Taller Práctico Regresión Lineal Múltiple (1) *

Estadística II *Universidad Nacional de Colombia, Sede Medellín*

Este documento corresponde al primer taller práctico del curso de **Estadística II** para la *Universidad Nacional de Colombia*, Sede Medellín, en el periodo 2025 - 1. Se brinda una introducción al análisis de regresión. El enfoque de este taller está la comprensión del modelo de regresión lineal múltiple a nivel matricial. **Monitor:** *Santiago Carmona Hincapié.*

Keywords: regresión múltiple

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. Considere el siguiente modelo de regresión lineal múltiple con k variables regresoras, $p = (k+1)$ parámetros asociados $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i, \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$.

   - (a) Escriba el modelo de forma matricial junto con sus supuestos. **Especifique las dimensiones de cada componente**.

   - (b) Demuestre que el estimador $\hat{\boldsymbol{\beta}}$ que se obtiene a través del método de mínimos cuadrados es un estimador insesgado para $\boldsymbol{\beta}$. Analice $\hat{\boldsymbol{\beta}}$.

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

   - (a) Los modelos de regresión múltiple (1).$\mathrm{Y}_i = \beta_0 + \sum_{j=1}^{k} \beta_j (X_{ij})^{1/2} + \varepsilon_i$; (2).$\ln(\mathrm{Y}_i) = \beta_0 + \sum_{j=1}^{k} \beta_j \ln(X_{ij}) + \varepsilon_i; \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ no son modelos lineales, puesto que se altera la forma de la superficie de respuesta.

   - (b) Bajo los supuestos del modelo de regresión lineal múltiple, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, el estimador para los parámetros $\hat{\boldsymbol{\beta}}$ obtenido a través del método de mínimos cuadrados ordinarios es el mismo que el de máxima verosimilitud, así como para la varianza $\hat{\sigma}^2$. **Explique**.

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

- (c) Se requiere que la matriz $(\mathbf{X^tX})$ sea singular. Además, la matriz $\mathbf{H}$ es simétrica e idempotente, al igual que $(\mathbf{I_n} - \mathbf{H})$.

- (d) La matriz de varianzas- covarianzas siempre es simétrica respecto a su diagonal principal, además, siempre tiene unos en su diagonal principal.

- (e) El parámetro β_0, como la respuesta media de Y es únicamente interpretable si el conjunto de observaciones $(X_1, X_2, ..., X_k) = (0, 0, ..., 0)$ está incluido en el rango experimental.

- (f) Los parámetros $\beta_j, j = 1, ..., k$ indican el cambio en la respuesta media de Y por unidad de incremento en la respectiva variable X_j.

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
71.98394	46.64215	62.54602	51.64761
69.90818	66.69022	67.26147	50.50888
64.82243	47.40409	55.15063	51.73342
72.17802	34.96857	69.47489	52.43953
74.75588	47.54257	57.83499	47.76798

Considere a *'Overall Performance'* como la variable respuesta. *Las covariables en análisis se especifican en la tabla mostrada con anterioridad.* **De respuesta a los siguientes planteamientos:**

1. Determine cuál es la matriz de diseño $(\mathbf{X})$ para este problema en específico. Calcule el vector de parámetros estimados $\hat{\boldsymbol{\beta}}$. Brinde una interpretación adecuada.

2. Calcule la estimación de la respuesta media $\hat{\boldsymbol{y}}$.

3. Calcule el vector de los errores estimados $\hat{\boldsymbol{\varepsilon}}$.

4. Construya el valor de la estimación para la varianza $\hat{\sigma}^2$.