

**Tarea 1. Introducción a Analítica. Departamento de Estadística.
Facultad de Ciencias. Universidad Nacional de Colombia, Sede
Medellín**

Modo y fecha de entrega: La entrega de las tareas es por medio electrónico, vía email y se deben enviar a jcsalaza@unal.edu.co el 30 de abril de 2025 hasta las 6:00pm (18:00 horas); se usará el reloj de mi PC para llevar el registro de entrega. Tareas que se entreguen entre las 6:01 y 6:30pm se calificarán sobre 4:00, y tareas que se entreguen después de las 6:30pm se calificarán como reprobadas. Se pueden conformar grupos de máximo 4 personas. **OJO:** 1) Se debe entregar SOLAMENTE un (1) archivo en formato pdf, múltiples archivos NO se recibirán ni calificarán, el código en R se anexa en el mismo archivo pdf, 2) Cada archivo debe estar marcado siguiendo este formato: Tarea1IAA_012025_NOMBRE DE AL MENOS UN INTEGRANTE DEL GRUPO.pdf. El archivo pdf debe estar marcado de esta forma, de lo contrario se devolverá.

1. (20 pts.) Demuestre la siguiente ecuación (conocida como Bias-Variance Tradeoff):

$$E \left(y_0 - \hat{f}(x_0) \right)^2 = Var \left(\hat{f}(x_0) \right) + \left[Bias \left(\hat{f}(x_0) \right) \right]^2 + Var(\varepsilon)$$

Aquí,

$$Bias(\hat{f}(x)) = E \left(\hat{f}(x) \right) - f(x)$$

y

$$Var \left(\hat{f}(x) \right) = E \left(\hat{f}(x) - E \left(\hat{f}(x) \right) \right)^2$$

¿Porqué entonces el Test-MSE, no puede ser nunca menor que $Var(\varepsilon)$?

2. (50 pts.) Este es un ejercicio de simulación similar a los vistos en clase.
 - a) Simule 100 datos de la siguiente función. Para evitar duplicidad con otras tareas, use `set.seed(NÚMERO DE SU DOCUMENTO DE IDENTIDAD)`, este número de documento se debe especificar y decir de quien de los integrantes es. Anexe el código usado. Con dicho número se debe replicar exactamente el ejercicio.

$$f(x) = \frac{x^3}{5} + \frac{4x^2}{5} - \frac{7x}{5} - 2$$

- b)* Ajuste una linea recta OLS, un loess con poco suavizamiento y un loess con poco suavizamiento. Grafique la curva original $f(x)$ y estas tres funciones, Deben estar superpuestas a los datos generados.
 - c)* Haga un gráfico del training-MSE, del test.MSE y de la $\text{var}(e)$, similar a los vistos en clase.
 - d)* Cuál modelo se debe seleccionar y porqué.
- 3. (30 pts.) Usando la misma función del numeral anterior, verifique, vía simulación la ecuación del Bias.Variance tradeoff. Puede tratar de adaptar el programa ilustrado en clase. Anexe el código usado.