PRÁCTICA 7: Tablas de contingencia y regresión (2).

DESARROLLO DE LA PRÁCTICA

En la práctica de esta semana, vamos a tomar varios conjuntos de datos y realizar tareas relativas a las dos prácticas anteriores, cuyo contenido será evaluado en el segundo cuestionario. A saber, tablas de contingencia para variables cualitativas, y estadísticos descriptivos de una variable continua para cada nivel de una variable cualitativa (factor); y de otro lado realizar diversos ajustes de regresión entre varias variables y con la posibilidad de realizarlo para determinadas subpoblaciones.

Primer conjunto de datos
DATA SET de la librería ISLR: **Auto**

Description
Gas mileage, horsepower, and other information for 392 vehicles.
Format
A data frame with 392 observations on the following 9 variables.
mpg: miles per gallon
cylinders: Number of cylinders between 4 and 8
displacement: Engine displacement (cu. inches)
horsepower: Engine horsepower
weight: Vehicle weight (lbs.)
acceleration: Time to accelerate from 0 to 60 mph (sec.)
year: Model year (modulo 100)
origin: Origin of car (1. American, 2. European, 3. Japanese)
name: Vehicle name

The original data contained 408 observations but 16 observations with missing values were removed.

Para este conjunto de datos, se pueden cambiar de unidades americanas a unidades del sistema métrico, las variables: mpg, displacement, weight y acceleration
(1 Milla por galón= 0.425144 km por litro, 1 libra= 0.453592 kg y 1 pulgada cúbica = 16.39 centímetros cúbicos)

CUESTIONES:
a) Entre las variables continuas (mpg, displacement, horsepower, weight) existe una relación lineal o parabólica que sea de interés debido a su fortaleza.
b) Entre las variables continua se mantienen los mismos valores (media, mediana, sd ,...) según la procedencia del vehículo
c) ¿El número de cilindros depende del año de fabricación?

Segundo conjunto de datos
DATA SET de la librería COUNT: **titanic**

Description
The data is an observation-based version of the 1912 Titanic passenger survival log,
Format
A data frame with 1316 observations on the following 4 variables.

Class: a factor with levels 1st class 2nd class 3rd class crew
Age: a factor with levels child adults
Sex: a factor with levels women man
Survived: a factor with levels no yes

Source
Found in many other texts

CUESTIONES:
a) ¿Las personas que murieron dependen de la clase en que estaban alojados?
b) ¿Las personas que murieron dependen de su edad?
c) ¿Las personas que murieron dependen de su sexo
d) Dependiendo de los resultados anteriores analizar la supervivencia con otras dos variables más (tres variables a analizar)

Tercer conjunto de datos
DATA SET de la librería datasets: **iris**

Description
This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are Iris setosa, versicolor, and virginica.
Format
iris is a data frame with 150 cases (rows) and 5 variables (columns) named Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

Sepal.Length: longitud del sépalo.
Sepal.Width: anchura del sépalo
Petal.Length: longitud del pétalo.
Petal.Width: anchura del pétalo.
Species: especie de lirio.

Source
Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, Part II, 179–188.

CUESTIONES:
a) Entre las variables continuas (Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) existe una relación lineal o parabólica que sea de interés debido a su fortaleza.
b) Entre las variables continua se mantienen los mismos valores (media, mediana, sd ,...) según la especie de lirio considerada
c) En base a los resultados del apartado b) puedes considerar que alguna(s) variable(s) continua se ve muy afectada por el tipo de lirio.