Regresión Lineal
Parte I

Profesor: Dr. Alejandro S. M. Santa Cruz
JTP: Dr. Juan Ignacio Manassaldi
Auxiliar: Srta. Amalia Rueda

Regresión Lineal

¿Qué es la regresión? Dados n datos: $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$

¿Cuál es la curva $y = f(x)$ que mejor ajusta a los datos?

El residuo en cada punto es: $E_i = y_i - f(x_i)$

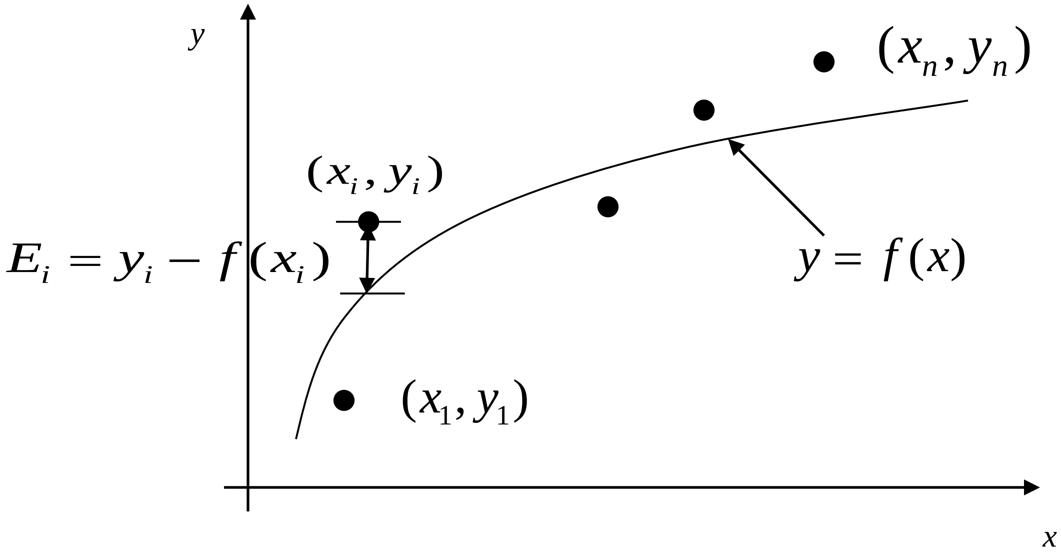


Fig. Modelo básico para regresión.

Regresión Lineal – Criterio#1

Dados n datos: $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ Ajustar la recta $y = a_0 + a_1 x$ a los datos

Minimizando $\sum_{i=1}^{n} E_i$ ¿Funciona como criterio de ajuste?

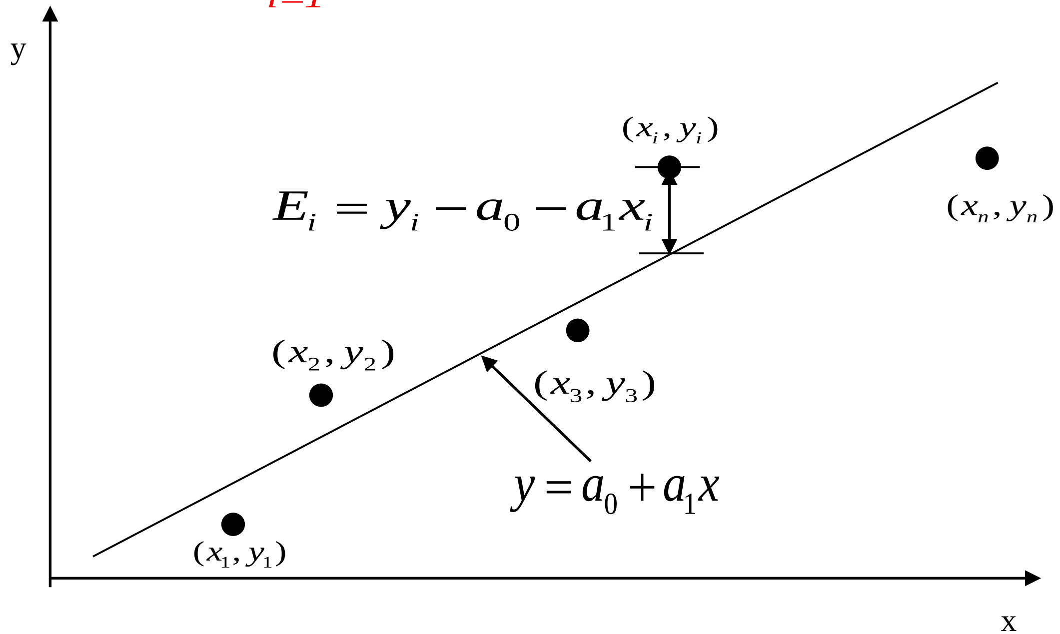


Fig. Regresión lineal de datos *y* vs *x* mostrando el residuo en un punto típico X_i.

Ejemplo para el Criterio#1

Ejemplo: Dados los datos (2,4), (3,6), (2,6) y (3,8), ajustarlos a una línea recta utilizando el Criterio#1

