

Ejercicio práctico: Estadística descriptiva

Sesión 2. Curso: ‘Análisis estadístico aplicado con R Commander’

Ejercicio 1.

En un estudio sobre el cáncer de próstata se desea estudiar la relación entre el antígeno prostático específico (PSA) y un amplio número de indicadores clínicos de pronóstico en hombres con cáncer de próstata en estado avanzado. Los datos se encuentran en el fichero `cancer.prostata.RData` y corresponden a 97 hombres a los que se les había realizado una prostatectomía radical. Cada fila del banco de datos corresponde a uno de estos pacientes y contiene el código del enfermo e información de aquellas variables consideradas inicialmente como relevantes:

- **caso**: código del paciente.
- **psa**: nivel de PSA (en mg/ml).
- **volumen**: volumen estimado del cáncer de próstata (en cc).
- **prostata**: peso de la próstata (en gramos).
- **edad**: edad del enfermo (en años).
- **hyper**: cantidad de hiperplasia prostática benigna (en cm^2).
- **invasion**: presencia o ausencia de invasión en la vesícula seminal (si/no).
- **penetracion**: grado de penetración capsular (en cm).
- **grado**: grado de la enfermedad (los valores altos indican peores pronósticos).

Genera un informe utilizando R Markdown que contenga lo siguiente:

- a) Describe las variables del estudio indicando su tipo.
- b) Realiza un análisis descriptivo numérico de cada una de las variables del banco de datos. Comenta los resultados obtenidos.
- c) Representa el histograma y el diagrama de cajas de la variable **psa**. ¿Cómo es la forma de la distribución de los niveles de PSA? ¿Hay individuos que presentan valores atípicos del nivel de PSA?
- d) Realiza un análisis descriptivo numérico y gráfico que permita evaluar la relación entre el nivel de PSA y el resto de variables del estudio. Comenta los resultados obtenidos. ¿Encuentras indicios de la posible existencia de una relación entre el nivel de PSA y el resto de variables?