

Máster Universitario en Ciencia de Datos – CUNEF Universidad

Fundamentos para el Análisis de Datos y la Investigación - Curso 2023-24

Directrices Trabajo 2 en Grupo de Fundamentos: PCA

Objetivo:

Realizar un Análisis en Componentes Principales, PCA, haciendo uso de un conjunto de variables cuantitativas relacionadas, que describan algún fenómeno en ciencia de datos. Se realizará eligiendo al menos 5 variables y como mínimo 17 observaciones por variable.

Algunos ejemplos:

1. Fenómeno turístico: variables que definen la actividad turística a partir de datos de municipios de una CCAA.

https://datos.comunidad.madrid/catalogo/dataset/municipio_comunidad_madrid/resource/e750429-1e05-411a-b026-a57ea452a34a

2. Estudio de un tipo de automóvil, según sus características.

<https://coches.idae.es/base-datos>

3. Estadísticas de equipos de fútbol a partir de una serie de variables:

<https://es.whoscored.com/statistics>

4. Estadísticas de jugadores de fútbol o baloncesto a partir de una serie de estadísticas del jugador

<https://es.whoscored.com/statistics>

<https://espndeportes.espn.com/basquetbol/nba/estadisticas/equipo>

5. Indicadores del sistema nacional de salud (por CCAA)

<http://inclasns.msssi.es/main.html>

6. Bases de datos abiertas:

<https://www.comunidad.madrid/gobierno/datos-abiertos>

<https://datos.madrid.es/portal/site/egob>

El análisis debe incluir los aspectos que aparecen en los scripts de clase sobre PCA, es decir:

- Análisis descriptivo de las variables: gráficos (diagramas matriciales, etc.) y numéricos (matrices correlaciones, etc.)
- Contrastes de validez: test de Bartlett, KMO
- Elegir número de factores (grafico de sedimentación, porcentaje de varianza,...)
- Interpretación de los factores y porcentajes de explicación de cada factor en términos de la varianza.
- Factores, cargas factoriales y puntuaciones

- Gráficos biplot
- Rotaciones
- Indices

Aspectos a valorar:

- Análisis PCA y gráficos
- Presentación
- Cualquier aspecto innovador o diferenciador en cuanto a las metodologías utilizadas