

# 1. Análisis de Componentes Principales

## Sparse y Descomposición CUR

En esta práctica analizamos los resultados del cuestionario MBI aplicado sobre un grupo de 51 profesionales del colectivo de los farmacéuticos. El objetivo del estudio es corroborar la existencia de las tres escalas definidas por teoría: agotamiento personal, despersonalización y autoestima profesional.

El **Burnout** (término que puede traducirse como “estar quemado”) es una patología severa, relativamente reciente (1974), que está relacionada con el ámbito laboral y el estilo de vida que se lleva. Se trata de un síndrome clínico provocado por el trabajo se traduce en una pérdida progresiva de energía que acaba desmotivando al profesional con su trabajo. Este trastorno emocional conlleva graves consecuencias físicas y psicológicas, como el agotamiento y la depresión, cuando el fenómeno se somatiza. La problemática suele ser padecida, principalmente, en profesionales del ámbito de los servicios sociales. Las profesiones relacionadas con el mundo sanitario, de la educación o la administración pública suelen ser las que más incidencia reflejan en las estadísticas, debido a los cambios de horario y largas jornadas de trabajo, y a que están en contacto con personas con problemas y suelen ser ellos los que deben solucionarlos, llegando hasta a involucrarse en ellos.

El *Maslach Burnout Inventory* (**MBI**) es el instrumento utilizado para medir la presencia del síndrome en distintos colectivos de profesionales. En él se plantea al sujeto una serie de enunciados sobre los sentimientos y pensamientos con relación a su interacción con el trabajo. Es el instrumento que ha generado un mayor volumen de estudios e investigaciones, formado por 22 ítems que se valoran con una escala tipo Likert. El sujeto valora, mediante un rango de 5 adjetivos que van de 'nunca' a 'diariamente', con qué frecuencia experimenta cada una de las situaciones descritas en los ítems. Los 22 ítems que conforman este cuestionario se muestran a continuación:

- **Item 1.** Me siento emocionalmente agotado por mi trabajo.
- **Item 2.** Me siento cansado al final de la jornada de trabajo.
- **Item 3.** Me siento fatigado cuando me levanto por la mañana y tengo que enfrentarme con otro día de trabajo.
- **Item 4.** Fácilmente comprendo cómo se sienten los clientes/pacientes.
- **Item 5.** Creo que trato a las otras personas como si fuesen objetos impersonales.
- **Item 6.** Trabajar todo el día con mucha gente es un esfuerzo.
- **Item 7.** Trato eficazmente los problemas de los demás.
- **Item 8.** Siento que mi trabajo me está desgastando.
- **Item 9.** Creo que estoy influyendo positivamente, con mi trabajo, en la vida de los demás.
- **Item 10.** Me he vuelto más insensible con la gente desde que ejerzo esta profesión.
- **Item 11.** Me preocupa el hecho de que este trabajo me esté endureciendo emocionalmente.
- **Item 12.** Me siento muy activo.
- **Item 13.** Me siento frustrado con mi trabajo.
- **Item 14.** Creo que estoy trabajando demasiado.
- **Item 15.** No me preocupa lo que le ocurra a los clientes/pacientes.
- **Item 16.** Trabajar directamente con los demás me produce estrés.
- **Item 17.** Fácilmente puedo crear una atmósfera relajada con los clientes/pacientes.
- **Item 18.** Me siento estimulado después de trabajar en contacto con los clientes/pacientes.
- **Item 19.** He conseguido muchas cosas útiles en mi profesión.
- **Item 20.** Me siento como si estuviera al límite de mis posibilidades.
- **Item 21.** En mi trabajo trato los problemas que se me presentan con mucha calma.
- **Item 22.** Creo que mis clientes/pacientes me culpan de alguno de sus problemas.

Tal y como mostraron [Maslach y Jackson](#) en 1981, la factorización de los 22 ítems arroja en la mayoría de los trabajos 3 factores: **agotamiento emocional**, **despersonalización** y **realización personal** en el trabajo. Los tres factores han sido obtenidos con muestras de diferentes colectivos profesionales. Estos factores constituyen las tres subescalas del MBI y los distintos ítems del cuestionario se relacionan con una de dichas subescalas, según el siguiente patrón:

Subescala	Ítems
Despersonalización	5, 10, 11, 15, 22
Agotamiento Emocional	1, 2, 3, 6, 8, 13, 14, 16, 20
Autoestima Profesional	4, 7, 9, 12, 17, 18, 19, 21

