencia de un hueco en el lado izquierdo de la elipse sugiere que en el espacio de las componentes principales, hay un área donde no hay observaciones representadas.
- Scree plot: proporciona una guía útil para seleccionar cuántos componentes principales utilizar en función de la cantidad de varianza que explican. En este caso, los primeros cinco componentes principales capturan parte significativa de la varianza.
- Contribution of variables to Dim-1: proporciona información sobre cómo las varibales individuales contribuyen a la primer dimensión principal del PCA. En este caso las variables LinTelf, ConsEner, MortInf, EmisCO2 y CrecPobl son las que tienen contribuciones significativamente más altas a Dim-1 en comparación con otras variables.