MODELOS LINEALES
Primera Relación de Problemas

1. Dados los siguientes modelos decir cuáles de ellos son lineales:

   a) $y = \beta x + \epsilon$.
   b) $y = \beta_0 + \beta_1 \operatorname{sen}(x) + \epsilon$.
   c) $y = \beta_0^2 + \beta_1^3 x_1 + \beta_2 x_2 + \epsilon$.
   d) $y = ax^b + \epsilon$.

2. Sean tres variables X, Y, y Z, sobre las cuales se conoce que $X + Y = Z$. Tomada una muestra de tamaño N de las tres variables se construyen las rectas de regresión de X sobre Z y de Y sobre Z. Llamemos $\widehat{\beta}_0$ y $\widehat{\beta}_1$ a las estimaciones de la ordenada en el origen y de la pendiente, respectivamente, de la recta de regresión de X sobre Z; y llamemos $\widehat{\beta}'_0$ y $\widehat{\beta}'_1$ a las estimaciones de la ordenada en el origen y de la pendiente, respectivamente, de la recta de regresión de Y sobre Z.

   - Demostrar que $\widehat{\beta}'_0 = -\widehat{\beta}_0$ y que $\widehat{\beta}'_1 = 1 - \widehat{\beta}_1$.
   - ¿Qué relación existe entre los residuos de ambas regresiones?
   - Estima la varianza de las perturbaciones en ambos modelos. ¿Qué relación existe entre dichas estimaciones?

3. Antes de conocerse el método de mínimos cuadrados se empleaba el siguiente procedimiento para la estimación de la pendiente de la recta de regresión: Se consideraban las rectas que unían cada punto (x_i, y_i) de la nube con el centro de gravedad de la misma y se calculaban las pendientes de dichas rectas. Una vez hecho esto se tomaba como estimador de la pendiente la media de las pendientes de las rectas consideradas. ¿Es insesgado este estimador? Calcular su varianza. ¿Es eficiente?

4. Dado el modelo lineal de regresión sin término constante

$$y_i = \beta x_i + \epsilon_i \quad , \quad i = 1, \ldots, N$$

   bajo los supuestos:

   - $\mathrm{E}[\epsilon_i] = 0$, $\forall\, i = 1, \ldots, N$.
   - $\mathrm{Var}[\epsilon_i] = \sigma^2$, $\forall\, i = 1, \ldots, N$.
   - $\mathrm{Cov}[\epsilon_i, \epsilon_j] = 0$, $\forall\, i, j = 1, \ldots, N$, $i \neq j$.

   Resolver las siguientes cuestiones:

   a) ¿Qué hipótesis ha de imponerse sobre la variable X para que este modelo sea de rango completo? ¿En qué se traduce dicha hipótesis?
   b) Obtener el estimador mínimo cuadrático de β. Verificar si es o no un estimador insesgado y obtener su varianza.
   c) ¿Es ahora cierto que $\sum_{i=1}^{N}(y_i - \overline{y})^2 = \sum_{i=1}^{N}(\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^{N} e_i^2$? ¿Por qué?

d) Comprobar la siguiente relación

$$\text{VNE} = \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum_{i=1}^{N} x_i y_i\right)^2}{\sum_{i=1}^{N} x_i^2}$$

A partir de ella establecer un estimador insesgado para σ^2.

e) ¿Qué expresión adopta en este caso el coeficiente de determinación?

5. Consideremos el modelo de regresión simple univariante sobre el que se suponen las hipótesis usuales. Algunos programas de ordenador no permiten estimar el modelo sin término constante. En tales casos se sugiere trabajar con el doble de observaciones, introduciendo los valores (x_i, y_i) y $(-x_i, -y_i)$, $i = 1, \ldots, N$, y considerando el modelo

$$y_i^* = \beta_0 + \beta_1 x_i^* + \epsilon_i^*, \qquad i = 1, \ldots, 2N,$$

donde

$$x_i^* = \begin{cases} x_i & \text{si } i = 1, \ldots, N, \\ -x_{i-N} & \text{si } i = N+1, \ldots, 2N. \end{cases}$$

e

$$y_i^* = \begin{cases} y_i & \text{si } i = 1, \ldots, N, \\ -y_{i-N} & \text{si } i = N+1, \ldots, 2N. \end{cases}$$

a) Empleando esta estrategia, estima los parámetros β_0 y β_1. Compara dicha estimación con la obtenida en el modelo sin término constante.

b) Obtén la expresión de la variabilidad no explicada para este modelo y relaciónala con la del modelo sin término constante.

c) Obtén la estimación de la varianza de las perturbaciones y relaciónala con la obtenida en el modelo sin término constante.