CENTRO UNIVERSITARIO UAEM ZUMPANGO

LICENCIATURA EN TURISMO

TEMA: 3.1 PRUEBAS DE ASOCIACIÓN PARAMÉTRICAS Y NO PARAMÉTRICAS.

“REGRESIÓN LINEAL Y CORRELACIÓN LINEAL”

M. EN C. LUIS ENRIQUE KU MOO

FECHA: SEPTIEMBRE DE 2018

UNIDAD DE APRENDIZAJE
“ESTADISTICA”

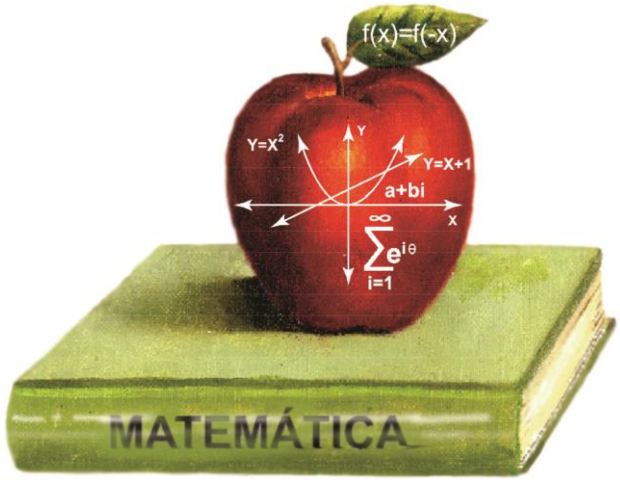
UNIDAD DE COMPETENCIA III:

“PRUEBAS ESTADÍSTICAS APLICADAS A UN CASO PRÁCTICO EN PARTICULAR RELACIONADO CON EL TURISMO”

1. Pruebas de asociación paramétricas y no paramétricas
2. Pruebas de comparación de medias y poblaciones, paramétricas y no paramétricas
3. Aplicación a un caso práctico, diseño, análisis, interpretación y presentación

OBJETIVO

Aplicar las pruebas estadísticas que correspondan al planteamiento y solución de problemas o casos relacionados con el Turismo.



CONCEPTOS

Regresión lineal. Método estadístico que sirve para estimar la relación de dependencia entre 2 o más variables. Una o varias llamadas independientes u otra llamada dependiente.

$$y = f\,(x_1, x_2, x_3, \ldots, x_n)$$

y = Variable dependiente (Predicha o explicada)
x = Variable independiente (Predictora o explicativa)

y = ventas, consumo, demanda, oferta, etc…
x = ingreso, precio, población, publicidad, etc…

REGRESIÓN LINEAL SIMPLE

Regresión lineal. Es un modelo donde la variable dependiente y se relaciona con una sola variable independiente x. Se ajusta este modelo a un conjunto de datos mediante el método de los mínimos cuadrados.

$$Y_i = a + b\,x_i + \xi_i$$

Parámetros

Error
- errores casuales
- errores de medición
- deficiencias del modelo

ε_i es la parte de y_i que no está explicada por la regresión lineal de Y sobre x_i .

REGRESIÓN LINEAL SIMPLE

A la derecha tenemos datos observados de dos variables.

No se muestran en ningún orden particular.

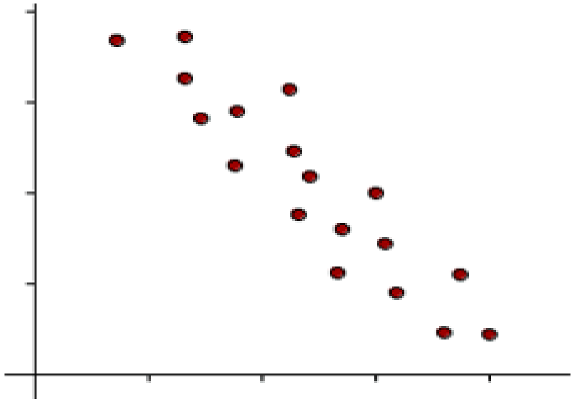
Dichas observaciones pueden ser representadas en un diagrama de dispersión.

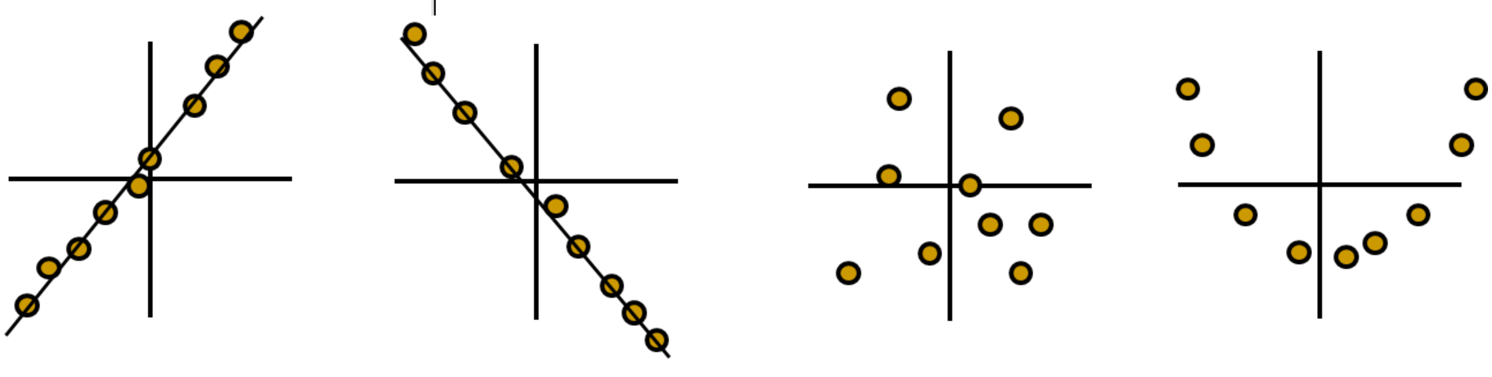
¿Cuál es la variable dependiente y cual la independiente?

