# Analisis factorial- Psicologia

Lino Oswaldo Sanchez

21/4/2022

## Paqueterías necesarías

```
library(psych)
library(polycor)
library(ggcorrplot)
```

## Extraccion de datos

Se encuentra dentro de la paquetería *psych* y es una base que esta conformada por datos de un estudio de psicología.

```
x <- bfi
```

### Exploracion de la matriz

```
dim(x)
```

```
## [1] 2800   28
```

Observamos la dimensión de la matriz esta esta conformada por 2800 observaciones y 28 variables

- Tipos de variables

```
str(x)
```

```
## 'data.frame':    2800 obs. of  28 variables:
##  $ A1       : int  2 2 5 4 2 6 2 4 4 2 ...
##  $ A2       : int  4 4 4 4 3 6 5 3 3 5 ...
##  $ A3       : int  3 5 5 6 3 5 5 1 6 6 ...
##  $ A4       : int  4 2 4 5 4 6 3 5 3 6 ...
##  $ A5       : int  4 5 4 5 5 5 5 1 3 5 ...
##  $ C1       : int  2 5 4 4 4 6 5 3 6 6 ...
##  $ C2       : int  3 4 5 4 4 6 4 2 6 5 ...
##  $ C3       : int  3 4 4 3 5 6 4 4 3 6 ...
```

```
##  $ C4       : int  4 3 2 5 3 1 2 2 4 2 ...
##  $ C5       : int  4 4 5 5 2 3 3 4 5 1 ...
##  $ E1       : int  3 1 2 5 2 2 4 3 5 2 ...
##  $ E2       : int  3 1 4 3 2 1 3 6 3 2 ...
##  $ E3       : int  3 6 4 4 5 6 4 4 NA 4 ...
##  $ E4       : int  4 4 4 4 4 5 5 2 4 5 ...
##  $ E5       : int  4 3 5 4 5 6 5 1 3 5 ...
##  $ N1       : int  3 3 4 2 2 3 1 6 5 5 ...
##  $ N2       : int  4 3 5 5 3 5 2 3 5 5 ...
##  $ N3       : int  2 3 4 2 4 2 2 2 2 5 ...
##  $ N4       : int  2 5 2 4 4 2 1 6 3 2 ...
##  $ N5       : int  3 5 3 1 3 3 1 4 3 4 ...
##  $ O1       : int  3 4 4 3 3 4 5 3 6 5 ...
##  $ O2       : int  6 2 2 3 3 3 2 2 6 1 ...
##  $ O3       : int  3 4 5 4 4 5 5 4 6 5 ...
##  $ O4       : int  4 3 5 3 3 6 6 5 6 5 ...
##  $ O5       : int  3 3 2 5 3 1 1 3 1 2 ...
##  $ gender   : int  1 2 2 2 1 2 1 1 1 2 ...
##  $ education: int  NA NA NA NA NA 3 NA 2 1 NA ...
##  $ age      : int  16 18 17 17 17 21 18 19 19 17 ...
```

**int** en R denota variables discretas - Nombre de las variables

```
colnames(x)
```

```
##  [1] "A1"        "A2"        "A3"        "A4"        "A5"        "C1"
##  [7] "C2"        "C3"        "C4"        "C5"        "E1"        "E2"
## [13] "E3"        "E4"        "E5"        "N1"        "N2"        "N3"
## [19] "N4"        "N5"        "O1"        "O2"        "O3"        "O4"
## [25] "O5"        "gender"    "education" "age"
```

- Creación de una nueva base de datos donde se incluyen las variables 1 a 25 y solo usamos 200 observaciones

```
x1<-bfi[1:200,1:25]
```

