ESTADÍSTICA
PROBABILIDAD Y ESTADÍSTICA

Unidad 7 - *Regresión y Correlación*

Ingenierías en: Recursos Hídricos, Informática, Ambiental, Agrimensura e Inteligencia Artificial

Año 2025

Prof. Juan Pablo Taulamet

consultas: taulamet@unl.edu.ar

2025 ~ 40° Aniversario de la Creación del Consejo Interuniversitario Nacional

ESTUDIO DE ASOCIACIÓN ENTRE VARIABLES

REGRESIÓN

ENCONTRAR UN MODELO

CORRELACIÓN

EXACTITUD DEL MODELO

ANÁLISIS DE REGRESIÓN

Permite encontrar un modelo que vincula a dos o más variables, brindando un mecanismo de pronóstico.

ANÁLISIS DE CORRELACIÓN

Determina la medida del grado de exactitud de la relación entre las variables

La siguiente tabla muestra valores de precipitación y escurrimiento en una cuenca urbana.

x (mm)	y (mm)
50	60
60	60
70	80
80	70
90	80
100	100
110	100
120	130
130	110
140	120
150	150
160	140

La siguiente tabla muestra valores de precipitación y escurrimiento en una cuenca urbana.

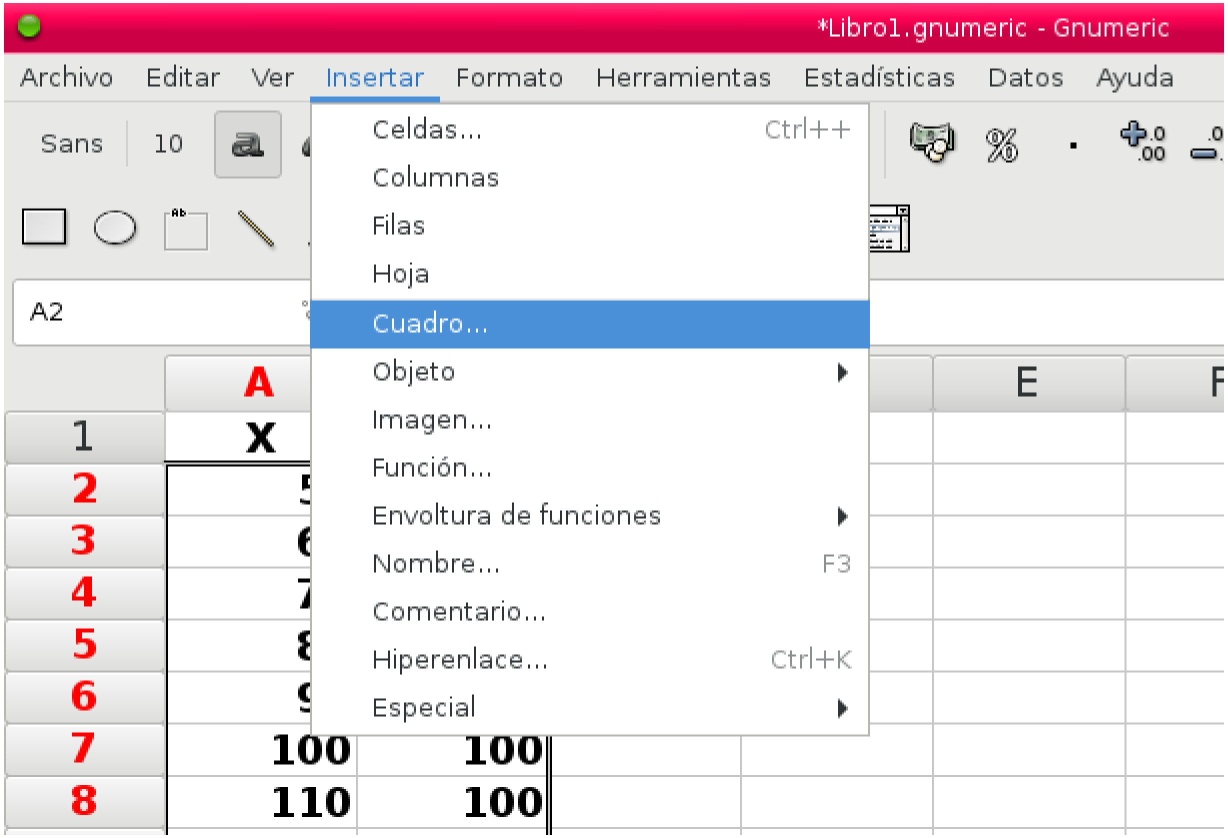
¿Existe relación entre X e Y?

x (mm)	y (mm)
50	60
60	60
70	80
80	70
90	80
100	100
110	100
120	130
130	110
140	120
150	150
160	140

