

# T4. Taller encuestas. Análisis factorial AF\*

VIII Jornadas R Albacete 2016\*\*

Álvaro Hernández Vicente, Elvira Ferre Jaén, Antonio José Perán Orcajada, Ana Belén  
Marín Valverde, Antonio Maurandi López\*\*\*

17 de noviembre de 2016

## Índice

<b>1. Lectura de datos</b>	<b>1</b>
<b>2. Comprobaciones previas</b>	<b>1</b>
2.1. Matriz de correlaciones . . . . .	2
2.2. Test de Bartlett . . . . .	3
2.3. Coeficiente Kaiser-Meyer-Olkin (KMO) . . . . .	4
2.4. Determinante de la matriz de correlaciones . . . . .	4
<b>3. Análisis de Componentes Principales</b>	<b>4</b>
3.1. Determinación del número de componentes principales . . . . .	4
3.2. Análisis . . . . .	6
3.3. Medidas de ajuste del modelo . . . . .	7
3.4. Rotación . . . . .	8
3.5. Interpretación . . . . .	10
3.6. Cargas de individuos . . . . .	11
3.7. Resumen . . . . .	11
<b>4. Análisis de Fiabilidad</b>	<b>12</b>
4.1. $\alpha$ de Cronbach para el Miedo a los computadores . . . . .	12
4.2. Alfa de Cronbach para el Miedo a la estadística . . . . .	13
4.3. $\alpha$ de Cronbach para el Miedo a las matemáticas . . . . .	14
4.4. $\alpha$ de Cronbach para el Evaluación por pares . . . . .	15
<b>Referencias y bibliografía</b>	<b>16</b>

## 1. Lectura de datos

Para llevar a cabo todo el procedimiento que se realiza en este documento seguimos el capítulo 17: Exploratory factor analysis de *Discovering Statistics Using R* (A. Field, Miles, and Field 2012).

```
load( "saeraq.RData" )
```

## 2. Comprobaciones previas

Es aconsejable realizar las siguientes comprobaciones antes de proceder con el análisis factorial.

---

\* doc:T4\_analisisAF.Rmd

\*\* <http://r-es.org/8jornadasR/>

\*\*\* Servicio de Apoyo Estadístico; alvarohv@um.es, elvira@um.es, antoniojose.peran@um.es, anabelen.marin4@um.es, amaurandi@um.es



- Observar la matriz de correlaciones: en busca de variables poco o demasiado correlacionadas.
- Test de Bartlett
- Coeficiente Kaiser-Meyer-Olkin (KMO)
- Determinante de la matriz de correlaciones.

## 2.1. Matriz de correlaciones

Podemos trabajar indistintamente con el conjunto de datos o con la matriz de correlaciones, pero en cualquier caso, siempre con variables numéricas, no con factores. Se aconseja trabajar con la matriz de correlaciones, ya que así evitamos recalcular dicha matriz cada vez que llamemos a alguna función del análisis.

