

Primera Tarea de Análisis Multivariante

Construir un conjunto de datos multivariantes heterogéneos o de tipo mixto, es decir, con variables cuantitativas, variables binarias y variables categóricas multi-estado. La información debe ser de corte transversal, es decir, las series temporales no son válidas. Por ejemplo, se puede recopilar información sobre:

- Las provincias españolas: variables socio-económicas, demográficas, inversión pública en distintos servicios para el bienestar de los ciudadanos (salud, educación, seguridad y justicia, medio ambiente, etc.), mortalidad por COVID en año concreto, tasa de vacunación en un año concreto, etc.
- Análogamente sobre los países de la Unión Europea.
- O también sobre un conjunto de (unos 30) países del mundo (por ejemplo, los de mayor GDP, los de mayor calidad democrática, los de mayor calidad de vida, etc.).
- Las empresas del IBEX-35: número de empleados, sueldo medio de los empleados, sueldo medio del consejo de dirección, cotización bursátil, capital de la empresa, etc.
- Y otras posibilidades que se os ocurran.

Es preferible que construyáis el conjunto de datos vosotros a que utilizéis unos datos ya disponibles. Por ejemplo, podéis encontrar variables interesantes en el INE (www.ine.es), Eurostat (www.eurostat.es), Our World in Data de la Universidad de Oxford (<https://ourworldindata.org/>), Banco Mundial (<https://datos.bancomundial.org/>), Naciones Unidas (<https://data.un.org/>), OCDE (<https://data.oecd.org/>).

1. Con las variables cuantitativas del conjunto de datos:
 - a) Realizar una descriptiva multivariante: vector de medias, matrices de covarianzas y correlaciones, gráfico de dispersión matricial.
 - b) Analizar si las variables siguen una ley normal. En caso contrario, aplicar transformaciones no lineales para intentar simetrizar las distribuciones de datos. Realizar de nuevo la descriptiva multivariante: vector de medias, matrices de covarianzas y correlaciones, gráfico de dispersión matricial.
 - c) Obtener medidas escalares de dispersión: variación total, variación generalizada y η^2 para las variables originales y para las variables transformadas.
2. Seleccionar una variable binaria:
 - a) Con todas (o varias) de las variables cuantitativas transformadas, realizar un gráfico de dispersión matricial donde se distingan los dos grupos que establece la variable binaria (usar la instrucción `gplotmatrix`).
 - b) Para los dos grupos, realizar un contraste de comparación de medias, mediante el estadístico T^2 de Hotelling.
3. Seleccionar una variable categórica multi-estado (es decir, de más de dos categorías):
 - a) Con todas (o varias) de las variables cuantitativas transformadas, realizar un gráfico de dispersión matricial donde se distingan los grupos que establece la variable categórica (usar la instrucción `gplotmatrix`).
 - b) Para los distintos grupos, realizar un contraste de comparación de medias, mediante el estadístico Lambda de Wilks.

Redactar un informe donde se describa el conjunto de datos y se expliquen los resultados y principales conclusiones de los ejercicios 1, 2 y 3.