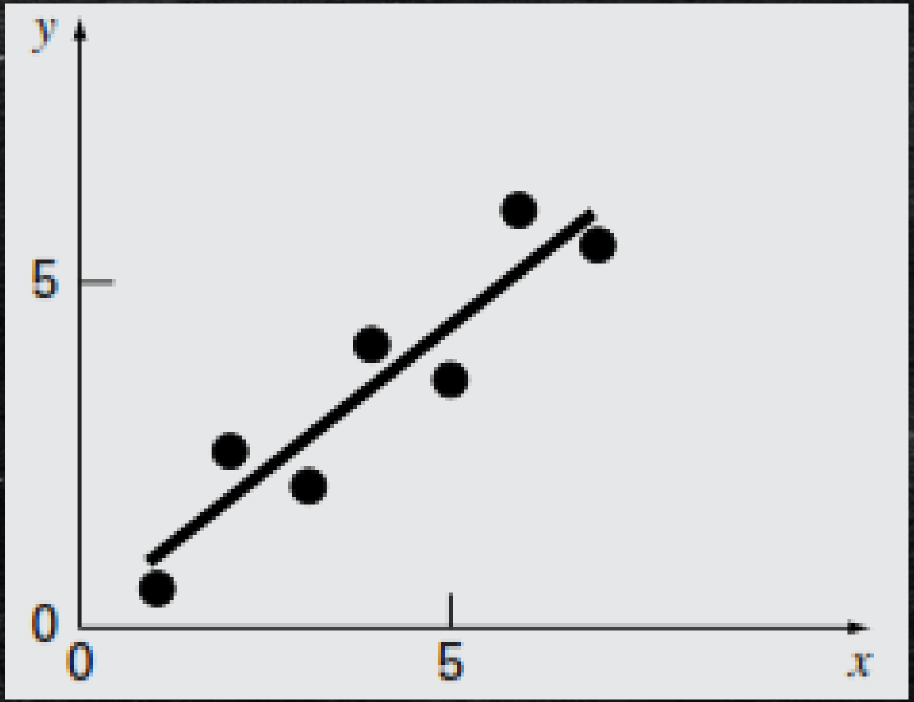
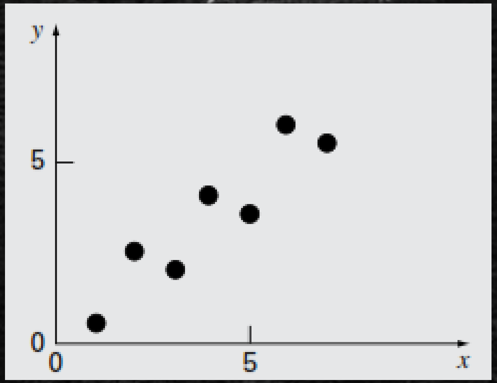
MÉTODOS NUMÉRICOS: TEMA: *Aproximación Funcional*

Subtema: *Regresión por los mínimos cuadrados*

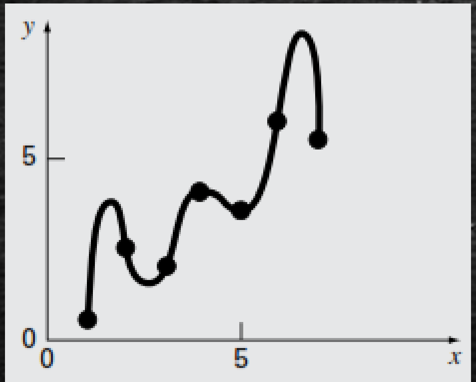


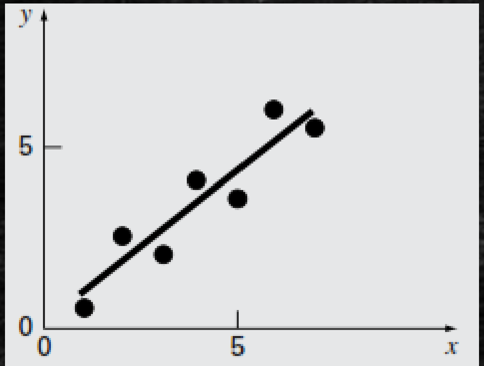
Franco Salcedo López.
2020

Regresión por los mínimos cuadrados



Si se encuentra un polinomio de interpolación que se ajusta a estos datos , este polinomio *pasará exactamente a través de todos los puntos.* Sin embargo, a causa de la variabilidad en los datos, la curva oscila mucho en el intervalo entre los puntos, por lo que esta interpolación es inapropiada





