



Técnicas de Agrupación y Reducción de la Dimensión

PRÁCTICA 2: Estructura temporal de los tipos de interés

V. Miguel Sempere Navarro

ÍNDICE

Introducción	3
Desarrollo del análisis.....	3
Conclusiones	10

Introducción

En el presente informe se va a llevar a cabo la explicación de un estudio a un conjunto de 978 observaciones de distintos productos financieros norteamericanos a distintos plazos entre el 2 de enero de 1995 y el 30 de septiembre de 1998 para responder a las exigencias de la práctica correspondiente.

En ésta, se plantean distintos objetivos o motivaciones, los cuales son:

1. **Averiguar si tiene sentido realizar un análisis de los componentes principales a los datos proporcionados.**
2. **Justificar con cuántos componentes, como mínimo, se podría explicar la estructura subyacente.**
3. **Explicar si tiene sentido llevar a cabo una rotación de las variables subyacentes.**
4. **Predecir el valor del bono a 10 años.**

Para poder contestar a los objetivos planteados, han sido utilizadas diferentes técnicas que justifican las respuestas. Se explicarán a continuación y se extraerán unas conclusiones finales.

Desarrollo del análisis

En este apartado se explicará cómo se ha desarrollado el análisis y se dará respuesta a los diferentes objetivos planteados.

El principio del análisis se sitúa en la importación de los datos, donde se averigua que varios datos carecen de observación. Para solucionarlo, se ha entrado en la página web de la Reserva Federal y se han importado los datos oficiales, los cuales han sido importados posteriormente por R: <https://fred.stlouisfed.org/series/USD1MTD156N>

Tras ser limpiados de datos que no aportaban valor al análisis, como las fechas, se ha tratado de contestar al **primer objetivo**:

En primer lugar, se realiza la Matriz de Correlaciones de las primeras 949 observaciones, eliminando los valores del bono a 10 años y el resultado es el siguiente:

	DEPO.1M	DEPO.3M	DEPO.6M	DEPO.12M	IRS.2Y	IRS.3Y	IRS.4Y	IRS.5Y	IRS.7Y
DEPO.1M	1.0000000	0.9220899	0.7423781	0.5071214	0.3554720	0.3101345	0.2816444	0.2601207	0.2268505
DEPO.3M	0.9220899	1.0000000	0.9279445	0.7643855	0.6333915	0.5895566	0.5591534	0.5341612	0.4954058
DEPO.6M	0.7423781	0.9279445	1.0000000	0.9440732	0.8607230	0.8245985	0.7953775	0.7693432	0.7281305
DEPO.12M	0.5071214	0.7643855	0.9440732	1.0000000	0.9731443	0.9515135	0.9292639	0.9074514	0.8714994
IRS.2Y	0.3554720	0.6333915	0.8607230	0.9731443	1.0000000	0.9948671	0.9836059	0.9699235	0.9443331
IRS.3Y	0.3101345	0.5895566	0.8245985	0.9515135	0.9948671	1.0000000	0.9966304	0.9891730	0.9719242
IRS.4Y	0.2816444	0.5591534	0.7953775	0.9292639	0.9836059	0.9966304	1.0000000	0.9977443	0.9875916
IRS.5Y	0.2601207	0.5341612	0.7693432	0.9074514	0.9699235	0.9891730	0.9977443	1.0000000	0.9956517
IRS.7Y	0.2268505	0.4954058	0.7281305	0.8714994	0.9443331	0.9719242	0.9875916	0.9956517	1.0000000

Ilustración 1. Matriz de Correlaciones. Fuente: elaboración propia.

Se observa que, a menor diferencia temporal, mayor es la correlación. Esto significa que cuando un depósito o un bono tienen un espacio temporal corto entre ellos, el comportamiento es similar. Esto ya nos indica, a priori, que el interés del bono a 10 años tendrá un comportamiento más parecido al bono de 7 años que a un depósito a 1 mes.

