

## Práctica 5. Modelos Lineales Generalizados.

1. En el archivo `chdage.dat` aparecen los datos de edad (`age`) y de presencia o ausencia de una enfermedad coronaria (`chd`) de 100 sujetos seleccionados para participar en un estudio. También aparece el número identificativo que se le asignó a cada sujeto. Nuestro interés se centra en estudiar la relación entre la edad y la presencia de la enfermedad coronaria. Ajusta a estos datos un modelo de regresión logística, dando las estimaciones de los parámetros, así como intervalos de confianza al 95% para los mismos. ¿Podemos decir que la edad es un factor de riesgo significativo en la presencia de la enfermedad coronaria? ¿Qué probabilidad predice el modelo de presentar la enfermedad para una persona de 28 años? ¿Y para uno de 76? ¿Qué podemos decir de la bondad del ajuste? Razona las respuestas.
2. El archivo `icu.dat` contiene los datos de 200 individuos que formaron parte de estudio mucho mayor sobre la supervivencia de los pacientes en una Unidad de Cuidados Intensivos (UCI). La descripción de cada una de las variables aparece en la hoja anexa. El objetivo del estudio era la predicción de la probabilidad de supervivencia de un paciente tras ser ingresado en la UCI. Ajustar un modelo de regresión logística de la variable `sta` sobre las variables `age`, `can`, `cpr` (Cardiopulmonary Resuscitation), `inf` y `race`. Obtener los intervalos de confianza al 95% para dichos parámetros. ¿Podemos decir que el modelo ajusta suficientemente bien? ¿De entre las variables que hemos elegido nosotros cuáles resultan significativas? Si utilizamos un procedimiento *forward* o *backward*, ¿qué modelo sería el resultante? ¿Qué probabilidad de morir predice el modelo para un paciente de 47 años, de raza blanca, que no tenga cáncer, pero que haya requerido de CPR y que tenga síntomas de infección? Razona las respuestas.
3. En el archivo `mortal.dat` se muestran los datos relativos al número de insectos muertos tras cinco horas de ser expuestos al carbono disulfido,  $CS_2$ , en varias concentraciones. Las variables que aparecen son `dos1s`,  $\log_{10}CS_2$  mg/l, `insectos`, número de insectos expuestos y `muertos`, número de insectos que murieron tras 5 horas de exposición. Ajustar modelos logit, probit y loglog complementarios y determinar cuál es a vuestro juicio el que mejor ajusta los datos.
4. El número de muertes por Sida en Australia en períodos de tres meses entre 1983 y 1986 aparecen en el archivo `sida.dat`. Ajustar un modelo de regresión de Poisson a dichos datos: determinar los coeficientes e intervalos de confianza al 95% para los mismos. Determinar la bondad del ajuste. ¿Resulta significativa la variable periodo en la mortalidad?
5. Para desarrollar un estudio sobre la relación existente entre el estrés y la posibilidad de adquirir una enfermedad, se eligió aleatoriamente un miembro de una familia, también elegida aleatoriamente en Oakland, California, USA. A cada uno de estos individuos se les preguntó, dentro de una lista de 41 sucesos estresantes, cuáles le habían ocurrido en los últimos 18 meses anteriores a la entrevista. Los resultados, que aparecen en el archivo `estres.dat`, muestran el mes contado hacia atrás desde el momento de la entrevista y el número de individuos que recordaban uno solo de dichos sucesos. Ajustar un modelo de regresión de Poisson a los datos, extrayendo toda la información que te sea posible.