En nuestro caso, calculamos la matriz de correlaciones:

```
corrMatrix <- cor(df)
round(corrMatrix, 3)
```

```
##      Q01  Q02  Q03  Q04  Q05  Q06  Q07  Q08  Q09  Q10
## Q01  1.000 -0.099 -0.337  0.436  0.402  0.217  0.305  0.331 -0.092  0.214
## Q02 -0.099  1.000  0.318 -0.112 -0.119 -0.074 -0.159 -0.050  0.315 -0.084
## Q03 -0.337  0.318  1.000 -0.380 -0.310 -0.227 -0.382 -0.259  0.300 -0.193
## Q04  0.436 -0.112 -0.380  1.000  0.401  0.278  0.409  0.349 -0.125  0.216
## Q05  0.402 -0.119 -0.310  0.401  1.000  0.257  0.339  0.269 -0.096  0.258
## Q06  0.217 -0.074 -0.227  0.278  0.257  1.000  0.514  0.223 -0.113  0.322
## Q07  0.305 -0.159 -0.382  0.409  0.339  0.514  1.000  0.297 -0.128  0.284
## Q08  0.331 -0.050 -0.259  0.349  0.269  0.223  0.297  1.000  0.016  0.159
## Q09 -0.092  0.315  0.300 -0.125 -0.096 -0.113 -0.128  0.016  1.000 -0.134
## Q10  0.214 -0.084 -0.193  0.216  0.258  0.322  0.284  0.159 -0.134  1.000
## Q11  0.357 -0.144 -0.351  0.369  0.298  0.328  0.345  0.629 -0.116  0.271
## Q12  0.345 -0.195 -0.410  0.442  0.347  0.313  0.423  0.252 -0.167  0.246
## Q13  0.355 -0.143 -0.318  0.344  0.302  0.466  0.442  0.314 -0.167  0.302
## Q14  0.338 -0.165 -0.371  0.351  0.315  0.402  0.441  0.281 -0.122  0.255
## Q15  0.246 -0.165 -0.312  0.334  0.261  0.360  0.391  0.300 -0.187  0.295
## Q16  0.499 -0.168 -0.419  0.416  0.395  0.244  0.389  0.321 -0.189  0.291
## Q17  0.371 -0.087 -0.327  0.383  0.310  0.282  0.391  0.590 -0.037  0.218
## Q18  0.347 -0.164 -0.375  0.382  0.322  0.513  0.501  0.280 -0.150  0.293
## Q19 -0.189  0.203  0.342 -0.186 -0.165 -0.167 -0.269 -0.159  0.249 -0.127
## Q20  0.214 -0.202 -0.325  0.243  0.200  0.101  0.221  0.175 -0.159  0.084
## Q21  0.329 -0.205 -0.417  0.410  0.335  0.272  0.483  0.296 -0.136  0.193
## Q22 -0.104  0.231  0.204 -0.098 -0.133 -0.165 -0.168 -0.079  0.257 -0.131
## Q23 -0.004  0.100  0.150 -0.034 -0.042 -0.069 -0.070 -0.050  0.171 -0.062
##      Q11  Q12  Q13  Q14  Q15  Q16  Q17  Q18  Q19  Q20
## Q01  0.357  0.345  0.355  0.338  0.246  0.499  0.371  0.347 -0.189  0.214
## Q02 -0.144 -0.195 -0.143 -0.165 -0.165 -0.168 -0.087 -0.164  0.203 -0.202
## Q03 -0.351 -0.410 -0.318 -0.371 -0.312 -0.419 -0.327 -0.375  0.342 -0.325
## Q04  0.369  0.442  0.344  0.351  0.334  0.416  0.383  0.382 -0.186  0.243
## Q05  0.298  0.347  0.302  0.315  0.261  0.395  0.310  0.322 -0.165  0.200
## Q06  0.328  0.313  0.466  0.402  0.360  0.244  0.282  0.513 -0.167  0.101
## Q07  0.345  0.423  0.442  0.441  0.391  0.389  0.391  0.501 -0.269  0.221
## Q08  0.629  0.252  0.314  0.281  0.300  0.321  0.590  0.280 -0.159  0.175
## Q09 -0.116 -0.167 -0.167 -0.122 -0.187 -0.189 -0.037 -0.150  0.249 -0.159
## Q10  0.271  0.246  0.302  0.255  0.295  0.291  0.218  0.293 -0.127  0.084
## Q11  1.000  0.335  0.423  0.325  0.365  0.369  0.587  0.373 -0.200  0.255
## Q12  0.335  1.000  0.489  0.433  0.332  0.408  0.333  0.493 -0.267  0.298
## Q13  0.423  0.489  1.000  0.450  0.342  0.358  0.408  0.533 -0.227  0.204
## Q14  0.325  0.433  0.450  1.000  0.380  0.418  0.354  0.498 -0.254  0.226
## Q15  0.365  0.332  0.342  0.380  1.000  0.454  0.373  0.343 -0.210  0.206
## Q16  0.369  0.408  0.358  0.418  0.454  1.000  0.410  0.422 -0.267  0.265
## Q17  0.587  0.333  0.408  0.354  0.373  0.410  1.000  0.376 -0.163  0.205
## Q18  0.373  0.493  0.533  0.498  0.343  0.422  0.376  1.000 -0.257  0.235
## Q19 -0.200 -0.267 -0.227 -0.254 -0.210 -0.267 -0.163 -0.257  1.000 -0.249
## Q20  0.255  0.298  0.204  0.226  0.206  0.265  0.205  0.235 -0.249  1.000
## Q21  0.346  0.441  0.374  0.399  0.300  0.421  0.363  0.430 -0.275  0.468
## Q22 -0.162 -0.167 -0.195 -0.170 -0.168 -0.156 -0.126 -0.160  0.234 -0.100
## Q23 -0.086 -0.046 -0.053 -0.048 -0.062 -0.082 -0.092 -0.080  0.122 -0.035
##      Q21  Q22  Q23
```



```
## Q01 0.329 -0.104 -0.004
## Q02 -0.205 0.231 0.100
## Q03 -0.417 0.204 0.150
## Q04 0.410 -0.098 -0.034
## Q05 0.335 -0.133 -0.042
## Q06 0.272 -0.165 -0.069
## Q07 0.483 -0.168 -0.070
## Q08 0.296 -0.079 -0.050
## Q09 -0.136 0.257 0.171
## Q10 0.193 -0.131 -0.062
## Q11 0.346 -0.162 -0.086
## Q12 0.441 -0.167 -0.046
## Q13 0.374 -0.195 -0.053
## Q14 0.399 -0.170 -0.048
## Q15 0.300 -0.168 -0.062
## Q16 0.421 -0.156 -0.082
## Q17 0.363 -0.126 -0.092
## Q18 0.430 -0.160 -0.080
## Q19 -0.275 0.234 0.122
## Q20 0.468 -0.100 -0.035
## Q21 1.000 -0.129 -0.068
## Q22 -0.129 1.000 0.230
## Q23 -0.068 0.230 1.000
```

Solo con esta matriz ya podemos extraer información importante de los datos que nos servirá para realizar el análisis correctamente. El propósito principal del **Análisis Factorial** y del **Análisis de Componentes Principales** es el de encontrar constructos formados por variables que correlacionan bien entre ellas, pero no perfectamente.

Por tanto, antes de llevar a cabo el análisis, se recomienda examinar la matriz de correlaciones en busca de variables que no correlacionen bien con ninguna otra, es decir, con coeficientes de correlación todos menor que 0.3, y variables que correlacionen demasiado bien con otras, es decir, variables que tengan algún coeficiente de correlación mayor que 0.9. Las primeras deberíamos eliminarlas del análisis, y las segundas podremos mantenerlas, pero teniendo en cuenta que quizá causen problemas de multicolinealidad.

```
cat("Nº de variables que no correlacionan bien: ",
    sum( sapply(as.data.frame(corrMatrix), function(x) all(x < 0.3)) ) )
```

```
## Nº de variables que no correlacionan bien: 0
```

```
cat("Nº de variables que podrían causar multicolinealidad: ",
    sum( sapply(as.data.frame(corrMatrix), function(x) any(x >= 0.9 & x != 1)) ) )
```

```
## Nº de variables que podrían causar multicolinealidad: 0
```

En nuestro caso, no tenemos problemas de este tipo.

