# LA RECTA DE LOS MÍNIMOS CUADRADOS

**Autor:** Mario Suárez

Se llama línea de mejor ajuste y se define como la línea que hace mínima la suma de los cuadrados de las desviaciones respecto a ella de todos los puntos que corresponden a la información recogida.

La recta de los mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots \dots (X_N, Y_N)$ tomando en cuenta a Y como variable dependiente tiene por ecuación

$$Y = a_0 + a_1 X$$

A esta ecuación suele llamarse recta de regresión de $Y$ sobre $X$, y se usa para estimar los valores de $Y$ para valores dados de $X$

Si a la recta de regresión $Y = a_0 + a_1 X$ se le suma en ambos lados $\sum Y = \sum(a_0 + a_1 X)$ se obtiene $\sum Y = a_0 N + a_1 \sum X$

Si a la recta de regresión $Y = a_0 + a_1 X$ se multiplica por $X$ a ambos lados y luego se suma $\sum X\,Y = \sum X(a_0 + a_1 X)$ se obtiene $\sum X\,Y = a_0 \sum X + a_1 \sum X^2$

Las constantes $a_0 \; y \; a_1$ quedan fijadas al resolver simultáneamente las ecuaciones anteriormente encontradas, es decir, al resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases}$$

Que se llaman las ecuaciones normales para la recta de mínimos cuadrados.

Las constantes $a_0$ y $a_1$ de las anteriores ecuaciones también se pueden calcular empleando las siguientes fórmulas:

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2} \qquad a_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$$

Otra ecuación para los mínimos cuadrados para $x = X - \bar{X}$ , $y = Y - \bar{Y}$ de la recta de regresión de Y sobre X es:

$$y = \left(\frac{\sum xy}{\sum \mathrm{x}^2}\right) x$$

La recta de los mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \dots \dots (X_N, Y_N)$ tomando en cuenta a X como variable dependiente tiene por ecuación:

$$X = b_0 + b_1 Y$$

A esta ecuación suele llamarse recta de regresión de X sobre Y, y se usa para estimar los valores de X para valores dados de Y. Las constantes $b_0$ y $b_1$ quedan fijadas al resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

Las constantes $b_0$ y $b_1$ del sistema de ecuaciones anterior se pueden calcular empleando las siguientes fórmulas:

$$b_0 = \frac{\sum X \cdot \sum Y^2 - \sum Y \cdot \sum XY}{N\sum Y^2 - (\sum Y)^2} \qquad b_1 = \frac{N\sum XY - \sum X \cdot \sum Y}{N\sum Y^2 - (\sum Y)^2}$$

Otra ecuación para los mínimos cuadrados para $x = X - \bar{X}$ , $y = Y - \bar{Y}$ es:

$$x = \left(\frac{\sum xy}{\sum \mathrm{y}^2}\right) y$$

El punto de intersección entre las rectas $Y = a_0 + a_1 X$ con $X = b_0 + b_1 Y$ se simboliza $(\bar{X}, \bar{Y})$ y se llama centroide o centro de gravedad

**Ejemplo ilustrativo**

Con los datos de la siguiente tabla sobre la altura en centímetros (X) y los pesos en kilogramos (Y) de una muestra de 8 estudiantes varones tomada al azar del segundo semestre de una universidad.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 152 | 157 | 162 | 167 | 173 | 178 | 182 | 188 |
| Y | 56 | 61 | 67 | 72 | 70 | 72 | 83 | 92 |

1) Ajustar la recta de mínimos cuadrados para Y como variable dependiente resolviendo el sistema:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases}$$

2) Ajustar la recta de mínimos cuadrados para Y como variable dependiente empleando las fórmulas:

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N\sum X^2 - (\sum X)^2} \qquad a_1 = \frac{N\sum XY - \sum X \cdot \sum Y}{N\sum X^2 - (\sum X)^2}$$

3) Ajustar la recta de mínimos cuadrados para Y como variable dependiente empleando la fórmula:

$$y = \left(\frac{\sum xy}{\sum \mathrm{x}^2}\right) x$$

4) Ajustar la recta de mínimos cuadrados para X como variable dependiente resolviendo el sistema:

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

5) Calcular el punto centroide.

6) Elaborar el diagrama de dispersión. Y en el mismo diagrama graficar las dos rectas de mínimos cuadrados obtenidas en los pasos anteriores.

7) Estimar el valor de Y cuando X = 200 en el diagrama de dispersión de Y como variable dependiente.

