

Modelos no paramétricos y de regresión/Estadística II Semestre 2018-1

Tarea 1

Fecha de entrega: 25 de agosto

En ningún ejercicio es necesario suponer normalidad en los errores

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

$$\begin{aligned} a) \sum_{i=1}^n e_i &= 0, & d) \sum_{i=1}^n \hat{y}_i &= \sum_{i=1}^n y_i, \\ b) \sum_{i=1}^n x_i e_i &= 0, & e) \sum_{i=1}^n y_i e_i &= \sum_{i=1}^n e_i^2, \\ c) \sum_{i=1}^n \hat{y}_i e_i &= 0, \end{aligned}$$

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

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

- 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 $j = 1, \dots, n$.
- Justificar que $E[(Y_j - \hat{Y}_j)^2 | \mathbf{X}] = V(Y_j - \hat{Y}_j | \mathbf{X}) = \sigma^2 \sum_{i=1}^n w_i^2(j)$.
- Mostrar que $\sum_{j=1}^n \sum_{i=1}^n w_i^2(j) = n - 2$.
- Con los resultados anteriores justificar que $E[\hat{\sigma}_{MCO}^2 | \mathbf{X}] = \sigma^2$.

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

$$E(Y | X) = \beta X,$$

con los mismos supuestos que el modelo RLS con intercepto. Asumir una muestra de variables independientes $(X_1, Y_1), \dots, (X_n, Y_n)$ del modelo anterior.

- Obtener el estimador de MCO de β .
 - Calcular el valor esperado y la varianza del estimador del inciso anterior.
 - 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](#)).
- 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.
 - 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?