

Taller Práctico Regresión Lineal Múltiple (5) *

Estadística II *Universidad Nacional de Colombia, Sede Medellín*

Este documento corresponde al octavo taller práctico del curso de **Estadística II** para la *Universidad Nacional de Colombia*, Sede Medellín, en el periodo 2024 - 2. Se brinda una introducción al análisis de regresión. El enfoque de este taller está la comprensión de los métodos de regresión secuenciales y modelos de regresión con interacciones. **Monitor:** *Santiago Carmona Hincapié*.

Keywords: regresión múltiple, secuencialidad, interacción

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

De respuesta a las preguntas formuladas a continuación en base a la teoría tratada en clase. **Provea una interpretación de ser necesario.**

1. Determine el valor de verdad de las siguientes afirmaciones.

- (a) El método '*Forward*' parte del modelo con todas las variables, eliminando secuencialmente de a una variable con el propósito de reducir la suma de cuadrados del error asociado. El método '*Backward*' sigue la lógica contraria.
- (b) En el modelo de regresión lineal simple de la clase $Y_i = \beta_0 + \beta_1 I_{i1} + \beta_2 I_{i2} + \dots + \beta_{i,c-1} I_{i,c-1} + \varepsilon_i; \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ las rectas generadas son horizontales.
- (c) En el modelo de regresión lineal simple de la clase $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \dots + \beta_{ic} I_{i,c-1} + \varepsilon_i; \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ las rectas generadas cambian su pendiente.
- (d) En el modelo de regresión lineal con interacción $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \dots + \beta_{ic} I_{i,c-1} + \beta_{11} X_{i1} I_{i1} + \beta_{12} X_{i1} I_{i2} + \dots + \beta_{1,c-1} X_{i1} I_{i,c-1} + \varepsilon_i; \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ las rectas generadas cambian su intercepto y pendiente.

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

Ejercicio con datos reales

Considere el siguiente conjunto de datos que agrupa una serie de métricas enfocadas en evaluar el rendimiento en educación física de estudiantes en una institución. **Se incluyen únicamente las métricas cuantitativas**, cuya descripción puede encontrarse **en el siguiente enlace**: <https://www.kaggle.com/datasets/ziya07/student-physical-education-performance>

Table 1: Información en análisis

Performance	Strength	Skills	Speed
337.4335	60.22040	37.55007	54.67685
418.0430	45.23407	72.26675	53.60737
413.5214	32.09499	86.78852	44.96376
378.6311	63.44574	44.25407	53.85836
428.8907	49.40831	93.26630	43.28241

Considere a ‘Overall Performance’ como la variable respuesta. *Las covariables en análisis se especifican en la tabla mostrada con anterioridad.* **Suponga que los supuestos del modelo se cumplen. De respuesta a los siguientes planteamientos:**

1. Escriba el modelo de regresión lineal múltiple, junto con sus supuestos. Use el método de selección ‘Forward’ para determinar el mejor modelo (paso por paso) y compárelo con el resultado dado por el método ‘Backward’ y ‘Stepwise’.
2. Considere el modelo de regresión lineal dado por la interacción $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 I_{i1} + \beta_3 I_{i2} + \beta_{11} X_{i1} I_{i1} + \beta_{12} X_{i1} I_{i2} + \varepsilon_i; \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$, empleado para evaluar el cambio de precio en automóviles en función de ‘Bore’ (numérica) y ‘BodyStyle’ (categórica), esta última con tres categorías ‘Sedan’, ‘Otros’ y ‘Hatchback’. **Se toma como referencia la categoría ‘Hatchback’.** **Responda:**
 - (a) Caracterize la información en análisis, reportando la ecuación de las rectas en función de los diferentes niveles, así como las rectas ajustadas.
 - (b) Interprete la estimación del parámetro β_{11} en términos del problema.
 - (c) Determine si existe diferencia entre las ordenadas en el origen de las rectas correspondientes a los carros con ‘BodyStyle’ de la clase ‘Hatchback’ y ‘Sedan’. Escriba la hipótesis correspondiente y el criterio de decisión.
 - (d) Determine si el cambio promedio ‘Price’ por unidad de cambio en ‘Bore’ es igual para los niveles de ‘BodyStyle’ de la clase ‘Sedan’ y ‘Otros’. Escriba la hipótesis correspondiente y el criterio de decisión.