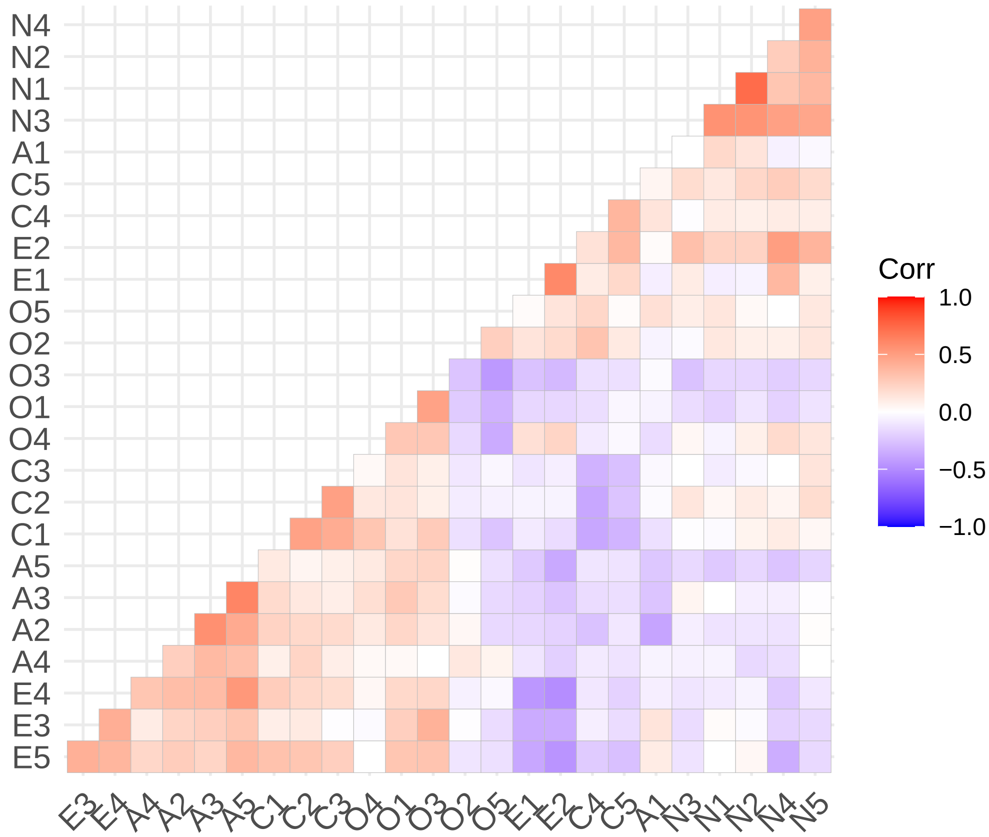
# Matriz de correlaciones

Se calcula la correlación entre las variables

```
R<- hetcor(x1)$correlations
```

- Gráfico de correlaciones

```
ggcorrplot(R,type="lower",hc.order= TRUE)
```




## Factorización de la matriz de correlaciones

Se utiliza la prueba de esfericidad de Bartlett.

```
prueba_Bartlett<- cortest.bartlett(R)
```

- Visualización de el p-valor

```
prueba_Bartlett$p.value
```

```
## [1] 5.931663e-60
```

H0 : variables correlacionadas H1 : las variables no están correlacionadas

No rechazo H0 con ayuda del p-valor

### Criterio Kaiser-Meyer-Olkin

Me permite identificar si los datos que voy a analizar son adecuados para un análisis factorial.

0.00 a 0.49 No adecuados 0.50 a 0.59 Poco adecuados 0.60 a 0.69 Aceptables 0.70 a 0.89 Buenos 0.90 a 1.00 Excelente

```
KMO(R)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = R)
## Overall MSA =  0.76
## MSA for each item =
##   A1   A2   A3   A4   A5   C1   C2   C3   C4   C5   E1   E2   E3   E4   E5   N1
## 0.66 0.77 0.69 0.73 0.75 0.74 0.79 0.76 0.76 0.74 0.80 0.81 0.79 0.81 0.83 0.70
##   N2   N3   N4   N5   O1   O2   O3   O4   O5
## 0.67 0.82 0.79 0.82 0.79 0.65 0.81 0.62 0.77
```

- Overall MSA = 0.76 son buenos para continuar con el análisis

# Extracción de factores

minres : minimo residuo mle : max verosimilitud pfa: ejesprincipales alpha: alfa minchi: minimos cuadrdos minrank: rango minimo

