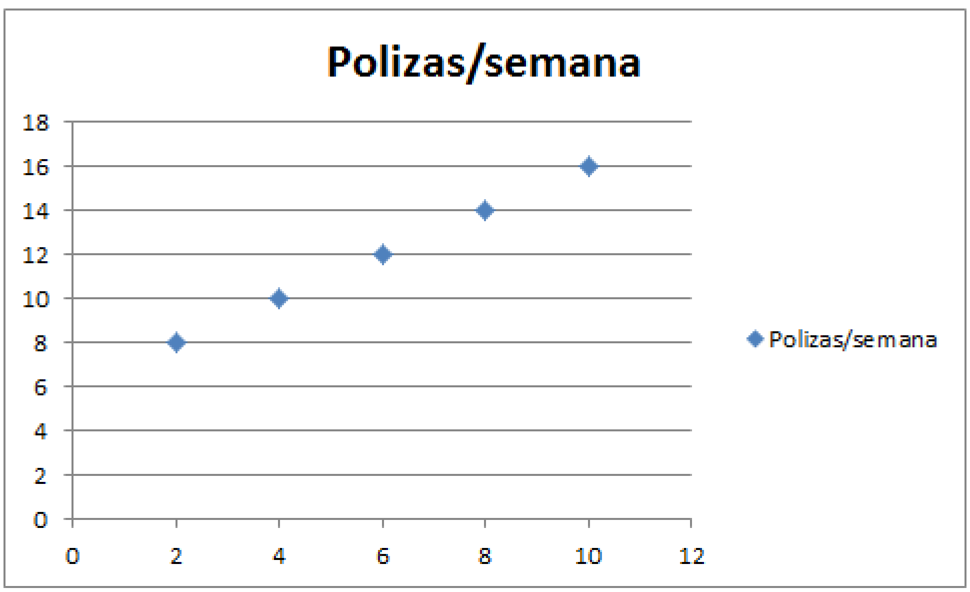
Correlación

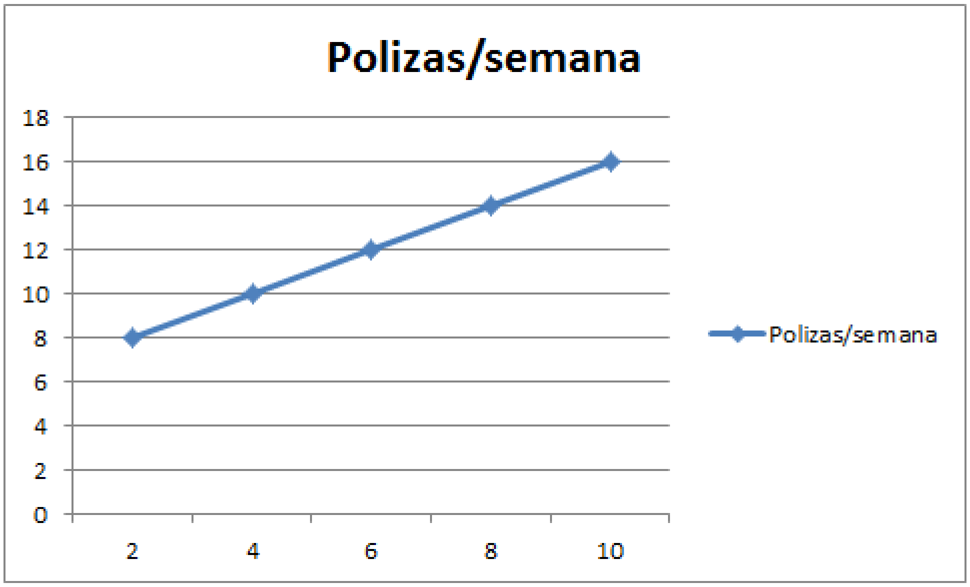
Tabla de datos

	(X)	(Y)(Ventas)
	(años)	
Agentes	Experiencia	Polizas/semana
A	2	8
B	4	10
C	6	12
D	8	14
E	10	16

Diagrama de dispersión



En gráfico superior realizado con Excel, parece sugerir que los datos están en una línea recta. Insertaremos una poligonal utilizando Excel



De nuevo observamos que los datos parecen estar en una recta.

