Universidad Rey Juan Carlos

UNIVERSIDAD
REY JUAN CARLOS

INGENIERÍA DEL SOFTWARE Y MATEMÁTICAS

Curso Académico 2018/2019

PRÁCTICA 6 - ANÁLISIS FACTORIAL

Autores : Sandra Gómez Gálvez y Sergio Casado López

Capítulo 1

Análisis Factorial

Realizaremos un análisis factorial a los datos para encontrar grupos de variables con un significado común y así reducir el número de dimensiones para explicar las respuestas de los sujetos. Para la realización de esta práctica utilizaremos el archivo *cereales2.txt* correspondientes a *Datos práctica parcial*

Cereales2SinCambiar: Bloc de notas

Archivo Edición Formato Ver Ayuda

```
Filling Natural Fibre Sweet Easy Salt Energy Fun Kids Soggy
5.00 5.00 5.00 1.00 2.00 1.00 4.00 1.00 4.00 5.00
1.00 2.00 2.00 1.00 5.00 2.00 1.00 1.00 5.00 3.00
5.00 4.00 5.00 5.00 5.00 3.00 5.00 5.00 5.00 3.00
5.00 5.00 5.00 3.00 5.00 2.00 5.00 5.00 5.00 3.00
4.00 5.00 3.00 2.00 5.00 2.00 4.00 5.00 5.00 1.00
4.00 4.00 4.00 2.00 5.00 2.00 4.00 5.00 5.00 1.00
4.00 4.00 3.00 2.00 5.00 1.00 5.00 5.00 5.00 1.00
4.00 3.00 3.00 2.00 5.00 1.00 4.00 4.00 5.00 1.00
4.00 3.00 3.00 2.00 5.00 1.00 4.00 4.00 5.00 1.00
4.00 3.00 3.00 2.00 5.00 1.00 4.00 4.00 5.00 1.00
```

Figura 1.1: cereales2.txt

Para ello, en R, debemos importar la librería psych:

```r
# Ejecutamos la libreria que tiene que estar instalada
library(psych)
```

Figura 1.2: Libreria psych

Cogeremos el fichero con los datos:

```
# Cogemos el fichero Datos1.txt
M <-read.table("C:/Users/sandr/Documents/Geometria Computancional/PRACTICA6/Cereales2SinCambiar.txt",header=T,sep="")
```

Figura 1.3: Fichero cereales2.txt

Sacamos su **determinante**:

```
det(data.matrix(M)) #Como su determinante es 0 la matriz no es invertible
```

Figura 1.4: Determinante

Al hacer run al archivo en R, observamos como su determinante es 0.

```
> det(data.matrix(M)) #Como su determinante es 0 la matriz no es invertible
[1] 0
```

Figura 1.5: Determinante=0

Esto implica que la matriz no será invertible, por ello, cambiaremos los datos en *cereales2.txt* para que el determinante sea distinto de cero. Para ello, quitaremos las 2 últimas filas y las 2 últimas columnas ya que las 3 últimas filas son iguales entre ellas.

```
Cereales2: Bloc de notas
Archivo  Edición  Formato  Ver  Ayuda
Filling Natural Fibre Sweet Easy Salt Energy Fun
5.00 5.00 5.00 1.00 2.00 1.00 4.00 1.00
1.00 2.00 2.00 1.00 5.00 2.00 1.00 1.00
5.00 4.00 5.00 5.00 5.00 3.00 5.00 5.00
5.00 5.00 5.00 3.00 5.00 2.00 5.00 5.00
4.00 5.00 3.00 2.00 5.00 2.00 4.00 5.00
4.00 4.00 4.00 2.00 5.00 2.00 4.00 5.00
4.00 4.00 3.00 2.00 5.00 1.00 5.00 5.00
4.00 3.00 3.00 2.00 5.00 1.00 4.00 4.00
```

Figura 1.6: Nuevo cereales2.txt

Por tanto, al hacer run al archivo R, nos sale determinante=216

```
> det(data.matrix(M))
[1] 216
```