Gracias a estas escalas, podemos realizar una interpretación de los resultados obtenidos por los profesionales en el cuestionario. El **síndrome de burnout** estará presente en los profesionales; es decir, los individuos estarán desgastados por su trabajo, cuando sus resultados en el MBI sean los siguientes: una **puntuación alta en Agotamiento Emocional** (debido a la tensión elevada que sufre el profesional, carente de recursos emocionales, se vuelve menos sensible con el resto de individuos y más duro consigo mismo), una **puntuación alta en Despersonalización** (el profesional adopta una actitud fría y distante hacia los demás, viéndolos como objetos), y una **puntuación baja en Autoestima Profesional** (baja confianza en la capacidad de enfrentar personalmente los desafíos de su trabajo).

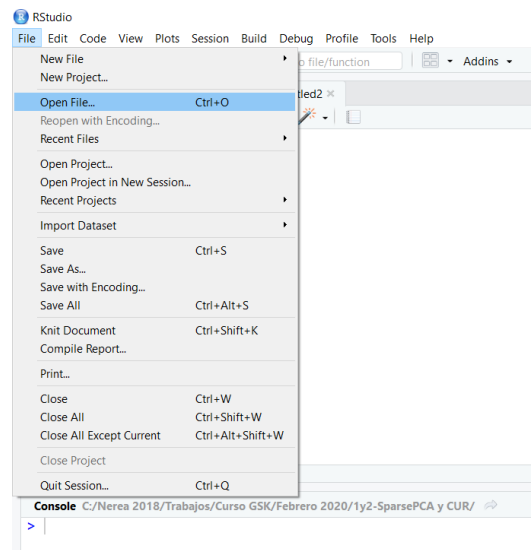
El análisis factorial de los 22 ítems produce en la mayoría de los trabajos 3 factores que son denominados agotamiento emocional, despersonalización y realización personal en el trabajo. Analizaremos la existencia de estas escalas a través del análisis de la respuesta media de un conjunto de 51 farmacéuticos a cada uno de los 22 ítems. Con el propósito de ver si la supuesta estructura a priori está en los datos, compararemos los resultados del Análisis Factorial y Análisis de Componentes Principales, con y sin rotación VARIMAX y una alternativa más reciente en la literatura conocida con el nombre de Análisis de Componentes Principales *Sparse*.

La matriz de datos está formada por 51 filas (51 individuos evaluados) y 22 columnas (los 22 ítems del cuestionario MBI). La reducción de la dimensión de esta matriz se lleva a cabo a través de las técnicas mencionadas, con el objetivo de encontrar las 3 componentes de las que hablan los profesionales: agotamiento personal, despersonalización y autoestima profesional. A través de las matrices de saturaciones/cargas analizaremos los ítems que conforman cada una de las componentes.

Para ello realice las siguientes tareas:

1. Cargue la base de datos en R.

File -> Open File -> Datos\_MBI.RData



2. Compruebe que la lectura de los datos se ha realizado correctamente. Para ello analice la dimensión y la clase del objeto, así como el nombre de las variables. ¿Cuál es el tamaño muestral del estudio?

Se trata de un objeto del tipo `data.frame`, de tamaño muestral  $n = 51$ , con los ítems almacenados en columnas.

```
dim(mbi)
[1] 51 22
colnames(mbi)
[1] "F1" "F2" "F3" "F6" "F8" "F13" "F14" "F16" "F20"
[10] "F4" "F7" "F9" "F12" "F17" "F18" "F19" "F21" "F5"
[19] "F10" "F11" "F15" "F22"
class(mbi)
"data.frame"
```

3. Cree un vector de colores que le ayude a identificar la dimensión latente a la que pertenecen los ítems recogidos (recuerde que las variables están ordenadas según la dimensión latente teórica de cada ítem definida por los autores).

