Bonos - RMD

Sergio Casares

8/11/2020

1. Introducción al trabajo

El objetivo que perseguimos en el presente trabajo es, simplemente, efectuar una comprobación empírica mediante la aplicación del ACP a un conjunto de 978 observaciones de los rendimientos de 10 bonos norteamericanos a distintos plazos entre el 2 de enero de 1995 y el 30 de septiembre de 1998. No pretendemos nada más que verificar si, tal y como plantean los estudios teóricos, puede establecerse una estructura subyecente que sintetice y agrupe los distintos plazos en virtud de sus características comunes.

2.Importamos librerias y el Dataset

```
library(readr)
library(glmnet)
library(tidyverse)
library(fBasics)
library(car)
library(dplyr)
library(ggplot2)
library(knitr)
library(MASS)
library(corrplot)
library(PerformanceAnalytics)
library(gvlma)
library(tinytex)
library(devtools)
library(rsample)
library(tidyr)
library(broom)
library(flextable)
library(mgcv)
library(reshape2)
library(rmdformats)

library(factoextra)
library(FactoMineR)

bonos <- read_delim("ACPTIUSD.csv", ";",
                    escape_double = FALSE, trim_ws = TRUE)
head(bonos)
```

X1 <chr>	DEPO 1M <dbl>	DEPO 3M <dbl>	DEPO 6M <dbl>	DEPO 12M <dbl>	IRS 2Y <dbl>	IRS 3Y <dbl>	IRS 4Y <dbl>	IRS 5Y <dbl>
02/01/1995	6.000	6.500	7.000	7.750	8.170	8.24	8.25	8.22
03/01/1995	5.938	6.500	7.000	7.813	8.220	8.28	8.29	8.29
04/01/1995	5.938	6.500	7.000	7.813	8.130	8.20	8.23	8.24
05/01/1995	5.898	6.438	6.938	7.688	8.145	8.22	8.26	8.27
06/01/1995	5.875	6.438	6.938	7.688	8.070	8.20	8.25	8.27
09/01/1995	5.875	6.375	6.875	7.688	8.135	8.25	8.29	8.31

6 rows | 1-9 of 11 columns

3. Limpieza y Transformación del Dataser y Analisis Exploratorio

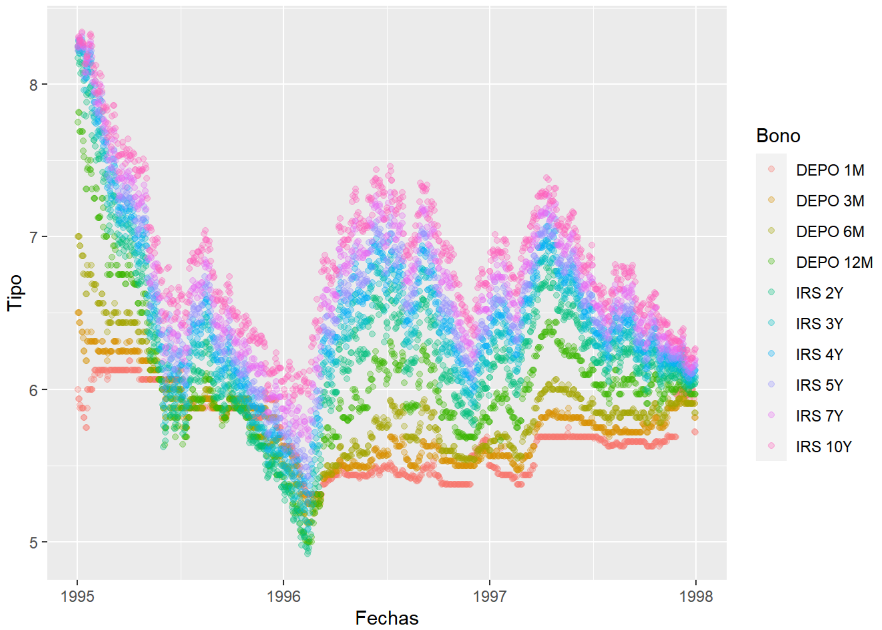
```
#Cambiamos el formato de las fechas del dataset

bonos = bonos[complete.cases(bonos), ]
bonos$Fechas = as.Date(bonos$X1, format = "%d/%m/%Y")
bonos=bonos[,2:12]

## Función melt de reshape2: "estira" el data frame

bonos_m <- melt( bonos )
View(bonos_m)

# Lo aplicamos al TIUSD2, empleando como indicador Fechas
data_long = melt(bonos, id="Fechas")
ggplot(data=data_long, aes(x= Fechas, y=value,  color=variable)) +
  geom_point(alpha = 0.3,  position = position_jitter()) +
  labs(y = "Tipo", colour="Bono")
```




```
#Estructura del Dataset

#Training - observaciones activas
bonos_act <- bonos[1:750,1:9]
fechas_act <- bonos[1:750, 11]
bonos_act <- cbind(bonos_act, fechas_act)
dim(bonos_act)
```

```
## [1] 750  10
```

```
View(bonos_act)

#Testing - observaciones suplementarias
bonos_sup <- bonos[750:783, 10]
dim(bonos_sup)
```

```
## [1] 34  1
```

```
View(bonos_sup)

#Summary
#1
summary(bonos_act)
```

```
##     DEPO 1M         DEPO 3M         DEPO 6M         DEPO 12M    
##  Min.   :5.313   Min.   :5.250   Min.   :5.121   Min.   :4.996  
##  1st Qu.:5.438   1st Qu.:5.551   1st Qu.:5.625   1st Qu.:5.813  
##  Median :5.656   Median :5.738   Median :5.813   Median :5.969  
##  Mean   :5.680   Mean   :5.758   Mean   :5.841   Mean   :6.026  
##  3rd Qu.:5.875   3rd Qu.:5.875   3rd Qu.:5.938   3rd Qu.:6.179  
##  Max.   :6.188   Max.   :6.500   Max.   :7.000   Max.   :7.813  
##      IRS 2Y          IRS 3Y          IRS 4Y          IRS 5Y     
##  Min.   :4.920   Min.   :5.080   Min.   :5.260   Min.   :5.430  
##  1st Qu.:5.950   1st Qu.:6.100   1st Qu.:6.201   1st Qu.:6.285  
##  Median :6.175   Median :6.315   Median :6.435   Median :6.530  
##  Mean   :6.222   Mean   :6.372   Mean   :6.486   Mean   :6.579  
##  3rd Qu.:6.445   3rd Qu.:6.625   3rd Qu.:6.744   3rd Qu.:6.840  
##  Max.   :8.220   Max.   :8.280   Max.   :8.290   Max.   :8.310  
##      IRS 7Y          Fechas          
##  Min.   :5.670   Min.   :1995-01-02  
##  1st Qu.:6.400   1st Qu.:1995-09-20  
##  Median :6.665   Median :1996-06-08  
##  Mean   :6.717   Mean   :1996-06-08  
##  3rd Qu.:6.985   3rd Qu.:1997-02-25  
##  Max.   :8.330   Max.   :1997-11-14  
```

```
#2
bonos_act_sum <- bonos_act[-10]

bonos_act_stats = data.frame(
  Min = apply(bonos_act_sum, 2, min, na.rm=TRUE), # mín
  Q1 = apply(bonos_act_sum, 2, quantile, 1/4, na.rm=TRUE), # 1er cuartil
  Med = apply(bonos_act_sum, 2, median, na.rm=TRUE), # mediana
  Mean = apply(bonos_act_sum, 2, mean, na.rm=TRUE), # media
  SD = apply(bonos_act_sum, 2, sd), # Desviación típica
  Q3 = apply(bonos_act_sum, 2, quantile, 3/4, na.rm =TRUE), # 3er cuartil
  Max = apply(bonos_act_sum, 2, max, na.rm=TRUE) # Máx
)
bonos_act_stats=as.data.frame(round(bonos_act_stats, 1))
bonos_act_stats
```

	Min <dbl>	Q1 <dbl>	Med <dbl>	Mean <dbl>	SD <dbl>	Q3 <dbl>	Max <dbl>
DEPO 1M	5.3	5.4	5.7	5.7	0.2	5.9	6.2
DEPO 3M	5.2	5.6	5.7	5.8	0.3	5.9	6.5
DEPO 6M	5.1	5.6	5.8	5.8	0.3	5.9	7.0
DEPO 12M	5.0	5.8	6.0	6.0	0.5	6.2	7.8
IRS 2Y	4.9	6.0	6.2	6.2	0.5	6.4	8.2
IRS 3Y	5.1	6.1	6.3	6.4	0.5	6.6	8.3
IRS 4Y	5.3	6.2	6.4	6.5	0.5	6.7	8.3
IRS 5Y	5.4	6.3	6.5	6.6	0.5	6.8	8.3
IRS 7Y	5.7	6.4	6.7	6.7	0.5	7.0	8.3

9 rows

Como se puede observar en el gráfico anterior, podemos ver la evolución del tipo de interés de cada bono o depósito a lo largo de 1995 a 1998. A partir de mediados de 1995 podemos ver una convergencia de los tipos de interés (aunque se puede observar que el tipo de interés es mayor en proporción a la duración del mismo). A partir de 1996 se observa una clara diferenciación de los tipos de interes y permanece esa tendencia hasta 1998 donde vuelven a converger.

