Componentes principales 2

Ileana Parra

13/10/2022

#PARTE 1

```
M=read.csv("C:/Users/ilean/Desktop/ITESM/7to Semestre/Parte
2/paises_mundo.csv")
```

##Matriz de varianza-covarianza

```
cov(M)

##                  CrecPobl       MortInf   PorcMujeres         PNB95
## CrecPobl     1.538298e+00  2.195026e+01 -6.078026e+00 -8.933379e+04
## MortInf      2.195026e+01  1.032859e+03 -9.249342e+00 -2.269332e+06
## PorcMujeres -6.078026e+00 -9.249342e+00  7.698322e+01  2.813114e+05
## PNB95       -8.933379e+04 -2.269332e+06  2.813114e+05  4.999786e+10
## ProdElec    -4.973964e+04 -1.043435e+06  2.260248e+05  2.247791e+10
## LinTelf     -1.369079e+02 -4.381366e+03  4.499750e+02  2.039550e+07
## ConsAgua    -4.827092e+01 -1.288211e+03 -1.568313e+03  1.097481e+07
## PropBosq    -3.887018e+00 -1.466316e+01  6.517895e+01  2.474311e+05
## PropDefor    3.361974e-01  1.276296e+01  2.680592e-01 -5.806203e+04
## ConsEner    -8.384169e+02 -4.442568e+04  2.855207e+02  1.415628e+08
## EmisCO2     -1.137877e+00 -9.485500e+01 -2.150132e+00  2.501673e+05
##                  ProdElec       LinTelf      ConsAgua      PropBosq
## CrecPobl    -4.973964e+04 -1.369079e+02 -4.827092e+01     -3.887018
## MortInf     -1.043435e+06 -4.381366e+03 -1.288211e+03    -14.663158
## PorcMujeres  2.260248e+05  4.499750e+02 -1.568313e+03     65.178947
## PNB95        2.247791e+10  2.039550e+07  1.097481e+07 247431.122807
## ProdElec     1.821909e+10  7.583050e+06  1.399817e+07  70359.785965
## LinTelf      7.583050e+06  3.841247e+04  1.193110e+04    248.715789
## ConsAgua     1.399817e+07  1.193110e+04  3.301981e+05  -2220.757895
## PropBosq     7.035979e+04  2.487158e+02 -2.220758e+03    401.003509
## PropDefor   -3.180340e+04 -9.940461e+01 -6.743793e+01      2.625263
## ConsEner     6.801296e+07  3.426262e+05  2.092242e+05  -5153.438596
## EmisCO2      1.392779e+05  6.385700e+02  4.869328e+02    -12.897193
##                 PropDefor      ConsEner       EmisCO2
## CrecPobl     3.361974e-01 -8.384169e+02     -1.137877
## MortInf      1.276296e+01 -4.442568e+04    -94.855000
## PorcMujeres  2.680592e-01  2.855207e+02     -2.150132
## PNB95       -5.806203e+04  1.415628e+08 250167.323509
## ProdElec    -3.180340e+04  6.801296e+07 139277.888640
## LinTelf     -9.940461e+01  3.426262e+05    638.570000
## ConsAgua    -6.743793e+01  2.092242e+05    486.932763
## PropBosq     2.625263e+00 -5.153439e+03    -12.897193
## PropDefor    1.817253e+00 -1.051522e+03     -2.632487
```

```
## ConsEner    -1.051522e+03  5.014395e+06  10286.159781
## EmisCO2     -2.632487e+00  1.028616e+04     27.268614
```