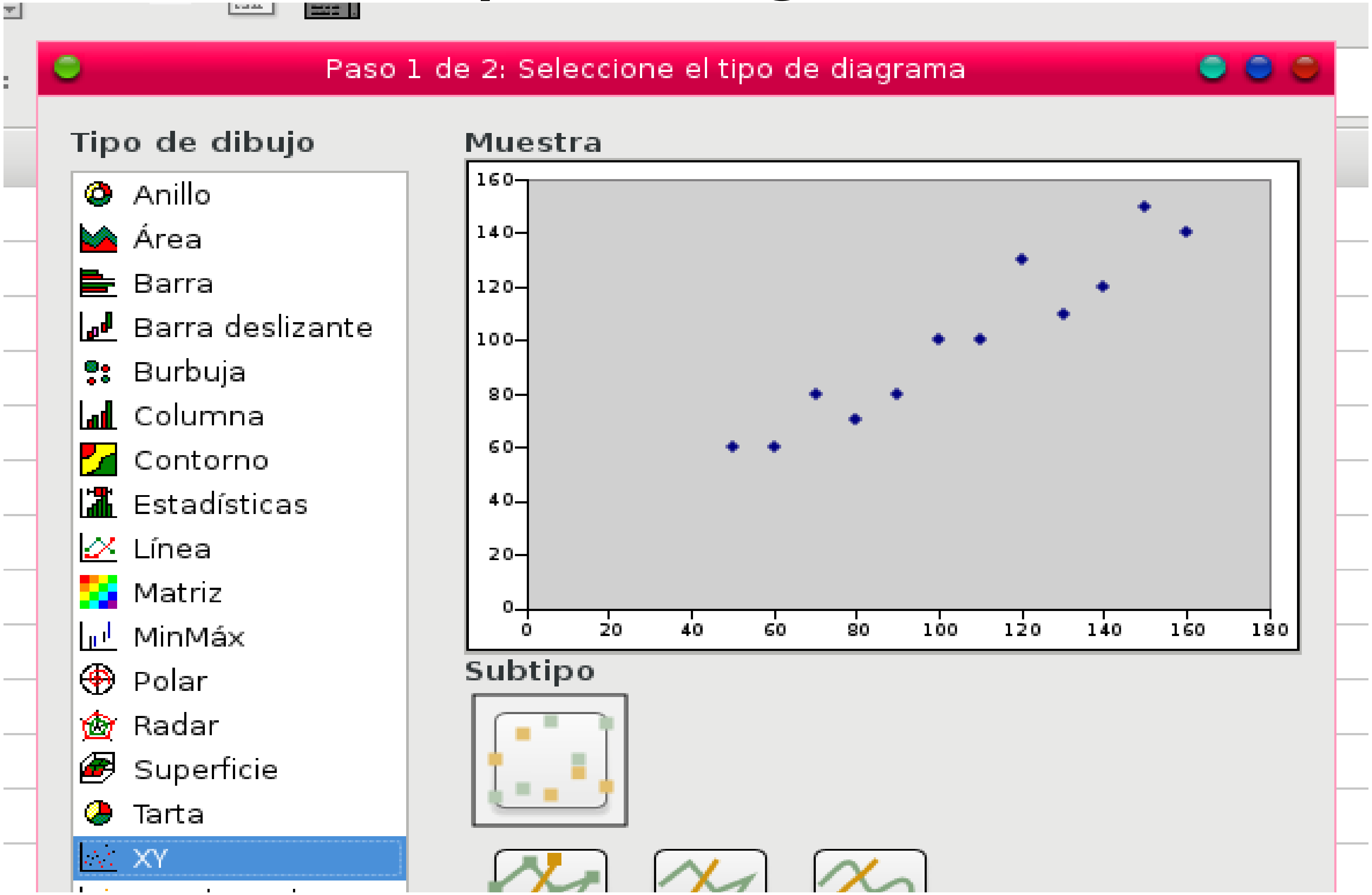
En Gnumeric

	A	B	C	D
1	X	Y		
2	50	60		
3	60	60		
4	70	80		
5	80	70		
6	90	80		
7	100	100		
8	110	100		
9	120	130		
10	130	110		
11	140	120		
12	150	150		
13	160	140		
14				

