

# Curso: Análisis estadístico aplicado con R Commander

## Ejercicio práctico: Regresión logística y regresión de Poisson

### Ejercicio 1.

Se desea estudiar la posible asociación entre el cáncer de vejiga, el consumo de café y el ambiente de residencia. Para ello se eligieron 100 pacientes de forma aleatoria de los que 50 tiene cáncer. Las variables del banco de datos son: CANCER (no/si), CAFE (no consume/si consumo) y MEDIO (rural y urbano). Los datos se encuentran en el fichero `cancer.RData`.

- Describe las variables del estudio indicando su tipo e identifica cuál es la variable respuesta y cuáles son las variables explicativas. ¿Qué tipo de distribución sigue la variable respuesta?
- Ajusta un modelo de regresión adecuado para evaluar si existe relación entre padecer cáncer de vejiga, consumir café y vivir en un entorno rural o urbano. ¿Sería posible eliminar del modelo alguna de las variables explicativas? En caso afirmativo, realiza un procedimiento de selección de variables paso a paso para elegir el mejor modelo.
- Interpreta los coeficientes del modelo final ajustado y su implicación.
- Valora la calidad del ajuste del modelo en base a su capacidad discriminatoria. Para ello:
  - Representa la curva ROC del modelo de regresión ajustado. ¿Cuánto vale el área bajo la curva ROC? ¿Qué representa esta medida?
  - ¿Cuál es el *cut-point* que permite una mayor discriminación entre los individuos que presentan cáncer y los que no?
  - Para el *cut-point* seleccionado, ¿cuál es la sensibilidad y la especificidad del modelo de regresión logística? Interpreta estos valores.
- Valora la calidad del ajuste del modelo mediante el test de Hosmer-Lemeshow.

### Ejercicio 2.

Un especialista en geriatría diseñó un experimento prospectivo para investigar los efectos de dos tratamientos en la frecuencia de caídas. Se asignó aleatoriamente los dos tratamientos (educación/ educación y entrenamiento aeróbico) a cien personas de al menos 65 años y en buena salud. Se observó en cada individuo el número de caídas en 6 meses. Se consideraron tres variables control de interés: sexo (mujer/hombre), un índice de equilibrio (mayor valor indicando mayor estabilidad) y un índice de fuerza (mayor índice indicando mayor fuerza). Los datos se encuentran en el fichero `caidas.RData`.

- Describe las variables del estudio indicando su tipo e identifica cuál es la variable respuesta y cuáles son las variables explicativas. ¿Qué tipo de distribución sigue la variable respuesta?
- Ajusta un modelo de regresión adecuado para evaluar si existe relación entre el número de caídas que sufren los individuos y el tratamiento recibido teniendo en cuenta el sexo, el índice de equilibrio y el índice de fuerza de dichos individuos. ¿Sería posible eliminar del modelo alguna de las variables explicativas? En caso afirmativo, realiza un procedimiento de selección de variables para elegir el mejor modelo.
- Interpreta los coeficientes del modelo final ajustado y su implicación. Por ejemplo, ¿el ejercicio aeróbico reduce las caídas?

- d) Valora la calidad del ajuste del modelo en base al valor de la DEVIANCE del modelo en comparación con la DEVIANCE de un modelo NULO y en base a la magnitud de los residuos DEVIANCE.