

# Tipos de Interés

Isabel Afán de Ribera

10/11/2020

## Introducción

El Análisis de Componentes Principales es una técnica de interdependencia que tiene como objetivo reducir el número de variables originales hasta un número menor de componentes de forma que cada uno de ellos forma un índice de las variables originales; el número de componentes mantenidos debe recoger la mayor parte posible de la varianza de los datos. Por su parte el Análisis Factorial trata de identificar la estructura interna, subyacente, que explica la interacción entre las variables. A través de esta práctica vamos a verificar la idoneidad entre estos dos métodos ACP – ANFAC.

## Descripción de la base de datos

Vamos a trabajar con el dataset ACPTIUSD compuesto por 978 observaciones y 11 variables relativas a la información sobre rendimientos de 10 bonos americanos a distintos plazos entre el 2 de enero de 1995 y el 30 de septiembre de 1998.

Variables:

- X: fecha
- DEPO 1M: depósito a 1 mes
- DEPO 3M: depósito a 3 mes
- DEPO 6M: depósito a 6 mes
- DEPO 12M: depósito a 12 mes
- IRS 2Y: Interest Rate Swap 2 años
- IRS 3Y: Interest Rate Swap 3 años
- IRS 4Y: Interest Rate Swap 4 años
- IRS 5Y: Interest Rate Swap 5 años
- IRS 7Y: Interest Rate Swap 7 años
- IRS 10Y: Interest Rate Swap 10 años

## Objetivo del trabajo

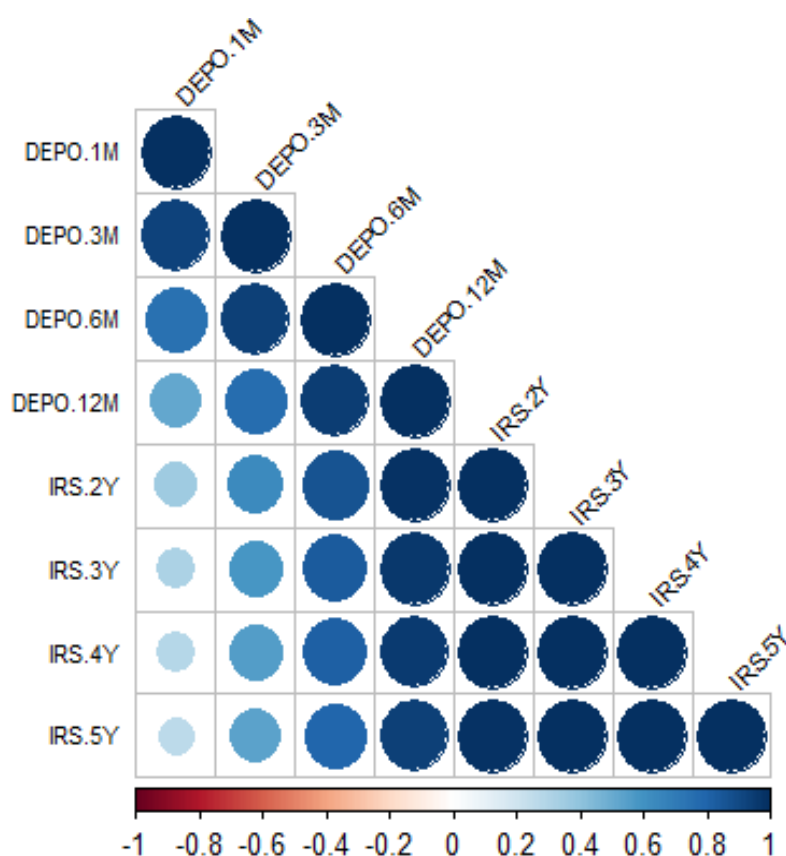
El objetivo de este análisis es efectuar una comprobación empírica mediante la aplicación del ACP al conjunto de datos descritos anteriormente. Pretendiendo verificar si, tal y como plantean los estudios teóricos, puede establecerse una estructura

subyacente que sintetice y agrupe los distintos plazos en virtud de sus características comunes. Para ello, utilizaremos las 949 primeras observaciones (denominadas observaciones activas) y las 9 primeras variables (las variables activas).