- modelo Varimax

```
modelo1<- fa(R,nfactor=3,rotate = "none",fm = "mle")
```

- modelo Mínimo Residuo

```
modelo2<- fa(R,nfactor=3,rotate = "none",fm = "minres")
```

Extraer el resultado de las Comunalidades, allí se encuentra la proporción de varianza explicada. Se interpreta de tal forma que números cercanos a 1 están bastante bien explicadas por los factores comunes, entre más Comunalidades altas aya en el factor este explica mejor la variable y el análisis en consecuencia será mejor.

```
C1<-sort(modelo1$communality,decreasing = TRUE)
```

```
C2<-sort(modelo2$communality,decreasing = TRUE)
```

combinar los resultados para comparar

```
head(cbind(C1,C2))
```

```
##           C1        C2
## N1 0.7576920 0.6809294
## E2 0.6802809 0.6564523
## N2 0.6797943 0.5866483
## E1 0.5219674 0.5394762
## N3 0.5198285 0.4942059
## N4 0.4839516 0.4744005
```

Extracción de unidades La unicidad es el cuadrado del coeficiente del factor único, y se expresa como la proporción de la varianza explicada por el factor único. es decir, no puede ser explicada por otros factores.

- Unicidad del modelo 1

```r
u1<- sort(modelo1$uniquenesses,decreasing = TRUE)
```

- Unicidad del modelo 2

```r
u2<- sort(modelo2$uniquenesses,decreasing = TRUE)
```

- Comparación

```r
head(cbind(u1,u2))
```

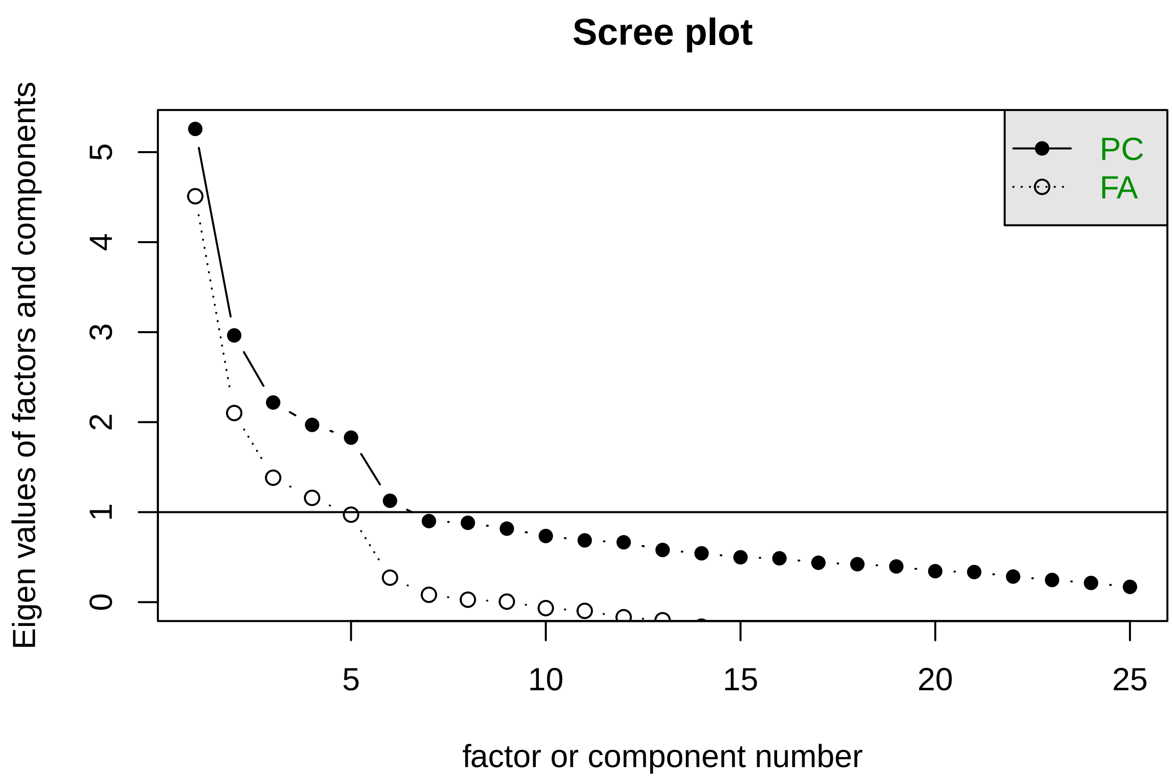
```
##           u1        u2
## O2 0.9460554 0.9293483
## A4 0.8928892 0.8908844
## A1 0.8607240 0.8822080
## O5 0.8533481 0.8272041
## C5 0.8136600 0.7931685
## O1 0.7986908 0.7904667
```