Minimizar $\sum_{i=1}^{n} E_i$

Tabla. Puntos datos

x	y
2.0	4.0
3.0	6.0
2.0	6.0
3.0	8.0

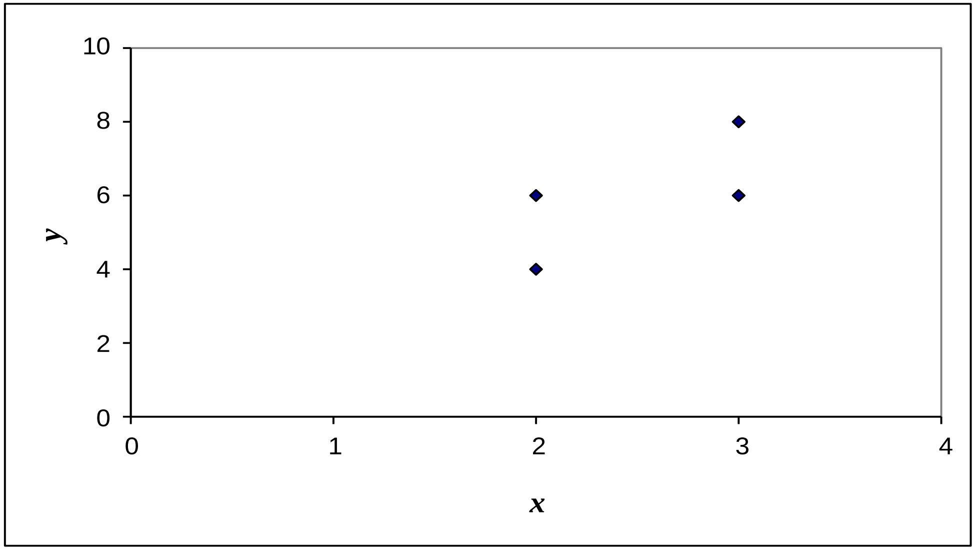


Fig. Datos *y* vs *x*.

Regresión Lineal – Criterio#1

Usando $y = 4x - 4$ como curva de regresión.

Tabla. Residuos en cada punto para el modelo de regresión $y = 4x - 4$

x	y	$y_{predicted}$	$E = y - y_{predicted}$
2.0	4.0	4.0	0.0
3.0	6.0	8.0	-2.0
2.0	6.0	4.0	2.0
3.0	8.0	8.0	0.0
			$\sum_{i=1}^{4} E_i = 0$

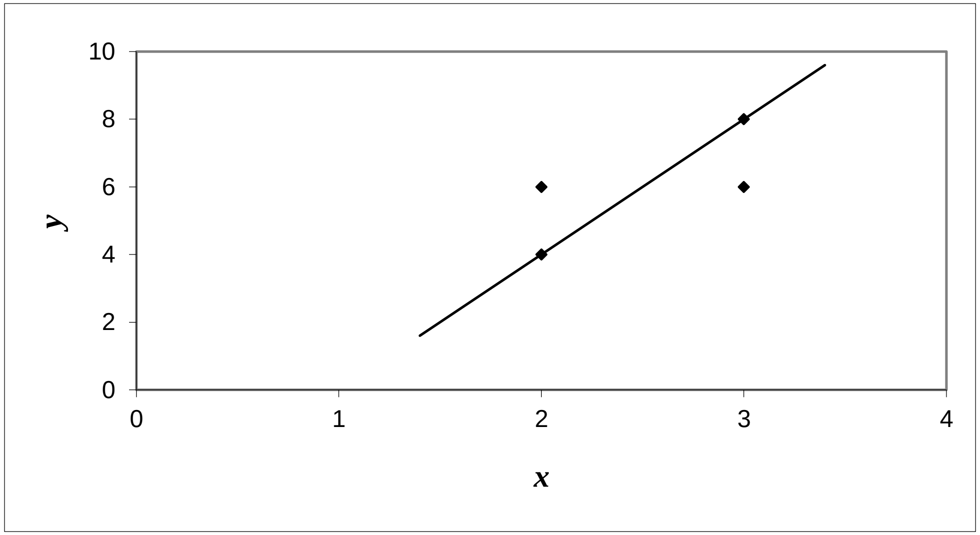


Fig. Curva de regresión $y = 4x - 4$ y datos y vs x.

Regresión Lineal – Criterio#1

Usando $y = 6$ como curva de regresión.

Tabla. Residuos en cada punto para el modelo de regresión $y = 6$

x	y	$y_{predicted}$	$E = y - y_{predicted}$
2.0	4.0	6.0	-2.0
3.0	6.0	6.0	0.0
2.0	6.0	6.0	0.0
3.0	8.0	6.0	2.0
			$\sum_{i=1}^{4} E_i = 0$

