

Análisis de Datos Multivariantes.

Hoja nº 2 de problemas y prácticas R

1. El fichero Familia.txt contiene el peso (kgs) y la altura (cms) de los integrantes de una familia.
 - i) Leer en R este fichero y calcular la media y la cuasidesviación típica de cada variable.
 - ii) Dibujar la nube de puntos (Peso, Altura) y superponer la recta de mínimos cuadrados. Calcular el coeficiente de correlación lineal entre ambas variables.
 - iii) ¿Qué observación es *outlier* para las variable peso y altura?
 - iv) Repetir el apartado ii) sin esa observación.
 - v) Calcular el Índice de Masa Corporal (IMC), definido como el cociente entre el peso y el cuadrado de la altura (en metros). Representarlo con un gráfico de barras.
 - vi) Ordenar los familiares de mayor a menor IMC.
 - vii) Seleccionar aquellos familiares que tengan un IMC en el intervalo (20,25). Mostrarlos ordenados de mayor a menor IMC.
2. En este problema se trabajará con el conjunto de datos Boston de la librería MASS.
 - i) Acceder a los datos, en particular, averiguar qué información contiene y cuál es la dimensionalidad de los datos.
 - ii) Ordenar las variables según el valor absoluto de su coeficiente de correlación lineal con *medv* (variable a predecir en este conjunto de datos).
 - iii) ¿Destaca algún distrito por su tasa de criminalidad? Similarmente, por los impuestos sobre la propiedad o por la ratio alumnos-profesor.
 - iv) ¿Cuántos distritos son limítrofes con el río? Calcular las medias de *crim* y *medv* según chas.
 - v) Analizar la relación lineal entre *lstat* y *medv*.
3. Acceder al fichero de datos USairpollution de la librería HSAUR2.
 - i) Generar las nubes de puntos para cada par de variables. Calcular la matriz de correlaciones.
 - ii) Obtener la nube de puntos para las variables *manu* y *popul*.
 - iii) Añadir a la nube de puntos anterior un gráfico caja y bigotes de *popul* y un histograma de *manu*.
4. Leer el fichero de datos decathlon1989.sav (formato SPSS).
 - i) Resumir numéricamente las variables.
 - ii) Dibujar todos los gráficos de caja y bigotes por separado.
 - iii) Generar las nubes de puntos y añadir histogramas en la diagonal, superponer rectas de minimos cuadrados o suavizaciones o bien mezclar las nubes de puntos con los valores de las correlaciones.
 - iv) Utilizar la librería *corrplot* para visualizar la matriz de correlaciones.
 - v) Obtener los principales estadísticos de forma directa y con funciones R.
5. Cargar el espacio de trabajo “Pisa2009.RData”. Dibujar representaciones multivariantes como las caras de Chernoff o los gráficos “star”.
6. Leer el fichero “dietas.dat”, donde se distinguen 4 dietas alimenticias, y se mide el peso durante 6 meses. Cada variable toma valores de 1 a 6, a mayor valor más lejos del peso ideal.
 - i) Representar gráficamente la evolución de los pesos medios según la dieta.
 - ii) Para cada individuo, calcular el número de meses en los que su distancia al peso ideal es superior 3. Tabular esta variable.
7. Ilustrar el Teorema de Fisher a partir de 5000 muestras de tamaño 10 de una ley $N(0,1)$.