Al observar la unicidad de los dos métodos de rotación la variación entre ellas es muy pequeña no disciernen mucho una de la otra,teniendo en cuneta que la unicidad es el cuadrado del coeficiente del factor único, que expresa la proporción de la varianza que queda explicada por el factor único que también podemos decir de otro modo: la varianza que no puede explicarse por los factores comunes(Comunalidades).

Por lo que apreciamos ya sea que rotemos por Máxima Verosimilitud o Mínimo residuo la varianza que no se puede explicar es casi la misma los coeficientes son muy parecidos y solo queda en decisión del investigador cual es el método que más le conviene para llegar a una conclusión aunque los resultados por cual quiera de estos dos métodos serán muy aproximados.

- Para elegir el numero de los factores

```r
scree(R)
```




En este Scree plot podemos escoger la cantidad de factores que podemos utilizar, ademas de que en este gráfico tenemos dos opciones por las cuales podemos decantarnos; ya sea por el método de Principal Componentes(PC) y Factorial Analysis(FA). Como en los componentes principales la cantidad de factores que seleccionamos es hasta donde se forme un codo, esa sera la cantidad optima de factores que se puede utilizar.

Aquí observamos que en los dos métodos nos orillan a usar casi la misma cantidad de factores pues la gráfica de cada uno no discrepa en solo uno, dependiendo de la metodología que usemos sera la cantidad de factores que se seleccionaran, aun que como en este trabajo estamos usando el Factorial Analysis(FA) la cantidad seleccionada de factores será 5.

Todo esto claro con los datos originales sin ser rotados.