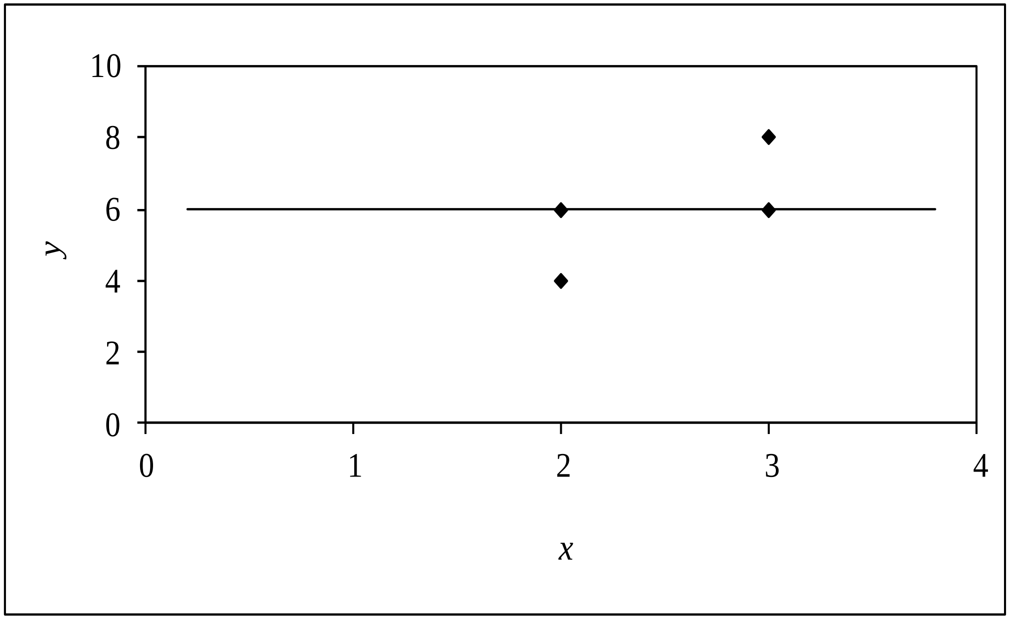


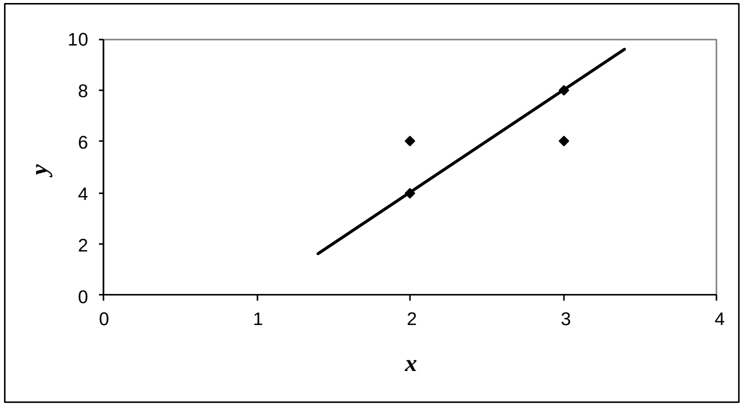
Fig. Curva de regresión $y = 6$ y datos y vs x.

Regresión Lineal – Criterio#1

$$\sum_{i=1}^{4} E_i = 0$$ Para ambos modelos de regresión $y = 4x-4$ e $y = 6$

La suma de los residuos está minimizada, en ambos casos es cero. Pero el modelo de regresión no es único.

De aquí que el criterio de minimizar la suma de los residuos es un criterio malo.



Regresión Lineal – Criterio#2

Dados n datos: $(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_n, y_n)$ Ajustar la recta $y = a_0 + a_1 x$ a los datos.

Minimizando $\sum_{i=1}^{n} |E_i|$ ¿Funciona mejor como criterio de ajuste?

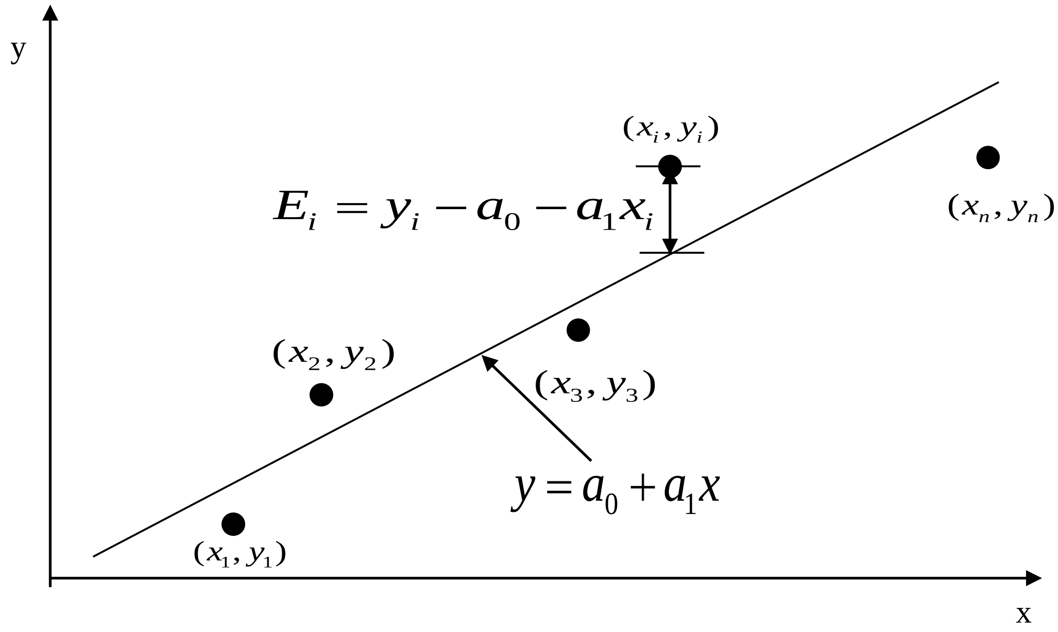


Fig. Regresión lineal de datos *y* vs *x* mostrando el residuo en un punto típico x_i.

Ejemplo para el Criterio#2

Ejemplo: Dados los datos (2,4), (3,6), (2,6) y (3,8), ajustarlos a una línea recta utilizando el Criterio#2

Minimizar $\sum_{i=1}^{n} |E_i|$

Tabla. Puntos datos

x	y
2.0	4.0
3.0	6.0
2.0	6.0
3.0	8.0

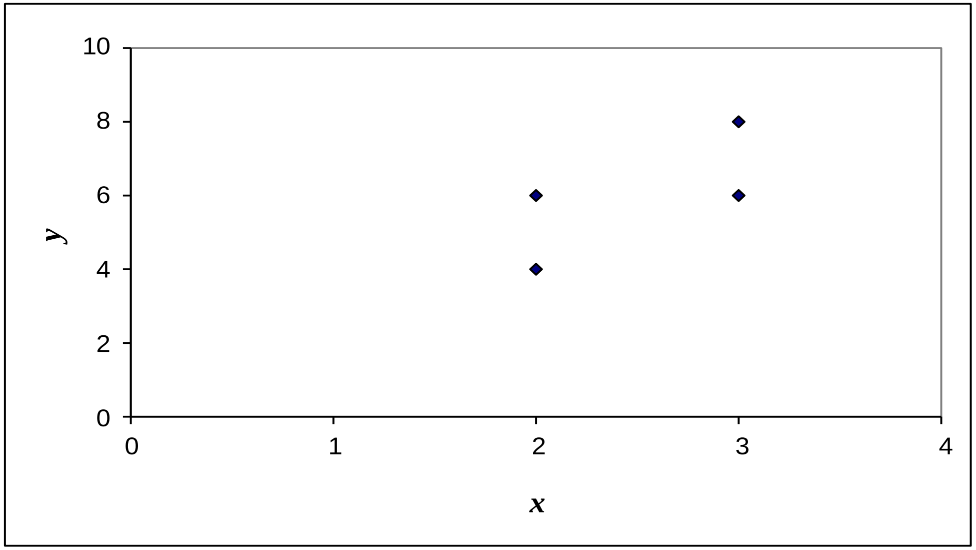


Fig. Datos *y* vs *x*.

