

Estadística en Ciencia de Datos

Ejercicio: Regresión Lineal Múltiple y Selección de Variables

Base de datos: *Student Performance* (UCI)

Contexto

Se dispone de información académica, familiar y socioeconómica de estudiantes de educación secundaria. El objetivo es analizar los factores que influyen en el desempeño académico final, medido por la `nota_final`.

Modelo de Regresión Lineal Múltiple

Considere el siguiente modelo de regresión lineal múltiple:

$$\begin{aligned} \text{nota_final}_i = & \beta_0 + \beta_1 \text{edad}_i + \beta_2 \text{tiempo_estudio}_i + \beta_3 \text{inasistencias}_i + \beta_4 \text{reprobaciones_previas}_i \\ & + \beta_5 \text{educacion_madre}_i + \beta_6 \text{educacion_padre}_i + \beta_7 \text{internet_casa}_i \\ & + \beta_8 \text{apoyo_familiar}_i + \varepsilon_i, \\ \varepsilon_i \sim & \mathcal{N}(0, \sigma^2). \end{aligned} \tag{1}$$

Preguntas

- a) Ajuste el modelo de regresión lineal múltiple utilizando mínimos cuadrados ordinarios. Presente la tabla de estimación de los parámetros e identifique las variables estadísticamente significativas al nivel $\alpha = 0,05$.
- b) Interprete los coeficientes asociados a las variables `tiempo_estudio`, `inasistencias` y `reprobaciones_previas`, manteniendo constantes las demás covariables.
- c) Reporte el coeficiente de determinación R^2 y el R^2 ajustado. Explique por qué el R^2 ajustado es más apropiado en este contexto.
- d) Evalúe los principales supuestos del modelo lineal (linealidad, normalidad de los residuos y homocedasticidad), incluyendo gráficos de diagnóstico y comentarios.
- e) Analice la posible presencia de multicolinealidad utilizando el factor de inflación de la varianza (VIF).

Selección de variables: método *Backward*

Aplique el método de eliminación hacia atrás (*backward elimination*) utilizando el criterio de información de Akaike (AIC), partiendo del modelo completo.

- a) Indique qué variables son eliminadas en el proceso y presente el modelo final seleccionado.

- b) Compare el valor del AIC del modelo completo con el del modelo reducido y comente si la reducción del modelo implica una mejora en términos de parsimonia.
- c) Analice el coeficiente de determinación ajustado del modelo final y discuta los cambios respecto al modelo inicial.
- d) Evalúe los supuestos del modelo seleccionado e interprete los resultados.

Indicaciones

- Trate las variables binarias como factores.
- Justifique cualquier decisión de exclusión de variables.
- Presente conclusiones claras y bien fundamentadas.