UNIVERSIDAD AUTONOMA DEL ESTADO DE MÉXICO

FACULTAD DE INGENIERÍA

MÉTODOS NUMÉRICOS

MODELOS DE REGRESIÓN LINEAL

POR:

LUIS CONRADO TOLEDO VEGA

FECHA DE ELABORACIÓN:

1ra quincena de Septiembre de 2015

Guión explicativo

El presente material didáctico constituye un apoyo para la presentación de los elementos teóricos y prácticos del método de Regresión lineal mediante por Mínimos Cuadrados en la unidad de aprendizaje Métodos Numéricos, perteneciente a la División de Materias Propedéuticas de la Facultad de Ingeniería de la UAEM.

El material proyectable se desarrollo en este año, en el periodo 2015 B, contribuye a la capacidad de comprensión de la metodología que permite ajustar rectas para modelar una serie de datos con tendencia lineal. Para su posterior análisis y explotación en casos prácticos y aplicaciones reales, como parte fundamental en la solución de múltiples problemas que resultan del ejercicio de la Ingeniería.

Para el desarrollo del tema se consultaron diversas fuentes, resultando que donde mejor se explican los conceptos teóricos es en los libros de métodos numéricos incluidos en la bibliografía del programa de la Unidad de Aprendizaje. Por tal razón los teoremas y algoritmos fueron tomados de dichos textos.

Antecedentes

En el análisis de datos de la ingeniería se presentan situaciones en las que se proporciona información con datos numéricos que constituyen un conjunto de puntos discretos de manera tabular.

Antecedentes

Al graficar los datos se construye el diagrama de dispersión y se observa que los datos tienden a ajustarse a la gráfica de una función conocida.

Datos

X	Y
23	36
76	97
65	78
25	78

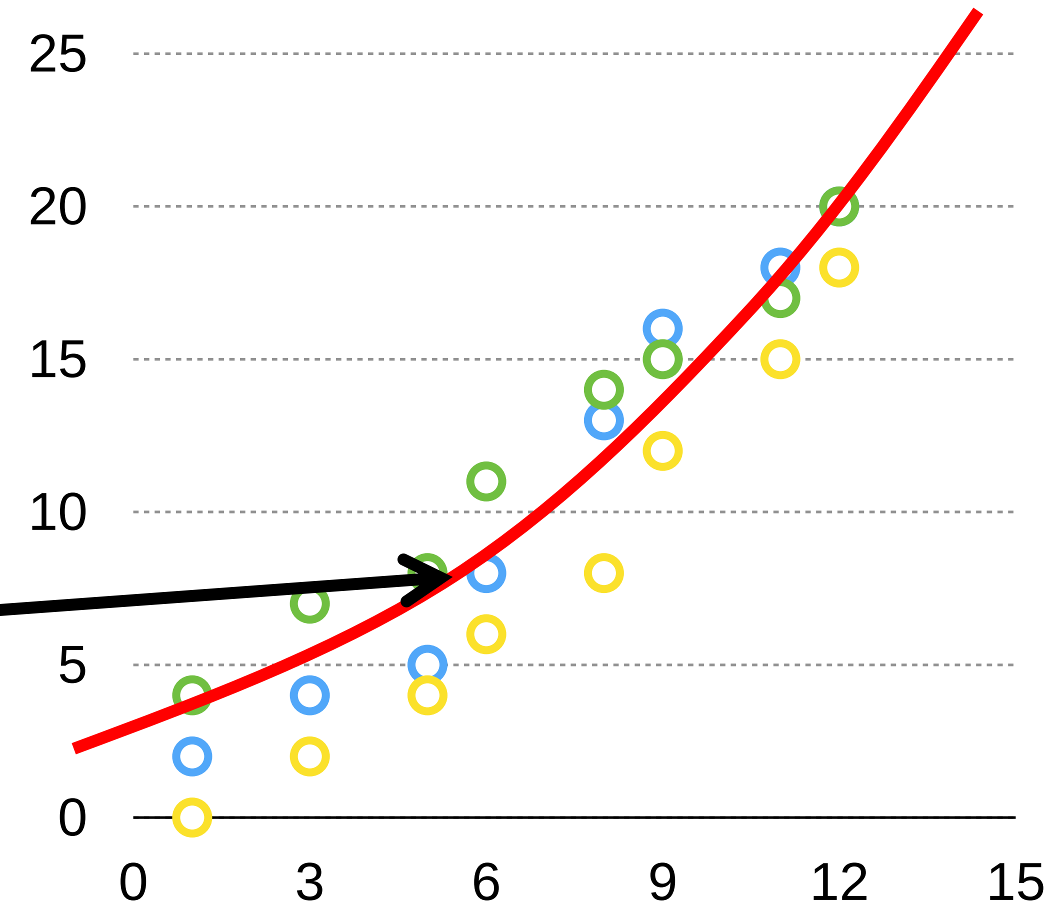
Donde:

X – Variable independiente

Y – Variable dependiente

Diagrama de dispersión

Curva que representa la tendencia general de los datos

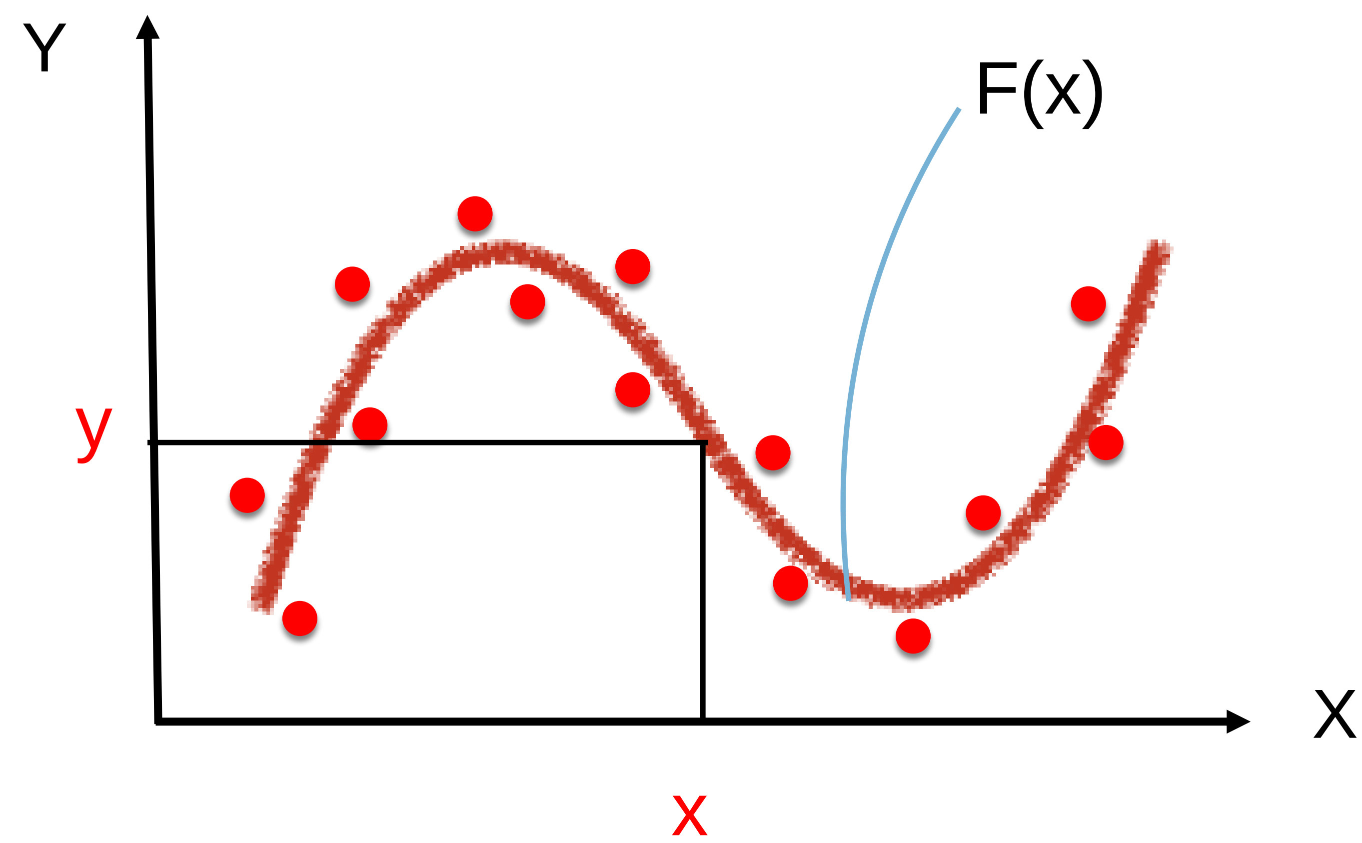


Antecedentes

Algunas veces se requieren de estimaciones entre los valores de X (variable independiente) de Y (variable dependiente) que no estén contenidos en la tabla de datos.