Ejemplo para el Criterio#2

Usando $y = 4x - 4$ como curva de regresión.

Tabla. Residuos en cada punto para el modelo de regresión $y = 4x - 4$

x	y	$y_{predicted}$	$E = y - y_{predicted}$
2.0	4.0	4.0	0.0
3.0	6.0	8.0	-2.0
2.0	6.0	4.0	2.0
3.0	8.0	8.0	0.0
			$\sum_{i=1}^{4} \lvert E_i \rvert = 4$

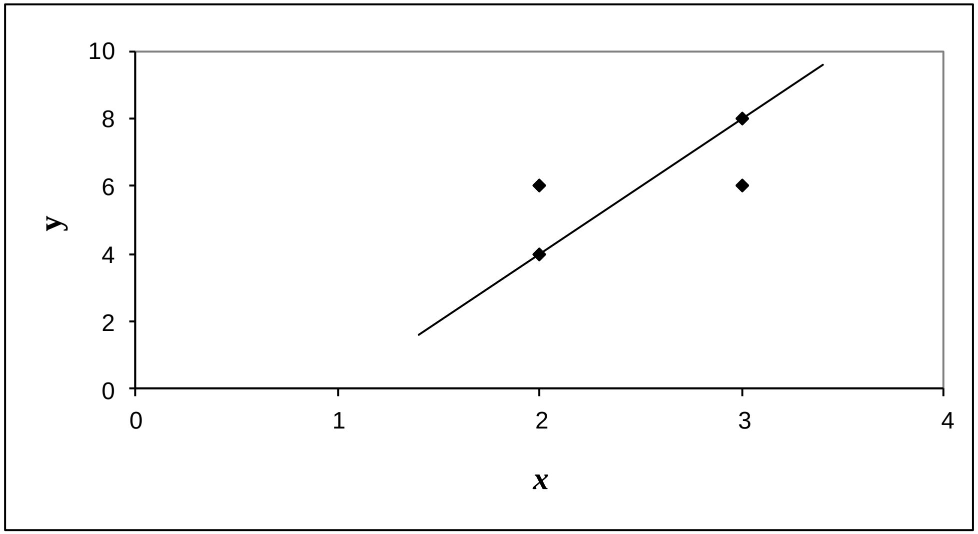


Fig. Curva de regresión $y = 4x - 4$ y datos y vs. x

Regresión Lineal – Criterio#2

Usando $y = 6$ como curva de regresión.

Tabla. Residuos en cada punto para el modelo de regresión $y = 6$

x	y	$y_{predicted}$	$E = y - y_{predicted}$
2.0	4.0	6.0	-2.0
3.0	6.0	6.0	0.0
2.0	6.0	6.0	0.0
3.0	8.0	6.0	2.0
			$\sum_{i=1}^{4} \lvert E_i \rvert = 4$

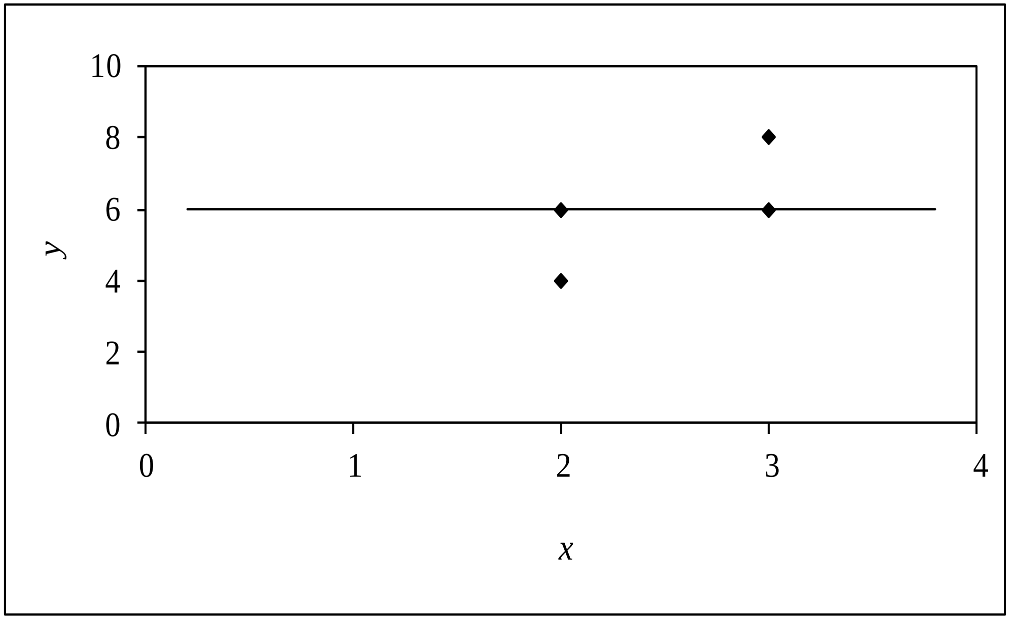


Fig. Curva de regresión $y = 6$ y datos y vs x.

Regresión Lineal – Criterio#2

$\sum_{i=1}^{4} |E_i| = 4$ Para ambos modelos de regresión $y = 4x - 4$ e $y = 6$.

La suma de los residuos absolutos se hizo tan pequeña como fue posible, esto es, 4, pero el modelo de regresión no es único.

Por consiguiente, el criterio de minimizar la suma de los valores absolutos de los residuos también es un criterio malo.