4.Preguntas del Trabajo

4.1. Preguntas del Trabajo ¿Tiene sentido llevar a cabo, en este caso, un análisis de componentes principales? Para justificarlo, deberá llevar a cabo las pruebas que estime oportunas, como, por ejemplo el análisis de la matriz de correlaciones, el del determinante de dicha matriz, la prueba de esfericidad de Bartlett, el KMO o el MSA.

El análisis de ACP tiene como finalidad crear un algoritmo de aprendizaje automático no supervisado que intenta reducir la dimensionalidad (número de características) dentro de un conjunto de datos mientras retiene toda la información posible. Esto se realiza buscando un nuevo conjunto de características denominado componentes, que son los compuestos de las características originales que no son correlativas entre sí.

Para realizar un debido análisis ACP, recurriremos a:

- Análisis de la matriz de correlación
- Análisis de determinantes de la matriz de correlación

- Prueba de esfericidad de Bartlett
- Análisis KMO

Análisis de la matriz de correlación

Sirve para determinar si el valor obtenido (del coeficiente de correlación) muestra que las variables X e Y están relacionadas en realidad o tan solo presentan dicha relación como consecuencia del azar.

Para este apartado crearemos la matriz de correlaciones y la visualizaremos con la finalidad de:

- Comprobar la relación que existe entre el resto de variables independientes con el fin de generar subgrupos de variables cuya correlación sea parecida entre ellas.

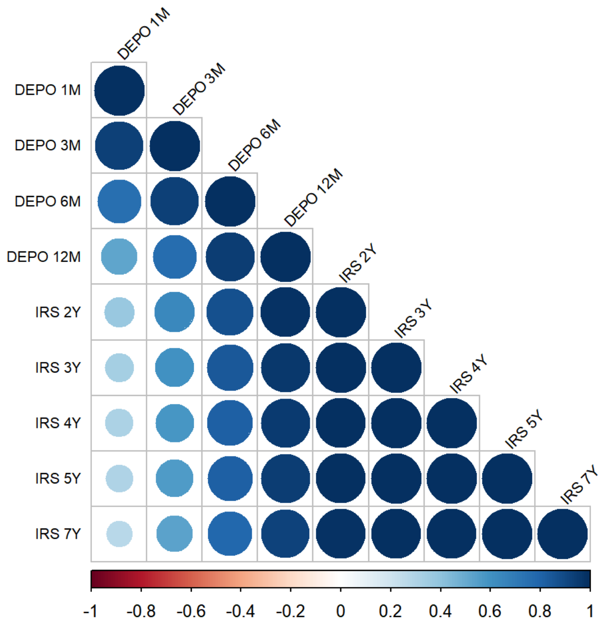
```
#Creamos la matriz de correlaciones

cor.mat = round(cor(bonos_act_sum),2)
cor.mat
```

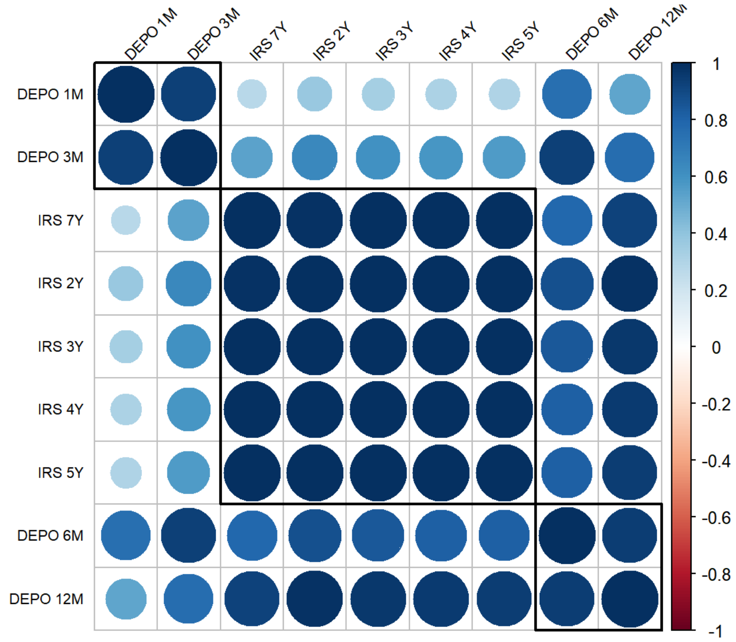
```
##          DEPO 1M DEPO 3M DEPO 6M DEPO 12M IRS 2Y IRS 3Y IRS 4Y IRS 5Y IRS 7Y
## DEPO 1M     1.00    0.93    0.75     0.52   0.37   0.33   0.31   0.30   0.27
## DEPO 3M     0.93    1.00    0.93     0.76   0.64   0.60   0.58   0.56   0.53
## DEPO 6M     0.75    0.93    1.00     0.94   0.87   0.84   0.82   0.81   0.78
## DEPO 12M    0.52    0.76    0.94     1.00   0.98   0.96   0.95   0.94   0.92
## IRS 2Y      0.37    0.64    0.87     0.98   1.00   1.00   0.99   0.99   0.98
## IRS 3Y      0.33    0.60    0.84     0.96   1.00   1.00   1.00   1.00   0.99
## IRS 4Y      0.31    0.58    0.82     0.95   0.99   1.00   1.00   1.00   0.99
## IRS 5Y      0.30    0.56    0.81     0.94   0.99   1.00   1.00   1.00   1.00
## IRS 7Y      0.27    0.53    0.78     0.92   0.98   0.99   0.99   1.00   1.00
```

```
#Visualizamos la matriz de correlaciones

#Visualización original
require(corrplot)
corrplot(cor.mat, type="lower", order="original",
         tl.col="black", tl.cex=0.7, tl.srt=45)  # las correlaciones positivas en azul, las negativas en rojo
```




```
#Visualización por cluster
corrplot(cor.mat, type="full", order="hclust", addrect = 3,
         tl.col="black", tl.cex=0.7, tl.srt=45) #permite visualizar clusters
```




```
#...  y también podemos visualizar un chart de correlaciones con el paquete PerformanceAnalytics, que cargamos
#install.packages("PerformanceAnalytics")
```

Como se puede observar en las gráficas anteriores, se puede observar que existe una clara correlación (positiva) entre las distintas variables.

Por otra parte, existe un aumento de la correlación entre variables que presentan duraciones parecidas, es por eso, que para el ACP, podemos afirmar que podríamos agrupas variables en función de su correlación (y por ense, de su duración).

