

# Componentes principales 2

Iker Ledesma - A01653115

18 de octubre de 2022

## Parte I

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

```
M = read.csv("países_mundo.csv")
summary(M)
```

```
##      CrecPobl      MortInf      PorcMujeres      PNB95
## Min.   :-0.600   Min.    :  4.00   Min.     :13.00   Min.    :  1353
## 1st Qu.: 0.875   1st Qu.: 11.00   1st Qu.:31.00   1st Qu.:  8014
## Median : 2.000   Median : 32.50   Median :39.00   Median : 29243
## Mean   : 1.871   Mean    : 39.06   Mean    :37.28   Mean    :116587
## 3rd Qu.: 2.725   3rd Qu.: 56.50   3rd Qu.:44.00   3rd Qu.:109222
## Max.    : 5.800   Max.     :124.00   Max.     :51.00   Max.     :1451051
##      ProdElec      LinTelf      ConsAgua      PropBosq
## Min.    :    6   Min.     :  2.0   Min.     :  7.0   Min.     :  0.00
## 1st Qu.: 4676   1st Qu.: 15.5   1st Qu.: 178.5   1st Qu.:10.00
## Median :21422   Median : 74.0   Median : 381.5   Median :25.50
## Mean    :69261   Mean    :165.1   Mean     :509.8   Mean     :27.33
## 3rd Qu.:65784   3rd Qu.:231.5   3rd Qu.: 662.0   3rd Qu.:42.25
## Max.     :928083   Max.     :681.0   Max.     :4575.0   Max.     :77.00
##      PropDefor      ConsEner      EmisCO2
## Min.   :-4.3000   Min.    :  20.0   Min.     :  0.100
## 1st Qu.: -0.0250   1st Qu.: 287.8   1st Qu.:  0.700
## Median : 0.1500   Median : 914.5   Median :  2.600
## Mean    : 0.5531   Mean    :1854.4   Mean     :  4.554
## 3rd Qu.: 1.2000   3rd Qu.:2553.5   3rd Qu.:  7.050
## Max.     : 5.1000   Max.     :10531.0   Max.     :33.900
```

```
cova = cov(M)
corre = cor(M)
```

### 2 Calcule los valores y vectores propios de cada matriz. La función en R es: eigen().

```
lambda_cov = eigen(cova)$values
lambda_cor = eigen(corre)$values
```

**3 Calcule la proporción de varianza explicada por cada componente. Se sugiere dividir cada lambda entre la varianza total (las lambdas están en `eigen(S)[1]`). La varianza total es la suma de las varianzas de la diagonal de S. Una forma es `sum(diag(S))`. La varianza total de los componentes es la suma de los valores propios (es decir, la suma de la varianza de cada componente), sin embargo, si sumas la diagonal de S (es decir, la varianza de cada x), te da el mismo valor (¡compruébalo!). Recuerda que las combinaciones lineales buscan reproducir la varianza de X.**

```
var_cov=sum(diag(cova))
vp_cov=lambda_cov/var_cov

var_corr=sum(diag(corre))
vp_corr=lambda_cor/var_corr
```