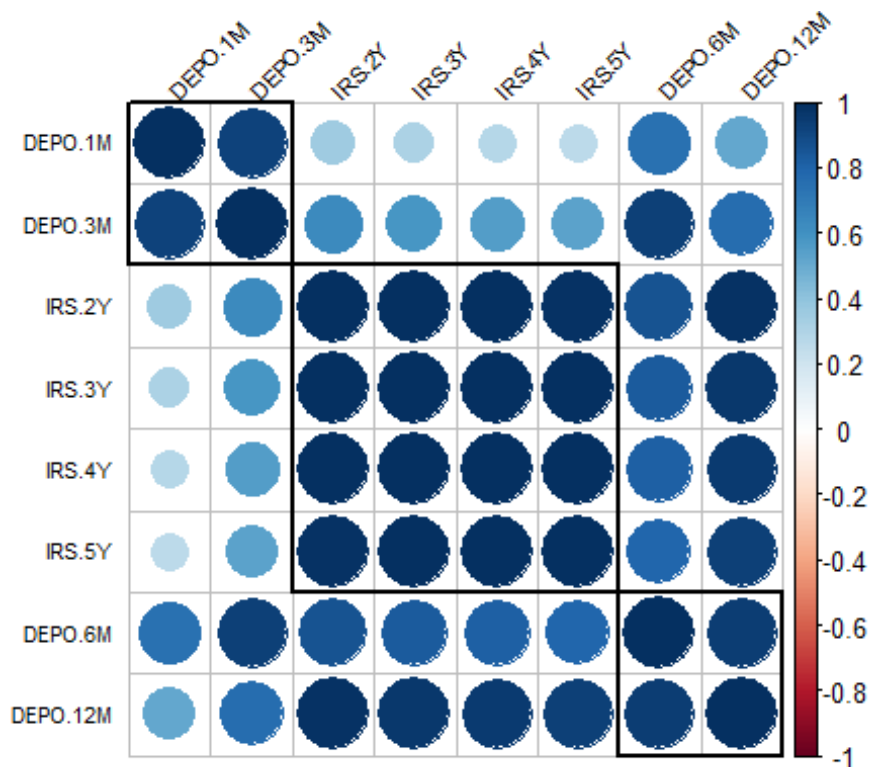
## ¿Tiene sentido llevar a cabo, en este caso, un análisis de componentes principales?

### Análisis de la matriz de correlación

Aquí el objetivo será comprobar si las características de la matriz son adecuadas o no para llevar a cabo el Análisis Factorial (ANFAC).