X	Y
23	36
76	97
x	?
65	78
25	78

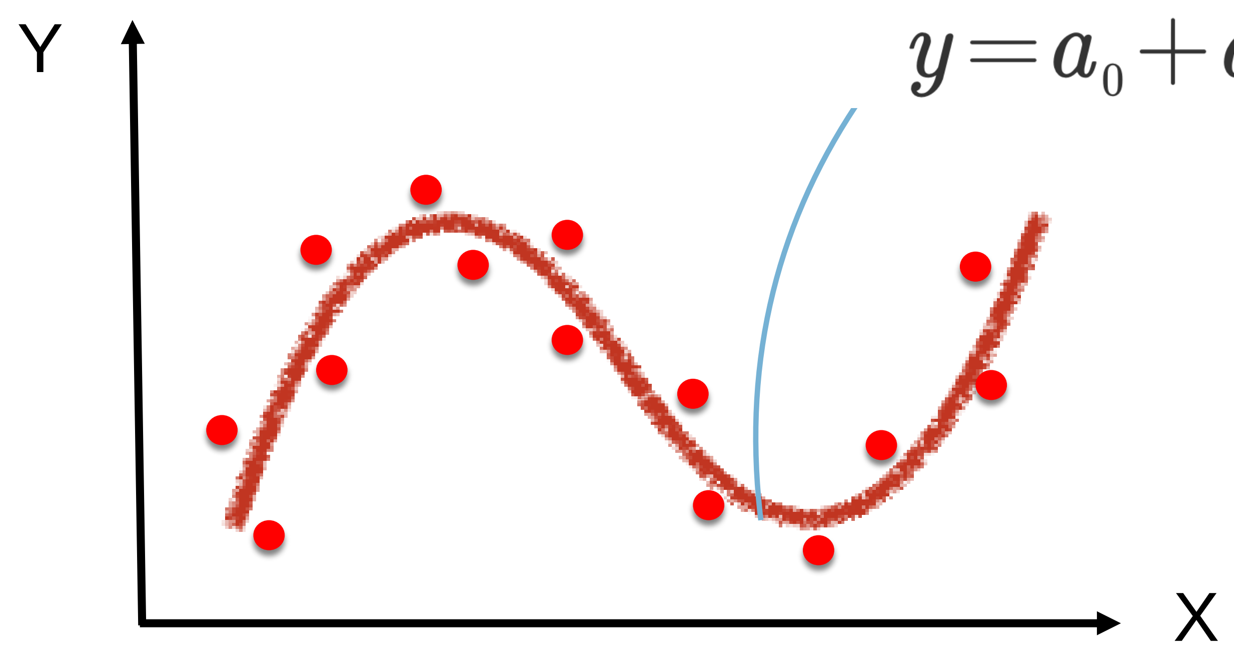
Si x es tal valor,
cuanto es y?



Antecedentes

O bien, se requiere una versión simplificada que represente el comportamiento general de los datos.

Se puede ajustar una expresión matemática que junto con una gráfica representan a los datos de una manera más descriptiva, continua y sintética



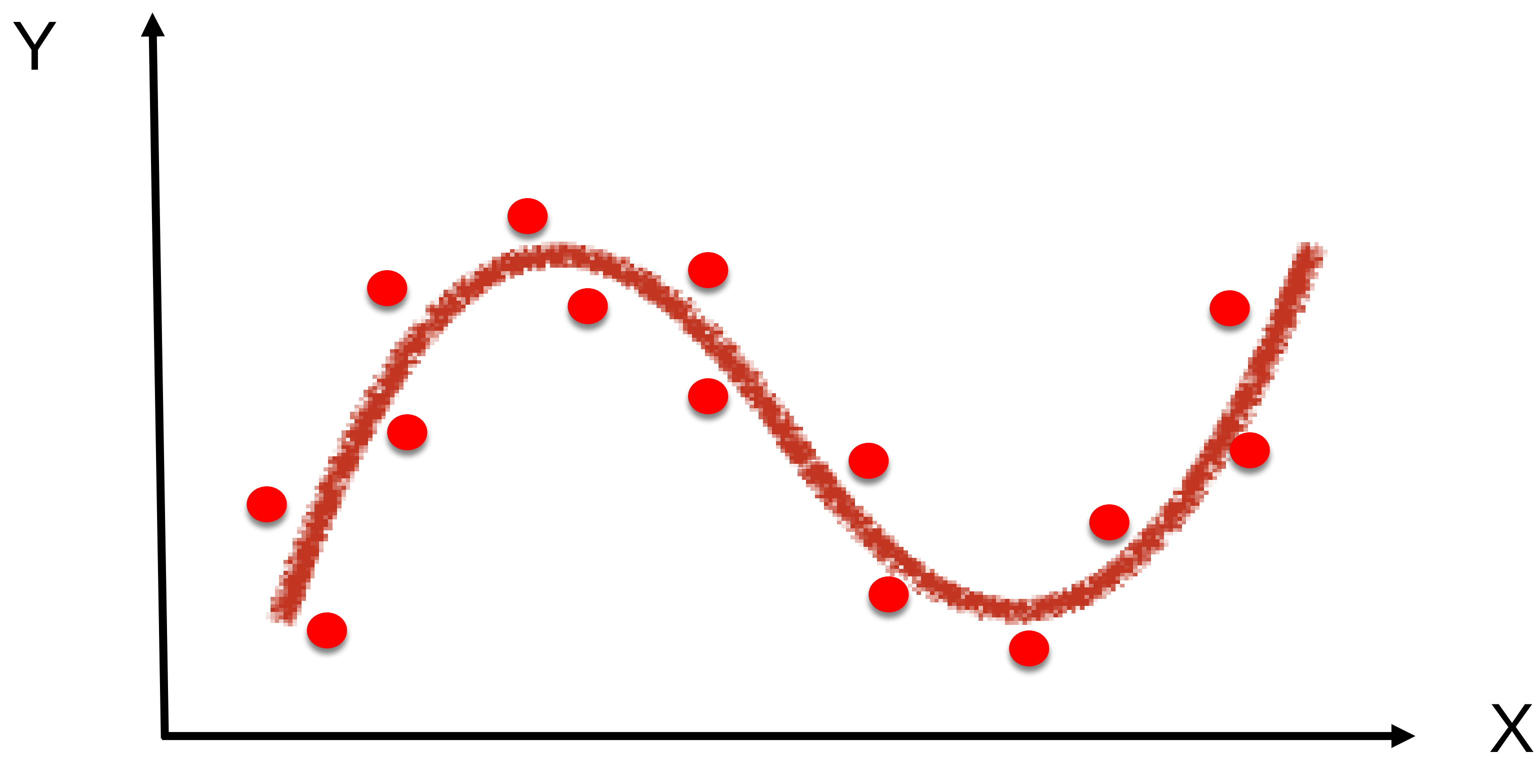
$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$$

$$f\ x = 2x^2 - x^3$$

$$g\ x = 3x$$

Ajuste de curvas

A diferencia de la interpolación la ecuación de la gráfica ajustada, no necesariamente pasa por cada uno los puntos del diagrama de dispersión, solo representará la tendencia general de crecimiento de los datos



