

# Taller Práctico Regresión Logística (1) \*

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

---

Este documento corresponde al sexto 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 logístico como método de clasificación, empleando variables continuas y categóricas. **Monitor:** *Santiago Carmona Hincapié*.

*Keywords:* regresión múltiple, secuencialidad, interacción

---

## Información general

Con el propósito de profundizar en los conceptos del modelo de regresión logística 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) En un modelo de regresión logística, la variable respuesta es de naturaleza binaria, tal que  $Y_i \sim \text{Ber}(\theta_i)$ ; donde  $\theta$  es la probabilidad de éxito en el  $i$ -ésimo ensayo.
- (b) Los coeficientes  $\underline{\beta}$  de un modelo de regresión logística se estiman numéricamente el método de mínimos cuadrados ordinarios, ya que la función de verosimilitud es no- lineal.
- (c) Al incluir variables categóricas al modelo, es necesario definir una categoría de referencia, y la interpretación de tales coeficientes se realiza en comparación con tal referencia.
- (d) El valor  $\exp(\beta_j)$  se interpreta como el cambio en el logaritmo de los 'odds' por cada incremento unitario en la variable predictora  $X_j$ .
- (e) El estadístico de prueba  $\chi_c = D_0^2 - D^2$ , con una región de rechazo asociada  $R_c = \{\chi_c > \chi_{p,\alpha}\}$  se emplea para probar la significancia global del modelo, mientras que se prefiere la prueba *t- student* para la significancia individual.

---

\*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 características enfocadas en *clasificar* la personalidad de múltiples individuos. **Se incluyen variables cuantitativas y cualitativas.** La información puede profundizarse en: <https://www.kaggle.com/datasets/rakeshkapilavai/extrovert-vs-introvert-behavior-data>

Table 1: Información en análisis

Personality	Time_spent_Alone	Stage_fear	Social_event_attendance	Going_outside
<b>Extrovert</b>	4	No	4	6
<b>Introvert</b>	9	Yes	0	0
<b>Introvert</b>	9	Yes	1	2
<b>Extrovert</b>	0	No	6	7
<b>Extrovert</b>	3	No	9	4

Considere a ‘*Personality*’ como la variable respuesta. *Algunas de las covariables en análisis se especifican en la tabla mostrada con anterioridad.* **De respuesta a los siguientes planteamientos:**

1. Ajuste un modelo de regresión logístico. Analice los componentes  $\theta_i, g(\theta_i), \psi(\mathbf{x}_i)$ . ¿Por qué se emplea la función ‘*logit*’ como función de enlace?
2. Realice una prueba de significancia general del modelo de regresión logístico propuesto. Interprete los resultados obtenidos.
3. Realice una prueba de significancia de los parámetros individuales. Provea una *interpretación directa* de los parámetros del modelo de regresión logística, así como una *interpretación en términos de la razón de probabilidades*.
4. Realice una predicción de futuras observaciones. Utilice las probabilidades obtenidas para la construcción un clasificador logístico con etiquetas.