Identificamos los 9 ítems del Agotamiento Emocional con el color verde, los 8 de Autoestima con el color azul y los ítems de Despersonalización con el color naranja.

```
colores<-c(rep("green",9),rep("blue",8),rep("orange",5))
colores
[1] "green" "green" "green" "green" "green" "green"
[7] "green" "green" "green" "blue" "blue" "blue"
[13] "blue" "blue" "blue" "blue" "blue" "blue"
[19] "orange" "orange" "orange" "orange"
```

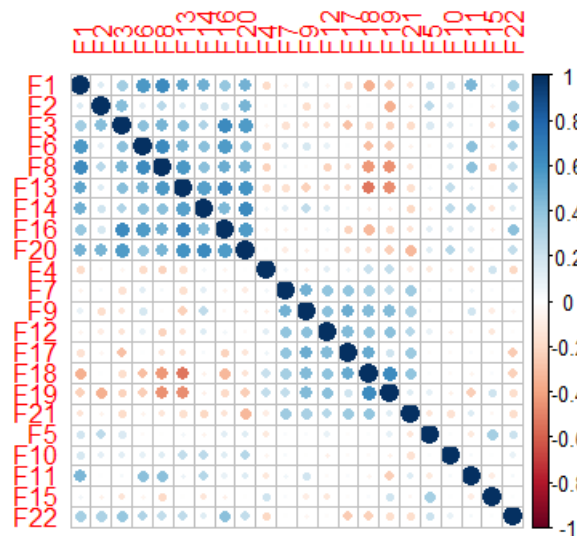
#### 4. Analice la información de la matriz de correlaciones.

Instalamos y cargamos la librería `corrplot` para hacer un análisis gráfico de la matriz de correlaciones.

```
library(corrplot)
install.packages("corrplot")
```

A continuación se calcula la matriz de correlaciones y se grafica:

```
cmbi<-cor(mbi)
corrplot(cmbi)
```



Se observa una relación directa entre los ítems del Agotamiento Emocional y entre los ítems de la Autoestima. Además, se observa una relación inversa entre ítems de Agotamiento y Autoestima. Sin embargo, los ítems de la Despersonalización presentan correlaciones muy bajas entre sí.

#### 5. Implemente el Análisis Factorial sin rotación. Analice el porcentaje de varianza explicada por cada uno de los factores y la estructura de la matriz de configuración. ¿Qué estructura puede adivinarse para el MBI en la muestra de farmacéuticos a partir de los resultados?

Instalamos y cargamos la librería `stats`:

```
install.packages("stats")
library(stats)
```

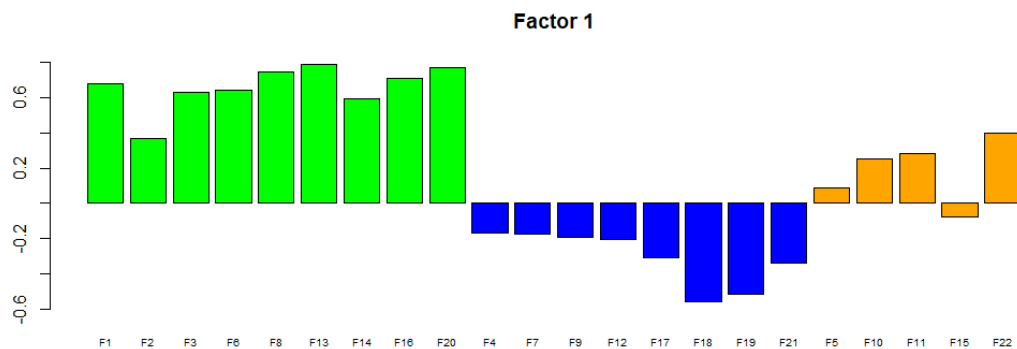
De forma exploratoria, generaremos una solución factorial no rotada, con 3 dimensiones. Almacenamos en un objeto al que llamaremos saturaciones las saturaciones de cada ítem en los 3 factores latentes retenidos.

```
FA_mbi<-factanal(mbi,factors=3,rotation="none")
FA_mbi
```

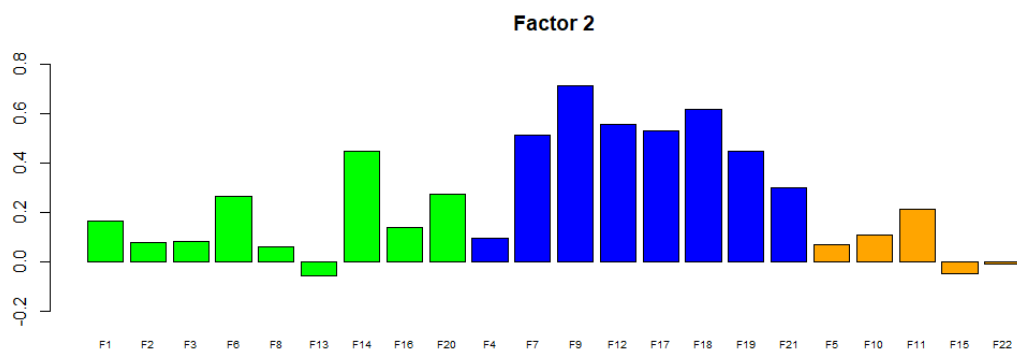
```
saturaciones<-FA_mbi$loadings[,1:3]
saturaciones
```