Una estrategia más apropiada consiste en obtener una función de aproximación que se ajuste a la tendencia general de los datos y que pase *entre los datos*. La forma de hacerlo es obtener una curva que minimice la discrepancia entre los puntos y la curva. Una técnica para lograr tal objetivo, es la llamada regresión por mínimos cuadrados.

Regresión por los mínimos cuadrados

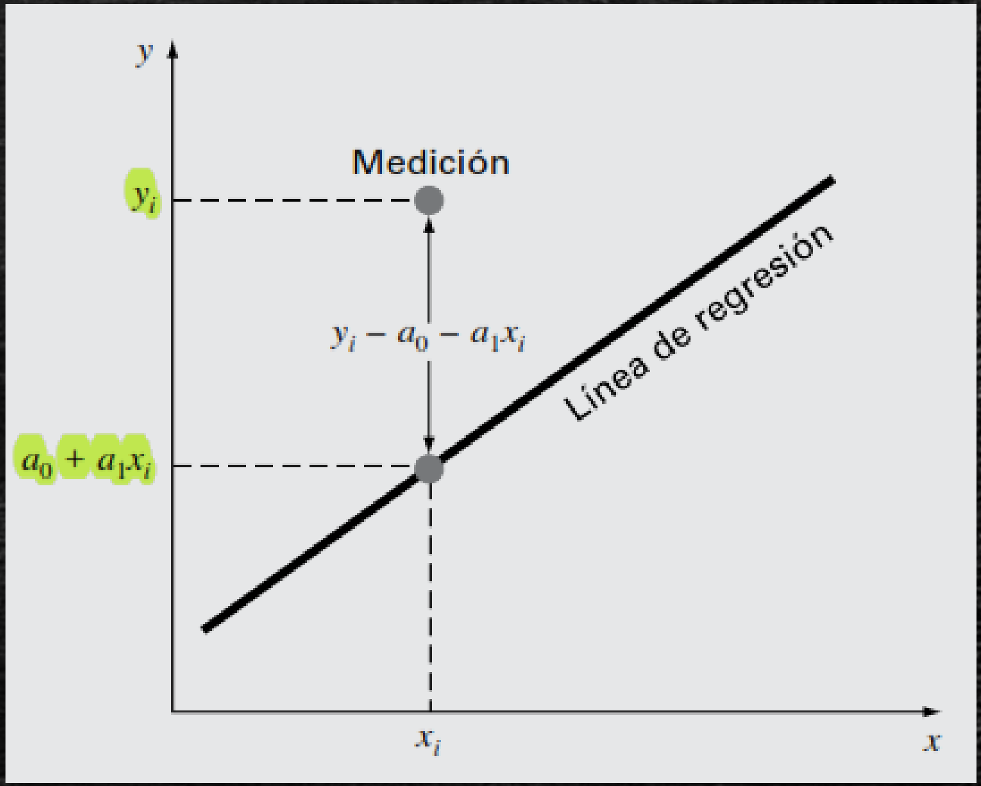
El ejemplo más simple de una aproximación por mínimos cuadrados es *ajustar una línea recta* a un conjunto de observaciones definidas por puntos: $(x_1,y_1),(x_2,y_2),\ldots,(x_n, y_n)$.