$ publicid en mill.	**Demanda mill.**
3	2
4	4
6	4
8	5
9	7
11	8
14	9

REGRESIÓN LINEAL SIMPLE

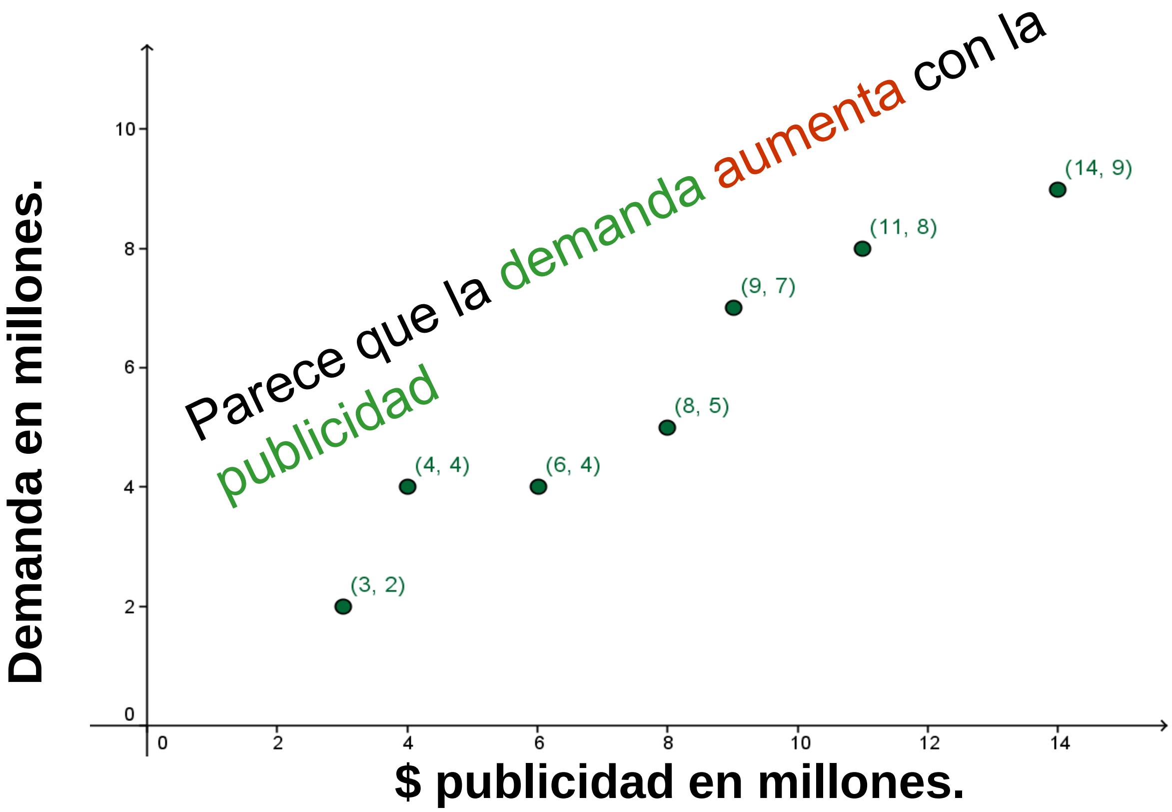
El diagrama de dispersión. Es una representación gráfica de la relación entre estas dos variables. Se representa cada par de valores como las coordenadas de un punto (x_i, y_i).





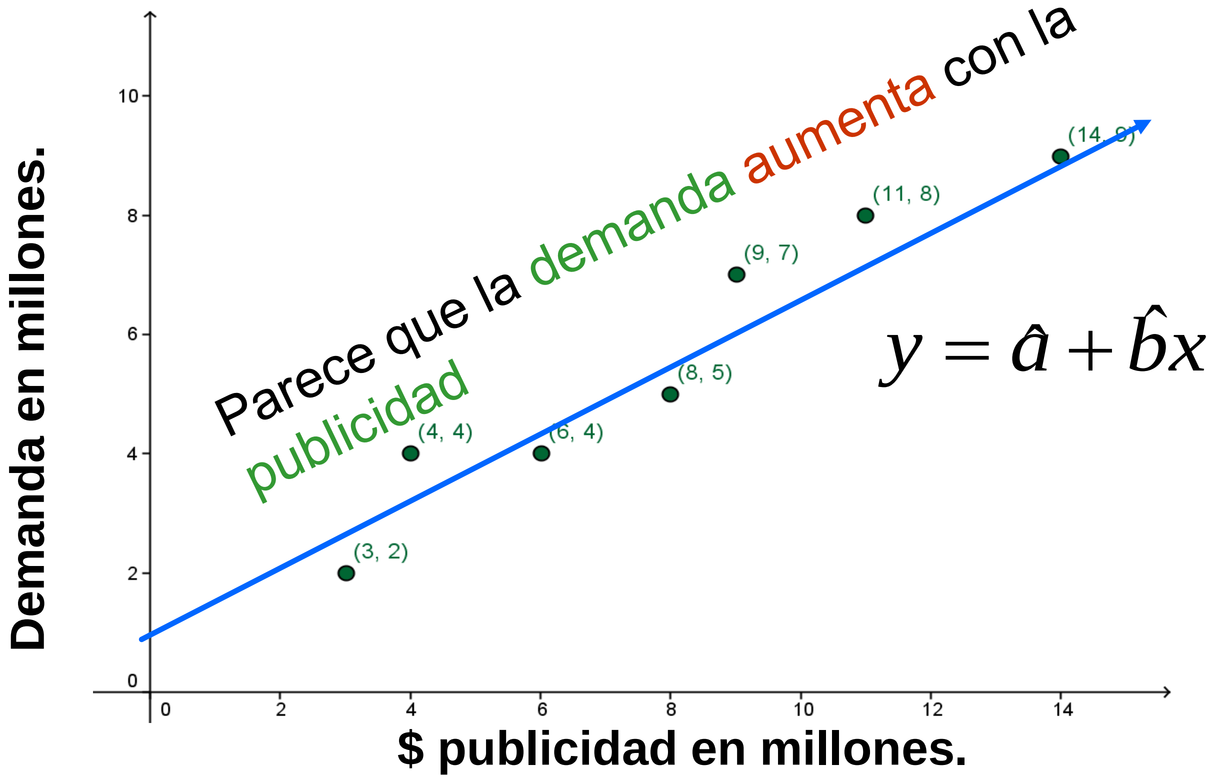
REGRESIÓN LINEAL SIMPLE

Nuestro objetivo será intentar reconocer a partir del mismo si hay relación entre las variables.



