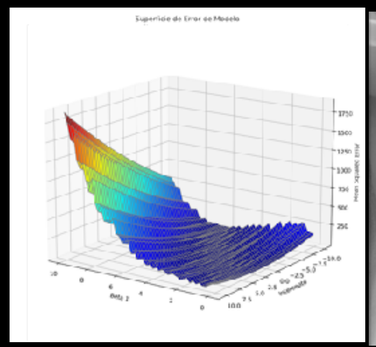
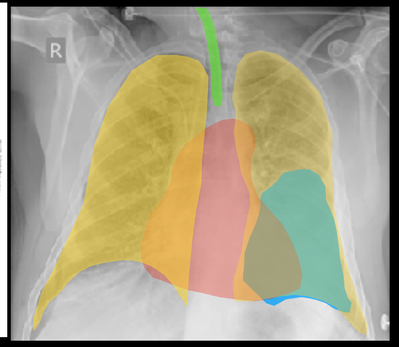
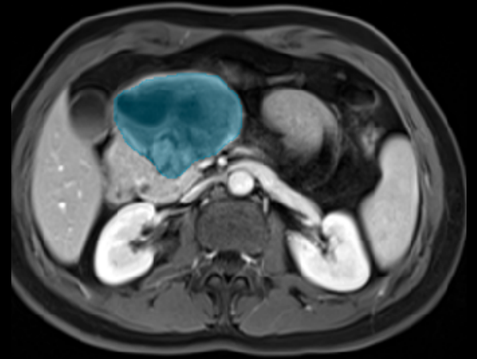
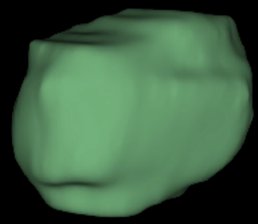
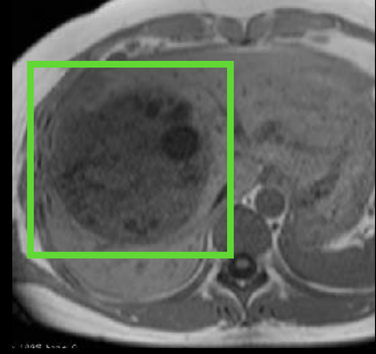







Estadistica parte 2

Curso de Investigación en Radiología

Héctor Henríquez Leighton MD, MS
hhenriquez@miuandes.cl
Profesor Asistente Universidad de los Andes
Investigador Principal CVLab U Andes
Clínica Santa Maria, Santiago de Chile

Clínica Santa María
Especialistas en ti

Universidad de los Andes

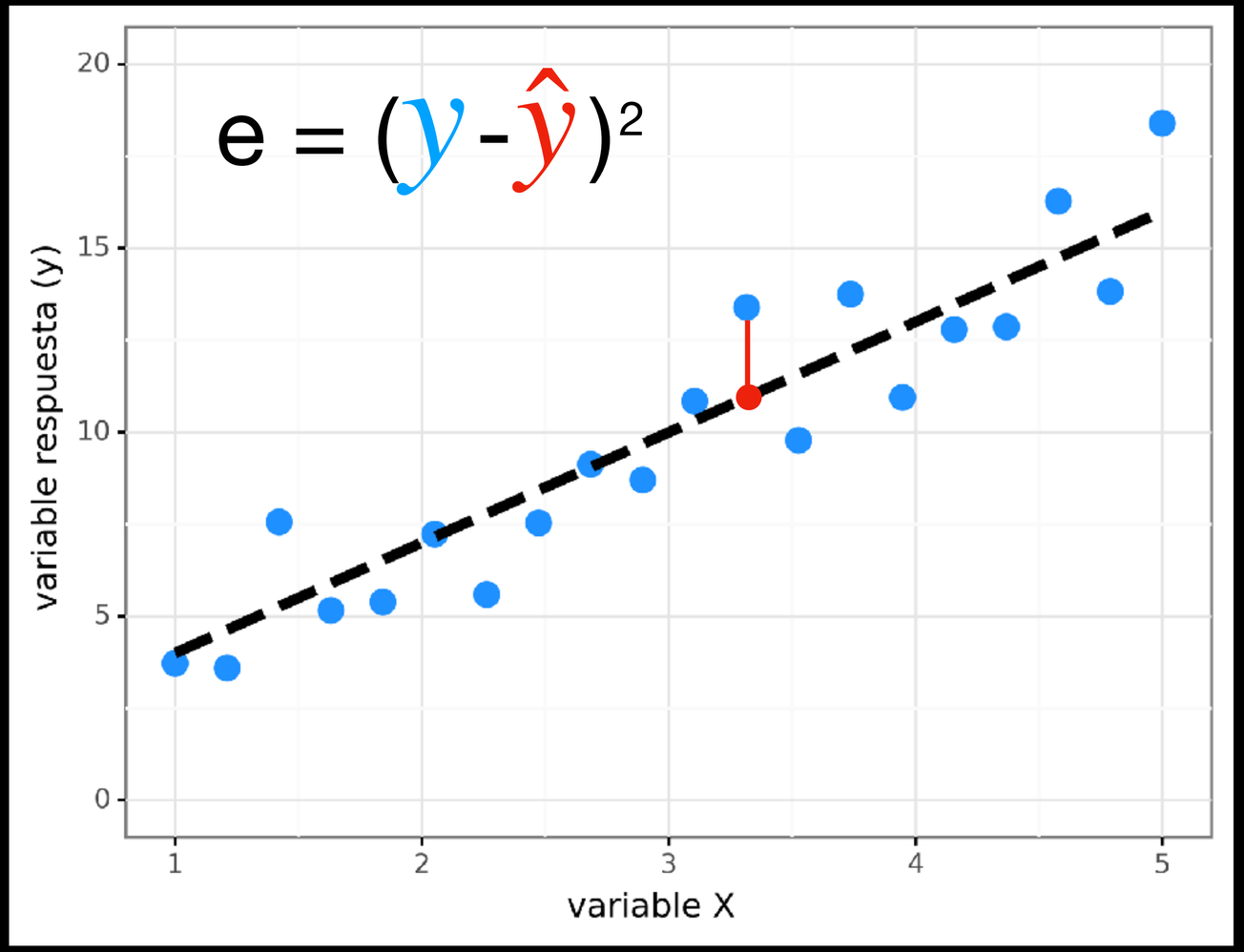
Clínica Santa María
Especialistas en ti

Contenidos

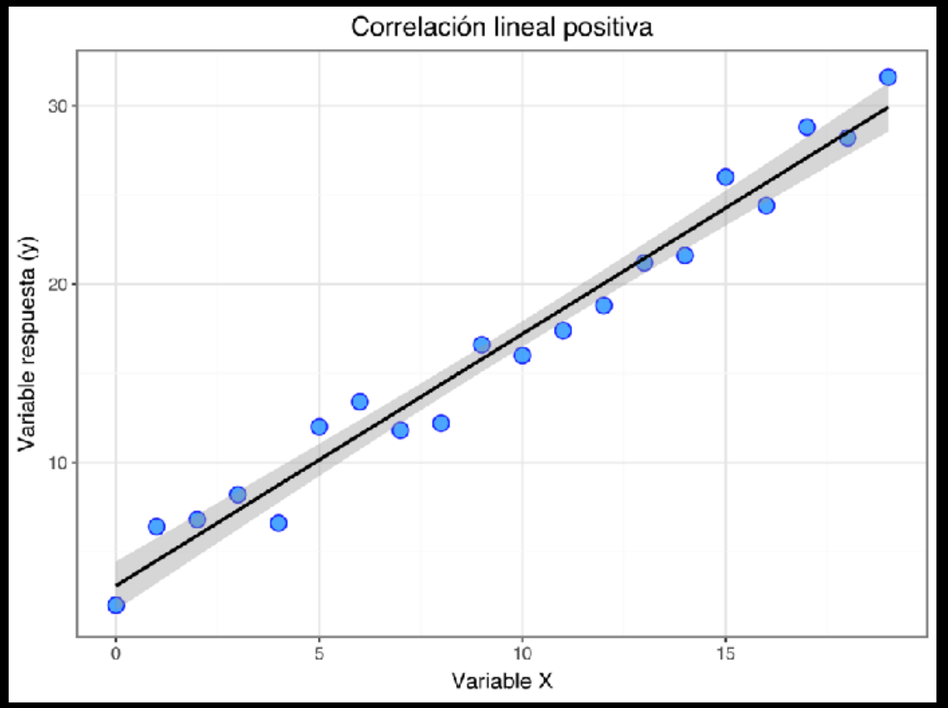
- Correlación entre variables
- Modelos de regresión lineal.
- Métricas de regresión lineal.
- Modelos de regresión logística.
- Métricas de clasificación.