8) Estimar el valor de X cuando Y= 100 en el diagrama de dispersión X como variable dependiente.

**Solución:**

Se llena la siguiente tabla:

| $X$ | $Y$ | $XY$ | $X^2$ | $Y^2$ |
|---|---|---|---|---|
| 152 | 56 | 8512 | 23104 | 3136 |
| 157 | 61 | 9577 | 24649 | 3721 |
| 162 | 67 | 10854 | 26244 | 4489 |
| 167 | 72 | 12024 | 27889 | 5184 |
| 173 | 70 | 12110 | 29929 | 4900 |
| 178 | 72 | 12816 | 31684 | 5184 |
| 182 | 83 | 15106 | 33124 | 6889 |
| 188 | 92 | 17296 | 35344 | 8464 |
| $\Sigma X = 1359$ | $\Sigma Y = 573$ | $\Sigma XY = 98295$ | $\Sigma X^2 = 231967$ | $\Sigma Y^2 = 41967$ |

1) Remplazando valores en el sistema se tiene:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases} \Rightarrow \begin{matrix} 573 = a_0 \cdot 8 + a_1 \cdot 1359 \\ 98295 = a_0 \cdot 1359 + a_1 \cdot 231967 \end{matrix} \Rightarrow \begin{cases} 8a_0 + 1359a_1 = 573 \\ 1359a_0 + 231967a_1 = 98295 \end{cases}$$

Resolviendo el sistema por determinantes (regla de Cramer) se obtiene:

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{\begin{vmatrix} 573 & 1359 \\ 98295 & 231967 \end{vmatrix}}{\begin{vmatrix} 8 & 1359 \\ 1359 & 231967 \end{vmatrix}} = \frac{573 \cdot 231967 - 98295 \cdot 1359}{8 \cdot 231967 - 1359 \cdot 1359} = \frac{-665814}{8855} = -75{,}191$$

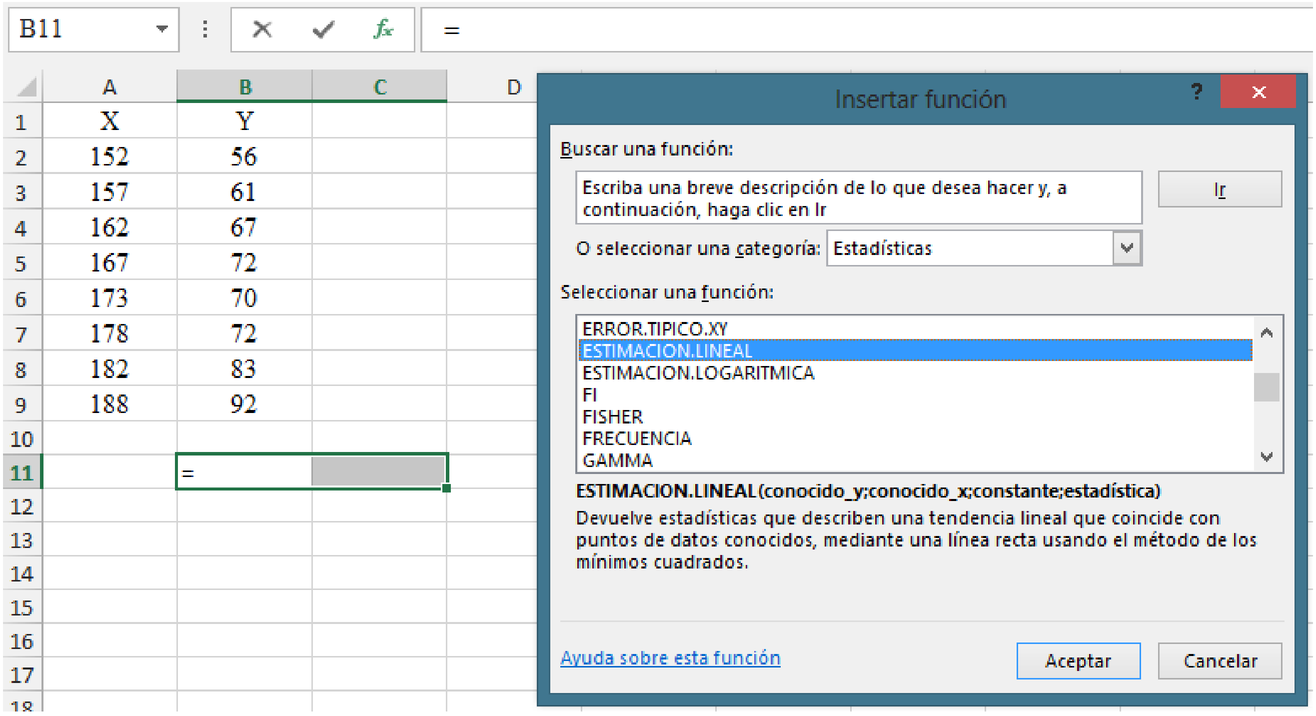
$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{\begin{vmatrix} 8 & 573 \\ 1359 & 98295 \end{vmatrix}}{8855} = \frac{8 \cdot 98295 - 1359 \cdot 573}{8855} = \frac{7653}{8855} = 0{,}864$$