En el segundo gráfico, hemos agrupado los 9 tipos de bonos y depósitos en tres grupos:

- Muy corto plazo: depósitos a uno y tres meses
- Corto plazo: depósitos a 6 y 12 meses
- Medio-Largo plazo: bonos a 2, 3, 4, 5 y 7 años

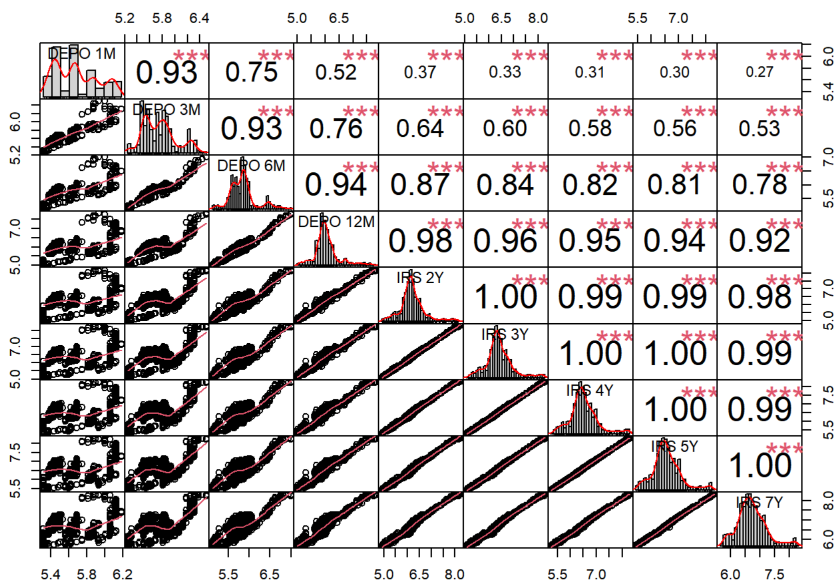
Mapa de correlacion compuesta:

En la siguiente ilustración, observamos en la diagonal la distribución de los tipos de interés de cada una de las variables. Por lo general presentan una distribución normal, aunque con cierta skewness para algunos casos, como el del bono a 7 años.

En el triángulo inferior, aparecen los diagramas de dispersión por cada uno de los pares de variables. En este apartado, los depósitos de uno y tres meses presentan más variabilidad de los datos frente al resto de variables, por el otro lado, los bonos que presentan una mayor duración tienen una distribucuión de los datos más homogénea (lo que podrían ser entre el conjunto de bonos que pertenecen al medio-largo plazo).

En el triángulo superior aparecen los valores de la correlación junto a unas estrellas que representan el nivel de significación (p-value) de cada una de las variables, a más distacia de duraciones, menor valor de correlación; pero en general, todos tienen un nivel alto de significación.

```
require(PerformanceAnalytics)
chart.Correlation(bonos_act_sum, histogram=TRUE, pch=19)
```

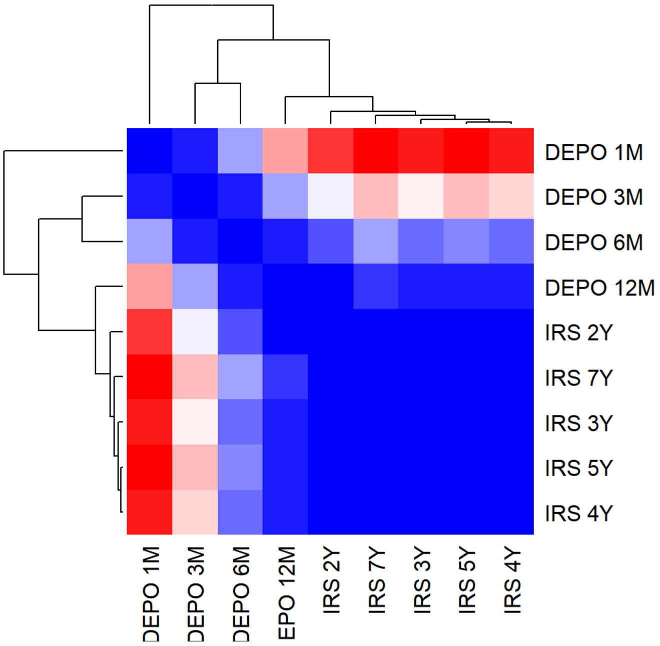


Mapa de calor de las correlaciones:

Este mapa de correlaciones trata de agrupar las diferentes variables según su nivel de correlación en función del número de divisiones que queremos realizar (hasta un máximo de 9, ya que carece de sentido agrupar las variables de una en una).

Al igual que en la gráfica de agrupación de variables, existe una clara agrupación en función de la duración de los activos.

```
col = colorRampPalette(c("red", "white", "blue"))(20) #definimos la paleta de colores;
heatmap(x = cor.mat, col = col, symm = TRUE)
```



Matriz de correlaciones parciales

La matriz de correlación parcial pretende calcular la relación entre dos variable, pero eliminando el efecto de una tercera variable.

La verdad es que no entiendo mucho su significado, pero se encarga de ilustrar el coecifiente de correlación de Pearson, junto con su p-valor y su estadístico t.

```
require(ppcor)

p.cor.mat=pcor(bonos_act_sum)
p.cor.mat
```

```
## $estimate
##             [,1]         [,2]          [,3]        [,4]        [,5]
##  [1,]  1.00000000  0.707921634 -0.0297992339 -0.24853227 -0.10366735
##  [2,]  0.70792163  1.000000000  0.6714673572 -0.09439923  0.05305015
##  [3,] -0.02979923  0.671467357  1.0000000000  0.62546084  0.12052413
##  [4,] -0.24853227 -0.094399233  0.6254608368  1.00000000  0.13622902
##  [5,] -0.10366735  0.053050150  0.1205241271  0.13622902  1.00000000
##  [6,]  0.11451356 -0.078402993 -0.1304108618  0.07393168  0.83334958
##  [7,] -0.09811168  0.030732122  0.0630240287 -0.06824871 -0.17691965
##  [8,]  0.07147222  0.005408987 -0.0009453282 -0.06006821 -0.14742675
##  [9,] -0.01941839 -0.013465519 -0.0029929731  0.10423106 -0.09413565
##             [,6]        [,7]          [,8]         [,9]
##  [1,]  0.11451356 -0.09811168  0.0714722190 -0.019418385
##  [2,] -0.07840299  0.03073212  0.0054089871 -0.013465519
##  [3,] -0.13041086  0.06302403 -0.0009453282 -0.002992973
##  [4,]  0.07393168 -0.06824871 -0.0600682090  0.104231062
##  [5,]  0.83334958 -0.17691965 -0.1474267530 -0.094135652
##  [6,]  1.00000000  0.59143124  0.0594809490 -0.172140050
##  [7,]  0.59143124  1.00000000  0.5790864941  0.040368805
##  [8,]  0.05948095  0.57908649  1.0000000000  0.690068875
##  [9,] -0.17214005  0.04036880  0.6900688750  1.000000000
##
## $p.value
##               [,1]          [,2]         [,3]         [,4]          [,5]
##  [1,]  0.000000e+00 5.166725e-114 4.173193e-01 6.366931e-12  4.674709e-03
##  [2,] 5.166725e-114  0.000000e+00 1.546203e-98 1.003689e-02  1.485632e-01
##  [3,]  4.173193e-01  1.546203e-98 0.000000e+00 6.566021e-82  9.957497e-04
##  [4,]  6.366931e-12  1.003689e-02 6.566021e-82 0.000000e+00  1.957350e-04
##  [5,]  4.674709e-03  1.485632e-01 9.957497e-04 1.957350e-04  0.000000e+00
##  [6,]  1.769316e-03  3.261404e-02 3.652621e-04 4.394697e-02 5.720708e-193
##  [7,]  7.444279e-03  4.028826e-01 8.602941e-02 6.297596e-02  1.217975e-06
##  [8,]  5.148650e-02  8.829811e-01 9.794772e-01 1.018267e-01  5.486847e-05
##  [9,]  5.971763e-01  7.140352e-01 9.350878e-01 4.453712e-03  1.024836e-02
##               [,6]         [,7]          [,8]          [,9]
##  [1,]  1.769316e-03 7.444279e-03  5.148650e-02  5.971763e-01
##  [2,]  3.261404e-02 4.028826e-01  8.829811e-01  7.140352e-01
##  [3,]  3.652621e-04 8.602941e-02  9.794772e-01  9.350878e-01
##  [4,]  4.394697e-02 6.297596e-02  1.018267e-01  4.453712e-03
##  [5,] 5.720708e-193 1.217975e-06  5.486847e-05  1.024836e-02
##  [6,]  0.000000e+00 2.691844e-71  1.052241e-01  2.363417e-06
##  [7,]  2.691844e-71 0.000000e+00  9.458846e-68  2.717815e-01
##  [8,]  1.052241e-01 9.458846e-68  0.000000e+00 3.799153e-106
##  [9,]  2.363417e-06 2.717815e-01 3.799153e-106  0.000000e+00
##
## $statistic
##            [,1]       [,2]        [,3]      [,4]      [,5]      [,6]
##  [1,]  0.0000000 27.2841622 -0.81153474 -6.984524 -2.837249  3.137851
##  [2,] 27.2841622  0.0000000 24.66583348 -2.581198  1.446131 -2.140823
##  [3,] -0.8115347 24.6658335  0.00000000 21.820885  3.304917 -3.580533
##  [4,] -6.9845245 -2.5811978 21.82088460  0.000000  3.743230  2.018040
##  [5,] -2.8372486  1.4461312  3.30491678  3.743230  0.000000 41.040296
##  [6,]  3.1378515 -2.1408226 -3.58053278  2.018040 41.040296  0.000000
##  [7,] -2.6836766  0.8369641  1.71901433 -1.862161 -4.893174 19.965806
##  [8,]  1.9505562  0.1472419 -0.02573309 -1.638094 -4.057486  1.622022
##  [9,] -0.5286937 -0.3665824 -0.08147303  2.852846 -2.573926 -4.756887
##            [,7]        [,8]        [,9]
##  [1,] -2.6836766  1.95055617 -0.52869367
##  [2,]  0.8369641  0.14724190 -0.36658236
##  [3,]  1.7190143 -0.02573309 -0.08147303
##  [4,] -1.8621615 -1.63809360  2.85284572
##  [5,] -4.8931738 -4.05748636 -2.57392607
##  [6,] 19.9658060  1.62202155 -4.75688705
##  [7,]  0.0000000 19.33539951  1.09978845
##  [8,] 19.3353995  0.00000000 25.95474373
##  [9,]  1.0997885 25.95474373  0.00000000
##
## $n
## [1] 750
##
## $gp
## [1] 7
##
## $method
## [1] "pearson"
```