Insertamos un Dispersiograma

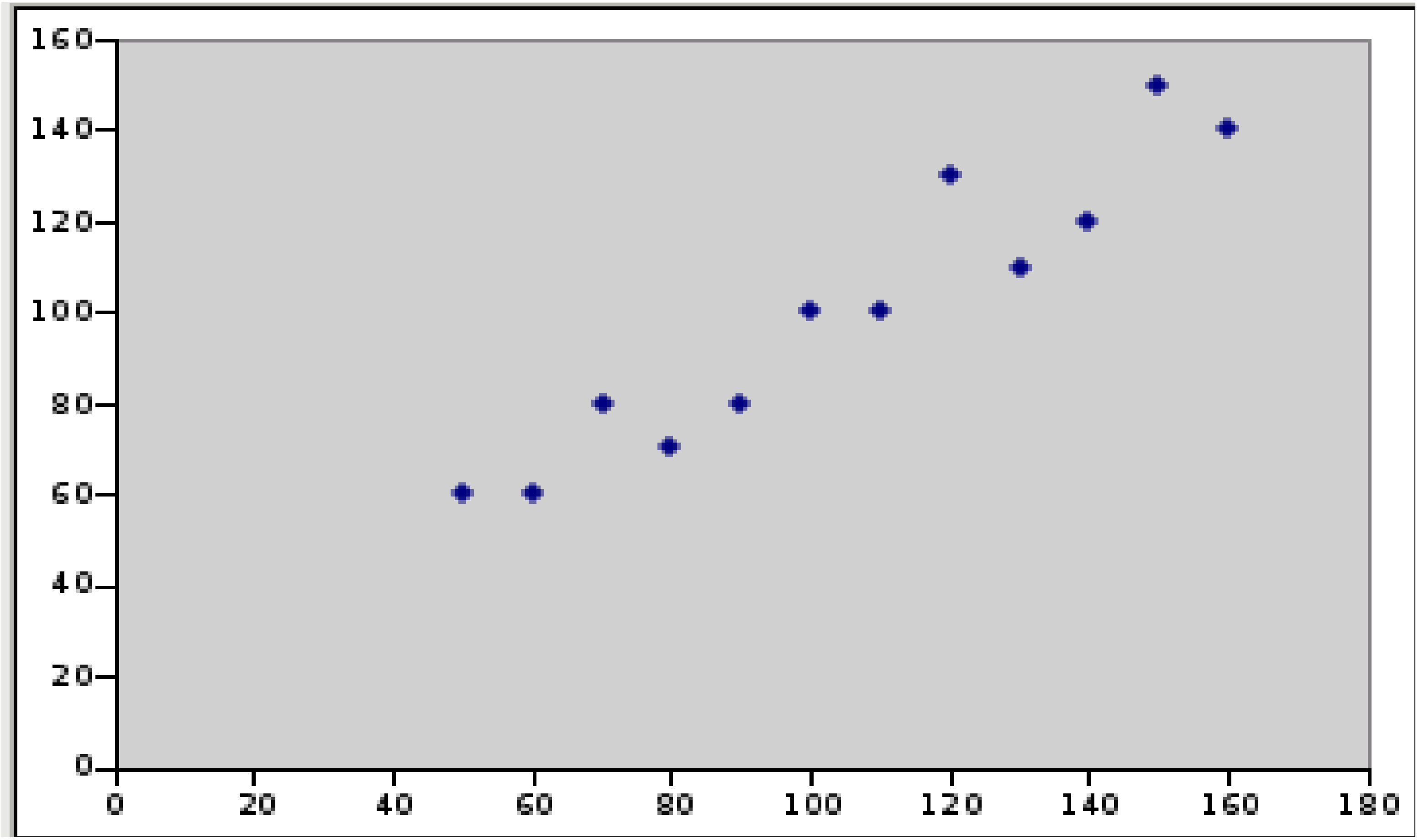


Insertamos un Dispersiograma



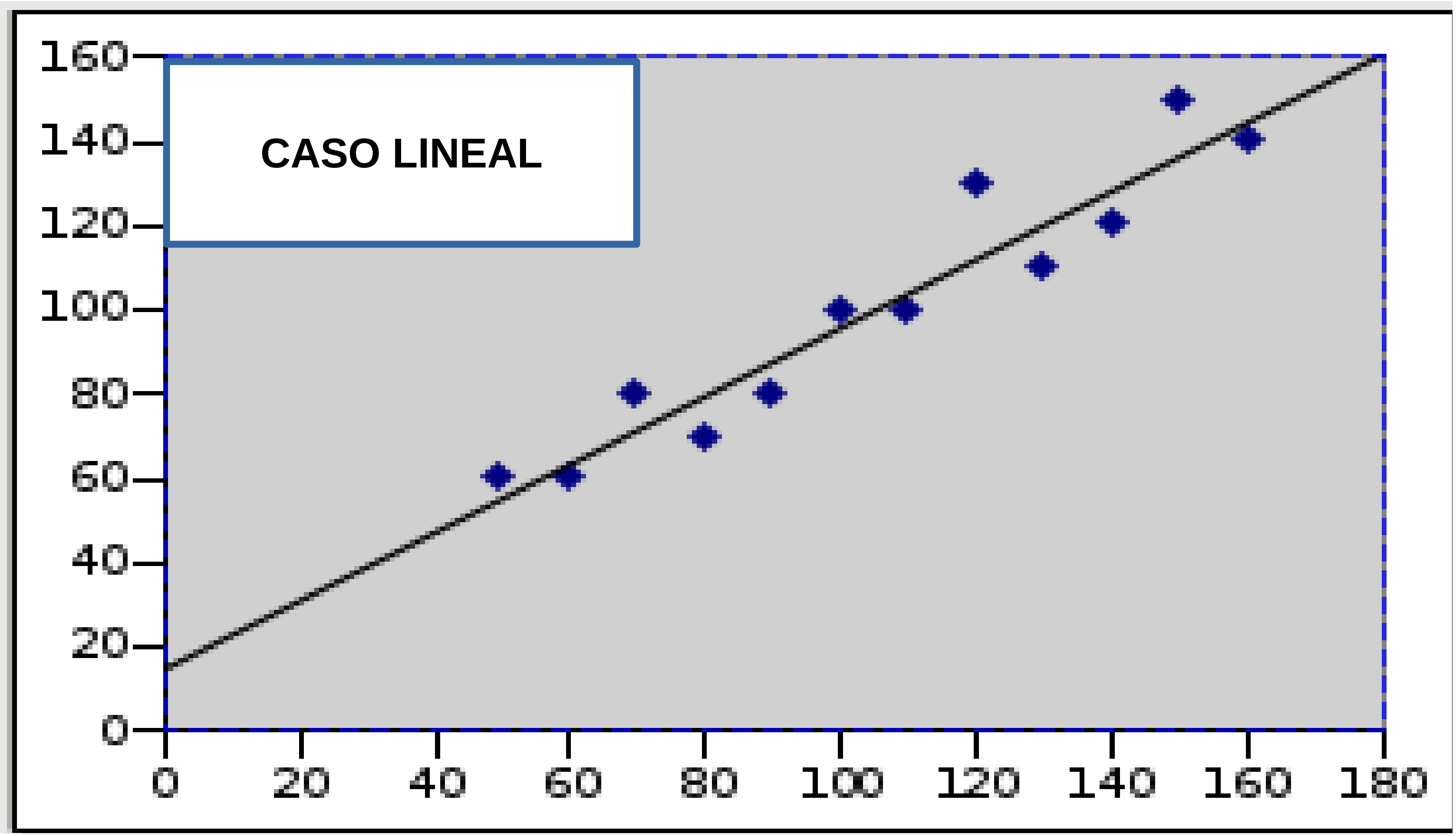
Dispersiograma

¿Existe relación entre X e Y? ¿Cómo es?



