

# Actividad 1\_7

Ricardo Kaleb Flores Alfonso

2024-10-08

## 0) Cargamos la base de datos

```
df = read.csv("datoslago.csv")
#Cargamos librería
library(psych)
```

## 1) Obtener la matriz de correlaciones y la matriz de valores p de significancia por pares.

```
corr.test(df,adjust="none")
```

```
## Call:corr.test(x = df, adjust = "none")
## Correlation matrix
##           Tamano  Temperatura  Proteinas  Oxigeno  LnProteinas  LnTamano
## Tamano      1.00      0.23      0.63    -0.34      0.84      0.97
## Temperatura 0.23      1.00     -0.22     0.34     -0.08     0.22
## Proteinas   0.63     -0.22      1.00    -0.37     0.78     0.57
## Oxigeno     -0.34     0.34     -0.37      1.00    -0.38    -0.31
## LnProteinas 0.84     -0.08     0.78    -0.38      1.00     0.86
## LnTamano    0.97     0.22     0.57    -0.31     0.86     1.00
## Sample Size
## [1] 90
## Probability values (Entries above the diagonal are adjusted for multiple tests.)
##           Tamano  Temperatura  Proteinas  Oxigeno  LnProteinas  LnTamano
## Tamano      0.00      0.03      0.00      0        0.00      0.00
## Temperatura 0.03      0.00      0.04      0        0.46      0.04
## Proteinas   0.00      0.04      0.00      0        0.00      0.00
## Oxigeno     0.00      0.00      0.00      0        0.00      0.00
## LnProteinas 0.00      0.46      0.00      0        0.00      0.00
## LnTamano    0.00      0.04      0.00      0        0.00      0.00
##
## To see confidence intervals of the correlations, print with the short=FALSE option
```

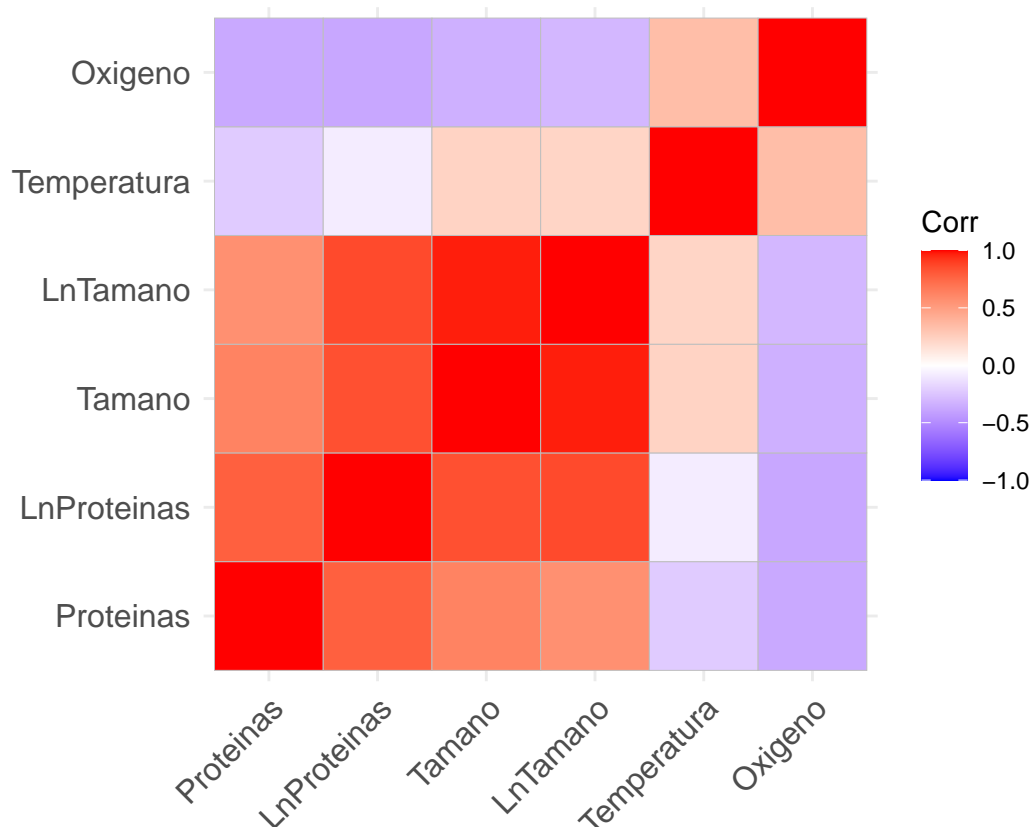
La correlación entre Tamaño y LnTamaño tiene un valor de 0.97, lo que indica que las variables tienen una alta correlación lineal. De igual manera Tamaño y Proteínas tiene una alta correlación con 0.63. Para los p valores, la mayoría de las correlaciones entre las variables son significativas menos la deLnproteinas con Temperatura. Lo que implica que existen relaciones entre las distintas variables

