ANÁLISIS NUMÉRICO

Mag. Carlos Alberto Ardila Albarracín

BLOQUE 2. AJUSTE DE CURVAS
2.4. EJEMPLOS COMPLEMENTARIOS SOBRE REGRESIÓN

EJEMPLO 1.

Use la regresión por mínimos cuadrados para ajustar una línea recta a:

X	6	7	11	15	17	21	23	**29**	**29**	37	39
Y	29	21	29	14	21	15	7	**7**	**13**	0	3

Además de la pendiente y la intersección, calcule el error estándar de la estimación y el coeficiente de correlación (r). Haga una gráfica de los datos y la línea de regresión.

¿Si otra persona hiciera una medición adicional de (x=10, y=10), usted pensaría que la medición era válida o inválida? Justifique su conclusión.

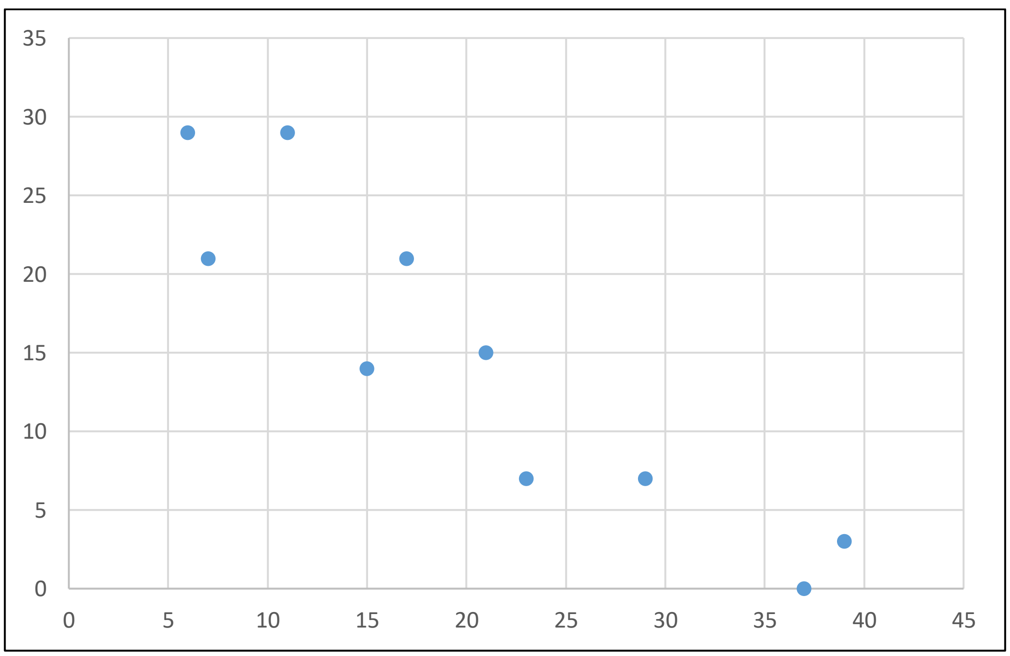
SOLUCIÓN:

Como puede verse se están reportando 2 resultados para el valor x=29. Entonces, en esos casos se sugiere hacer dos modelos: uno que incluya la pareja (x=29, y=7) y excluya (x=29, y=13) y otro que incluya la pareja (x=29, y=13) y excluya la pareja (x=29, y=7). Y se evalúa cuál es mejor.

Modelo 1:

X	6	7	11	15	17	21	23	**29**	37	39
Y	29	21	29	14	21	15	7	**7**	0	3

Graficando:



Al observar la gráfica, se aprecia que hay una razonable tendencia de linealidad así como de homocedasticidad.

Utilizando el programa de regresión lineal:

REGRESION LINEAL

n	10
ST	960.4
SR	156.05039

Sy	10.33011
Sy/x	4.41659

Interpretación al leer Sy y Sy/x:
(Sy/x **<** Sy**) → Aprox. SÍ se considera aceptable**

