José Carlos Girón Márquez

1064718

Hoja de Trabajo – Clustering

En la siguiente hoja de deberá utilizar las fuentes de datos indicadas para analizar la información usando R

Para cada uno de los ejercicios:

* Suba el archivo .R indicado
* Adjunte los PLOT como imagen en esta hoja de trabajo
* Responda a las preguntas indicadas

1. Ejercicio # 1
   1. Cree un nuevo script .R llamado “Clustering”
   2. Lea el archivo subido al portal llamado “Protein”
      1. Este archivo contiene un listado de diferentes observaciones y posee 10 variables que indican el consumo de proteínas en Europa y sus calificaciones según las variables.
   3. Debe de realizar un script que permita por medio de la técnica de agrupamiento jerárquico (con el método a su elección) y de K-Means realizar los cluster que mejor agrupen la información dada.
   4. Debe de crear y adjuntar los siguientes plots:
      1. Un plot con el dendograma del modelo jerárquico

Chart

Description automatically generated

* + 1. Un plot con los diferentes grupos clusters definidos (cuttree)

Diagram

Description automatically generated with low confidence

* + 1. Un plot utilizando factoextra para representar los k-means

Scatter chart

Description automatically generated

* 1. Debe de responder lo siguiente:
     1. ¿Cuál es el mejor número de clusters? ¿Por qué?

3 clústeres, ya que denota una mayor diferencia entre los datos disponibles, separando claramente áreas geográficas con centroides separados.

* + 1. ¿Cuáles serían las etiquetas para cada cluster?

El algoritmo parece haber detectado las tendencias regionales, por lo tanto, cada etiqueta representa partes de Europa:

* Europa Central (Rojo)
* Europa Oriental (Verde)
* Europa Occidental (Azul)

Shape

Description automatically generated

1. Ejercicio 2
   1. Comente dentro del script de “Clustering” la elaboración de este ejercicio.
   2. Obtengas un resumen de las principales variables estadísticas del dataset.
   3. Calcule la media, moda y desviación estardart de una variable a su elección.