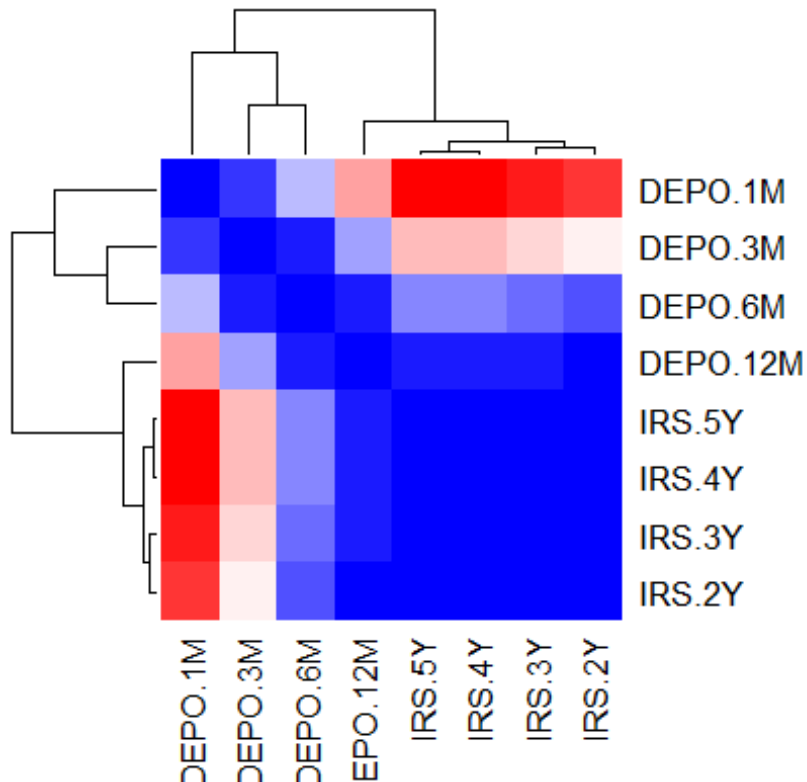


Ninguna correlación entre las variables es negativa, por pequeña que sea la asociación siempre es directa. En general existe una alta correlación entre las variables excepto en el caso de los bonos a 1 y 3 meses con correlaciones bajas con respecto a los bonos de 12 meses a 5 años.



Este segundo correlograma nos devuelve 3 grupos o clústeres dividiendo los bonos por plazos. El primer grupo, es el relativo a los bonos a corto plazo incluyendo los depósitos a 1 mes y 3 meses. El segundo grupo, relativo al medio plazo incluye los depósitos a 6 y 12 meses. El tercero, hace referencia al largo plazo e incluye los swaps a 2, 3, 4 y 5 años. En cada uno de los grupos se aprecia que las variables están muy asociadas, comportamiento común.

Estos resultados nos llevan a concluir la alta correlación entre las variables y, por tanto, el Análisis Factorial puede tener sentido.



En el mapa de calor pueden apreciarse tres separaciones: los bonos de 1, 3 y 6 meses; los bonos de 12 meses 5 y 4 años; los bonos de 4 a 2 años, pero claramente la división de grupos más clara es la que separa los depósitos a 6 meses con los de 12.

### Determinante de la matriz de correlaciones

El determinante de la matriz es muy pequeño ( $5.5021e-12$ ), lo cual indica alta asociación entre las variables, siendo por ello adecuado llevar a cabo el ANFAC.

### KMO

Es una medida de adecuación de la muestra; este índice permite comparar las magnitudes de los coeficientes de correlación observados con las magnitudes de los coeficientes de correlación parcial.

KMO de 0.83799 (siempre está entre 0 y 1). Se trata de un valor alto, mayor a 0.7, es conveniente el uso del ANFAC. En el caso de la matriz de adecuación de la muestra (MSA) el valor de los coeficientes es alto (cercano a 1), lo que indica que puede aplicarse ANFAC.

### Prueba de esfericidad de Bartlett

Se emplea para contratar la hipótesis de que la matriz de correlaciones es una matriz identidad, I. Siendo la  $H_0: R = I$

Un determinante próximo a cero implica que una o más variables podrían expresarse como combinación lineal de otras variables. En este caso se rechaza la hipótesis nula, p-value igual a 0, las variables si están relacionadas por tanto una o más variables pueden expresarse como combinación lineal de otras variables. Hay asociación entre las variables y, por tanto, es adecuado el empleo del ANFAC.