*Para calcular los valores de $a_1$ y $a_0$ en Excel se calcula de la siguiente manera:*

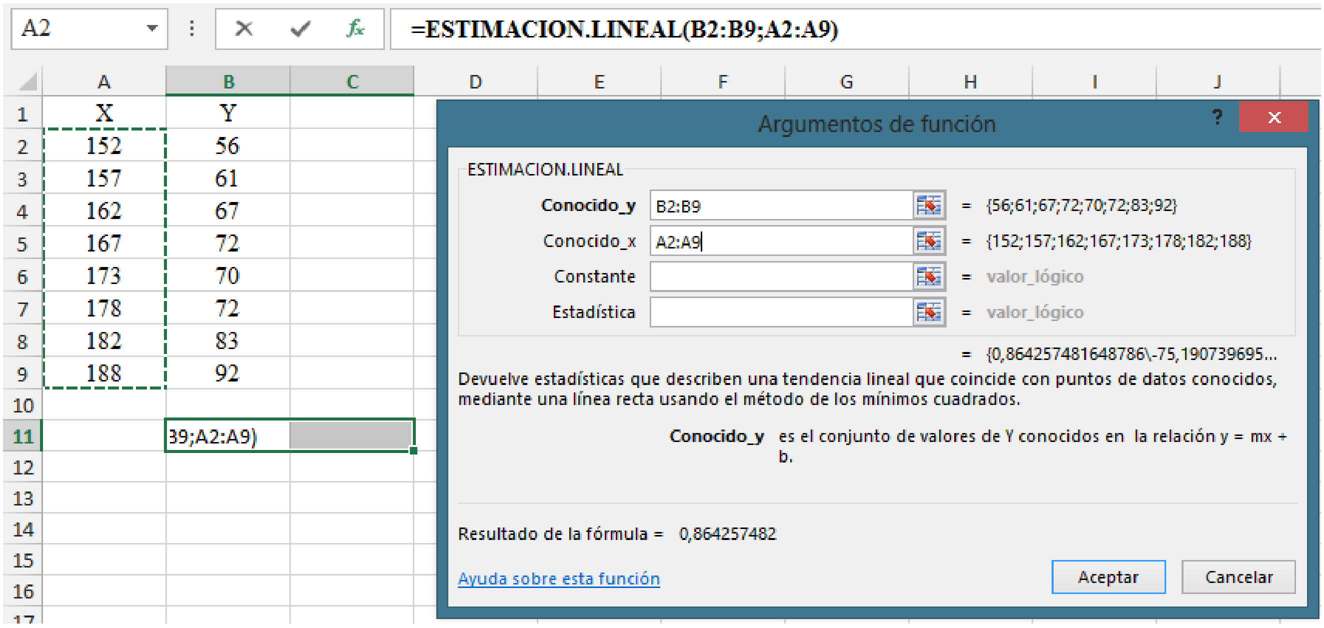
a) Escribir los datos. Seleccionar las celdas donde aparecerá la respuesta

| | A | B | C |
|---|---|---|---|
| 1 | X | Y | |
| 2 | 152 | 56 | |
| 3 | 157 | 61 | |
| 4 | 162 | 67 | |
| 5 | 167 | 72 | |
| 6 | 173 | 70 | |
| 7 | 178 | 72 | |
| 8 | 182 | 83 | |
| 9 | 188 | 92 | |
| 10 | | | |
| 11 | | | |