MODELO DE REGRESIÓN BIVARIADO

$$E(Y/x) = \mu_{Y/x} = \alpha + \beta X_i$$



consultas: taulamet@unl.edu.ar

MODELO DE REGRESIÓN LINEAL SIMPLE

PARÁMETROS DEL MODELO

VARIABLE EXPLICATIVA INDEPENDIENTE

VARIABLE DEPENDIENTE

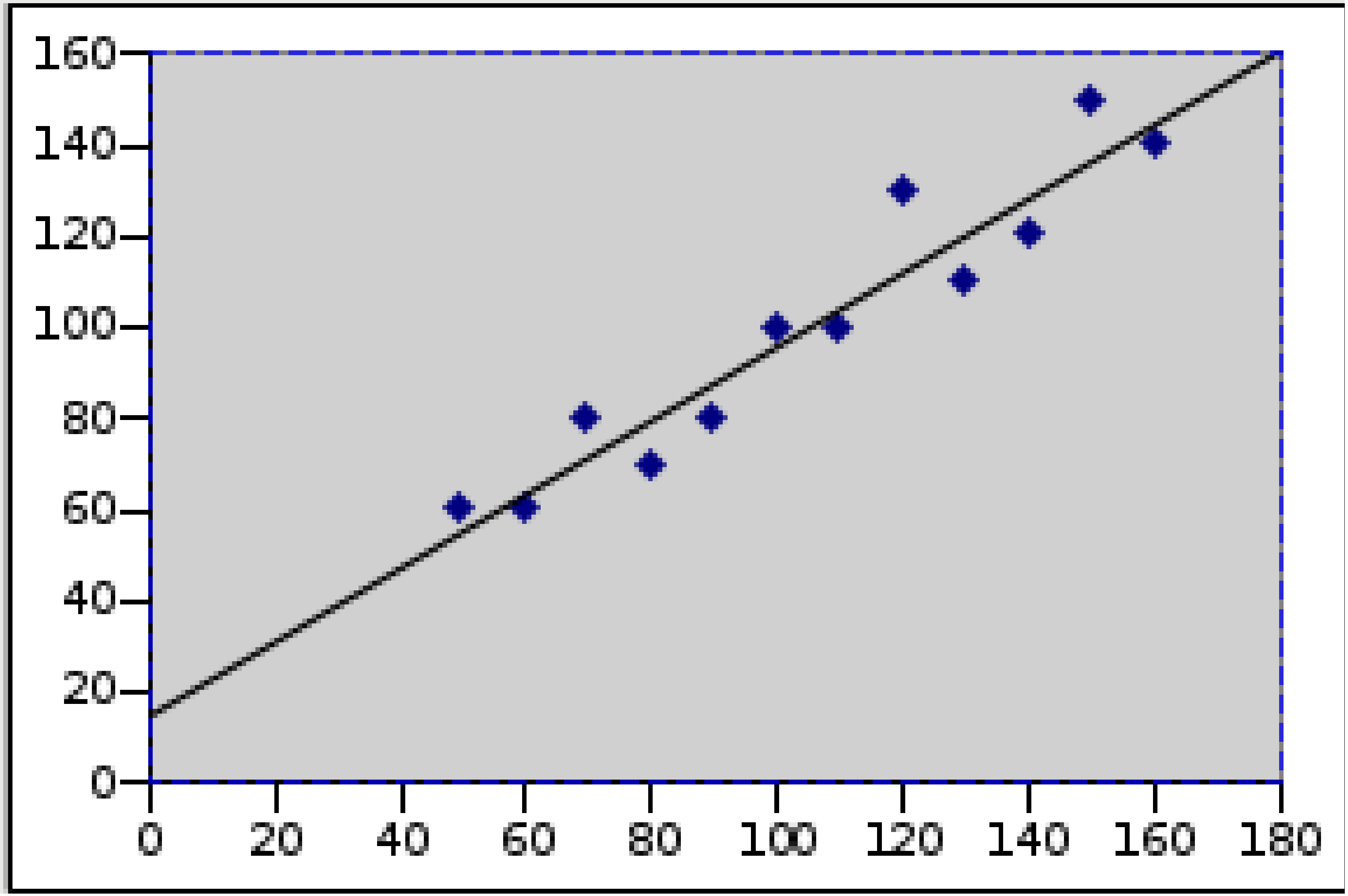
ERROR ALEATORIO

$$Y_i = \underset{I}{\underline{\alpha + \beta X_i}} + \underset{II}{\underline{\in_i}}$$