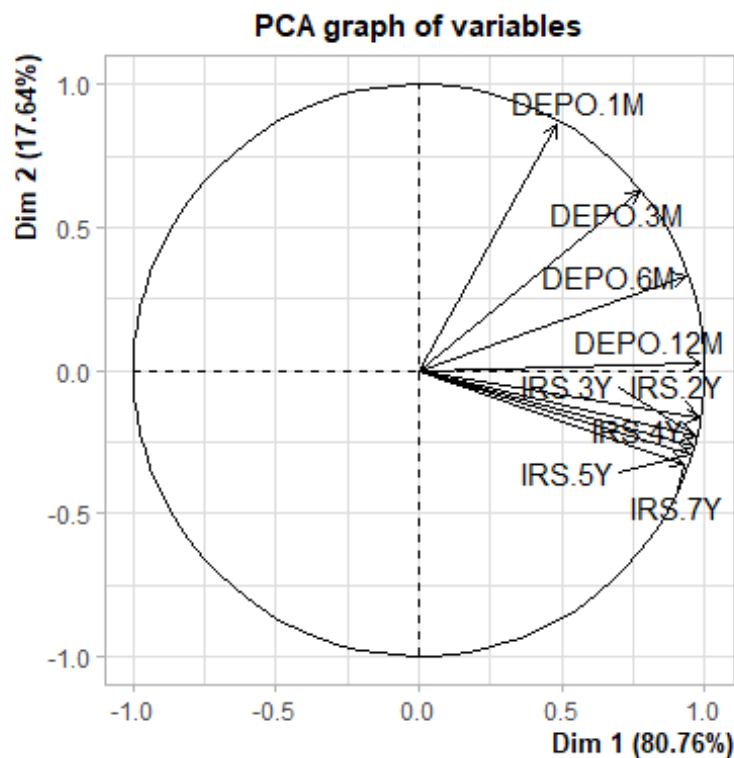
### Correlación anti-imagen

En esta prueba se emplea la matriz de correlaciones parciales, la cual permite determinar el grado de asociación exclusivo entre dos valores, eliminando la influencia que el resto de las variables pueda tener sobre cualquiera de las dos cogidas.

De nuestra matriz de coeficientes parciales concluimos que la mayoría de los coeficientes son bajos e incluso negativos (anti-imagen), siendo adecuado el uso del ANFAC.

**¿Cuántos componentes permitirían explicar, adecuadamente, la estructura subyacente de los tipos de interés aquí analizados?**  
**Análisis de componentes principales.**

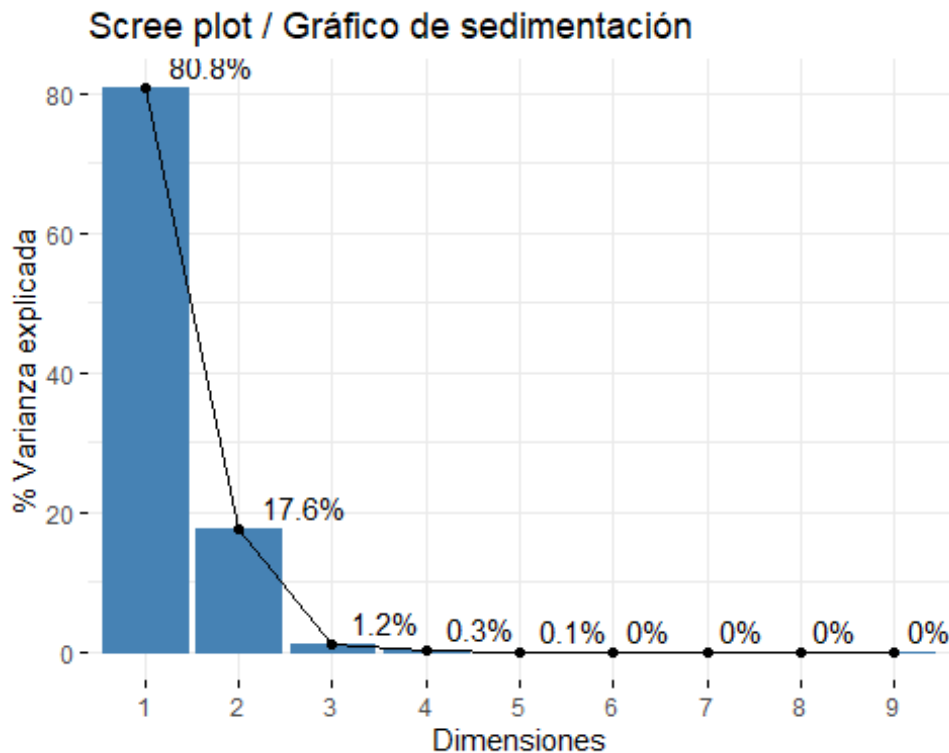
### Varianza explicada



Según se aprecia en el gráfico la dimensión 1 explica el 80.76% de la varianza mientras la segunda dimensión explica un 17.64 %, y ambas dimensiones explican un total del