Criterio de Mínimos Cuadrados

El criterio de mínimos cuadrados minimiza la suma de los cuadrados de los residuos en el modelo y también produce una única línea recta.

$$S_r = \sum_{i=1}^{n} E_i^{\,2} = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i\right)^2$$

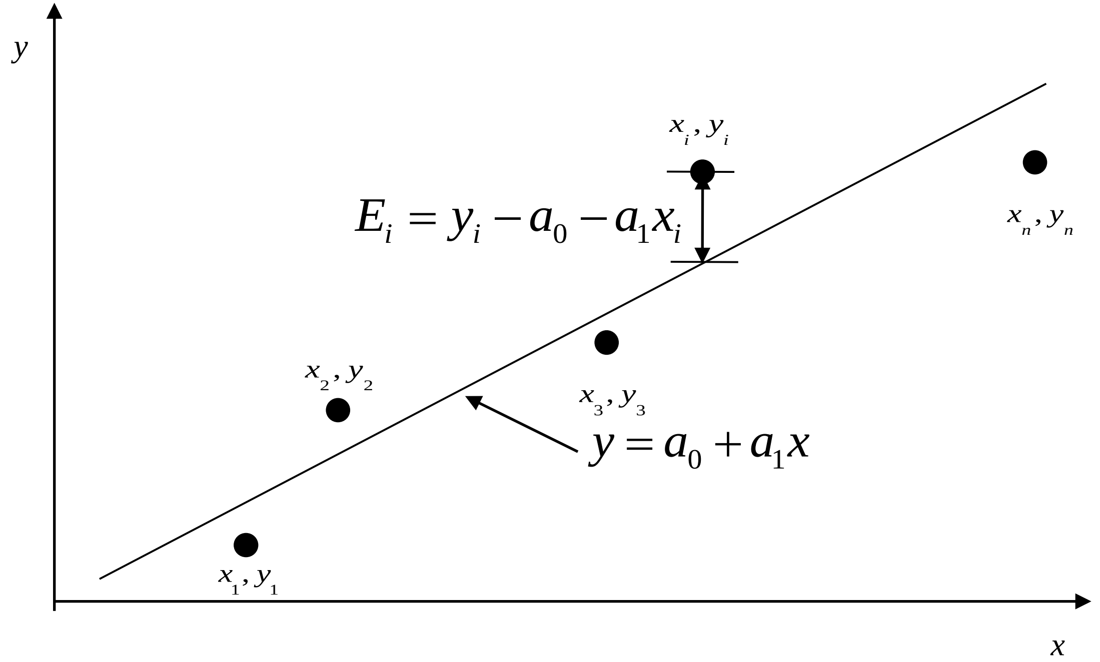


Fig. Regresión lineal de los datos *y* vs. *x* data mostrando el residuo en un típico punto x_i.

Determinación de las Constantes del Modelo Lineal

Minimizar la suma de los cuadrado de los residuos: $S_r = \sum_{i=1}^{n} E_i^{\,2} = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i\right)^2$

Para encontrar a_0 y a_1 minimizamos S_r con respecto a_0 y a_1: .

$$\frac{\partial S_r}{\partial a_0} = 2\sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)\left(-1\right) = 0$$

$$\frac{\partial S_r}{\partial a_1} = 2\sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)\left(-x_i\right) = 0$$

Generando el siguiente SEAL:

$$\sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i = \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} a_0 x_i + \sum_{i=1}^{n} a_1 x_i^{\,2} = \sum_{i=1}^{n} y_i x_i$$

Determinación de las Constantes del Modelo Lineal

Resolviendo directamente para a_0 y a_1 se obtiene:

$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

y

$$a_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Ejemplo 1

El torque, T, que se necesita para doblar un ángulo θ un resorte de torsión de una trampera se expresa como:

$$T = k_1 + k_2\theta$$

Tabla: Torque vs. ángulo para un resorte de torsión.

Ángulo, θ	Torque, T
Radianes	*N-m*
0.698132	0.188224
0.959931	0.209138
1.134464	0.230052
1.570796	0.250965
1.919862	0.313707

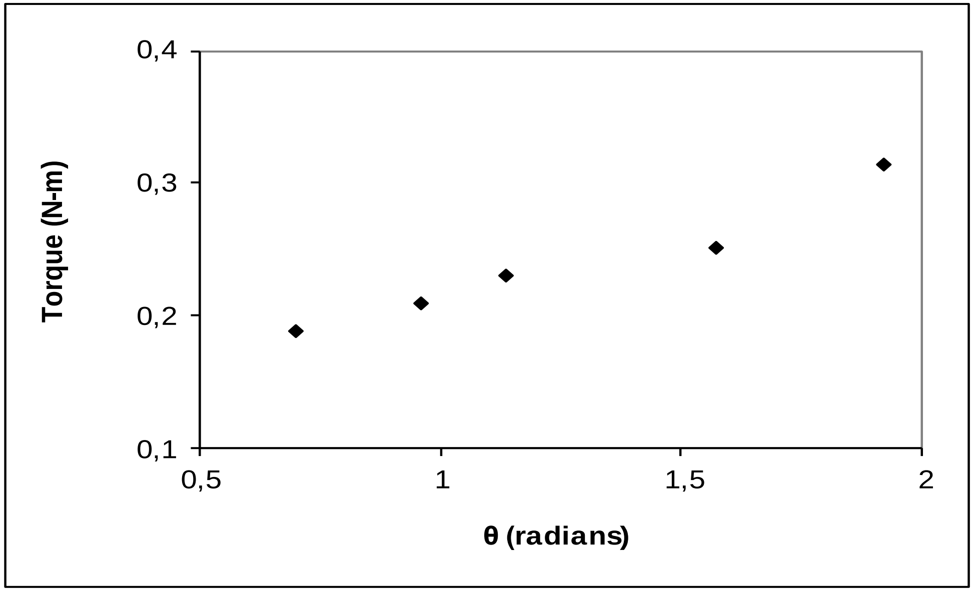


Fig. Datos de torques vs. ángulos de torsión

Ejemplo 1 (cont.)