## 2.2. Test de Bartlett

Buscamos que el test de Bartlett salga significativo, es decir, que nuestra matriz no sea similar a una matriz identidad.

```
library(psych)
cortest.bartlett(corrMatrix, nrow(df))
```

```
## $chisq
## [1] 19334.49
##
## $p.value
## [1] 0
##
```



```
## $df
## [1] 253
```

### 2.3. Coeficiente Kaiser-Meyer-Olkin (KMO)

Ver: <http://tolstoy.newcastle.edu.au/R/e2/help/07/08/22816.html>

- Es una medida de la comparación de los coeficientes de correlación observados con los coeficientes de correlación parcial que asume valores entre 0 y 1.
- No deben aceptarse valores menores a 0.5 según Kaiser(1974); si el valor está entre 0.5 y 0.7 se considera mediocre, si está entre 0.7 y 0.9 se considera bueno, y si es mayor que 0.9 se considera muy bueno, según Hutcheson & Sofroniou (1999).
- El KMO puede calcularse de forma global o para cada una de las variables. Deberemos considerar eliminar las variables con KMO individual inferior a 0.5, y una vez sustraídas, recalcarlo.

```
kmo(df)[1:3]
```

```
## $overall
## [1] 0.9302245
##
## $report
## [1] "The KMO test yields a degree of common variance marvelous."
##
## $individual
##      Q01      Q02      Q03      Q04      Q05      Q06      Q07
## 0.9297610 0.8747754 0.9510378 0.9553403 0.9600892 0.8913314 0.9416800
##      Q08      Q09      Q10      Q11      Q12      Q13      Q14
## 0.8713055 0.8337295 0.9486858 0.9059338 0.9548324 0.9482270 0.9671722
##      Q15      Q16      Q17      Q18      Q19      Q20      Q21
## 0.9404402 0.9336439 0.9306205 0.9479508 0.9407021 0.8890514 0.9293369
##      Q22      Q23
## 0.8784508 0.7663994
```

En nuestro caso, el KMO es muy bueno, y todas las variables tienen un KMO individual considerablemente alto.

### 2.4. Determinante de la matriz de correlaciones

Para poder realizar el análisis factorial, la matriz de correlaciones ha de ser definida positiva, es decir, al menos su determinante debe ser mayor que 0. Se recomienda que sea mayor que 0.00001 (Andy Field?)

```
cat("Valor del Determinante: ", det(corrMatrix))
```

```
## Valor del Determinante: 0.0005271037
```

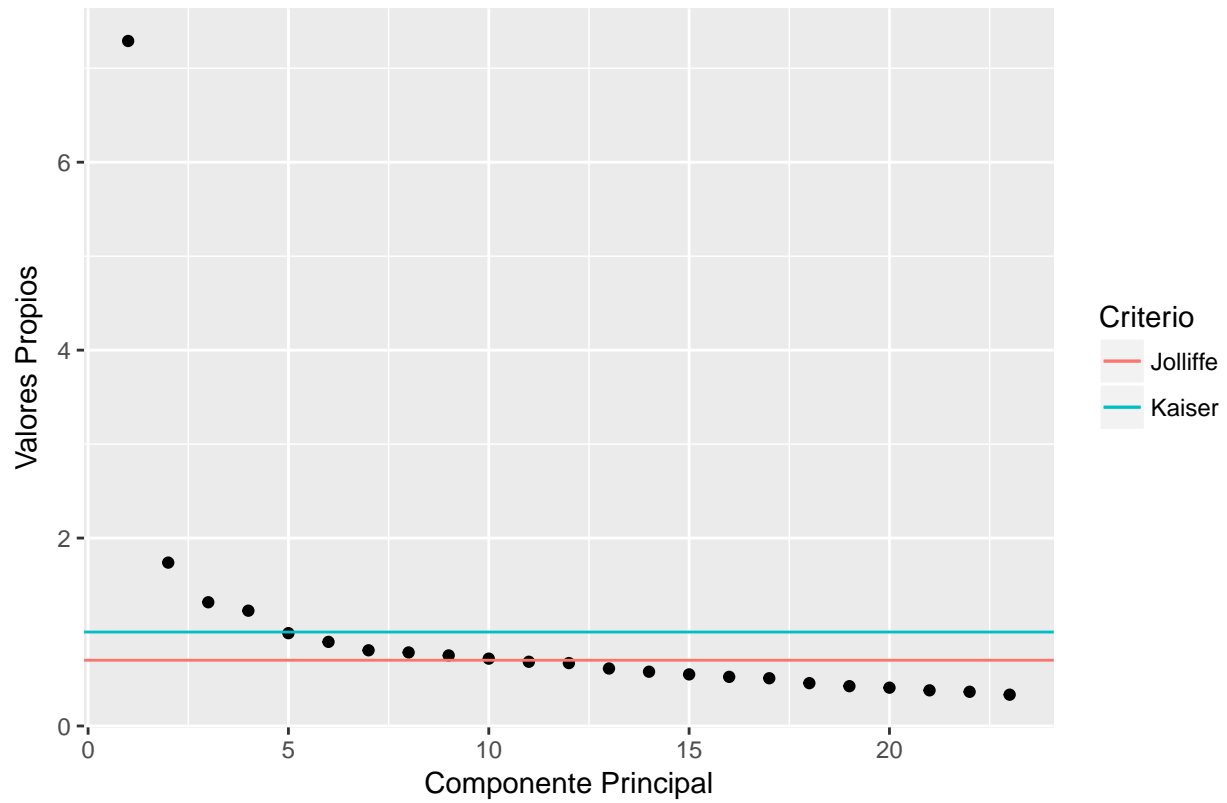
## 3. Análisis de Componentes Principales

### 3.1. Determinación del número de componentes principales

Para determinar el número de componentes principales a mantener, realizamos un análisis *a priori* en el que el número de componentes sea igual al número de variables, y elegimos el número de componentes a extraer de acuerdo a alguno de los siguientes criterios.

- Criterio de Kaiser: en azul, escogeríamos tantas componentes como valores propios superen el valor 1.
- Criterio de Jolliffe: en rojo, las que superen el 0.7.