**4 Acumule los resultados anteriores. (`cumsum()` puede servirle).**

```
cumsum(vp_cov)

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

cumsum(vp_corr)

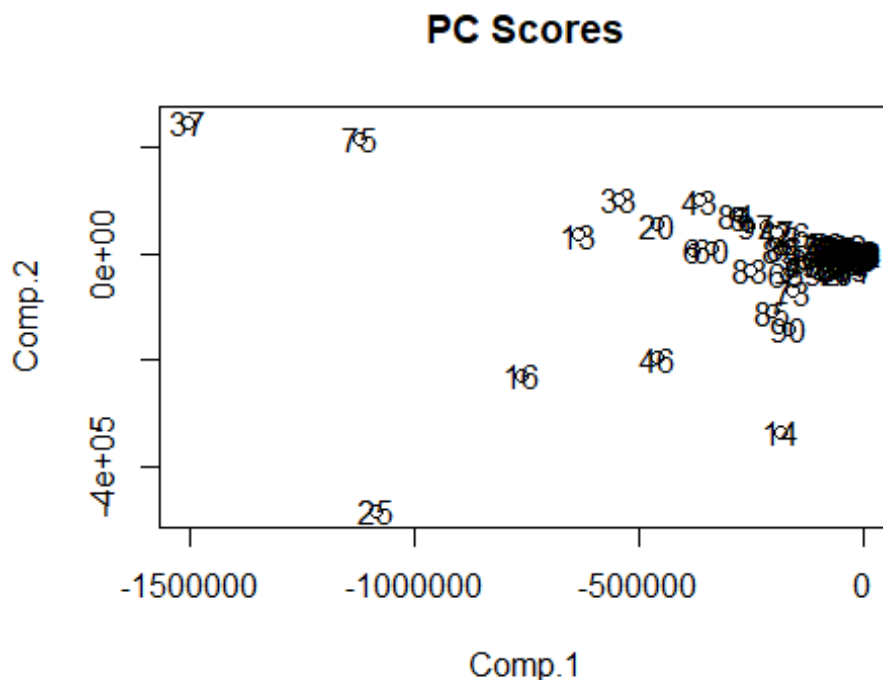
## [1] 0.3663526 0.5418065 0.6663893 0.7449816 0.8171762 0.8834671
## [8] 0.9651132 0.9803921 0.9936947 1.0000000
```

**5 Según los resultados anteriores, ¿qué componentes son los más importantes? ¿qué variables son las que más contribuyen a la primera y segunda componentes principales? ¿por qué lo dice? ¿influyen las unidades de las variables?**

Por lo general se eligen el número de componentes a partir del cual la proporción de la varianza no incrementa significativamente, en este caso se podrían tomar los primeros dos componentes principales puesto que explican el 0.9999273 de la varianza total. Si las variables se miden en diferentes unidades si influyen en el resultado final puesto que las variables que contengan mayor varianza representarán mayor importancia frente a otras puesto que pueden estar medidas con escalas diferentes. Por otro lado por lo generado a través de la varianza acumulada se puede visualizar que a partir del componente principal número 7 ya no se aprecian incrementos significativos en la varianza que recolecta cada componente por lo que se podrían usar los primeros 6 componentes principales y obtener un 93.54% de varianza explicada. # Parte II ## Obtenga las gráficas de respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras

componentes e interprete los resultados en término de agrupación de variables (puede ayudar “índice de riqueza”, “índice de ruralidad”)

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

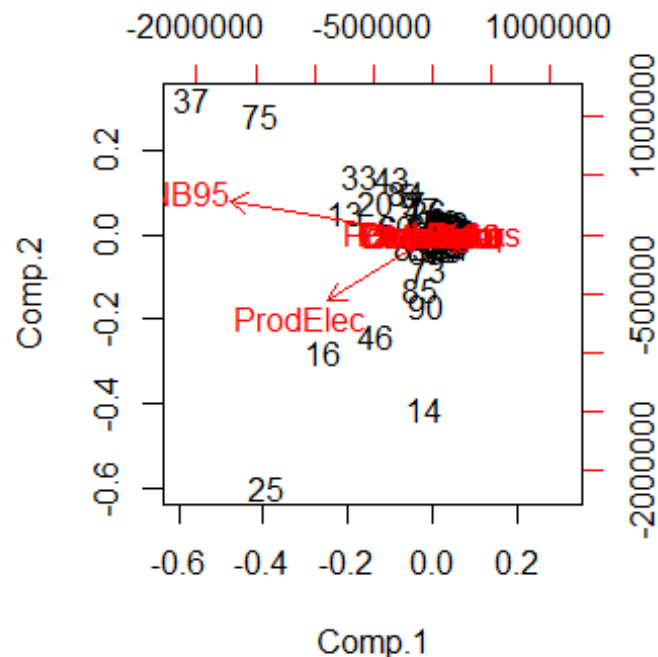
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L],
length
## = arrow.len): zero-length arrow is of indeterminate angle and so
skipped

## 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

```



# Parte III ##

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

```
#No se pudo correr este apartado porque aparecía el error Error in  
library(FactoMineR) : there is no package called 'FactoMineR'  
#library(FactoMineR)  
#datos=M  
#cp3 = PCA(datos, graph = FALSE)  
#fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)  
#fviz_sceplot(cp3)  
#fviz_contrib(cp3, choice = c("var"))
```