

# Estructura temporal

*Mayra Goicochea Neyra*

*28/10/2019*

## Caso ACP

### Introducción

Para analizar el rendimiento de bonos norteamericanos, se recoge la información de 978 observaciones de los rendimientos de 10 bonos a distintos plazos entre el 2 de enero de 1995 y el 30 de septiembre de 1998. Se tiene la información en 10 variables.

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(ggplot2)
library(readr)
library(kknn)
library(factoextra)

## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ

library(FactoMineR)

raw_data <- read.csv("../ACPTIUSD.csv", sep = ";")
colnames(raw_data[, -1])

## [1] "DEPO.1M" "DEPO.3M" "DEPO.6M" "DEPO.12M" "IRS.2Y" "IRS.3Y"
## [7] "IRS.4Y" "IRS.5Y" "IRS.7Y" "IRS.10Y"
```

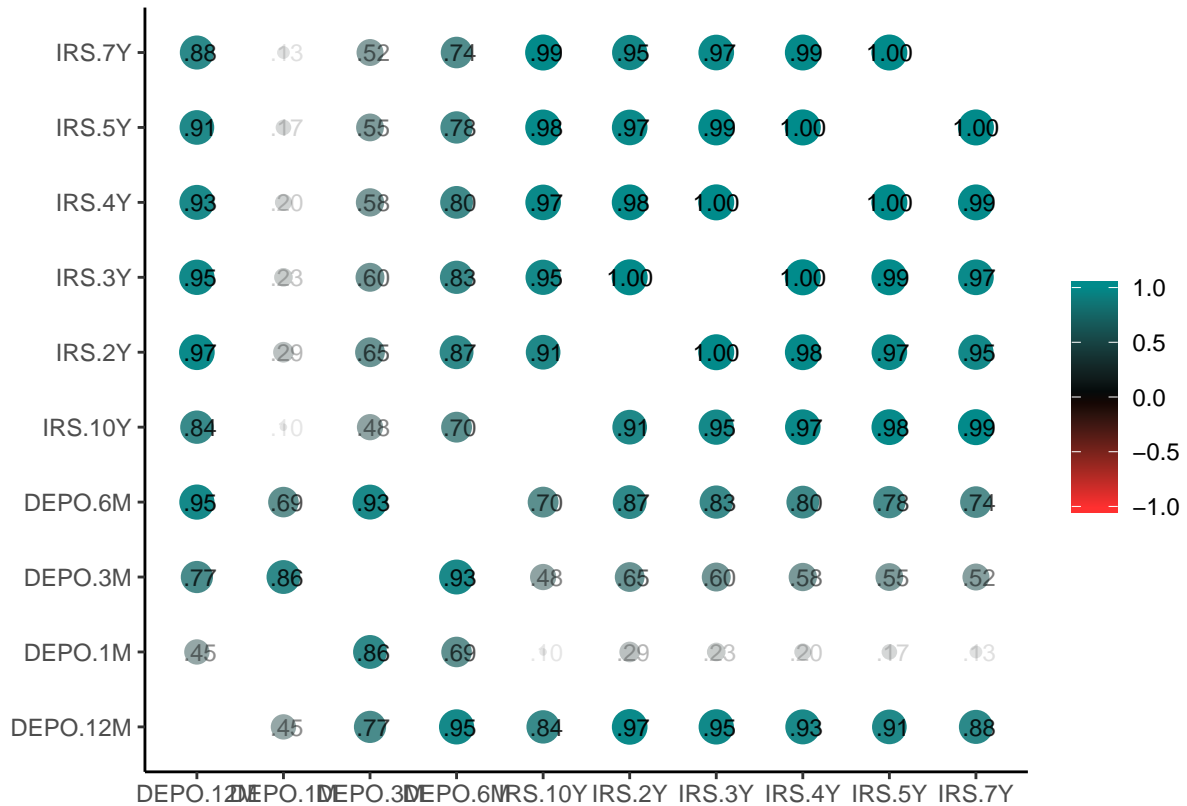
El objetivo del siguiente análisis es predecir el valor de un bono a 10 años, para cumplir con esa finalidad, se consideró la muestra de las primeras 949 observaciones como activas, y las siguientes 29 observaciones como suplementarias para comprobar el modelo resultante.

### Análisis Exploratorio de Datos (EDA)