### Scree Plot



En nuestro caso, retendremos 4 componentes principales de acuerdo al Criterio de Kaiser.



### 3.2. Análisis

```
pc1 <- principal(corrMatrix, nfactors = 4, rotate = "none")
pc1

## Principal Components Analysis
## Call: principal(r = corrMatrix, nfactors = 4, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      PC1  PC2  PC3  PC4  h2  u2 com
## Q01  0.59  0.18 -0.22  0.12  0.43  0.57  1.6
## Q02 -0.30  0.55  0.15  0.01  0.41  0.59  1.7
## Q03 -0.63  0.29  0.21 -0.07  0.53  0.47  1.7
## Q04  0.63  0.14 -0.15  0.15  0.47  0.53  1.3
## Q05  0.56  0.10 -0.07  0.14  0.34  0.66  1.2
## Q06  0.56  0.10  0.57 -0.05  0.65  0.35  2.1
## Q07  0.69  0.04  0.25  0.10  0.55  0.45  1.3
## Q08  0.55  0.40 -0.32 -0.42  0.74  0.26  3.5
## Q09 -0.28  0.63 -0.01  0.10  0.48  0.52  1.5
## Q10  0.44  0.03  0.36 -0.10  0.33  0.67  2.1
## Q11  0.65  0.25 -0.21 -0.40  0.69  0.31  2.2
## Q12  0.67 -0.05  0.05  0.25  0.51  0.49  1.3
## Q13  0.67  0.08  0.28 -0.01  0.54  0.46  1.4
## Q14  0.66  0.02  0.20  0.14  0.49  0.51  1.3
## Q15  0.59  0.01  0.12 -0.11  0.38  0.62  1.2
## Q16  0.68  0.01 -0.14  0.08  0.49  0.51  1.1
## Q17  0.64  0.33 -0.21 -0.34  0.68  0.32  2.4
## Q18  0.70  0.03  0.30  0.13  0.60  0.40  1.4
## Q19 -0.43  0.39  0.10 -0.01  0.34  0.66  2.1
## Q20  0.44 -0.21 -0.40  0.30  0.48  0.52  3.2
## Q21  0.66 -0.06 -0.19  0.28  0.55  0.45  1.6
## Q22 -0.30  0.47 -0.12  0.38  0.46  0.54  2.8
## Q23 -0.14  0.37 -0.02  0.51  0.41  0.59  2.0
##
##
##          PC1  PC2  PC3  PC4
## SS loadings      7.29  1.74  1.32  1.23
## Proportion Var   0.32  0.08  0.06  0.05
## Cumulative Var   0.32  0.39  0.45  0.50
## Proportion Explained 0.63  0.15  0.11  0.11
## Cumulative Proportion 0.63  0.78  0.89  1.00
##
## Mean item complexity = 1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.06
##
## Fit based upon off diagonal values = 0.96
```

En la primera matriz que muestra el resultado anterior tenemos, en las distintas columnas:

- Las cargas de cada una de las componentes en cada una de las variables (PC1, PC2, PC3, PC4).
- Las comunales (h<sup>2</sup>), que se interpreta como el porcentaje de varianza de cada variable que explican las componentes seleccionadas.
- La varianza única o uniquenesses (u<sup>2</sup>), que se interpreta como lo que queda sin explicar.
- Los índices de complejidad de Hoffman (com).



Hay una serie de pautas que pueden ayudarnos a la hora de comprobar que nuestra elección del *Criterio de Kaiser* para la extracción de factores ha sido buena. Por ejemplo, si el tamaño muestral es menor de 30 y las communalidades son todas mayores a 0.7; o cuando el tamaño muestral es superior a 250 y la media de las communalidades es superior a 0.6.

En nuestro caso, dicha media es 0.5031663, próxima a 0.6, luego podemos considerarla buena debido al gran tamaño de nuestra muestra: 2571 individuos.

### 3.3. Medidas de ajuste del modelo

Las medidas de ajuste del modelo se basan en estimar de distintas maneras el tamaño de los residuos que resultan al comparar la matriz de correlaciones original con la matriz de correlaciones reproducida por el modelo.

Lo ideal sería que la matriz de correlaciones reproducida por el modelo fuese lo más parecida posible a la calculada con los datos.

Una buena medida de ajuste viene dada por la función `principal`,

```
## Fit based upon off diagonal values = 0.9645252
```

y resulta de dividir la suma de los cuadrados de los residuos por la suma de los cuadrados de los coeficientes de la matriz de correlación original.