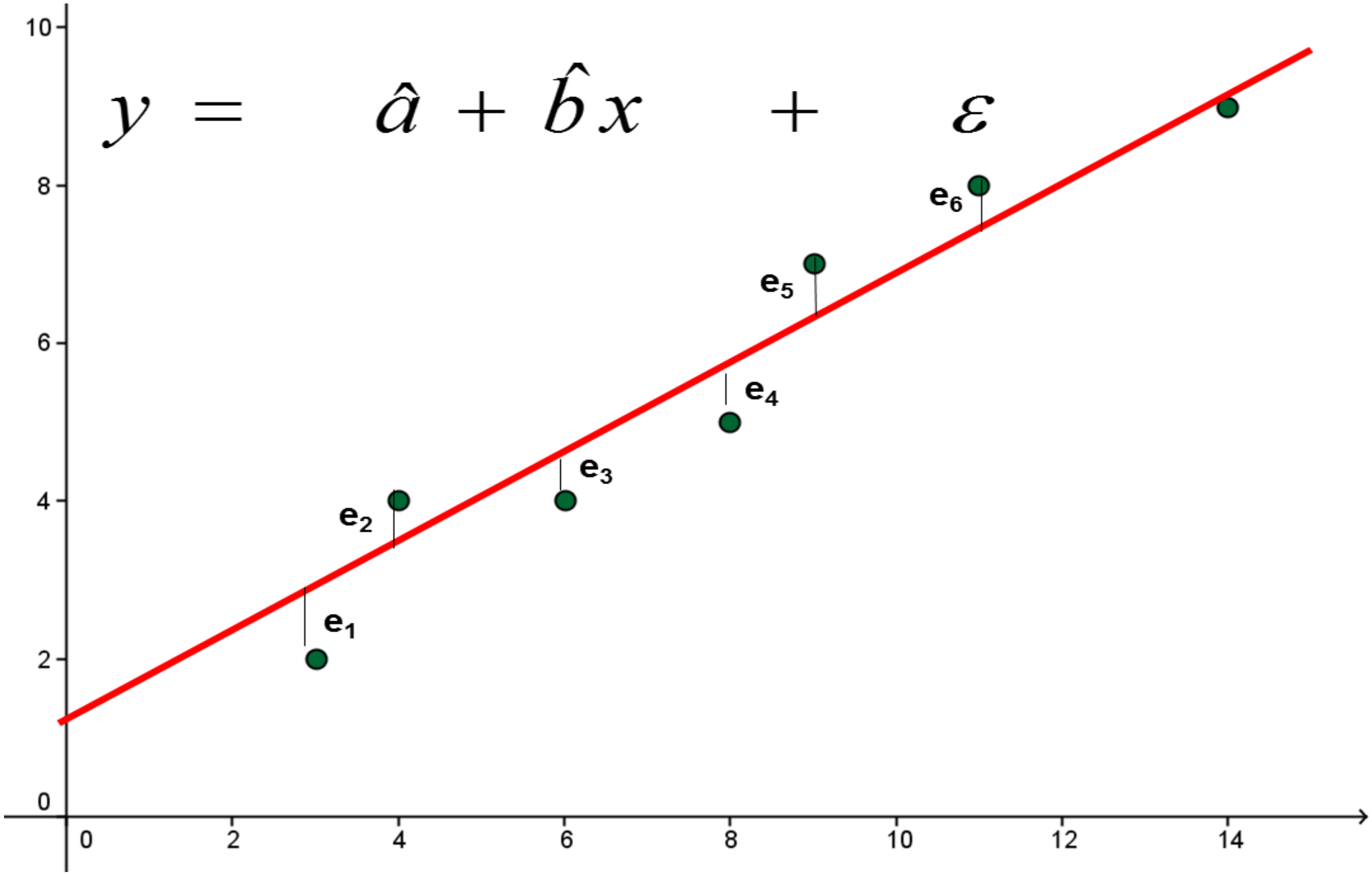
REGRESIÓN LINEAL SIMPLE

Entonces, podríamos escoger un modelo que relacione a *y* con *x* trazando una línea recta a través de los puntos de la figura.



REGRESIÓN LINEAL SIMPLE

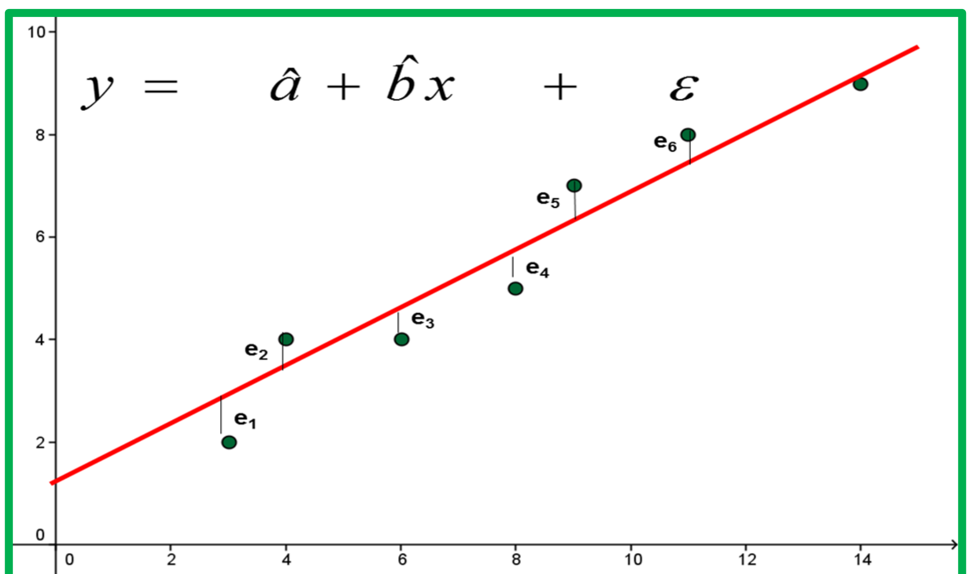
Algunas desviaciones positivas y otras negativas. Con $E(\varepsilon) = 0$.



REGRESIÓN LINEAL SIMPLE

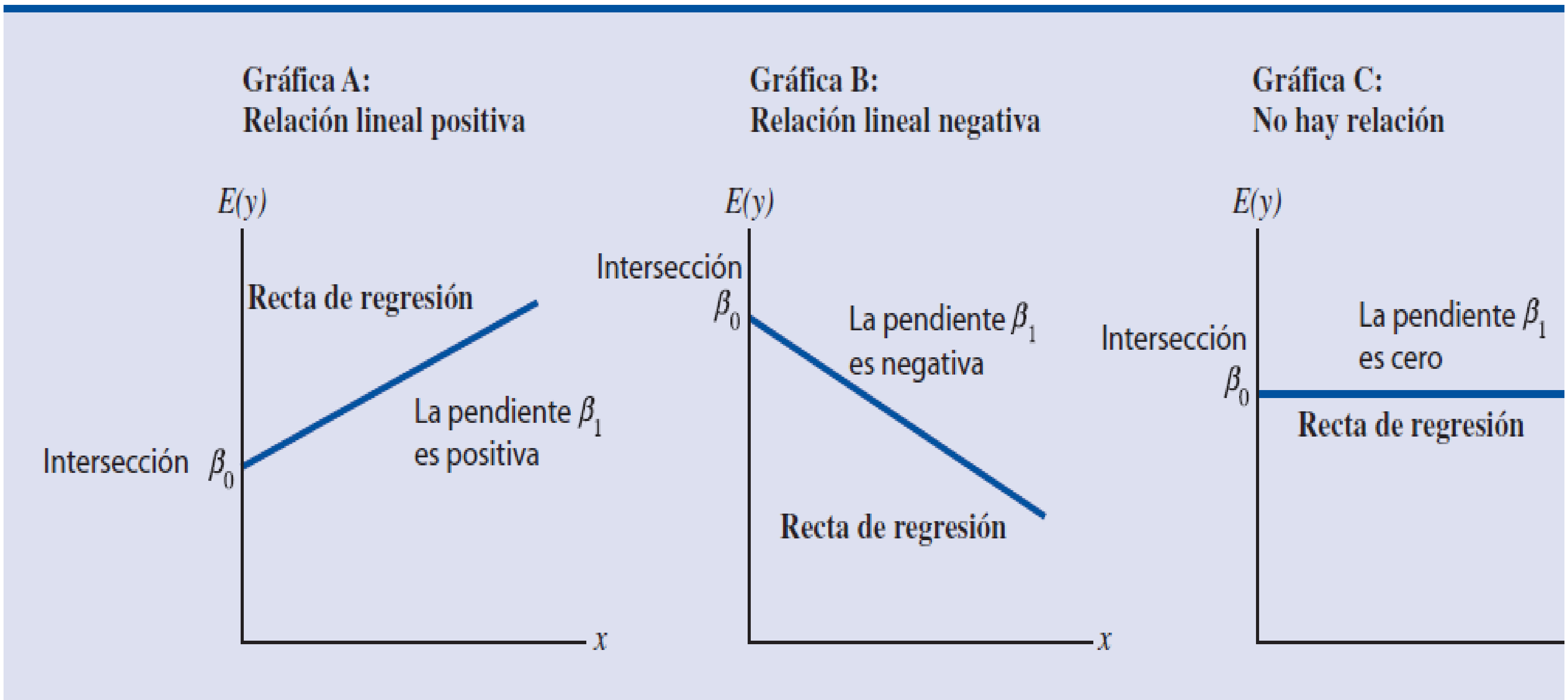
Para ajustar un modelo de regresión lineal simple a un conjunto de datos, debemos encontrar estimadores para los parámetros desconocidos *a* y *b*