Ajuste de curvas

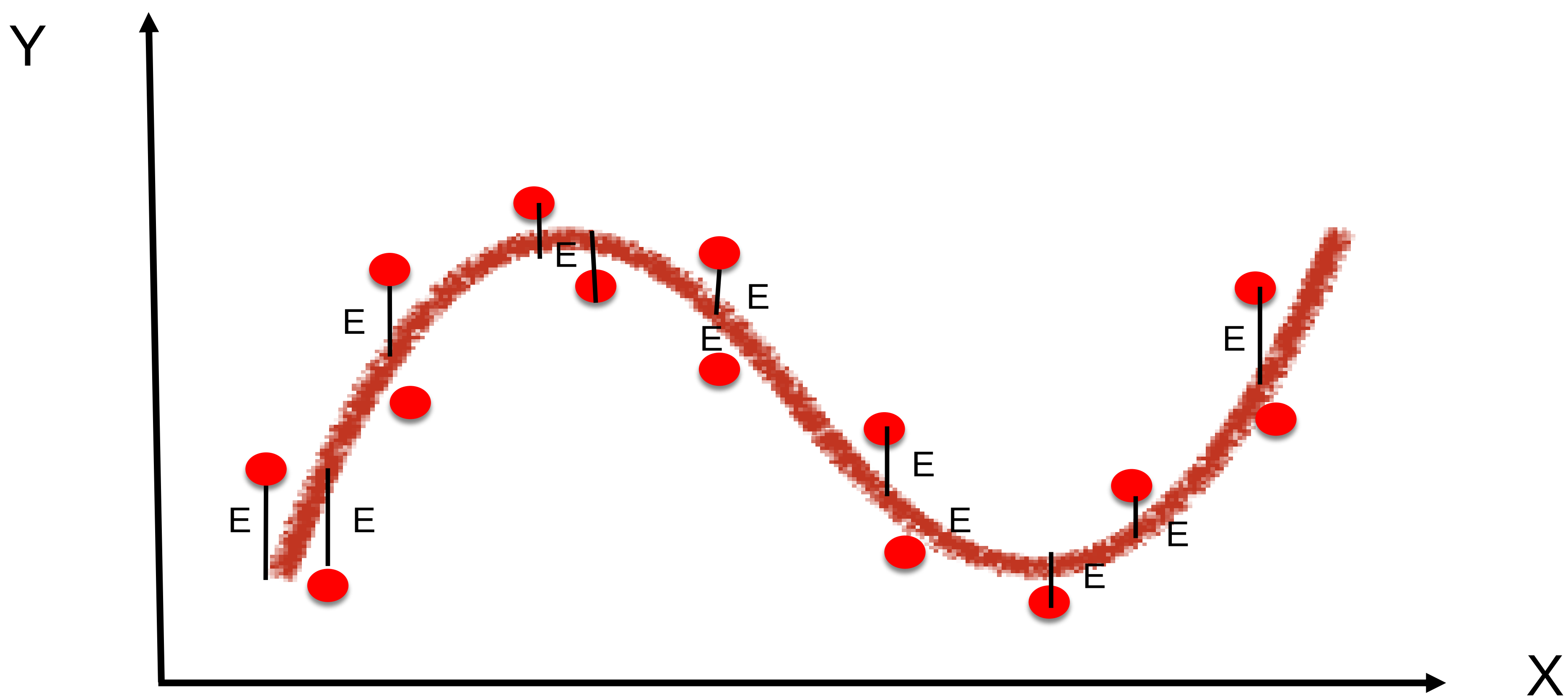
- Modelos de regresión con mínimos cuadrados.

Cuando un conjunto de datos muestren un grado significativo de error, se puede derivar una curva simple siguiendo un patrón sobre los puntos tomados como un todo. Una inspección visual sugiere una relación entre las variables X, Y que pueden ser cualesquiera variables de la ingeniería mecánica.

Por la variabilidad de los datos se obtendrá una expresión matemática que represente el comportamiento general de los datos y que se pueda emplear para propósitos práctico, como para interpolar valores o para hacer pronósticos.

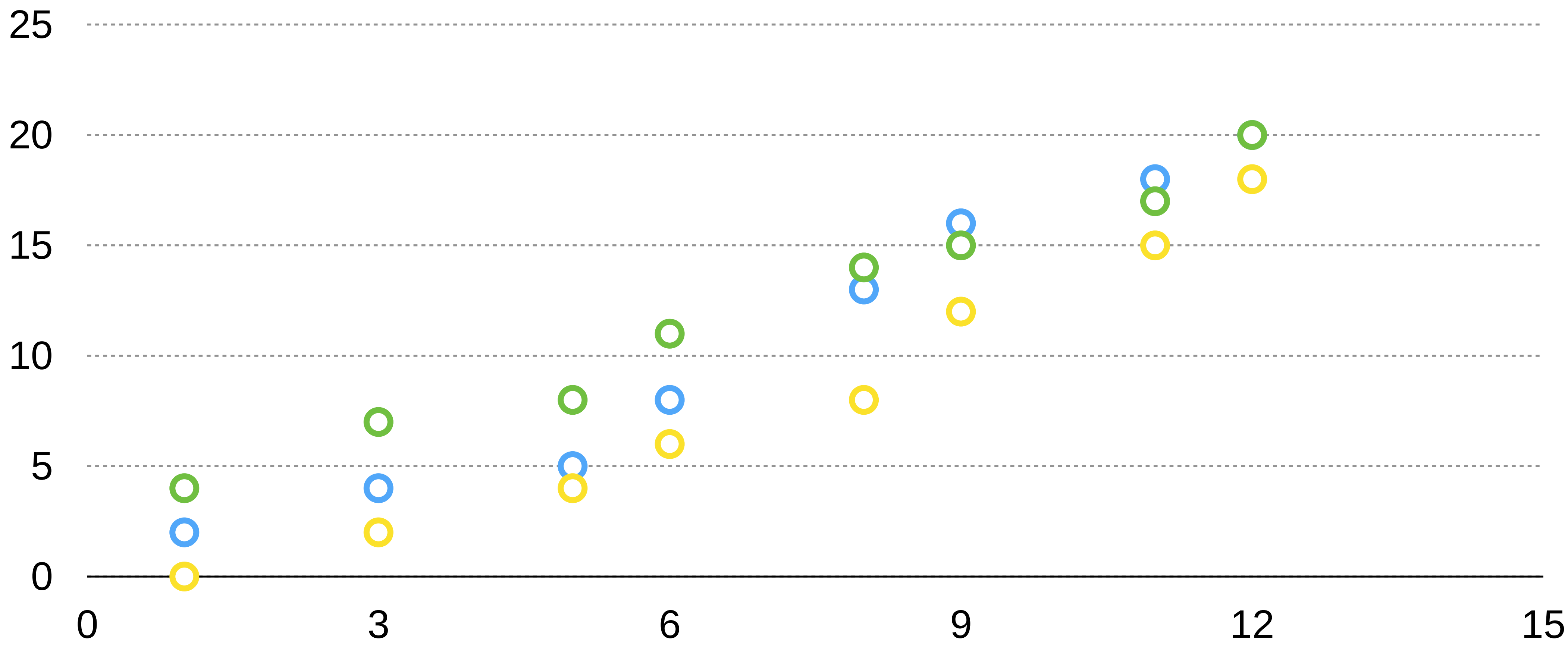
Regresión con mínimos cuadrados

Para encontrar la curva que mejor se ajuste a la tendencia general de crecimiento de los datos, se utilizan técnicas de regresión con mínimos cuadrados. Que consisten en minimizar los errores entre el modelos de ajuste y los datos.