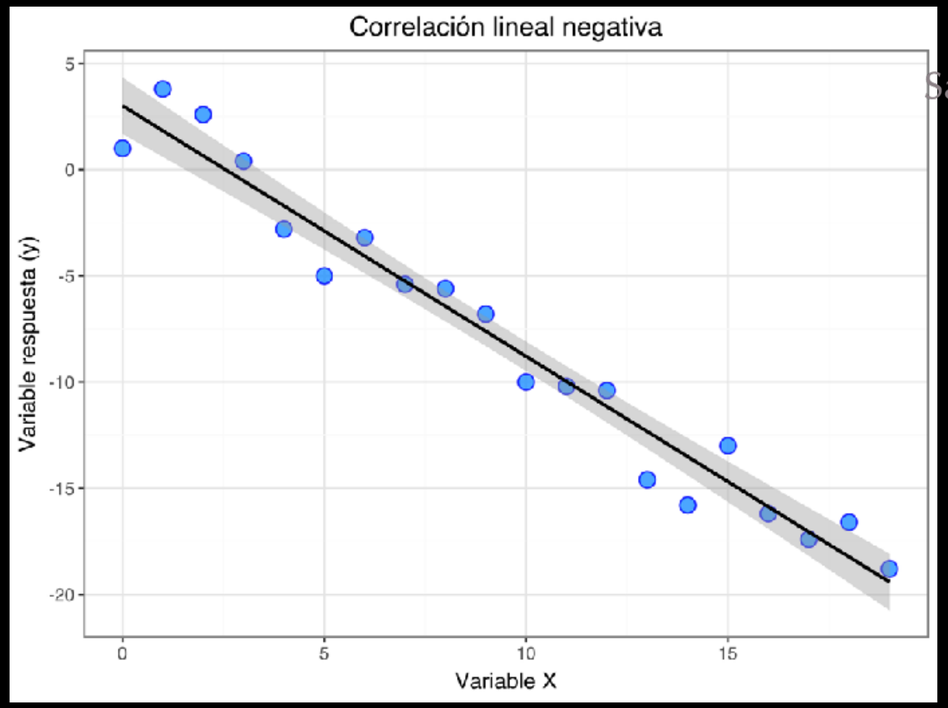
ABC de modelos predictivos

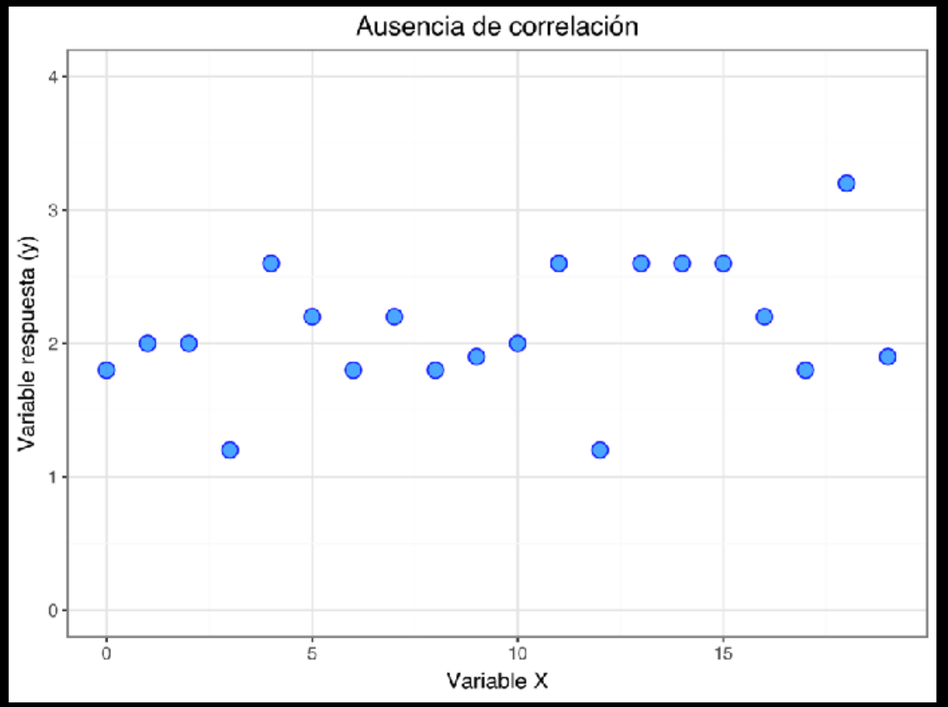
$$f(x) = y \quad \rightarrow \quad \hat{y} = \beta_0 + \beta_1 X + \epsilon$$

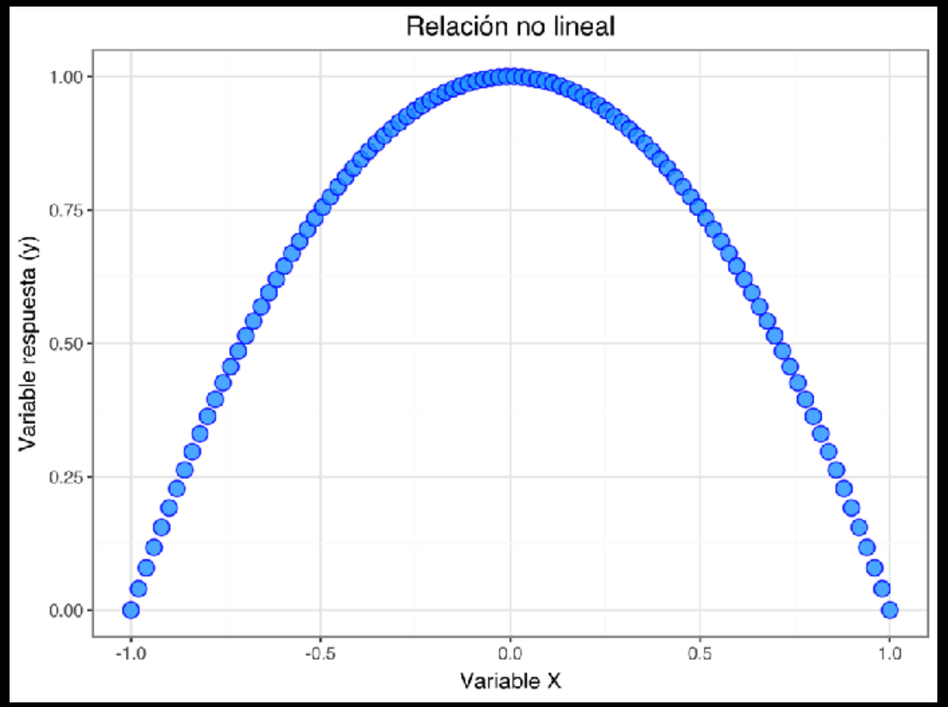


- *Regresión*
- *Clasificación*









Clínica Santa María
Especialistas en ti

Medidas de dependencia lineal

- Covarianza:
  - Si hay relación lineal positiva, la covarianza es positiva y grande.
  - Si hay relación lineal negativa, la covarianza es negativa y valor absoluto grande.
  - Si no hay relación entre las variables o marcadamente no lineal, la covarianza es próxima a cero.
  - El problema: covarianza depende de las unidades de medida de las variables.

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Clínica Santa María
Especialistas en ti

Medidas de dependencia lineal

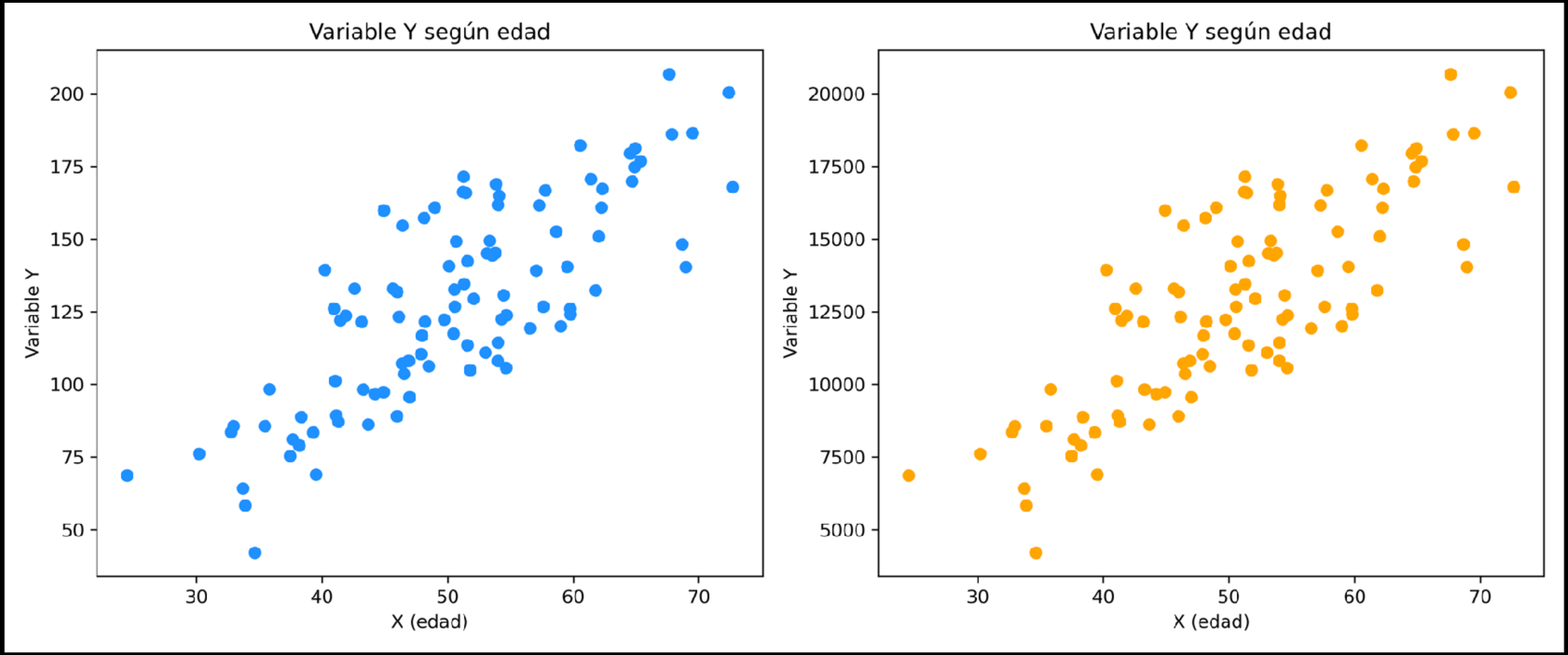
- Coeficiente de correlación lineal:
  - Medida de dependencia lineal que no depende de las unidades de medida de las variables.
  - Valores entre -1 y 1