Otras formas de medir el ajuste interesantes podemos realizarlas con la siguiente función provista en *Discovering Statistics Using R*, Andy Field (2012).

```
# library(psych)

residual.stats <- function(corrMatrix, model){
  matrix <- factor.residuals(corrMatrix, model$loadings)
  residuals <- as.matrix(matrix[upper.tri(matrix)])
  large.resid <- abs(residuals) > 0.05
  numberLargeResids <- sum(large.resid)
  propLargerResid <- numberLargeResids/nrow(residuals)
  rmsr <- sqrt(mean(residuals^2))

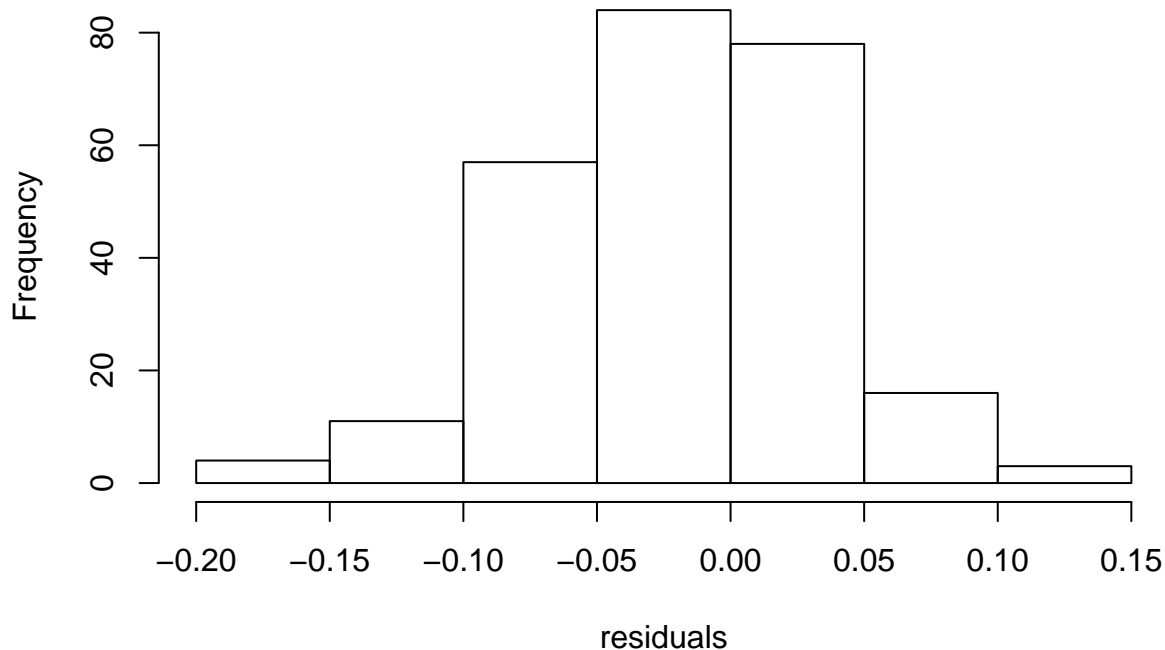
  cat("Root means squared residual = ", rmsr, "\n")
  cat("Number of absolute residuals > 0.05 = ", numberLargeResids, "\n")
  cat("Proportion of absolute residuals > 0.05 = ", propLargerResid, "\n")
  hist(residuals)
}
```

```
library(psych)
residual.stats(corrMatrix, pc1)

## Root means squared residual = 0.05549286
## Number of absolute residuals > 0.05 = 91
## Proportion of absolute residuals > 0.05 = 0.3596838
```



## Histogram of residuals



### 3.4. Rotación

El procedimiento de rotar los factores permite una mejor interpretación de la estructura subyacente en los datos. El método consiste en aumentar las cargas de las variables sobre alguna componente y minimizarlas sobre otros sin alterar las comunalidades, aunque sí la varianza explicada por cada componente.

```
pc1 <- principal(corrMatrix, nfactors = 4, rotate = "varimax")
pc1
```

```
## Principal Components Analysis
## Call: principal(r = corrMatrix, nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##      RC3  RC1  RC4  RC2  h2  u2  com
## Q01  0.24  0.50  0.36  0.06  0.43  0.57  2.4
## Q02 -0.01 -0.34  0.07  0.54  0.41  0.59  1.7
## Q03 -0.20 -0.57 -0.18  0.37  0.53  0.47  2.3
## Q04  0.32  0.52  0.31  0.04  0.47  0.53  2.4
## Q05  0.32  0.43  0.24  0.01  0.34  0.66  2.5
## Q06  0.80 -0.01  0.10 -0.07  0.65  0.35  1.0
## Q07  0.64  0.33  0.16 -0.08  0.55  0.45  1.7
## Q08  0.13  0.17  0.83  0.01  0.74  0.26  1.1
## Q09 -0.09 -0.20  0.12  0.65  0.48  0.52  1.3
## Q10  0.55  0.00  0.13 -0.12  0.33  0.67  1.2
## Q11  0.26  0.21  0.75 -0.14  0.69  0.31  1.5
## Q12  0.47  0.52  0.09 -0.08  0.51  0.49  2.1
## Q13  0.65  0.23  0.23 -0.10  0.54  0.46  1.6
## Q14  0.58  0.36  0.14 -0.07  0.49  0.51  1.8
## Q15  0.46  0.22  0.29 -0.19  0.38  0.62  2.6
## Q16  0.33  0.51  0.31 -0.12  0.49  0.51  2.6
```





```
## Q17  0.27  0.22  0.75 -0.04  0.68  0.32  1.5
## Q18  0.68  0.33  0.13 -0.08  0.60  0.40  1.5
## Q19 -0.15 -0.37 -0.03  0.43  0.34  0.66  2.2
## Q20 -0.04  0.68  0.07 -0.14  0.48  0.52  1.1
## Q21  0.29  0.66  0.16 -0.07  0.55  0.45  1.5
## Q22 -0.19  0.03 -0.10  0.65  0.46  0.54  1.2
## Q23 -0.02  0.17 -0.20  0.59  0.41  0.59  1.4
##
##                               RC3  RC1  RC4  RC2
## SS loadings                   3.73 3.34 2.55 1.95
## Proportion Var                 0.16 0.15 0.11 0.08
## Cumulative Var                 0.16 0.31 0.42 0.50
## Proportion Explained           0.32 0.29 0.22 0.17
## Cumulative Proportion          0.32 0.61 0.83 1.00
##
## Mean item complexity = 1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.06
##
## Fit based upon off diagonal values = 0.96
```