PARTE SISTEMÁTICA

PARTE ESTOCÁSTICA

PENSANDO ESTA RELACIÓN ESTADÍSTICA



consultas: taulamet@unl.edu.ar

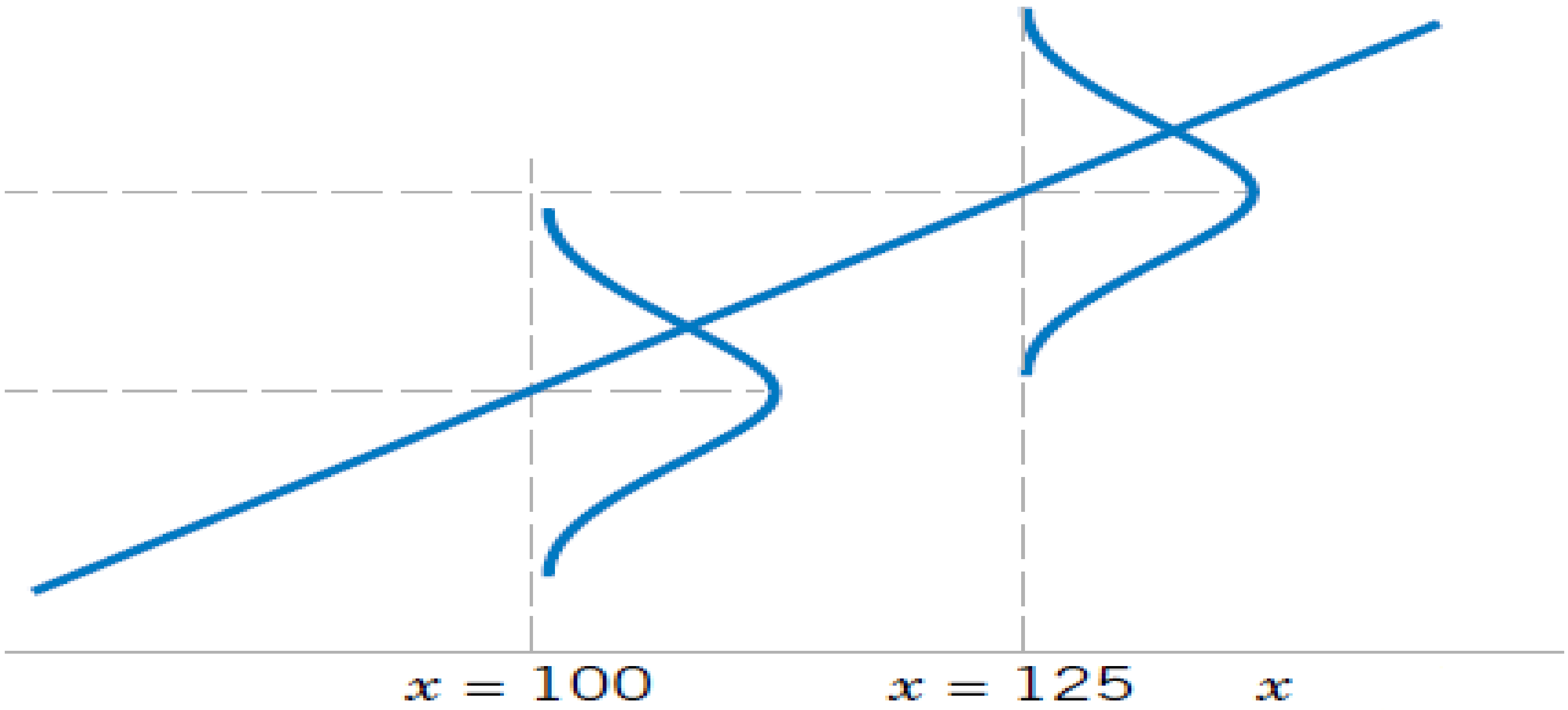
PENSANDO ESTA RELACIÓN ESTADÍSTICA

La variable dependiente tiende a variar con la variable independiente de una manera sistemática.

Existe dispersión de los puntos entorno a la curva que relaciona las variables.

PENSANDO ESTA RELACIÓN ESTADÍSTICA

Existe una distribución de probabilidades de la variable dependiente (Y) para cada valor de X. Existen grupos de valores de Y para cada valor de X que llamaremos sub-poblaciones. El modelo de regresión es el que *relaciona los valores medios* de dichas sub-poblaciones:



MODELOS DE REGRESIÓN

LINEAL
NO-LINEAL
(Polinomial, exponencial, etc.)

SIMPLE
(2 Variables)
MÚLTIPLE
(Más de 2 Variables)

MODELO DE REGRESIÓN LINEAL SIMPLE

$$Y_i = \underbrace{\alpha + \beta X_i}_{I} + \underbrace{\in_i}_{II}$$

Para conocer más del modelo supondremos que la X se fija en un valor y hallaremos la esperanza y varianza del mismo.

MODELO DE REGRESIÓN LINEAL SIMPLE

Supongamos que la media y la varianza de $\in$ son 0 y σ^2

$$E(Y/x) = E(\alpha + \beta x + \varepsilon) = \alpha + \beta x + E(\varepsilon) = \alpha + \beta x$$

$$Var(Y/x) = Var(\alpha + \beta x + \varepsilon) = Var(\alpha + \beta x) + Var(\varepsilon) = 0 + \sigma^2 = \sigma^2$$