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

Medidas de dependencia lineal

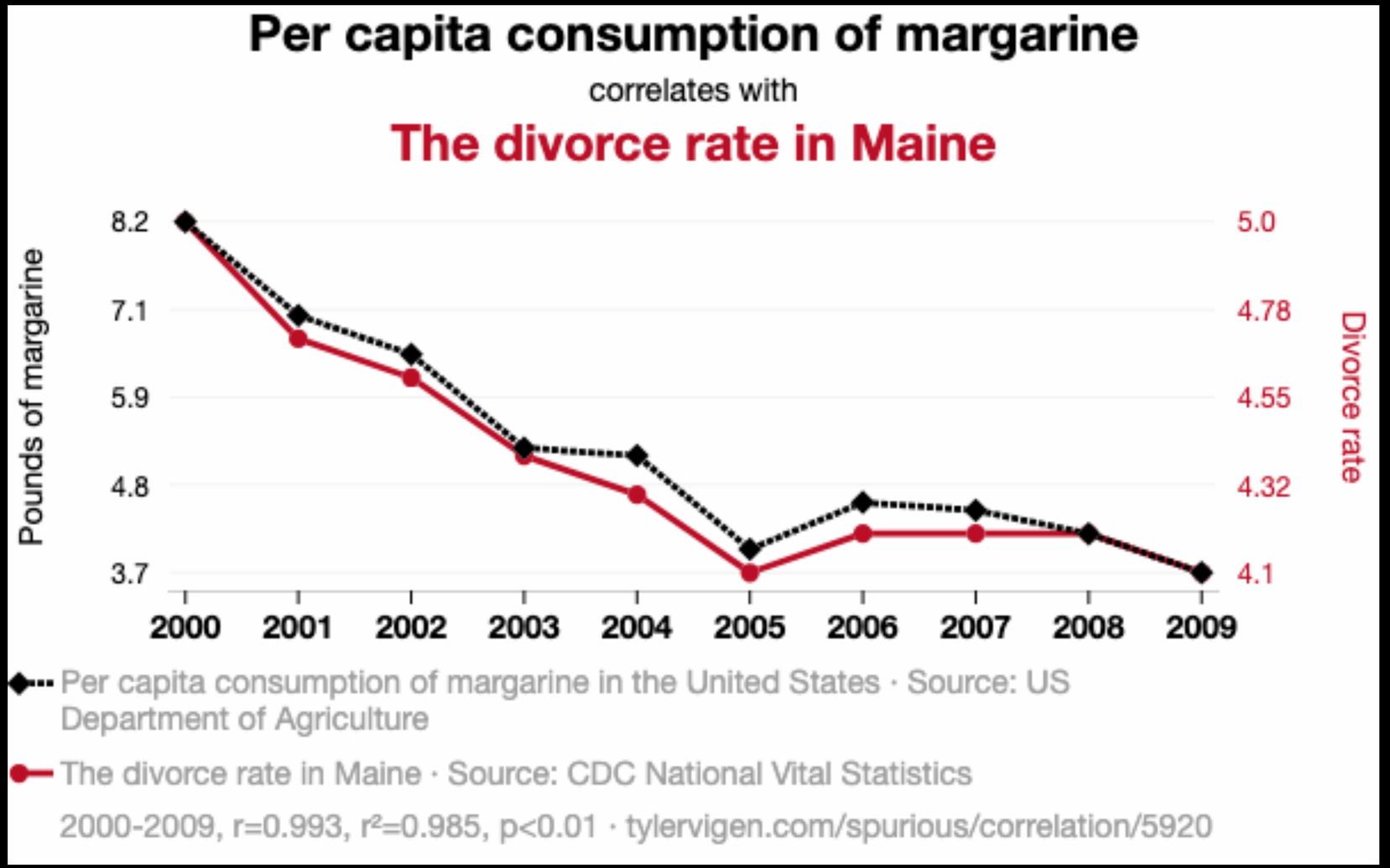
Covarianza: 280.05
Correlación lineal: 0.80