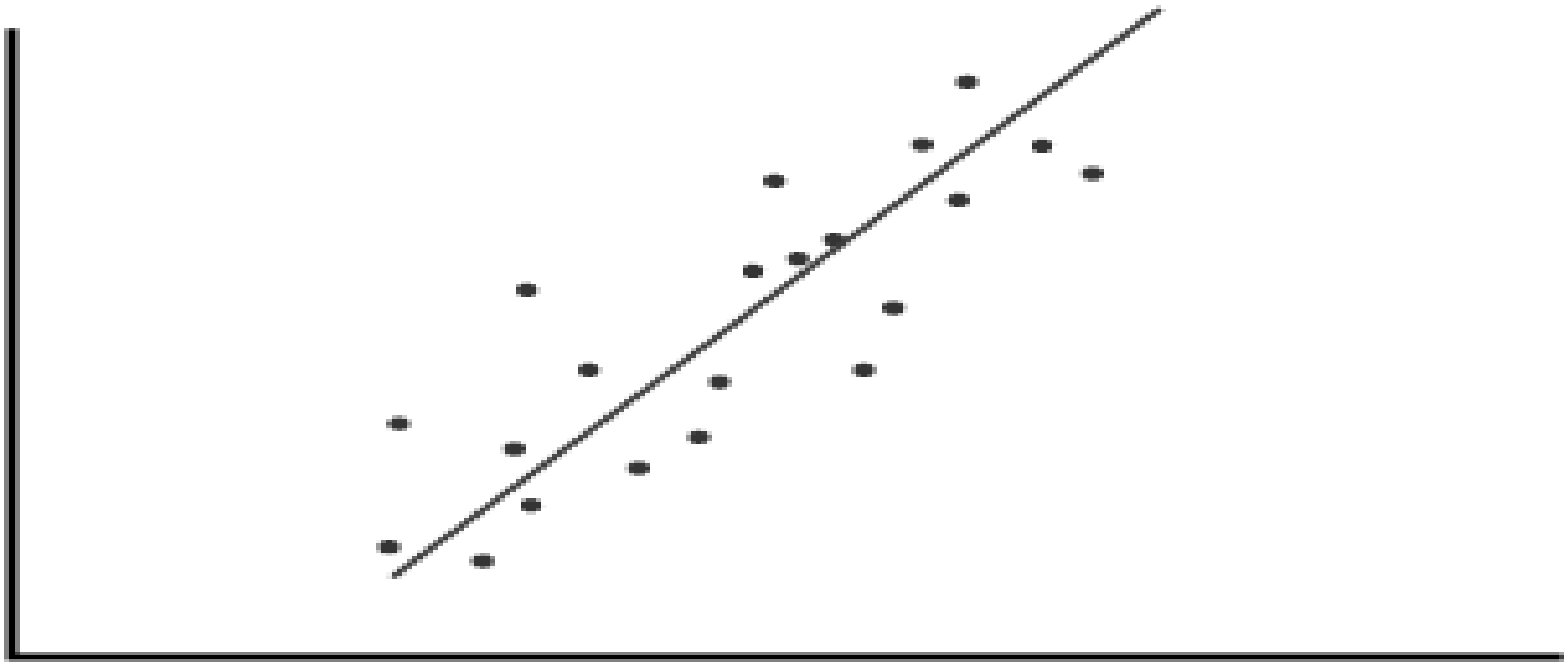


Modelos de regresión lineal

Una inspección visual de los datos siguientes, sugiere la tendencia a alinearse, se observa que a medida que X crece Y lo hace de manera proporcional, pudiendo ajustar una línea recta al comportamiento de los datos.

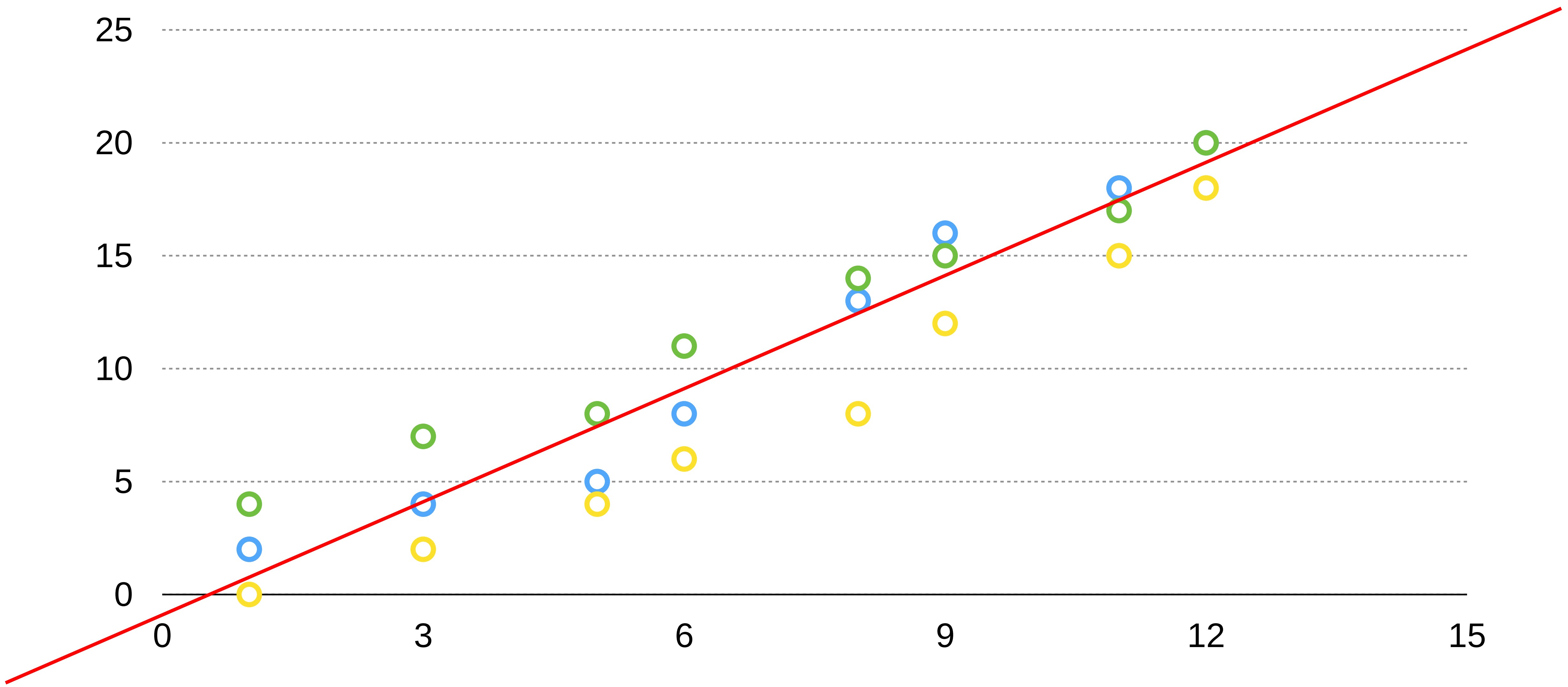


Modelos de regresión lineal



Modelos de regresión lineal

La aproximación más simple para ajustar un conjunto de parejas de datos por mínimos cuadrados es el ajuste de una línea recta.



Modelos de regresión lineal

La expresión de una línea recta es:

$$y = a_0 + a_1 x$$

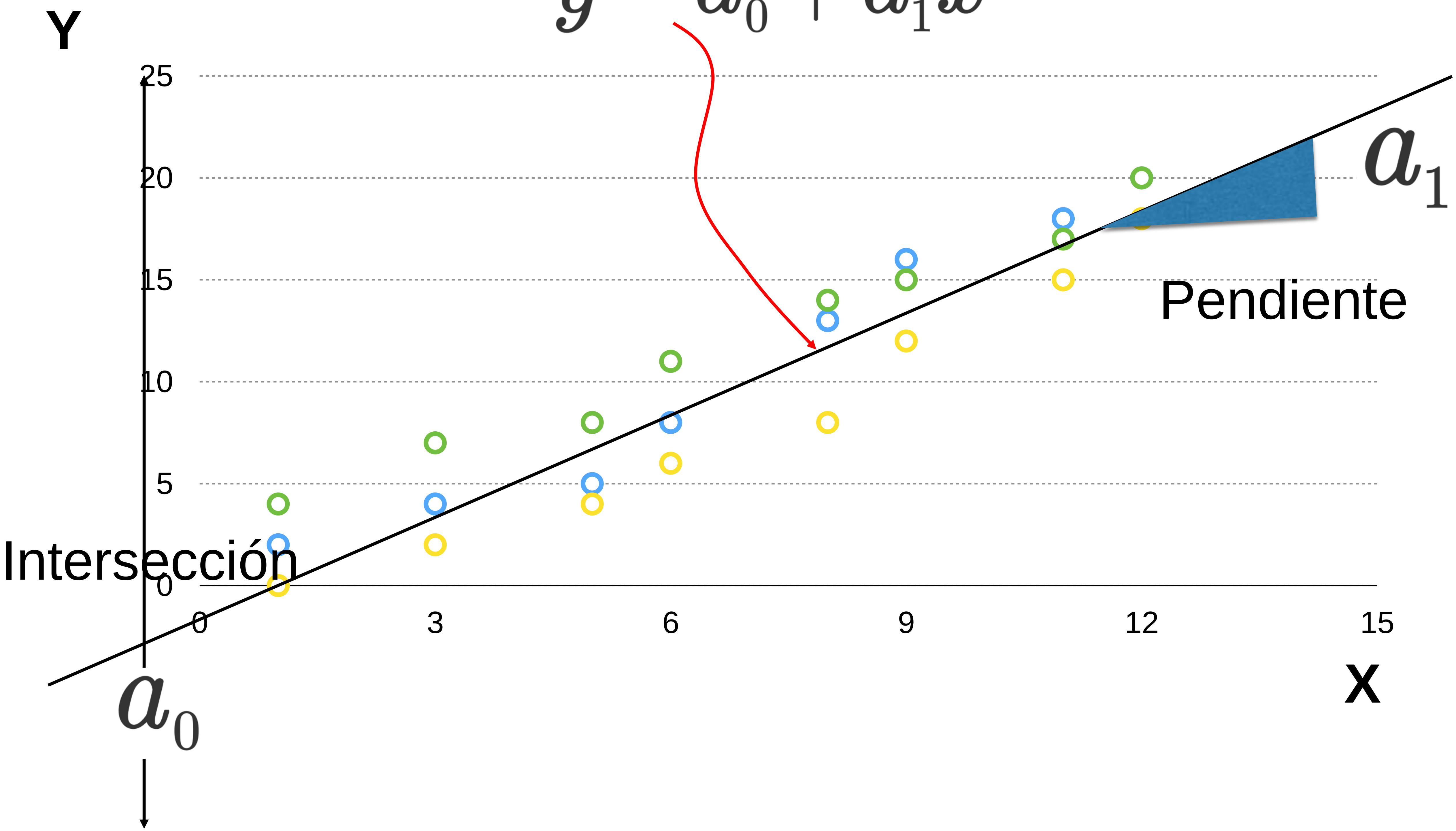
Donde los coeficientes son:

a_0 Representa la intersección con el eje Y

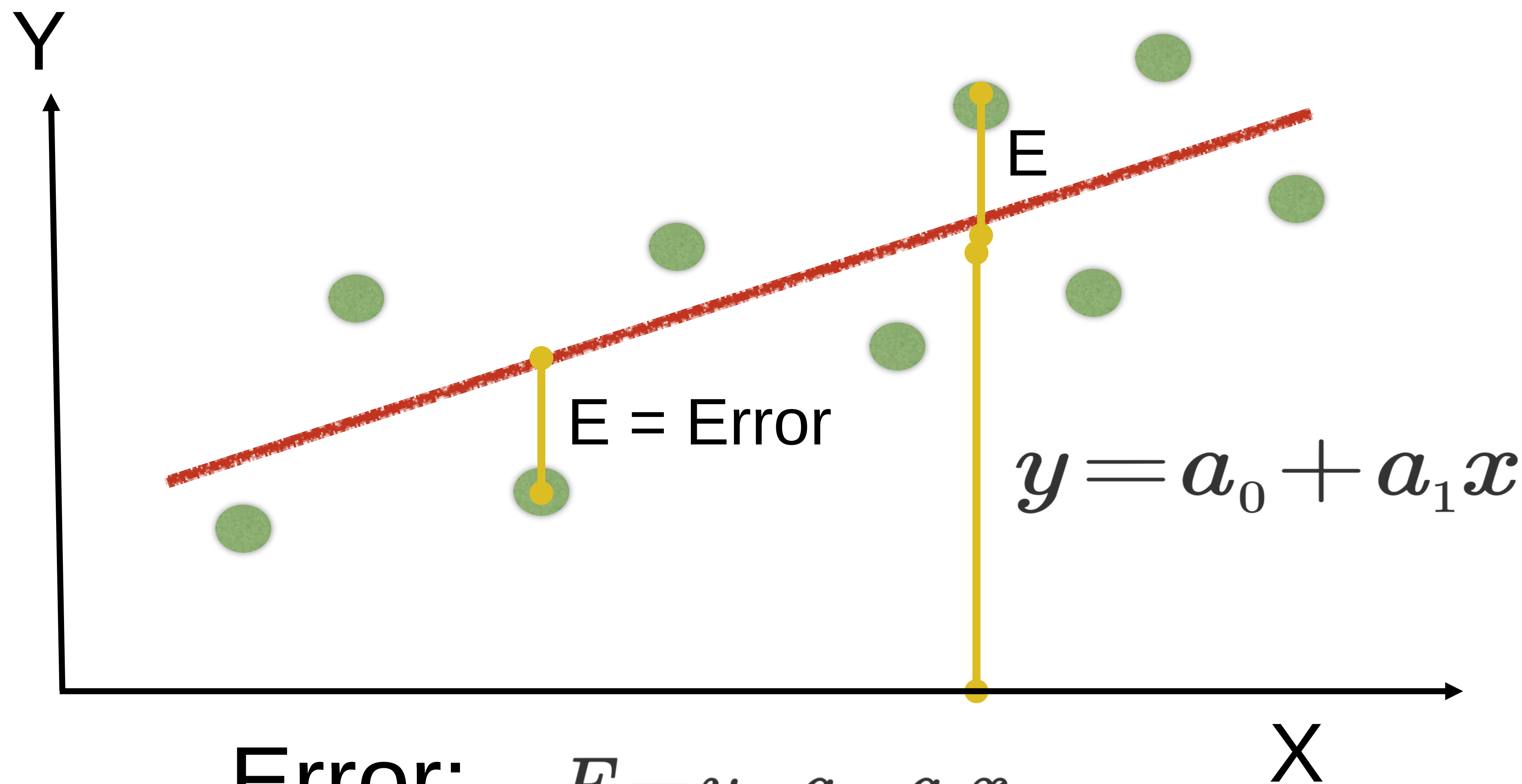
a_1 Representa la pendiente de la recta

Gráfica de datos con Ecuación Lineal

$$y = a_0 + a_1 x$$



Error en la Regresión lineal

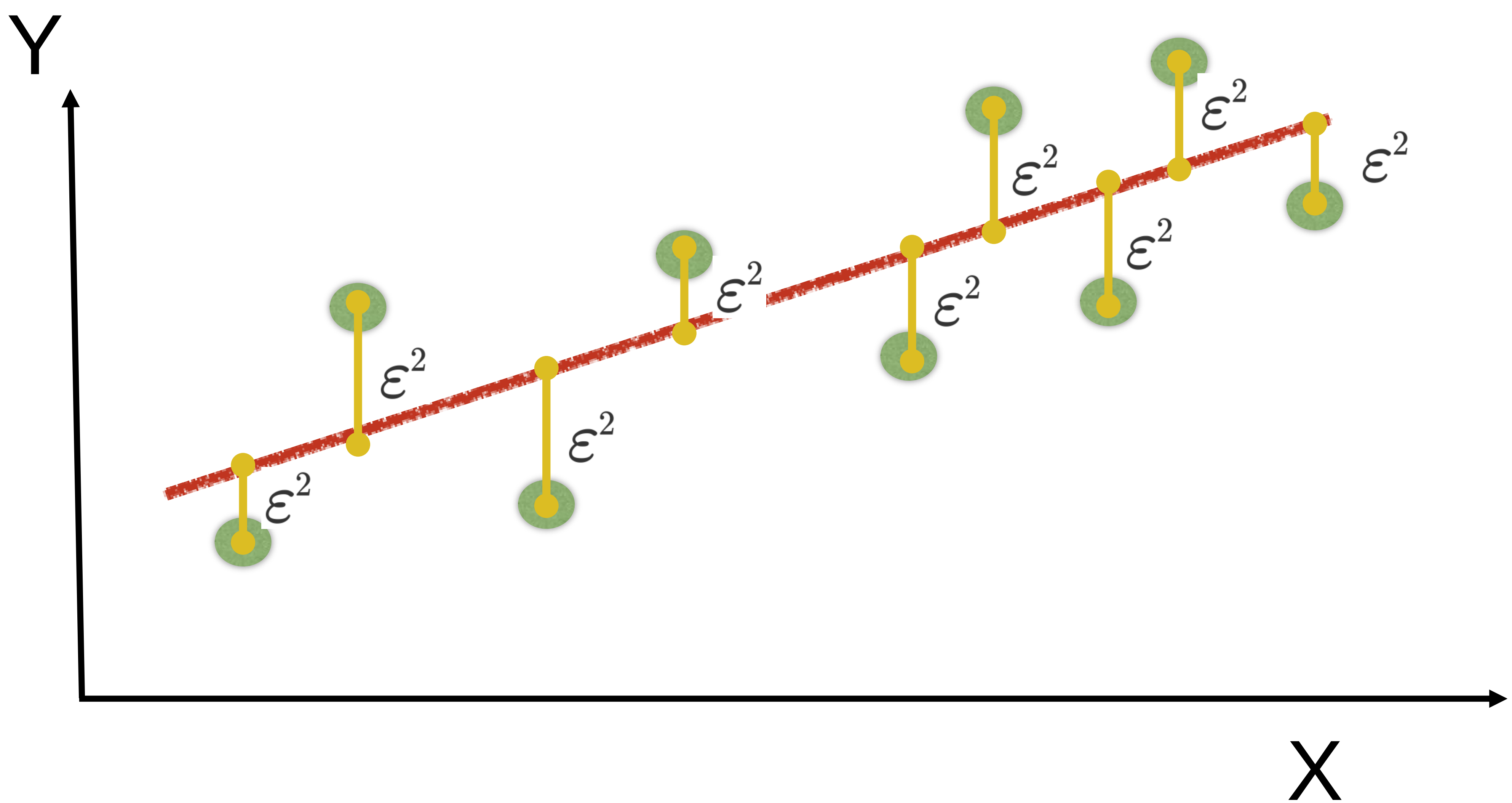


Error: $E = y_i - a_0 - a_1 x_i$

El error es la discrepancia entre el modelo de ajuste y los datos

Suma de los errores al cuadrado (Sr)

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



Modelos de regresión lineal

Con el objetivo de encontrar los coeficientes a_0 y a_1, debemos minimizar la suma de los errores al cuadrado (Sr).

Para ello derivamos parcialmente Sr con respecto a cada uno de los coeficientes.

Para generar un mínimo de Sr, se igualan a cero cada una de las derivadas parciales.

Reordenando lo resultante como un conjunto de ecuaciones lineales simultáneas, tenemos:

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

Derivando Sr parcialmente con respecto a los coeficientes constantes:

$$\frac{\partial Sr}{\partial a_0} = -2\Sigma(y_i - a_0 - a_1 x_i)^2$$

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

Igualando a cero las derivadas parciales para generar un mínimo

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

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

considerando:

$$\Sigma a_0 = n a_0$$

Las ecuaciones se expresan como un conjunto
de ecuaciones simultáneas con dos incognitas:

$$a_{0,} a_1$$

$$n a_0 + \Sigma x_i a_1 = \sum y_i$$

$$\sum x_i a_0 + \sum x_i^2 a_1 = \sum x_i y_i$$

Resolviendo:

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

$$a_0 = \frac{\sum y}{n} - \frac{\sum x}{n} a_1$$

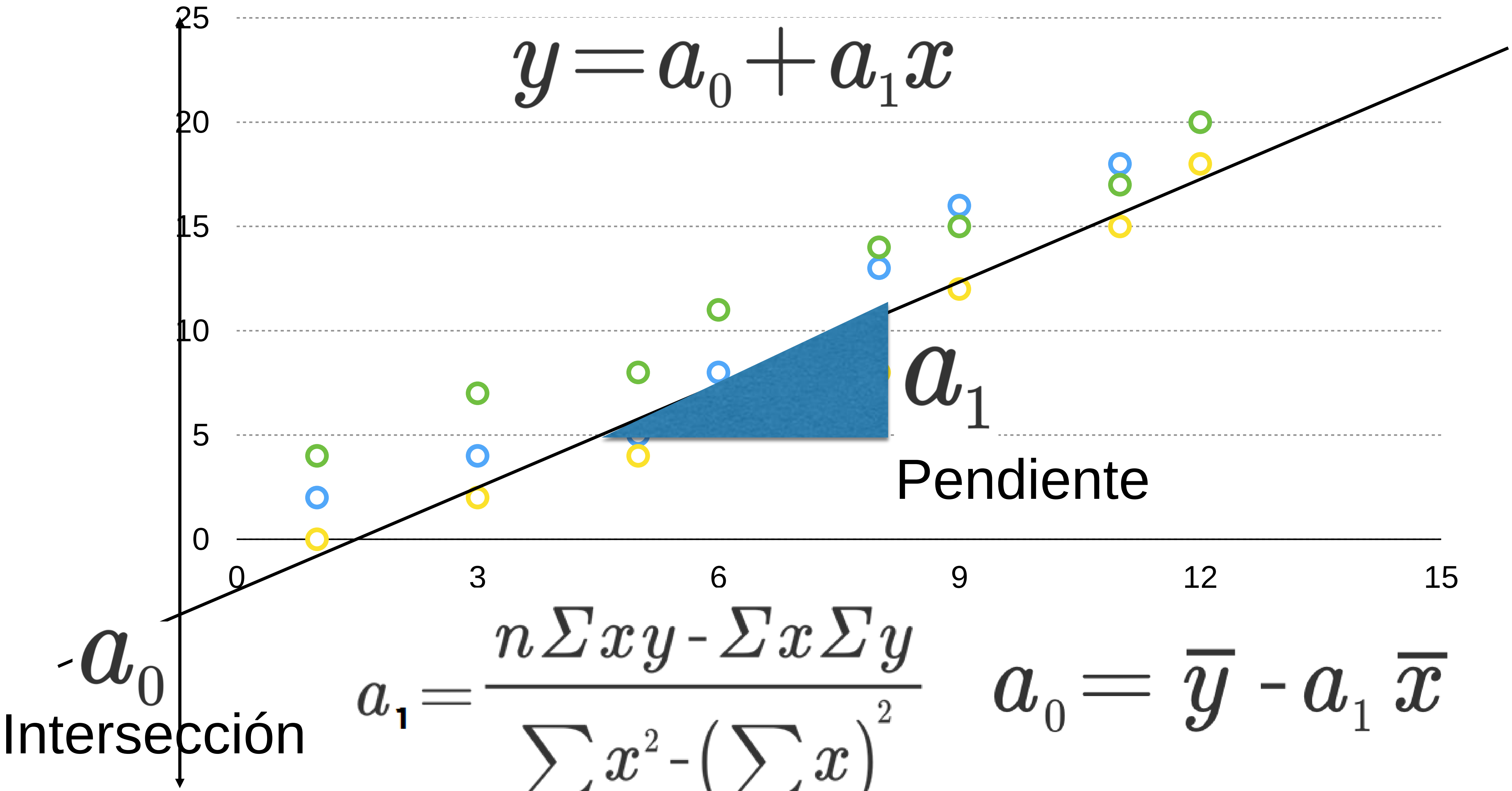
$$a_0 = \overline{y} - a_1 \overline{x}$$

Una vez encontrados los coeficientes se sustituyen en la ecuación de la recta

$$y = a_0 + a_1 x$$

Que podrá utilizarse en propósitos practicos

Gráfica de datos con Ecuación Lineal



$$y = a_0 + a_1 x$$

$$a_1 = \frac{n \Sigma x y - \Sigma x \Sigma y}{\sum x^2 - \left(\sum x\right)^2} \qquad a_0 = \bar{y} - a_1 \bar{x}$$

Cuantificación del error en la regresión lineal

Error estándar de la aproximación

$$S_{\frac{y}{x}} = \sqrt{\frac{Sr}{n-2}}$$

Desviación estándar total

$$S_y = \sqrt{\frac{\sum \left(y_i - \bar{y}\right)^2}{n-1}}$$

Cuantificación del error en la regresión lineal

Suma total de los cuadrados

$$S_T = \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2$$

Coeficiente de correlación

$$r = \sqrt{\frac{S_t - S_r}{S_t}}$$

Algoritmo para la regresión lineal

1. Verificar que los datos tienen o presentan una tendencia lineal
2. Sumar todos los valores de x
3. Sumar todos los valores de y
4. Sumar los productos xy
5. Sumar los valores de x^2
6. Obtener el valor promedio de x, dividiendo la suma de los valores entre el número de datos
7. Obtener el promedio de los valores de y, dividiendo la suma de los valores de y entre el número de datos
8. Calcular los valores de las constantes a_0 y a_1
9. De ser necesario calcular el error.