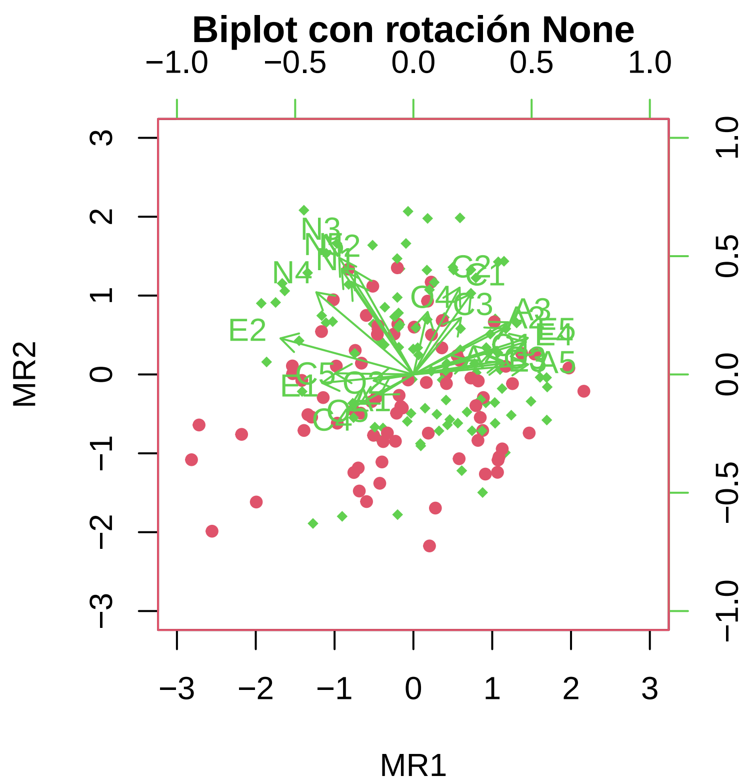
## Rotación de la matriz

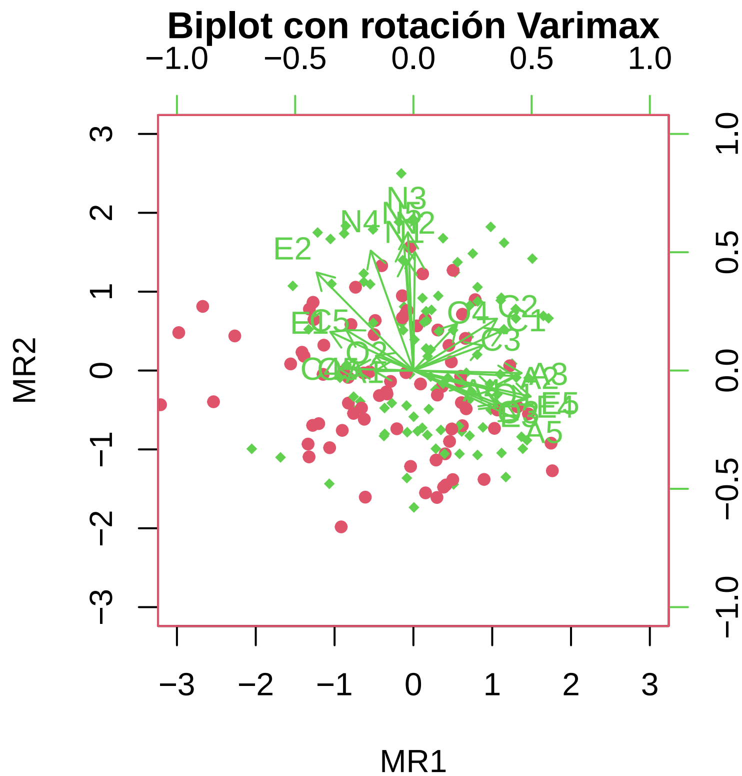
Para la rotación de la matriz usaremos un Baiplot con cuatro diferentes rotaciones, en el primer gráfico **no tendrá rotación**, para el segundo se usara la rotación **Varimax**, en el tercero usaremos la rotación **Quartimax** y en el curto la **Promax**, todos estos con el uso de 2 factores que le indicamos en la función que se ve descrita abajo al iguala que la extracción sera por el método de Mínimo Residuo.