Haciendo uso de la función `print.psych` del paquete `psych` podemos mostrarlos de forma ordenada y seleccionando únicamente las cargas que nos interesen, por ejemplo, las mayores que 0.3.

```
library(psych)
print.psych(pc1, cut = 0.3, sort = TRUE)
```

```
## Principal Components Analysis
## Call: principal(r = corrMatrix, nfactors = 4, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##   item  RC3  RC1  RC4  RC2  h2  u2 com
## Q06   6  0.80
## Q18  18  0.68  0.33
## Q13  13  0.65
## Q07   7  0.64  0.33
## Q14  14  0.58  0.36
## Q10  10  0.55
## Q15  15  0.46
## Q20  20      0.68
## Q21  21      0.66
## Q03   3      -0.57      0.37  0.53  0.47  2.3
## Q12  12  0.47  0.52
## Q04   4  0.32  0.52  0.31
## Q16  16  0.33  0.51  0.31
## Q01   1      0.50  0.36
## Q05   5  0.32  0.43
## Q08   8      0.83
## Q17  17      0.75
## Q11  11      0.75
## Q09   9      0.65  0.48  0.52  1.3
## Q22  22      0.65  0.46  0.54  1.2
## Q23  23      0.59  0.41  0.59  1.4
## Q02   2      -0.34
## Q19  19      -0.37      0.43  0.34  0.66  2.2
```



```
##
##          RC3  RC1  RC4  RC2
## SS loadings      3.73 3.34 2.55 1.95
## Proportion Var   0.16 0.15 0.11 0.08
## Cumulative Var   0.16 0.31 0.42 0.50
## Proportion Explained 0.32 0.29 0.22 0.17
## Cumulative Proportion 0.32 0.61 0.83 1.00
##
## Mean item complexity = 1.8
## Test of the hypothesis that 4 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.06
##
## Fit based upon off diagonal values = 0.96
```

### 3.5. Interpretación

Basándonos en la salida anterior, podemos ver que los ítems 6, 18, 13, 7, 14, 10, y 15, que se corresponden con los enunciados

```
## [1] "6. Mis amigos son mejores que yo usando R."
## [2] "18. Me despierto bajo el edredón pensando que estoy atrapado en una distribución normal."
## [3] "13. Los ordenadores me odian."
## [4] "7. Si soy bueno en estadística, mis amigos creen que soy un friki."
## [5] "14. Los ordenadores tienen mente propia y se rompen intencionadamente cada vez que los uso."
## [6] "10. Tengo poca experiencia con los ordenadores."
## [7] "15. Los ordenadores solo sirven para jugar."
```

los cuales tienen en común que todos representan alguna forma de expresar cierto miedo hacia los computadores, por ello, podemos pensar que esta componente representa exactamente eso, el Miedo a los computadores.

La siguiente componente carga fuertemente sobre los ítems 20, 21, 3, 12, 4, 16, 1 y 5, que se corresponden con los enunciados

```
## [1] "20. La gente intenta decirte que R hace la estadística más fácil. Es mentira."
## [2] "21. Sueño que Pearson me ataca con coeficientes de correlación."
## [3] "3. Entro en coma cada vez que veo una ecuación."
## [4] "12. Me temo que causaré daños irreparables por mi incompetencia con los ordenadores."
## [5] "4. En la escuela era realmente malo en matemáticas."
## [6] "16. Los ordenadores me la tienen jurada."
## [7] "1. No entiendo la estadística."
## [8] "5. Mis amigos son mejores que yo en estadística."
```

los cuales representan el Miedo a la estadística.

La tercera componente carga sobre los ítems 8, 17 y 11, que se corresponden con los enunciados

```
## [1] "8. Mis amigos pensarán que soy estúpido por no ser capaz de enfrentarme a R."
## [2] "17. No puedo dormir pensando en vectores propios."
## [3] "11. R siempre se colapsa cuando intento usarlo."
```

los cuales representan el Miedo a las matemáticas.

Y por último, la cuarta componente está formada por los ítems 9, 22, 23, 2, y 19, que se corresponden con los enunciados

```
## [1] "9. Todo el mundo me mira cuando uso R."
## [2] "22. Llora intensamente cada vez que mencionan una medida de tendencia central."
```

```
## [3] "23. La estadística me hace llorar."
## [4] "2. Nunca he sido bueno en matemáticas."
## [5] "19. Las desviaciones estándar me ponen."
```

los cuales representan la Evaluación por pares, pues todos contienen cierta componente de opinión que los demás puedan tener sobre uno mismo .

### 3.6. Cargas de individuos

Para obtener las cargas de cada individuo sobre cada componente, es decir, las “coordenadas” de cada individuo, solo es necesario ejecutar

```
pc1 <- principal(df, nfactors = 4, rotate = "varimax", scores = TRUE)
pc1$scores[1:10, ]
```

```
##          RC3          RC1          RC4          RC2
## [1,] -0.1066865  0.93003207  1.826653434  0.4592970
## [2,]  0.5827147  0.18940296  0.041391495 -0.2934021
## [3,]  0.5473302 -0.02775712 -0.199814251  0.9714979
## [4,] -0.7460157 -0.72254964  0.686965559  0.1867694
## [5,] -0.2518823  0.51410587  0.633631733 -0.6834459
## [6,] -1.9164897  0.27320273  0.681588482  0.5265475
## [7,]  0.2595842  1.40833282  0.003891988 -0.9120978
## [8,]  0.2852544  0.91636639  0.089483877 -1.0386790
## [9,] -1.7164596 -1.15589264 -3.155156531 -0.8116503
## [10,] 0.6913221  0.73141449 -0.183353670 -1.4964086
```

### 3.7. Resumen

En la siguiente tabla se muestra un resumen de la interpretación que hemos llevado a cabo.