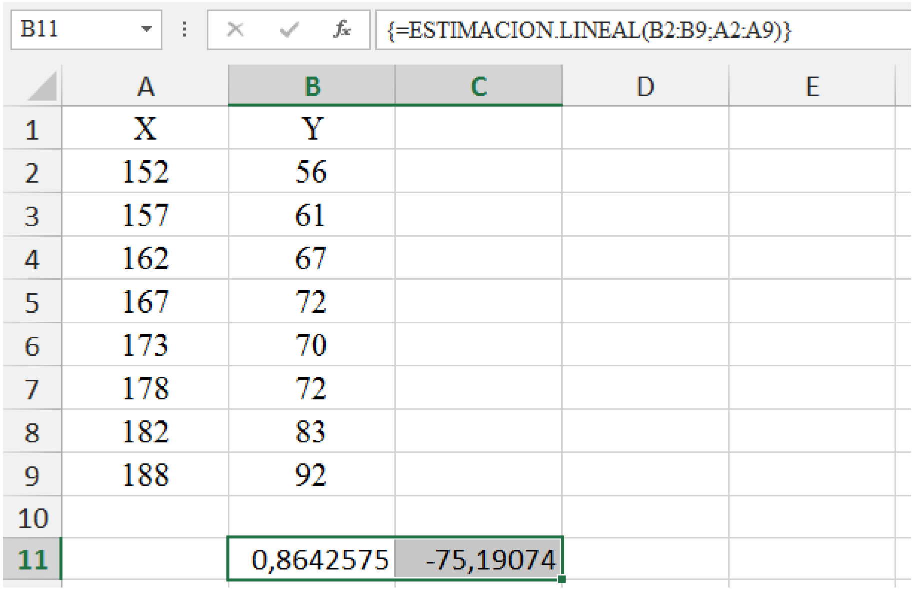
b) Insertar la función ESTIMACION.LINEAL




c) En la ventana Argumentos de Función, en la casilla Conocido_y seleccione los datos de Y, es decir, B2:B9 y en la casilla Conocido_x seleccione los datos de X, es decir, A2:A9

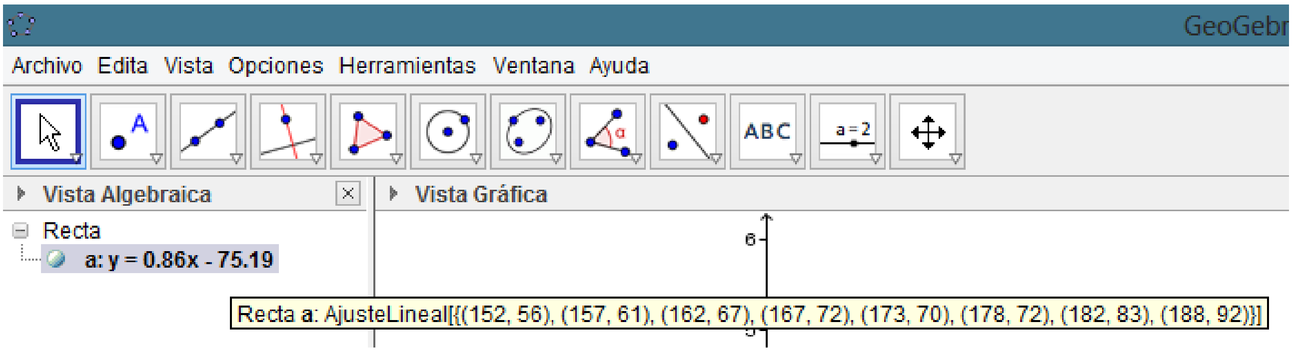


d) Presione CTRL+SHIFT+ENTER


B11 {=ESTIMACION.LINEAL(B2:B9;A2:A9)}

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | X | Y | | | |
| 2 | 152 | 56 | | | |
| 3 | 157 | 61 | | | |
| 4 | 162 | 67 | | | |
| 5 | 167 | 72 | | | |
| 6 | 173 | 70 | | | |
| 7 | 178 | 72 | | | |
| 8 | 182 | 83 | | | |
| 9 | 188 | 92 | | | |
| 10 | | | | | |
| 11 | | 0,8642575 | -75,19074 | | |

*Los cálculos en GeoGebra se muestran en la siguiente figura:*




Remplazando valores en la ecuación respectiva se obtiene:

$Y = a_0 + a_1X \Rightarrow Y = -75{,}191 + 0{,}864X$

**Interpretación:**

- El valor $a_1 = 0{,}864$ indica que la recta tiene una pendiente positiva aumentando a razón de 0,864

- El valor de $a_0 = -75{,}191$ indica el punto en donde la recta interseca al eje Y cuanto X = 0

2) Con los datos de la tabla anterior se substituye valores en las siguientes ecuaciones:

