

# MODELO DE CONGLOMERADOS PARA PAÍSES DE EUROPA SEGÚN SU NIVEL DE DESARROLLO HUMANO

LEONARDO MARTIN Y LAURA VILLALBA<sup>1</sup>

<sup>1</sup>Universidad Simón Bolívar  
Caracas – Venezuela

marzo del 2020

[Abstract] En este trabajo se presentan estudios sobre métodos de conglomerados para establecer una agrupación entre los países europeos, adscritos a las naciones unidas, que nos permitan conocer que países cuentan con características similares. Se utilizarán métodos jerárquicos y no jerárquicos, específicamente el método de *K-means*. Se llega a la conclusión de que los países europeos se pueden agrupar en 3 diferentes grupos.

## I. Análisis Exploratorio de los datos

Se realiza un análisis exploratorio de los datos verificando la clasificación de los países según: el nivel de *Human Development*, la correlación entre las variables y la distancia de Mahalanobis entre los países. En *Human Development* encontramos que los países se encuentran divididos en: *Very High Human Development (VHHD)*, *High Human Development (HHD)* y *Medium Human Development (MHD)*.

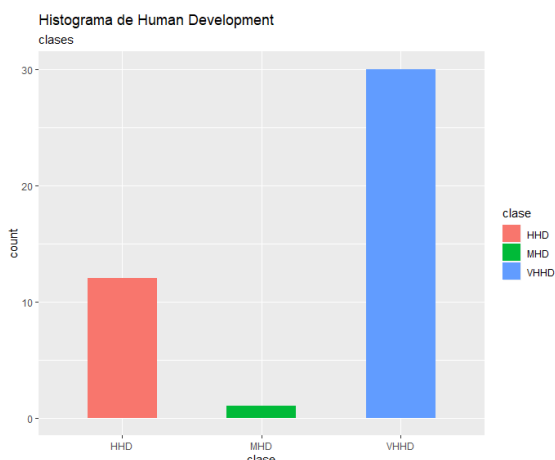


Figura 1: Frecuencia de los países en cada nivel del Human Development

Como podemos observar en la Fig.1 casi un 70 % de los países cuentan con un nivel muy alto de desarrollo humano, menos del 30 % se encuentran en un nivel alto de desarrollo humano y menos del 1 %, es decir, un solo país, tiene un desarrollo humano medio. Por lo tanto estamos en presencia de un continente con un alto desarrollo humano.

Por otro lado en la Fig.2 se observa que las variables de población y de labor, tanto de hombres como de mujeres, solo tienen correlación entre ellas mientras que las demás variables tienen una alta correlación lineal directa o inversa, es decir, que estas combinaciones de variables aportarán una mayor información en la formación de los grupos.

Por último al estudiar la distancia de Mahalanobis obtuvimos que: Estonia es el país con menor distancia, 26 países tienen una distancia de Mahalanobis menor que la distancia

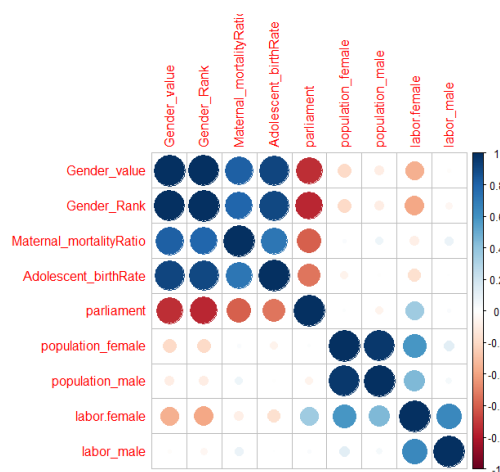


Figura 2: Correlación entre las variables

promedio, 17 países están por encima de la distancia promedio y Turquía es el país con mayor distancia de Mahalanobis.

