

EJERCICIO DE CLASE

ANÁLISIS DE COMPONENTES PRINCIPALES

El fichero BARRIOS contiene información socio-económica de algunos barrios de Madrid.

Las variables contenidas en el fichero son:

P_TOT	Población Total en miles
P_14	Porcentaje de población menor de 14 años
P_65	Porcentaje de población mayor de 65 años
ANAL	Porcentaje de población analfabeta
NES	Porcentaje de población con nivel de estudios superiores
OCU	Porcentaje de población ocupada
OCUIN	Porcentaje de población ocupada en industria
OCUSER	Porcentaje de población ocupada en sector servicios
TEC	Porcentaje de población en puestos técnicos
PD	Porcentaje de población en puestos directivos
TM	Porcentaje de población ocupada en trabajos manuales

Para reducir el número de variables e intentar encontrar relaciones, tanto entre variables como entre barrios, realizar los siguientes apartados.

1. Calcular los estadísticos básicos de todas las variables. Comparar sus medias y varianzas.
1. Representar los gráficos de dispersión de las variables (NES,ANAL) y (ANAL,TM).
Comentar el sentido de la nube de puntos
2. **Calcular** la matriz de correlaciones, y su representación gráfica ¿Cuáles son las variables más correlacionadas? ¿Cómo es el sentido de esa correlación?
3. 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 para explicar adecuadamente los datos? ¿Qué porcentaje de variabilidad queda explicado?
4. 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) Mostrar 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. ¿Qué grupos de variables contribuyen más a cada Componente?
- f) Sobre los gráficos que representan las observaciones en los nuevos ejes, y ayudándonos del biplot, teniendo en cuenta la posición de los barrios en el gráfico. Comentar las características socioeconómicas de algunos grupos de barrios
- g) ¿Qué valor tiene Salamanca en la Componente 2?, ¿Y Villaverde?, ¿Qué barrio tiene un valor más alto de la Componente 3?