

# Taller de Repaso - Regresión Lineal Múltiple \*

**Estadística II**    *Universidad Nacional de Colombia, Sede Medellín*

---

Este documento corresponde a un taller de repaso del curso de **Estadística II** para la *Universidad Nacional de Colombia, Sede Medellín*, en el periodo 2025 - 1. Se profundizan aspectos claves del modelo de regresión lineal múltiple, incluyendo su verificación práctica y comprensión teórica. **Monitor:** *Santiago Carmona Hincapié*.

*Keywords:* regresión múltiple, supuestos, inferencia, diagnóstico

---

## Información general

Con el propósito de profundizar en los conceptos del modelo de regresión lineal múltiple vistos en clase, se propone afrontar este taller en dos partes: una de teoría básica y otra práctica.

**La solución para cada uno de los problemas se efectúa a partir del software estadístico R.**

---

### *Parte teórica*

**Instrucciones:** Para cada afirmación, indique si es verdadera o falsa seleccionando una de las cuatro opciones dadas.

1. Bajo los supuestos del modelo, el estimador  $\hat{\beta}$  por MCO coincide con el estimador de máxima verosimilitud.
  - (A) Verdadera, porque ambos procedimientos minimizan la suma de cuadrados.
  - (B) Verdadera, si y solo si se asume normalidad de los errores.
  - (C) Falsa, ya que los métodos son conceptualmente distintos.
  - (D) Falsa, porque MCO requiere menos supuestos que máxima verosimilitud.
2. Rechazar  $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$  implica que el modelo es significativo.
  - (A) Verdadera, y puede comprobarse con el estadístico F.
  - (B) Verdadera, solo si el número de predictores es mayor a 2.
  - (C) Falsa, ya que el test F solo prueba significancia individual.

---

\*El material asociado a este taller puede encontrarse en el repositorio del curso, (<https://github.com/Itssach/Estadistica-II>)

- (D) Falsa, debe emplearse el estadístico t para esta prueba.

3. La matriz  $H$  y  $I - H$  son simétricas e idempotentes.

- (A) Verdadera, dado que  $H$  es una proyección ortogonal.
- (B) Falsa, solo  $H$  es idempotente.
- (C) Verdadera, siempre y cuando  $X$  tenga rango completo.
- (D) Falsa, ninguna tiene esas propiedades.

4. El estadístico F para evaluar la regresión global es:

$$F = \frac{SSR/k}{SSE/(n-k)} \sim F_{k,n-k}$$

- (A) Verdadera, representa la relación entre variabilidad explicada y residual.
- (B) Falsa, el denominador debe ser  $n - p$ .
- (C) Falsa, porque el numerador es incorrecto.
- (D) Verdadera, pero solo con dos regresores.

5. Una observación es influyente si cumple:  $D_i > 1$ ,  $|DFBETAS_j(i)| > 2/\sqrt{n}$ ,  $|DFFITs_i| > 2\sqrt{p/n}$ .

- (A) Verdadera, esos umbrales indican influencia significativa.
- (B) Falsa, basta con que uno se cumpla.
- (C) Verdadera solo si  $n > 50$ .
- (D) Falsa, esos umbrales no son estándares.

---

### Parte práctica

El presente taller se basa en un contexto aplicado al análisis del rendimiento académico universitario. Se pretende modelar el **promedio acumulado** de los estudiantes (escala 0 a 5.0) en función de tres variables explicativas:

- **Horas\_Estudio**: número de horas promedio dedicadas al estudio semanal.
- **Estres**: nivel de estrés percibido, en una escala de 0 a 10.
- **Sueno**: número promedio de horas de sueño por noche.

Table 1: Información en análisis

Promedio	Horas_Estudio	Estres	Sueno
4.187733	22.19762	3.579187	9.138572
3.811590	23.84911	5.513767	8.074896
4.464927	32.79354	4.506616	6.181826
3.961398	25.35254	4.304915	7.151833
4.357778	25.64644	3.096763	6.002792
4.929359	33.57532	4.909945	5.928504

Utilice el conjunto anterior con los siguientes objetivos:

1. Verifique los supuestos del modelo de regresión lineal múltiple ajustado:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

- Verifique: normalidad (prueba gráfica y prueba Shapiro-Wilk), homocedasticidad, independencia.
2. Determine la significancia global del modelo ajustado usando el test F.
  3. Estime los parámetros individuales, interprete sus  $p$ -valores e incluya intervalos de confianza al 95
  4. Identifique puntos atípicos, de balanceo e influencias mediante: -  $|r_i| > 2$  -  $h_{ii} > 2p/n$   
-  $D_i$ ,  $DFBETAS_j$ ,  $DFFITS_i$
  5. Construya el intervalo de confianza para la respuesta media y el intervalo de predicción para el punto:

$$\mathbf{x}_0 = [1, 55, 32, 19]'$$

Verifique previamente que no es extrapolación.