##Matriz de correlaciones

```
cor(M)

##                CrecPobl     MortInf PorcMujeres       PNB95
ProdElec
## CrecPobl     1.00000000  0.55067948 -0.55852711 -0.32212154 -
0.29711119
## MortInf      0.55067948  1.00000000 -0.03280139 -0.31579250 -
0.24053689
## PorcMujeres -0.55852711 -0.03280139  1.00000000  0.14338826
0.19085114
## PNB95       -0.32212154 -0.31579250  0.14338826  1.00000000
0.74476081
## ProdElec    -0.29711119 -0.24053689  0.19085114  0.74476081
1.00000000
## LinTelf     -0.56321228 -0.69558922  0.26167018  0.46539599
0.28664508
## ConsAgua    -0.06772953 -0.06975563 -0.31106243  0.08541500
0.18047653
## PropBosq    -0.15650281 -0.02278415  0.37096694  0.05525919
0.02603078
## PropDefor    0.20107881  0.29459348  0.02266339 -0.19262327 -
0.17478434
## ConsEner    -0.30187731 -0.61731132  0.01453216  0.28272492
0.22501894
## EmisCO2     -0.17568860 -0.56520778 -0.04692837  0.21425123
0.19760017
##                 LinTelf    ConsAgua    PropBosq   PropDefor
ConsEner
## CrecPobl    -0.56321228 -0.06772953 -0.15650281  0.20107881 -
0.30187731
## MortInf     -0.69558922 -0.06975563 -0.02278415  0.29459348 -
0.61731132
## PorcMujeres  0.26167018 -0.31106243  0.37096694  0.02266339
0.01453216
## PNB95        0.46539599  0.08541500  0.05525919 -0.19262327
0.28272492
## ProdElec     0.28664508  0.18047653  0.02603078 -0.17478434
0.22501894
## LinTelf      1.00000000  0.10593934  0.06337138 -0.37623801
0.78068385
## ConsAgua     0.10593934  1.00000000 -0.19299225 -0.08705811
0.16259804
## PropBosq     0.06337138 -0.19299225  1.00000000  0.09725032 -
0.11492480
## PropDefor   -0.37623801 -0.08705811  0.09725032  1.00000000 -
0.34833836
```

```
## ConsEner     0.78068385  0.16259804 -0.11492480 -0.34833836
1.00000000
## EmisCO2      0.62393719  0.16227447 -0.12333592 -0.37396154
0.87965517
##                 EmisCO2
## CrecPobl    -0.17568860
## MortInf     -0.56520778
## PorcMujeres -0.04692837
## PNB95        0.21425123
## ProdElec     0.19760017
## LinTelf      0.62393719
## ConsAgua     0.16227447
## PropBosq    -0.12333592
## PropDefor   -0.37396154
## ConsEner     0.87965517
## EmisCO2      1.00000000
```

##Valores y vectores propios de S

```
lambda=eigen(cov(M))
lambda$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
```

```
lambda$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 -
0.0944030203
##  [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  2.258020e-02 -7.891128e-02 -1.637836e-02
4.264872e-04
##  [3,]  8.552992e-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 -3.416812e-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,]  4.944397e-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 la matriz de correlaciones

```
lambda2=eigen(cor(M))
lambda$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

lambda$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 -
0.0944030203
##  [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  2.258020e-02 -7.891128e-02 -1.637836e-02
4.264872e-04
##  [3,]  8.552992e-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 -3.416812e-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,]  4.944397e-04
##  [7,]  4.780416e-04
##  [8,] -3.748976e-03
##  [9,] -1.052934e-01
## [10,]  5.906241e-05
## [11,] -8.221371e-02
```

Cálculo de la proporción de varianza explicada por cada componente con matriz S.

```
S=cov(M)
varianza=sum(diag(S))
v=sum(lambda$values)
print("La varianza total es:")

## [1] "La varianza total es:"

print(varianza)

## [1] 68222335253

lambda$values/varianza

##  [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

print("La suma acumulada es:")

## [1] "La suma acumulada es:"

cumsum(lambda$values/varianza)

##  [1] 0.9034543 0.9999273 0.9999953 0.9999998 1.0000000 1.0000000
1.0000000
##  [8] 1.0000000 1.0000000 1.0000000 1.0000000
```

El primer componente explica el 90.34% de la varianza, por lo que este es el componente más importante. Las variables que más influyen en los componentes, son las variables con coeficientes grandes. *Las variables que más influyen en el primer componente son: PorcMujeres y PNB95.* Las variables que más influyen en el segundo componente son: ProbBosques y ProdElec.

Cálculo de la proporción de varianza explicada por cada componente con matriz de correlaciones.

