

Estadística Aplicada II

Tarea 1. (base de datos Yarn)

- a. Describa las variables e individuos del subgrupo asignado en la base *Yarn*. Realice el gráfico de dispersión entre las variables $Y = \text{densidad}$ y la variable $x = \text{NIRj}$ que presentan la mayor correlación.
- b. Ajuste un ‘buen’ modelo de regresión lineal simple tomando *densidad* como variable de respuesta y *NIRj* como variable predictora. Juzgue la bondad del modelo interpretando:
 - b1. El R^2 , los coeficientes de regresión $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}$
 - b2. Las pruebas t , F y el ANOVA asociado (SCReg?)
 - b3. Las gráficas y las pruebas formales en la validación de supuestos
 - b4. Los residuales estandarizados, studentizados, distancia de cook, ..
 - b5. Intervalos de confianza al 95% para β_1 y para $E(Y|x)$ tomando un valor específico x .
- c. Realice el modelo simple como en b., pero con ambas variables estandarizadas y compare los resultados con los escenarios b1, b2, b3, b4 y b5; explique.
- d. ¿Un modelo polinómico de grado 2 mejora las estimaciones?

Nota: Apóyese con *Rmarkdown* en el informe, y registre los principales resultados (originales) derivados del uso del software libre R. Suba el informe en archivo *pdf* únicamente vía campus virtual antes de abril 13 de 2023.

Víctor M González R.

Profesor Titular