$\mathbf{R^2}$	0.83751

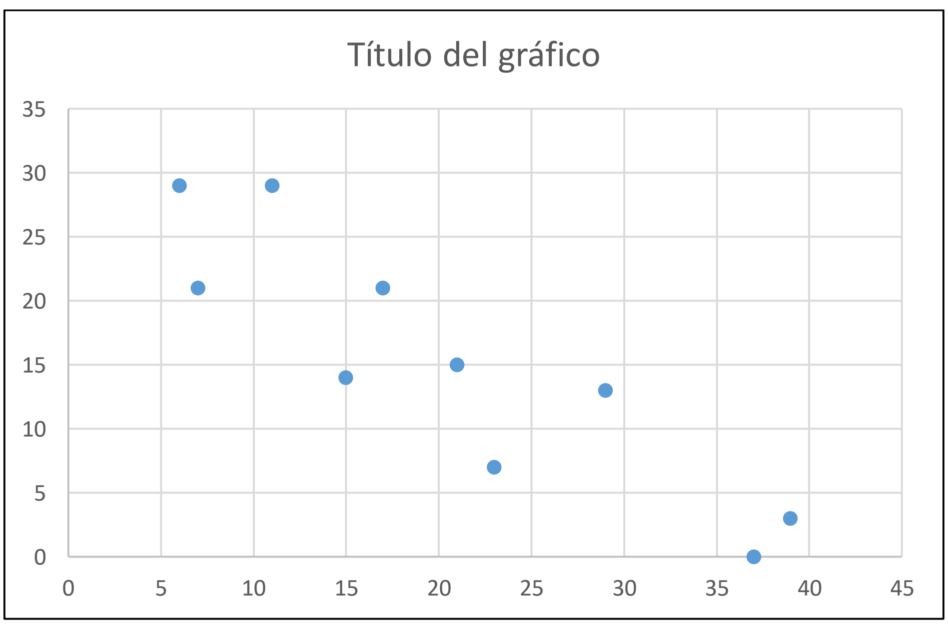
Interpretación de $\mathbf{R^2}$
El 83.75% de la incertidumbre original se explica mediante el modelo lineal

Ecuación de la recta obtenida:
b_0 = 31.25572
b_1 = -0.81247

$Y = - 0.81247\mathbf{x} + 31.25572$

Modelo 2:

X	6	7	11	15	17	21	23	**29**	37	39
Y	29	21	29	14	21	15	7	**13**	0	3



Al observar la gráfica, se aprecia que hay una razonable tendencia de linealidad así como de homocedasticidad.

Utilizando el programa de regresión lineal:

REGRESION LINEAL

n	10
ST	901.6
SR	177.98818

Sy	10.00888
Sy/x	4.71683

Interpretación al leer Sy y Sy/x:
(Sy/x **<** Sy**) → Aprox. SÍ se considera aceptable**

$\mathbf{R^2}$	0.80259

Interpretación de $\mathbf{R^2}$
El 80.26% de la incertidumbre original se explica mediante el modelo lineal

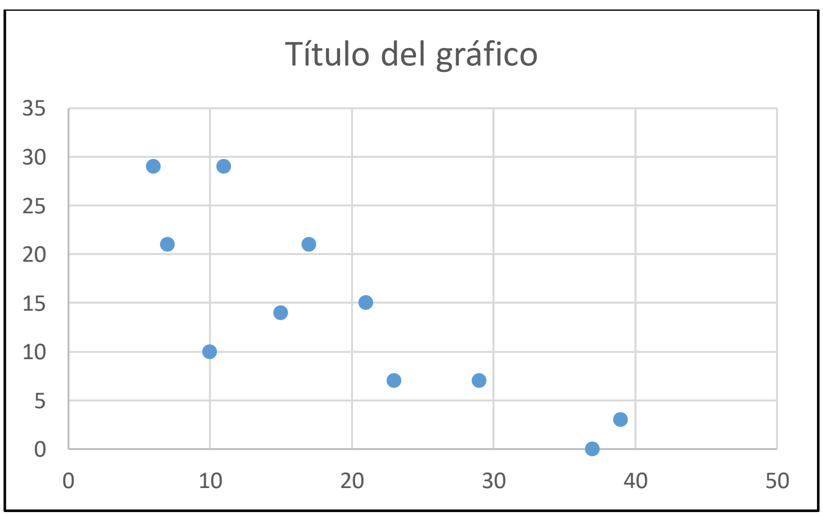
Ecuación de la recta obtenida:
b_0 = 30.99770
b_1 = -0.77062

Y = **-** 0.77062**x** + 30.99770

Como puede verse al evaluar el valor R^2, es mejor el modelo 1, es decir, el que incluye la pareja (x=29, y=7) y excluye (x=29, y=13).

Entonces, sobre el modelo escogido, se agrega la nueva pareja (x=10, y=10).