El análisis continúa realizando el determinante de dicha matriz:

```
> det(cor_datosbonosOK)
[1] 2.376064e-18
```

Ilustración 2. Determinante de la Matriz de Correlaciones. Fuente: elaboración propia.

El determinante es muy reducido. Esto significa que las diferencias entre las correlaciones son muy pequeñas entre sí: comportamiento similar.

El siguiente paso realizado ha sido el Test de Cortest Barlett:

```
> cortest.bartlett(cor_datosbonosOK, n = length(datos_bonosOK))
$chisq
[1] 169.0879

$p.value
[1] 3.875322e-19

$df
[1] 36
```

Ilustración 3. Test de Cortest Barlett. Fuente: elaboración propia.

El valor que más nos interesa de dicho test es el “P Value” o coeficiente de significación. Este tiene un valor muy bajo, según se esperaba, y se complementa estudiando el Test KMO:

```
> KMO(cor_datosbonosOK)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = cor_datosbonosOK)
Overall MSA = 0.87
MSA for each item =
  DEPO.1M  DEPO.3M  DEPO.6M  DEPO.12M  IRS.2Y  IRS.3Y  IRS.4Y  IRS.5Y  IRS.7Y
    0.80    0.79    0.85    0.92    0.88    0.85    0.89    0.87    0.90
```

Ilustración 4. Test KMO. Fuente: elaboración propia.

Todos los valores a excepción del depósito a 3 meses son mayores a 0.8, pero se considera que todos lo son puesto que 0.79 es un valor muy próximo al mínimo aceptable para considerar oportuno realizar un análisis de los componentes principales; pues todos los indicadores necesarios para comprobar su necesidad de ser utilizado reclaman su uso: **no se perderá apenas información realizando un ACP, por tanto, conviene hacerlo.**

Una vez contestado el primer objetivo, se tratará de contestar al **segundo objetivo**, el cual está ligado con el primero pues es necesaria la realización del ACP para poder contestar:

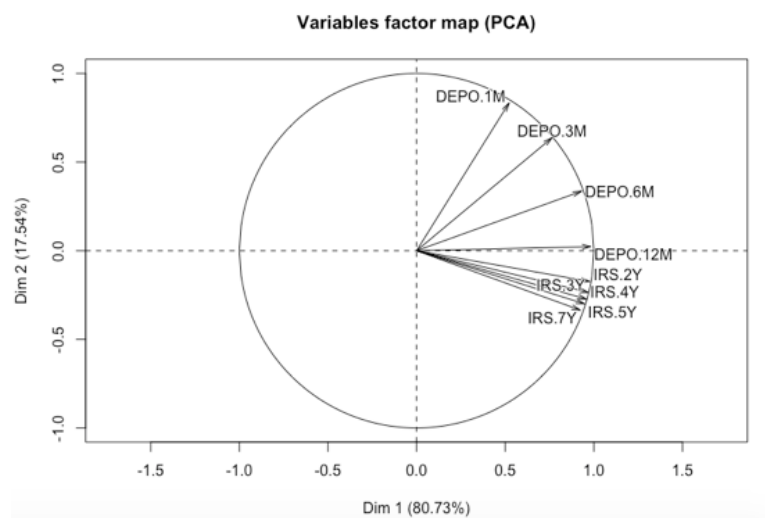


Ilustración 5. Análisis de los Componentes Principales. Fuente: elaboración propia.

Tras realizarse dicho análisis, el gráfico aporta información muy valiosa: el 80'73% de la varianza acumulada es representada en la Dimensión 1 y el 17,54% en la Dimensión 2. Representándose así un 98'27% de la varianza total. También se demuestra analíticamente con los autovalores y el \cos^2 :

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	7.2660647033	8.073405e+01	80.73405
comp 2	1.5781896616	1.753544e+01	98.26949
comp 3	0.1228101098	1.364557e+00	99.63405
comp 4	0.0232990822	2.588787e-01	99.89293
comp 5	0.0059055614	6.561735e-02	99.95855
comp 6	0.0031328553	3.480950e-02	99.99336
comp 7	0.0003920010	4.355567e-03	99.99771
comp 8	0.0001280494	1.422772e-03	99.99913
comp 9	0.0000779759	8.663989e-04	100.00000

Ilustración 6. Autovalores. Fuente: elaboración propia.