Así, la expresión matemática para la línea recta es:

$$y = a_0 + a_1 x + e$$

donde a_0 y a_1 son coeficientes que representan la intersección con el eje y, y la pendiente, respectivamente, *e* es el error, o diferencia, entre el modelo y las observaciones. Así:

$$e = \underbrace{y}_{\text{Valor verdadero}} - \underbrace{a_0 - a_1 x}_{\text{Aproximación}}$$



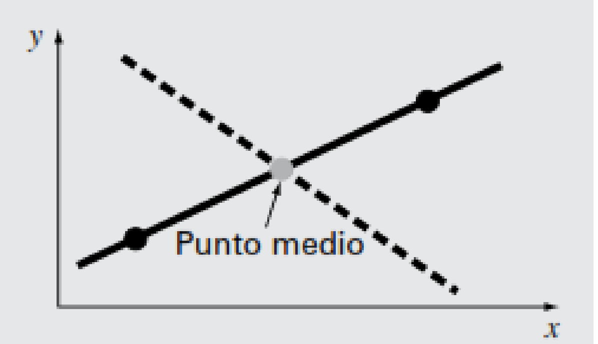
Franco Salcedo López.
2020

Regresión por los mínimos cuadrados

Criterios para un mejor ajuste

Una estrategia para ajustar una "mejor" línea a través de los datos será *minimizar la suma de los errores residuales de todos los datos disponibles*, como sigue:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$



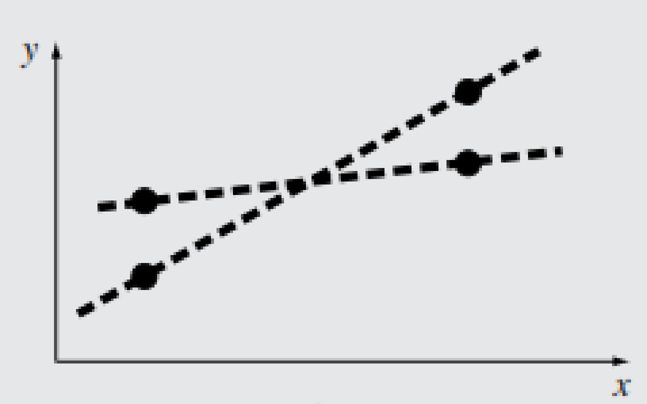
Éste es un *criterio inadecuado*. Obviamente, en este caso el mejor ajuste es la línea que une los 2 puntos. Sin embargo, cualquier línea recta que pase a través del punto medio que une la línea da como resultado un valor mínimo de la ecuación igual a cero, debido a que los errores se cancelan.

Regresión por los mínimos cuadrados

Criterios para un mejor ajuste

Una 2da estrategia para ajustar una "mejor" línea a través de los datos será *minimizar la suma de los valores absolutos de las diferencias*, como sigue:

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

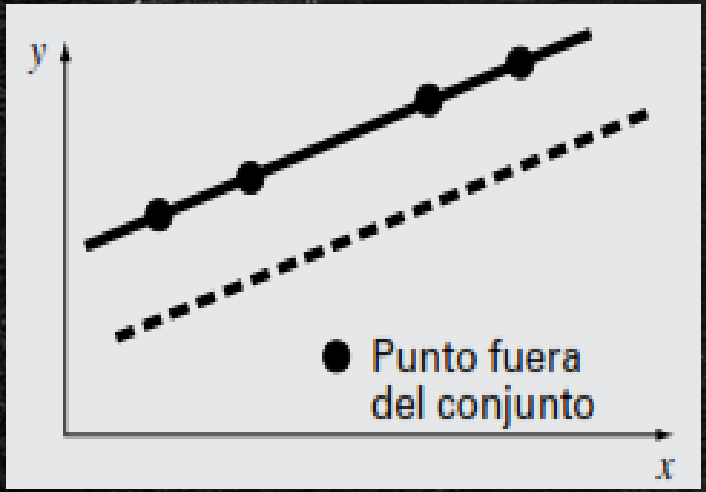


Éste también es un *criterio inadecuado.* Para los cuatro puntos dados, cualquier línea recta que esté dentro de las líneas punteadas minimizará el valor absoluto de la suma.