$$y = \underbrace{\hat{a} + \hat{b}x}_{\text{Valor medio de } y \text{ para una } x \text{ dada}} + \underbrace{\varepsilon}_{\text{Error aleatorio}}$$



REGRESIÓN LINEAL SIMPLE

Para ajustar el modelo de regresión lineal se usa el método de mínimos cuadrados para obtener la ecuación de regresión estimada.



REGRESIÓN LINEAL SIMPLE: MODELO

Modelo de regresión: $y = \hat{a} + \hat{b}x$

Donde: *y* = variable dependiente y *x* = variable independiente.

$\hat{a}$ = punto en que la línea corta el eje *y.*

$\hat{b}$= pendiente de la línea.

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \; ; \qquad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

$\Sigma x_i y_i$= Suma del producto x por y

$\bar{x} = \text{promedio}$ de x

$\bar{y} = \text{promedio}$ de y

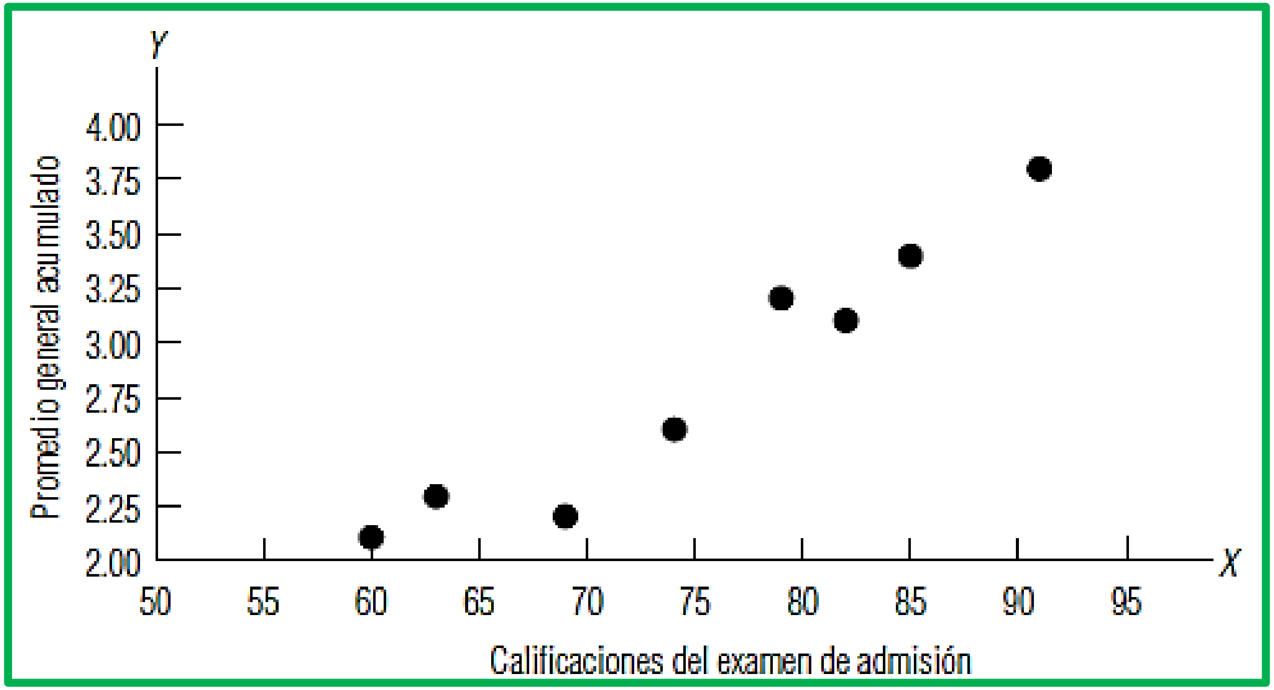
Σx_i^2 = Suma de los cuadrados de x

n = Número de datos

REGRESIÓN LINEAL SIMPLE: Ejemplo.

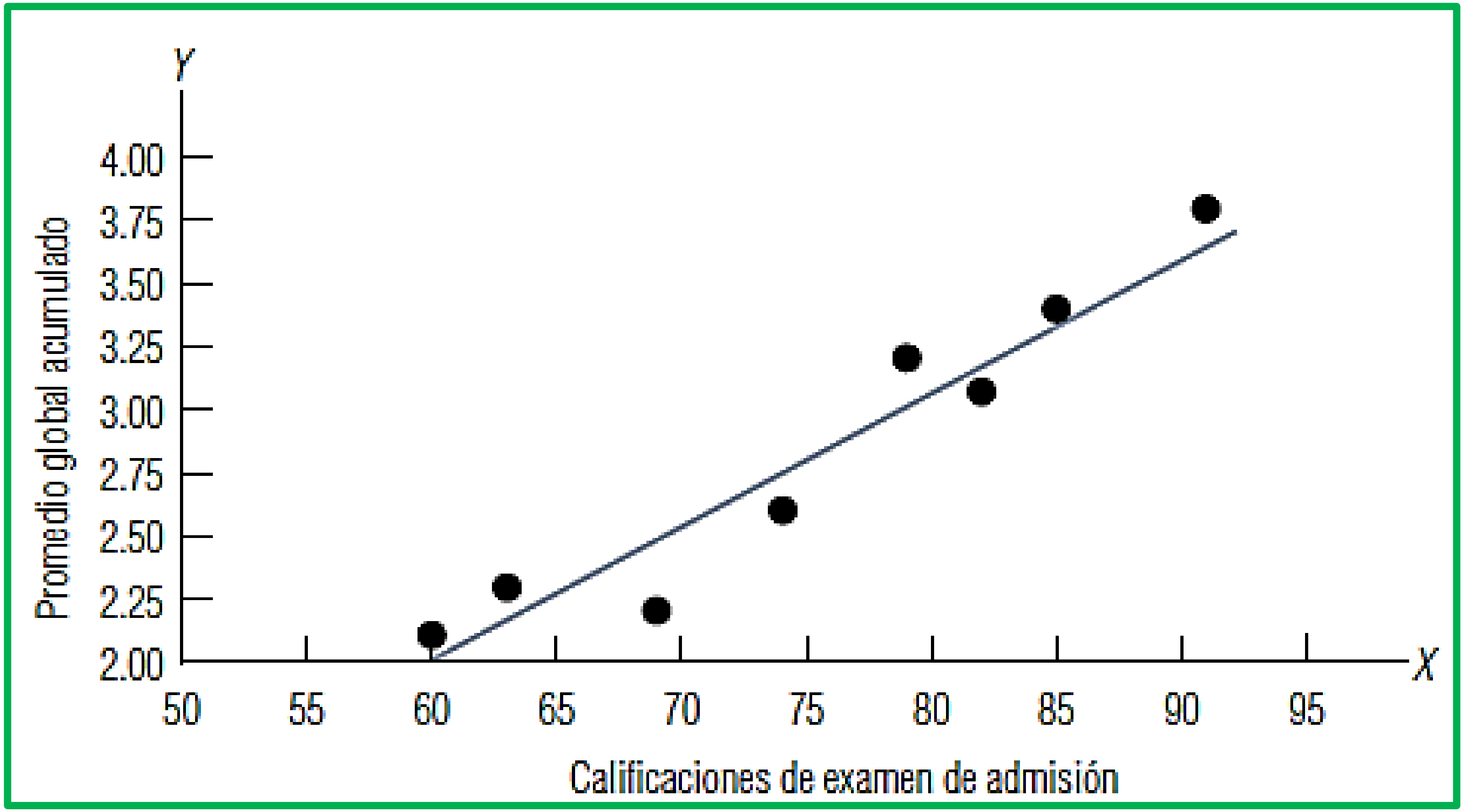
Se pretende determinar si existe una relación entre las calificaciones de un estudiante en su examen de admisión y su promedio general al graduarse