Tabla 1: Tabla Resumen de las Componentes Principales

	Miedo a los computadores	M
My friends are better at R than I am.	0.800	
I wake up under my duvet thinking that I am trapped under a normal distribution.	0.684	
All computers hate me.	0.647	
If I'm good at statistics my friends will think I'm a nerd.	0.638	
Computers have minds of their own and deliberately go wrong whenever I use them.	0.578	
I have little experience of computers.	0.550	
Computers are useful only for playing games.	0.459	
People try to tell you that R makes statistics easier to understand but it doesn't.	-0.039	
I dream that Pearson is attacking me with correlation coefficients.	0.286	
I slip into a coma whenever I see an equation.	-0.202	
I worry that I will cause irreparable damage because of my incompetence with computers.	0.472	
I did badly at mathematics at school.	0.319	
Computers are out to get me.	0.333	
I don't understand statistics.	0.240	
My friends are better at statistics than me.	0.319	
My friends will think I'm stupid for not being able to cope with R.	0.131	
I can't sleep for thoughts of eigen vectors.	0.271	
R always crashes when I try to use it.	0.262	
Everybody looks at me when I use R.	-0.094	




---

	Miedo a los computadores
I weep openly at the mention of central tendency.	-0.191
Statistics makes me cry.	-0.024
I have never been good at mathematics.	-0.005
Standar deviations excite me.	-0.146

---

## 4. Análisis de Fiabilidad

El coeficiente de fiabilidad más ampliamente utilizado es el  $\alpha$  de Cronbach. Suele hablarse, en ocasiones y de forma errónea, sobre un  $\alpha$  de Cronbach general para una encuesta. si tenemos claro que en cierta encuesta subyacen distintos constructos, como es nuestro caso, debe calcularse un  $\alpha$  para cada uno de ellos, teniendo en cuenta, además, que todos lo ítems apunten hacia el mismo sentido.

Los constructos de nuestra encuesta son

```
miedoComputadores <- c(6, 7, 10, 13, 14, 15, 18)
miedoEstadistica   <- c(1, 3, 4, 5, 12, 16, 20, 21)
miedoMatematicas   <- c(8, 11, 17)
evaluacionPares    <- c(2, 9, 19, 22, 23)
```

### 4.1. $\alpha$ de Cronbach para el Miedo a los computadores

```
library( psych )
psych::alpha( df[ , miedoComputadores] )

##
## Reliability analysis
## Call: psych::alpha(x = df[, miedoComputadores])
##
##   raw_alpha std.alpha G6(smc) average_r S/N   ase mean  sd
##     0.82     0.82    0.81     0.4 4.6 0.0052  3.4 0.71
##
## lower alpha upper      95% confidence boundaries
## 0.81 0.82 0.83
##
## Reliability if an item is dropped:
##   raw_alpha std.alpha G6(smc) average_r S/N alpha se
## Q06     0.79     0.79    0.77     0.38 3.7  0.0063
## Q07     0.79     0.79    0.77     0.38 3.7  0.0063
## Q10     0.82     0.82    0.80     0.44 4.7  0.0053
## Q13     0.79     0.79    0.77     0.39 3.8  0.0062
## Q14     0.80     0.80    0.77     0.39 3.9  0.0060
## Q15     0.81     0.81    0.79     0.41 4.2  0.0056
## Q18     0.79     0.78    0.76     0.38 3.6  0.0064
##
## Item statistics
##   n raw.r std.r r.cor r.drop mean  sd
## Q06 2571 0.75 0.74 0.68 0.62 3.8 1.12
## Q07 2571 0.75 0.73 0.68 0.62 3.1 1.10
## Q10 2571 0.54 0.57 0.44 0.40 3.7 0.88
## Q13 2571 0.72 0.73 0.67 0.61 3.6 0.95
```



```
## Q14 2571 0.70 0.70 0.64 0.58 3.1 1.00
## Q15 2571 0.64 0.64 0.54 0.49 3.2 1.01
## Q18 2571 0.76 0.76 0.72 0.65 3.4 1.05
##
## Non missing response frequency for each item
##      1      2      3      4      5 miss
## Q06 0.06 0.10 0.13 0.44 0.27 0
## Q07 0.09 0.24 0.26 0.34 0.07 0
## Q10 0.02 0.10 0.18 0.57 0.14 0
## Q13 0.03 0.12 0.25 0.48 0.12 0
## Q14 0.07 0.18 0.38 0.31 0.06 0
## Q15 0.06 0.18 0.30 0.39 0.07 0
## Q18 0.06 0.12 0.31 0.37 0.14 0
```

La salida anterior nos ofrece mucha información sobre la fiabilidad de nuestro constructo midiendo lo que pretende. Por un lado, `raw_alpha` y `std.alpha` son los  $\alpha$  de Cronbach basado en la covarianza y la correlación respectivamente. El coeficiente `G6(smc)` es el Guttman's Lambda 6 reliability?? y `average_r` es la correlación media inter-ítem.

La tabla `Reliability if an item is dropped` nos indica cómo cambiarían cada uno de los coeficientes anteriores al eliminar cierto ítem. Si alguno de estos coeficientes mejorase al eliminar algún ítem, deberíamos plantearnos excluirlo de este constructo, pues no lo estaría reflejándolo de forma precisa.

En la tabla `Item Statistics` tenemos, entre otras cosas, la media y desviación típica de las puntuaciones de cada ítem en las columnas `mean` y `sd` respectivamente. Por otro lado, `raw.r` y `std.r` nos dan la correlación de cada ítem con todos los demás ítems de la escala (incluyendo el ítem en cuestión), cuando los ítems están o no estandarizados respectivamente.

La última tabla es simplemente una tabla de frecuencias, que indica el porcentaje de individuos que han respondido cada respuesta en cada uno de los ítems que conforman el constructo.