Determinante de la matriz de correlación

Un determinante bajo, indica alta multicolinealidad entre las variables. De ser igual a cero, indicaría que algunas de las variables son linealmente dependientes y no se podrían realizar ciertos cálculos necesarios para los procedimientos multivariados.

```
det(cor.mat)
```

```
## [1] 4.682e-15
```

En este caso observamos que es muy cercano a 0, lo que sugiere un alto nivel de colinealidad en el conjunto de variables involucradas en la matriz.

Prueba de esfericidad de Bartlett

Se utiliza para probar la Hipótesis Nula que afirma que las variables no están correlacionadas en la población. Es decir, comprueba si la matriz de correlaciones es una matriz de identidad.

Prueba de esfericidad de Bartlett:

Si Sig. (p-valor) < 0.05 aceptamos H0 (hipótesis nula) > se puede aplicar el análisis factorial.

Si Sig. (p-valor) > 0.05 rechazamos H0 > no se puede aplicar el análisis factorial.

```
bartlett.test(bonos_act_sum)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  bonos_act_sum
## Bartlett's K-squared = 1068.9, df = 8, p-value < 2.2e-16
```

En este caso el valor es cercano a cero, por lo tanto, se rechaza la Hipótesis Nula y se continúa con el Análisis.

KMO o el MSA.

El test KMO relaciona los coeficientes de correlación, observados entre las variables. Cuanto más cerca de 1 tenga el valor obtenido del test KMO, implica que la relación entre las variables es alta.

- Si KMO ≥ 0.9, el test es muy bueno;
- notable para KMO ≥ 0.8;
- mediano para KMO ≥ 0.7;
- bajo para KMO ≥ 0.6;
- uy bajo para KMO < 0.5.

```
library(psych)
KMO(bonos_act_sum)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: KMO(r = bonos_act_sum)
## Overall MSA =  0.87
## MSA for each item = 
##  DEPO 1M  DEPO 3M  DEPO 6M DEPO 12M   IRS 2Y   IRS 3Y   IRS 4Y   IRS 5Y 
##     0.79     0.80     0.87     0.93     0.88     0.84     0.89     0.87 
##   IRS 7Y 
##     0.92 
```

La media de la relación entre las variables es de 0.87 por lo que se puede afirmar que la relación entre variables es alta, lo que implica que se puede aplicar el analisis factorial a las variables utilizadas.

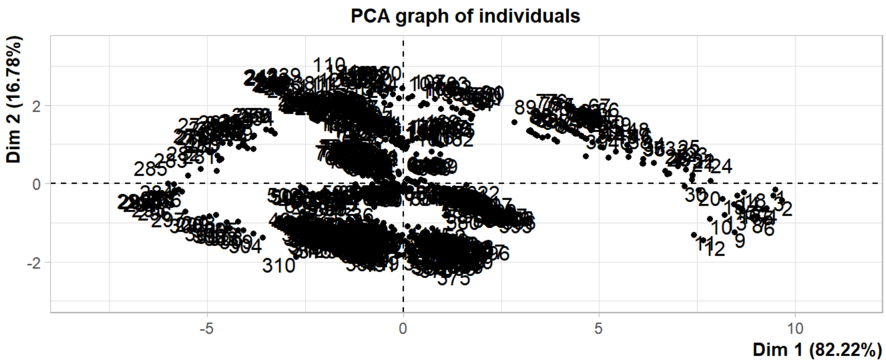
4.2 ¿Cuántos componentes permitirían explicar, adecuadamente, la estructura subycente de los tipos de interés aquí analizados? Justifique su respuesta empleando, por ejemplo, las pruebas de la varianza explicada o del gráfico de sedimentación.

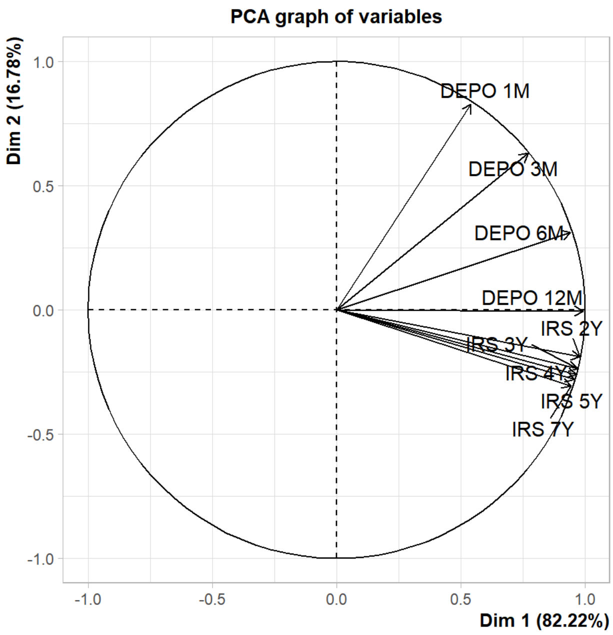
PRUEBA DE LA VARIANZA EXPLICADA

Este apartado tiene como finalidad el poder explicar el porcentaje varianza explicada de cada componente (variables o tipo de activo) con respecto del total con el fin de poder elegir el número de componentes que pueden explicar la mayor varianza posible con el menor número de variables.

También hablaremos de las componentes principales, que no son más que los nuevos ejes proyectados a partir de los ejes originales. Los nuevos puntos generados (los puntos de los ejes principales que se transforman en puntos de los ejes proyectados) se denominan puntuaciones de las componentes principales.

```
acp = PCA(bonos_act_sum, graph=T)
```






```
acp
```

```
## **Results for the Principal Component Analysis (PCA)**
## The analysis was performed on 750 individuals, described by 9 variables
## *The results are available in the following objects:
##
##    name               description
## 1  "$eig"             "eigenvalues"
## 2  "$var"             "results for the variables"
## 3  "$var$coord"       "coord. for the variables"
## 4  "$var$cor"         "correlations variables - dimensions"
## 5  "$var$cos2"        "cos2 for the variables"
## 6  "$var$contrib"     "contributions of the variables"
## 7  "$ind"             "results for the individuals"
## 8  "$ind$coord"       "coord. for the individuals"
## 9  "$ind$cos2"        "cos2 for the individuals"
## 10 "$ind$contrib"     "contributions of the individuals"
## 11 "$call"            "summary statistics"
## 12 "$call$centre"     "mean of the variables"
## 13 "$call$ecart.type" "standard error of the variables"
## 14 "$call$row.w"      "weights for the individuals"
## 15 "$call$col.w"      "weights for the variables"
```

```
# extraemos los autovalores y observamos la varianza explicada

acp$eig # con FacotMineR
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## comp 1 7.399851e+00           8.222057e+01                          82.22057
## comp 2 1.510459e+00           1.678287e+01                          99.00344
## comp 3 6.789677e-02           7.544086e-01                          99.75785
## comp 4 1.239343e-02           1.377047e-01                          99.89556
## comp 5 5.774959e-03           6.416621e-02                          99.95972
## comp 6 2.951861e-03           3.279845e-02                          99.99252
## comp 7 4.533614e-04           5.037349e-03                          99.99756
## comp 8 1.415903e-04           1.573225e-03                          99.99913
## comp 9 7.823462e-05           8.692735e-04                         100.00000
```

```
get_eig(acp) #con factoextra
```