```
dfData <- rminer::imputation("hotdeck", raw_data, "DEPO.1M")
dfData <- dfData[, -1]
dfData.act <- dfData[1:949,]
dfData.sup <- dfData[950:978,]

m.corr <- cor(dfData)
corrr::rplot(m.corr, legend = TRUE, colours = c("firebrick1", "black", "darkcyan"), print_cor = TRUE)
```

```
## Don't know how to automatically pick scale for object of type noquote. Defaulting to continuous.
## Don't know how to automatically pick scale for object of type noquote. Defaulting to continuous.
```



```
det(m.corr)
```

```
## [1] 3.244602e-21
```

La matriz de correlaciones muestra alta relacion entre las variables. La determinante es muy cercana a cero, no se puede concluir que sean linealmente dependientes, pero si que hay cierta multicolinealidad.

###Prueba de Bartlett,

```
psych::cortest.bartlett(m.corr)
```

```
## Warning in psych::cortest.bartlett(m.corr): n not specified, 100 used
```

```
## $chisq
## [1] 4473.98
##
## $p.value
## [1] 0
##
## $df
## [1] 45
```

Según la prueba de bartlett, aparentemente hay multicolinealidad entre las variables.

## Prueba KMO (Kaiser-Meyer-Olkin)

```
psych::KMO(dfData[, -10])
```

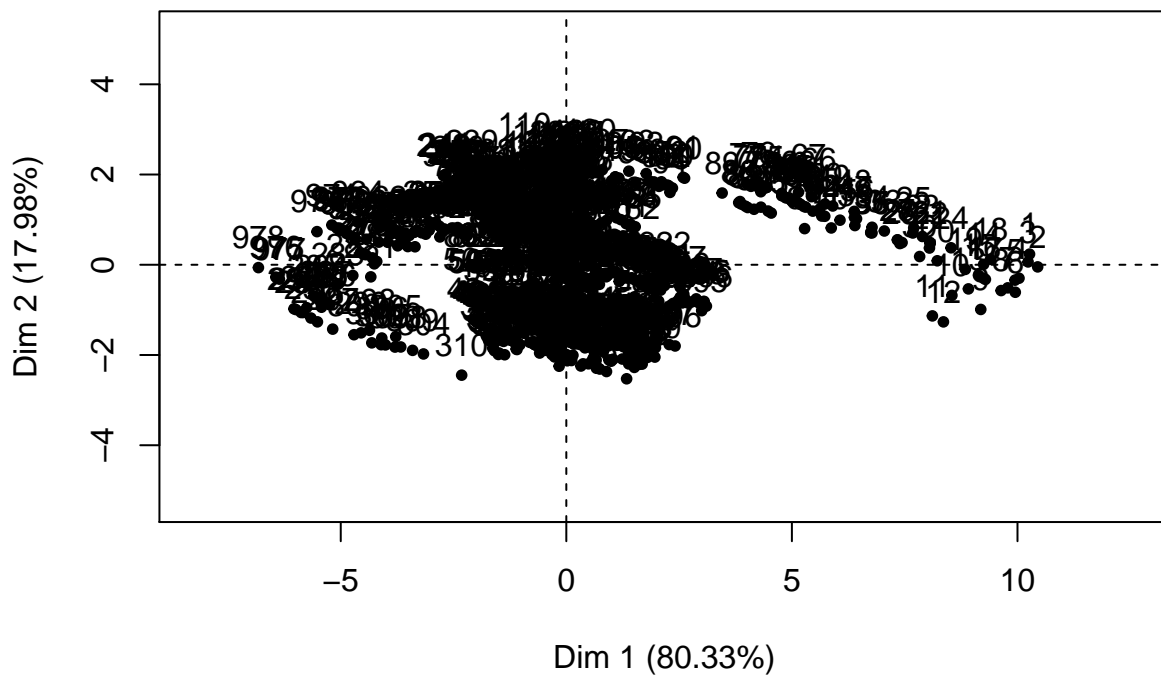
```
## Kaiser-Meyer-Olkin factor adequacy
## Call: psych::KMO(r = dfData[, -10])
## Overall MSA = 0.87
## MSA for each item =
##  DEPO.1M  DEPO.3M  DEPO.6M  DEPO.12M  IRS.2Y  IRS.3Y  IRS.4Y  IRS.5Y
##    0.87    0.81    0.83    0.91    0.88    0.85    0.90    0.87
##  IRS.7Y
##    0.91
```

Mediante la prueba de Kaiser-Meyer-Olkin, se verifica que la mayoría de variables pueden ser explicadas por otras, es así que un análisis de componentes principales puede ser adecuado para reducir la cantidad de variables.

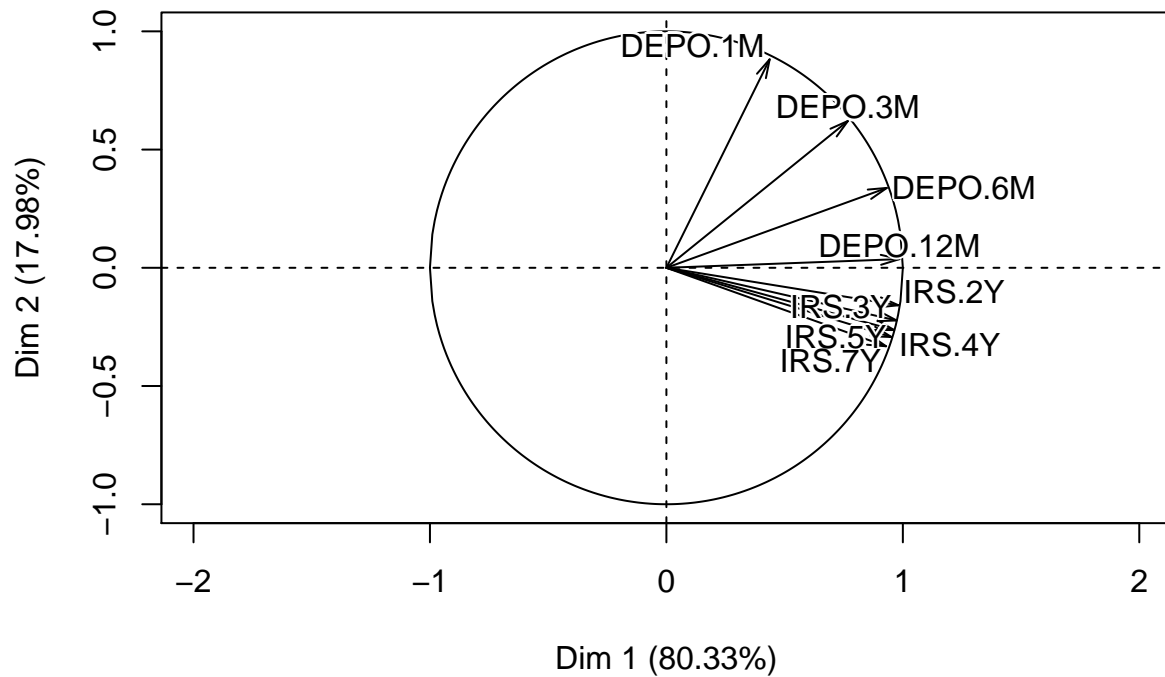
## Análisis de Componentes Principales

