

# Clase 1: Modelos Lineales

Justo Andrés Manrique Urbina

27 de agosto de 2019

Existe un modelo  $Y_1 = B_0 + B_1X + \varepsilon$ . Asumimos que el error tiene distribución normal  $N \sim (1, \sigma^2)$  e independiente.  $X$ , por el momento, es conocido fijo y observable. Por otro lado  $Y_i$  es una variable aleatoria. Bajo estas condiciones,  $Y_i$  tiene distribución normal  $Y_i \sim N(B_0 + B_i * x_i, \sigma^2)$ .

Desde el punto de vista de los modelos lineales generalizados,  $Y_i$  tiene distribución normal  $Y_i \sim N(\mu_i, \sigma^2)$  dónde  $\mu_i = N_i$  y  $N_i = B_0 + B_i * x_i$ . Lo anteriormente expuesto se conoce como los tres componentes del modelo lineal generalizado:

- **Parte aleatoria:** Es la distribución de la variable respuesta, en este caso  $Y_i \sim N(\mu_i, \sigma^2)$ .
- **Predictor lineal:** El predictor lineal  $N_i = B_0 + B_i * x_i$
- **Enlace:** La función de enlace es  $\mu_i = N_i$

Una de las ventajas del modelo lineal es bien flexible, lo cual permite modelar un montón de distribuciones que pueden ser complejas. Por ejemplo, imaginemos que se desea modelar el salario con edad. Esta relación sigue una curva no lineal cóncava. Utilizando lo explicado anteriormente, se tendría lo siguiente:

- **Parte aleatoria:**  $Y_i \sim N(\mu_i, \sigma^2)$
- **Predictor lineal:**  $N_i = B_0 + B_i * x_i + x_i^2$
- **Enlace:**  $\mu_i = N_i$

Incluso a partir de la función cóncava del predictor lineal puede ser modelada particionando los datos y modelando la función mediante funciones indicadoras.

Por ejemplo, imaginemos que queremos modelar una variable de conteo. Utilizando lo explicado anteriormente, se tendría lo siguiente:

- **Parte aleatoria:**  $Y_i \sim Poisson(\mu_i)$
- **Predictor lineal:**  $N_i = B_0 + B_i * x_i$
- **Enlace:**  $\log(\mu_i) = N_i$

El modelo logístico tendría la siguiente estructura:

- **Parte aleatoria:**  $Y_i \sim \text{Bernoulli}(\mu_i)$
- **Predictor lineal:**  $N_i = B_0 + B_i * x_i$
- **Enlace:**  $\log(\frac{\mu_i}{1-\mu_i}) = N_i$

Dichas estructuras tienen la estructura de un modelo lineal generalizado. En donde  $Y_i \sim FE(\mu_i, \phi)$ , en donde FE significa familia exponencial y  $\phi$  es un parámetro de dispersión. Las distribuciones relacionadas a la familia exponencial: Normal, Gamma, Normal Inversa, Poisson, Bernoulli y toda familia doblemente diferenciable y monótona.

Los modelos que serán estudiados sirven para modelar la respuesta media. Existen otros métodos de regresión que modelan otros aspectos de la distribución de la variable respuesta (ejemplo: la regresión cuantílica sirve para modelar cuantiles).

Se tiene un modelo  $Y_i = B_0 + B_1 * X_2 * X_{i2} + \dots + B_k * X_{ik} + \varepsilon_i, i = 1, 2, \dots, n$ . La notación para manejar estos ejemplos más extensos se realiza de forma matricial.

Se tiene entonces  $Y = (Y_1, Y_2, \dots, Y_n)^T$  y  $B = (B_0, B_1, \dots, B_k)^T$  y

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & x_{i2} & \dots & x_{ik} \end{pmatrix}$$

y  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$

Bajo esta notación, todo se puede evaluar como  $Y = XB + \varepsilon; \varepsilon \sim N(1, \sigma^2 I)$ . El estimador de mínimos cuadrados ordinarios del siguiente modelo es  $\sum_{i=1}^n \varepsilon_i^2 = E^T E = (Y - XB)^T (Y - XB)$  y el estimador de B es  $LS(B) = Y^T T - 2Y^T X B + B^T X^T X B$ .

Derivamos la suma de cuadrados en relación a  $\beta$ , obteniendo  $-2X^T + 2X^T X B = 0$ . Posteriormente, tenemos que  $(X^T X)^{-1} X^T X B = (X^T X)^{-1} X^T Y$ . Se cancelan los argumentos y se tiene que  $\beta = (X^T X)^{-1} X^T Y$ .

Finalmente, debemos verificar que el estimador es insesgado y tiene varianza mínima (es el más eficiente). Hallamos el valor esperado de este estimador y obtenemos lo siguiente, que  $E(\hat{B}) = E((X^T X)^{-1} X^T Y) = (X^T X)^{-1} X^T E(Y) = (X^T X)^{-1} X^T X B = B$ . Por lo tanto, el estimador  $\hat{B}$  es insesgado.

