Modelos no paramétricos y de regresión/Estadística II Semestre 2018-2

Tarea 7

Fecha de entrega: 20 de abril

En ningún ejercicio es necesario suponer normalidad en los errores

1. Suponer que se ajusta un modelo RLS por MCO a partir de las observaciones (x_i, y_i), $i = 1, \ldots, n$. Verificar que se cumplen las siguientes igualdades

   a) $\sum_{i=1}^n e_i = 0$,

   b) $\sum_{i=1}^n x_i e_i = 0$,

   c) $\sum_{i=1}^n \hat{y}_i e_i = 0$,

   d) $\sum_{i=1}^n \hat{y}_i = \sum_{i=1}^n y_i$,

   e) $\sum_{i=1}^n y_i e_i = \sum_{i=1}^n e_i^2$,

   donde e_i es el i-ésimo residuo, es decir, $e_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, $i = 1, \ldots, n$.

2. Mostrar que $\hat{\sigma}^2_{MCO} = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$, es insesgado para σ^2, donde $\hat{\beta}_0$ y $\hat{\beta}_1$ son los estimadores de MCO de β_0 y β_1.

   a) Escribir $\hat{Y}_j$ como combinación lineal de las Y_i, esto es, encontrar constantes $w_i(j)$ tales que $\hat{Y}_j = \sum_{i=1}^n w_i(j) Y_i$, para $j = 1, \ldots, n$.

   b) Justificar que $E[(Y_j - \hat{Y}_j)^2 \mid \mathbf{X} = \mathbf{x}] = V(Y_j - \hat{Y}_j \mid \mathbf{X} = \mathbf{x}) = \sigma^2 \sum_{i=1}^n w_i^2(j)$.

   c) Mostrar que $\sum_{j=1}^n \sum_{i=1}^n w_i^2(j) = n - 2$.

   d) Con los resultados anteriores justificar que $E[\hat{\sigma}^2_{MCO} \mid \mathbf{X} = x] = \sigma^2$.

3. El modelo RLS **sin intercepto** establece que

   $$E(Y \mid X = x) = \beta x,$$

   con los mismos supuestos que el modelo RLS con intercepto. Asumir una muestra $(x_1, Y_1), \ldots, (x_n, Y_n)$ del modelo anterior.

   a) Obtener el estimador de MCO de β.

   b) Calcular el valor esperado y la varianza del estimador del inciso anterior.

   c) Considerar $\tilde{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \tilde{\beta} x_i)^2$, donde $\tilde{\beta}$ es el estimador de β del inciso *a)*. Mostrar que $\tilde{\sigma}^2$ es insesgado para la varianza del modelo σ^2.

4. Con los resultados del ejercicio 2, ajustar un modelo RLS sin intercepto a los datos de estaturas de Galton (disponibles aquí).

   a) Reportar las estimaciones de β, σ^2 y la varianza del estimador de β, incluir los valores de las expresiones (sumas de cuadrados o productos cruzados) empleadas.

   b) Si utilizamos la suma de cuadrados de los residuos como criterio de comparación de modelos, ¿qué modelo *ajusta* mejor a los datos? ¿RLS con o sin intercepto?

5. Considerar el modelo RLS con errores normales $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $i = 1, \ldots, n$, y los estimadores de MV. ¿Qué pasa con $\hat{\beta}_0$ y $\hat{\beta}_1$ cuando se aplican las siguientes transformaciones a los datos?

   a) $Y^* = Y + c$, $c \in \mathbb{R}$.

   b) $Y^* = cY$, $c \in \mathbb{R} \setminus \{0\}$.

   c) $Y^* = cY + d$, $c \in \mathbb{R} \setminus \{0\}$ y $d \in \mathbb{R}$.

   d) $X^* = X + c$, $c \in \mathbb{R}$

   e) $X^* = cX$, $c \in \mathbb{R} \setminus \{0\}$.

   f) $X^* = cX + d$, $c \in \mathbb{R} \setminus \{0\}$ y $d \in \mathbb{R}$.