```
acp <- PCA(dfData[, -10], scale.unit = TRUE, graph = T)
```

### Individuals factor map (PCA)



## Variables factor map (PCA)

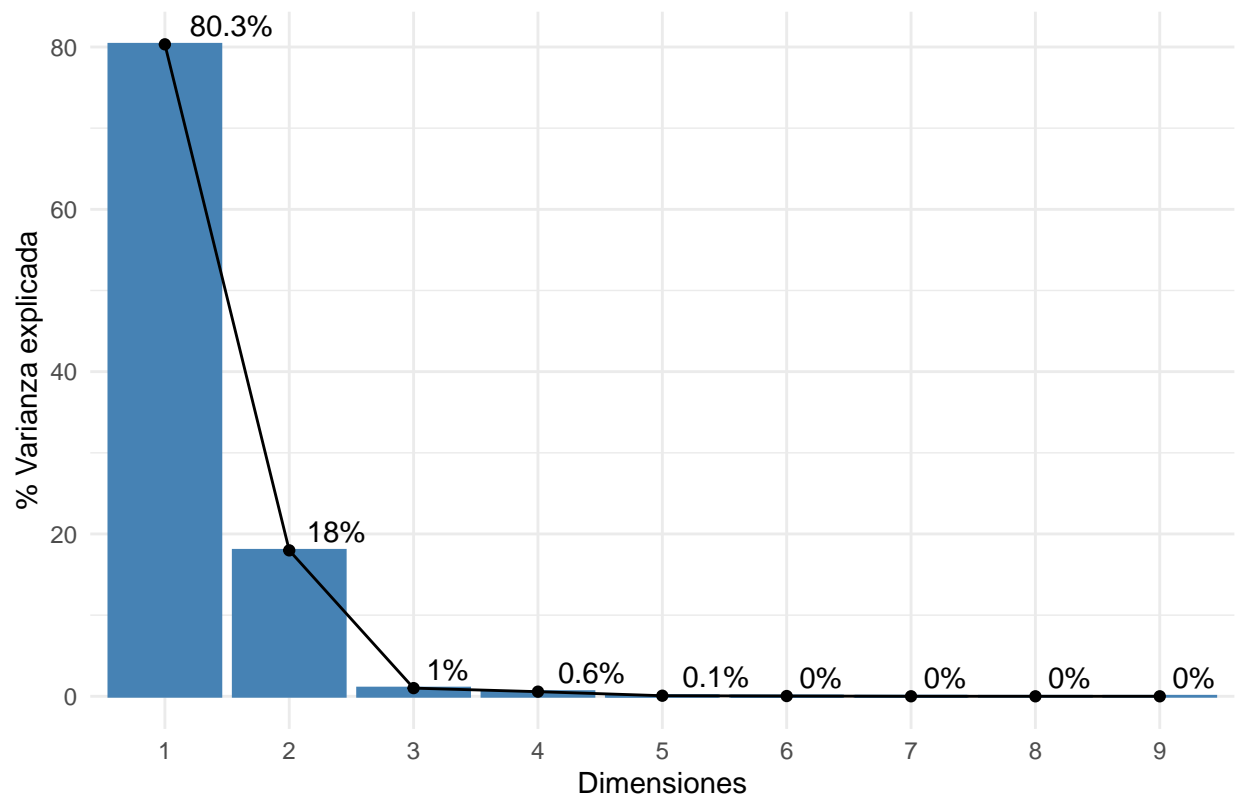


Según el PCA, se puede representar la información al 98.31% con dos componentes. Dejando el 77.78% de las variables solo se perdería el 1.69% de la información. También se pueden diferenciar dos grupos de variables. En cuanto a las variables, DEPO.1M estaría mejor explicada por la dimensión 2, a diferencia de las otras, que están mas asociadas a la dimensión 1.

### Varianza Explicada