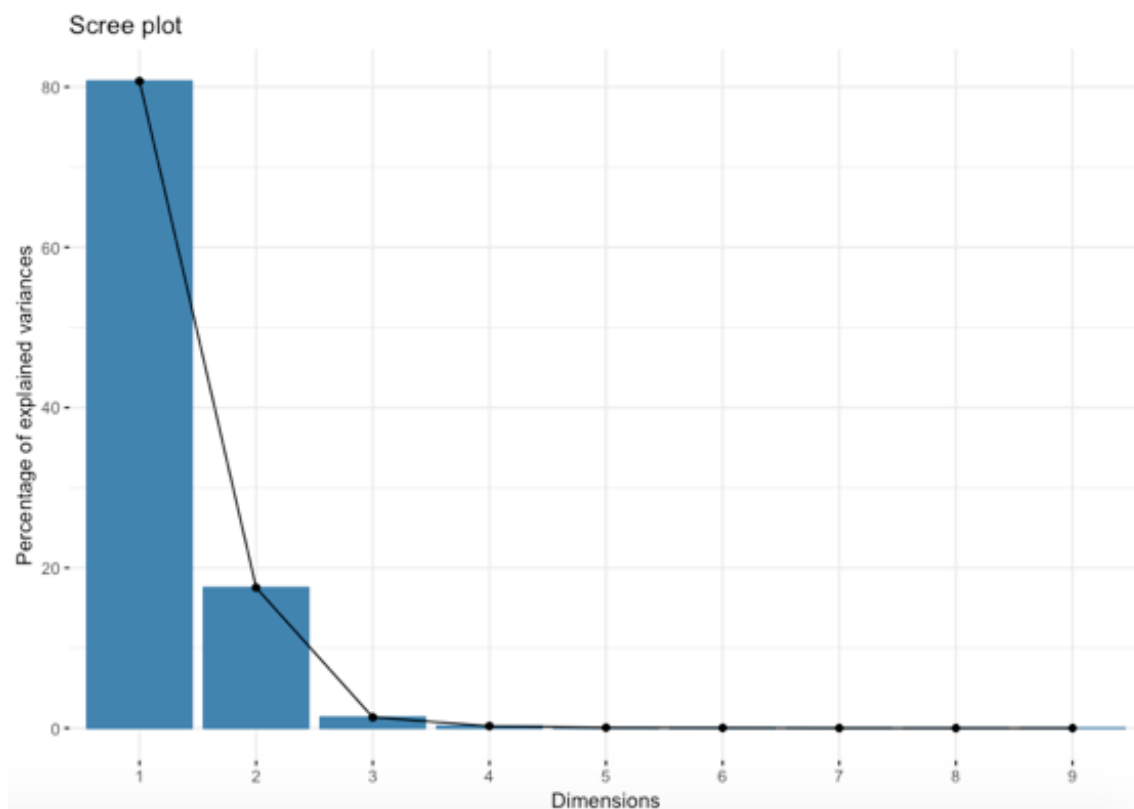


Ilustración 7. Porcentaje de varianza explicada. Fuente: elaboración propia.

Como se observa, “comp 1” y “comp 2” representan 80,73% de la varianza y 17,74%, respectivamente. Es decir, un 98,27% de la varianza total. Además, según la regla de

Kaiser, solo se aceptarán los componentes con autovalor ≥ 1 . (comp1 y comp2 en este caso).