## 4.2. Alfa de Cronbach para el Miedo a la estadística

Observar que aquí cambiamos la dirección de ítem 3, cuyo enunciado es

```
dicc$spanish[9]
```

```
## [1] "3. Entro en coma cada vez que veo una ecuación."
# library( psych )
psych::alpha( df[ , miedoEstadistica], keys = c(1, -1, 1, 1, 1, 1, 1, 1) )

##
## Reliability analysis
## Call: psych::alpha(x = df[, miedoEstadistica], keys = c(1, -1, 1, 1,
##      1, 1, 1, 1))
##
##      raw_alpha std.alpha G6(smc) average_r S/N      ase mean  sd
##      0.82      0.82      0.81      0.37 4.7 0.0053    3 0.64
##
## lower alpha upper      95% confidence boundaries
## 0.81 0.82 0.83
##
## Reliability if an item is dropped:
##      raw_alpha std.alpha G6(smc) average_r S/N alpha se
## Q01      0.80      0.80      0.79      0.37 4.1 0.0060
## Q03-      0.80      0.80      0.79      0.37 4.1 0.0061
```



```
## Q04      0.80      0.80      0.78      0.36 4.0  0.0062
## Q05      0.81      0.81      0.80      0.38 4.2  0.0058
## Q12      0.80      0.80      0.79      0.36 4.0  0.0061
## Q16      0.79      0.80      0.78      0.36 3.9  0.0062
## Q20      0.82      0.82      0.80      0.40 4.6  0.0055
## Q21      0.79      0.80      0.78      0.36 3.9  0.0063
##
## Item statistics
##      n raw.r std.r r.cor r.drop mean  sd
## Q01 2571 0.65 0.67 0.60 0.54 3.6 0.83
## Q03- 2571 0.69 0.67 0.60 0.55 2.6 1.08
## Q04 2571 0.69 0.70 0.64 0.58 3.2 0.95
## Q05 2571 0.63 0.63 0.55 0.49 3.3 0.96
## Q12 2571 0.69 0.69 0.63 0.57 2.8 0.92
## Q16 2571 0.71 0.71 0.67 0.60 3.1 0.92
## Q20 2571 0.58 0.56 0.47 0.42 2.4 1.04
## Q21 2571 0.72 0.71 0.67 0.61 2.8 0.98
##
## Non missing response frequency for each item
##      1 2 3 4 5 miss
## Q01 0.02 0.07 0.29 0.52 0.11 0
## Q03 0.03 0.17 0.34 0.26 0.19 0
## Q04 0.05 0.17 0.36 0.37 0.05 0
## Q05 0.04 0.18 0.29 0.43 0.06 0
## Q12 0.09 0.23 0.46 0.20 0.02 0
## Q16 0.06 0.16 0.42 0.33 0.04 0
## Q20 0.22 0.37 0.25 0.15 0.02 0
## Q21 0.09 0.29 0.34 0.26 0.02 0
```

### 4.3. $\alpha$ de Cronbach para el Miedo a las matemáticas

```
# library( psych )
psych::alpha( df[, miedoMatematicas] )

##
## Reliability analysis
## Call: psych::alpha(x = df[, miedoMatematicas])
##
##      raw_alpha std.alpha G6(smc) average_r S/N ase mean  sd
##      0.82      0.82      0.75      0.6 4.5 0.0062 3.7 0.75
##
## lower alpha upper      95% confidence boundaries
## 0.81 0.82 0.83
##
## Reliability if an item is dropped:
##      raw_alpha std.alpha G6(smc) average_r S/N alpha se
## Q08      0.74      0.74      0.59      0.59 2.8 0.010
## Q11      0.74      0.74      0.59      0.59 2.9 0.010
## Q17      0.77      0.77      0.63      0.63 3.4 0.009
##
## Item statistics
##      n raw.r std.r r.cor r.drop mean  sd
## Q08 2571 0.86 0.86 0.76 0.68 3.8 0.87
```



```
## Q11 2571 0.86 0.86 0.75 0.68 3.7 0.88
## Q17 2571 0.85 0.85 0.72 0.65 3.5 0.88
##
## Non missing response frequency for each item
##      1      2      3      4      5 miss
## Q08 0.03 0.06 0.19 0.58 0.15 0
## Q11 0.02 0.06 0.22 0.53 0.16 0
## Q17 0.03 0.10 0.27 0.52 0.08 0
```

#### 4.4. $\alpha$ de Cronbach para el Evaluación por pares

```
# library( psych )
psych::alpha( df[, evaluacionPares] )

##
## Reliability analysis
## Call: psych::alpha(x = df[, evaluacionPares])
##
##      raw_alpha std.alpha G6(smc) average_r S/N ase mean sd
##      0.57      0.57      0.53      0.21 1.3 0.013 3.4 0.65
##
## lower alpha upper      95% confidence boundaries
## 0.54 0.57 0.6
##
## Reliability if an item is dropped:
##      raw_alpha std.alpha G6(smc) average_r S/N alpha se
## Q02      0.52      0.52      0.45      0.21 1.07 0.015
## Q09      0.48      0.48      0.41      0.19 0.92 0.017
## Q19      0.52      0.53      0.46      0.22 1.11 0.015
## Q22      0.49      0.49      0.43      0.19 0.96 0.016
## Q23      0.56      0.57      0.50      0.25 1.32 0.014
##
## Item statistics
##      n raw.r std.r r.cor r.drop mean sd
## Q02 2571 0.56 0.61 0.45 0.34 4.4 0.85
## Q09 2571 0.70 0.66 0.53 0.39 3.2 1.26
## Q19 2571 0.61 0.60 0.42 0.32 3.7 1.10
## Q22 2571 0.64 0.64 0.50 0.38 3.1 1.04
## Q23 2571 0.53 0.53 0.31 0.24 2.6 1.04
##
## Non missing response frequency for each item
##      1      2      3      4      5 miss
## Q02 0.01 0.04 0.08 0.31 0.56 0
## Q09 0.08 0.28 0.23 0.20 0.20 0
## Q19 0.02 0.15 0.22 0.33 0.29 0
## Q22 0.05 0.26 0.34 0.26 0.10 0
## Q23 0.12 0.42 0.27 0.12 0.06 0
```



## Referencias y bibliografía

Field, Andy, Jeremy Miles, and Zoe Field. 2012. *Discovering Statistics Using R*. 1st edition. Sage Publications Ltd.