```
C=cor(M)
varianza2=sum(diag(C))
v2=sum(lambda2$values)
print("La varianza total es:")
```

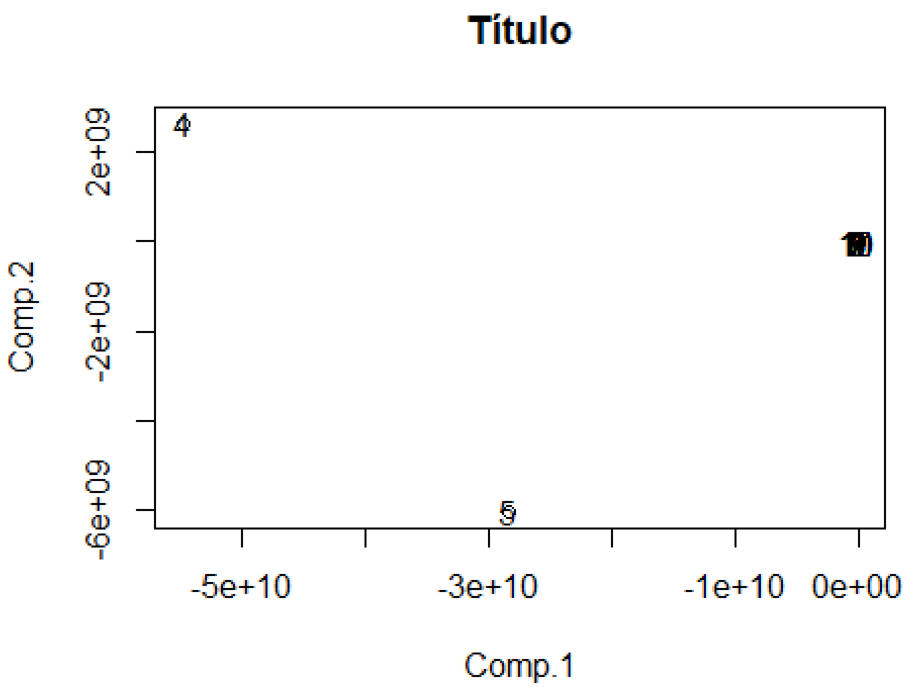
```
## [1] "La varianza total es:"

print(varianza2)

## [1] 11

lambda2$values/varianza2

##  [1] 0.366352638 0.175453813 0.124582832 0.078592361 0.072194597
0.066290906
##  [7] 0.051936828 0.029709178 0.015278951 0.013302563 0.006305332

print("La suma acumulada es:")

## [1] "La suma acumulada es:"

cumsum(lambda2$values/varianza)

##  [1] 5.906979e-11 8.735953e-11 1.074470e-10 1.201190e-10 1.317595e-10
##  [6] 1.424481e-10 1.508222e-10 1.556124e-10 1.580760e-10 1.602209e-10
## [11] 1.612375e-10
```

Se concluye que los componentes 1 y 2 son los componentes más importantes ya que explican la mayor proporción de varianza. Con la matriz S, es más fácil de entender los resaultados ya que su suma acumulada es 1.

#Parte 2

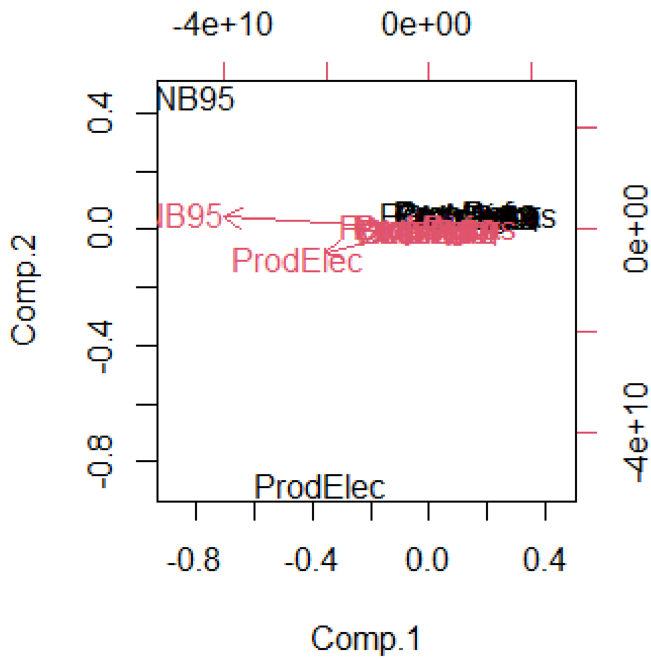
##Gráfica de S

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

## Loading required package: ggplot2

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

library(ggplot2)

datos=cov(M)
cpS=princomp(datos,cor=FALSE)
cpaS=as.matrix(datos)%*%cpS$loadings
plot(cpaS[,1:2],type="p", main = "Título")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```