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
DEPO.1M	0.2724194	0.694975895	2.573255e-02	6.732903e-03	1.392226e-04
DEPO.3M	0.5844445	0.406151747	1.317317e-05	7.629129e-03	1.518785e-03
DEPO.6M	0.8669085	0.112720847	1.645706e-02	1.583391e-03	9.129592e-04
DEPO.12M	0.9671203	0.000526677	2.945751e-02	5.153567e-04	1.052497e-03
IRS.2Y	0.9573429	0.030424248	7.685698e-03	3.218843e-03	1.136872e-03
IRS.3Y	0.9423580	0.055725878	1.976393e-04	1.138986e-03	4.782109e-04
IRS.4Y	0.9222535	0.074926913	2.537836e-03	5.671283e-05	7.079469e-05
IRS.5Y	0.8988163	0.090101671	1.074219e-02	1.932470e-04	1.443243e-05
IRS.7Y	0.8544014	0.112635786	2.998645e-02	2.230513e-03	5.817874e-04

Ilustración 8. Cos^2 . Fuente: elaboración propia.

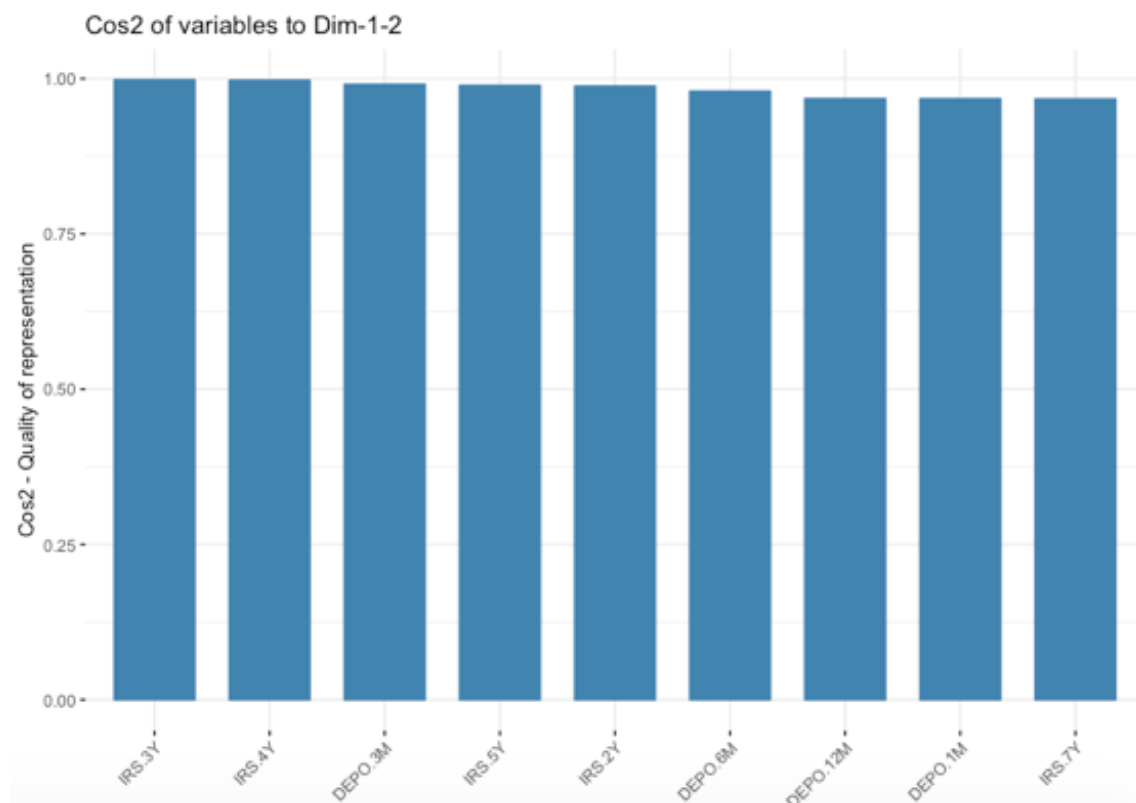


Ilustración 9. Calidad de la representación. Fuente: elaboración propia.

Los datos de la anterior ilustración muestran como cada variable es representada por cada dimensión.

