

## PROBABILIDADES Y ESTADÍSTICA (C)

### PRÁCTICA 10

1. Se tiene  $Y$  variable respuesta y  $x$  covariable y se obtiene una muestra  $(x_i; y_i)$ ,  $1 \leq i \leq n$ . Obtenga, para los siguientes modelos, los estimadores de cuadrados mínimos:
  - (a)  $Y = \beta x + \varepsilon$
  - (b)  $Y = \beta_0 + \beta_1 x + \varepsilon$
2. Se tiene en el archivo **autos.txt** una serie de modelos de autos de los que se conoce su precio y su calidad.
  - (a) Grafique un diagrama de dispersión. Tiene sentido asumir que hay una relación lineal entre precio y su calidad? En caso afirmativo, plantee un modelo lineal que describa esta relación, definiendo claramente las variables aleatorias involucradas en el modelo.
  - (b) Bajo el modelo de regresión lineal obtenga los estimadores de mínimos cuadrados. ¿Qué observa del coeficiente de determinación? Interprete.
  - (c) Estime la calidad de un auto de valor 50000.
  - (d) Dar un estimador de la varianza de los errores. ¿Cuál es su valor con los datos observados?
  - (e) Suponiendo normalidad de los errores, ¿es la pendiente significativa? Escribir las hipótesis, el estadístico del test y su distribución bajo  $H_0$ . ¿Cuál es la conclusión del test? Realice el cálculo del p-valor a mano y compárelo con lo obtenido con el R.
3.
  - (a) Con el comando **rnorm**, genere 100 observaciones provenientes de una distribución normal estándar. Guarde esto en la variable **x**.
  - (b) Con el mismo comando, genere 100 observaciones provenientes de una  $N(0, 0.25)$ , guarde esto en la variable **eps**.
  - (c) Usando **x** y **eps**, construya la variable dependiente **y** según el modelo  $Y = -1 + 0.5x + \varepsilon$ . ¿Quiénes serán los valores de  $\beta_0$  y  $\beta_1$  en este modelo?
  - (d) Efectúe un scatter plot entre las variables **x** e **y**.
  - (e) Efectúe un ajuste de cuadrados mínimos para obtener  $\hat{\beta}_0$  y  $\hat{\beta}_1$ , compárelos con los valores verdaderos  $\beta_0$  y  $\beta_1$ .
  - (f) Grafique en un mismo plot el scatter plot, la verdadera recta de **y** vs **x** y la recta obtenida por cuadrados mínimos, en distinto color.
  - (g) Repetiremos el procedimiento de antes, pero ahora habrá aproximadamente un 10% de los datos contaminados. Para eso, ahora el modelo que usaremos es el siguiente,  $Y = V(-1 + 0.5x + \varepsilon) + (1 - V)W$  donde  $\varepsilon \sim N(0, 0.25)$ ,  $V \sim Bi(1, 9/10)$  (es decir, vale 1 con probabilidad 9/10) y  $W \sim N(50, 1)$ .
4. Se tiene en el archivo **girasol.txt** el rendimiento de diversas parcelas de girasol (en toneladas) según la cantidad de dinero invertida en fertilizantes (en miles de pesos).

- (a) Bajo un modelo de regresión lineal obtenga los coeficientes con mínimos cuadrados.
  - (b) Grafique un diagrama de dispersión de los datos junto con la recta de regresión obtenida, ¿detecta algo sospechoso?
  - (c) ¿Encuentra evidencia en estos datos de que la inversión en fertilizante influye en el rendimiento de las parcelas de girasol?
  - (d) ¿Que proporción de la variabilidad en el rendimiento de las parcelas de girasol queda explicada por la inversión en fertilizante?
  - (e) Efectúe una “limpieza” de los datos y repita el procedimiento.  
*Sugerencia:* utilice la instrucción `identify`.
5. Al observar el gráfico de dispersión de las variables metros vs precio de los datos **inmuebles.txt** observamos que los puntos parecen ajustarse mejor a una parábola que a una recta. Usar el comando **lm** para realizar un ajuste lineal adecuado. ¿Cuáles son los coeficientes de la cuadrática que describe la relación?