Figura 1.7: Determinante=216

Con el determinante ya distinto de 0, comenzaremos realizando el **Test de esfericidad de Bartlett**, con ello veremos si la matriz de correlación de las p variables observadas es la identidad. Contrastaremos si la matriz de correlación es la identidad

```
# Realizamos el Test de esfericidad de Bartlett
bartlett.test(M)
#Por tanto, podemos realizar el análisis factorial

```

Figura 1.8: Test de esfericidad de Barlett

Que resulta en:

```
> bartlett.test(M)

        Bartlett test of homogeneity of variances

data:  M
Bartlett's K-squared = 6.1477, df = 7, p-value = 0.5226

> #Por tanto, podemos realizar el análisis factorial
```

Figura 1.9: Resultados Test de esfericidad de Barlett

Desde la salida podemos ver que el valor de p de 0.5226 no es menor que el nivel de significación de 0.05. Esto significa que no podemos rechazar la hipótesis nula de que la varianza es la misma para todos los tipos de cereales.

Por tanto, podemos continuar con el análisis factorial. **Analizaremos los componentes principales sin rotar y los introduciremos en la variable** *Modelo1*:

```
# princomp( ): análisis de componentes principales sin rotar
Modelo1 <- princomp(M, cor=TRUE)
```

Figura 1.10: Ánalisis de componentes principales

Realizaremos la **varianza** de cada componente:

```
# Realizamos la varianza de cada factor
summary(Modelo1 )
```

Figura 1.11: summary

Que resultará:

```
> # Realizamos la Varianza de cada factor
> summary(Modelo1 )
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5      Comp.6      Comp.7 Comp.8
Standard deviation     2.0228057 1.5274408 1.0348143 0.60954864 0.29320518 0.170048364 0.133812038      0
Proportion of Variance 0.5114679 0.2916344 0.1338551 0.04644369 0.01074616 0.003614556 0.002238208      0
Cumulative Proportion  0.5114679 0.8031023 0.9369574 0.98340108 0.99414724 0.997761792 1.000000000      1
```

Figura 1.12: Varianza de cada componente

Observamos que estamos trabajando con 8 dimensiones, queremos saber si podemos trabajar con meno, ¿con cuántas menos?

Para ello, le **daremos puntuaciones a las componentes y crearemos un gráfico de sedimentación:**

```
# Estamos trabajando con 8 dimensiones, pero podemos trabajar con menos ¿pero cuántas menos?
    ##Puntuaciones de las componentes y gráfico de sedimentación:
# Puntuaciones en los factores
loadings(Modelo1)
# Gráfico de sedimentación
plot(Modelo1,type="lines")
```

Figura 1.13: Puntuaciones y gráfico

Las puntuaciones y gráfico serán:

```
> loadings(Modelo1)

Loadings:
        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
Filling  0.461  0.207         0.172         0.393  0.674  0.317
Natural  0.384  0.278        -0.734  0.207 -0.385  0.119 -0.155
Fibre    0.405  0.217 -0.392  0.149 -0.733 -0.178 -0.126 -0.169
Sweet    0.360 -0.360 -0.280  0.406  0.468 -0.269  0.131 -0.435
Easy           -0.618  0.283        -0.363 -0.462  0.349  0.257
Salt     0.148 -0.432 -0.622 -0.391         0.285 -0.149  0.377
Energy   0.459         0.309  0.240  0.173 -0.186 -0.559  0.508
Fun      0.336 -0.365  0.439 -0.172 -0.181  0.515 -0.206 -0.438

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.125  0.125  0.125  0.125  0.125  0.125  0.125  0.125
Cumulative Var  0.125  0.250  0.375  0.500  0.625  0.750  0.875  1.000
```