Calificación de examen de admisión	**74**	**69**	**85**	**63**	**82**	**60**	**79**	**91**
Promedio general acumulado (4.0=A)	2.6	2.2	3.4	2.3	3.1	2.1	3.2	3.8



REGRESIÓN LINEAL SIMPLE: Ejemplo.

Línea de recta "ajustada" que representa la relación entre las calificaciones del examen de admisión y el promedio global acumulado.



REGRESIÓN LINEAL SIMPLE: Ejemplo

x = Edad de autos en años

y = Gastos de reparación en cientos

Dato	x	y	xy	x^2
1 2 3 4	5 3 3 1	7 7 6 4	35 21 18 4	25 9 9 1
Σ	12	24	78	44

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{78-(4)(3)(6)}{44-((4)(3^2)} = \frac{6}{8} = .75$$

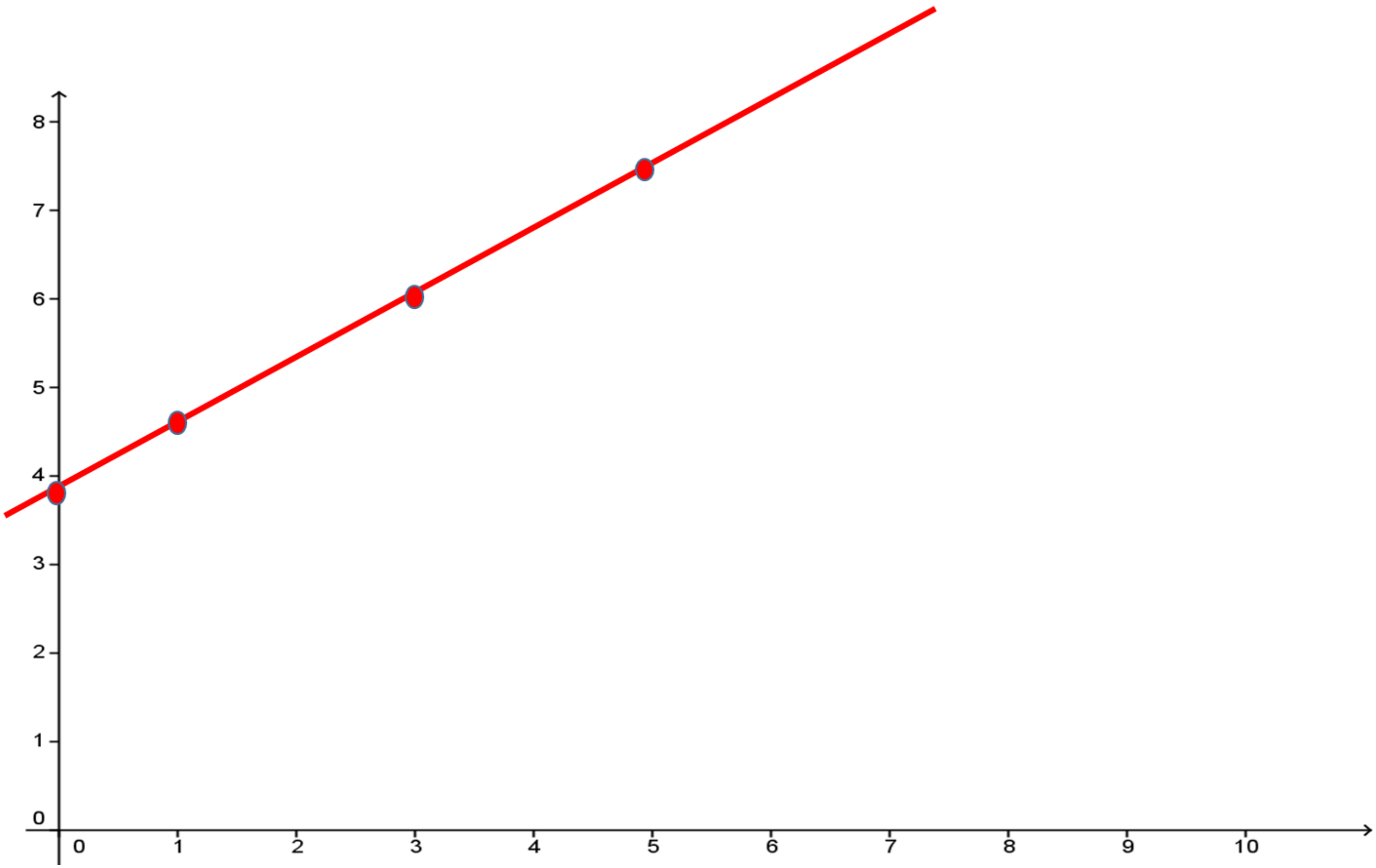
$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 6-(.75)(3) = 3.75$$

$y = 3.75 + 0.75x$ es la ecuación de la recta

Los gastos en reparación de autos asciende a 375 pesos anuales más 75 pesos por cada año de edad del automóvil.