```
biplot(cpS)

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

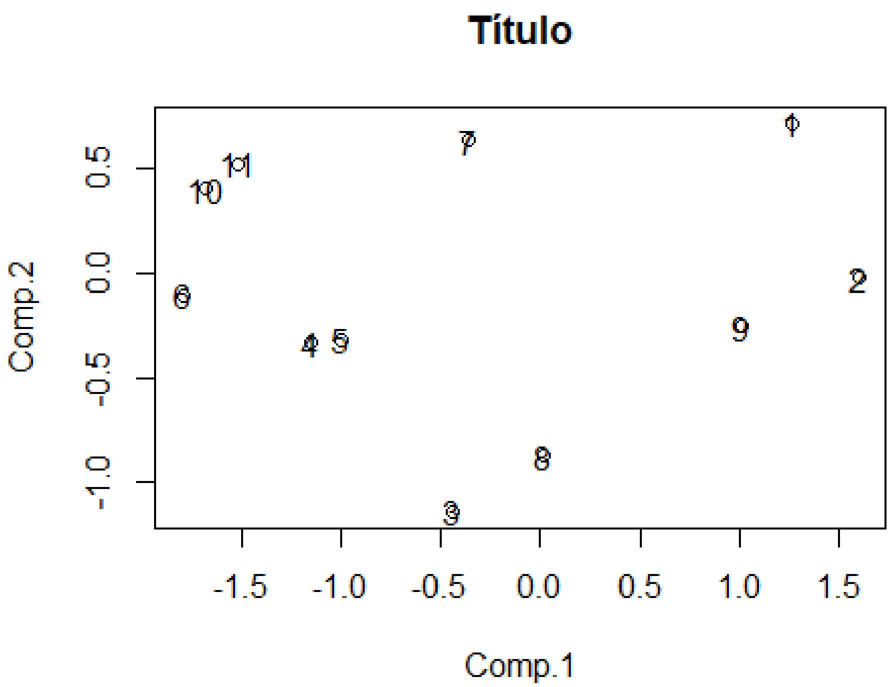
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```



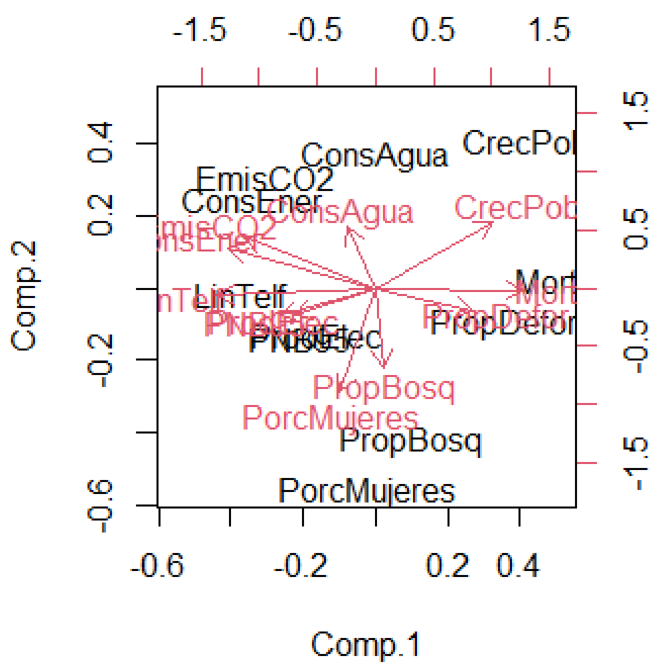
*Prod Elec y PnB95 influyen negativamente en el componente 1