Regresión por los mínimos cuadrados

Criterios para un mejor ajuste

Una 3era estrategia para ajustar una "mejor" línea a través de los datos será usar el *criterio minimax*, como sigue:



En esta técnica, la línea se elige de manera que minimice la máxima distancia a que un punto se encuentra de la línea. Tal estrategia es inadecuada para la regresión, ya que da excesiva influencia a puntos fuera del conjunto; es decir, a un solo punto con un gran error.

Regresión por los mínimos cuadrados

Criterios para un mejor ajuste

Una *4ta estrategia y la mejor* para ajustar una "mejor" línea a través de los datos será *minimizar la suma de los cuadrados de los residuos entre la* ***y*** *medida y la* ***y*** *calculada con el modelo lineal:*

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{medida}} - y_{i,\text{modelo}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Este criterio tiene varias ventajas, entre ellas el hecho de que se obtiene una línea única para cierto conjunto de datos.

Regresión por los mínimos cuadrados

Ajuste de una línea recta por mínimos cuadrados

Para determinar los valores de a_0 y a_1, de la ecuación S_r, se deriva con respecto a cada uno de los coeficientes:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{medida}} - y_{i,\text{modelo}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Así se tiene:

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i) x_i]$$

Al derivar considerar las constantes

Que al igualar estas derivadas a cero, dará como resultado un S_r mínimo

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

Regresión por los mínimos cuadrados

Ajuste de una línea recta por mínimos cuadrados

Ahora, ya que $\sum a_0 = na_0$, expresamos las ecuaciones como un conjunto de dos ecuaciones lineales simultáneas, con dos incógnitas (a_0 y a_1):

$$na_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

Éstas se llaman *ecuaciones normales*, y se resuelven en forma simultánea, con los métodos clásicos: Gauss-Jordan, Determinantes, Inversa, Etc..

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Reemplazando este valor en el sistema de ecuaciones normales, se tiene:

$$na_0 + \left(\sum x_i\right) a_1 = \sum y_i \quad \rightarrow \quad a_0 = \bar{y} - a_1 \bar{x}$$

Regresión por los mínimos cuadrados

Ejercicio base:

Ajuste a una línea recta los valores x y y, en las dos primeras columnas de la tabla

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
Σ	24.0	22.7143	2.9911

$n = 7 \qquad \sum x_i y_i = 119.5 \qquad \sum x_i^2 = 140$

$\sum x_i = 28 \qquad \bar{x} = \frac{28}{7} = 4$

$\sum y_i = 24 \qquad \bar{y} = \frac{24}{7} = 3.428571$

Entonces:

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} \rightarrow a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$

$$a_0 = \bar{y} - a_1\bar{x} \rightarrow a_0 = 3.428571 - 0.8392857(4) = 0.07142857$$

Recta que q e.

