TP 1: Regresión ordinal

Nicolás Celie, Martín Peralta, Nicolás Ian Rozenberg

2025-05-25

Ejercicio 3

Una forma de plantear el problema es mediante una regresión lineal donde la variable dependiente sea la respuesta en la escala Likert, y las covariables la edad y variables indicadoras del género. Es decir,

$$Likert_i = \beta_{\text{masc}}\text{masc}_i + \beta_{\text{fem}}\text{fem}_i + \beta_{\text{edad}}\text{edad}_i + \epsilon_i$$

El problema de este enfoque es que $Liker_i$ toma $1, ..., 5$ como posibles valores. Un problema es que el modelo lineal común asume normalidad en los errores, y por lo tanto en la variable dependiente. En este caso, es una variable discreta con un rango de sólo 5 valores. Por otra parte, se asume que las distancias entre las respuestas son iguales. Por ejemplo, podría no tener sentido considerar que la diferencia entre "Totalmente en desacuerdo" y "En desacuerdo" sea la misma que "En desacuerdo" y "Neutro".

Otra forma es modelarlo mediante una regresión multinomial. En esta, se modela a la probabilidad de que el individuo i responda la opción j como

$$\mathbb{P}(Likert_{i,j} = 1) = \text{Softmax}(z_i)_j$$

donde z_i es un vector en $\mathbb{R}^5$ tal que

$$z_{i,j} = \beta_{\text{masc},j}\text{masc}_i + \beta_{\text{fem},j}\text{fem}_i + \beta_{\text{edad},j}\text{edad}_i$$

y $\text{Softmax} : \mathbb{R}^p \to \mathbb{R}^p$ es la función

$$\text{Softmax}(z)_j = \frac{e^{z_j}}{\sum_{j=1}^{p} e^{z_j}}$$

Sin embargo, dicho enfoque no es el más apropiado tampoco. Esto se debe a que la variable dependiente no es una variable categórica nominal. Las posibles respuestas tienen un orden, que no se está modelando.

```
print("Hello World")
```

```
## [1] "Hello World"
```