X	6	7	**10**	11	15	17	21	23	29	37	39
Y	29	21	**10**	29	14	21	15	7	7	0	3



Al observar la gráfica, se aprecia un comportamiento heterocedástico. De modo que no se recomendaría usar el modelo de regresión lineal.

Sin embargo, y solo como verificación, se muestra lo que se obtiene al analizar esos datos:

Utilizando el programa de regresión lineal:

REGRESION LINEAL

n	11
ST	979.63636
SR	300.88494

Sy	9.89766
Sy/x	5.78201

Interpretación al leer Sy y Sy/x:
(Sy/x **<** Sy**) → Aprox. SÍ se considera aceptable**

$\mathbf{R^2}$	0.69286

Interpretación de $\mathbf{R^2}$
El 69.28% de la incertidumbre original se explica mediante el modelo lineal

Ecuación de la recta obtenida:
$b_0 = 28.20426$
$b_1 = -0.71743$

$Y = \mathbf{-}\ 0.71743\mathbf{x} + 28.20426$

Los resultados reflejan el comportamiento heterocedástico en la considerable reducción del valor del coeficiente de determinación R^2.

Por consiguiente, al apreciar la degradación del modelo, se puede considerar que el valor (x=10, y=10) ***NO sería válido***.

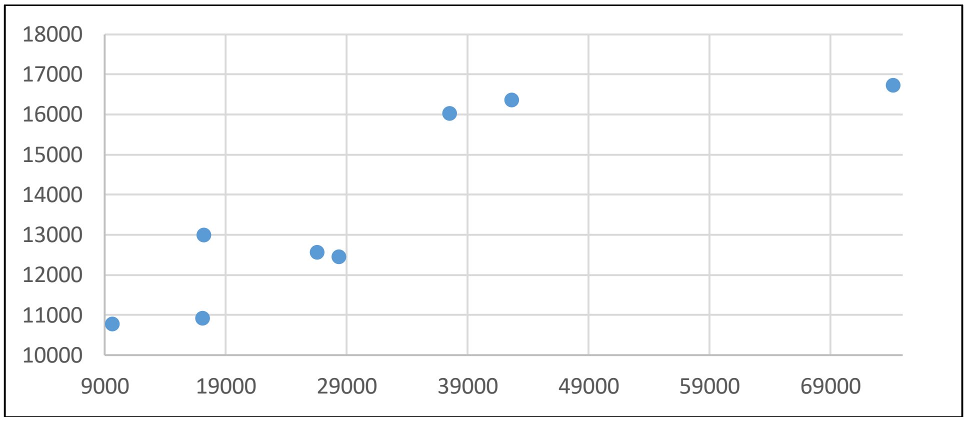
EJEMPLO 2.

La siguiente tabla muestra el consumo mensual de energía como función de unidades producidas por una fábrica:

x = unidades producidas	9596	17073	26569	28377	74181	17162	37541	42630
y = consumo (kilovatios-hora)	10784	10924	12565	12464	16736	12995	16036	16367

La pregunta es:
¿Cuál modelo ajusta mejor esos datos: Lineal, Función Potencia o Cuadrático?

Se procede a generar el modelo lineal:



Utilizando el programa de regresión lineal:

REGRESION LINEAL

n	8
ST	41242828.88
SR	10932001.8

Sy	2427.30977
Sy/x	1349.81491

Interpretación al leer Sy y Sy/x:
(Sy/x **<** Sy**) → Aprox. SÍ se considera aceptable**

$\mathbf{R^2}$	0.73494

Interpretación de $\mathbf{R^2}$
El 73.49% de la incertidumbre original se explica mediante el modelo lineal