Covarianza (y escalado): 28005.99
Correlación lineal(y escalado): 0.80



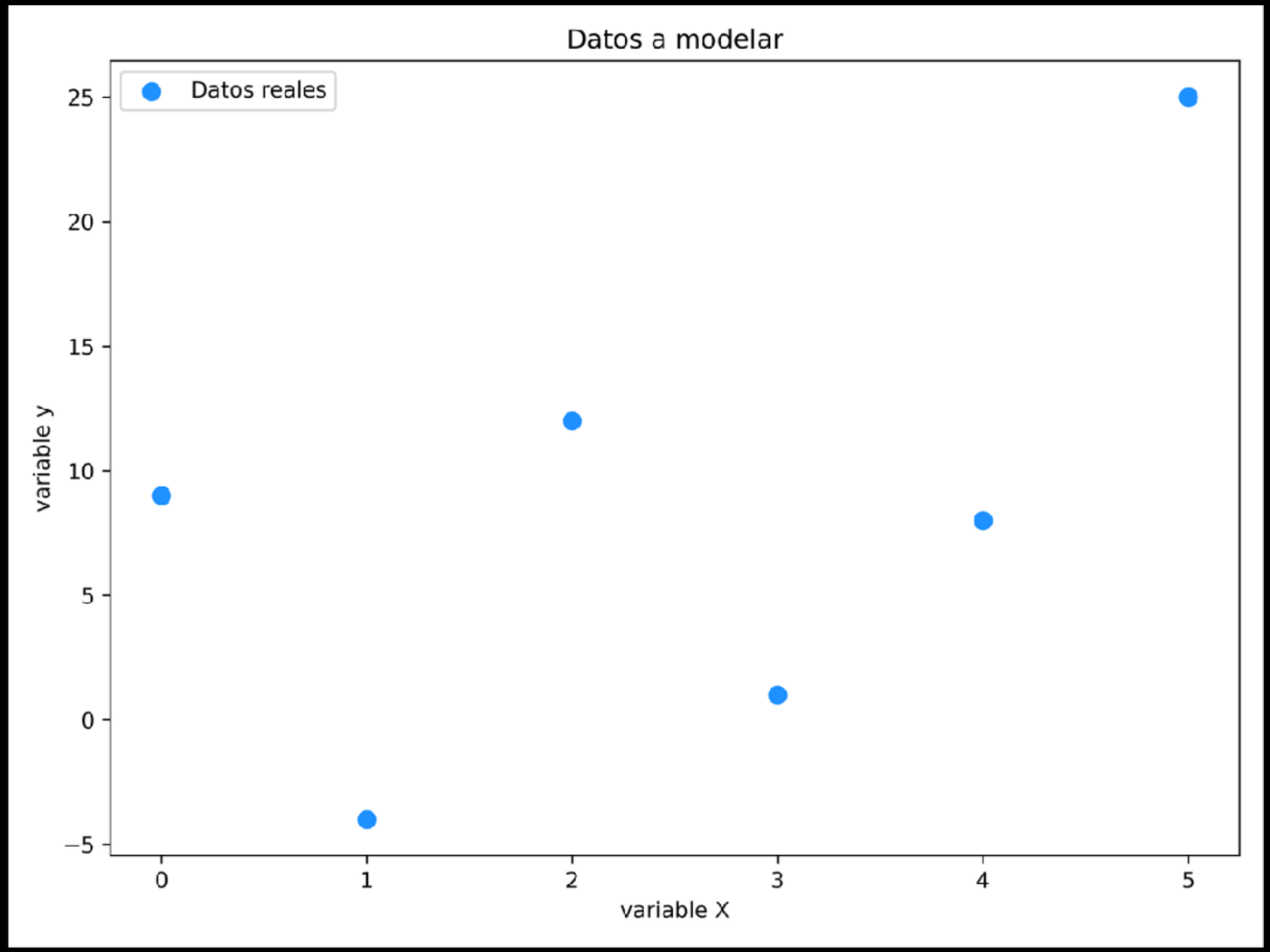
Clínica Santa María
Especialistas en ti

Correlación no es causalidad



Regresión lineal simple

Clínica Santa María
Especialistas en ti



Regresión lineal simple

$$\hat{y} = \beta_0 + \beta_1 X_1 + \epsilon$$

Clínica Santa María
Especialistas en ti

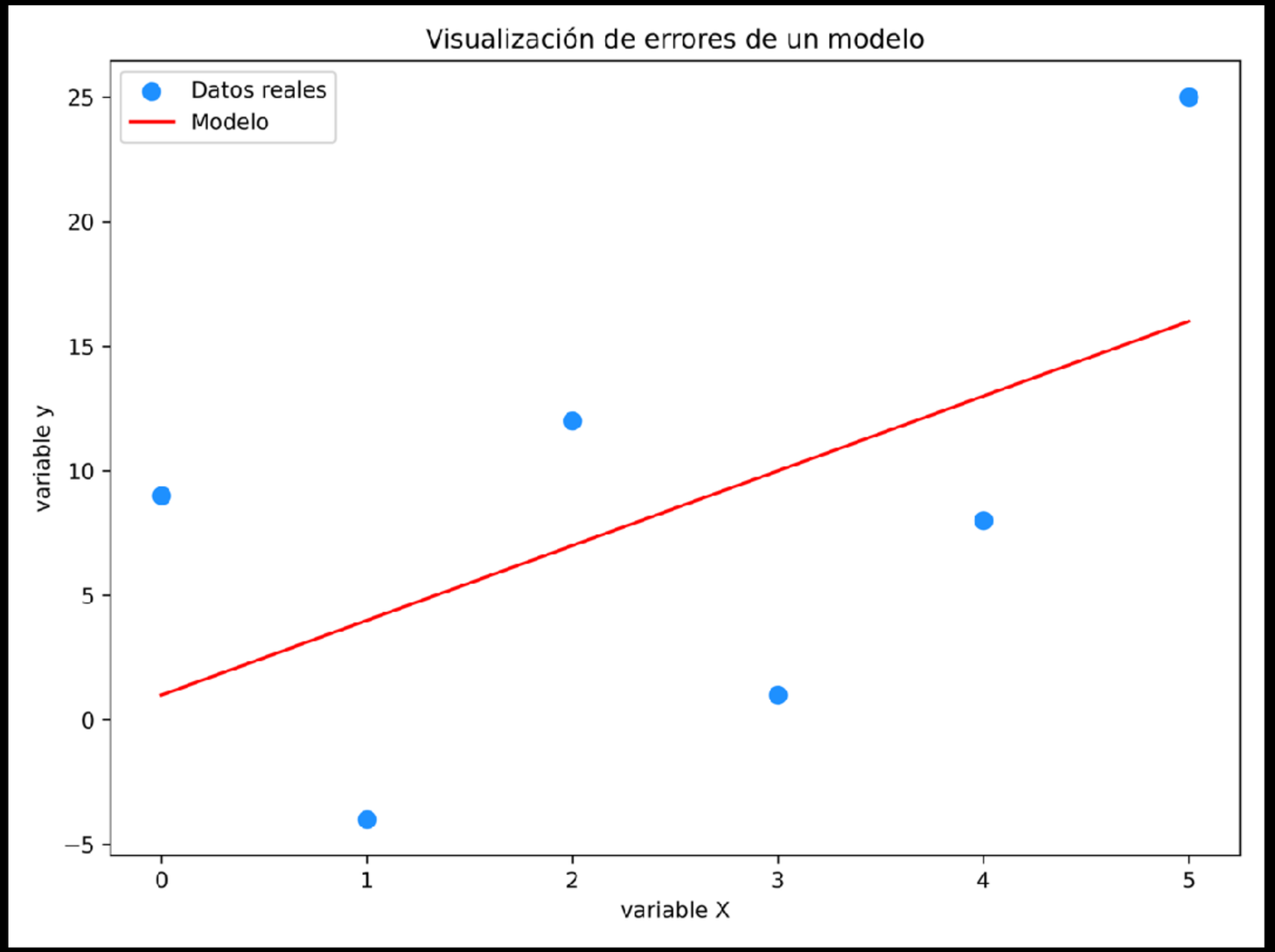
$\hat{y}$: *Predicción*

β_0 : *Intercepto*

β_1 : *Pendiente*

X_1 : *Variable independiente*

ϵ : *Ruido*

