Ejercicio :
MÍNIMOS CUADRADOS

En una empresa biotecnológica se ha fabricado cierto fármaco que lleva ya siete años en el mercado. Uno de los científicos que allí trabajan está pasando por momentos económicamente difíciles y desea saber cuántos beneficios obtendrá la empresa por la venta de este medicamento el próximo año 2021, pues su sueldo es directamente proporcional a estas cantidades.

Para ello, ha indagado en los archivos de contabilidad y ha elaborado la siguiente tabla, que contiene las ganancias de la empresa por la venta de este medicamento en los últimos siete años. Para predecir los resultados del nuevo año, decide emplear el método de mínimos cuadrados que le enseñaron en primero de carrera para hallar los coeficientes de la recta de regresión que mejor se aproxima a los datos. Además, planea hacer un algoritmo para que los cálculos sean más sencillos… ¿Podrías ayudarle?

AÑOS	EUROS
2014	5000
2015	13000
2016	10000
2017	13500
2018	11000
2019	14000
2020	18000

Almacenado en el vector s= (1, 2, 3, 4, 5, 6, 7) abscisas

Almacenado en el vector t ordenadas

n = 7

RESOLUCIÓN

Queremos hallar los coeficientes de la reca de regresión: R(x) = a + bx

Para ello expresamos la función de dos variables como:

$$f\,(a,b) = \sum_{j=1}^{n} d_j{}^2 = \sum_{j=1}^{n} (y_j - (a + b \cdot s_j))^2$$

Derivamos e igualamos a 0:

$$\frac{\partial f}{\partial a} = 2\sum_{j=1}^{n} [(y_j - (a + b \cdot s_j))(-1)] = 0$$

$$\sum_{j=1}^{n} (a + b \cdot s_j) = \sum_{j=1}^{n} (y_j)$$

$$n \cdot a + b \sum_{j=1}^{n} (s_j) = \sum_{j=1}^{n} (y_j)$$

$$\frac{\partial f}{\partial b} = 2\sum_{j=1}^{n} [(y_j - (a + b \cdot s_j))(-s_j)] = 0$$

$$\sum_{j=1}^{n} [(a + b \cdot s_j)(s_j) = \sum_{j=1}^{n} (y_j \cdot s_j)$$

$$a \sum_{j=1}^{n} (s_j) + b \sum_{j=1}^{n} (s_j)^2 = \sum_{j=1}^{n} (y_j \cdot s_j)$$

$$n \cdot a + b \sum_{j=1}^{n} (s_j) = \sum_{j=1}^{n} (y_j)$$

$$a \sum_{j=1}^{n} (s_j) + b \sum_{j=1}^{n} (s_j)^2 = \sum_{j=1}^{n} (y_j \cdot s_j)$$

$$\mathbf{SX} = \sum s_j$$

$$\mathbf{SY} = \sum y_j$$

$$\mathbf{SX2} = \sum (s_j)^2$$

$$\mathbf{SXY} = \sum (y_j \cdot s_j)$$

$$n \cdot a + b \cdot \mathbf{SX} = \mathbf{SY}$$
$$a \cdot \mathbf{SX} + b \cdot \mathbf{SX2} = \mathbf{SXY}$$

$$A = \begin{pmatrix} n & \mathbf{SX} \\ \mathbf{SX} & \mathbf{SX2} \end{pmatrix} \qquad X = \begin{bmatrix} a \\ b \end{bmatrix} \qquad F = \begin{pmatrix} \mathbf{SY} \\ \mathbf{SXY} \end{pmatrix}$$

m2/2020

RESOLUCIÓN

Hallamos el determinante (**D**) de A:

$$\mathbf{D} = \begin{vmatrix} n & \mathbf{SX} \\ \mathbf{SX} & \mathbf{SX2} \end{vmatrix} = n \cdot \mathbf{SX2} - (\mathbf{SX})^2 = 7 \cdot (1^2+2^2+3^2+4^2+5^2+6^2+7^2) - (1+2+3+4+5+6+7)^2 = 196$$