##Gráfica de R

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

datos=cor(M)
cpS=princomp(datos,cor=FALSE)
cpaS=as.matrix(datos)%*%cpS$loadings
plot(cpaS[,1:2],type="p", main = "Título")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```



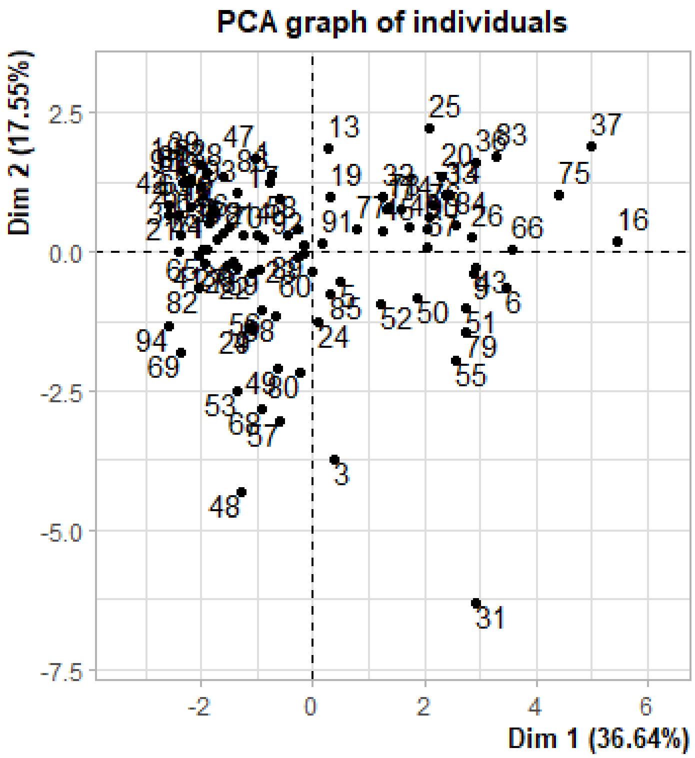

```
biplot(cpS)
```

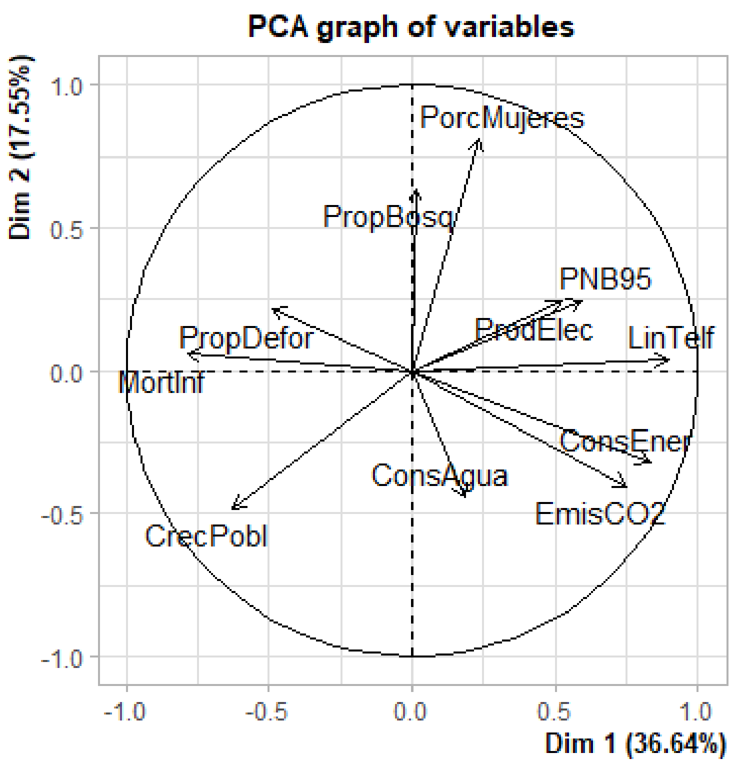


MortInf y LinTelf son las variables que más influyen en el componente 1, sin embargo LinTelf tiene *una correlación negativa.* El componente 2 se enfoca más en proporciones ya que las variables que más influyen son PropBosq y PorcMujeres.