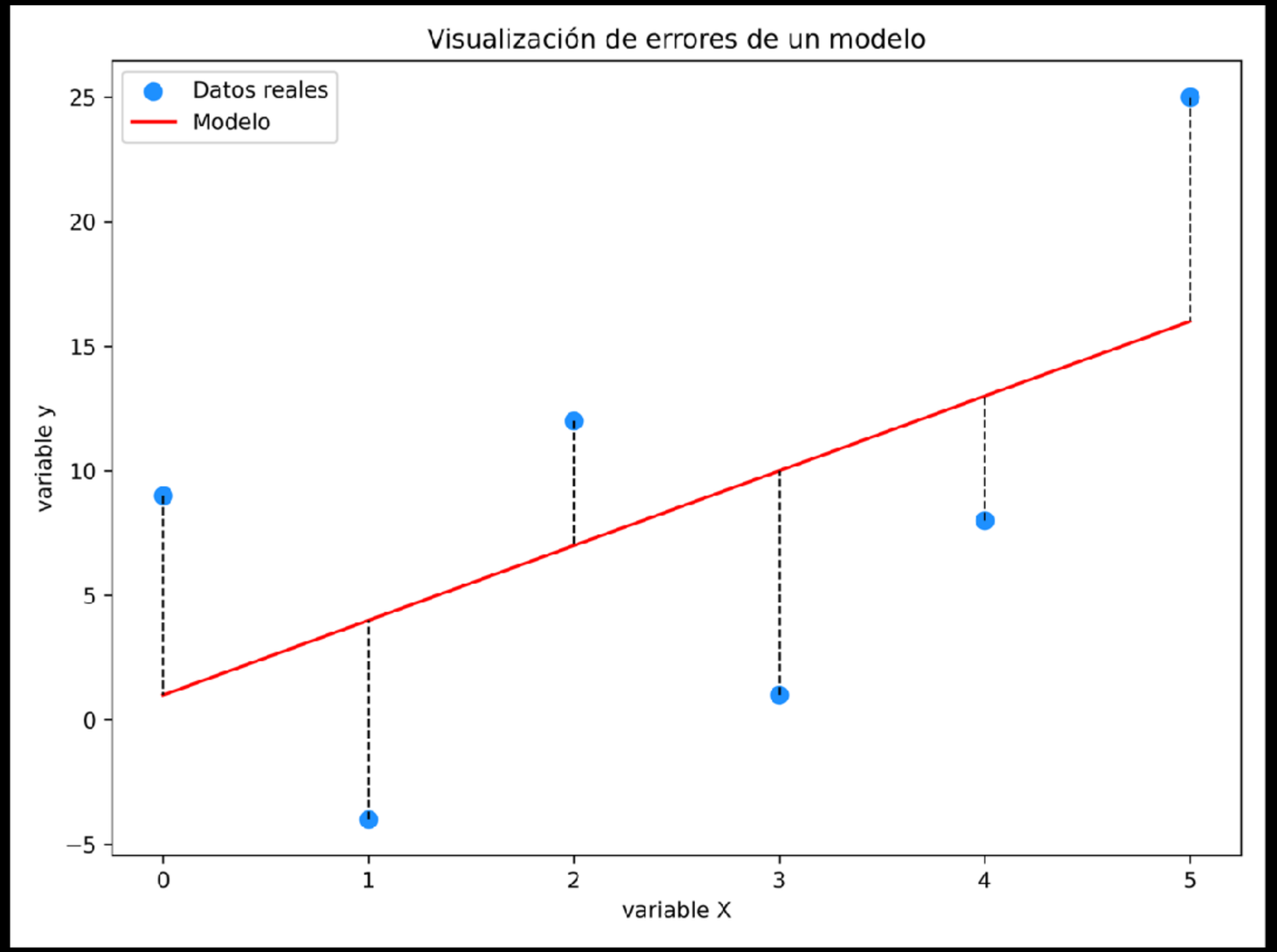


Regresión lineal simple

$$\hat{y} = 1 + 3X_1$$

Clínica Santa María
Especialistas en ti

x	y	Pred
0	9	1
1	-4	4
2	12	7
3	1	10
4	8	13
5	25	16

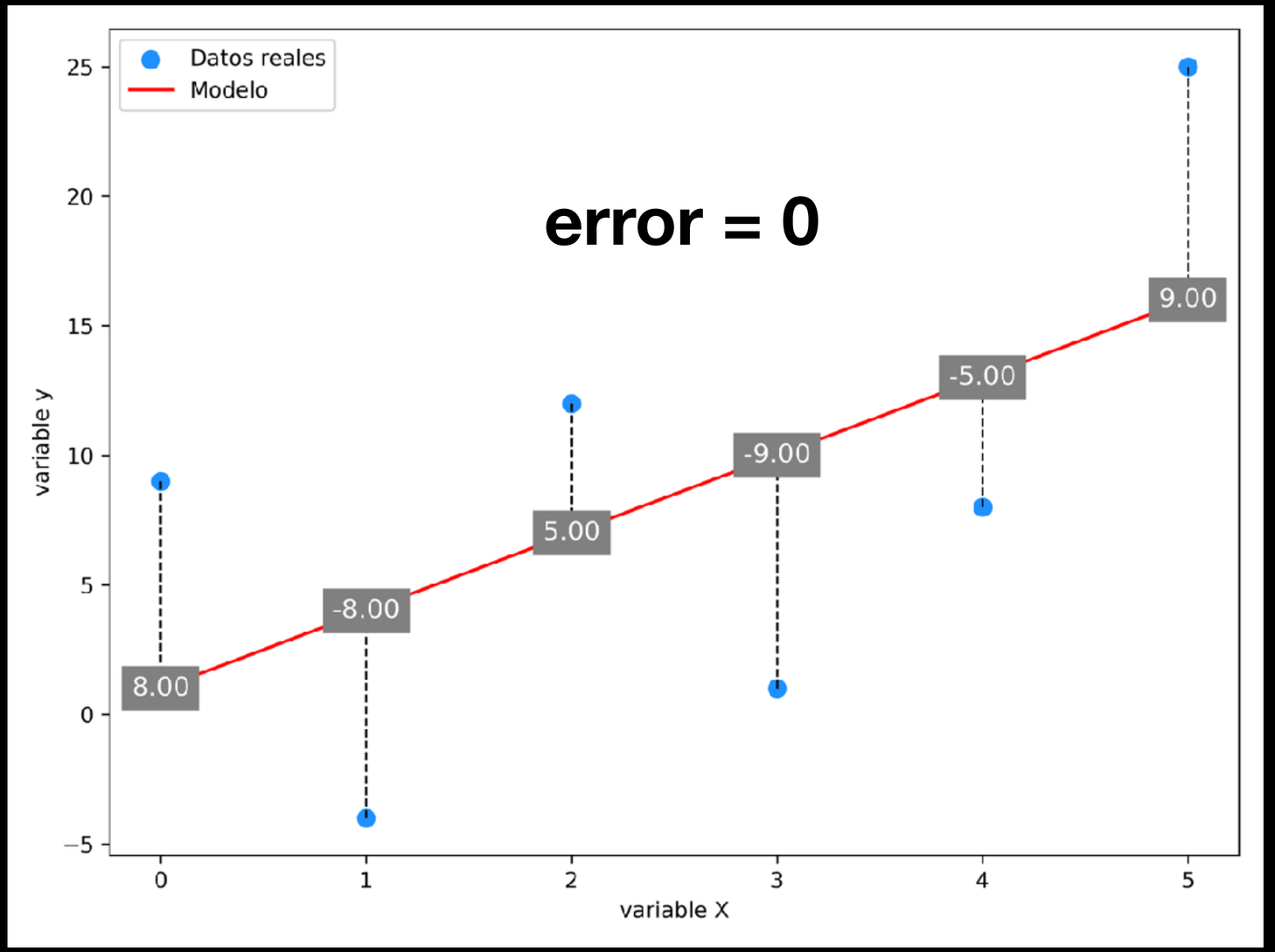


Regresión lineal simple

$$\hat{y} = 1 + 3X_1$$

Clínica Santa María
Especialistas en ti

x	y	Pred
0	9	1
1	-4	4
2	12	7
3	1	10
4	8	13
5	25	16



Métricas

$$\hat{y} = 1 + 3X_1$$

Error Medio Absoluto

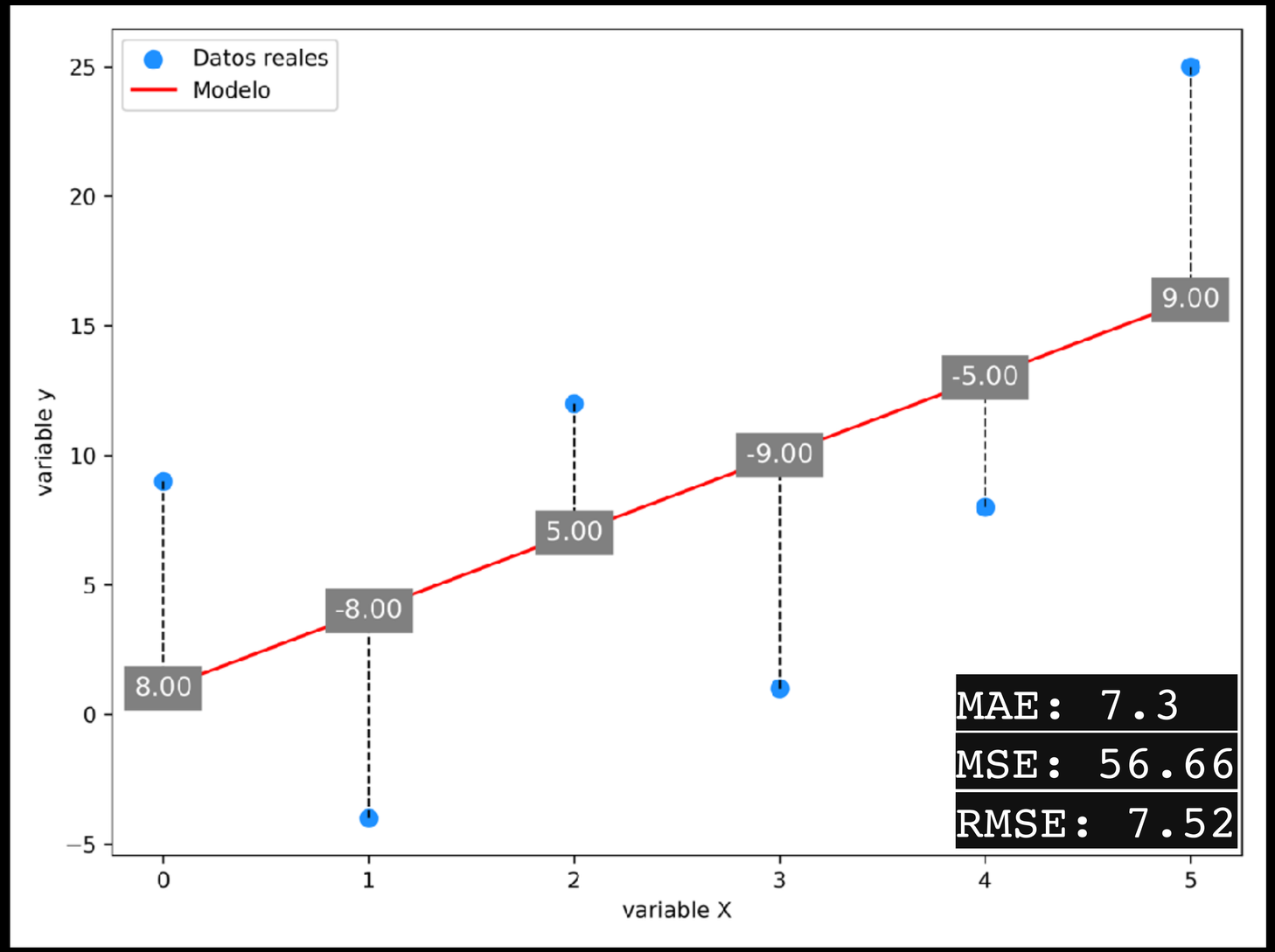
$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|$$