ESTIMACIÓN POR MÍNIMOS CUADRADOS

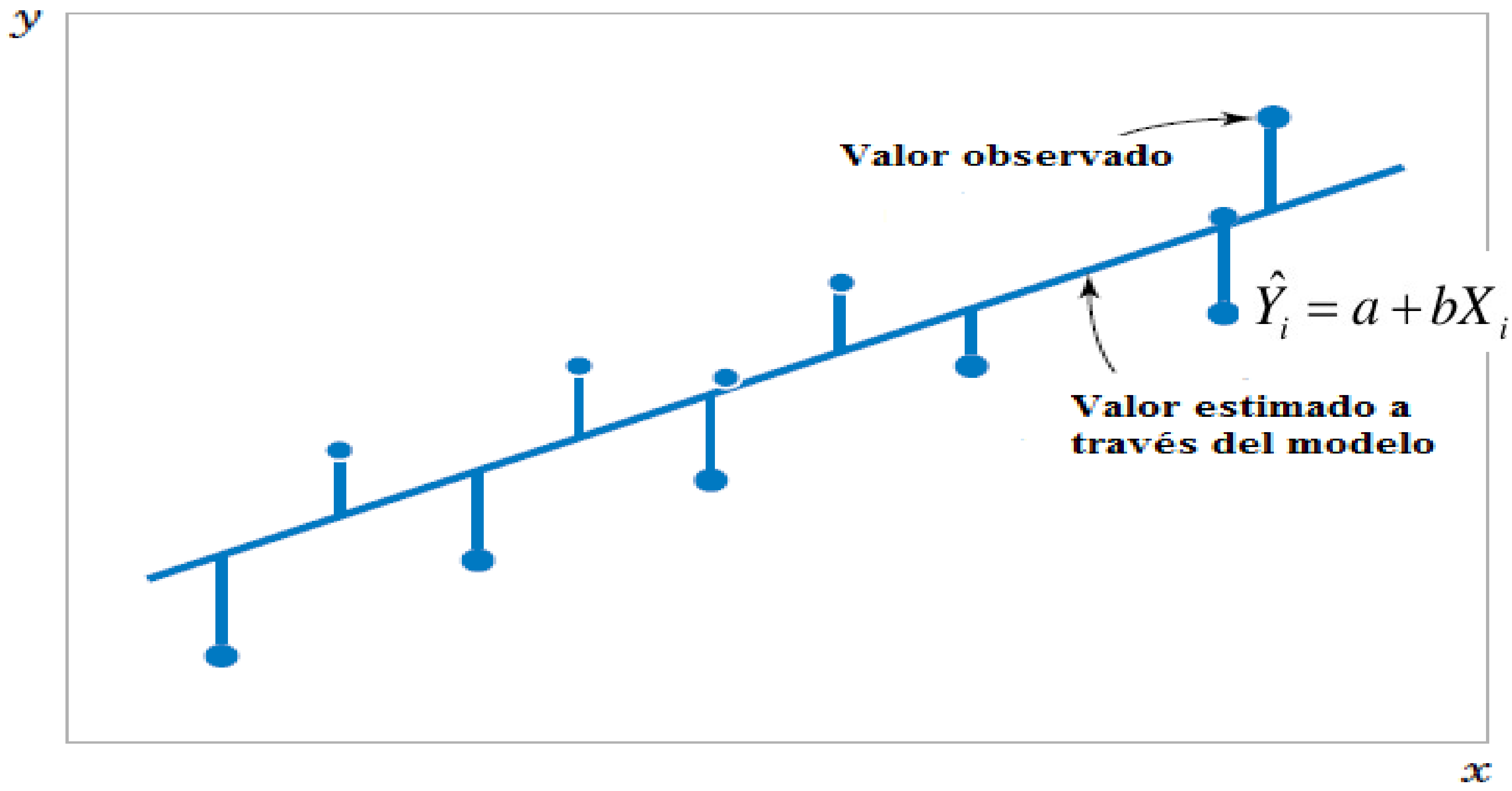
SUPUESTOS:

**La VA del error es independiente.*

**La distribución del error es normal*

**X es un conjunto de números fijos*

ESTIMACIÓN POR MÍNIMOS CUADRADOS



$$S = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2 \qquad \frac{\partial S}{\partial \alpha} = 0 \quad \text{y} \quad \frac{\partial S}{\partial \beta} = 0$$