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N\sum X^2 - (\sum X)^2} = \frac{573 \cdot 231967 - 1359 \cdot 98295}{8 \cdot 231967 - (1359)^2} = \frac{-665814}{8855} - 75{,}191$$

$$a_1 = \frac{N\sum XY - \sum X \cdot \sum Y}{N\sum X^2 - (\sum X)^2} = \frac{8 \cdot 98295 - 1359 \cdot 573}{8 \cdot 231967 - (1359)^2} = \frac{7653}{8855} = 0{,}864$$

Remplazando valores en la ecuación respectiva se obtiene:

$Y = a_0 + a_1X \Rightarrow Y = -75{,}191 + 0{,}864X$

3) Se calcula las medias aritméticas de X y Y para llenar la siguiente tabla:

$$\bar{X} = \frac{1359}{8} = 169{,}875\ ;\ \bar{Y} = \frac{573}{8} = 71{,}625$$

| $X$ | $Y$ | $x = X - \bar{X}$ | $y = Y - \bar{Y}$ | $xy$ | $x^2$ | $y^2$ |
|---|---|---|---|---|---|---|
| 152 | 56 | -17,88 | -15,625 | 279,297 | 319,516 | 244,141 |
| 157 | 61 | -12,88 | -10,625 | 136,797 | 165,766 | 112,891 |
| 162 | 67 | -7,875 | -4,625 | 36,422 | 62,016 | 21,391 |
| 167 | 72 | -2,875 | 0,375 | -1,078 | 8,266 | 0,141 |
| 173 | 70 | 3,125 | -1,625 | -5,078 | 9,766 | 2,641 |
| 178 | 72 | 8,125 | 0,375 | 3,047 | 66,016 | 0,141 |
| 182 | 83 | 12,125 | 11,375 | 137,922 | 147,016 | 129,391 |
| 188 | 92 | 18,125 | 20,375 | 369,297 | 328,516 | 415,141 |
| $\Sigma X = 1359$ | $\Sigma Y = 573$ | | | $\Sigma xy = 956{,}625$ | $\Sigma x^2 = 1106{,}875$ | $\Sigma y^2 = 925{,}875$ |

Remplazando valores en la fórmula respectiva se obtiene:

$$y = \left(\frac{\sum xy}{\sum \mathrm{x}^2}\right) x \Rightarrow y = \frac{956{,}625}{1106{,}875} x \Rightarrow Y - \bar{Y} = \frac{956{,}625}{1106{,}875}(X - \bar{X})$$

$$Y - 71{,}625 = \frac{956{,}625}{1106{,}875}(X - 169{,}875) \Rightarrow 1106{,}875(Y - 71{,}625) = 956{,}625(X - 169{,}875)$$

$$1106{,}875Y - 79280{,}20838 = 956{,}625X - 162510{,}4984$$

$$1106{,}875Y = 956{,}625X - 162510{,}4984 + 79280{,}20838$$

$$1106{,}875Y = 956{,}625X - 83230{,}29$$

$$Y = \frac{956{,}625X - 83230{,}29}{1106{,}875} \Rightarrow Y = \frac{956{,}625X}{1106{,}875} - \frac{83230{,}29}{1106{,}875} \Rightarrow Y = 0{,}864X - 75{,}19$$

$$Y = -75{,}19 + 0{,}864X$$

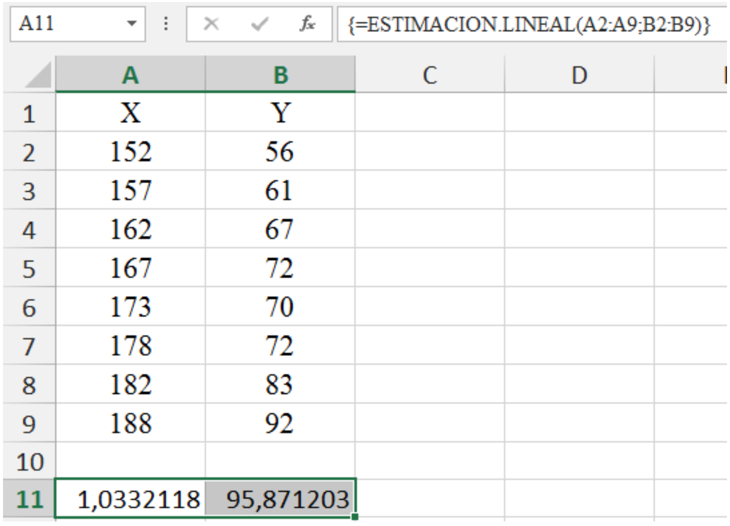
4) Remplazando valores en sistema respectivo se obtiene:

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases} \Rightarrow \begin{matrix} 1359 = b_0 \cdot 8 + b_1 \cdot 573 \\ 98295 = b_0 \cdot 573 + b_1 \cdot 41967 \end{matrix} \Rightarrow \begin{matrix} 8b_0 + 573b_1 = 1359 \\ 573b_0 + 41967b_1 = 98295 \end{matrix}$$

Resolviendo el sistema se obtiene:

$$b_0 = 95{,}871;\ b_1 = 1{,}033$$

*Los cálculos en Excel se muestran en la siguiente figura:*



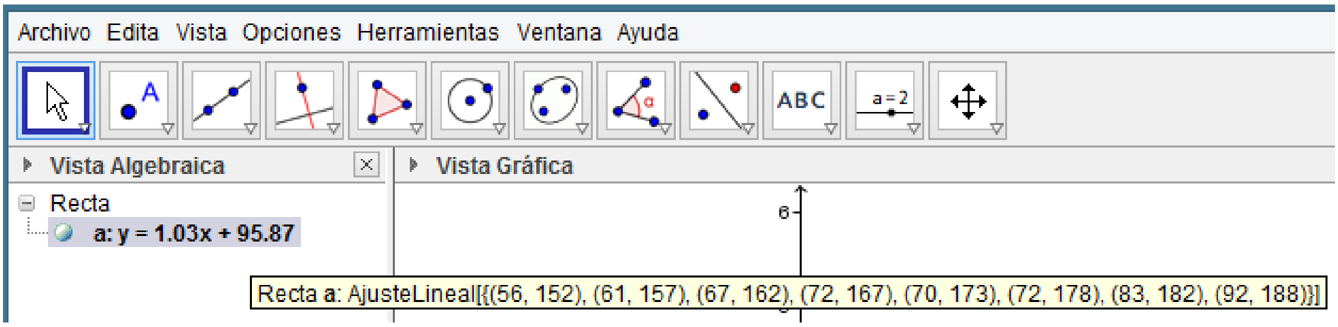
A11 {=ESTIMACION.LINEAL(A2:A9;B2:B9)}

| | A | B | C | D |
|---|---|---|---|---|
| 1 | X | Y | | |
| 2 | 152 | 56 | | |
| 3 | 157 | 61 | | |
| 4 | 162 | 67 | | |
| 5 | 167 | 72 | | |
| 6 | 173 | 70 | | |
| 7 | 178 | 72 | | |
| 8 | 182 | 83 | | |
| 9 | 188 | 92 | | |
| 10 | | | | |
| 11 | 1,0332118 | 95,871203 | | |

Remplazando valores en la ecuación de la recta de mínimos cuadrados se obtiene:

$X = b_0 + b_1 Y \Longrightarrow X = 95{,}871 + 1{,}033Y$

*Los cálculos en GeoGebra insertando Ajuste Lineal se muestran en la siguiente figura:*




**Interpretación:**

- El valor $b_1 = 1{,}033$ indica que la recta tiene una pendiente positiva aumentando a razón de 1,033
- El valor de $b_0 = 95{,}871$ indica el punto en donde la recta interseca al eje X cuanto Y = 0

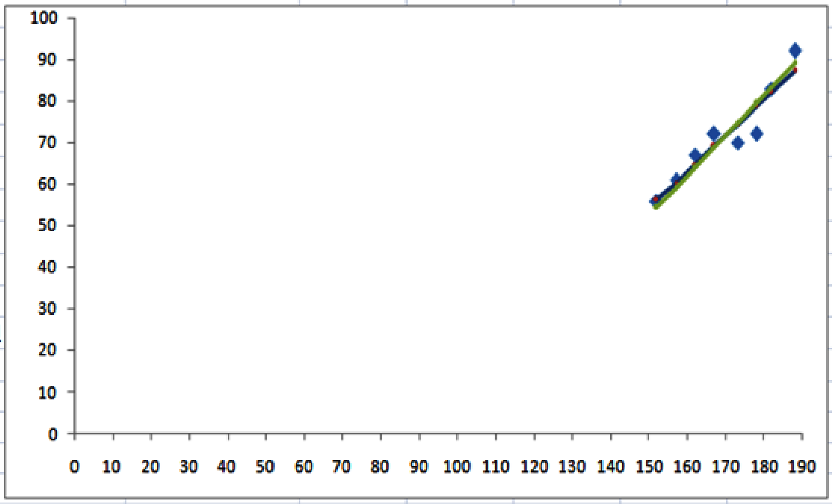
5) Para calcular el centroide $(\bar{X}, \bar{Y})$ se resuelve el sistema formado por las dos rectas de los mínimos cuadrados en donde X es $\bar{X}$ y Y es $\bar{Y}$.