Para hallar la varianza de B tenemos que  $cov(\hat{B}) = cov((X^T X)^{-1} X^T Y)$ , esto es  $(X^T X)^{-1} X^T cov(Y) = (X^T X)^{-1} \sigma^2 I$ . Multiplicamos ambos segmentos por  $(X^T X)^{-1}$ . Teniendo finalmente que la varianza de B es  $\sigma^2 (X^T X)^{-1}$ . Posteriormente, debemos probar que el estimador de B, el cual se halló mediante mínimos cuadrados ordinarios, es el mejor que existe. Para eso ver la siguiente demostración:

Sea el siguiente estimador lineal:

$$\tilde{B} = ((X^T X)^{-1} X^T + B)Y + b.$$

dónde  $\tilde{B}$  es una matriz  $p \times n$  cualquiera.

El esperado de esta matriz es:

$$\begin{aligned} E(\tilde{B}) &= E(((X^T X)^{-1} X^T + B)Y + b). \\ E((X^T X)^{-1} X^T Y + BY + b). \\ (X^T X)^{-1} X^T E(Y) + B * E(Y) + b. \\ (X^T X)^{-1} X^T XB + BXB + b. \\ B + BXB + b. \end{aligned}$$

Para que  $\tilde{B}$  sea insesgado se debe cumplir que tanto  $b = 0$  como  $BX = 0$ .

Una vez definido el valor esperado, observaremos la varianza del estimador mediante lo siguiente:

$$\begin{aligned} cov(\tilde{B}) &= cov((X^T X)^{-1} X^T + B)Y + b). \\ ((X^T X)^{-1} X^T + B)cov(Y)(X(X^T X)^{-1} + B^T). \\ ((X^T X)^{-2} X^T + B)\sigma^2 I(X(X^T X)^{-1} + B^T). \\ \sigma^2(X^T X)^{-1} X^T X(X^T X)^{-1} + \sigma B X(X^T X)^{-1} + \sigma^2(X^T X)^{-1} X^T B^T + \sigma^2 B B^T. \\ \sigma^2(X^T X)^{-1} + \sigma^2 B B^T. \end{aligned}$$

Posteriormente, hallaremos la varianza de una combinación lineal de  $\tilde{B}$ . Definimos entonces  $l$  como una matriz con dimensión  $p \times 1$ . Se tiene entonces lo siguiente:

$$\begin{aligned} Var(l^T \tilde{B}) &= l^T var(\tilde{B})l. \\ l^T (\sigma^2(X^T X)^{-1} + \sigma^2 B B^T)l. \\ l^T (\sigma^2(X^T X)^{-1}l + \sigma l^T B B^T l. \\ l^T var(\tilde{B})l + \sigma^2 l^T l. \\ var(l^T \tilde{B}) + \sigma^2 l^T l. \end{aligned}$$

El término  $\sigma^2 l^T l$  es mayor igual a cero. Por lo tanto, se tiene que:

$$var(l^T \tilde{B}) \leq var(l^T \hat{B}).$$

$\hat{B}$  es el mejor estimador lineal insesgado.

Posteriormente, analizaremos el residual. **Ojo:** Se conoce como el residual como el estimador del error. El residual es definido como  $e = Y - \hat{Y}$ , lo cual tiene lo siguiente:

$$\begin{aligned} Y - X\hat{B}. \\ Y - X(X^T X)^{-1} X^T Y. \\ Y - HY. \\ e = (I - H)Y. \end{aligned}$$

Posteriormente, definiremos  $e^T e$  como la suma de cuadrados de los residuales y hallaremos su valor esperado. Esto es:

$$E(e^T e) = E(Y^T (I - H)(I - H)Y).$$

$$E(Y^T (I - H)Y).$$

$$tr((I - H)cov(Y)) + E(Y)^T (I - H)E(Y).$$

$$tr((I - H)\sigma^2 I) + B^T X^T (I - H)XB.$$

$$\sigma^2 tr(I - H) + B^T X^T XB - B^T X^T HXB.$$

$$\sigma^2(n - p) + B^T X^T XB - B^T X^T XB.$$

$$E(e^T e) = \sigma^2(n - p).$$

Se tiene entonces que:

$$E\left(\frac{e^T e}{n - p}\right) = \sigma^2.$$

Por lo tanto, podemos definir  $\hat{\sigma}^2 = \frac{e^T e}{n - p}$  como un estimador insesgado de  $\sigma^2$ .

Se definen, en base a lo mencionado, distintas propiedades. Estas son:

- $\hat{Y}^T e = 0$
- $\hat{X}^T e = 0$
- $\sum_{i=1}^n e_i = 0$
- $\sum_{i=1}^n \hat{Y}_i = \sum_{i=1}^n Y_i$
- La recta de regresión estimada pasa por el punto  $(1, \bar{X}_1, \bar{X}_2, \dots, \bar{X}_K, \bar{Y})$