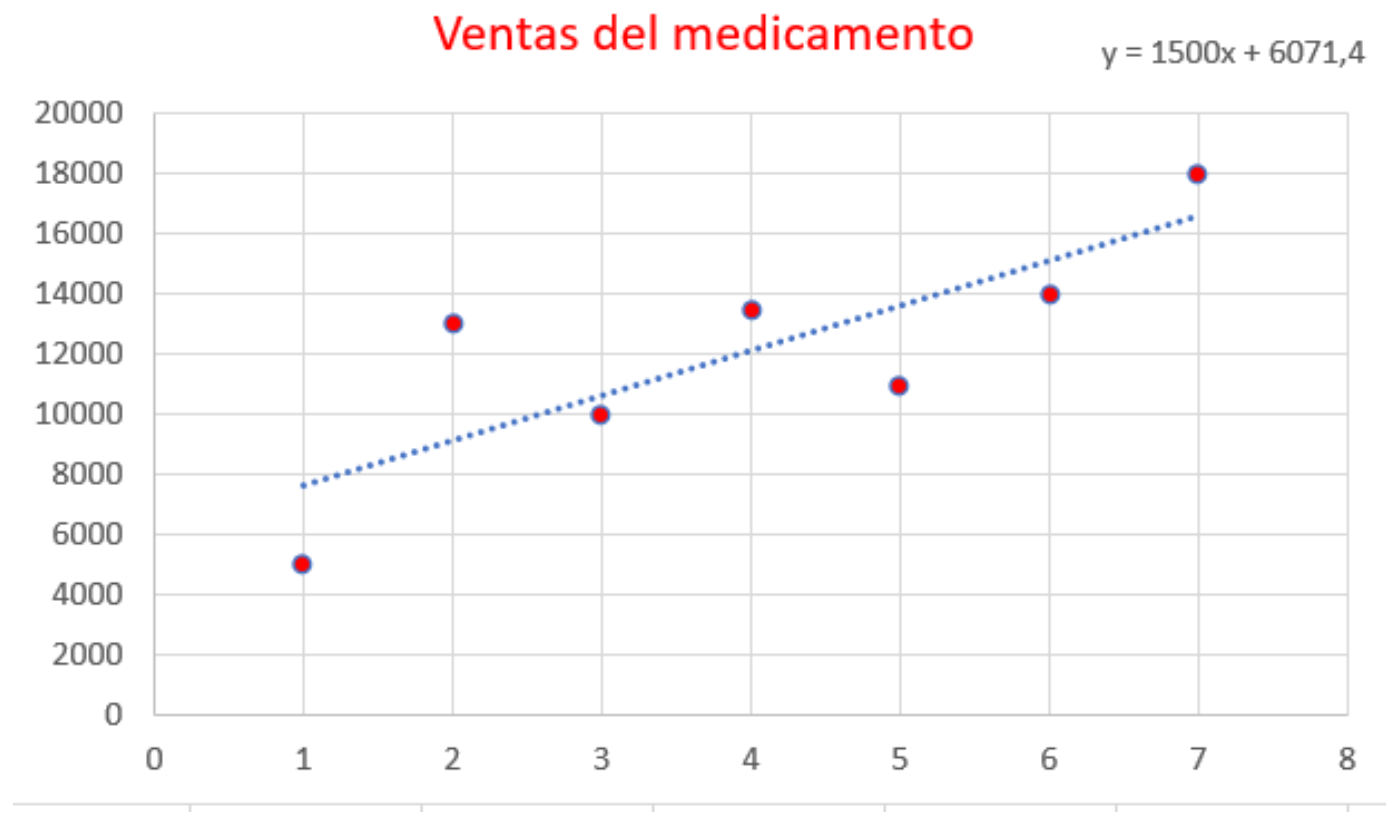
Hallamos los coeficientes a y b:

$$\mathbf{a} = \frac{\begin{vmatrix} \mathbf{SY} & \mathbf{SX} \\ \mathbf{SXY} & \mathbf{SX2} \end{vmatrix}}{\mathbf{D}} = \frac{1}{\mathbf{D}} (\mathbf{SY} \cdot \mathbf{SX2} - \mathbf{SX} \cdot \mathbf{SXY}) =$$

$$\frac{1}{196} [(5000+13000+10000+13500+11000+14000+18000) \cdot (1^2+2^2+3^2+4^2+5^2+6^2+7^2) - (1+2+3+4+5+6+7) \cdot (1*5000 + 2*13000 + 3*10000 + 4*13500 + 5*11000 + 6*14000 + 7*18000)] = 6071{,}4$$

$$\mathbf{b} = \frac{\begin{vmatrix} n & \mathbf{SY} \\ \mathbf{SX} & \mathbf{SXY} \end{vmatrix}}{\mathbf{D}} = \frac{1}{\mathbf{D}} (\mathrm{n} \cdot \mathbf{SXY} - \mathbf{SY} \cdot \mathbf{SX}) =$$

$$\frac{1}{196} [7 \cdot (1*5000 + 2*13000 + 3*10000 + 4*13500 + 5*11000 + 6*14000 + 7*18000) - (5000+13000+10000+13500+11000+14000+18000) \cdot (1+2+3+4+5+6+7)] = 1500$$



Por último, calculamos las ganancias previstas para el 2021 sustituyendo la x por 8, pues 2021 es el octavo año de venta:

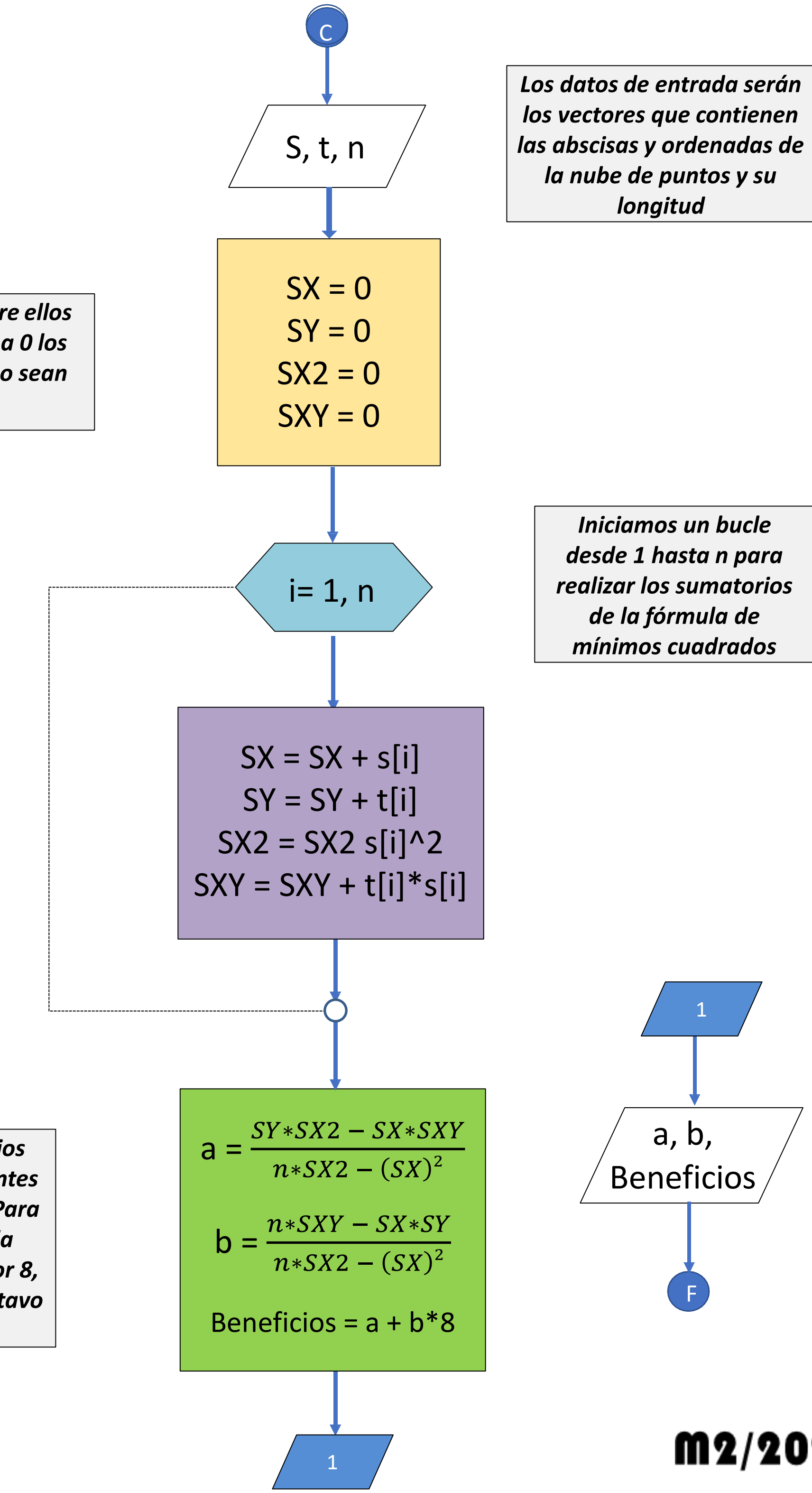
$y = a + bx = 6071{,}4 + 1500*8 =$ **18071,4** € se prevén que consiga la empresa por la venta del medicamento en 2021

M2/2020

ALGORITMO

En algunos lenguajes, entre ellos R, es necesario inicializar a 0 los vectores y matrices que no sean datos iniciales

Hallaremos los beneficios tras obtener los coeficientes mediante las fórmulas. Para ello sustituiremos en la ecuación la variable x por 8, puesto que 2021 es el octavo año de venta.



Los datos de entrada serán los vectores que contienen las abscisas y ordenadas de la nube de puntos y su longitud

Iniciamos un bucle desde 1 hasta n para realizar los sumatorios de la fórmula de mínimos cuadrados

M2/2020