$$\begin{cases} Y = -75{,}191 + 0{,}864X \\ X = 95{,}871 + 1{,}033Y \end{cases}$$

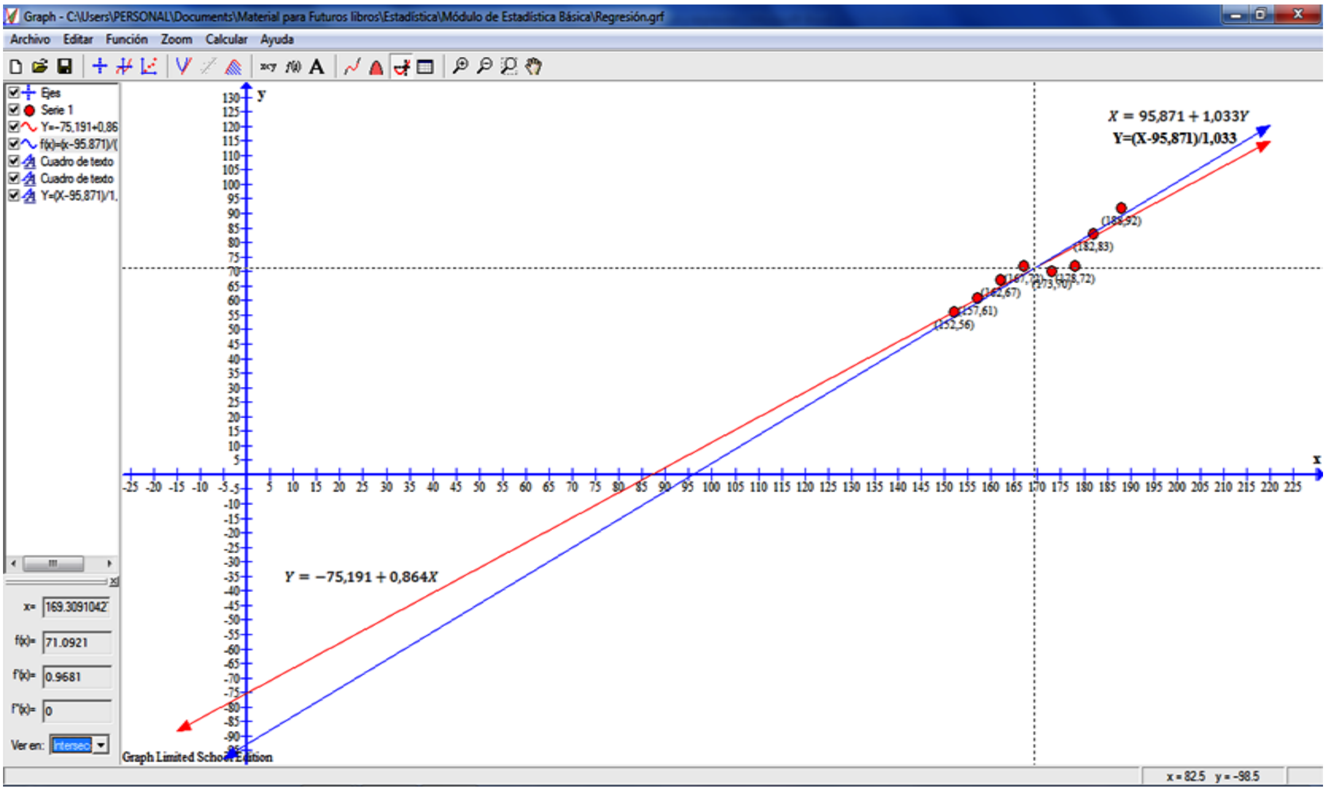
Al resolver el sistema se obtiene el centroide: X = 169,3 y Y = 71,092

6) *En Excel, insertando gráfico de dispersión se obtiene la siguiente figura:*

| X | Y | Y=-75,191+0,864X | y=(X- 95,871)/1,033 | |
|---|---|---|---|---|
| 152 | 56 | 56,137 | 54,336 | |
| 157 | 61 | 60,457 | 59,176 | |
| 162 | 67 | 64,777 | 64,016 | |
| 167 | 72 | 69,097 | 68,857 | |
| 173 | 70 | 74,281 | 74,665 | |
| 178 | 72 | 78,601 | 79,505 | |
| 182 | 83 | 82,057 | 83,378 | |
| 188 | 92 | 87,241 | 89,186 | |
| | | | | |
| | $a_0$ | -75,191 | $b_0$ | 95,871 |
| | $a_1$ | 0,864 | $b_1$ | 1,033 |




