REGRESIÓN LINEAL

DADO UN CONJUNTO DE CASOS CON UNA V.O. CONTINUA Y UN GRUPO DE VARIABLES DE SOPORTE ASIMISMO CONTINUAS:

$$V.O. = Y \in [y_{min}, y_{max}]$$

$$V.S = \{X_1, X_2, X_3, \dots X_n\}$$

EL CASO i-ÉSIMO ES

$$y_i,\ x_{i,1}, x_{i,2}, \dots x_{i,n}$$

COMO "MODELO" SE PROPONE

$$Y = A_0 + A_1X_1 + A_2X_2 + \dots + A_nX_n$$

ESTO LO PODEMOS EXPRESAR COMO

$$Y = [A_0, A_1, \dots A_n] \cdot [1, X_1, X_2, \dots X_n]$$

¿CUÁLES SON LOS VALORES ADECUADOS DE LAS A's?

PARA ELLO DEBEMOS DEFINIR UN CRITERIO DE OPTIMALIDAD.

VEAMOS UNA ALTERNATIVA:

POR LO GENERAL, TENDREMOS QUE

$$y_i \neq A_0 + A_1 x_{i,1} + A_2 x_{i,2} + \dots + A_n x_{i,n}$$

O BIEN

$$\varepsilon_i = A_0 + A_1 x_{i,1} + A_2 x_{i,2} + \dots + A_n x_{i,n} - y_i$$

VEAMOS QUE OCURRE SI SOLAMENTE TOMAMOS COMO MODELO A_0:

$$\varepsilon_i = A_0 - y_i$$

CANTIDAD DE CASOS

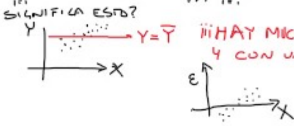
SUMANDO LOS ERRORES ε_i PARA $i = 1, \dots, m$

$$SE = \sum_{i=1}^{m} (A_0 - y_i) = mA_0 - \sum_{i=1}^{m} y_i$$

PODEMOS CALCULAR A_0 TAL QUE $SE = 0$:

$$mA_0 - \sum_{i=1}^{m} y_i = 0 \Rightarrow A_0 = \frac{1}{m}\sum_{i=1}^{m} y_i \quad \dots \text{ PROMEDIO DE } Y$$

¿QUÉ SIGNIFICA ESTO?



¡¡¡HAY MUCHA DISPERSIÓN Y CON UN PATRÓN!!!

NECESITAMOS UN "MEJOR" MODELO.

AGREGUEMOS OTRA A. PARA QUE SE INCLUYA UN EFECTO POR EL VALOR DE X:

$$Y = A_0 + A_1 X$$

EN ESTE CASO, $\varepsilon_i = A_0 + A_1 X_1 - y_i$

TOMANDO $SE = \sum_{i=1}^{m} \varepsilon_i = mA_0 + A_1 \sum_{i=1}^{m} X_{i,1} - \sum_{i=1}^{m} y_i = 0$

TENEMOS 2 INCÓGNITAS Y SOLAMENTE UNA ECUACIÓN. LE PODEMOS ASIGNAR CUALQUIER VALOR A A_0 Y DESPEJAR A_1 O VICEVERSA, PERO ELLO NO GARANTIZA QUE EL MODELO SEA EL CORRECTO.

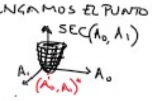
CONSIDEREMOS UN CRITERIO DE OPTIMALIDAD QUE MINIMICE LA DISPERSIÓN EN ε_i.

$$SEC = \sum_{i=1}^{m} \varepsilon_i^2 = \sum_{i=1}^{m} (A_0 + A_1 x_{i,1} - y_i)^2$$

Y HAY QUE ELEGIR A_0 Y A_1 QUE MINIMICEN SEC.

PARA HACER ESTO, OBTENGAMOS EL PUNTO CRÍTICO DE $SEC(A_0, A_1)$.

$$SEC(A_0, A_1) \geq 0$$



DERIVANDO RESPECTO A A_0:

$$\frac{\partial SEC}{\partial A_0} = 0 \Rightarrow \frac{\partial \sum_{i=1}^{m}}{\partial A_0}(A_0 + A_1 x_{i,1} - y_i)^2 =$$

$$= 2\sum_{i=1}^{m}(A_0 + A_1 x_{i,1} - y_i)\frac{\partial \sum}{\partial A_0}(A_0 + A_1 x_{i,1}) =$$

$$= \sum_{i=1}^{m}(A_0 + A_1 x_{i,1} - y_i) = 0$$

$$\Rightarrow mA_0 + A_1 \sum_{i=1}^{m} x_{i,1} = \sum_{i=1}^{m} y_i \quad \dots (1)$$

VEAMOS AHORA $\frac{\partial SEC}{\partial A_1} = 0$

$$\frac{\partial SEC}{\partial A_1} = \frac{\partial}{\partial A_1}\sum(A_0 + A_1 x_{i,1} - y_i)^2 = 0$$

$$\Rightarrow 2\sum(A_0 + A_1 x_{i,1} - y_i)\underbrace{\frac{\partial \sum}{\partial A_1}(A_0 + A_1 x_{i,1} - y_i)}_{x_{i,1}} = 0$$

$$\Rightarrow \sum_{i=1}^{m}(A_0 x_{i,1} + A_1 x_{i,1}^2 - y_i x_{i,1}) = 0$$

ó

$$A_0 \sum x_{i,1} + A_1 \sum_{i=1}^{m} x_{i,1}^2 = \sum_{i=1}^{m} y_i x_{i,1} \quad \dots (2)$$

TENEMOS ASÍ 2 ECUACIONES LINEALES CON 2 INCÓGNITAS.

ESTE SISTEMA PODEMOS ESCRIBIRLO COMO

$$\begin{bmatrix} m & \sum_{i=1}^{m} x_{i,1} \\ \sum_{i=1}^{m} x_{i,1} & \sum_{i=1}^{m} x_{i,1}^2 \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_{i,1} y_i \end{bmatrix}$$

REPRESENTANDO LAS SUMAS DE MANERA MÁS SIMPLE:

$$\begin{bmatrix} m & SX_1 \\ SX_1 & SX_1^2 \end{bmatrix} \begin{bmatrix} A_0 \\ A_1 \end{bmatrix} = \begin{bmatrix} SY \\ SX_1Y \end{bmatrix}$$

Y TIENE SOLUCIÓN ÚNICA SI

$$m SX_1^2 \neq (SX_1)^2$$