2) Hacer una gráfica de la matriz de correlaciones. Hacer un comentario entre todos los del equipo de la gráfica.

```
#Cargamos librerias
library(polycor)

##
## Adjuntando el paquete: 'polycor'
## The following object is masked from 'package:psych':
##
##   polyserial
library(ggcorrplot)

## Cargando paquete requerido: ggplot2
##
## Adjuntando el paquete: 'ggplot2'
## The following objects are masked from 'package:psych':
##
##   %+%, alpha
mat_cor <- hetcor(df)$correlations
ggcorrplot(mat_cor, hc.order = T)
```



En este caso tambien se observa la alta correlación lineal que existe, tanto positiva como negativa entre las variables.

### 3) Aplicar una prueba de correlación conjunta a los datos para verificar si es aplicable el Análisis Factorial y concluir.

$H_0 : p \leq 0.05$  las variables son ortogonales, no son correlacionadas  $H_1 : p > 0.05$  No es cierto, la matriz de correlación difiere significativamente de la matriz identidad, por tanto, sí hay correlación conjunta significativa.

```
#Cargamos librería
library(performance)

check_sphericity_bartlett(df)

## # Test of Sphericity
##
## Bartlett's test of sphericity suggests that there is sufficient significant correlation in the data :
# Para obtener valor del estadístico (chi-cuadrada), del parámetro (grados de libertad) y del valor p):
b = check_sphericity_bartlett(df)
b$chisq

## [1] 557.8585

b$p

## [1] 3.135832e-109

b$dof

## [1] 15
```

Al obtener un valor de p extremadamente bajo y menor que 0.05, hay suficiente evidencia para rechazar la hipótesis nula, por lo que se acepta la hipótesis alternativa y si existe suficiente correlación para hacer el análisis factorial.

### 4) Otra prueba para, para comprobar si el análisis factorial es viable, y muy citada, es la prueba KMO. Aplíquela a estos datos, ¿contradice los resultados del inciso anterior?

0.90 a 1.00 maravilloso 0.80 a 0.89 meritorio 0.70 a 0.79 medio 0.60 a 0.69 mediocre 0.50 a 0.59 miserable 0.00 a 0.49 inaceptable

```
R = cor(df)
K = KMO(R)
cat("El valor del estadístico es: ", K$MSA)
```

```
## El valor del estadístico es: 0.6297281
```

El valor del estadístico fue de 0.6297, lo que es un resultado mediocre en la escala, eso significa que el análisis factorial es viable, pero no es ideal.

### 5) Si los datos pasaron la prueba de los puntos anteriores 3 y 4, hacer un análisis factorial usando el criterio de máxima verosimilitud y el de mínimo residuo

```

R = cor(df)
modelo1 = fa(R, nfactors =2, rotate = "none", fm = "mle")
modelo2 = fa(R, nfactors = 2, rotate = "none", fm = "minres")

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

M1_commd = sort(modelo1$communality,decreasing = T)
M2_commd= sort(modelo2$communality,decreasing = T)
cbind(M1_commd,M2_commd)

##           M1_commd  M2_commd
## LnTamano    0.9950455 0.9748176
## Tamano      0.9558283 0.9238582
## LnProteinas 0.9181230 0.9030553
## Proteinas   0.7708383 0.8453996
## Temperatura 0.3963254 0.6029645
## Oxigeno     0.2071029 0.2835476

```

La variables que explican de mejor manera la varianza de los datos son LnTamaño, Tamaño y Lnproteinas. Tanto para el analisis de maxima verosimilitud y minimo residuo.