Error Cuadrático Medio

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

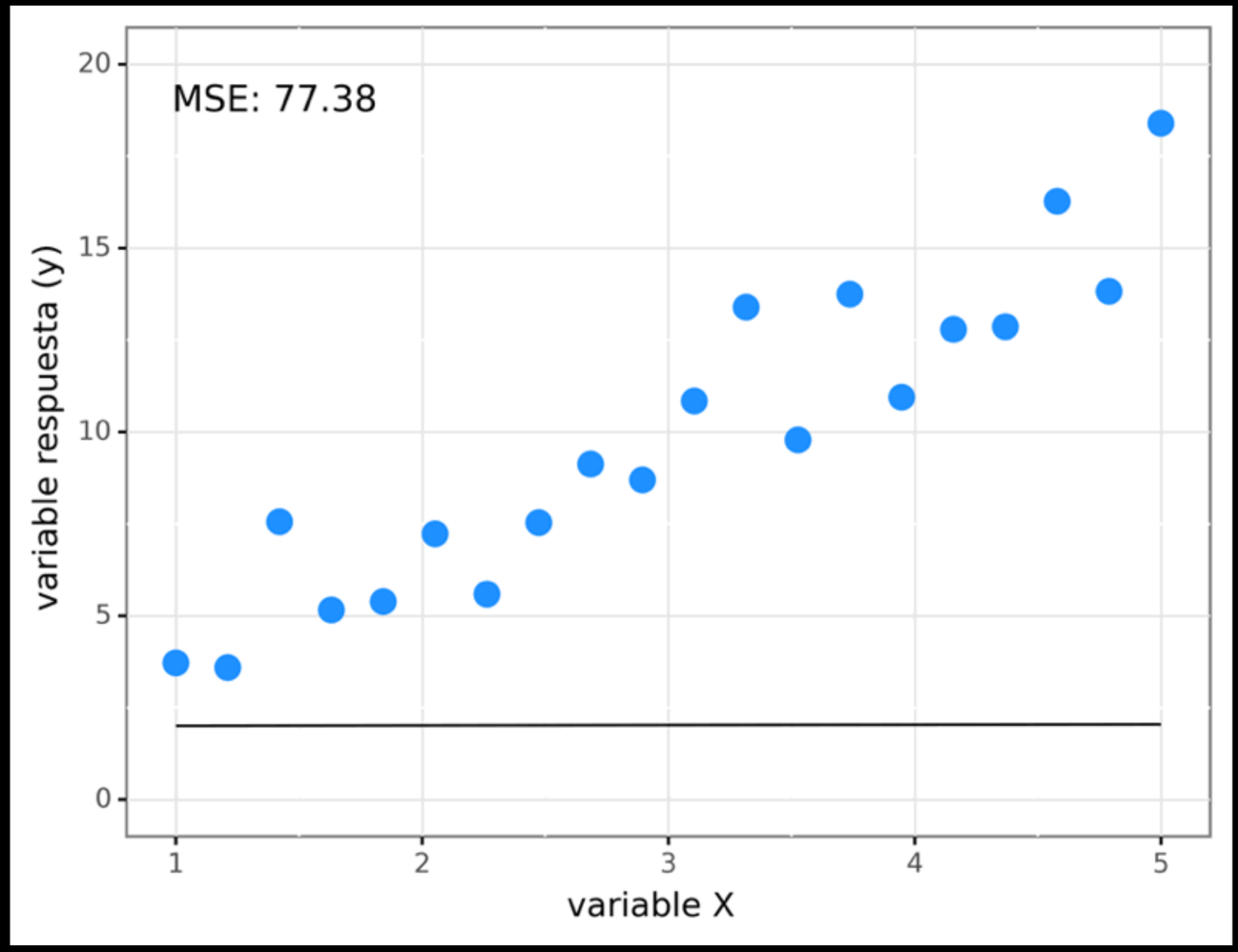
Raíz Cuadrada del Error Cuadrático Medio

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}$$



Entrenamiento del modelo

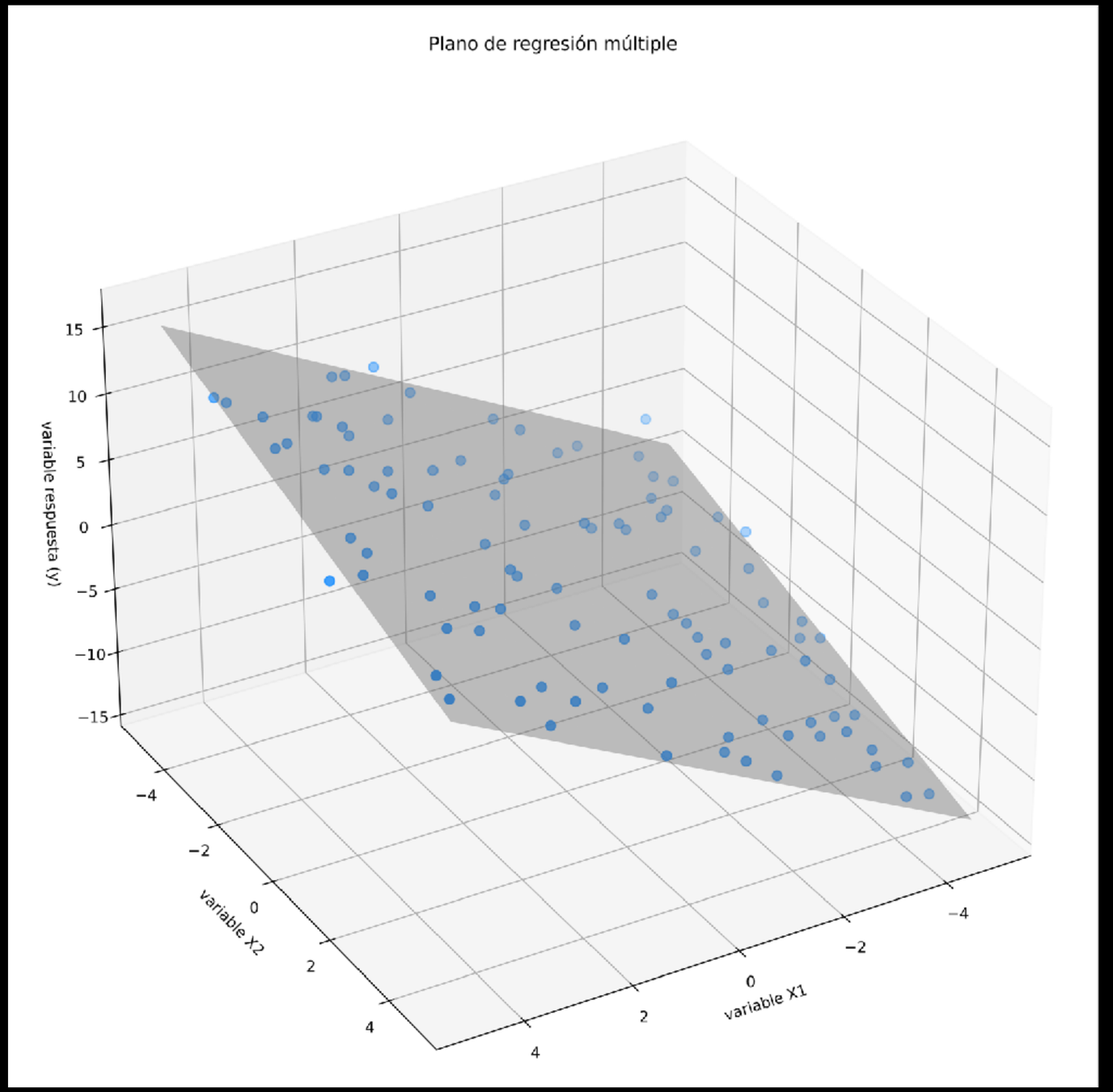
$$f(x) = y \quad \rightarrow \quad \hat{y} = \beta_0 + \beta_1 X + \epsilon$$



Iteración	MSE
1	77,38
2	
3	
4	
5	
6	
7	
8	
9	
10	

Regresión lineal múltiple

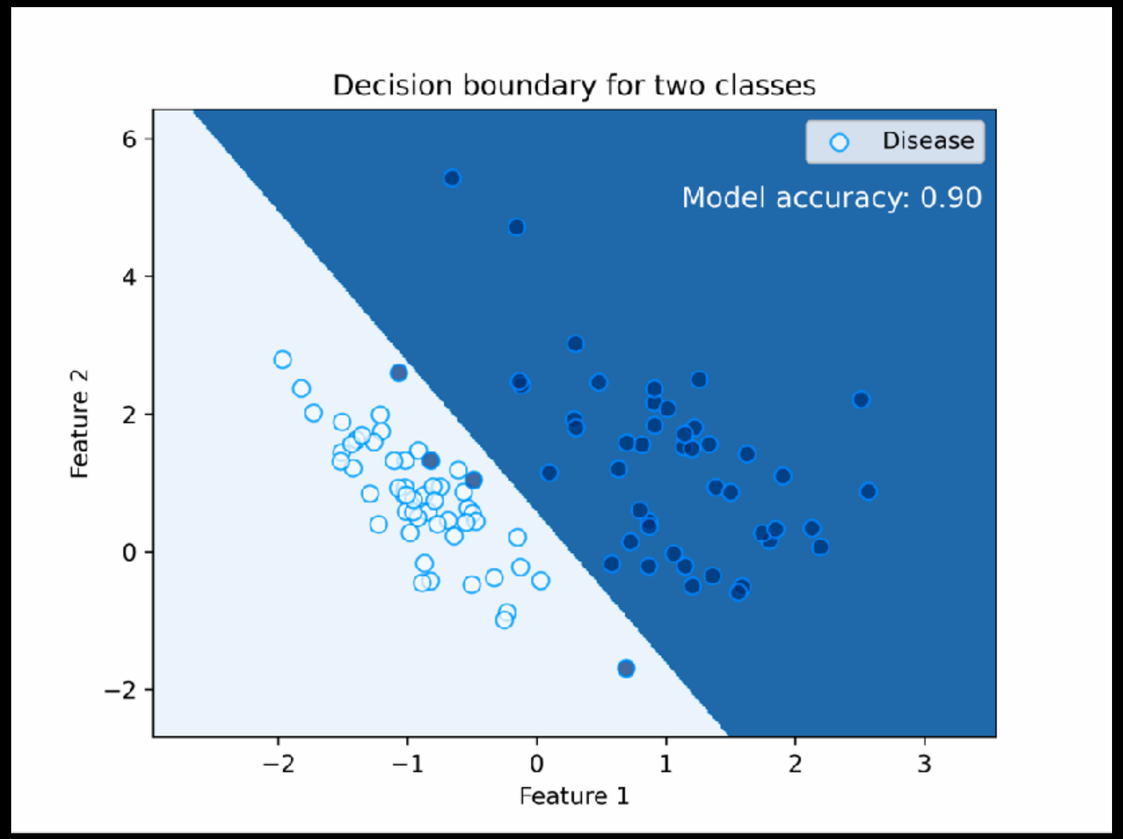
$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$