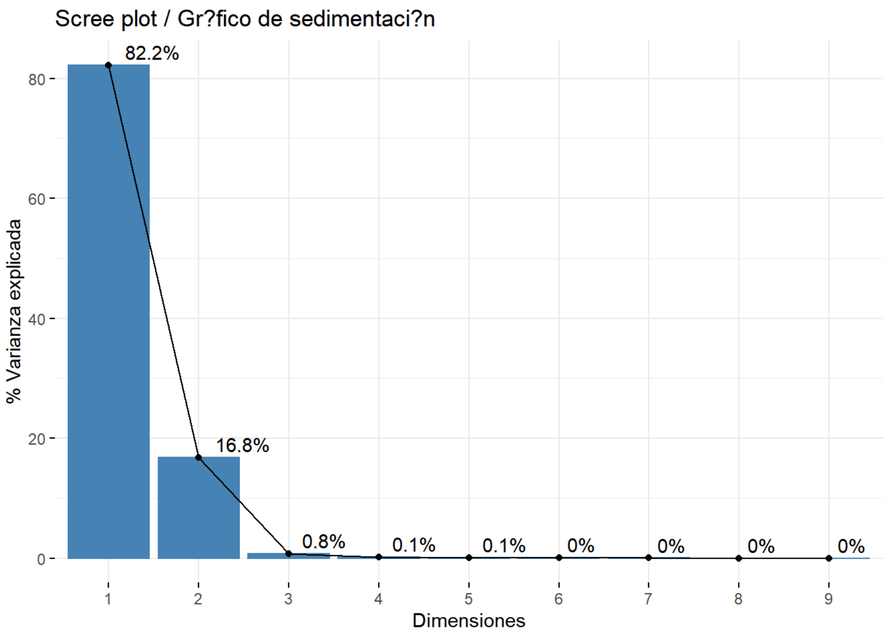
```
##         eigenvalue variance.percent cumulative.variance.percent
## Dim.1 7.399851e+00     8.222057e+01                    82.22057
## Dim.2 1.510459e+00     1.678287e+01                    99.00344
## Dim.3 6.789677e-02     7.544086e-01                    99.75785
## Dim.4 1.239343e-02     1.377047e-01                    99.89556
## Dim.5 5.774959e-03     6.416621e-02                    99.95972
## Dim.6 2.951861e-03     3.279845e-02                    99.99252
## Dim.7 4.533614e-04     5.037349e-03                    99.99756
## Dim.8 1.415903e-04     1.573225e-03                    99.99913
## Dim.9 7.823462e-05     8.692735e-04                   100.00000
```

Información clave de las gráficas anteriores:

- La gráfica de puntos individual explica para cada punto, el pocentaje explicado por la dimension uno y dos.
- El gráfico de 'PCA graph of Variables' indica las coordenadas de cada vaiable respecto de cada variables, en el caso del depósito a un mes, la coordenada respecto de la dimensión 1 es del 0,52 y del 0,80 respecto de la segunda dimensión.
- Con una sola dimensión (o eje) (x), se puede explicar el 82,22% de la varianza total, pero no se llega a explicar toda la varianza, para ello, utilizamos la siguiente dimensión (o eje) que recoja el máximo de varianza no explicada por la dimensión 1; la dimensión 2 (eje x e y), recoge el 16,78% del total de la varianza que la dimensión 1 no ha podido explicar. Uniendo ambas dimensiones se llega al 99,003% de la varianza explicada.

Hacemos el screeplot del porcentaje de varianza explicada por cada una de las dimensiones.

```
fviz_eig(acp, addlabels=TRUE, hjust = -0.3)+
  labs(title="Scree plot / Gr?fico de sedimentaci?n", x="Dimensiones", y="% Varianza explicada")
```



Se puede afirmar que con utilizar solo dos dimensiones podemos explicar la varianza del modelo, porque la tercera dimensión aporta poca varianza explicada al total (menos del 1%).

Relación de las variables con los CCPP:

- var$coord <- indica las coordenadas de cada variable respecto de su dimensión
- var$contrib <- indica la contribución de cada variable a los Componentes Principales
- var$cor <- correlaciones de las variables
- var$cos2 <- calidad de la representación para las variables sobre el mapa factorial

```
var=get_pca_var(acp) #factoextra
var
```

```
## Principal Component Analysis Results for variables
##  ===================================================
##   Name       Description                                    
## 1 "$coord"   "Coordinates for the variables"                
## 2 "$cor"     "Correlations between variables and dimensions"
## 3 "$cos2"    "Cos2 for the variables"                       
## 4 "$contrib" "contributions of the variables"
```

```
var$coord
```

```
##              Dim.1        Dim.2        Dim.3        Dim.4        Dim.5
## DEPO 1M  0.5415951  0.826710181  0.144735692  0.045916754 -0.012964813
## DEPO 3M  0.7713679  0.630663373 -0.046762429 -0.056896675  0.039436818
## DEPO 6M  0.9425262  0.312069637 -0.108081095 -0.020475324 -0.027051215
## DEPO 12M 0.9916011 -0.005098406 -0.116294316  0.028099181 -0.035405376
## IRS 2Y   0.9795366 -0.188990539 -0.036888058  0.047452064  0.031343481
## IRS 3Y   0.9711878 -0.235529523  0.003715993  0.027412700  0.021498611
## IRS 4Y   0.9641979 -0.261706492  0.039663436  0.001999582  0.009500856
## IRS 5Y   0.9572758 -0.279697658  0.069991624 -0.018115370 -0.001933562
## IRS 7Y   0.9440004 -0.306506929  0.108208980 -0.050037228 -0.022971553
```

```
var$contrib
```

```
##              Dim.1        Dim.2       Dim.3       Dim.4       Dim.5
## DEPO 1M   3.963934 45.247828582 30.85333765 17.01182747  2.91060728
## DEPO 3M   8.040817 26.332154195  3.22066084 26.12055404 26.93114434
## DEPO 6M  12.005047  6.447542318 17.20482810  3.38275200 12.67140122
## DEPO 12M 13.287737  0.001720917 19.91901438  6.37082856 21.70648567
## IRS 2Y   12.966368  2.364674111  2.00411411 18.16848941 17.01161481
## IRS 3Y   12.746279  3.672669731  0.02033765  6.06334436  8.00335133
## IRS 4Y   12.563465  4.534403497  2.31702938  0.03226169  1.56306315
## IRS 5Y   12.383722  5.179273341  7.21511095  2.64790884  0.06473921
## IRS 7Y   12.042631  6.219733308 17.24556694 20.20203363  9.13759298
```

```
var$cor
```

```
##              Dim.1        Dim.2        Dim.3        Dim.4        Dim.5
## DEPO 1M  0.5415951  0.826710181  0.144735692  0.045916754 -0.012964813
## DEPO 3M  0.7713679  0.630663373 -0.046762429 -0.056896675  0.039436818
## DEPO 6M  0.9425262  0.312069637 -0.108081095 -0.020475324 -0.027051215
## DEPO 12M 0.9916011 -0.005098406 -0.116294316  0.028099181 -0.035405376
## IRS 2Y   0.9795366 -0.188990539 -0.036888058  0.047452064  0.031343481
## IRS 3Y   0.9711878 -0.235529523  0.003715993  0.027412700  0.021498611
## IRS 4Y   0.9641979 -0.261706492  0.039663436  0.001999582  0.009500856
## IRS 5Y   0.9572758 -0.279697658  0.069991624 -0.018115370 -0.001933562
## IRS 7Y   0.9440004 -0.306506929  0.108208980 -0.050037228 -0.022971553
```