REGRESIÓN LINEAL SIMPLE

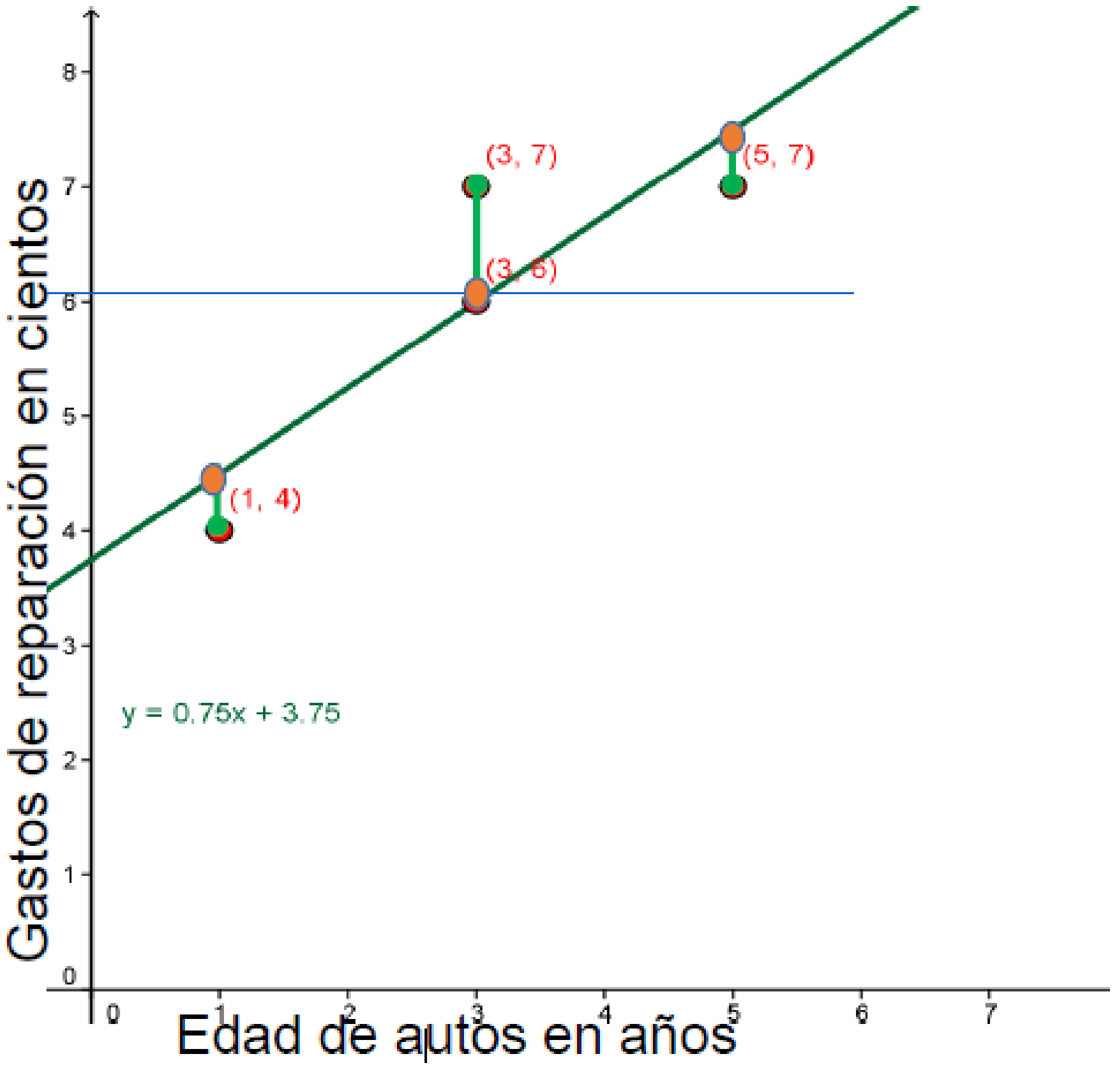
$y = 3.75 + 0.75x$ es la ecuación de la recta