Figura 1.14: Puntuaciones

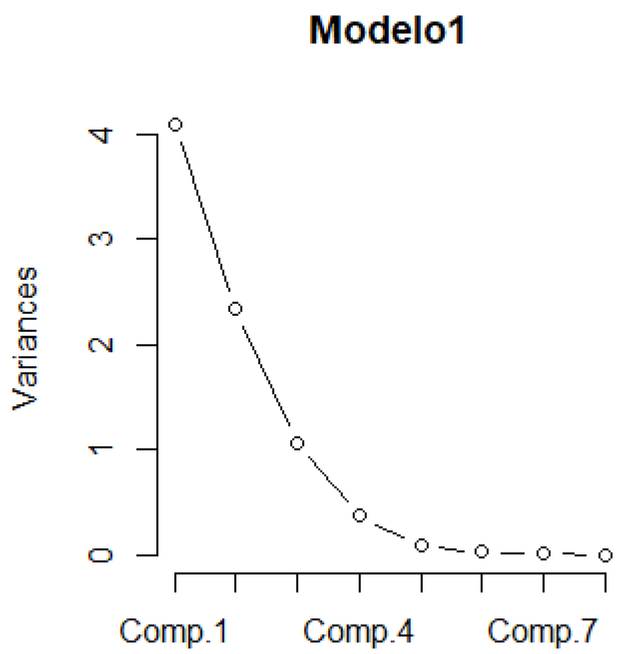

Figura 1.15: Gráfico

Estableceremos el número de factores en 2. Y utilizaremos una rotación *varimax*. Podríamos haber establecido la rotación como *none, varimax,quartimax,promax,oblimin,simplimax, o cluster.* Pero elegimos *varimax* porque así nos ajustaremos lo máximo posible a la realidad.
Por tanto, ahora, usando la libreria *GPArotation* obtendremos un *Modelo2* con número de factores =2 y rotación varimax.

```
library(GPArotation)
Modelo2 <- principal(M, nfactors=2, rotate="varimax")
# Imprime los resultados
Modelo2
```

Figura 1.16: Factores=2 y Varimax

Que resultará en:

```
> # Imprime los resultados
> Modelo2
Principal Components Analysis
Call: principal(r = M, nfactors = 2, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
         RC1   RC2   h2    u2 com
Filling  0.98  0.10 0.97 0.029 1.0
Natural  0.88 -0.06 0.78 0.216 1.0
Fibre    0.88  0.04 0.78 0.220 1.0
Sweet    0.43  0.80 0.83 0.166 1.5
Easy    -0.42  0.85 0.89 0.107 1.5
Salt     0.00  0.72 0.53 0.475 1.0
Energy   0.87  0.33 0.87 0.133 1.3
Fun      0.39  0.79 0.77 0.229 1.5

                       RC1  RC2
SS loadings           3.79 2.64
Proportion Var        0.47 0.33
Cumulative Var        0.47 0.80
Proportion Explained  0.59 0.41
Cumulative Proportion 0.59 1.00

Mean item complexity =  1.2
Test of the hypothesis that 2 components are sufficient.

The root mean square of the residuals (RMSR) is  0.12
 with the empirical chi square  6.9  with prob <  0.91

Fit based upon off diagonal values = 0.94
```

Figura 1.17: Modelo2

Además, podemos ver en que nos dice que la hipótesis de 2 factores es suficiente, por tanto es correcta.

Volvemos a realizar la **varianza de cada factor, las puntuaciones en los factores, las puntuaciones en cada caso, y un biplot para el *Modelo2*:**

```
# Varianza de cada factor
summary(Modelo2)
# Puntuaciones en los factores
loadings(Modelo2)
# Puntuaciones de los casos
Modelo2 $scores
biplot(Modelo2)

```

Figura 1.18: Varianza, puntuaciones factores, puntuaciones cada caso y Biplot

Que resultan:

```
> summary(Modelo2)

Factor analysis with Call: principal(r = M, nfactors = 2, rotate = "varimax")

Test of the hypothesis that 2 factors are sufficient.
The degrees of freedom for the model is 13  and the objective function was  24.55
The number of observations was  8  with Chi Square =  53.2  with prob <  8.3e-07

The root mean square of the residuals (RMSA) is  0.12
```

Figura 1.19: Resultado Varianza

```
> # Puntuaciones en los factores
> loadings(Modelo2)

Loadings:
        RC1    RC2   
Filling  0.980  0.101
Natural  0.883       
Fibre    0.882       
Sweet    0.434  0.803
Easy    -0.418  0.847
Salt            0.725
Energy   0.870  0.333
Fun      0.385  0.789

                 RC1   RC2
SS loadings    3.787 2.638
Proportion Var 0.473 0.330
Cumulative Var 0.473 0.803
```