```
fviz_eig(acp, addlabels = TRUE, hjust = -0.3) +  
  labs(title = "Scree plot / Gráfico de sedimentación", x = "Dimensiones", y = "% Varianza explicada") +  
  theme_minimal()
```

### Scree plot / Gráfico de sedimentación

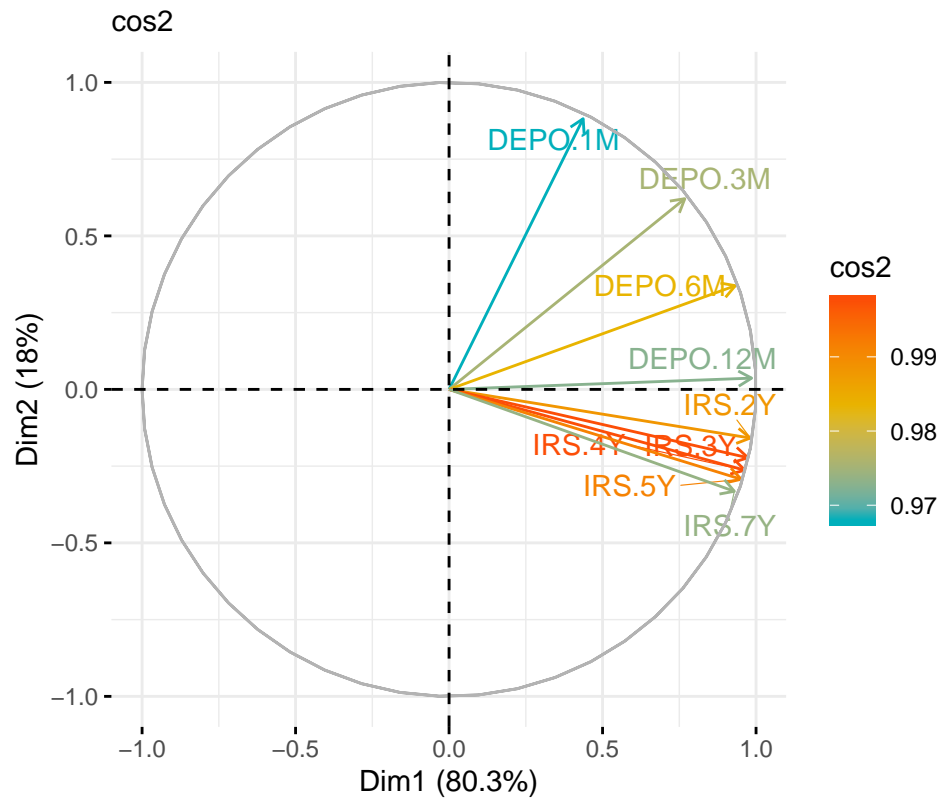


Los dos primeros componentes explican el 98.3% de la varianza total. Se puede concluir que los otros 7 componentes no son significativos para la varianza.

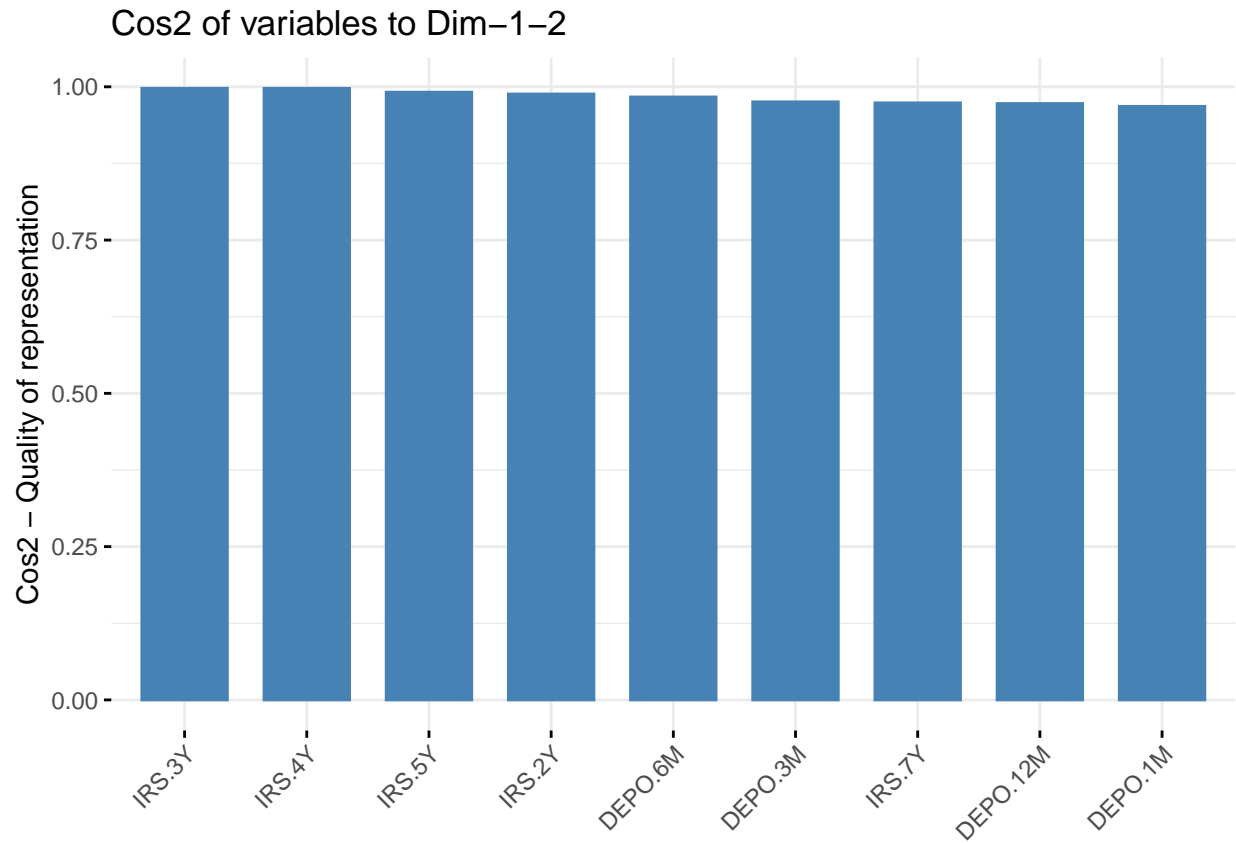
### Calidad de Representación