Representamos mediante un diagrama de barras las saturaciones en cada dimensión:

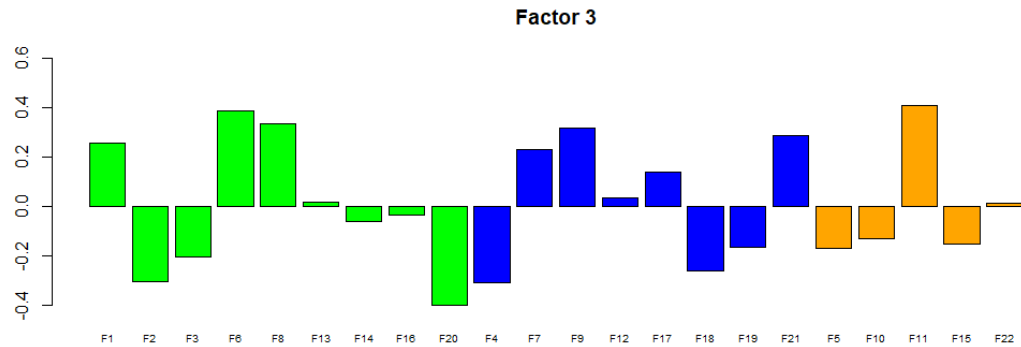
```
FA1<-barplot(saturaciones[,1],ylim=c(-0.6,0.8),names.arg=colnames(mbi),cex.names=0.6,col = colores,main = "Factor 1")
```



```
FA2<-barplot(saturaciones[,2],ylim=c(-0.2,0.8),names.arg=colnames(mbi),cex.names=0.6,col = colores,main = "Factor 2")
```



```
FA3<-barplot(saturaciones[,3],ylim=c(-0.4,0.6),names.arg=colnames(mbi),cex.names=0.6,col = colores,main = "Factor 3")
```

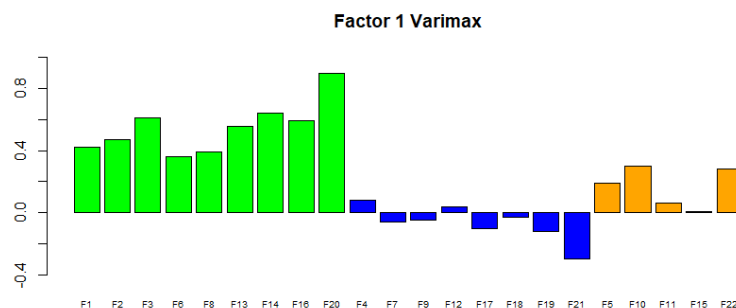


Prácticamente todas las saturaciones más altas están en el primer factor. No aparece ninguna variable especialmente relevante en la formación de los factores. No existe una interpretación clara de los ejes comunes que nos lleve a identificar de forma segura los ítems que conforman cada una de las escalas de las que hablaban los autores. Aun así, podemos identificar el primer factor como el agotamiento emocional y el segundo como la autoestima. Sin embargo, en todas ellas aparecen variables (ítems) con saturaciones también “importantes” que no se tendrían en cuenta si siguiésemos la estructura latente teórica. El análisis Factorial sobre esta matriz de datos no identifica de manera clara la estructura teórica buscada no se trata de un buen modelo factorial. La despersonalización no aparece recogida en ninguna dimensión latente.

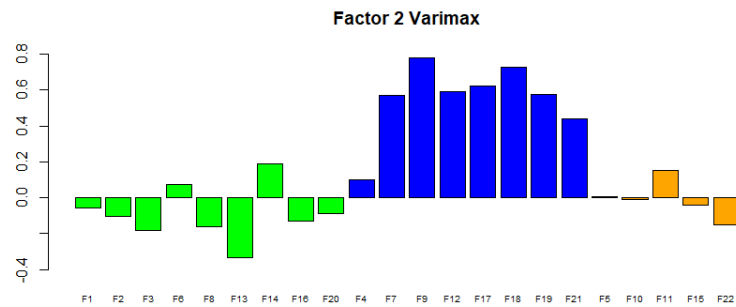
6. Implemente el Análisis Factorial con rotación Varimax. ¿Qué estructura puede adivinarse para el MBI en la muestra de farmacéuticos ahora?