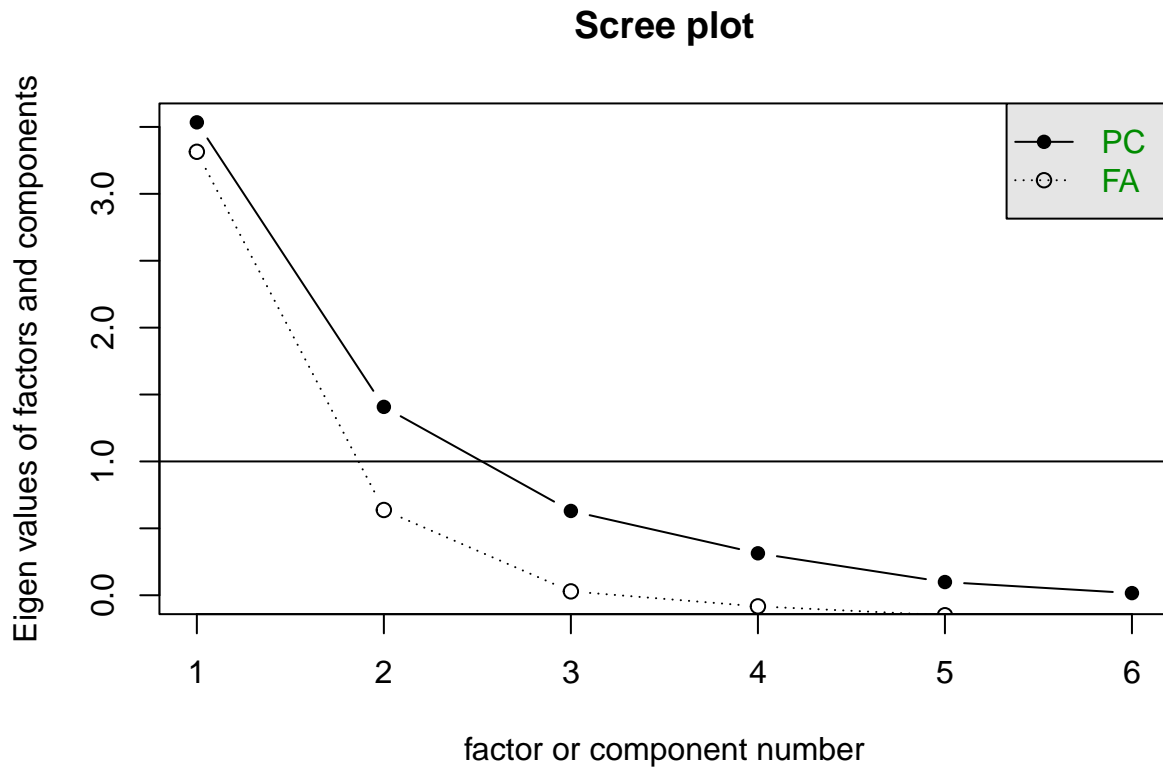
## 6) Determine el número de factores adecuado según el criterio del gráfico de Cattell

```

R = cor(df)
scree(R)

## Warning in fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs, :
## The estimated weights for the factor scores are probably incorrect. Try a
## different factor score estimation method.