#Parte 3

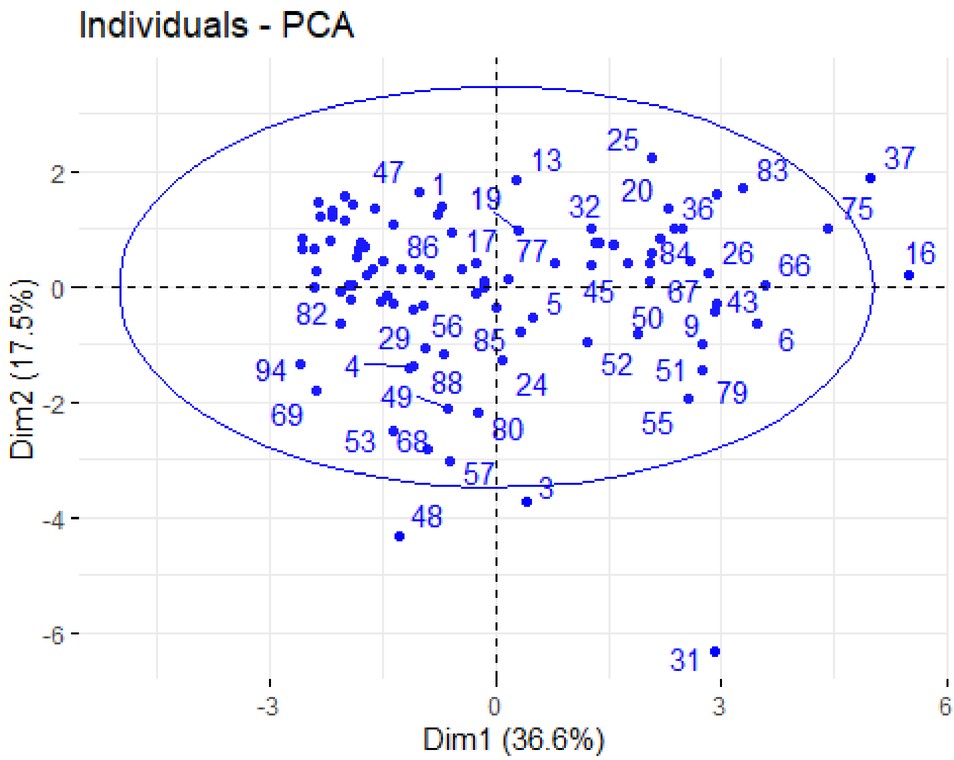
```
library(stats)
library(FactoMineR)
library(factoextra)
library(ggplot2)
datos=M
cp3 = PCA(datos)
```