## II. Análisis de Conglomerados

### Método Jerárquico

Estudiamos las agrupaciones de los países, en 4 grupos, a través de métodos de conglomerado jerárquicos utilizando la medida de proximidad del vecino mas lejano

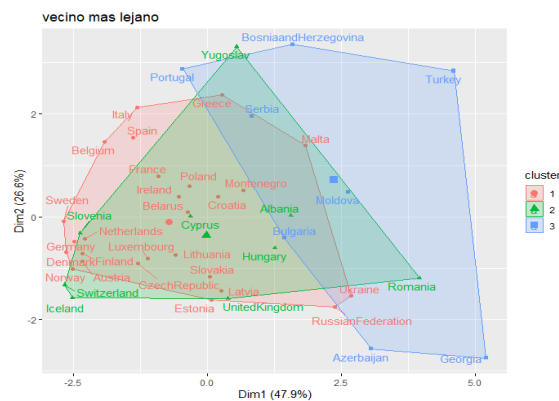


Figura 3: Agrupación de países por método jerárquico,  $k = 4$

Al ver la Fig.3 es claro que no se logran distinguir de for-

ma clara la división de los 4 grupos ya que países cercanos se encuentran agrupados con países más distantes interceptándose cada una de las agrupaciones.

Método No Jerárquico

Dada la falta de información que los métodos jerárquicos nos proporcionan para este problema utilizaremos el método no jerárquico, *K-means*, el cual utiliza el criterio de la incercia indicando el modelo óptimo a partir de una cantidad prefijada de grupos hallando una distancia intra grupal mínima y una distancia inter grupal máxima.

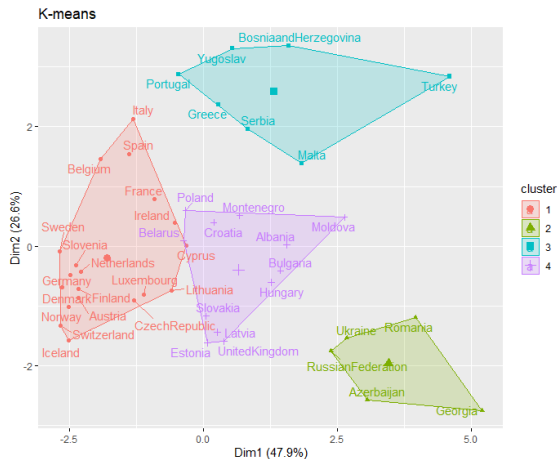


Figura 4: Agrupación de países por método K-means, k = 4

Ahora la Fig.4 nos muestra una clara agrupación para los países. Para este gráfico hemos dividido el continente en 4 grupos observando la mayor conglomeración de países en el sector rojo y que a su vez se encuentran cerca de ciertos países del sector morado y a una gran distancia de los países del sector amarillo. Finalmente, concluimos con el sector verde el cual posee a Turquía más alejado del resto de los países esto debido a que es el país con mayor distancia de Mahalanobis respecto al resto.

Selección óptima de clusters

Según diferentes índices como los de Hubert, Silhouette, Scott, entre otros, podemos identificar cuál es el número óptimo de clusters o de conglomerados con los que debemos trabajar el conjunto de datos.