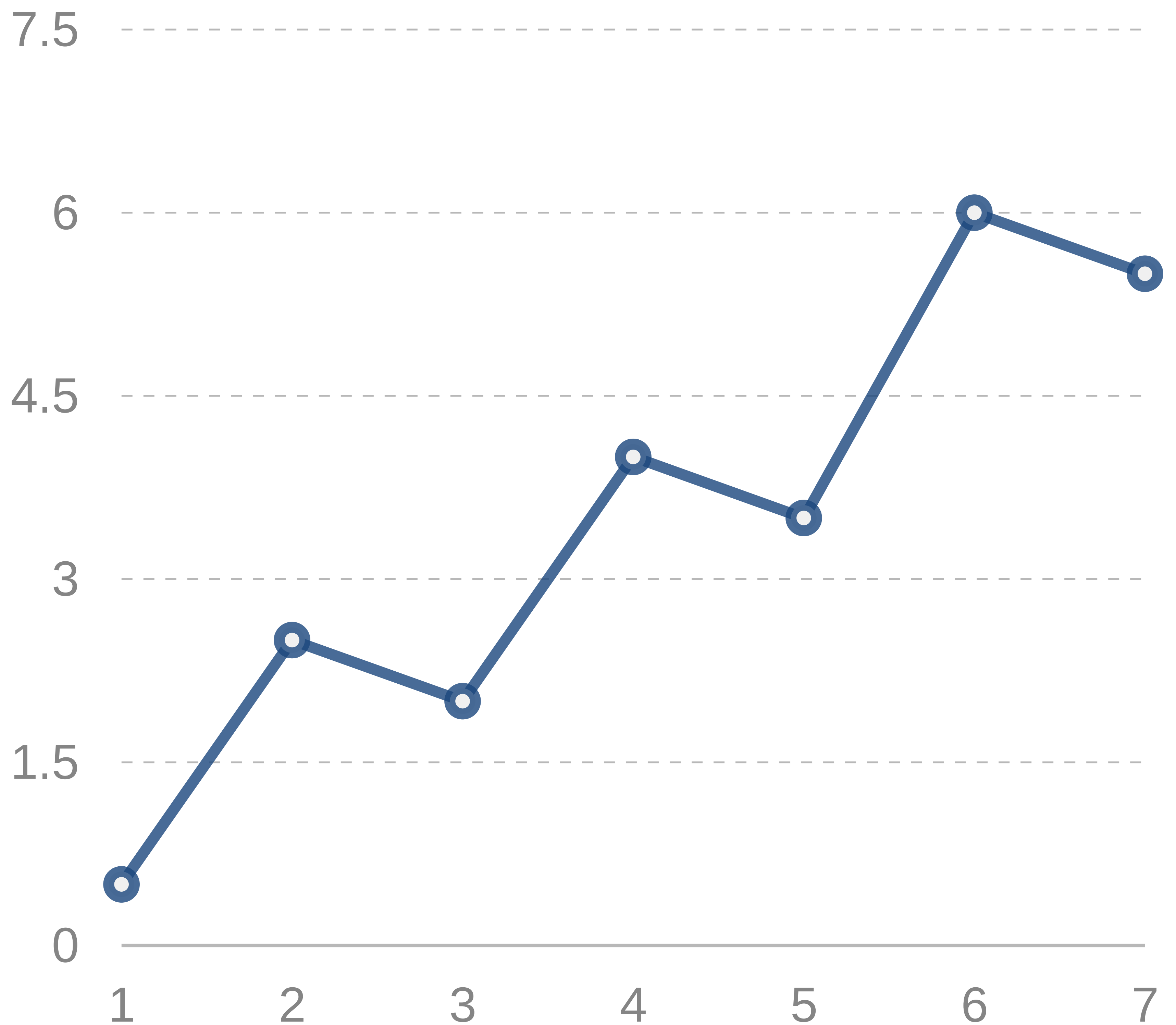
Seudocódigo para realizar un programa para el modelo de regresión lineal:

```
SUB Regress (x,y,n,a)
Sumx=0:        sumxy=0:
Sumy=0:        sumx2=0:
DOFOR i=1,n
Sumx=sumx+x;
Sumy=sumy+y;
Sumxy=sumxy+x*y;
Sumx2=sumx2+x*x;
END DO
Xm=sumx/n
Ym=sumy/n
A1=(n*sumxy-sumx*sumy)/(n*sumx2-sumx*sumx)
A0=ym-a1*xm
```

Es importante señalar, que éste algoritmo es el mismo para los modelos de regresión no lineal como el exponencial, logarítmico, etc.

Ejemplo: ajustar la ecuación de una recta al siguiente conjunto de datos.

x	y
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5



Haciendo los cálculos necesarios para encontrar los coeficientes:

i	x	y	xy	x^2	St	Sr
1	1	0.5	0.5	1	8.5765	0.1687
2	2	2.5	5	4	0.8622	0.5625
3	3	2.0	6	9	2.0408	0.3473
4	4	4.0	16	16	0.3265	0.3265
5	5	3.5	17.5	25	0.0051	0.5896
6	6	6.0	36	36	6.6122	0.7972
7	7	5.5	38.5	49	4.2908	0.1993
Sumas	28	24	119.5	140	22.7143	2.9911

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

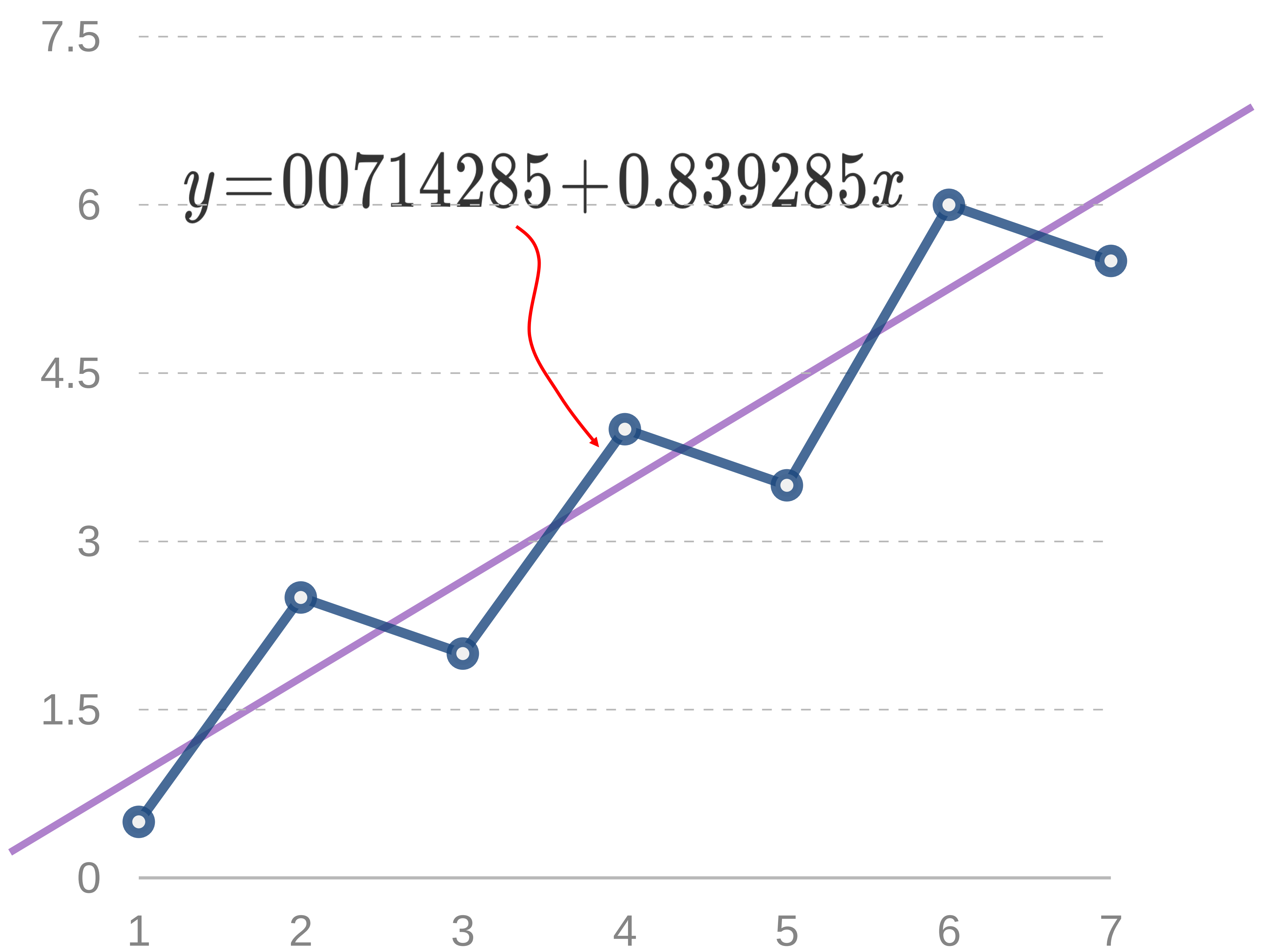
$$\bar{x} = \frac{\sum x}{n} = \frac{28}{7} = 4.0$$