Se observa con los anteriores datos y gráficos lo mismo que en el gráfico del ACP, lo cual contesta al segundo objetivo: **se justifica que con dos componentes se puede explicar**

la estructura subyacente, pues con los dos primeros se explica un 98'27% de la varianza acumulada.

Para contestar al **tercer objetivo** se ha realizado el siguiente análisis factorial:

```
Call:
factanal(x = datos_bonosOK, factors = 2, rotation = "varimax")

Uniquenesses:
DEPO.1M DEPO.3M DEPO.6M DEPO.12M IRS.2Y IRS.3Y IRS.4Y IRS.5Y IRS.7Y
0.174 0.050 0.005 0.021 0.014 0.005 0.005 0.006 0.021

Loadings:
          Factor1 Factor2
DEPO.1M          0.907
DEPO.3M 0.344 0.912
DEPO.6M 0.624 0.779
DEPO.12M 0.821 0.552
IRS.2Y 0.923 0.365
IRS.3Y 0.953 0.295
IRS.4Y 0.968 0.246
IRS.5Y 0.975 0.207
IRS.7Y 0.978 0.152

          Factor1 Factor2
SS loadings 5.791 2.913
Proportion Var 0.643 0.324
Cumulative Var 0.643 0.967

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 12186.7 on 19 degrees of freedom.
The p-value is 0
```

Ilustración 10. Análisis factorial. Fuente: elaboración propia.

Este análisis tiene como objetivo encontrar una dimensión en la cual la gran mayoría de las variables puedan ser lo máximo posible representadas en una dimensión. **Por los datos obtenidos, con la segunda dimensión se podría representar hasta un máximo de un 96'7% de la varianza, lo cual es inferior al 98'27% del caso anterior.** Sin embargo, aunque no es tan alto el porcentaje de varianza explicada, es un número a considerar. Es necesaria más información externa para poder tomar la decisión.

El cuarto y **último objetivo** de la práctica es realizar una predicción del bono a 10 años con los valores suplementarios que se otorgan. Gracias al análisis anterior y al uso de funciones de R, se ha obtenido el siguiente resultado:

	IRS.10Y	Prediccion
950	6.065	6.071920
951	6.085	6.070502
952	6.105	6.124059
953	6.055	6.069333
954	5.935	5.940717
955	5.915	5.932817
956	5.905	5.947481
957	5.875	5.880950
958	5.885	5.910524
959	5.825	5.823147
960	5.755	5.763296
961	5.745	5.767543
962	5.745	5.757979
963	5.705	5.709204
964	5.575	5.581874
965	5.565	5.565412
966	5.665	5.662167
967	5.635	5.603030
968	5.675	5.650909
969	5.605	5.591048
970	5.565	5.550395
971	5.485	5.485484
972	5.525	5.524465
973	5.555	5.548443
974	5.495	5.507316
975	5.385	5.388330
976	5.375	5.379340
977	5.395	5.398776
978	5.245	5.257924

Ilustración 11. Predicción bono 10 años. Fuente: elaboración propia.

Los resultados obtenidos de la predicción realizada, gracias al estudio previo a dicha predicción, son muy parecidos a los otorgados por el archivo original con un error muy reducido. **La predicción es buena.**

Conclusiones

La práctica estaba compuesta por cuatro objetivos, pero, aunque algunos de ellos obligaban a realizar algunas técnicas que se debían afianzar, estos no eran los objetivos principales, a mi juicio.

Una de las principales aplicaciones reales de este trabajo es la predicción del valor de un bono a 10 años y, probablemente, hubiésemos acertado; tal y como se ha demostrado.

Esto nos quiere decir que un ACP puede ser de gran utilidad, pues si realizando una predicción con los datos obtenidos del ACP nos da tal nivel de precisión –y no muy distinta, probablemente, de la posible predicción obtenida con los datos originales sin haber realizado un ACP- compensa realizar un ajuste de dimensiones para eliminar información que no nos aporta valor, ocupa espacio y ralentiza los procesos.