Clínica Santa María
Especialistas en ti

Modelos de clasificación

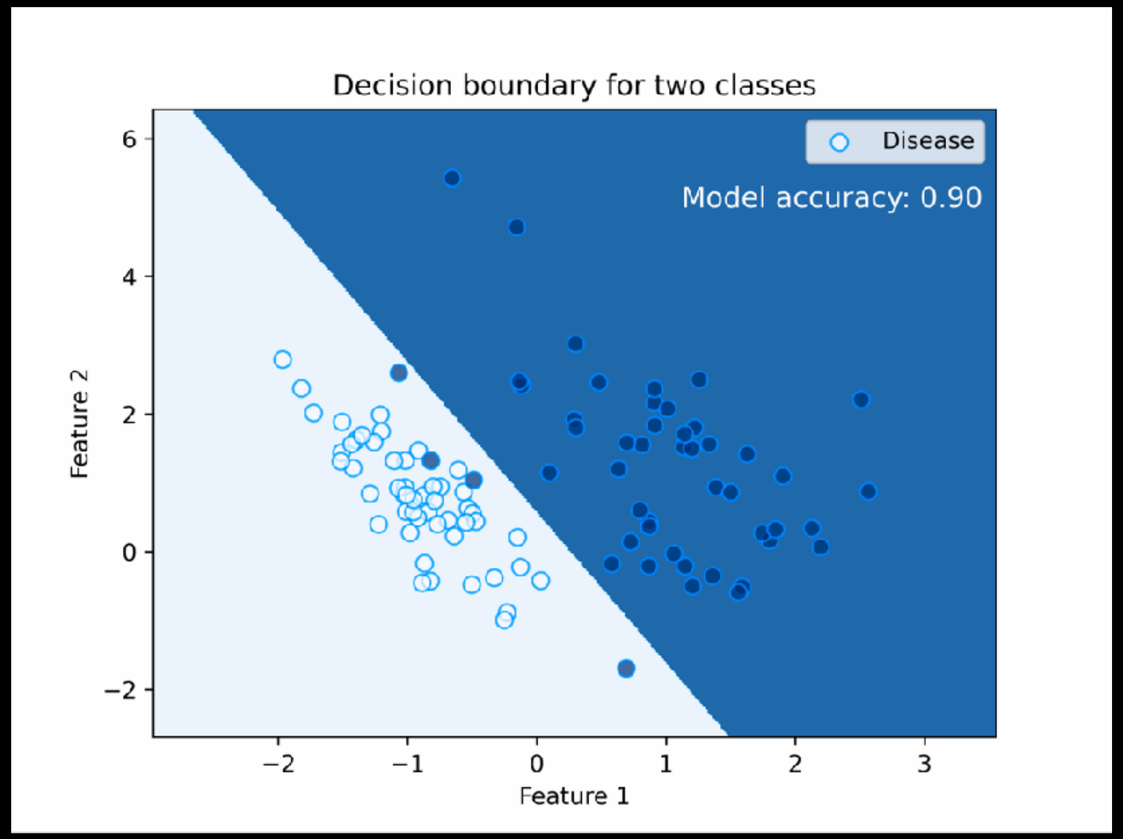
Modelo lineal



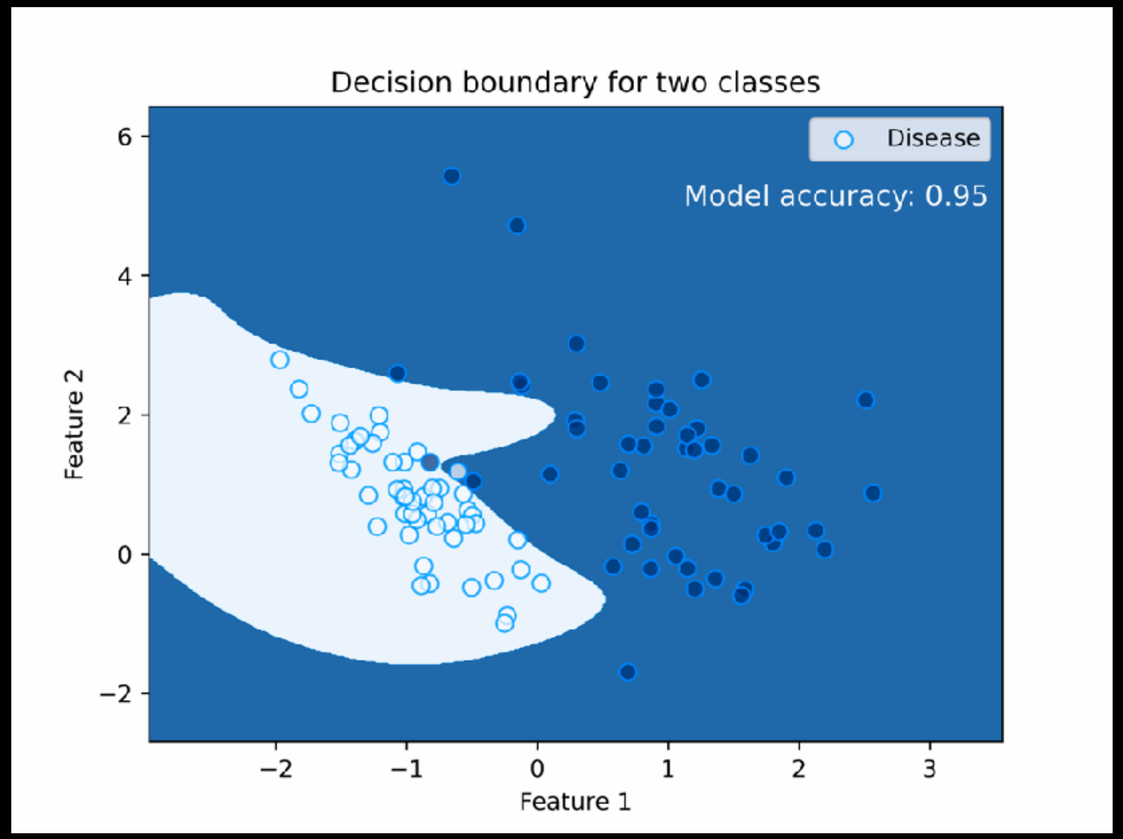
Clínica Santa María
Especialistas en ti

Modelos de clasificación

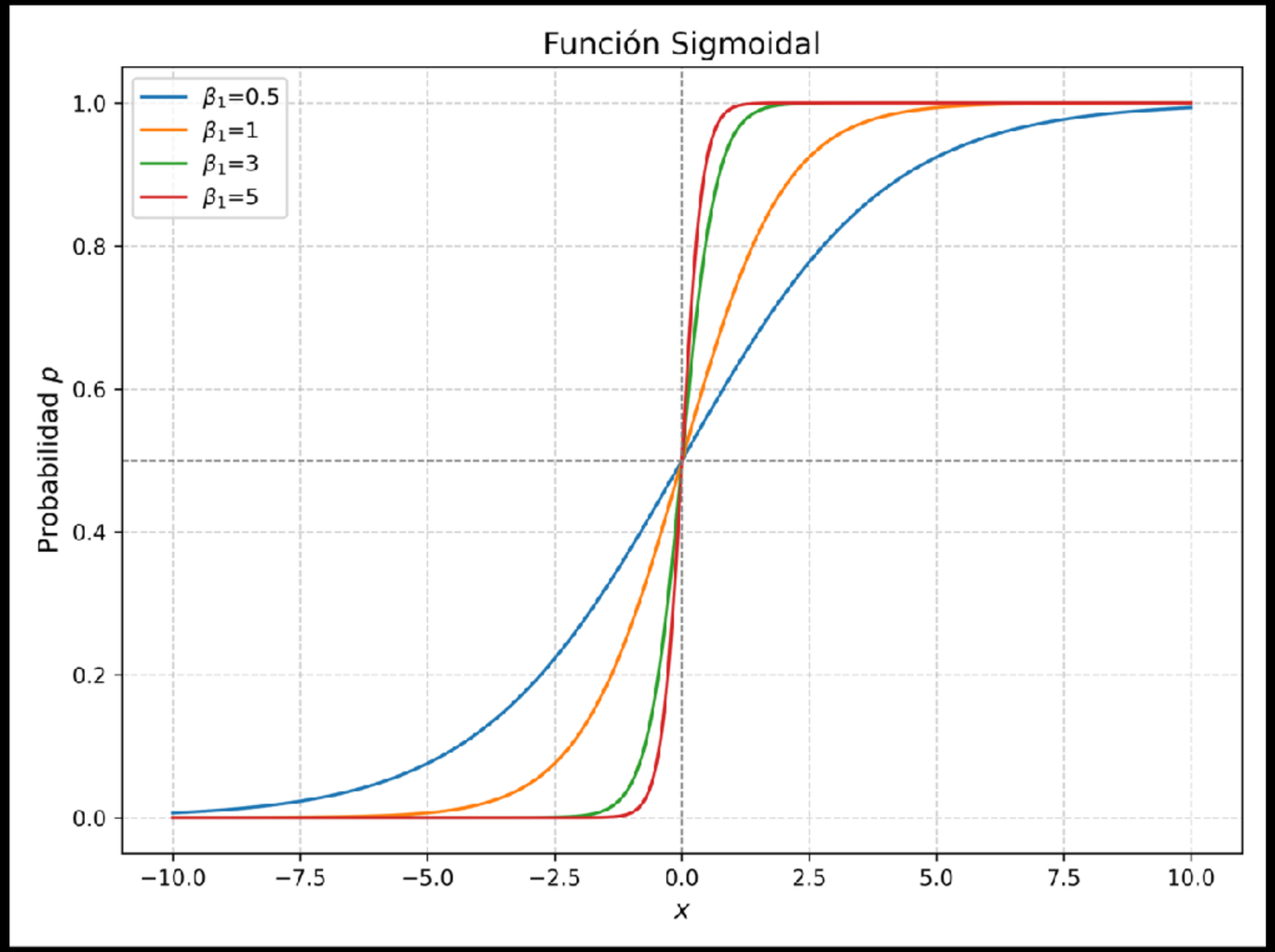
Modelo lineal



Support Vector Machine



Regresión logística



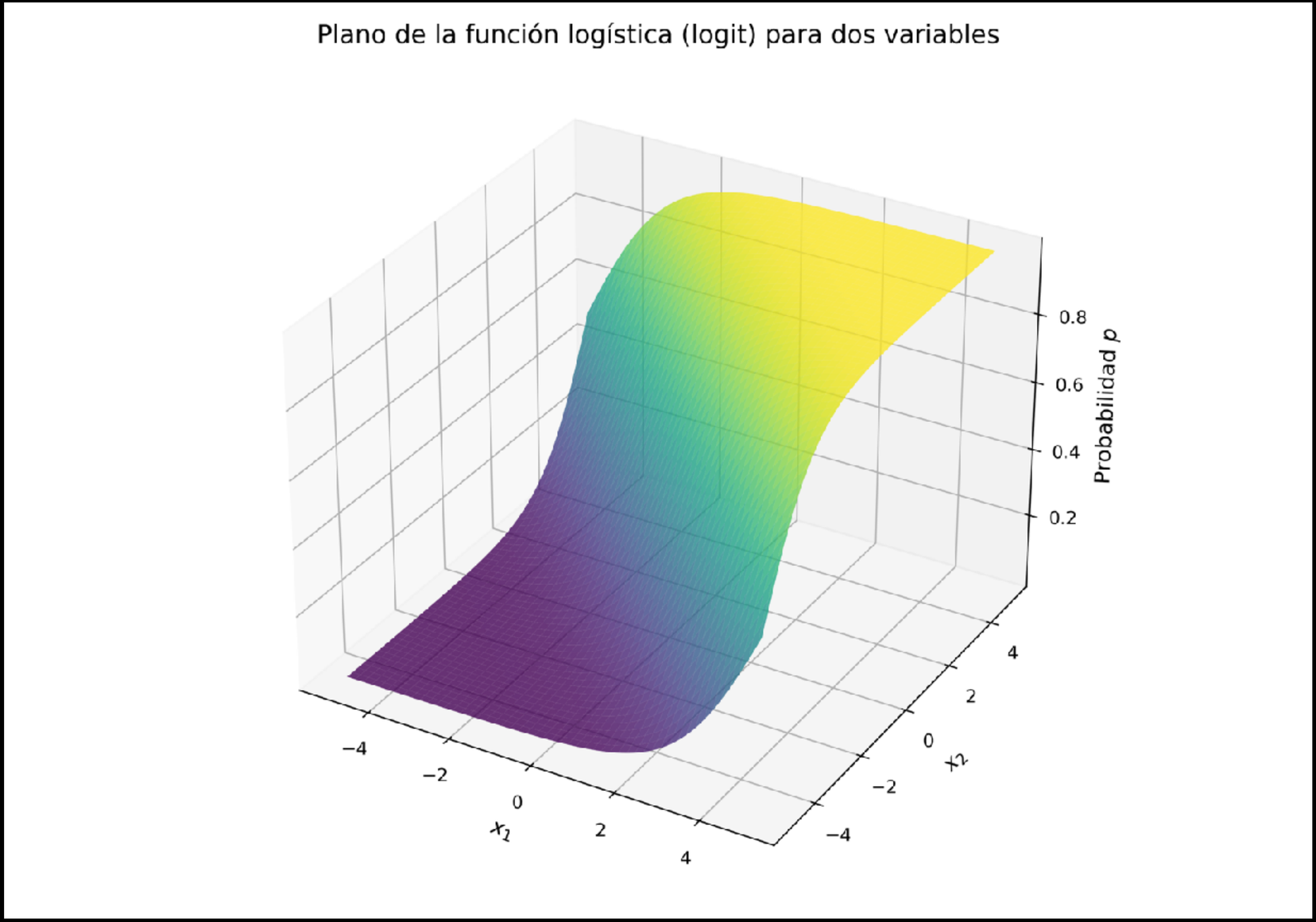
$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

$$p(x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k)}}$$