*Empleando el programa Graph se obtiene la siguiente figura:*




7) Remplazando X = 200 en la ecuación solicitada se obtiene:
$Y = -75{,}191 + 0{,}864X = -75{,}191 + 0{,}864 \cdot 200 = -75{,}191 + 172{,}8 = 97{,}609$

8) Remplazando Y = 100 en la ecuación solicitada se obtiene:
$X = 95{,}871 + 1{,}033Y = X = 95{,}871 + 1{,}033 \cdot 100 = X = 95{,}871 + 103{,}3 = 199{,}171$

## TAREA DE INTERAPRENDIZAJE

1) Consulte sobre la biografía de Francis Galton y de Cramer, y realice un organizador gráfico de cada una.

2) Dada la siguiente tabla sobre la altura en centímetros (X) y los pesos en kilogramos (Y) de una muestra de 8 estudiantes varones tomada al azar del segundo semestre de una universidad.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 150 | 155 | 160 | 165 | 170 | 175 | 180 | 185 |
| Y | 55 | 60 | 63 | 67 | 70 | 74 | 79 | 85 |

2.1) Ajuste la recta de mínimos cuadrados para Y como variable dependiente resolviendo el siguiente sistema y empleando Excel y GeoGebra.

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 \end{cases}$$

$$Y = -66{,}869 + 0{,}812X$$

2.2) Ajuste la recta de mínimos cuadrados para Y como variable dependiente empleando las fórmulas

$$a_0 = \frac{\sum Y \cdot \sum X^2 - \sum X \cdot \sum XY}{N \sum X^2 - (\sum X)^2} \qquad a_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum X^2 - (\sum X)^2}$$

$$a_0 = -66{,}869\,; \quad a_1 = 0{,}812$$

2.3) Ajuste la recta de mínimos cuadrados para Y como variable dependiente empleando la fórmula

$$y = \left(\frac{\sum xy}{\sum \mathrm{x}^2}\right) x$$

$$Y = -66{,}869 + 0{,}812X$$

2.4) Ajuste la recta de mínimos cuadrados para X como variable dependiente resolviendo el siguiente sistema y empleando Excel y GeoGebra.

$$\begin{cases} \Sigma X = b_0 N + b_1 \Sigma Y \\ \Sigma XY = b_0 \Sigma Y + b_1 \Sigma Y^2 \end{cases}$$

$$X = 83{,}18 + 1{,}22Y$$

2.5) Ajuste la recta de mínimos cuadrados para X como variable dependiente empleando las fórmulas

$$b_0 = \frac{\sum X \cdot \sum Y^2 - \sum Y \cdot \sum XY}{N \sum Y^2 - (\sum Y)^2} \qquad b_1 = \frac{N \sum XY - \sum X \cdot \sum Y}{N \sum Y^2 - (\sum Y)^2}$$

$$b_0 = 83.18\,; \quad b_1 = 1{,}22$$

2.6) Ajuste la recta de mínimos cuadrados para X como variable dependiente empleando la fórmula

$$x = \left(\frac{\sum xy}{\sum \mathrm{y}^2}\right) y$$

$$X = 83{,}18 + 1{,}22Y$$

2.7) Calcule el punto centroide.

$$\bar{X} = 170{,}9\,; \ \bar{Y} = 71{,}9$$

2.8) Calcule el coeficiente de determinación.

0,99

2.9) Elabore el diagrama de dispersión. Y en el mismo diagrama graficar las dos rectas de mínimos cuadrados obtenidas en los pasos anteriores. Elabore de manera manual, empleando Excel y el programa Graph.

2.10) Estime el valor de Y cuando X = 173 en el diagrama de dispersión de Y como variable dependiente.

73,6

2.11) Estime el valor de X cuando Y = 73 en el diagrama de dispersión de Y como variable dependiente.

172,2

3) Cree y resuelva un ejercicio similar al anterior con datos obtenidos de 10 amigas suyas.

4) Consulte en la biblioteca o en el internet sobre un ejercicio de aplicación de la rectas de los mínimos cuadrados. Presente ejercicio resuelto en forma manual y empleando Excel y Graph.