

Análisis Estadístico Multivariado | UTDT

Problem Set 3

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

Utilizando la base de datos *firmas.dta*, la cual contiene información financiera referida a dos tipos de firmas: 12 firmas consideradas de buena performance, y 12 firmas no tan buenas en ese sentido; y dispone de dos características relevada: EBITASS (gcias. despues de impuestos e intereses sobre activos; y ROTC (retorno sobre el capital), se pide:

- (a) Realizar un analisis de clusters empleando el algoritmo de las k -medias. Determinar el número de grupos a partir de un test F de reducción de variabilidad.
- (b) Realizar un análisis de conglomerados empleando métodos jerárquicos. Trabajar con distintas medidas de distancias entre clusters, comentando los pros y las contras de cada una de ellas.

Ejercicio 2

Utilice la base de datos *iris* para realizar un análisis de conglomerados. La base contiene datos de especies de flores de iris con registros de las medidas del sépalo y pétalo de cada una. Originalmente los datos ya están agrupados por especie, pero en este ejercicio se propone utilizar *k-means clustering* para aplicar el método y ver cómo se compara con las categorías verdaderas.

Nota: la base *iris* ya está precargada en R, porque podemos acceder a ella directamente.

Ejercicio 3

El archivo “govt_bond_returns.xlsx” contiene los rendimientos anuales para los inversores (a partir de septiembre de 2010) de bonos del gobierno, enumerados por duración y emisor. Estos retornos están influenciados por las respectivas tasas de inflación y fortalezas de las diversas monedas, así como las percepciones del mercado para estas tasas en el futuro. Realizar un agrupamiento jerárquico (*hierarchical clustering*) de estos retornos a identificar países con situaciones financieras similares. Comente como el dendrograma agrupa países en regiones geográficas similares.