

Estadística en Ciencia de Datos

Ejercicio: Análisis de Componentes Principales (ACP)

Base de datos: *Student Performance* (UCI)

Contexto

La base de datos *Student Performance* contiene información académica, familiar y socioeconómica de estudiantes de educación secundaria. Dado el gran número de variables explicativas potencialmente correlacionadas, resulta de interés aplicar técnicas de reducción de dimensionalidad.

El objetivo de este ejercicio es analizar la estructura de dependencia entre las variables mediante el *Análisis de Componentes Principales* (ACP).

Variables consideradas

Considere las siguientes variables cuantitativas:

- edad
- tiempo_estudio
- inasistencias
- reprobaciones_previas
- educacion_madre
- educacion_padre
- salud
- tiempo_libre
- salidas_sociales
- alcohol_dias_semana
- alcohol_fin_semana

Todas las variables deben ser tratadas como numéricas.

Preguntas

- a) Explique por qué es necesario estandarizar las variables antes de aplicar el Análisis de Componentes Principales en este contexto.
 - b) Calcule la matriz de correlaciones entre las variables y comente brevemente la presencia de relaciones lineales relevantes.
 - c) Aplique el Análisis de Componentes Principales utilizando la matriz de correlaciones.
 - d) Presente el *scree plot* e indique cuántos componentes principales deberían retenerse según el criterio de la varianza explicada acumulada (80 % o 90 %).
 - e) Interprete los dos primeros componentes principales en términos de las variables originales, analizando los signos y magnitudes de las cargas (*loadings*).
 - f) Represente gráficamente los individuos en el plano definido por los dos primeros componentes principales e identifique posibles patrones o agrupamientos.
 - g) Discuta si los componentes principales obtenidos podrían ser utilizados posteriormente como variables explicativas en un modelo de regresión.
-

Indicaciones técnicas

- Utilice variables estandarizadas.
 - Justifique el número de componentes retenidos.
 - Acompañe la interpretación con gráficos adecuados.
-

Objetivo pedagógico

Este ejercicio evalúa la capacidad del estudiante para:

- Comprender la motivación del ACP,
- Aplicar correctamente la técnica,
- Interpretar componentes principales,
- Integrar reducción de dimensionalidad con modelos estadísticos.