```
fviz_pca_var(acp, col.var = "cos2",  
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),  
             repel = TRUE ) +  
  labs(title = "Mapa de ejes principales", subtitle = "cos2")
```

## Mapa de ejes principales



```
fviz_cos2(acp, choice = "var", axes = 1:2)
```

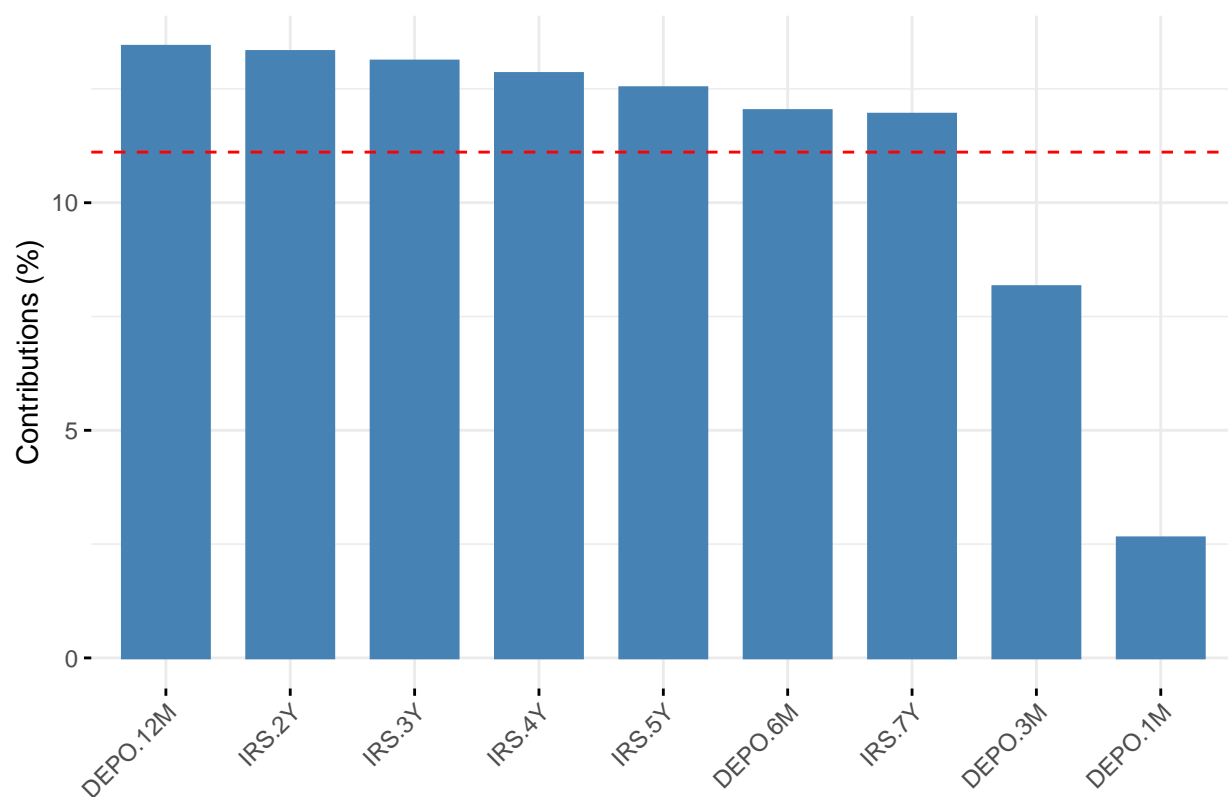


Las variables IRS.3Y, IRS.4Y e IRS.5Y son las mejores representadas por el componente principal 1. Las otras variables también son adecuadamente representadas dado su coeficiente de mayor e igual a 0.98.