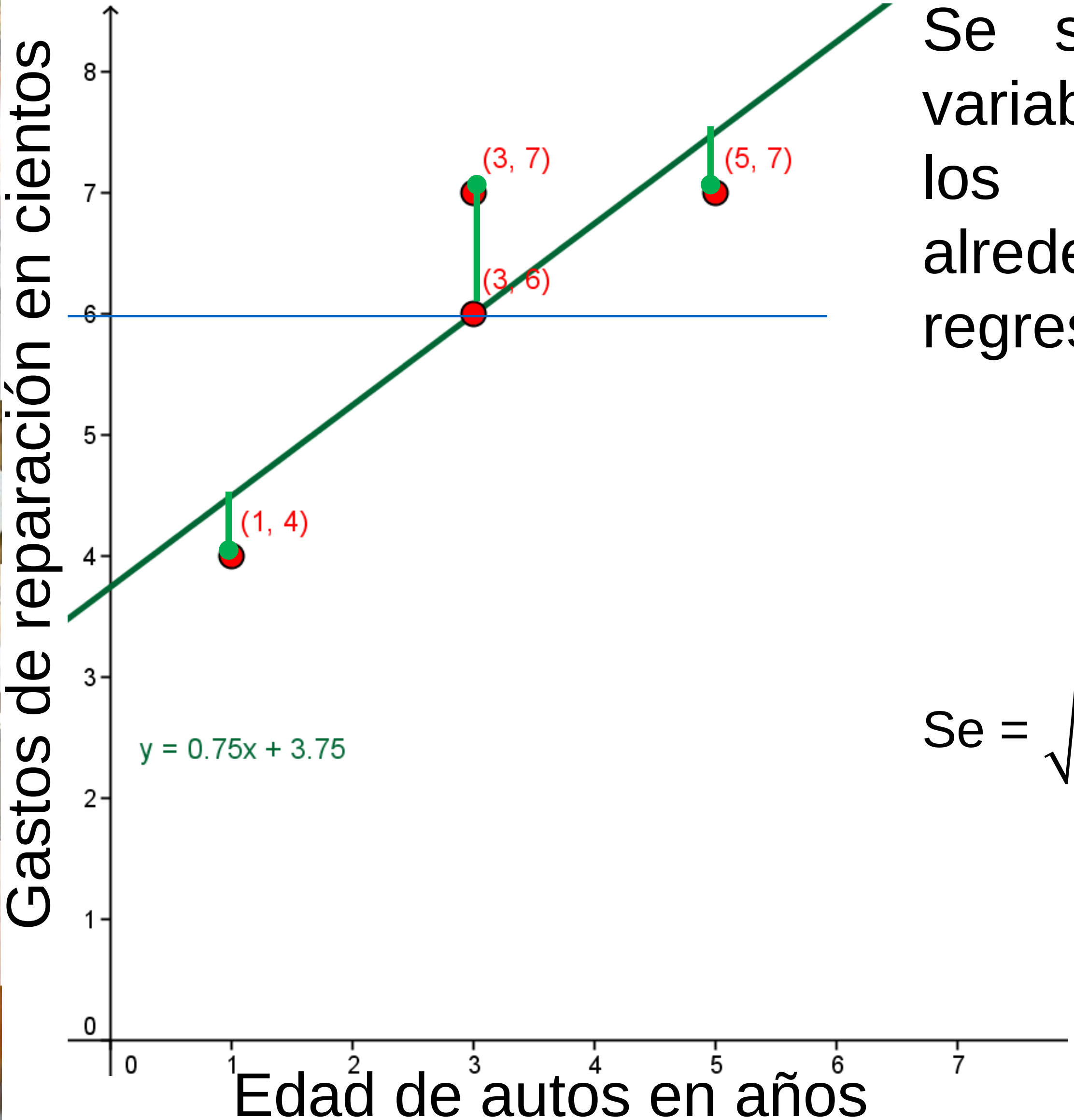
REGRESIÓN LINEAL SIMPLE

Dato	x	$\hat{y}$
1	5	7.5
2	3	6.0
3	3	6.0
4	1	4.5

Dato	x	y	$\hat{y}$
1	5	7	7.5
2	3	7	6.0
3	3	6	6.0
4	1	4	4.5



ERROR ESTANDAR DE ESTIMACIÓN



Se simboliza Se. Mide la variabilidad, o dispersión de los valores observados alrededor de la línea de regresión.

$$Se = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum y^2 - \hat{a}\sum y - \hat{b}\sum xy}{n-2}}$$

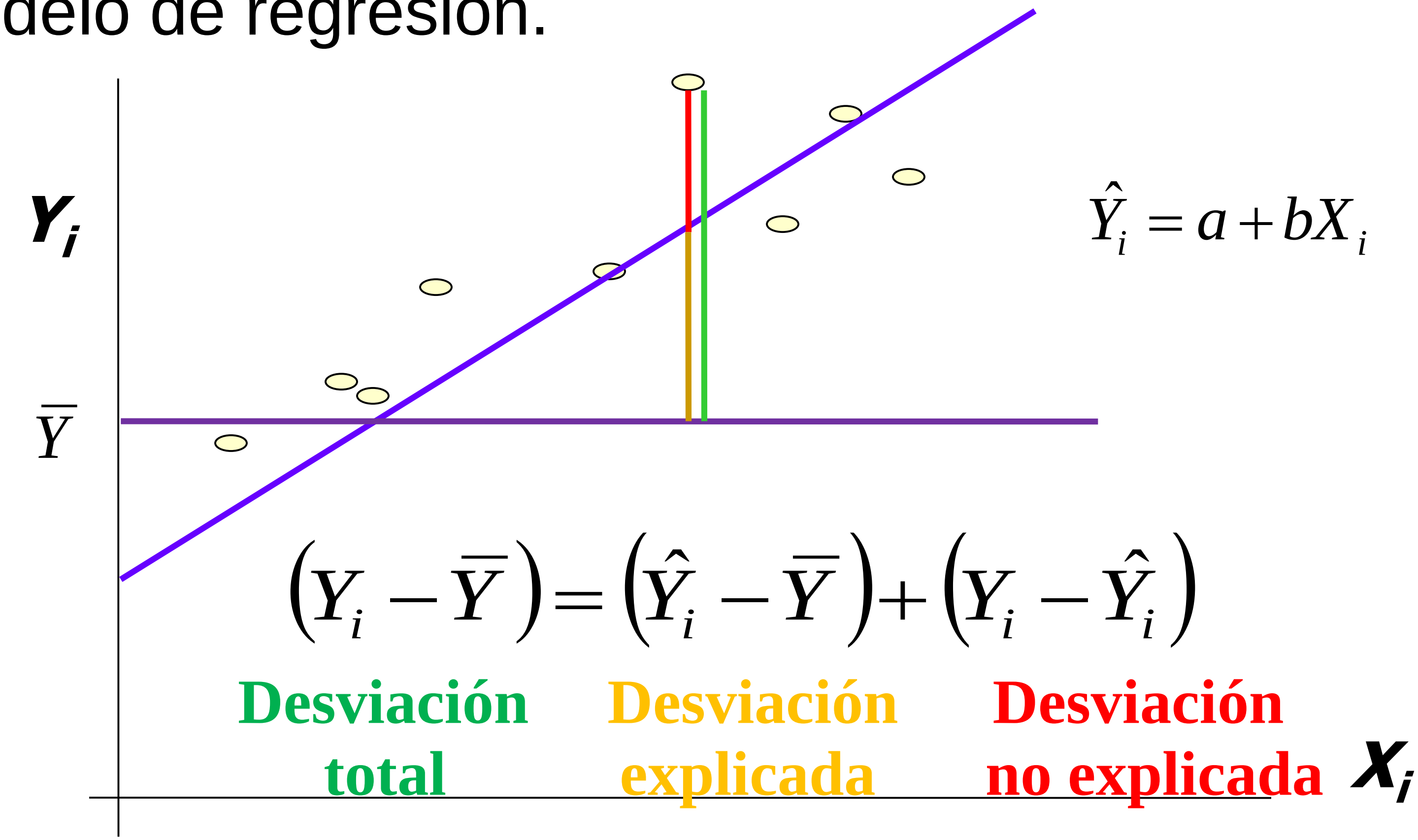
ERROR ESTANDAR DE ESTIMACIÓN

Dato	x	y_i	$\hat{y}$	$e=(y_i\text{-}\hat{y}_i)$	e^2
1	5	7	7.5	-0.5	.25
2	3	7	6.0	1.0	1
3	3	6	6.0	0.0	0
4	1	4	4.5	-0.5	.25
sumas			24	0.0	1.5

$$Se = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{1.5}{4-2}} = \sqrt{.75} = .8660$$