$$y = 0.07142857 + 0.8392857x$$

Franco Salcedo López.
2020

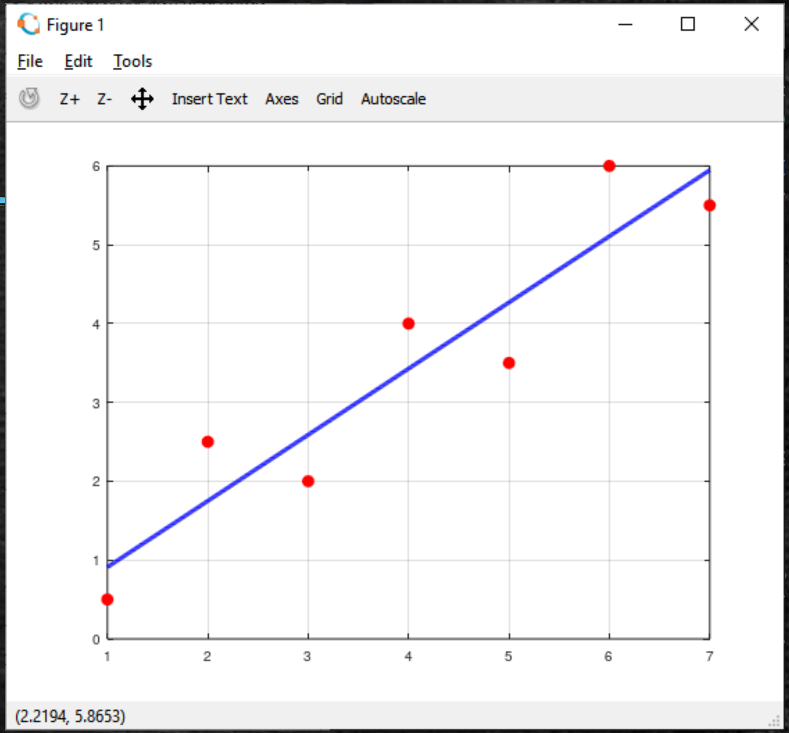
Regresión por los mínimos cuadrados

Ejercicio base:

Uso del Software OCTAVE:

```
>> x1=[1:1:7]
x1 =

   1   2   3   4   5   6   7
>> y1=[0.5 2.5 2 4 3.5 6 5.5]
y1 =

   0.50000   2.50000   2.00000   4.00000   3.50000   6.00000   5.50000
>>[a1,a0] = LinearReqression(x1,y1)
a1 =  0.83929
a0 =  0.071429
```

$$y = 0.071429 + 0.83929\ x$$



Regresión por los mínimos cuadrados

Ejercicios en clase

Ejemplo 1 Encuentre la línea de mínimos cuadrados que se aproxima a los datos en la tabla

x_i	y_i
1	1.3
2	3.5
3	4.2
4	5.0
5	7.0
6	8.8
7	10.1
8	12.5
9	13.0
10	15.6
55	81.0

Franco Salcedo López.
2020

Regresión por los mínimos cuadrados

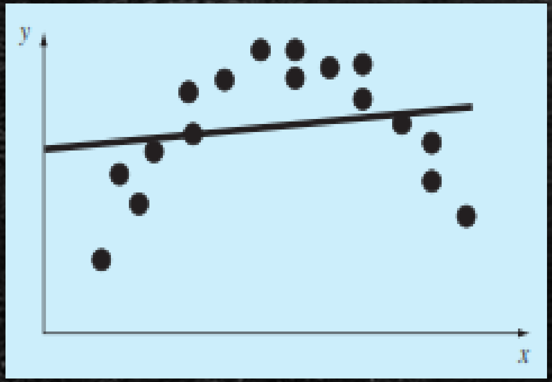
Ejercicios en clase

En la tabla siguiente se presentan los alargamientos de un resorte, correspondientes a fuerzas de diferente magnitud que lo deforman.

Puntos	1	2	3	4	5
Fuerza (kgf): x	0	2	3	6	7
Longitud del resorte (m): y	0.120	0.153	0.170	0.225	0.260

Determine por mínimos cuadrados el mejor polinomio de primer grado (recta) que represente la función dada.

Regresión Polinomial



En ingeniería, aunque algunos datos exhiben un patrón marcado, como el que se advierte en la figura, son pobremente representados por una línea recta, entonces, una curva podrá ser más adecuada para ajustarse a los datos.

Una alternativa es ajustar polinomios a los datos mediante regresión polinomial.

Regresión por los mínimos cuadrados

Regresión Polinomial

El procedimiento de mínimos cuadrados se puede extender fácilmente al ajuste de datos con un polinomio de grado superior. Por ejemplo, suponga que ajustamos un polinomio de segundo grado o cuadrático:

$$y = a_0 + a_1 x + a_2 x^2 + e$$

En este caso, la suma de los cuadrados de los residuos es:

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

Derivando con respecto a cada uno de los coeficientes desconocidos del polinomio, se tiene:

Regresión por los mínimos cuadrados

Regresión Polinomial

$$\frac{\partial S_r}{\partial a_0} = -2\sum (y_i - a_0 - a_1 x_i - a_2 x_i^2)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum x_i (y_i - a_0 - a_1 x_i - a_2 x_i^2)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum x_i^2 (y_i - a_0 - a_1 x_i - a_2 x_i^2)$$

Igualado a cero y reordenando, se tiene:

$$(n)a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum x_i y_i$$

$$\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum x_i^2 y_i$$

Con lo que se ha logrado obtener un sistema de ecuaciones lineales con 3 incógnitas, a_0, a_1, a_2.