#### Contribución a los CP

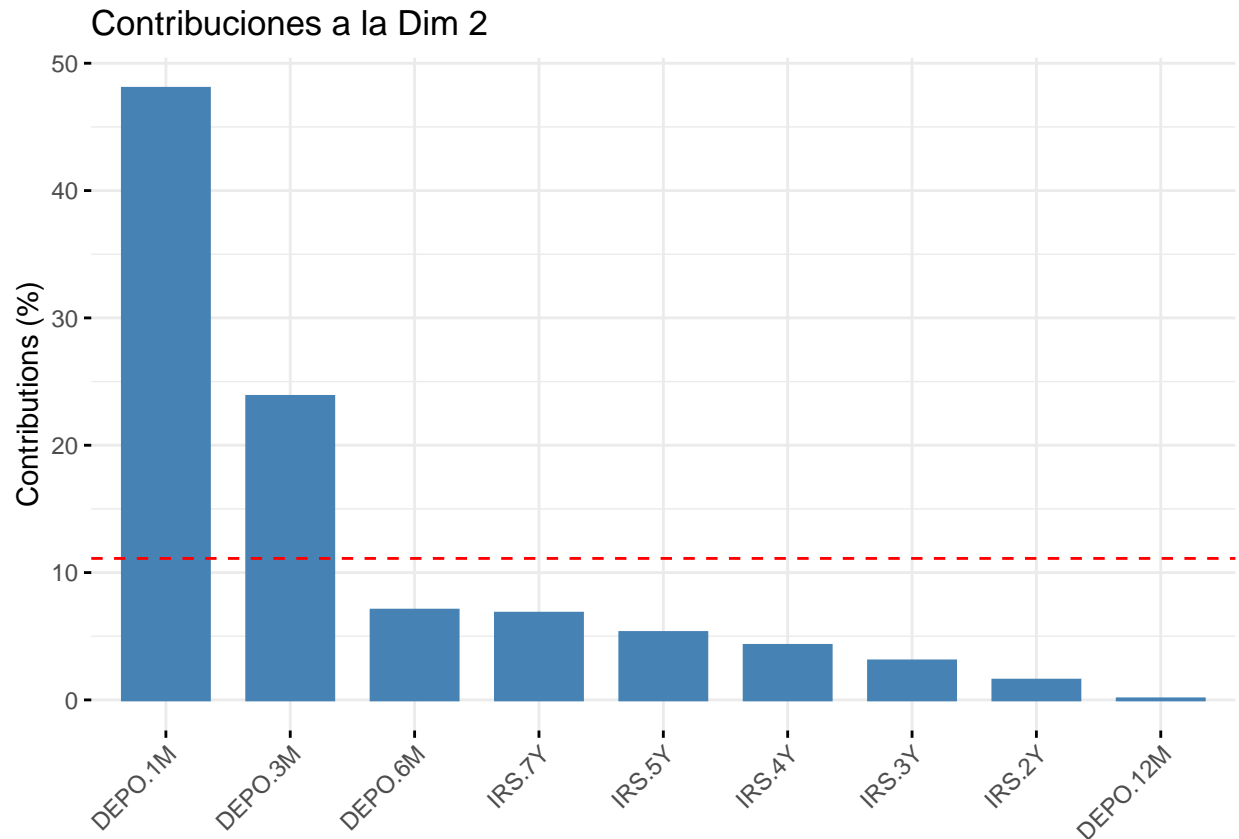
```
fviz_contrib(acp, choice = "var", axes = 1 ) +  
  labs(title = "Contribuciones a la Dim 1")
```

Contribuciones a la Dim 1



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





Las variables DEPO.12M, IRS.2Y, IRS.3Y, IRS.5Y, DEPO.6M, e IRS.7Y son muy bien explicadas por la dimensión 1. En cambio, las variables DEPO.1M y DEPO.3M son mejor representadas por la dimensión 2. Todas las variables contribuyen a demostrar la variabilidad del dataset.

## No Rotación o Ajuste de Rotación