```

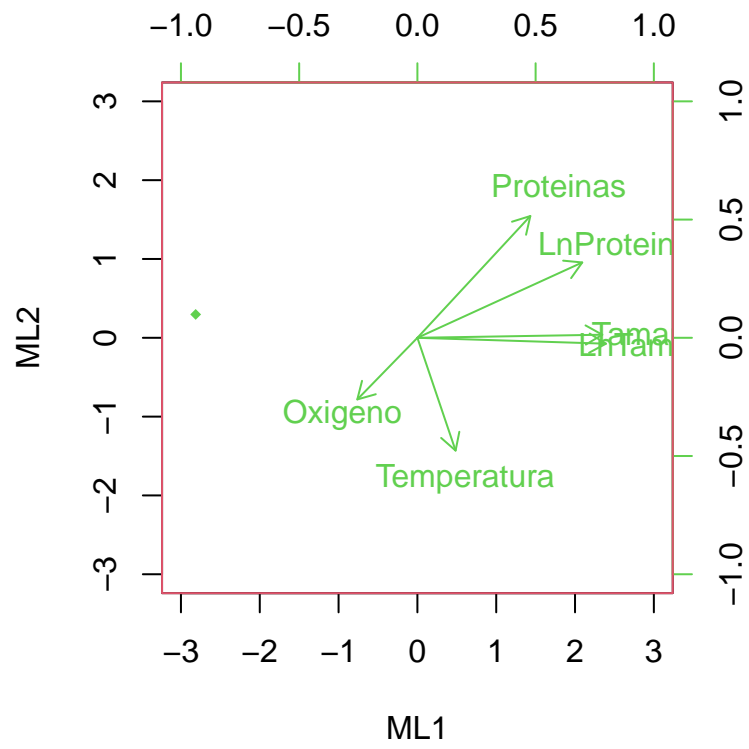


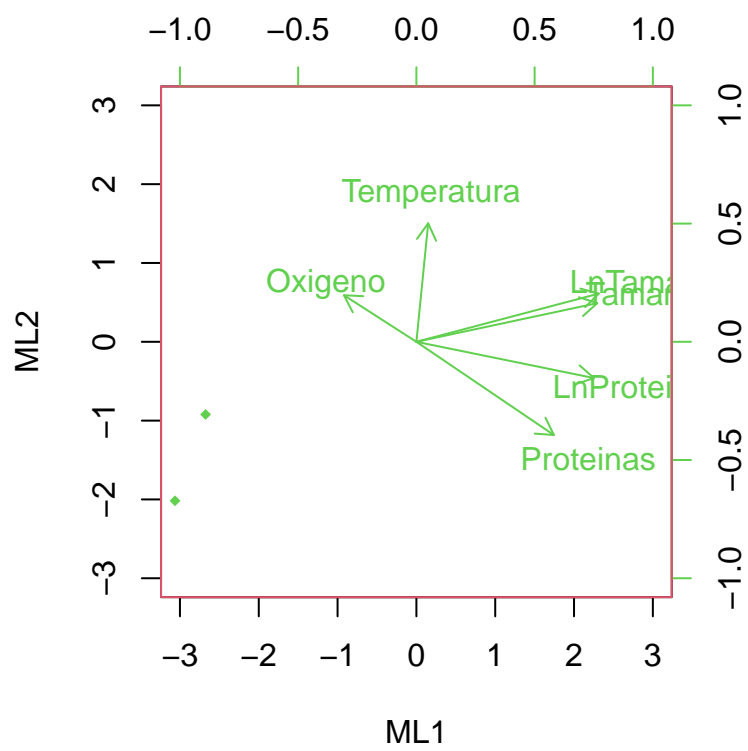
El gráfico muestra una que después del segundo factor, la ganancia de información es menor, por lo que dos factores son suficientes para explicar la mayor parte de la varianza.

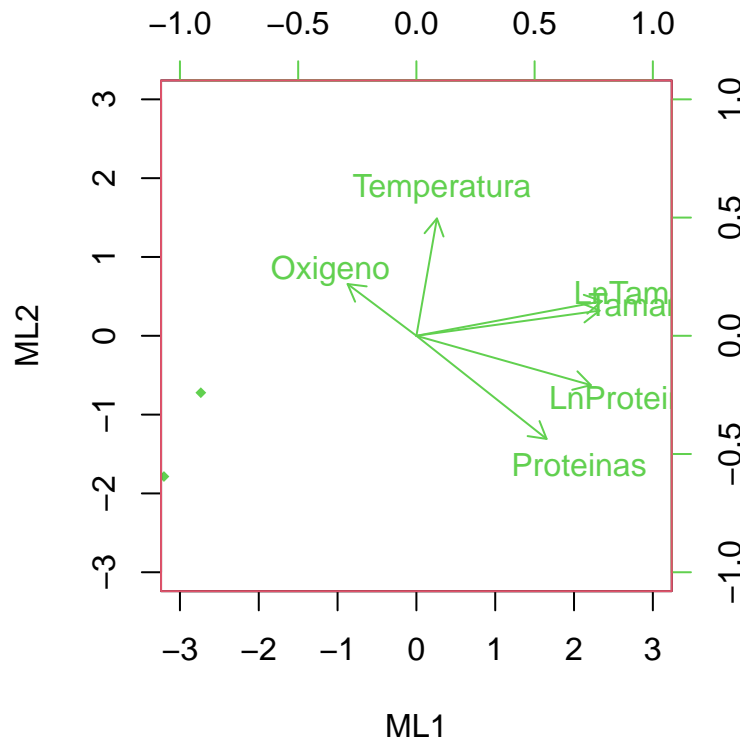
**7) Realicen los gráficos correspondientes a la rotación Varimax y quartimax de los datos e interpreten en equipo los resultados.**

```
library(GPARotation)
```

```
##
## Adjuntando el paquete: 'GPARotation'
## The following objects are masked from 'package:psych':
##
##     equamax, varimin
R = cor(df)
rot = c("none", "varimax", "quartimax")
bi_mod = function(tipo){
  biplot.psych(fa(df,nfactors = 2,fm="mle",rotate = tipo),main = "",col=c(2,3,4),pch = c(21,18),group = d
  sapply(rot,bi_mod)
```







```
## $none
## NULL
##
## $varimax
## NULL
##
## $quartimax
## NULL
```

La rotación nos permite observar de que manera se distribuye la varianza, esto pues hace que las relaciones entre variables y factores se observe mejor.

## 8) ¿Qué pueden concluir? ¿Resultó razonable para este caso el modelo de análisis factorial? Expliquen.

El análisis factorial es razonable para este conjunto de datos, a través del gráfico de Scree sabemos que, dos factores explican la mayor parte de la varianza. De igual manera sabemos según la prueba de kmo, que es posible aplicar el análisis de factores, sin embargo no es óptimo. Sin embargo vemos que es válido el análisis dada la prueba de Bartlett.