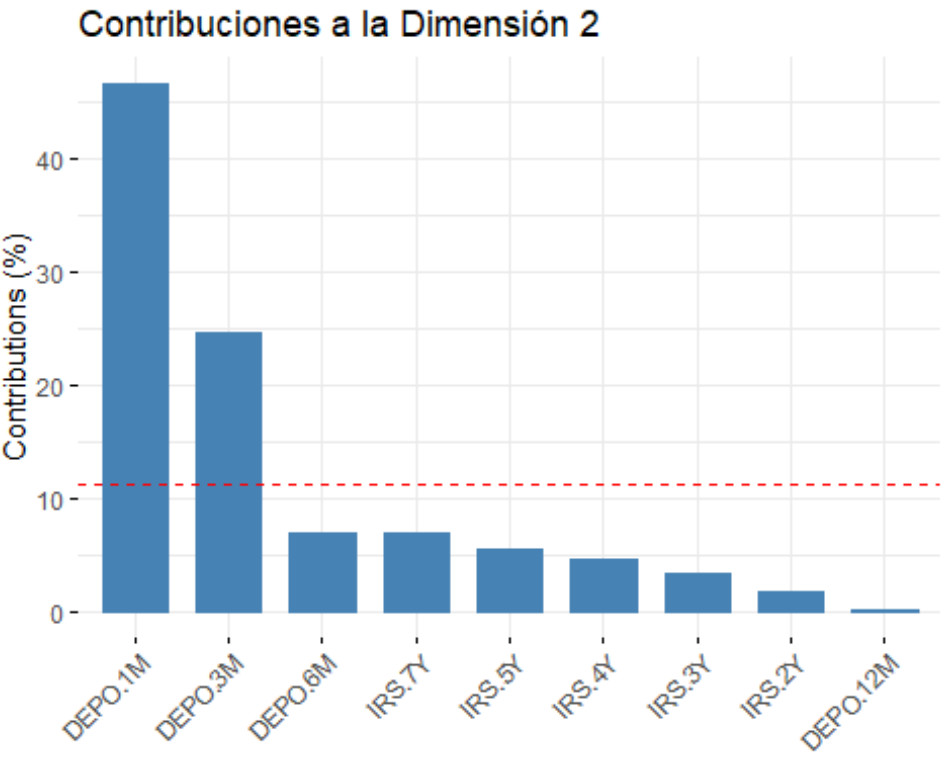
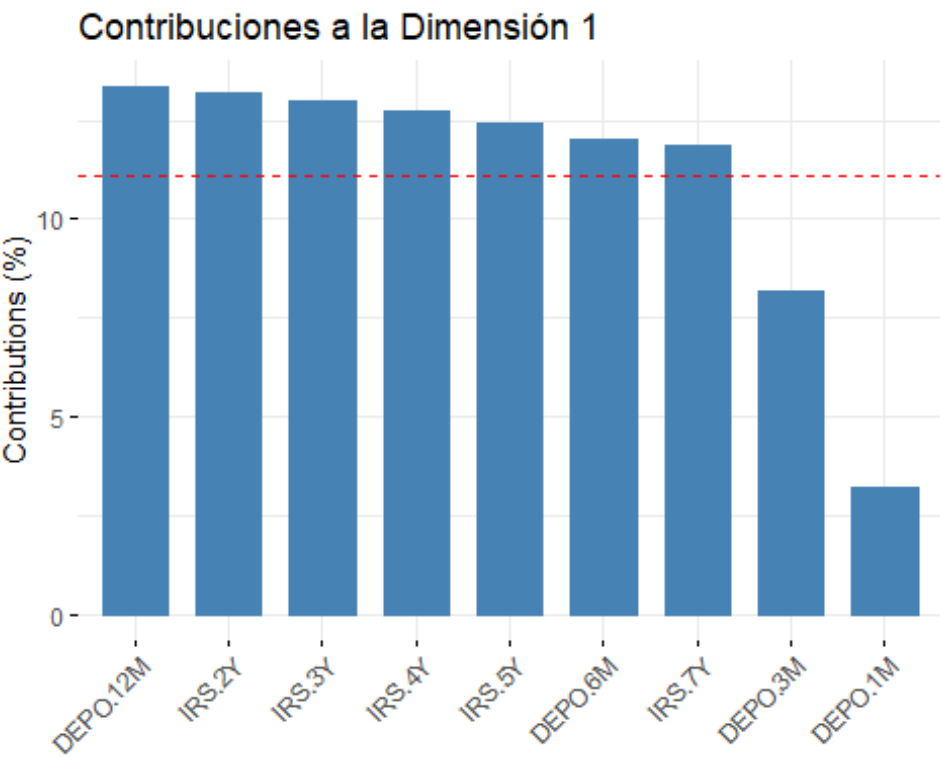
98.4% de la varianza. Además, si observamos los ejes veremos que hay una clara división en dos grupos de bonos, los del corto y los del largo plazo. En la parte derecha del eje vertical se encuentran los bonos de 1 mes a 12 meses siendo este último el divisor con los del largo plazo (bonos de 2 a 7 años) situados en la parte inferior derecha del eje horizontal.