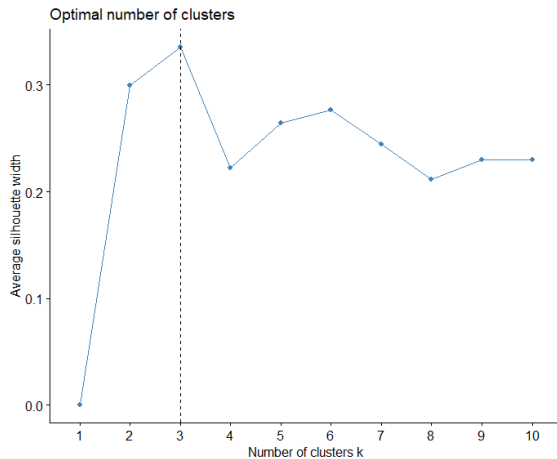


Figura 5: Índice de Silhouette

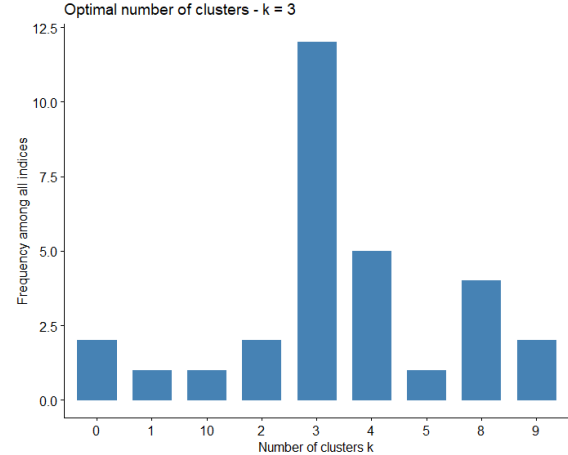


Figura 6: Frecuencia de decisión de número de clusters

Viendo las Figuras5-6 concluimos que la cantidad óptima de clusters es de 3 por consiguiente se realiza el análisis de *K-means* con k = 3.

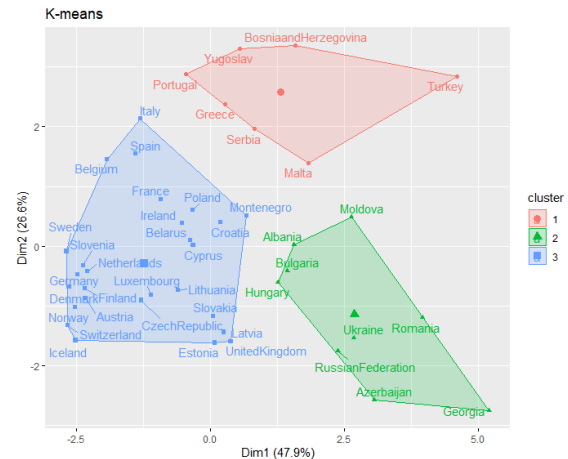


Figura 7: Agrupación de países por método K-means, k = 3

Al hacer la agrupación en 3 conglomerados (Fig.7) observamos como el sector de la parte superior no se modificó mientras que los países del sector izquierdo se juntaron con algunos países del sector morado de la Fig.4, por otro lado los países del sector inferior derecho de la Fig.4 se juntaron con los países restantes del sector morado.

Realizando un mapa de calor podemos confirmar los datos obtenidos con el análisis del método no jerárquico de *K-means*.

Europa	
Grupo	País
1	Norway, Switzerland, Denmark, Netherlands, Ireland
	Germany, UnitedKingdom, Sweden, Iceland, Luxembourg
	Belgium, France, Austria, Finland, Slovenia, Spain, Italy
	CzechRepublic, Estonia, Cyprus, Slovakia, Poland
	Lithuania, Latvia, Croatia, Montenegro, Belarus
2	Hungary, RussianFederation, Romania, Bulgaria, Ukraine
	Albania, Azerbaijan, Moldova, Georgia
3	Greece, Malta, Portugal, Serbia, Yugoslav, Bosnia, Turkey

TABLA 1: Agrupación de los países.

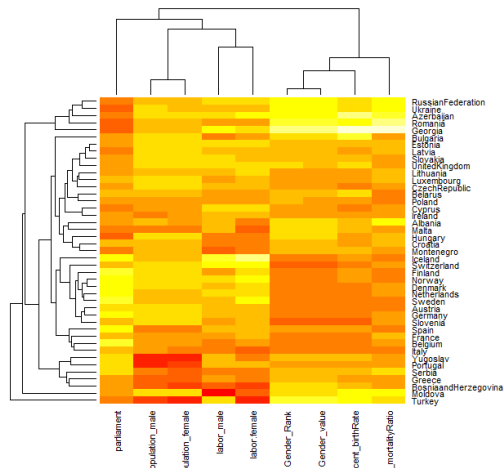


Figura 8: Mapa de calor

Si comparamos el mapa de calor (Fig.8) con el método de *K-means* (Fig.7) podemos confirmar las agrupaciones realizadas por este último método.

### III. Conclusión

Podemos concluir, gracias al método no jerárquico de *K-means*, que la mejor forma de agrupar al continente europeo es a través 3 grupos (ver Tabla.1) Como nos encontramos en un continente que es de alto nivel de desarrollo humano no es posible dividir los grupos según esa clasificación por lo tanto la relación en la agrupación está totalmente influenciada por las variables y cómo los países están correlacionadas con cada una de ellas, razón por la cual el mapa de calor (Fig.8) fue el gráfico final que verificó al método de *K-means*.