

# Regresión Lineal Múltiple

## Ejercicio 1: Conceptual (Interpretación *Ceteris Paribus*)

Un analista ajusta dos modelos para predecir el consumo de un coche (`mpg`):

1. `lm(mpg ~ wt)` obtiene un coeficiente para `wt` de -5.3.
2. `lm(mpg ~ wt + hp)` obtiene un coeficiente para `wt` de -3.8.

Explica detalladamente por qué el coeficiente para la variable `wt` (peso) cambia al añadir la variable `hp` (caballos de fuerza). ¿Cuál de los dos coeficientes representa el efecto “puro” o “aislado” del peso? Fundamenta tu respuesta en el principio de *ceteris paribus*.

## Ejercicio 2: Práctico (Ajuste e Interpretación de un Modelo Múltiple)

Usa el conjunto de datos `iris` de R. Queremos modelar la anchura del pétalo (`Petal.Width`) en función de la longitud del pétalo (`Petal.Length`) y la anchura del sépalo (`Sepal.Width`).

- a) Ajusta un modelo de regresión lineal múltiple: `lm(Petal.Width ~ Petal.Length + Sepal.Width, data = iris)`.
- b) Interpreta el coeficiente estimado para `Petal.Length`.
- c) Interpreta el coeficiente estimado para `Sepal.Width`.
- d) Interpreta el intercepto del modelo. ¿Tiene un significado práctico en este contexto biológico?

## Ejercicio 3: Conceptual ( $R^2$ vs. $R^2$ Ajustado)

Cuando pasamos de un modelo simple a uno múltiple, introducimos el  **$R^2$  ajustado** como medida de bondad de ajuste.

- a) ¿Cuál es el principal problema de usar el  $R^2$  tradicional para comparar modelos con diferente número de predictores?
- b) ¿Cómo soluciona el  $R^2$  ajustado este problema? Explica qué “penalización” introduce en su fórmula.

## Ejercicio 4: Interpretación de Salidas de R

Te presentan el siguiente resumen de un modelo que predice el prestigio de una ocupación (`prestige`) en función de los ingresos (`income`) y el nivel educativo (`education`).

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-6.0647	4.2750	-1.419	0.1595
income	0.0013	0.0003	4.524	1.9e-05 ***
education	4.1832	0.3887	10.762	< 2e-16 ***

Multiple R-squared: 0.79, Adjusted R-squared: 0.785  
F-statistic: 185.6 on 2 and 99 DF, p-value: < 2.2e-16

- ¿Es el modelo globalmente significativo? ¿En qué te basas?
- ¿Son los predictores `income` y `education` individualmente significativos, después de controlar por el efecto del otro? Justifica tu respuesta.
- Explica la diferencia conceptual entre lo que evalúa el **test F global** y lo que evalúan los **tests t individuales** en este modelo.

## Ejercicio 5: Conceptual (Multicolinealidad)

Describe con tus propias palabras qué es la **multicolinealidad**. Menciona tres consecuencias negativas que puede tener la multicolinealidad severa en un modelo de regresión y si afecta más a la **predicción** o a la **inferencia**.

## Ejercicio 6: Práctico (Diagnóstico de Multicolinealidad)

Usa el dataset `mtcars`. Ajusta un modelo para predecir el consumo (`mpg`) usando como predictores el número de cilindros (`cyl`), la cilindrada (`disp`), los caballos de fuerza (`hp`) y el peso (`wt`).

- Observa el `summary()` del modelo. ¿Hay alguna variable que, a pesar de tener una alta correlación simple con `mpg`, no resulte significativa en el modelo múltiple?
- Carga la librería `car` y calcula el **Factor de Inflación de la Varianza (VIF)** para cada predictor.
- Basándote en los valores del VIF, ¿qué variables presentan un problema de multicolinealidad? ¿Cuál es tu recomendación para simplificar el modelo?

### **Ejercicio 7: Teórico (Notación Matricial)**

- a) Escribe la fórmula del estimador de Mínimos Cuadrados Ordinarios ( $\hat{\gamma}$ ) en notación matricial.
- b) ¿Qué supuesto fundamental del modelo de regresión múltiple garantiza que la matriz  $(\mathbf{X}^T \mathbf{X})$  sea invertible?

### **Ejercicio 8: Práctico (Gráficos de Regresión Parcial)**

Usa el dataset `Prestige` de la librería `car`.

- a) Ajusta el modelo `lm(prestige ~ income + education + women, data = Prestige)`.
- b) Genera los gráficos de regresión parcial (o “added-variable plots”) para este modelo usando la función `avPlots(tu_modelo)`.
- c) Explica qué representa el gráfico para la variable `education`. ¿Qué significan los ejes X e Y de ese gráfico específico? ¿A qué corresponde la pendiente de la línea en ese gráfico?

### **Ejercicio 9: Inferencia (F-test vs. t-tests)**

Describe un escenario hipotético en el que el **test F global** de un modelo de regresión múltiple sea altamente significativo ( $p < 0.001$ ), pero **ninguno de los tests t individuales** para los coeficientes sea significativo. ¿Cuál es la causa estadística más probable de este fenómeno?

### **Ejercicio 10: Práctico (Comparación de Modelos Anidados)**

Usa el dataset `swiss`.

- a) Ajusta un **modelo reducido** para predecir `Fertility` usando solo `Agriculture` y `Education`.
- b) Ajusta un **modelo completo** que, además de las variables anteriores, incluya `Catholic` y `Infant.Mortality`.
- c) Utiliza la función `anova()` para comparar formalmente los dos modelos. ¿Aportan las variables `Catholic` y `Infant.Mortality` una mejora estadísticamente significativa al modelo? Interpreta el p-valor del test F resultante.