### Gráfico de sedimentación



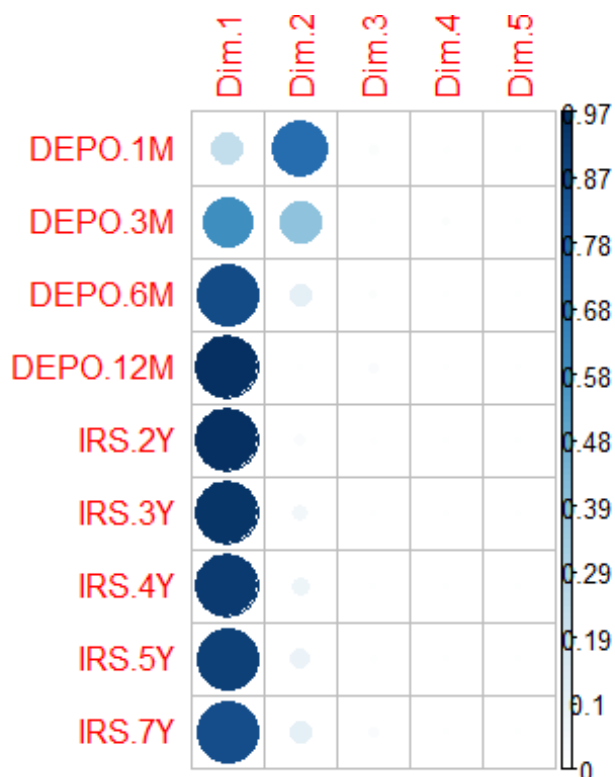
Esta gráfica de sedimentación muestra que las dos primeras dimensiones explican la mayor parte de la variabilidad total en los datos (dada por los valores propios). El resto de las dimensiones explican una proporción muy pequeña de la variabilidad y probablemente no son importantes.

Contribución de las variables a cada dimensión



En el caso de la dimensión 1 prácticamente todas las variables contribuyen al mismo nivel excepto la variable de bono a 1 y 3 meses que contribuyen en menor medida. En el caso de la dimensión 2 ocurre lo contrario, siendo la variable de bono a 1 y 3 meses las que más contribuyen en la dimensión con gran diferencia con respecto al resto de variables.

## Calidad de la representación



Como se observa de los resultados de la tabla cos2 (cosenos altos) y del gráfico de correlaciones (intensidad de color) existe una buena representación de las variables bonos a 6 meses hasta la variable bono a 7 años en el componente principal (Dim. 1). Mientras que en el componente principal (Dim. 2) la mejor representación viene dada por la variable bono a 1 mes con un coseno cuadrado de 0.738.

## ¿Tiene sentido llevar a cabo una rotación de las variables subyacentes?

La rotación factorial consiste en hacer girar los ejes de coordenadas que representan a los factores, hasta conseguir que se aproximen al máximo a las variables en que están saturados.