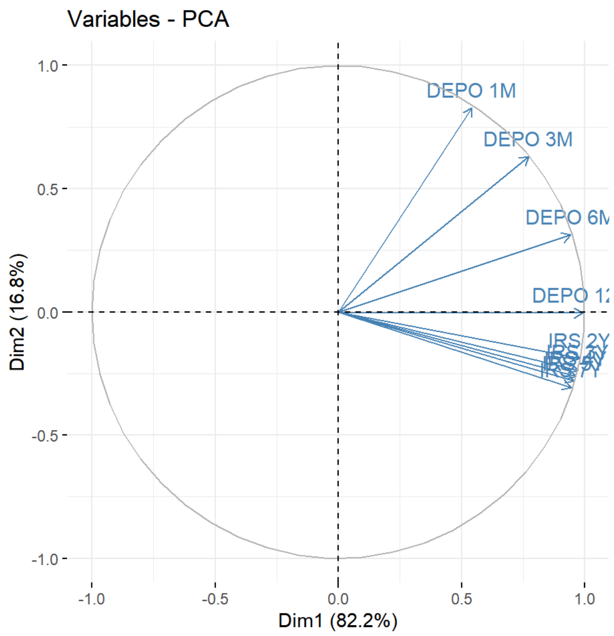
```
var$cos2
```

```
##              Dim.1        Dim.2        Dim.3        Dim.4        Dim.5
## DEPO 1M  0.2933252 6.834497e-01 2.094842e-02 2.108348e-03 1.680864e-04
## DEPO 3M  0.5950085 3.977363e-01 2.186725e-03 3.237232e-03 1.555263e-03
## DEPO 6M  0.8883556 9.738746e-02 1.168152e-02 4.192389e-04 7.317682e-04
## DEPO 12M 0.9832728 2.599375e-05 1.352437e-02 7.895640e-04 1.253541e-03
## IRS 2Y   0.9594919 3.571742e-02 1.360729e-03 2.251698e-03 9.824138e-04
## IRS 3Y   0.9432057 5.547416e-02 1.380861e-05 7.514561e-04 4.621903e-04
## IRS 4Y   0.9296777 6.849029e-02 1.573188e-03 3.998328e-06 9.026626e-05
## IRS 5Y   0.9163770 7.823078e-02 4.898827e-03 3.281666e-04 3.738663e-06
## IRS 7Y   0.8911368 9.394650e-02 1.170918e-02 2.503724e-03 5.276923e-04
```

Gráficas de:

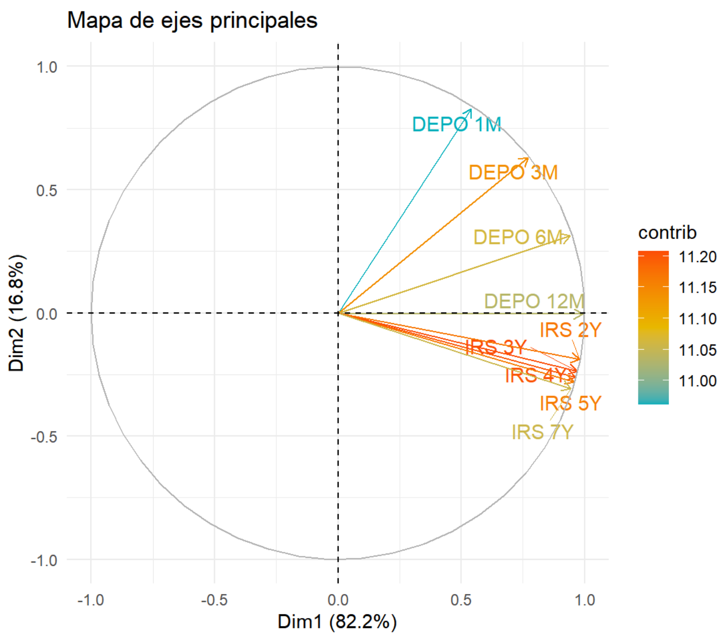
- Coordenadas de cada variable respecto de sus dimensiones
- Contribución de las variables a cada dimensión
- Calidad de la representación para las variables sobre el mapa factorial

```
# Gr?fico por defecto de las variables en el espacio de dos CCPP ...
fviz_pca_var(acp, col.var = "steelblue")
```

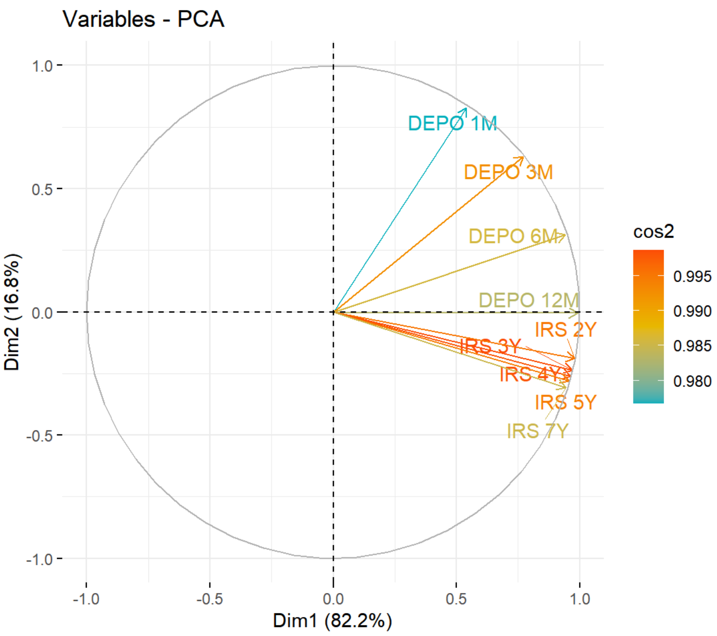



```
## ... del que modificamos los colores para visualizar mejor la contribuci?n de la variable en el eje principal

fviz_pca_var(acp, col.var="contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), repel=TRUE) +
  labs(title="Mapa de ejes principales")+
  theme_minimal()
```



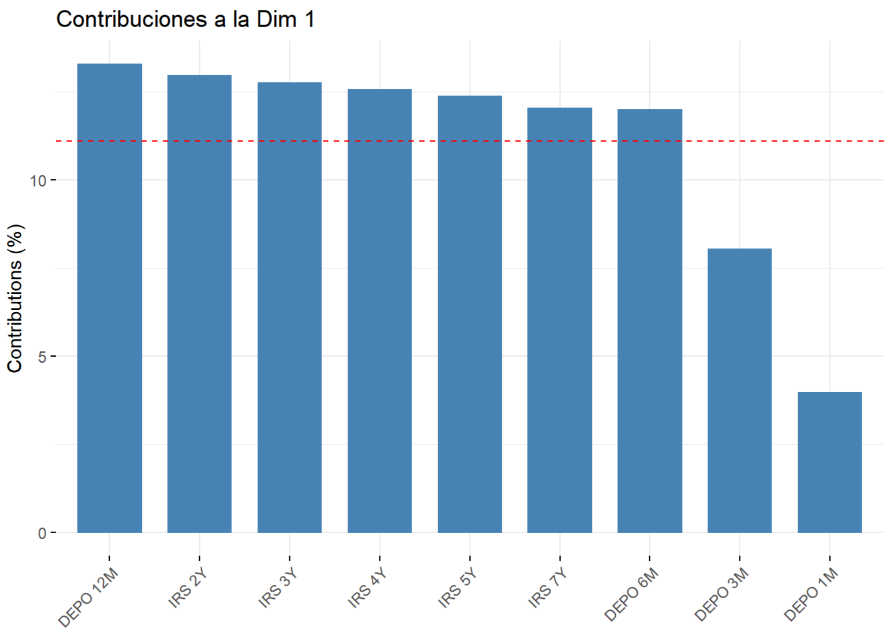

```
fviz_pca_var(acp, col.var = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE # Avoid text overlapping
)
```