La siguiente Tabla muestra las sumatorias que se necesitan para los cálculos de las constantes en el modelo de regresión:

Tabla. Tabulación de los datos para el cálculo de sumatorias importantes.

	θ	T	θ^2	$T\theta$
	Radianes	*N-m*	*Radianes*^{2}	*N-m-Radianes*
	0.698132	0.188224	0.487388	0.131405
	0.959931	0.209138	0.921468	0.200758
	1.134464	0.230052	1.2870	0.260986
	1.570796	0.250965	2.4674	0.394215
	1.919862	0.313707	3.6859	0.602274
$\sum_{i=1}^{5} =$	**6.2831**	**1.1921**	**8.8491**	**1.5896**

Utilizando las ecuaciones descriptas para a_0 y a_1 con $n = 5$:

$$k_2 = \frac{n\sum_{i=1}^{5}\theta_i T_i - \sum_{i=1}^{5}\theta_i \sum_{i=1}^{5} T_i}{n\sum_{i=1}^{5}\theta_i^2 - \left(\sum_{i=1}^{5}\theta_i\right)^2} =$$

$$= \frac{5(1.5896)-(6.2831)(1.1921)}{5(8.8491)-(6.2831)^2}$$

$$= 9.6091\times 10^{-2}$$

Ejemplo 1 (cont.)

Utilizar el torque promedio y el ángulo promedio para calcular las constantes del modelo:

$$\bar{T} = \frac{\sum_{i=1}^{5} T_i}{n} = \frac{1.1921}{5} = 2.3842 \times 10^{-1}$$

$$\bar{\theta} = \frac{\sum_{i=1}^{5} \theta_i}{n} = \frac{6.2831}{5} = 1.2566$$

Utilizando,

$$\begin{aligned} k_1 &= \bar{T} - k_2\,\bar{\theta} \\ &= 2.3842 \times 10^{-1} - \left(9.6091 \times 10^{-2}\right)\left(1.2566\right) \\ &= 1.1767 \times 10^{-1} \end{aligned}$$

Resultados del Ejemplo 1

Utilizando regresión lineal, se encuentra una línea de tendencia a partir de los datos experimentales:

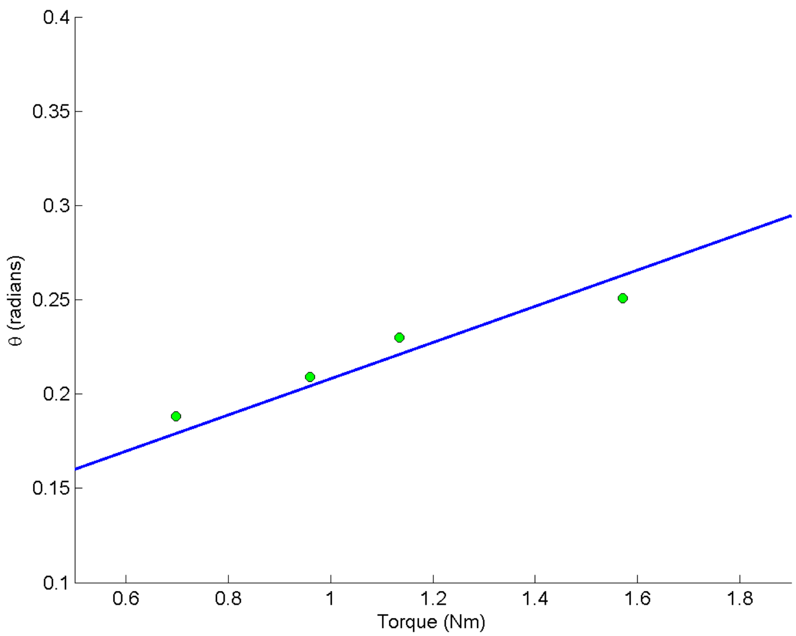


Fig. Regresión lineal del Torque vs. Ángulo

¿Puede determinar la energía en el resorte si éste es girado desde 0 a 180 grados?

Regresión Lineal – Caso Especial

Dados n datos:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

Determinar el mejor ajuste del modelo:

$$y = a_1 x$$

a los datos.

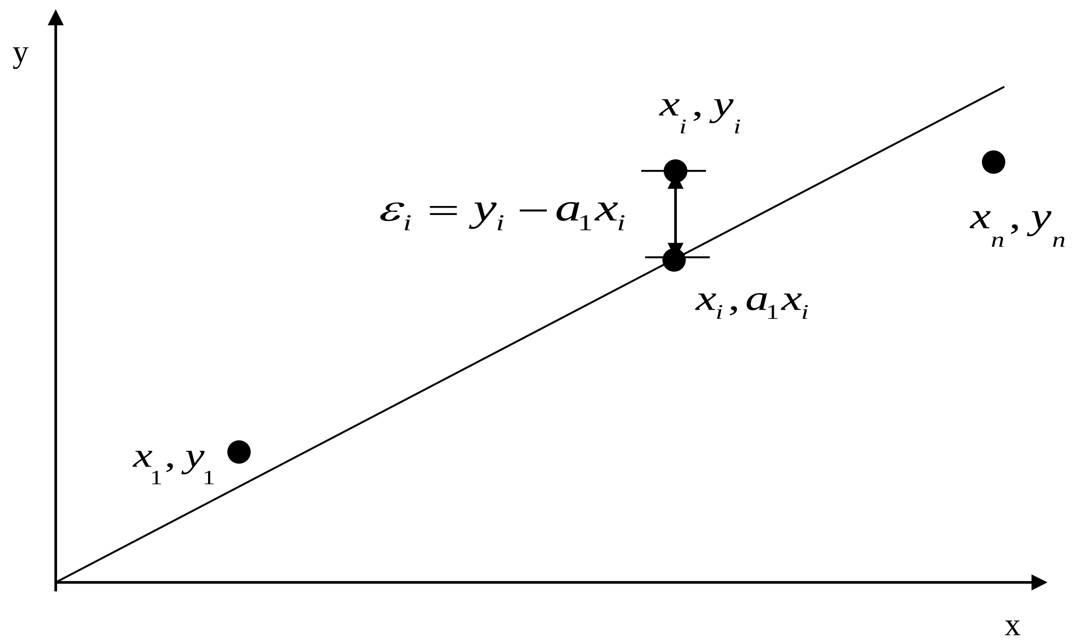
Regresión Lineal – Caso Especial (cont.)

$$y = a_1 x$$

$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

¿Es esto correcto?

