

## Taller de modelos no lineales

Giovany Babativa-Márquez, PhD

### Modelos supervisados: Regresión logística

El Ministerio de Educación Nacional ha venido desarrollando la Encuesta Nacional de Deserción (ENDE), en donde uno de los propósitos es profundizar en las condiciones de los desertores y sus hogares, esta encuesta es ejecutada en el año 2011 por el DANE y fue agregada como un módulo adicional en la GEIH para identificar niños que estuvieran fuera del sistema escolar (desertores). Asimismo, se seleccionó a un niño del mismo municipio, estrato y clase pero que asistiera a la escuela o colegio para que fuera parte del grupo contrafactual (grupo de control).

Uno de los aspectos de mayor relevancia consiste en identificar los factores de riesgo y de protección asociados con la deserción escolar; para ello se aplicó un instrumento que permitiera identificar los factores determinantes en la deserción desde la perspectiva del niño, el hogar y su entorno. El conjunto de datos que se le ha suministrado es una pequeña muestra de la ENDE con 5000 encuestas y que puede encontrarla en el repositorio de GitHub.

Para cargar el conjunto de datos siga estas instrucciones:

```
options(scipen = 999)
library(pacman)

p_load(tidyverse, broom, modelr, haven, labelled,
       skimr, corrplot, psych, gt, gtsummary, pROC)

url <- "https://github.com/jgbabativam/AnaDatos/raw/main/datos/Deser_DANE.sav"

datos <- read_sav(url)
```

### Ejercicios

1. Realice un análisis descriptivo y exploratorio que le permita comprender los datos.
  - a) ¿Cuántos registros por grupo (control/desertores) se tienen?

- b) Dentro de cada grupo, ¿cuántos son hombres y cuántas son mujeres?, ¿cree que puede existir alguna asociación entre esas variables? Use una prueba estadística para argumentar su conclusión.
- c) Considere que, si el niño debe trabajar, afecta el nivel de deserción. Explique y use una prueba estadística para soportar su conclusión.
- d) ¿El modelo educativo está relacionado con la deserción escolar?
- e) ¿La edad está relacionada con la deserción escolar?
- f) ¿La tenencia de libros está relacionada con la deserción escolar?
- g) ¿El nivel educativo de la madre está relacionada con la deserción escolar?
- h) Haga un boxplot del índice de calidad de vida (ICV) por grupo (control/desertor), ¿qué puede decir al respecto?

2. Ajuste el siguiente modelo para el riesgo de deserción:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 ICV$$

En donde  $\pi$  es la probabilidad de desertar,  $ICV$  es el índice de calidad de vida.

- 3. Presente los resultados del modelo anterior y explique el OR (coeficiente exponencial  $e^{\beta_1}$ ). ¿Cómo interpretaría el intervalo de confianza del OR?
- 4. Ahora ajuste un modelo en donde las variables independientes sean: ICV, Nivel educativo de la madre, trabajo infantil y cantidad de libros.
- 5. Interprete los coeficientes del modelo ajustado previamente.
- 6. ¿Que estrategias o políticas propondría con base en los resultados obtenidos?
- 7. Realice una curva ROC que compare los dos modelos. Concluya.