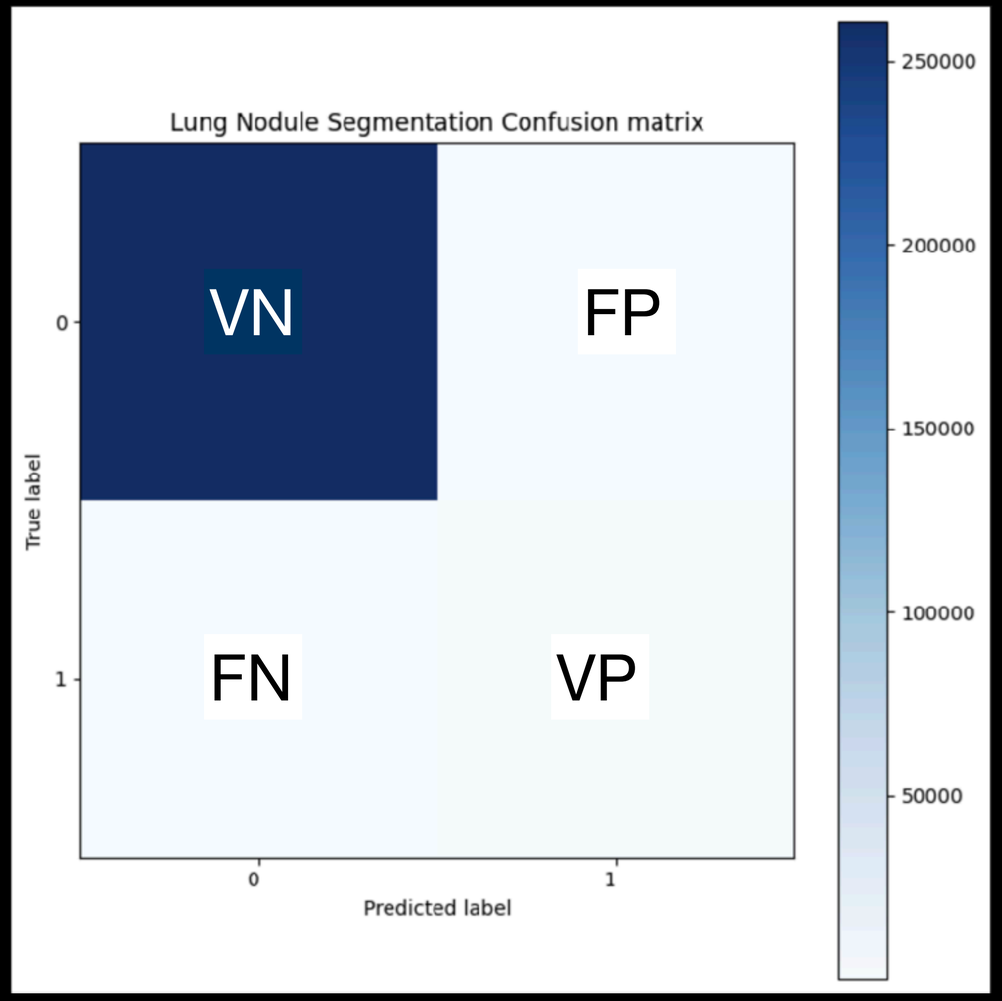
$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Clínica Santa María
Especialistas en ti

Regresión logística



Métricas



$$exactitud = \frac{VP + VN}{VP + VN + FN + FP}$$

$$\text{Sensibilidad} = \frac{\text{VP}}{\text{VP + FN}}$$
(TPR)

$$especificidad = \frac{VN}{VN + FP}$$
(TNR)

$$precision = \frac{VP}{VP + FP}$$
(PPV)

$$F1score = 2 \cdot \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$$