Regresión Lineal – Caso Especial (cont.)



Regresión Lineal – Caso Especial (cont.)

El residuo en cada punto dato es:

$$\varepsilon_i = y_i - a_1 x_i$$

La suma de los cuadrados de los residuos es:

$$S_r = \sum_{i=1}^{n} \varepsilon_i^{\,2}$$

$$= \sum_{i=1}^{n} \left(y_i - a_1 x_i\right)^2$$

Regresión Lineal – Caso Especial (cont.)

Derivando con respecto a a_1:

$$\frac{dS_r}{da_1} = \sum_{i=1}^{n} 2\left(y_i - a_1 x_i\right)\left(-x_i\right)$$

$$= \sum_{i=1}^{n} \left(-2 y_i x_i + 2 a_1 x_i^{\,2}\right)$$

$$\frac{dS_r}{da_1} = 0$$

Se obtiene:

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Regresión Lineal – Caso Especial (cont.)

¿Este valor de a_1 corresponde a un mínimo local o a un máximo local?

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

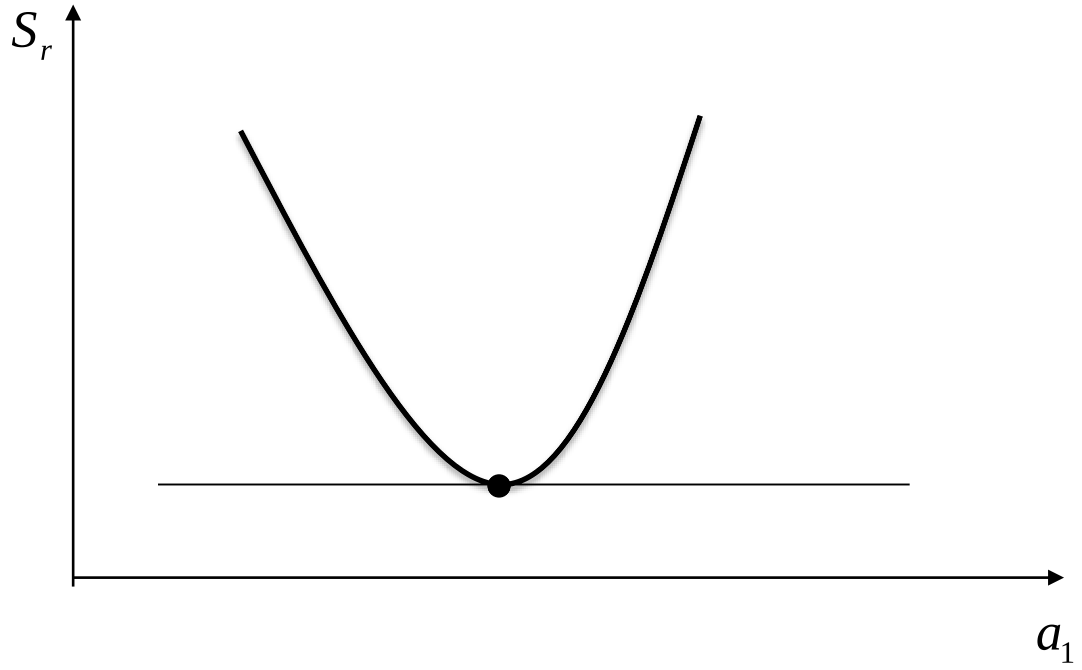
$$\frac{dS_r}{da_1} = \sum_{i=1}^{n} \left(-2 y_i x_i + 2 a_1 x_i^{\,2}\right)$$

$$\frac{d^2 S_r}{da_1^{\,2}} = \sum_{i=1}^{n} 2 x_i^{\,2} > 0$$

Corresponde a un mínimo local.

Regresión Lineal – Caso Especial (cont.)

¿Es el mínimo local de S_r un mínimo absoluto de S_r?



Ejemplo 2

Para determinar el módulo longitudinal de deformación de un compuesto, se recolectaron los siguientes datos. Determinar el módulo longitudinal E utilizando el modelo de regresión lineal $\sigma = E\,\varepsilon$ y el método de los Mínimos Cuadrados.

Tabla. Datos de tensión vs. deformación

Deformación	Tensión
(%)	*(MPa)*
0	0
0.183	306
0.36	612
0.5324	917
0.702	1223
0.867	1529
1.0244	1835
1.1774	2140
1.329	2446
1.479	2752
1.5	2767
1.56	2896