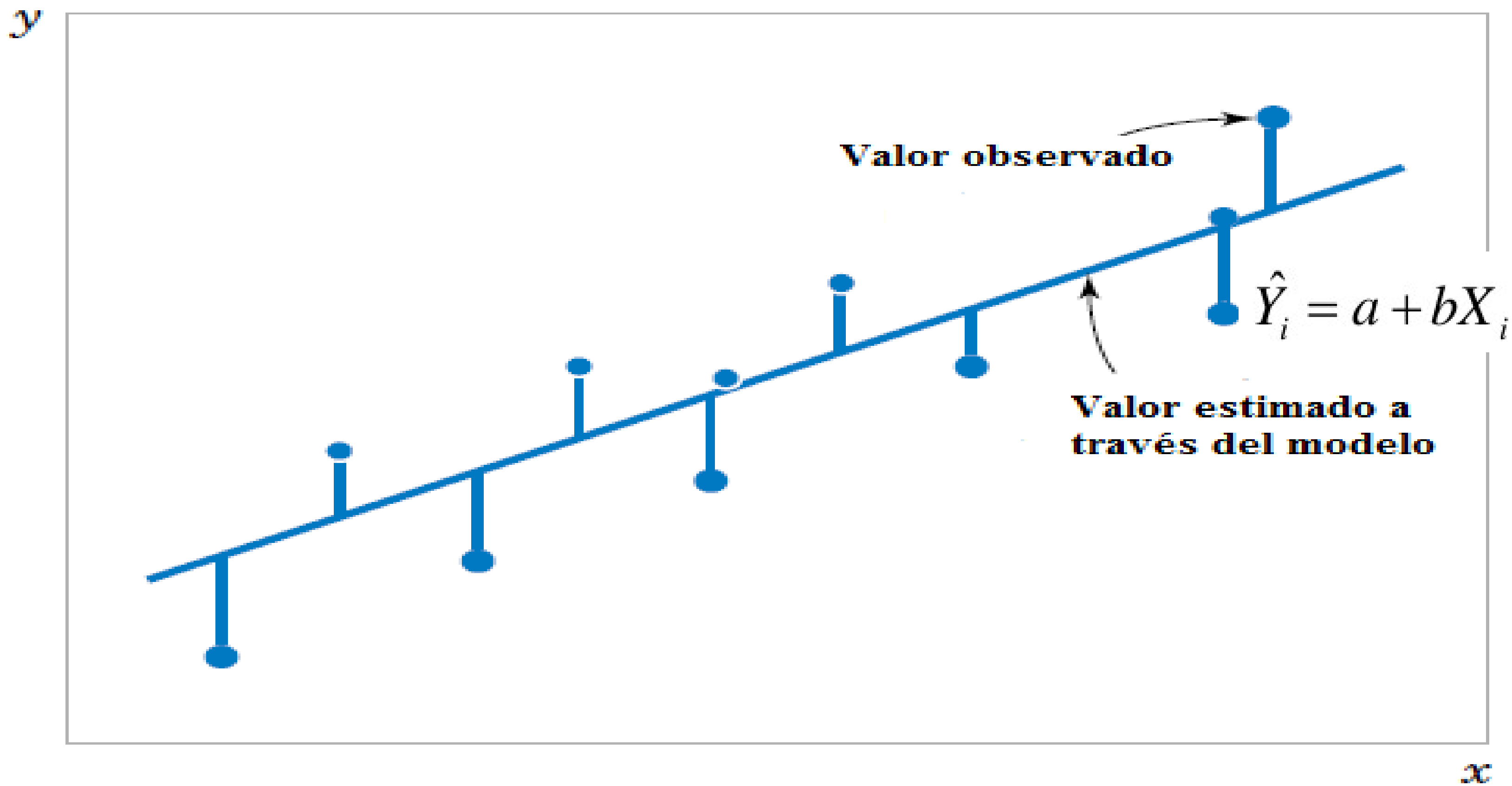
ESTIMACIÓN POR MÍNIMOS CUADRADOS

$$\hat{\beta} = b = \frac{m_{1,1}}{S_x^2} = \frac{\mathrm{cov}}{S_x^2}$$

$$\hat{\alpha} = a = \overline{Y} - \hat{\beta}\overline{X}$$

Este método minimiza las distancias verticales

VARIANZA DE LA REGRESIÓN



$$S^2_{y/x} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$$

ANÁLISIS DE CORRELACIÓN

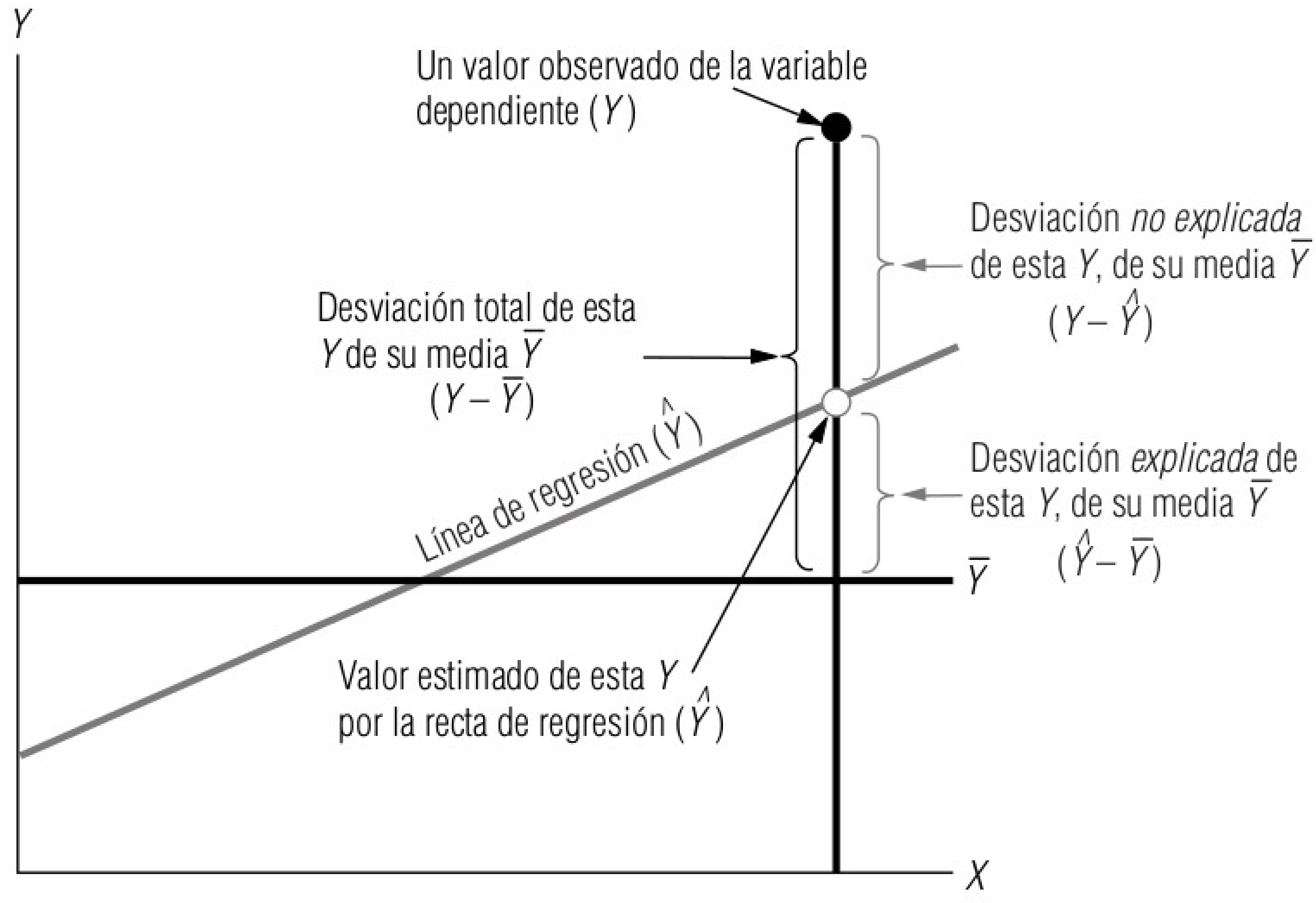
CASO LINEAL
Se utiliza el coeficiente de correlación r:

$$r = \hat{\rho} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_i - \bar{y})^2}} = \frac{m_{1,1}}{S_x S_y}$$

$$-1 \leq r \leq 1$$

Cuanto más cerca se encuentre del 0 menor relación lineal habrá entre la X y la Y. En la medida que se acerque a 1 o a -1, expresará un mayor grado de asociación lineal ya sea relación directa o inversamente proporcional.

PARTICIÓN DE LOS CUADRADOS



ANÁLISIS DE CORRELACIÓN

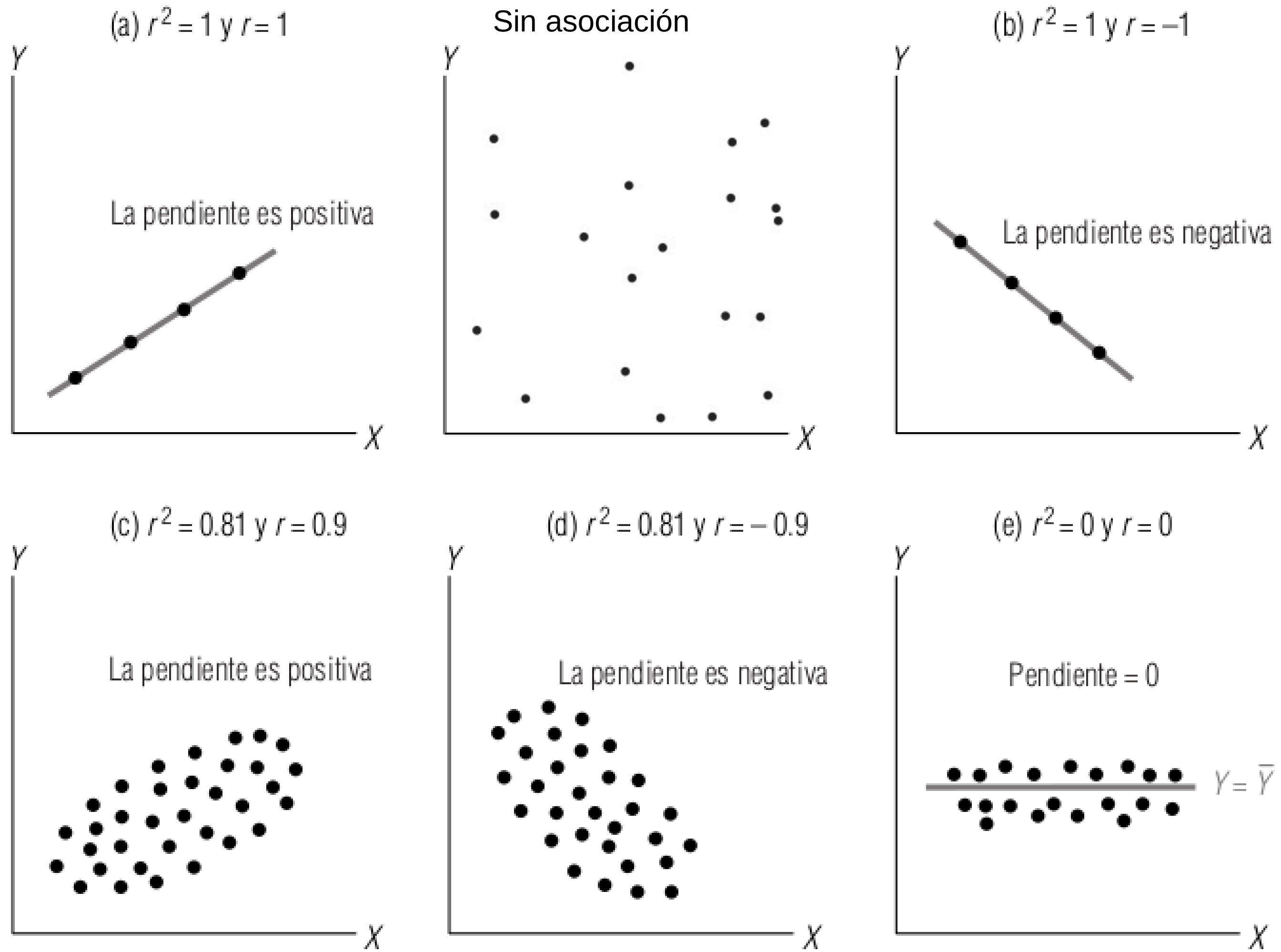
CASO NO LINEAL
Se utiliza el coeficiente de determinación r²:

$$r^2 = 1 - \frac{\Sigma(Y - \hat{Y})^2}{\Sigma(Y - \bar{Y})^2}$$

$$-0 \leq r^2 \leq 1$$

Cuanto más se acerque a 1, mayor será la proporción de suma de cuadrados explicados por la regresión en relación a la suma de cuadrados totales.

EJEMPLOS DE DISPERSIOGRAMAS



consultas: taulamet@unl.edu.ar