Un análisis algebraico de los datos nos lleva a concluir que los datos están sobre la recta $y = x + 6$.

Al reemplazar en tal recta los valores $x = 2,4,6,8,10$ obtenemos respectivamente los valores $y = 8,10,12,14,16$ que son “exactamente” los valores que da la tabla.

a) Al estar estos valores sobre una recta de pendiente positiva, el coeficiente de correlación es $R^2 = 1$.
   Verifiquemos con la fórmula:
   Como $\bar{x} = \frac{2+4+6+8+10}{5}$=6 , $\bar{y} = \frac{8+10+12+14+16}{5}$=12

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$= \frac{(2-6)(8-12) + (4-6)(10-12) + (6-6)(12-12) + (8-6)(14-12) + (10-6)(16-12)}{4}$$

$$= \frac{(-4)(-4) + (-2)(-2) + (0)(0) + (2)(2) + (4)(4)}{4} = 10$$

$$s_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{4}} = \sqrt{\frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{4}} = \sqrt{10}$$

$$s_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{4}} = \sqrt{\frac{(8-12)^2+(10-12)^2+(12-12)^2+(14-12)^2+(16-12)^2}{4}} = \sqrt{10}$$

Luego el coeficiente de correlación muestral o correlación de Pearson es:

$$r_{xy} = \frac{s_{xy}}{s_x . s_y} = \frac{10}{\sqrt{10}.\sqrt{10}} = \frac{10}{10} = 1$$

Por lo tanto la línea de regresión es la misma ecuación $y = x + 6$. Por lo tanto se espera que cuando el agente tenga 14 años de experiencia $x = 4$, entonces $y = x + 6 = 14 + 6 = 20$. Entonces:

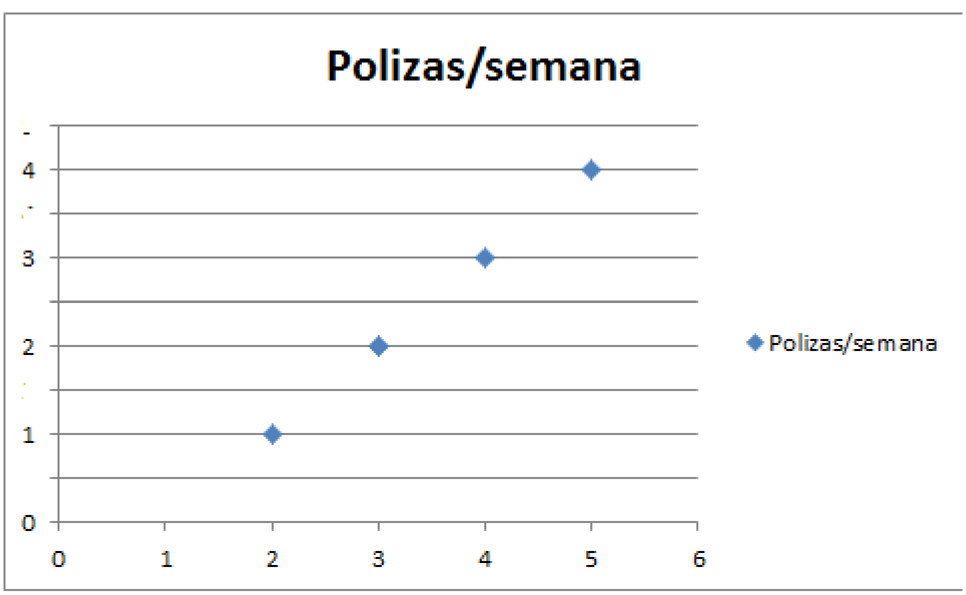
b) Se estima pueda realizar 20 pólizas

2. Se resolverá siguiendo el mismo orden que el ejercicio anterior.

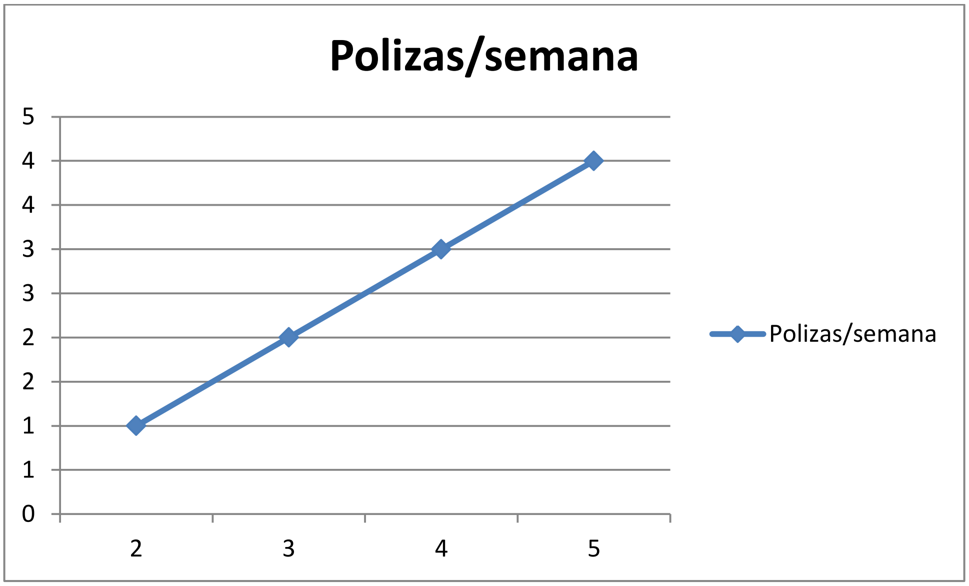
Tabla de datos

	(X)	(Y)(Ventas)
	(años)	
Agentes	Experiencia	Polizas/semana
A	2	1
B	3	2
C	3	2
D	4	3
E	5	4

Diagrama de dispersión



Es evidente que los datos están sobre una recta



Un análisis algebraico de los datos nos lleva a concluir que los datos están sobre la recta $y = x - 1$.

Al reemplazar en tal recta los valores $x = 2{,}3{,}3{,}4{,}5$ obtenemos respectivamente los valores $y = 1{,}2{,}2{,}3{,}4$ que son "exactamente" los valores que da la tabla.

a) La relación de los datos es evidentemente lineal. El grado de asociación lo mide el coeficiente de correlación de Pearson que de nuevo será igual a 1. Es decir que la relación lineal es perfecta.
Verifiquemos:

$\bar{x} = \frac{2+3+3+4+5}{5} = 3{,}4$, $\bar{y} = \frac{1+2+2+3+4}{5} = 2{,}4$

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$
$$= \frac{(2-3{,}4)(1-2{,}4) + (3-3{,}4)(2-2{,}4) + (3-3{,}4)(2-2{,}4) + (4-3{,}4)(3-2{,}4) + (5-3{,}4)(4-2{,}4)}{4}$$
$$= \frac{(-1{,}4)(-1{,}4) + (-0{,}4)(-0{,}4) + (-0{,}4)(-0{,}4) + (0{,}6)(0{,}6) + (1{,}6)(1{,}6)}{4} = 1{,}3$$

$$s_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{4}} = \sqrt{\frac{(2-3{,}4)^2+(3-3{,}4)^2+(3-3{,}4)^2+(4-3{,}4)^2+(5-3{,}4)^2}{4}} = \frac{\sqrt{130}}{10}$$

$$s_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{4}} = \sqrt{\frac{(1-2{,}4)^2+(2-2{,}4)^2+(2-2{,}4)^2+(3-2{,}4)^2+(4-2{,}4)^2}{4}} = \frac{\sqrt{130}}{10}$$

Luego el coeficiente de correlación muestral o correlación de Pearson es:

$$r_{xy} = \frac{s_{xy}}{s_x . s_y} = \frac{1,3}{\frac{\sqrt{130}}{10} \times \frac{\sqrt{130}}{10}} = \frac{1,3}{1,3} = 1$$

Por lo tanto la recta de regresión que ajusta los datos es la misma recta:

$$y = x - 1.$$

El grado de asociación es $r_{xy} = 1$

b) La correlación es lineal “alta” ya que da el máximo valor posible.