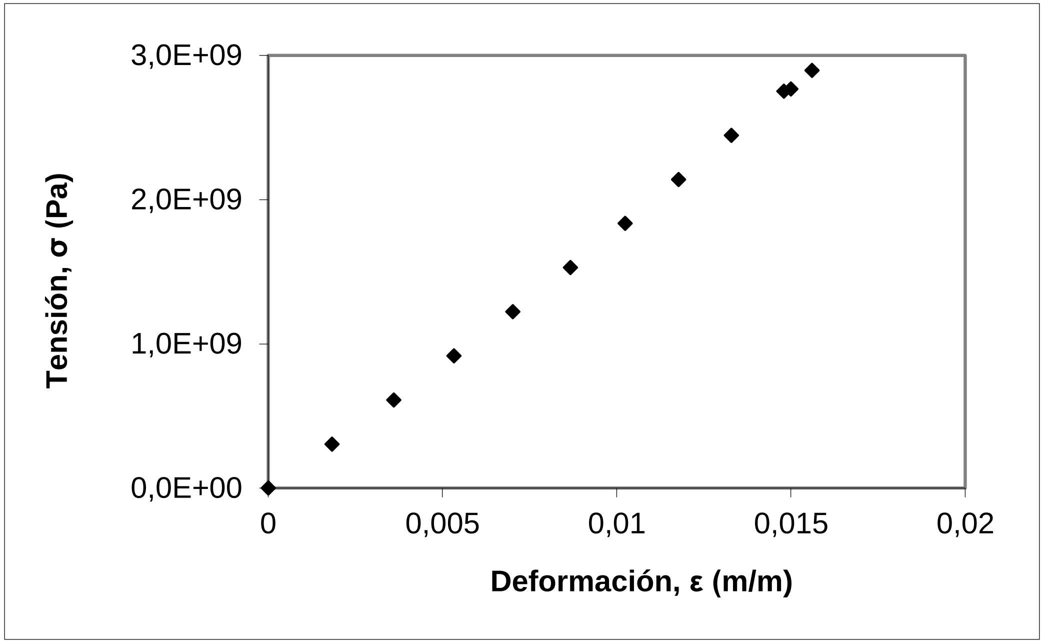


Fig. Datos de tensión vs. deformación.

Ejemplo 2 (cont.)

Tabla. Sumatoria de datos para el modelo de regresión.

i	ε	σ	ε²	εσ
1	0.0000	0.0000	0.0000	0.0000
2	1.8300×10^{-3}	3.0600×10^{8}	3.3489×10^{-6}	5.5998×10^{5}
3	3.6000×10^{-3}	6.1200×10^{8}	1.2960×10^{-5}	2.2032×10^{6}
4	5.3240×10^{-3}	9.1700×10^{8}	2.8345×10^{-5}	4.8821×10^{6}
5	7.0200×10^{-3}	1.2230×10^{9}	4.9280×10^{-5}	8.5855×10^{6}
6	8.6700×10^{-3}	1.5290×10^{9}	7.5169×10^{-5}	1.3256×10^{7}
7	1.0244×10^{-2}	1.8350×10^{9}	1.0494×10^{-4}	1.8798×10^{7}
8	1.1774×10^{-2}	2.1400×10^{9}	1.3863×10^{-4}	2.5196×10^{7}
9	1.3290×10^{-2}	2.4460×10^{9}	1.7662×10^{-4}	3.2507×10^{7}
10	1.4790×10^{-2}	2.7520×10^{9}	2.1874×10^{-4}	4.0702×10^{7}
11	1.5000×10^{-2}	2.7670×10^{9}	2.2500×10^{-4}	4.1505×10^{7}
12	1.5600×10^{-2}	2.8960×10^{9}	2.4336×10^{-4}	4.5178×10^{7}
$\sum_{i=1}^{12}$			**1.2764×10^{-3}**	**2.3337×10^{8}**

$$E=\frac{\sum_{i=1}^{n}\sigma_i\varepsilon_i}{\sum_{i=1}^{n}\varepsilon_i^{2}}$$

$$\sum_{i=1}^{12}\varepsilon_i^{2}=1.2764\times10^{-3}$$

$$\sum_{i=1}^{12}\sigma_i\varepsilon_i=2.3337\times10^{8}$$

$$E=\frac{\sum_{i=1}^{12}\sigma_i\varepsilon_i}{\sum_{i=1}^{12}\varepsilon_i^{2}}=\frac{2.3337\times10^{8}}{1.2764\times10^{-3}}=182.84\ GPa$$

Ejemplo 2 - Resultados

La ecuación $\sigma = (182.84\ 10^9\ Pa)\ \varepsilon$ describe los datos.

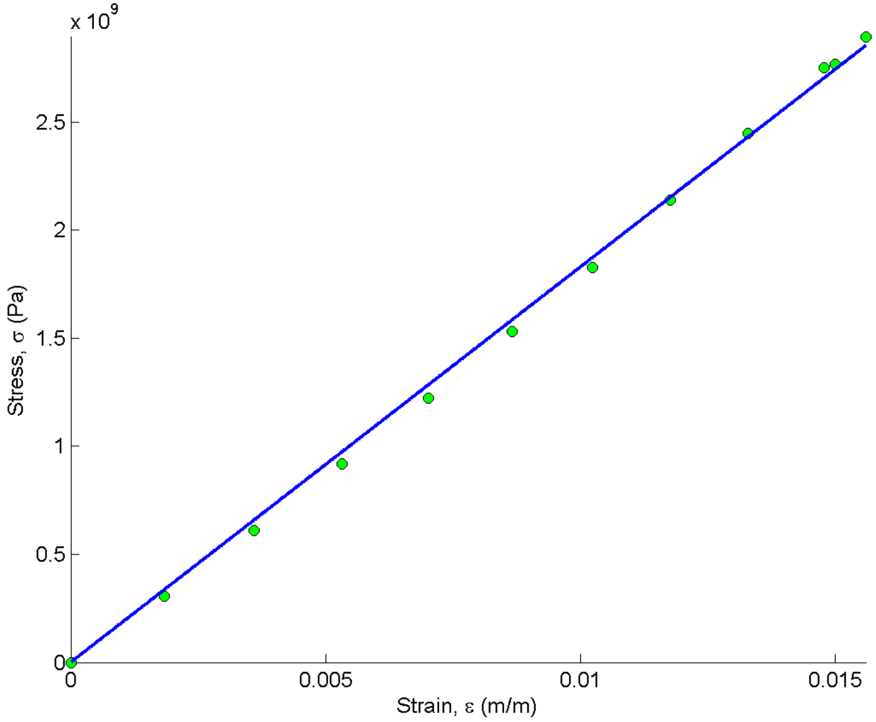


Fig. Regresión lineal de los datos de Tensión vs. Deformación