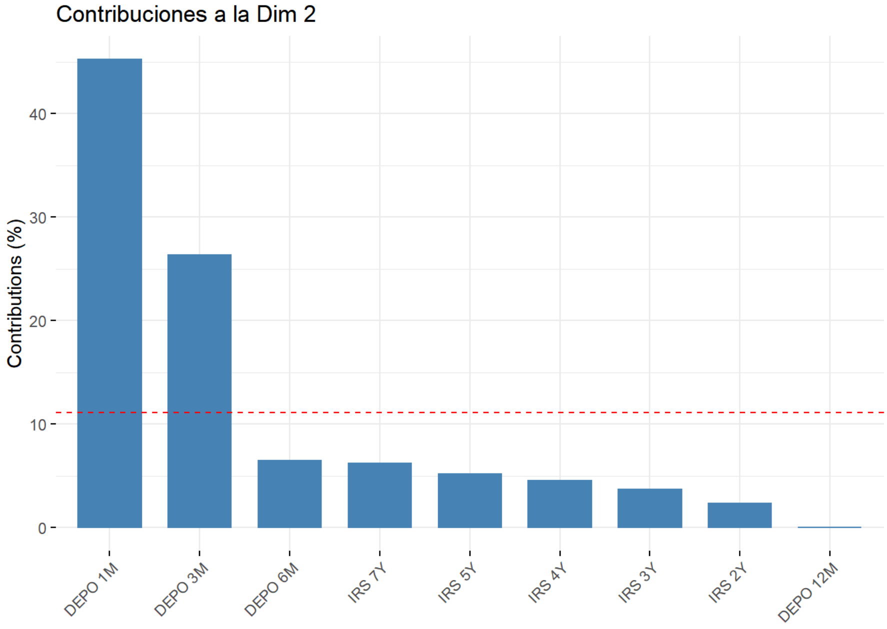
Gráficas de:

- Contribución de las variables a la dimensión 1
- Contribución de las variables a la dimensión 2
- Contribución acumulada de las variables a ambas dimensiones

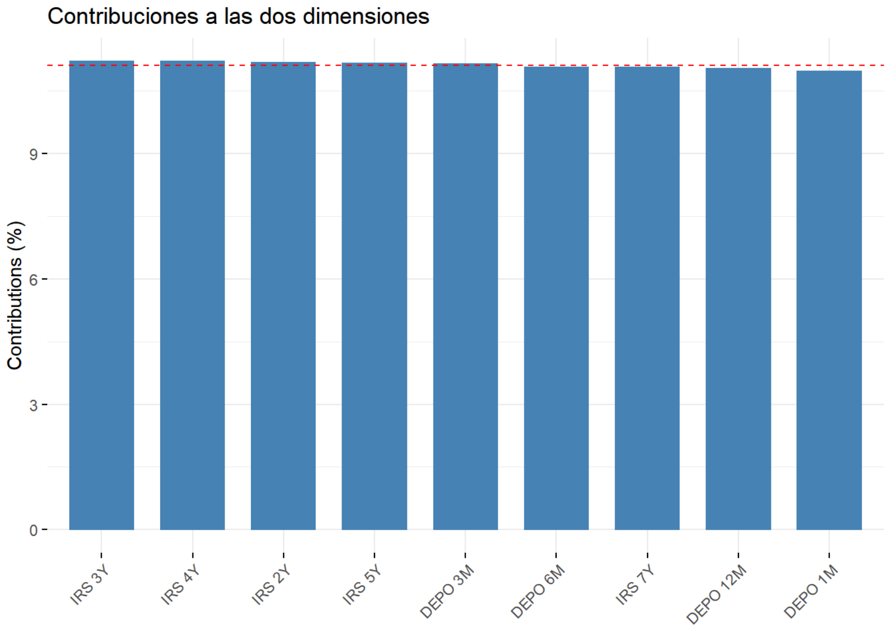
```
# Contribuci?n de las variables a cada uno de los ejes principales
## eje 1
fviz_contrib(acp, choice="var", axes = 1 )+
  labs(title = "Contribuciones a la Dim 1")
```




```
## eje 2
fviz_contrib(acp, choice="var", axes = 2 )+
  labs(title = "Contribuciones a la Dim 2")
```




```
## ambos ejes
fviz_contrib(acp, choice="var", axes = 1:2)+
  labs(title = "Contribuciones a las dos dimensiones")
```



Coordenadas de los individuos (observaciones) respecto de cada dimensión

```
# Extracción de resultados para individuos
#----------------------------------------
ind = get_pca_ind(acp)
ind
```

```
## Principal Component Analysis Results for individuals
##  ===================================================
##   Name       Description
## 1 "$coord"   "Coordinates for the individuals"
## 2 "$cos2"    "Cos2 for the individuals"
## 3 "$contrib" "contributions of the individuals"
```

```
#Coordenadas de los individuos / observaciones en el plano de los ejes principales
head(ind$coord)
```

```
##      Dim.1      Dim.2      Dim.3      Dim.4       Dim.5
## 1 9.477480 -0.1497580 -0.7817695 -0.0591335 0.089964900
## 2 9.650647 -0.4254062 -0.8993346 -0.1463232 0.097879300
## 3 9.438370 -0.3067392 -0.9536963 -0.2033538 0.001010778
## 4 9.257406 -0.6364965 -0.7571066 -0.1982038 0.109587059
## 5 9.169490 -0.6678323 -0.7933781 -0.3050298 0.056313067
## 6 9.190861 -0.9289690 -0.6210234 -0.1202699 0.058518120
```

Visualización del apartado anterior:

```
fviz_pca_ind(acp, repel = T, col.ind = "cos2")+
  scale_color_gradient2(low="blue", mid="white",
                      high="red", midpoint=0.6)+
  theme_minimal()
```

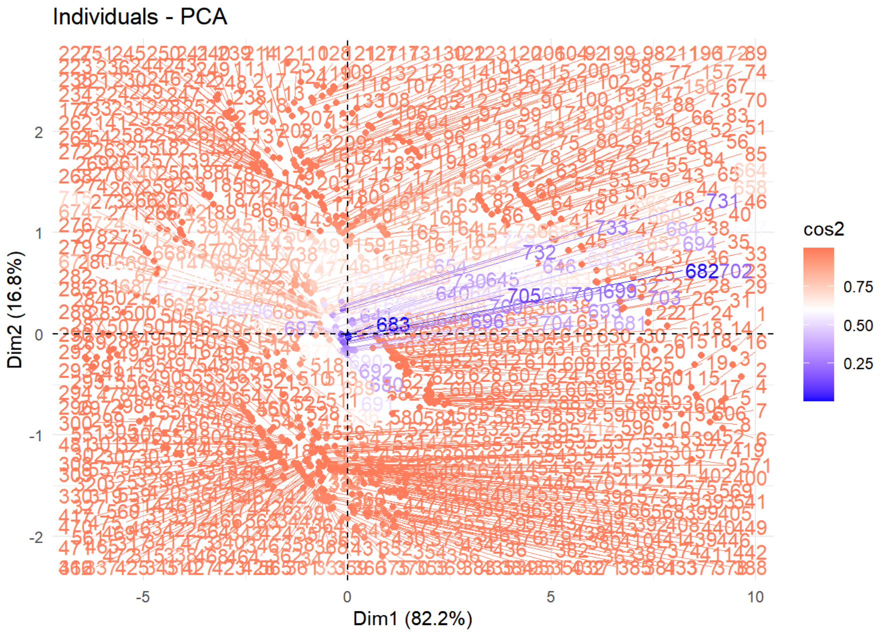
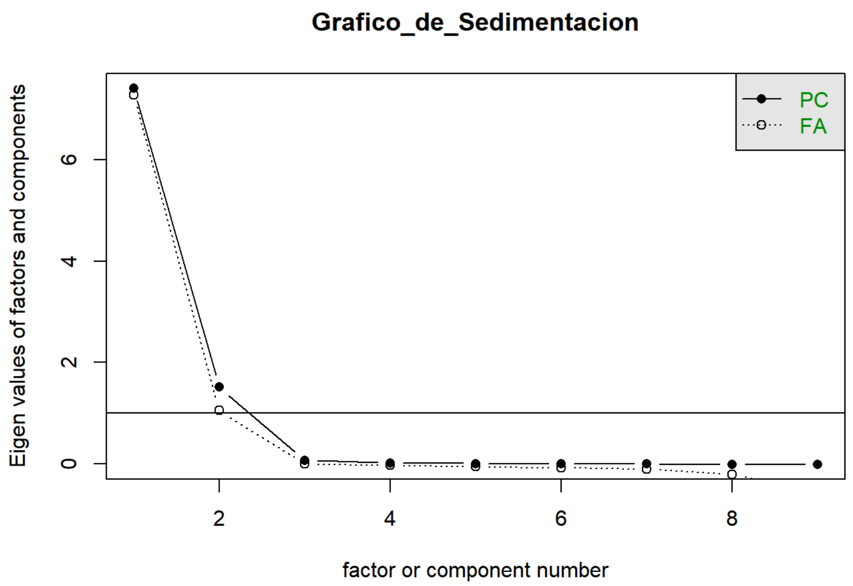


GRAFICO DE SEDIMENTACION

La gráfica de sedimentación muestra el número del componente principal versus su valor propio correspondiente. La gráfica de sedimentación ordena los valores propios desde el más grande hasta el más pequeño. Los valores propios de la matriz de correlación son iguales a las varianzas de los componentes principales.