## Rotación varimax

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

	Dim.1	Dim.2
Proportion Var	0.67	0.314
Cumulative Var	0.67	0.984
	Dim.1	Dim.2
DEPO.1M		0.985
DEPO.3M	0.391	0.913
DEPO.6M	0.671	0.729
DEPO.12M	0.859	0.482
IRS.2Y	0.946	0.308
IRS.3Y	0.967	0.250
IRS.4Y	0.976	0.213
IRS.5Y	0.978	0.184
IRS.7Y	0.975	0.142

La varianza explicada por el primer factor (Dim.1), 80.8% ha disminuido al 67% mientras que el segundo factor (Dim.2) que explicaba un 17.6% ha pasado a explicar un 31.4% con esta rotación de la matriz. Sin embargo, la varianza acumulada sigue siendo la misma que en el ACP 98.4%.

Los factores principales ya no se orientan en las direcciones de máxima variación, sino que se reorientan en función de otros objetivos y esta se distribuye más homogéneamente entre todos los factores.

Por otra parte, cambia también la contribución de cada variable a las dimensiones. Por ejemplo, en la dimensión 2, donde antes la gran parte de la contribución provenía del

bono a 1 y 3 meses, ahora también contribuye en alta proporción con un 0.729. Mientras en la dimensión 2 las contribuciones se mantienen bastante similares, pero ya no es el bono a 12 meses el que más contribuye sino los bonos de entre 2 y 7 años. Se aprecia por tanto una división más clara en los factores entre los bonos a corto y más largo plazo.

## Predicción del bono a 10 años

Predecimos mediante regresión de componentes principales. Primero, para poder evaluar la capacidad predictiva del modelo, dividimos las observaciones en dos grupos: uno de training para ajustar el modelo y otro de test para predecir.

El método PCR es un ajuste lineal por mínimos cuadrados que emplea componentes principales como predictores. Utilizamos la función `pcr` -principal components regression- del paquete `pls` e incluimos cross validation para identificar el número óptimo de componentes con el que se minimiza el MSE.

El número óptimo de componentes principales identificado por cross validation para minimizar el MSE es de 8. La estimación del error de predicción obtenido mediante el Mean Square Error (MSE) da es de 0.000236.

## Conclusiones

Tras el análisis realizado puede concluirse. En primer lugar, que si tiene sentido llevar a cabo un análisis de componentes principales pues tras las diversas pruebas realizadas se ha podido corroborar la alta asociación entre las variables.

En segundo lugar, y en relación con los componentes que permiten explicar la estructura subyacente de los tipos de interés, se ha concluido que son suficientes dos dimensiones para explicar esta estructura pues representan el 98.4% de la varianza. Además, dentro de cada dimensión se ha comprobado que en el caso de la dimensión 1 prácticamente todas las variables contribuyen al mismo nivel excepto la variable de bono a 1 y 3 meses que contribuyen en mucha menor medida. Con respecto a la dimensión 2 ocurre lo contrario, son las variables bono a 1 y 3 meses las que más contribuyen.

En tercer lugar, con respecto a la rotación de las variables subyacentes puede decirse que no sería necesaria dicha rotación pues el porcentaje de varianza acumulada explicada por los factores se mantiene igual, 98.4%.

Por último, cabe concluir que a través de un modelo de regresión de componentes principales hemos podido realizar la predicción del bono a 10 años en el que hemos obtenido que el número óptimo de componentes principales para minimizar el error es de 8 con un error de 0.000236.

## Referencias

- Análisis de Componentes Principales (Principal Component Analysis, PCA) y t-SNE. Disponible en: [https://rpubs.com/Joaquin\\_AR/287787](https://rpubs.com/Joaquin_AR/287787)
- Estadística y Machine Learning con R. Disponible en: <https://rpubs.com/PacoParra/293407#:~:text=La%20prueba%20de%20esfericidad%20de,modelo%20factorial%20no%20ser%C3%ADa%20pertinente.>
- Interpretar todos los estadísticos y gráficas para Análisis factorial. Disponible en: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/factor-analysis/interpret-the-results/all-statistics-and-graphs/>
- Introducción a los Métodos Multivariantes. Disponible en: [https://rpubs.com/marcelo-chavez/multivariado\\_1](https://rpubs.com/marcelo-chavez/multivariado_1)
- Principal Component Methods in R: Practical Guide. Disponible en: <http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/>