Basta con modificar el argumento rotation de la función *factanal*:

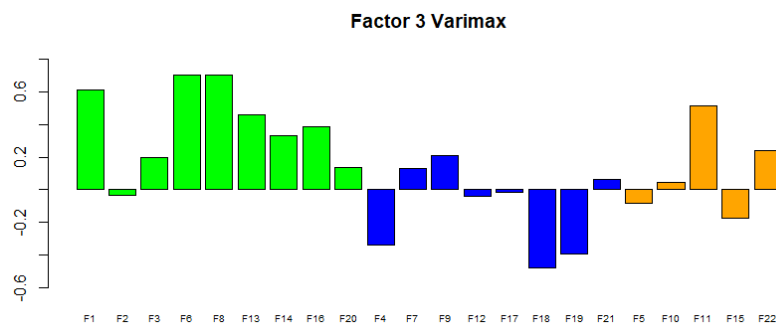
```
FA_mbi_varimax<-factanal(mbi,factors=3,rotation="varimax")
FA_mbi_varimax
barplot(FA_mbi_varimax$loadings[,1],ylim=c(-0.4,1),names.arg=col
names(mbi),cex.names=0.6,col = colores,main = "Factor 1 Varimax"
)
```



```
barplot(FA_mbi_varimax$loadings[,2],ylim=c(-0.4,0.8),names.arg=c(
olnames(mbi),cex.names=0.6, col = colores,main = "Factor 2 Varimax"))
```



```
barplot(FA_mbi_varimax$loadings[,3],ylim=c(-0.6,0.8),names.arg=c(
olnames(mbi),cex.names=0.6, col = colores,main = "Factor 3 Varimax"))
```



Si observamos la matriz de saturaciones rotadas, aunque la interpretación no es del todo fácil, se ha mejorado con respecto al caso anterior. Puede identificarse el primer factor de agotamiento y el segundo de autoestima. Sin embargo, este método falla en la identificación del tercer factor como la escala de la despersonalización, pues las saturaciones que definen el factor no coinciden con los ítems que definían la escala según los autores. La dimensión de la despersonalización no es clara en el colectivo de los farmacéuticos. Esto se debe a que no existe variabilidad en los ítems de dicha latente en esta muestra de sujetos.

7. Implemente el Análisis de Componentes Principales sin rotación. Analice el porcentaje de varianza explicada por cada una de las componentes y la estructura de la matriz de cargas. ¿Qué puede concluir acerca de la estructura factorial del cuestionario?

El ACP se implementa mediante la función `prcomp`:

```
PCAmbi<-prcomp(mbi)
```



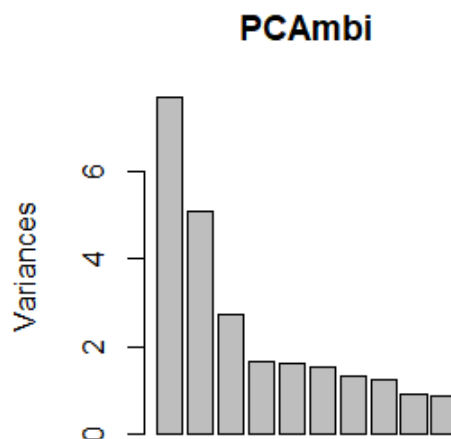
La matriz de cargas está almacenada bajo el nombre de *rotation*:

```
matrizcargas<-PCAmbi$rotation[,1:3]
```

El *screeplot* puede obtenerse gráficamente como sigue:

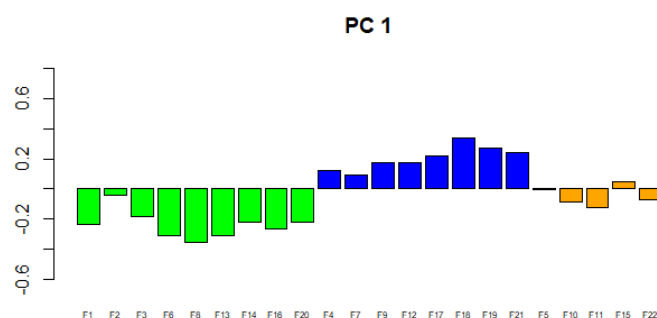
```
plot(PCAmbi)
```

Aunque en nuestro caso hemos seleccionado inicialmente 3 factores, existen distintos para seleccionar el número adecuado. El hecho de que un valor propio esté claramente separado del que le precede y del que le sigue, es señal de que el eje que le corresponde está bien separado y es razonable buscar la interpretación de ese eje. Con este razonamiento respaldamos nuestra elección inicial de tres dimensiones factoriales:

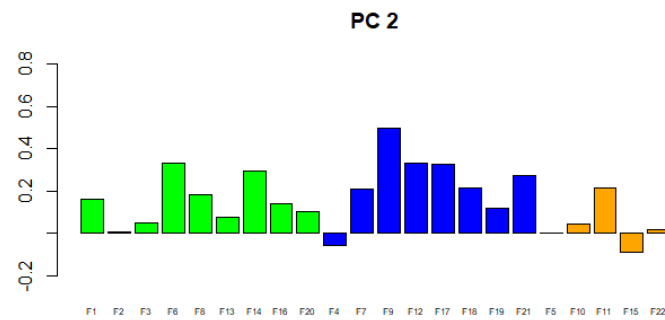


Se analiza la matriz de cargas factoriales, representando gráficamente el valor de las mismas para cada componente principal mediante un diagrama de barras:

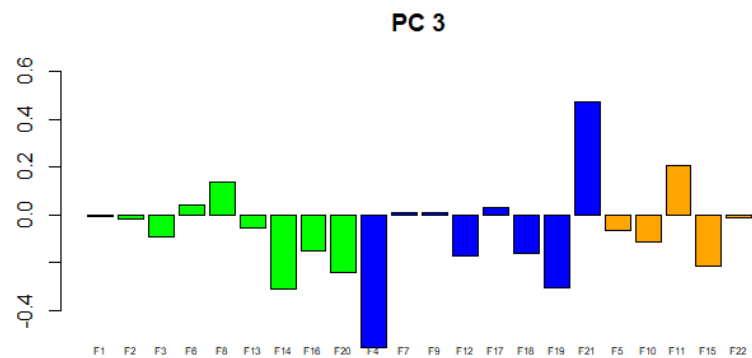
```
PCA1<-barplot(PCAmbi$rotation[,1],ylim=c(-0.6,0.8),names.arg=col  
names(mbi),cex.names=0.5,col = colores,main = "PC 1")
```



```
PCA2<-barplot(PCAmbi$rotation[,2],ylim=c(-0.2,0.8),names.arg=colnames(
mbi),cex.names=0.5,col = colores,main = "PC 2")
```



```
PCA3<-barplot(PCAmbi$rotation[,3],ylim=c(-0.4,0.6),names.arg=colnames(
mbi),cex.names=0.5,col = colores,main = "PC 3")
```



Al igual que ocurría en el Análisis Factorial, la interpretación de las cargas obtenidas tampoco es clara. Para facilitar la interpretación, podría establecerse un umbral por debajo del cual considerar cargas como nulas sin serlo (es decir, realizar una selección manual de las características, a diferencia del Análisis de Componentes Principales Sparse que veremos más adelante). Sin embargo, no todos los ítems establecidos por los autores cargan en la escala definida. La estructura vuelve a no ser clara.

**8.** Instale y cargue la librería *elasticnet* con la que se ejecutará el Análisis de Componentes Principales Sparse.

```
install.packages("elasticnet")
library(elasticnet)
```

9. Realice un Análisis de Componentes Principales Sparse, haciendo uso de la función *spca* de la librería *elasticnet*, reteniendo 3 componentes principales e imponiendo el número de cargas no nulas que debe aparecer en cada una de las componentes sparse (9,8 y 5 respectivamente).

La función *spca* trabaja sobre la matriz de correlaciones de los datos de partida. El número de componentes sparse a retener se indica mediante el argumento *K*, el tipo de datos con el que trabajamos mediante el argumento *type* (*type*="gram" indica que la matriz de partida es una matriz de correlaciones). Como se desea restringir el número de cargas no nulas de cada componente debe utilizarse el argumento *sparse*="varnum". Por último, el número de cargas no nulas deseadas en cada dimensión se indica en el argumento *para*:

```
spca1<-spca(cor(mbi),K=3,type="Gram",sparse="varnum",trace=TRUE,
para=c(9,8,5))
```