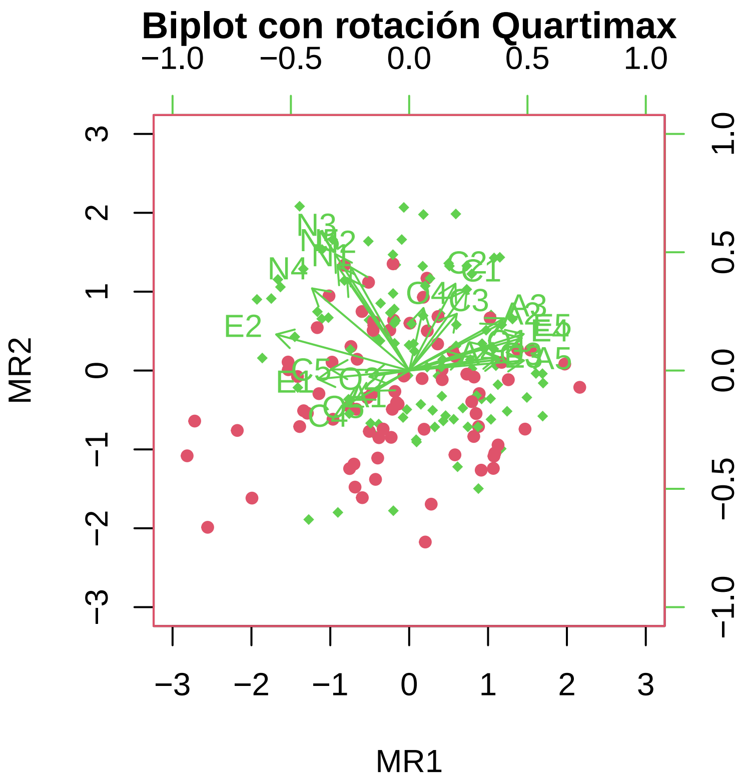
```
library(GPArotation)
```

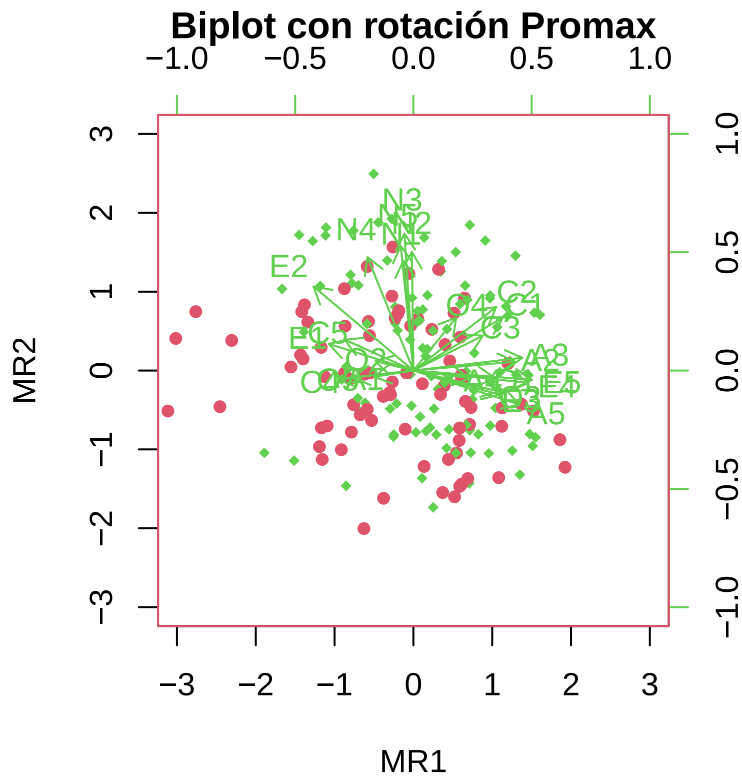
```
rot<-c("None", "Varimax", "Quartimax", "Promax")
bi_mod<-function(tipo){
  biplot.psych(fa(x1, nfactors = 2,
  fm= "minres", rotate=tipo),
```

```
  main = paste("Biplot con rotación", tipo),
  col=c(2,3,4), pch=c(21,18), group=bfi[,"gender"])
}
sapply(rot,bi_mod)
```




