

## Escuela de Ingeniería Informática

### Métodos Estadísticos, Curso 2019-20

### Ejercicios Prácticos Lab7

**Ejercicio 1:** El fichero “*Aloe\_Vera.txt*” contiene datos de cuatro variedades de plantas de Aloe obtenidas de una plantación experimental.

- a) Estudiar las variedades que dan más rendimiento desde el punto de vista de su masa y masa seca.
- b) Analizar las dependencias entre la masa y la altura de la variedad “*barbadensis*”
- c) Estimar el modelo de regresión con la función *lm*.
- d) Analizar el modelo estimado con la función *summary* y obtener un posible intervalo de confianza para las conclusiones de los distintos parámetros.
- e) Evaluar una predicción para una masa de  $x_0=5.1$  gramos y encontrar un intervalo de confianza para la misma.
- f) Encontrar el coeficiente de determinación  $R^2$
- g) Realizar un análisis de varianza para estudiar la bondad del ajuste y la linealidad de la regresión. Explicar los resultados obtenidos.
- h) **(Opcional)** Analizar si fuera posible aplicar el estudio anterior y la suposición de homocedasticidad (varianza constante a lo largo de las observaciones) al caso de analizar las relaciones entre la masa y la masa seca para la variedad “*saponaria*”. Utilizar el test de White (variedad del test de Breusch-Pagan *bptest*, del paquete *lmtest*). Explicar las conclusiones.

**Ejercicio 2. (Opcional)** Un comerciante minorista de la zona de Triana quiere analizar la influencia de los costos de publicidad en sus ventas. Durante 3 meses evalúa los costos semanales correspondientes, que se detallan en la siguiente tabla:

Costos de Publicidad (€)	40	20	25	30	30	50	40	20	50	40	25	50
Ventas/Semana (€)	385	400	395	365	475	440	490	420	560	525	480	510

- a) Visualizar los datos para disponer de una visión clara de su evolución y estimar posibles relaciones entre las variables implicadas.

- b) Calcular la ecuación de la recta de regresión para pronosticar las ventas semanales a partir de los gastos de publicidad.
- c) Analizar el modelo estimado con la función **summary** y obtener un posible intervalo de confianza para las conclusiones de los distintos parámetros.
- d) Evaluar una predicción para unos costes de publicidad de 35€ y encontrar un intervalo de confianza para la misma.
- e) Visualizar los intervalos de confianza para la respuesta media y las predicciones del modelo establecido en b)
- f) Realizar un análisis de varianza para estudiar la bondad del ajuste y la linealidad de la regresión. Explicar los resultados.

**Ejercicio 3 (Opcional).** El data set “*Seatbelts*” de la librería “*glmnet*” contiene datos de series temporales referidas a conductores fallecidos o con lesiones graves en UK entre los años 1966 y 1984. En enero de 1983 entró en vigor la ley que obliga a la utilización del cinturón de seguridad. Entre otras variables se dispone de las siguientes:

- *DriversKilled* : conductores de automóvil muertos.
- *front*: Pasajeros asientos delanteros muertos o gravemente heridos.
- *rear*: Pasajeros asientos delanteros muertos o gravemente heridos.
- *VanKilled*: número de conductores de furgonetas
- *law*: vigencia (0/1) de obligatoriedad del cinturón

Se pide:

- a) Analizar la serie temporal de fallecidos en accidentes, encontrar sus zonas de máximo valor y visualizar el efecto de entrada en vigor de la ley.
- b) Analizar las relaciones existentes entre los conductores fallecidos y las víctimas según estuvieran en los asientos delanteros o traseros. Explicar y estudiar en detalle el alcance de las suposiciones establecidas en los posibles modelos.
- c) Realizar un análisis de varianza para estudiar la bondad del ajuste y la linealidad de la regresión. Explicar los resultados
- d) Analizar y evaluar el efecto que tienen las furgonetas ligeras (tipo Van) en el conjunto de accidentes mortales antes y después de la aplicación de la ley. Justificar las respuestas