Ecuación de la recta obtenida:
b_0 = 10379.29938
b_1 = 0.10207

Y = 0.10207**x** + 10379.29938

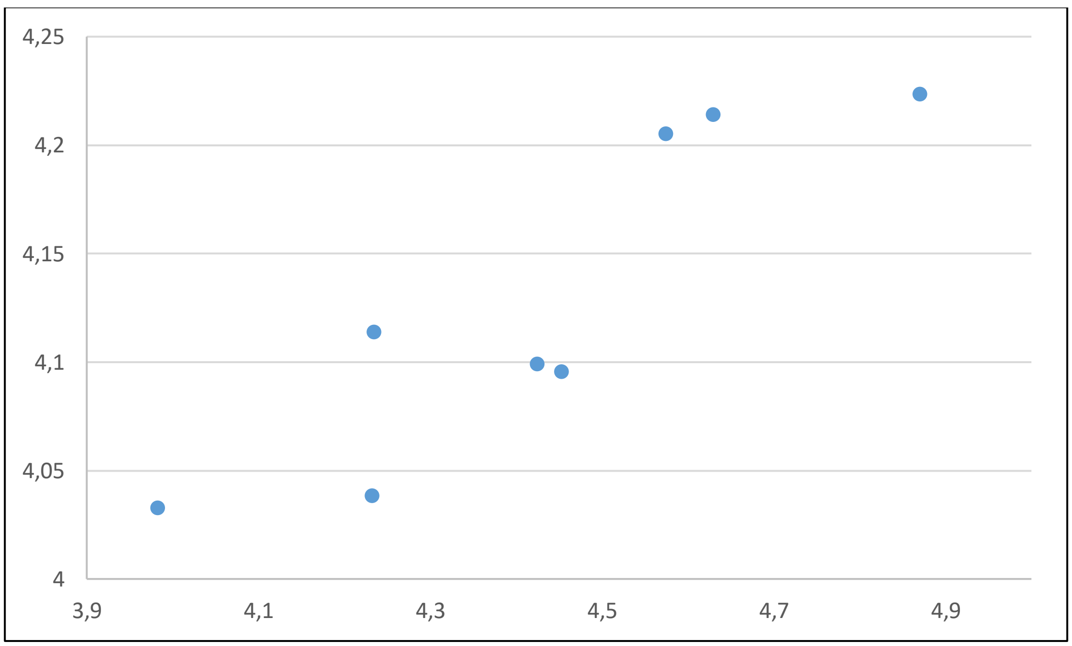
Se procede a generar el modelo para la función potencia:

Transformamos los datos:

Nueva X	Nueva Y
3,98209024	4,03277988
4,23230984	4,03838169
4,42437521	4,09916249
4,45296648	4,09565744
4,87029268	4,22365167
4,2345679	4,11377628
4,57450584	4,20509605
4,62971533	4,21396908

Con esos datos generamos el modelo lineal intermedio. Con los coeficientes que se obtengan, se podrán estimar los términos de la función potencia.

Gráfica de los datos transformados:



Los resultados obtenidos son los siguientes:

Utilizando el programa de regresión lineal:

REGRESION LINEAL

n	8
ST	0.04166
SR	0.00848

Sy	0.07715
Sy/x	0.03759

Interpretación al leer Sy y Sy/x:
(Sy/x **<** Sy**) → Aprox. SÍ se considera aceptable**

$\mathbf{R^2}$	0.79656

Interpretación de $\mathbf{R^2}$
El 79.66% de la incertidumbre original se explica mediante el modelo lineal

Ecuación de la recta obtenida:
b_0 = 3.02352
b_1 = 0.24955

Y = 0.24955**x** + 3.02352

Se arma la función potencia: $y = c * x^{a}$

potencial y=c* x a

$c = 10^{(bo)}$
a = b1

$y = 10^{(bo)} * x^{(b1)}$

y para nuestro ejemplo:

$c = 10^{(3.02352)}$
a = 0.24955

$y = 10^{(3.02352)} * x^{(0.24955)}$

$$\boldsymbol{y = 1055.65 * x^{(0.24955)}}$$

El valor $\mathbf{R^2}$ Se obtuvo al usar el modelo lineal de apoyo y es igual a 0.79656.

Se procede a generar el modelo para regresión cuadrática:

n	8
ST	41242828.88
SR	6683435.102

Sy	2427.30977
Sy/x	1156.15182

Interpretación al leer Sy y Sy/x:
(Sy/x **<** Sy**) → Aprox. SÍ se considera aceptable**

$\mathbf{R^2}$	0.83795

Interpretación de $\mathbf{R^2}$
El 83.79% de la incertidumbre original se explica mediante el modelo **cuadrático**

Ecuación de la parábola obtenida:
$b_0 = 7978.59827$
$b_1 = 0.25721$
$b_2 = -1.83764 * 10^{-6}$

$$Y = -1.83764 * 10^{-6} \mathbf{x}^2 + 0.25721 \mathbf{x} + 7978.59827$$

Resumiendo:

Modelo	Coeficiente de determinación
Lineal	0.73494
Potencial	0.79656
Cuadrático	0.83795

Como se puede apreciar el modelo con el mayor coeficiente de determinación (0.83795) es el obtenido mediante regresión cuadrática, de modo que es el modelo que ajusta mejor esos datos.

--- FIN DEL DOCUMENTO