La gráfica de sedimentación se utiliza para seleccionar el número de componentes que se usarán con base en el tamaño de los valores propios.

```
scree(cor.mat,main ="Grafico_de_Sedimentacion")
```



Esta gráfica de sedimentación muestra que los valores propios comienzan a formar una línea recta después del tercer componente principal. Por lo tanto, los componentes principales restantes explican una proporción muy pequeña de la variabilidad (cercana a cero) y probablemente carezcan de importancia.

4.3. Finalmente, ¿tiene sentido llevar a cabo una rotación de las variables subyacentes? Para responder, lleva a cabo una rotación Varimax, por ejemplo.

Varimax

Método de rotación ortogonal que minimiza el número de variables que tienen saturaciones altas en cada factor. Simplifica la interpretación de los factores.

```
factanal(bonos_act_sum, factors = 3, rotation = 'varimax')
```

```
##
## Call:
## factanal(x = bonos_act_sum, factors = 3, rotation = "varimax")
##
## Uniquenesses:
##  DEPO 1M  DEPO 3M  DEPO 6M DEPO 12M   IRS 2Y   IRS 3Y   IRS 4Y   IRS 5Y
##    0.033    0.005    0.005    0.005    0.005    0.005    0.005    0.005
##   IRS 7Y
##    0.005
##
## Loadings:
##          Factor1 Factor2 Factor3
## DEPO 1M          0.978
## DEPO 3M   0.365   0.924
## DEPO 6M   0.665   0.726   0.161
## DEPO 12M  0.861   0.470   0.181
## IRS 2Y    0.944   0.305   0.107
## IRS 3Y    0.962   0.263
## IRS 4Y    0.970   0.239
## IRS 5Y    0.974   0.221
## IRS 7Y    0.978   0.194
##
##                Factor1 Factor2 Factor3
## SS loadings      5.988   2.864   0.089
## Proportion Var   0.665   0.318   0.010
## Cumulative Var   0.665   0.984   0.994
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 5565.25 on 12 degrees of freedom.
## The p-value is 0
```

Rotaciones como las que produce varimax facilitan la interpretabilidad de los factores. Tienden a aplastar los coeficientes pequeños y a hacer crecer los grandes de manera que sus perfiles pueden asociarse más fácilmente a un subconjunto concreto de variables.

```
library(pls)
library(imputeTS)

rm(list=ls())

bonos <- read_delim("ACPTIUSD.csv", ";",
                    escape_double = FALSE, trim_ws = TRUE)

bonos = bonos[complete.cases(bonos), ]
bonos
```

X1 <chr>	DEPO 1M <dbl>	DEPO 3M <dbl>	DEPO 6M <dbl>	DEPO 12M <dbl>	IRS 2Y <dbl>	IRS 3Y <dbl>	IRS 4Y <dbl>	IRS 5Y <dbl>
02/01/1995	6.000	6.500	7.000	7.750	8.170	8.240	8.250	8.220
03/01/1995	5.938	6.500	7.000	7.813	8.220	8.280	8.290	8.290
04/01/1995	5.938	6.500	7.000	7.813	8.130	8.200	8.230	8.240
05/01/1995	5.898	6.438	6.938	7.688	8.145	8.220	8.260	8.270
06/01/1995	5.875	6.438	6.938	7.688	8.070	8.200	8.250	8.270
09/01/1995	5.875	6.375	6.875	7.688	8.135	8.250	8.290	8.310
10/01/1995	5.875	6.375	6.875	7.688	8.090	8.210	8.250	8.250
11/01/1995	5.875	6.336	6.875	7.625	8.070	8.200	8.230	8.240
12/01/1995	5.813	6.250	6.758	7.563	8.020	8.150	8.190	8.220
13/01/1995	5.813	6.250	6.750	7.523	7.800	7.960	8.030	8.050

1-10 of 783 rows | 1-9 of 11 columns Previous **1** 2 3 4 5 6 ... 79 Next

```
bonos$Fechas = as.Date(bonos$X1, format = "%d/%m/%Y")

TIUSD <- bonos

training <- TIUSD[1:750, 2:11]
test <- TIUSD[750:783, 2:11]

colnames(training)[10] <- 'IRS_10Y'
colnames(test)[10] <- 'IRS_10Y'


set.seed(123)
modelo_pcr <- pcr(formula = IRS_10Y ~ ., data = training, scale. = TRUE, validation = "CV")
modelo_pcr_CV <- MSEP(modelo_pcr, estimate = "CV")
min(modelo_pcr_CV$val)
```

```
## [1] 0.0003467249
```

```
# Test-MSE
predicciones <- predict(modelo_pcr, newdata = test, ncomp = 2)
test_mse <- mean((predicciones - test$IRS_10Y)^2)
test_mse
```

```
## [1] 0.07101432
```

4.4.Por último, deberá elaborar las oportunas conclusiones.

PREGUNTA 1:

El ACP es una técnica utilizada para describir un conjunto de datos en términos de nuevas variables («componentes») no correlacionadas. El ACP se emplea sobre todo en análisis exploratorio de datos y para construir modelos predictivos.

Como hemos podido observar, las componentes no se encontraban correlacionadas, por lo que se podría efectuar un ACP sin problemas, además la finalidad de este trabajo era la de predecir los valores del bono a 10 años (finalidad compartida también por la técnica ACP).

- Observamos que se pueden agrupar las variables en función de su correlación
- Según el Determinante de la matriz de correlación, que es cercano a 0 indicaría que existe multicolinealidad y no se podría seguir con el estudio, pero hemos usado otros modelos de validación que si nos indican que el método ACP se puede realizar.
- La Prueba de esfericidad de Bartlett nos indica, con un valor cercano a 0 que existe correlación con la variable independiente.
- El test KMO tiene un valor alto, lo que implica que se puede aplicar el analisis factorial a las variables utilizadas.

PREGUNTA 2:

- Como se ha podido comprobar en el estudio de la varianza y la gráfica de sedimentación, el número idóneo de componentes debe ser 2, los cuales son capaces de ecplicar el 99,003% de la varianza total.

PREGUNTA 3:

- Varimax rota los factores obtenidos, al rotarlos, hemos pasado de una varianza explicada con la primera dimensión de 82,2% al 66,5% (lo que perdemos un 15,7% de varianza explicada). Y si tenemos en cuenta el segundo factor, la varianza explicada (acumulada) baja del 99,003% al 98,4%. Por lo qu eno tiene sentido realizar la rotación Varimax.

5. Bibliografía

Análisis de Componentes Principales en R:

- http://rstudio-pubs-static.s3.amazonaws.com/2387_ee73f69e5ba04498a93e5bb12f636591.html (http://rstudio-pubs-static.s3.amazonaws.com/2387_ee73f69e5ba04498a93e5bb12f636591.html)
- https://www.cienciadedatos.net/documentos/35_principal_component_analysis (https://www.cienciadedatos.net/documentos/35_principal_component_analysis)
- https://rpubs.com/Csanchez15/551258 (https://rpubs.com/Csanchez15/551258)
- https://rpubs.com/Joaquin_AR/287787 (https://rpubs.com/Joaquin_AR/287787)

Estadística y Machine Learning con R:

- https://rpubs.com/PacoParra/293407# (https://rpubs.com/PacoParra/293407#):~:text=La%20prueba%20de%20esfericidad%20de,modelo%20factorial%20no%20ser%C3%ADa%20pertinente.
- Introducción a los Métodos Multivariantes:
- https://rpubs.com/marcelo-chavez/multivariado_1 (https://rpubs.com/marcelo-chavez/multivariado_1)

Interpretar todos los estadísticos y gráficas para Análisis de componentes principales:

- https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/# (https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/#):~:text=La%20gr%C3%A1fica%20de%20sedimentaci%C3%B3n%20muestra%20el%20n%C3%BAmero%20del%20componente%20principal,varianzas%