10. ¿Qué porcentaje de varianza retienen ahora dichas componentes? ¿Qué estructura puede adivinar analizando la matriz de cargas?

```
spca1
Call:
spca(x = cor(mbi), K = 3, para = c(9, 8, 5), type = "Gram", sparse = "varnum",
      trace = TRUE)
```

```
3 sparse PCs
Pct. of exp. var. : 20.2 14.0 6.4
Num. of non-zero loadings : 9 8 5
Sparse loadings
```

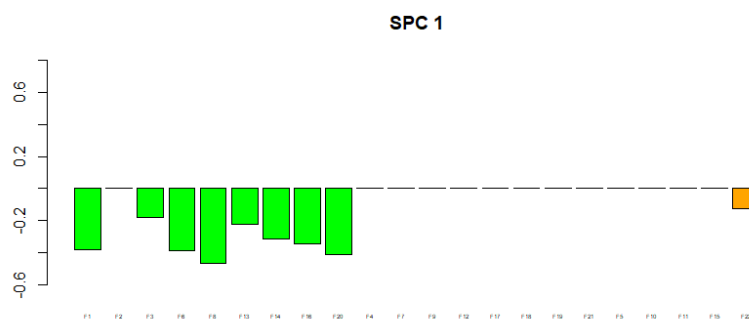
	PC1	PC2	PC3
F1	-0.383	0.000	0.000
F2	0.000	0.000	0.000
F3	-0.179	-0.050	0.000
F6	-0.388	0.000	0.000
F8	-0.466	0.000	0.000
F13	-0.225	0.000	0.000
F14	-0.313	0.000	0.000
F16	-0.346	0.000	0.000
F20	-0.412	0.000	0.000
F4	0.000	0.000	0.000
F7	0.000	0.351	0.000
F9	0.000	0.522	0.000
F12	0.000	0.414	0.000
F17	0.000	0.344	0.000
F18	0.000	0.454	0.000
F19	0.000	0.277	0.000
F21	0.000	0.170	-0.003
F5	0.000	0.000	0.780
F10	0.000	0.000	0.000
F11	0.000	0.000	-0.376
F15	0.000	0.000	0.476
F22	-0.126	0.000	0.151

El porcentaje de varianza explicada puede observarse en:

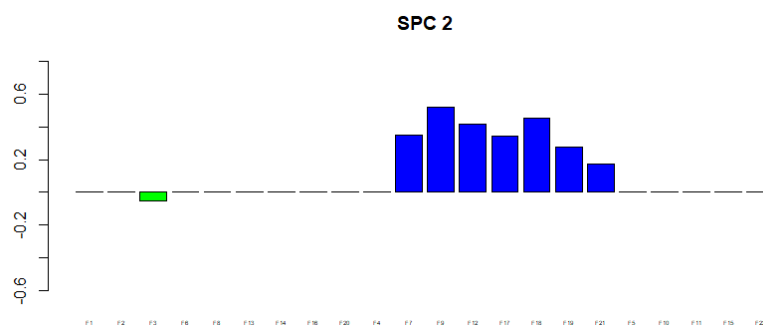
```
spca1$pev
```

Las componentes sparse retienen un 20.2%, 14% y 6.4% de varianza respectivamente. En total absorben un 40.6% de la varianza de los datos. Este porcentaje es menor al de las técnicas clásicas. Sin embargo, la interpretación de la matriz de cargas es más sencilla. Representando las cargas de cada componente:

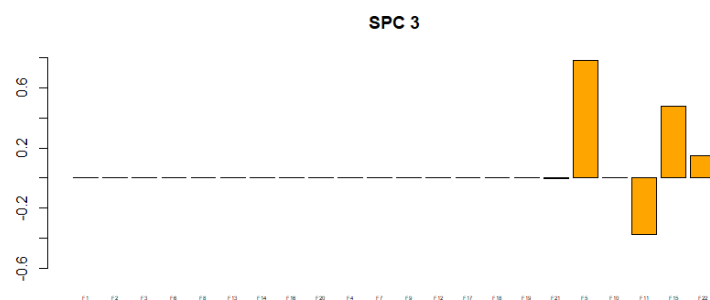
```
SPC1<-barplot(spca1$loadings[,1],ylim=c(-0.6,0.8),
names.arg=colnames(mbi),cex.names=0.4,col = colores,
main = "SPC 1")
```



```
SPC2<-barplot(spca1$loadings[,2],ylim=c(-0.6,0.8),
names.arg=colnames(mbi),cex.names=0.4, col = colores,
main = "SPC 2")
```



```
SPC3<-barplot(spca1$loadings[,3],ylim=c(-0.6,0.8),
names.arg=colnames(mbi),cex.names=0.4,col = colores,
main = "SPC 3")
```



Ahora, la primera componente sparse está conformada claramente por ítems del Agotamiento, la segunda representa la Autoestima y la tercera la Despersonalización. A diferencia de las técnicas clásicas, que trabajaban maximizando la varianza, el sparse ACP es capaz de identificar la despersonalización.