Regresión Polinomial GENERALIZADO

Se puede GENERALIZAR con facilidad para un polinomio de m-ésimo grado como sigue:

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m + e$$

Definiendo la suma de los cuadrados de los residuos, derivando con respecto a cada uno de los coeficientes desconocidos del polinomio e igualando a cero, se obtiene un sistema de ecuaciones lineales de m+1 ecuaciones que se debe resolver para obtener los coeficientes del polinomio.

Regresión por los mínimos cuadrados

Regresión Polinomial GENERALIZADO

Por ejemplo para obtener un polinomio de grado 3:

$$na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\,2}\right) a_2 + \left(\sum_{i=1}^{n} x_i^{\,3}\right) a_3 = \sum_{i=1}^{n} y_i$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^{\,2}\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\,3}\right) a_2 + \left(\sum_{i=1}^{n} x_i^{\,4}\right) a_3 = \sum_{i=1}^{n} x_i\, y_i$$

$$\left(\sum_{i=1}^{n} x_i^{2}\right) a_0 + \left(\sum_{i=1}^{n} x_i^{\,3}\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\,4}\right) a_2 + \left(\sum_{i=1}^{n} x_i^{\,5}\right) a_3 = \sum_{i=1}^{n} x_i^{2}\, y_i$$

$$\left(\sum_{i=1}^{n} x_i^{3}\right) a_0 + \left(\sum_{i=1}^{n} x_i^{\,4}\right) a_1 + \left(\sum_{i=1}^{n} x_i^{\,5}\right) a_2 + \left(\sum_{i=1}^{n} x_i^{\,6}\right) a_3 = \sum_{i=1}^{n} x_i^{3}\, y_i$$

Regresión por los mínimos cuadrados

Ejercicio base:

Ajustar a un polinomio de segundo grado los datos dados en la siguiente tabla:

x_i	y_i
0	2.1
1	7.7
2	13.6
3	27.2
4	40.9
5	61.1
Σ	152.6

$$(n)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 = \sum y_i$$
$$\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 = \sum x_i y_i$$
$$\left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 = \sum x_i^2 y_i$$

$m = 2 \qquad \sum x_i = 15 \qquad \sum x_i^4 = 979$

$n = 6 \qquad \sum y_i = 152.6 \qquad \sum x_i y_i = 585.6$

$\bar{x} = 2.5 \qquad \sum x_i^2 = 55 \qquad \sum x_i^2 y_i = 2\,488.8$

$\bar{y} = 25.433 \qquad \sum x_i^3 = 225$

Entonces:

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2\,488.8 \end{Bmatrix}$$

Resolviendo el sistema:

se tiene $a_0 = 2.47857$, $a_1 = 2.35929$ y $a_2 = 1.86071$.

$$y = 2.47857 + 2.35929x + 1.86071x^2$$

Recta que q e.

Franco Salcedo López.
2020

Regresión por los mínimos cuadrados

Ejercicio base:

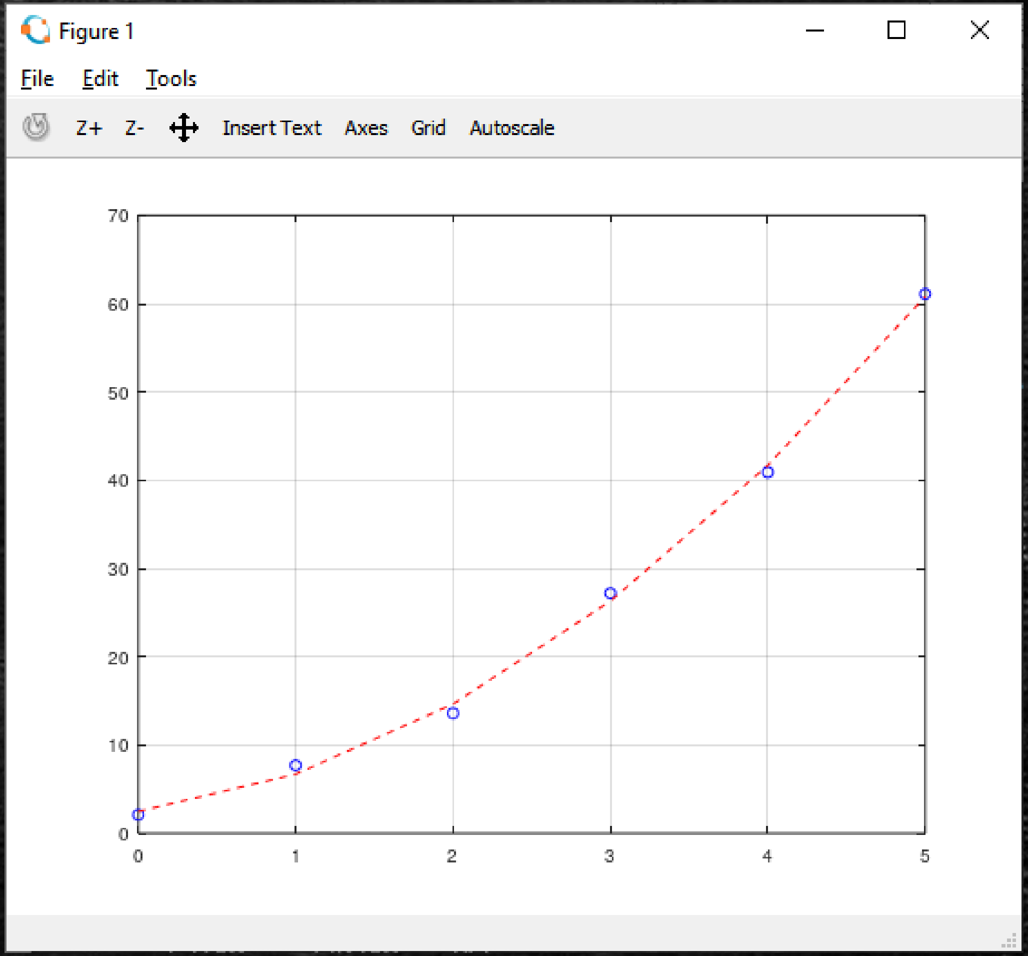
Uso del Software OCTAVE:

```
>> x2=[1:1:5]
x2 =

   0   1   2   3   4   5
>> y2=[2.1 7.7 13.6 27.2 40.9 61.1]
y2 =

    2.1000    7.7000   13.6000   27.2000   40.9000   61.1000

>> [A,pol] = regrelineal (x2, y2, 2)
```

Regresión por los mínimos cuadrados

Ejercicio base:

```
poli =

    521/280  3303/1400    347/140

A =

          6         15         55      763/5
         15         55        225     2928/5
         55        225        979    12444/5

pol = (sym)

         2
   521*x     3303*x    347
   ------ + ------ + ---
    280       1400    140
```



Regresión por los mínimos cuadrados

Ejercicios en clase

Ejemplo 1 Encuentre la línea de mínimos cuadrados que se aproxima a los datos en la tabla

x_i	y_i
1	1.3
2	3.5
3	4.2
4	5.0
5	7.0
6	8.8
7	10.1
8	12.5
9	13.0
10	15.6
55	81.0

Regresión por los mínimos cuadrados

Ejercicios en clase

En la tabla siguiente se presentan los alargamientos de un resorte, correspondientes a fuerzas de diferente magnitud que lo deforman.

Puntos	1	2	3	4	5
Fuerza (kgf): x	0	2	3	6	7
Longitud del resorte (m): y	0.120	0.153	0.170	0.225	0.260

Determine por mínimos cuadrados el mejor polinomio de primer grado (recta) que represente la función dada.