Biplot con rotación Varimax
MR1
MR2
−1.0
−0.5
0.0
0.5
1.0
−3
−2
−1
0
1
2
3


Biplot con rotación Quartimax
MR1
MR2
E2
N3
N4
C2
C1
C3
C4
A3
E5
E4
A5
C5
E1

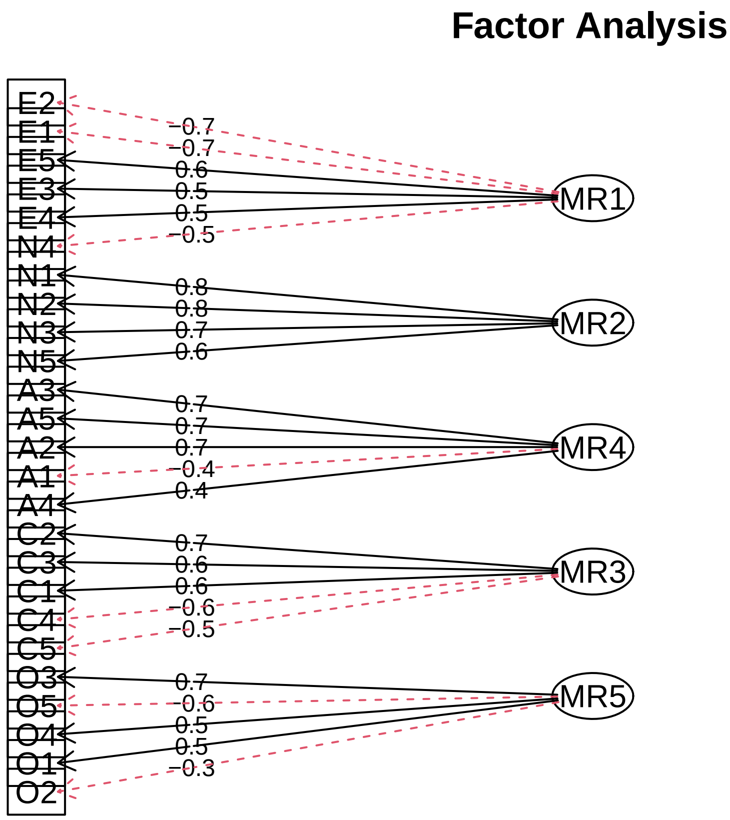



```
## $None
## NULL
##
## $Varimax
## NULL
##
## $Quartimax
## NULL
##
## $Promax
## NULL
```

## interpretación

Para esto utilizamos un gráfico de árbol

```
modelo_varimax<-fa(R,nfactor = 5,
                   rotate = "varimax",
                   fm="minres")
fa.diagram(modelo_varimax)
```




Donde usamos la matriz de correlaciones previa mente calculada con 5 factores que nos mostró la gráfica de Scree Plot y la rotamos mediante**Varimax** y el método de extracción de factores será por Mínimo Residuo.

Lineas rojas cargas positivas , lineas negras cargas negativas

## Visualización de la matriz de carga rotada

```
print(modelo_varimax$loadings,cut=0)
```

```
##
## Loadings:
##    MR1    MR2    MR4    MR3    MR5
## A1  0.234  0.106 -0.422 -0.072 -0.092
## A2  0.112 -0.032  0.653  0.190  0.113
## A3  0.198  0.066  0.744  0.051  0.169
## A4  0.163 -0.048  0.413  0.137 -0.142
## A5  0.328 -0.154  0.692 -0.009  0.115
## C1  0.054  0.089  0.140  0.634  0.287
## C2  0.052  0.174  0.114  0.690  0.050
## C3  0.032  0.018  0.076  0.642  0.016
## C4 -0.058  0.087 -0.090 -0.559 -0.159
## C5 -0.241  0.228 -0.040 -0.459  0.014
## E1 -0.691 -0.006 -0.066 -0.084 -0.017
## E2 -0.713  0.345 -0.138 -0.133 -0.025
```

```
## E3  0.546  0.003  0.157 -0.008  0.221
## E4  0.522 -0.027  0.416  0.167  0.048
## E5  0.588 -0.009  0.148  0.308  0.159
## N1  0.131  0.802 -0.150 -0.074 -0.133
## N2  0.088  0.800 -0.151 -0.038 -0.008
## N3 -0.183  0.701  0.005  0.037 -0.087
## N4 -0.513  0.491 -0.006  0.004  0.034
## N5 -0.274  0.571  0.059  0.096 -0.082
## O1  0.203 -0.107  0.148  0.076  0.535
## O2 -0.099  0.096  0.144 -0.191 -0.330
## O3  0.326 -0.159  0.034  0.062  0.680
## O4 -0.240  0.122  0.169  0.105  0.548
## O5 -0.004  0.061 -0.074 -0.077 -0.636
##
##                  MR1   MR2   MR4   MR3   MR5
## SS loadings    2.823 2.667 2.223 2.103 1.867
## Proportion Var 0.113 0.107 0.089 0.084 0.075
## Cumulative Var 0.113 0.220 0.309 0.393 0.467
```

Con esto podemos ver como están agrupados en el gráfico de árbol deben ser congruentes, ademas de obtener la varianza acumulada de los factor así como su varianza individual por cada factor.