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{7(119.5) - (28)(24)}{7(140) - (28)^2} = 0.8392857143$$

$$a_0 = \frac{\sum y}{n} - \frac{\sum x}{n} a_1 \quad = 3.428571 - (0.8392857143)(4) = 0.0714285718$$

Sustituyendo en la ecuación de la recta:

$$y = 00714285 + 0.839285x$$



7.5
6
4.5
3
1.5
0
1
2
3
4
5
6
7
$y = 00714285 + 0.839285x$

Cuantificación del error

Error estándar de la aproximación

$$S_{\frac{y}{x}} = \sqrt{\frac{Sr}{n-2}} = \sqrt{\frac{2.9911}{7-2}} = 0.77344683$$

Desviación estándar total

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{22.7143}{7-1}} = 1.94569182$$

Cuantificación del error

Suma total de los cuadrados

$$S_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad = 22.7143$$

Coeficiente de correlación

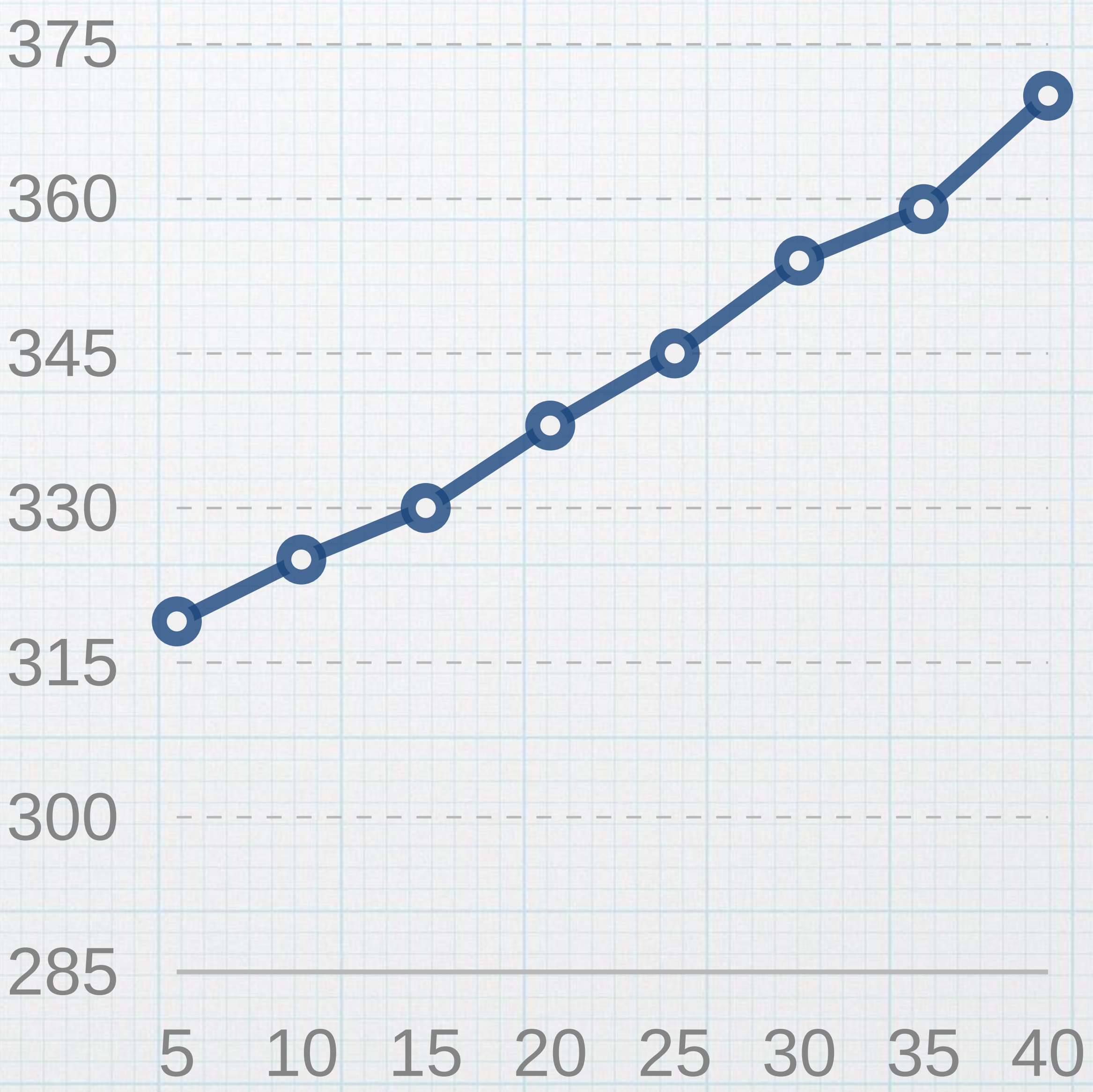
$$r = \sqrt{\frac{S_t - S_r}{S_t}} \quad = \sqrt{\frac{22.7143 - 2.9911}{22 - 7143}} = 0.931834$$

Ejercicio: En la tabla se enlistan los esfuerzos cortantes, en kilopascales (kPa), de nueve especímenes de suelo tomados a distintas profundidades de un estrato arcilloso. Estime el esfuerzo cortante a la profundidad de 4.5 m. Utilizando regresion lineal simple.

Profundidad, m	Esfuerzo, kPa
1.9	14.4
3.1	28.7
4.2	19.2
5.1	43.1
5.8	33.5
6.9	52.7
8.1	71.8
9.3	62.2
10.0	76.6

Ejercicio: El observatorio Mauna Loa, Hawai, registra la concentración de dióxido de carbono (en partes por millón) en la atmósfera terrestre. En la tabla se muestran los registros correspondientes al mes de enero de varios años

Año	CO2 (partes por millon)
5	319
10	325
15	330
20	338
25	345
30	354
35	359
40	370



MUCHAS GRACIAS

REFERENCIAS

Burden, R.L., Faires, J.D., 2011. *Análisis numérico.* 9ª Ed. México: Thompson Learning.

Chapra, S.C. Canale, R.P., 2011. *Métodos numéricos para ingenieros.* 6ª Ed. México. D.F.:McGraw-Hill.

Cheney, E.W. Kincaid, D., 2011. *Métodos numéricos y computación.* 6ª Ed. México: Cengage Learning.

Nieves, A. Domínguez, F.C., 2012. *Métodos numéricos aplicados a la ingeniería.* 4ª Ed. México, D.F.: Patria.

Toledo, L.C., 2014. *Apuntes de la unidad de aprendizaje Métodos Numéricos de la Facultad de Ingeniería de la UAEM.*