**11.** Realice ahora un Análisis de Componentes Principales Sparse, reteniendo 3 componentes principales e imponiendo ahora una penalización de 0.15 a la primera componente, 0.2 a la segunda y 0.3 a la tercera, para conseguir anular algunas de las cargas. Interprete los resultados obtenidos.

Ahora es necesario cambiar el argumento *sparse*, para indicarle que en este caso se penalizará mediante un factor de regularización las cargas de cada componente. En el argumento *para* se indica ahora el factor de penalización para cada componente:

```
spca2<-spca(cor(mbi),k=3,type="Gram",sparse="penalty",trace=TRUE,para=c(0.15,0.2,0.3))
```

La estructura factorial obtenida es similar y siguen identificándose las tres componentes sparse.

**12.** Modifique el factor de penalización de cada una de las componentes y analice las diferencias que encuentra en sus resultados. Tenga en cuenta que el factor de penalización siempre debe ser positivo: una penalización de 0 equivaldrá a un Análisis de Componentes Principales clásico y un factor de penalización muy alto hará que todas las cargas sean nulas.

**13.** Instale y cargue la librería rCUR.

```
install.packages("rCUR")  
library(rCUR)
```

**14.** Identifique los ítems que mayor variabilidad presentan en la muestra de farmacéuticos a partir de los leverage de la descomposición CUR, mediante el método “top.scores” en las 3 primeras componentes. ¿A qué dimensión latente pertenecen los ítems con mayor variabilidad? ¿Qué ocurre con los ítems pertenecientes a Despersonalización? Ayúdese de la representación gráfica de los leverage mediante la función `plotLeverage`.

La función *CUR* trabaja sobre la matriz de datos, cuya clase debe ser *matrix* y no *data.frame*. El número de componentes sobre las que va a calcular la variabilidad de cada variable se indica mediante el argumento *k*. El argumento *c* indica el número de columnas a seleccionar (debe ser siempre, como máximo, ( $n^{\circ}$  de variables - 1)) y el argumento *weighted* permite ponderar las componentes calculadas por los valores propios de cada componente. Por último, el argumento *method* permite determinar el método para seleccionar las columnas y/o filas en la descomposición. Se selecciona en esta ocasión el método "top.scores": se seleccionan las filas y/o columnas con mayor leverage.

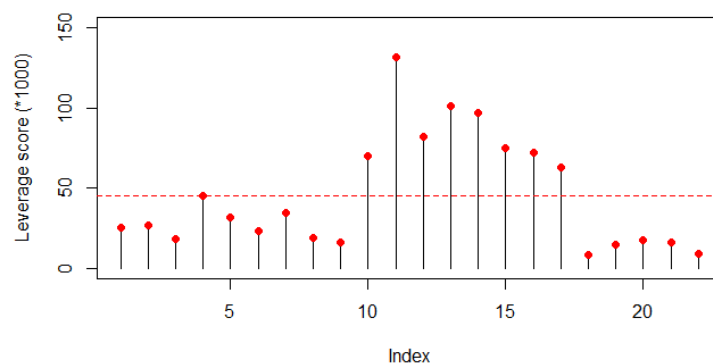
```
cur1<-CUR(as.matrix(mbi), k=3, c=21,method="top.scores",weighted
= TRUE,)
```

Podemos acceder a los leverage de cada ítem mediante la función:

```
leverage(cur1)
```

O graficarlos:

```
plotLeverage(cur1,ylim=c(0,150))
```



Los ítems de despersonalización, en las tres componentes retenidas, son los ítems que presentan una menor variabilidad y es por eso que las técnicas clásicas como el AF o el ACP presentan dificultades a la hora de identificar dicho factor latente. En el caso del colectivo de farmacéuticos se produce un efecto suelo en la variabilidad de despersonalización: de manera general, los farmacéuticos no sufren de despersonalización pues no trabajan siempre con un contacto directo con pacientes. Los ítems con mayor variabilidad en la muestra son los correspondientes a la subescala de la Autoestima.