```
psych::principal(dfData[, -10], nfactors = 2, rotate = "none")
```

```
## Principal Components Analysis
## Call: psych::principal(r = dfData[, -10], nfactors = 2, rotate = "none")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          PC1   PC2   h2   u2 com
## DEPO.1M  0.44  0.88  0.97  0.0320 1.5
## DEPO.3M  0.77  0.62  0.98  0.0246 1.9
## DEPO.6M  0.93  0.34  0.98  0.0166 1.3
## DEPO.12M 0.99  0.04  0.97  0.0273 1.0
## IRS.2Y   0.98 -0.16  0.99  0.0118 1.1
## IRS.3Y   0.97 -0.22  1.00  0.0025 1.1
## IRS.4Y   0.96 -0.26  1.00  0.0025 1.1
## IRS.5Y   0.95 -0.29  0.99  0.0087 1.2
## IRS.7Y   0.93 -0.33  0.97  0.0262 1.3
##
##          PC1   PC2
## SS loadings      7.23 1.62
## Proportion Var    0.80 0.18
## Cumulative Var    0.80 0.98
## Proportion Explained 0.82 0.18
```

```
## Cumulative Proportion 0.82 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.01
## with the empirical chi square 7.2 with prob < 0.99
##
## Fit based upon off diagonal values = 1
```

Con dos componentes se puede explicar el 98% de la varianza. El primer componente es el que tiene más porcentaje de interpretación de la varianza.

```
#Varimax
psych::principal(dfData[, -10], nfactors = 2, rotate = "varimax")
```

```
## Principal Components Analysis
## Call: psych::principal(r = dfData[, -10], nfactors = 2, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##          RC1  RC2  h2    u2 com
## DEPO.1M -0.01 0.98 0.97 0.0320 1.0
## DEPO.3M 0.40 0.90 0.98 0.0246 1.4
## DEPO.6M 0.68 0.72 0.98 0.0166 2.0
## DEPO.12M 0.86 0.48 0.97 0.0273 1.6
## IRS.2Y 0.95 0.30 0.99 0.0118 1.2
## IRS.3Y 0.97 0.24 1.00 0.0025 1.1
## IRS.4Y 0.98 0.20 1.00 0.0025 1.1
## IRS.5Y 0.98 0.17 0.99 0.0087 1.1
## IRS.7Y 0.98 0.12 0.97 0.0262 1.0
##
##          RC1  RC2
## SS loadings      6.08 2.76
## Proportion Var    0.68 0.31
## Cumulative Var    0.68 0.98
## Proportion Explained 0.69 0.31
## Cumulative Proportion 0.69 1.00
##
## Mean item complexity = 1.3
## Test of the hypothesis that 2 components are sufficient.
##
## The root mean square of the residuals (RMSR) is 0.01
## with the empirical chi square 7.2 with prob < 0.99
##
## Fit based upon off diagonal values = 1
```

Con una rotación Varimax, se tiene un mejor modelo, donde los pesos de las variables se diferencian mejor, y los componentes principales explican la varianza de forma más equilibrada.

## Conclusiones

- Dado que la muestra tiene multicolinealidad, nos permite reducir las dimensiones mediante el Análisis de Componentes Principales.
- Se observó mediante los graficos y calculos generados que se puede explicar el modelo mediante 2 componentes.
- El ajuste Varimex, permite identificar más fácilmente que componente tiende a asociarse con cada variable. Además, permite que se equilibren los autovalores.

## Bibliografía

- STHDA, en sitio web: <http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/#data-standardization>
- Análisis de Componentes Principales, rpub por Joaquín Amat en sitio web: [https://rpubs.com/Joaquin\\_\\_AR/287787](https://rpubs.com/Joaquin__AR/287787)
- Análisis de Componentes Principales, rpub por Cristina Gil en sitio web: [https://rpubs.com/Cristina\\_\\_Gil/PCA](https://rpubs.com/Cristina__Gil/PCA)