Figura 1.20: Puntuación en los factores

```
> # Puntuaciones de los casos
> Modelo2 $scores
            RC1         RC2
[1,]  2.5223393 -4.84277477
[2,] -8.4004076 -2.42105088
[3,]  3.3789131  4.16929439
[4,]  3.5306794  1.83305259
[5,]  0.2645172  0.80707572
[6,]  0.1958479  0.89881954
[7,]  0.1057380  0.09437238
[8,] -1.5976274 -0.53878896
```

Figura 1.21: Puntuación de los casos

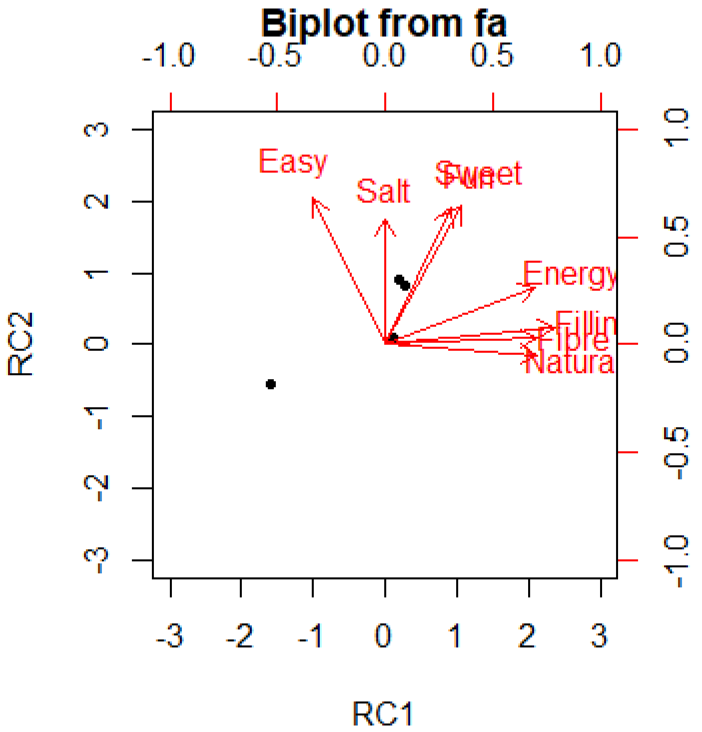

Figura 1.22: Biplot Modelo2

Por tanto, hemos podido reducir el Modelo a 2 casos funcionales.

Ahora, queremos obtener el **análisis factorial basado en máxima verosimilitud.**

Para ello, realizaremos la función factanal con 4 factores y rotación varimax. No la podríamos realizar con más de 4 factores pues excedería.

```
# Factanal produce el análisis factorial basado en máxima verosimilitud
Modelo3<- factanal(M,4, rotation="varimax")
```

Figura 1.23: Función factanal

Al hacer run sobre esta función encontraremos un error:

```
> Modelo3<- factanal(M,4, rotation="varimax")
Error in solve.default(cv) :
  system is computationally singular: reciprocal condition number = 2.97868e-17
```

Figura 1.24: Error Función factanal

Esto indica que la matriz tiene que ser semidefinida positiva, para ello tendremos que cambiar/quitar algunas variables como hicimos anteriormente para que el determinante fuese distinto de 0.

Tras varios intentos de cambios, esto no ha sido conseguido. En el caso de ser conseguido, los pasos siguientes serían:

```
print(Modelo3, digits=2, cutoff=.3, sort=TRUE)
# plot factor 1 y factor 2
load <- Modelo3$loadings[,1:2]
plot(load,type="n") # Establecemos el plot
text(load,labels=names(M),cex=.7) # añadimos los nombres de las variables
```

Figura 1.25: Pasos siguientes a la Función factanal

A pesar de no poder realizar el análisis factorial basado en máxima verosimilitud, hemos realizado un buen análisis factorial reduciendo todo a solamente 2 casos.