



Introducción al modelo lineal

Francesc Carmona y Alex Sánchez

19 de septiembre de 2020

Los ejercicios con (*) son opcionales, con (**) además son difíciles.

Ejercicios del libro de Faraway

1. (Ejercicio 1 cap. 1 pág. 12)

The dataset **teengamb** concerns a study of teenage gambling in Britain. Make a numerical and graphical summary of the data, commenting on any features that you find interesting. Limit the output you present to a quantity that a busy reader would find sufficient to get a basic understanding of the data.

2. (Ejercicio 2 pág. 12)

The dataset **uswages** is drawn as a sample from the Current Population Survey in 1988. Make a numerical and graphical summary of the data as in the previous question.

3. (*) (Ejercicio 3 pág. 12)

The dataset **prostate** is from a study on 97 men with prostate cancer who were due to receive a radical prostatectomy. Make a numerical and graphical summary of the data as in the first question.

4. (*) (Ejercicio 4 pág. 12)

The dataset **sat** comes from a study entitled “Getting What You Pay For: The Debate Over Equity in Public School Expenditures.” Make a numerical and graphical summary of the data as in the first question.

5. (*) (Ejercicio 5 pág. 12)

The dataset **divusa** contains data on divorces in the United States from 1920 to 1996. Make a numerical and graphical summary of the data as in the first question.

Ejercicios del libro de Carmona

1. (*) (Ejercicio 1.1 del Capítulo 1 página 24)

Hallar las estimaciones de los parámetros en un modelo de regresión lineal simple, minimizando la suma de los cuadrados de los errores:

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Hallar también una expresión para las predicciones \hat{y}_i y los residuos $e_i = y_i - \hat{y}_i$.

Nota: En el apartado 1.4 del libro de Carmona se indica que para minimizar S debemos verla como una función $S = f(\beta_0, \beta_1)$ de los parámetros ya que los valores de x_i y y_i son constantes conocidas

(las observaciones). Entonces debemos derivar S respecto a los dos parámetros, igualar a cero ambas derivadas y resolver el sistema formado por las dos ecuaciones.

$$\begin{cases} \partial S / \partial \beta_0 = 0 \\ \partial S / \partial \beta_1 = 0 \end{cases}$$

2. (**) (Ejercicio 1.2 del Capítulo 1 página 24)

Hallar las estimaciones de los parámetros en un modelo de regresión parabólico, minimizando la suma de los cuadrados de los errores:

$$S = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$$

Hallar también una expresión para las predicciones \hat{y}_i y los residuos $e_i = y_i - \hat{y}_i$.

3. (Ejercicio 1.3 del Capítulo 1 página 24)

Consideremos el problema de tráfico planteado en el apartado 1.2 de este capítulo, con la variable independiente *densidad* y la variable dependiente *raíz cuadrada de la velocidad*. Con los datos proporcionados en la tabla 1.1

```
dens <- c(12.7, 17.0, 66.0, 50.0, 87.8, 81.4, 75.6, 66.2, 81.1, 62.8, 77.0, 89.6,
          18.3, 19.1, 16.5, 22.2, 18.6, 66.0, 60.3, 56.0, 66.3, 61.7, 66.6, 67.8)
vel <- c(62.4, 50.7, 17.1, 25.9, 12.4, 13.4, 13.7, 17.9, 13.8, 17.9, 15.8, 12.6,
         51.2, 50.8, 54.7, 46.5, 46.3, 16.9, 19.8, 21.2, 18.3, 18.0, 16.6, 18.3)
rvel <- sqrt(vel)
```

realizar el siguiente proceso:

- Dibujar la nube de puntos y la recta que pasa por los puntos $(12.7, \sqrt{62.4})$ y $(87.8, \sqrt{12.4})$. Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos.
- Hallar la recta de regresión simple. Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos.
- Mejorar el modelo anterior considerando una regresión parabólica. Dibujar el gráfico de los residuos con la densidad y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos.
- Calcular la capacidad de la carretera o punto de máximo flujo. Recordar que $\text{flujo} = \text{vel} \times \text{densidad}$.

4. (Ejercicio 1.4 del Capítulo 1 página 24)

La siguiente tabla contiene los mejores tiempos conseguidos en algunas pruebas de velocidad en atletismo en los Juegos Olímpicos de Atlanta:

distancia	tiempo	
	hombres	mujeres
100	9.84	10.94
200	19.32	22.12
400	43.19	48.25
800	102.58	117.73
1500	215.78	240.83
5000	787.96	899.88
10000	1627.34	1861.63
42192	7956.00	8765.00

Si tomamos como variable regresora o independiente la distancia (metros) y como variable respuesta o dependiente el tiempo (segundos):

- (a) Calcular la recta de regresión simple con los datos de los hombres y dibujarla. Dibujar el gráfico de los residuos con la distancia y el gráfico con las predicciones. Calcular la suma de cuadrados de los residuos y el R^2 .
- (b) Repetir el apartado anterior utilizando los logaritmos de las variables tiempo y distancia.
- (c) Repetir los dos apartados anteriores utilizando los datos de las mujeres.

Otros ejercicios

1. Con los datos de la tensión arterial sistólica y la edad de los 69 pacientes que podemos encontrar en la web de www.fisterra.com
https://www.fisterra.com/mbe/investiga/regre_lineal_simple/regre_lineal_simple.asp
calcular los coeficientes de regresión de la recta mínimo cuadrática.