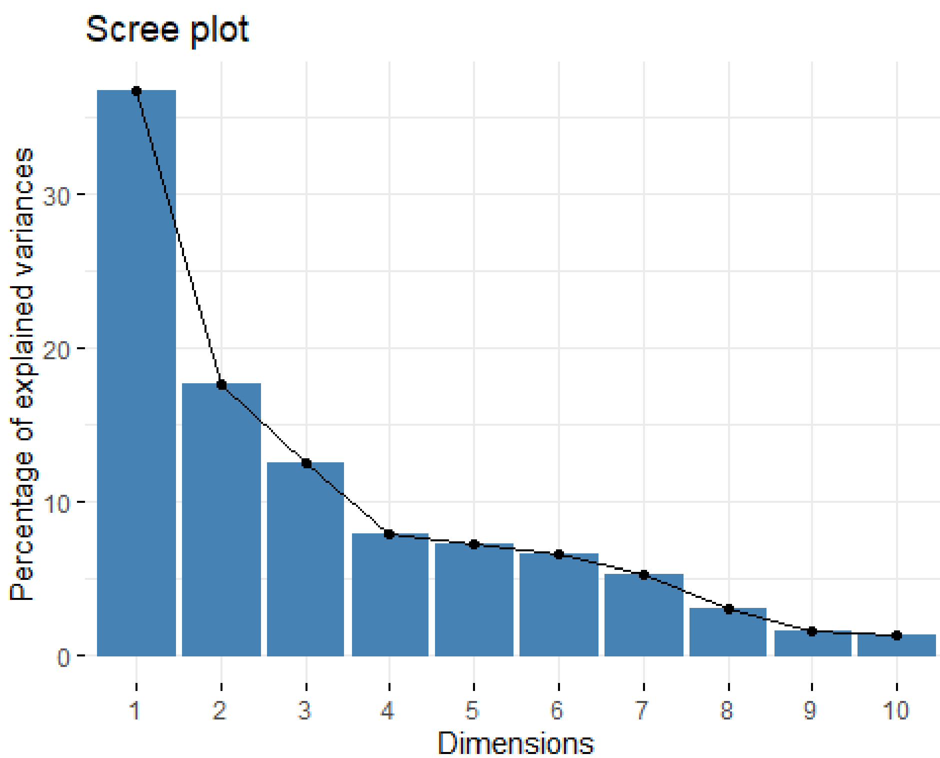



```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

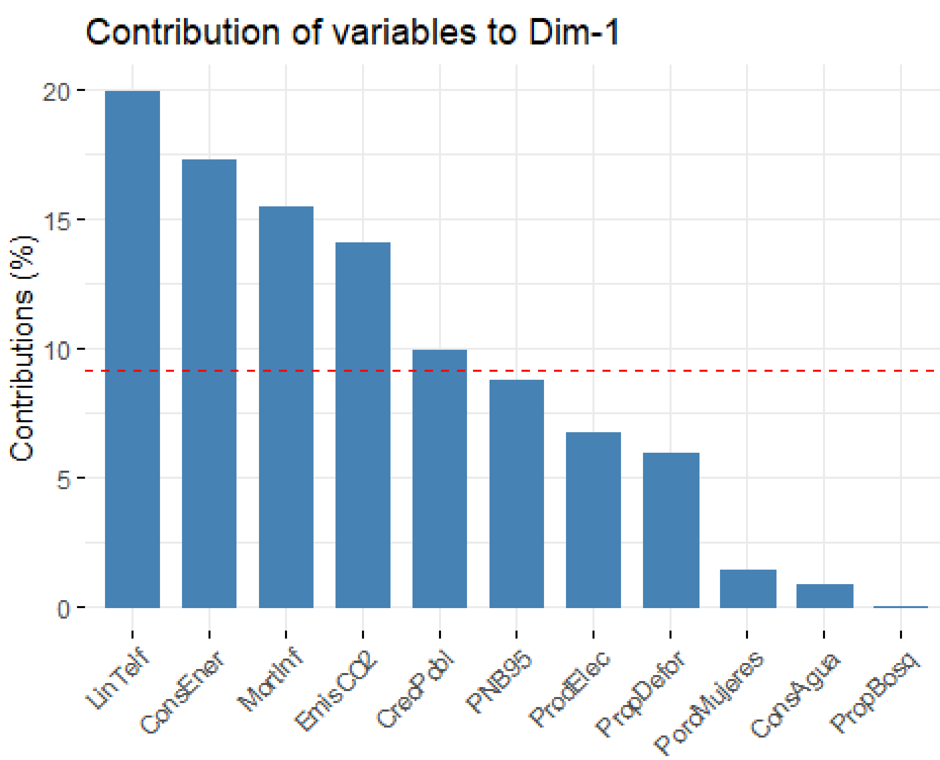
```
## Warning: ggrepel: 50 unlabeled data points (too many overlaps).
Consider
## increasing max.overlaps
```




```
fviz_screeplot(cp3)
```




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



*La primera gráfica indica que los datos no siguen una distribución normal ya que no se distribuyen aleatoriamente alrededor del 0. Se pueden ver algunos outliers como el 31 y un cluster de variables en el segundo cuadrante.

*La segunda gráfica muestra las variables que tienen mayor influencia en los componentes. Se puede ver que LinTelf es la variable con mayor influencia en el primer componente y PorcMujeres para el segundo componente.

*La gráfica de elipses muestra que se puede reducir la dimensionalidad de los datos. Los componentes principales están bien determinados.

*El gráfico de sedimentación muestra que el primer componente explica la mayor parte de la varianza.

*La última gráfica muestra que la variable LinTelf es la que más contribuye al componente 1.