EJERCICIO DE EVALUACIÓN I

ANÁLISIS DE COMPONENTES PRINCIPALES Y CLUSTER

El fichero Provincias contiene información socio-económica de las provincias españolas. (Se han modificado los nombres de las variables). Para reducir el número de variables e intentar encontrar relaciones, tanto entre variables como entre provincias, realizar los siguientes apartados.

1. **Calcular l**a matriz de correlaciones, y su representación gráfica ¿Cuáles son las variables más correlacionadas de forma inversa?
2. Realizar un análisis de componentes principales sobre la matriz de correlaciones, **calculando 7 componentes**. Estudiar los valores de los autovalores obtenidos y las gráficas que los resumen. ¿Cuál es el número adecuado de componentes?
3. Hacer de nuevo el análisis sobre la matriz de correlaciones pero ahora **indicando el número de componentes principales que hemos decidido retener** (Que expliquen aproximadamente el 90%)**.** Sobre este análisis contestar los siguientes apartados.
    - **a.** Mostrar los coeficientes para obtener las componentes principales ¿Cuál es la expresión para calcular la primera Componente en función de las variables originales?
    - b. Mostar una tabla con las correlaciones de las Variables con las Componentes Principales. Para cada Componente indicar las variables con las que está más correlacionada
    - **c.** Comentar los gráficos que representan las variables en los planos formados por las componentes, intentando explicar lo que representa cada componente
    - **d.** Mostrar la tabla y los gráficos que nos muestran la proporción de la varianza de cada variable que es explicado por cada componente. ¿Cuál de las variables es la que está peor explicada?
    - **e.** Mostrar la tabla y los gráficos que nos muestran el porcentaje de la varianza de cada Componente que es debido a cada variable. ¿Que variables contribuyen más a cada Componente?
    - f. Sobre los gráficos que representan las observaciones en los nuevos ejes y el gráfico Biplot., teniendo en cuenta la posición de las provincias en el gráfico Comentar las provincias que tienen una posición más destacada en cada componente, en positivo o negativo, ¿Qué significa esto en términos socioeconómicos para estas provincias?
    - g. Si tuviéramos que construir un índice que valore de forma conjunta el desarrollo económico de una provincia, como se podría construir utilizando

una combinación lineal de todas las variables. ¿Cuál sería el valor de dicho índice en Madrid? ¿Cual sería su valor en Melilla?

4. Representar un mapa de calor de la matriz de datos, estandarizado y sin estandarizar para ver si se detectan inicialmente grupos de provincias.
5. Realizar un análisis Jerárquico de clusters para determinar si existen grupos de provincias con comportamiento similar.
    a. A la vista del dendrograma ¿Cuántos clusters recomendarías?
    b. Representar los individuos agrupados según el número de clusters elegido.
    c. ¿Qué número óptimo de clusters nos indican los criterios Silhoutte y de Elbow?
    d. Con el número de clusters decidido en el apartado anterior realizar un agrupamiento no jerárquico.
        i. Representar los clusters formados en los planos de las Componentes principales. Relacionar la posición de cada cluster en el plano con lo que representa cada componente principal.
        ii. Evaluar la calidad de los clusters
    e. Explicar las provincias que forman cada uno de los clusters y comentar cuales son las características socioeconómicas que las hacen pertenecer a dicho cluster.