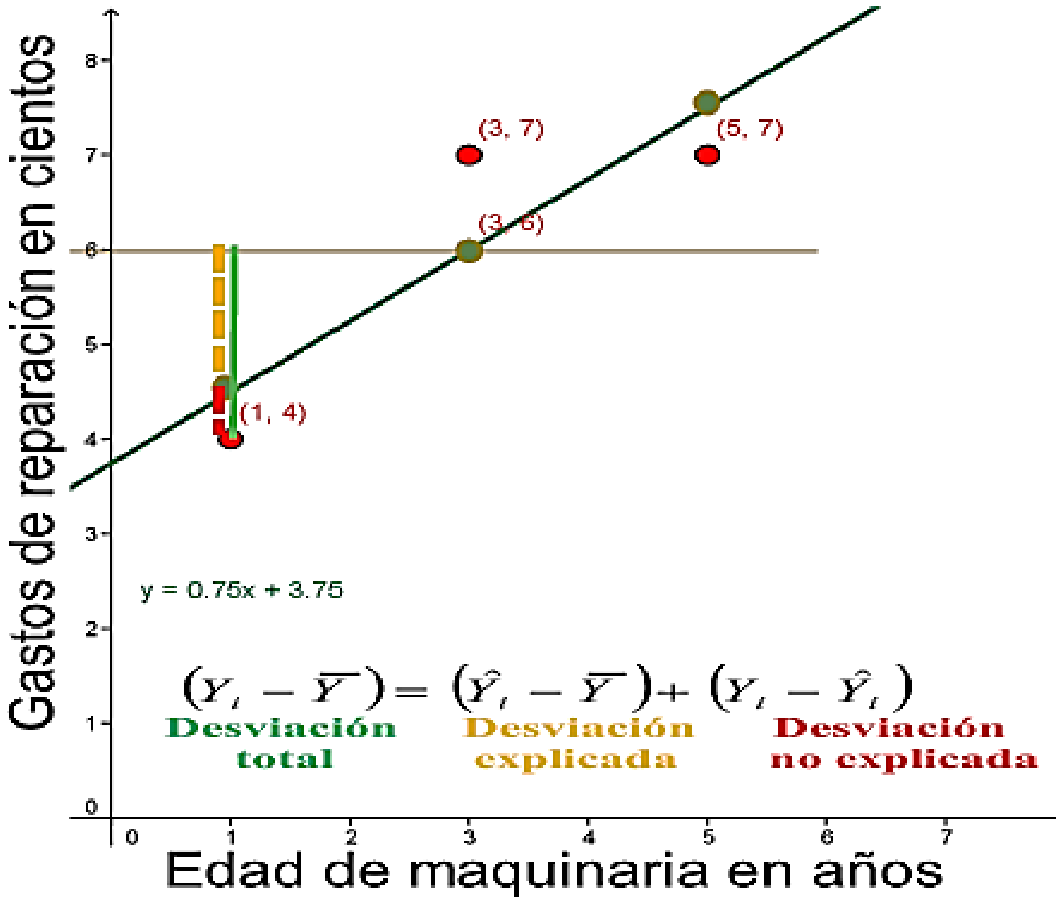
COEFICIENTE DE DETERMINACION

Se simboliza R^2. El coeficiente de determinación mide la proporción de variabilidad total de la variable dependiente respecto a su media que es explicada por el modelo de regresión.



COEFICIENTE DE DETERMINACION

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \qquad R^2 = \frac{\hat{a}\sum y + \hat{b}\sum xy - n\bar{y}^2}{\sum y^2 - n\bar{y}^2}$$



COEFICIENTE DE DETERMINACION

x_i	y_i	$\hat{y}$	$(\hat{y}-\bar{y})$	$(\hat{y}-\bar{y})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$
5	7	7.5	1.5	2.25	1.0	1.0
3	7	6	0.0	0.0	1.0	1.0
3	6	6	0.0	0.0	0.0	0.0
1	4	4.5	-1.5	2.25	-2.0	4.0
12	24	24	0.0	**4.5**	0.0	**6.0**

$$R^2 = \frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2} = \frac{\hat{a}\sum y+\hat{b}\sum xy-n\bar{y}^2}{\sum y^2-n\bar{y}^2}$$

$$R^2 = \frac{4.5}{6.0} = .75$$

CORRELACIÓN LINEAL:

El **coeficiente de correlación lineal de Pearson** (se denota r ó ρ) es una **medida de asociación lineal** entre dos variables aleatorias X e Y:

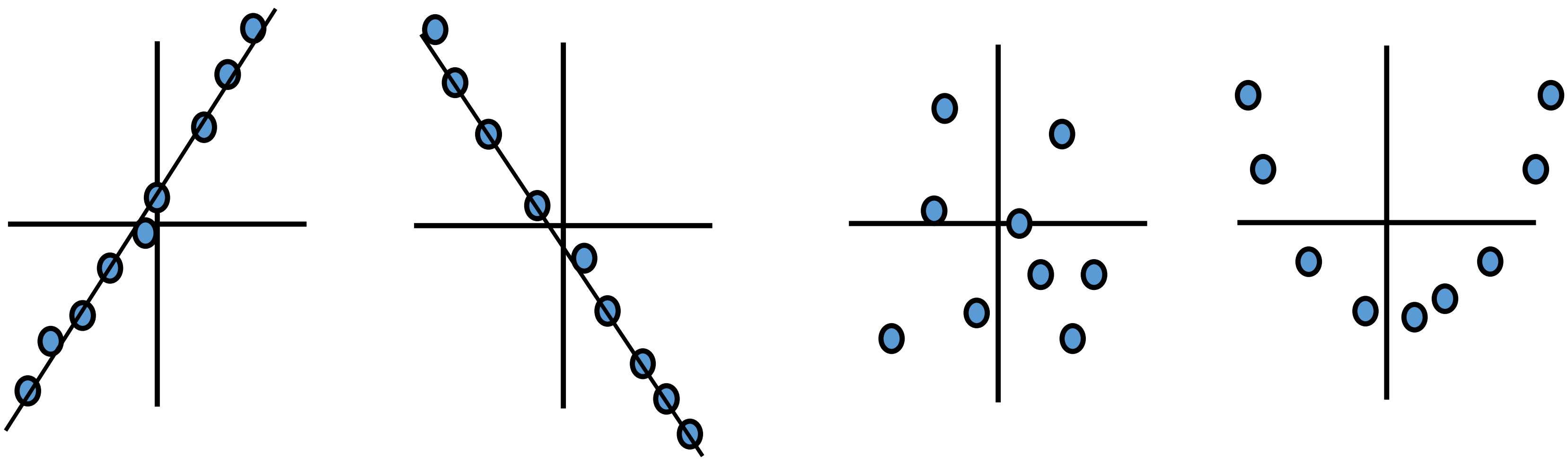
Cuando r es cercano a +1, hay una buena correlación positiva entre las variables, será creciente.

Cuando r es cercano a -1, hay una buena correlación negativa entre las variables, es decreciente.

$$r = \frac{S_{xy}}{\sqrt{S_x^2 S_y^2}} = \frac{S_{xy}}{S_x S_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (X_i - \bar{x})^2}\sqrt{\sum (y_i - \bar{y})^2}}$$

$$r = \frac{\frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y}}{\sqrt{\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right)\left(\frac{\sum y_i^2}{n} - \bar{y}^2\right)}} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\left(\sum x^2\right) - \left(\sum x\right)^2\right]\left[n\left(\sum y^2\right) - \left(\sum y\right)^2\right]}}$$

CORRELACIÓN LINEAL:



Correlación lineal positiva

$r \approx +1$

Correlación lineal negativa

$r \approx -1$

No hay correlación

$r \approx 0$

Hay correlación no lineal

$r \approx 0$

CORRELACIÓN LINEAL:

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{\left[n\left(\sum x^2\right)-\left(\sum x\right)^2\right]\left[n\left(\sum y^2\right)-\left(\sum y\right)^2\right]}}$$

x	y	x^2	y^2	xy
5	7	25	49	35
3	7	9	49	21
3	6	9	36	18
1	4	1	16	4
12	24	44	150	78

$$r=\frac{4(78)-(12)(24)}{\sqrt{\left[4(44)-(12)^2\right]\left[4(150)-(24)^2\right]}}=\frac{312-288}{\sqrt{(176-144)(600-576)}}$$

$$r=\frac{24}{\sqrt{(32)(24)}}=\frac{24}{\sqrt{768}}=\frac{24}{27.713}=.866$$

$$\boxed{r=86.6\%}$$

REGRESIÓN LINEAL SIMPLE

Realizar ejercicio de regresión lineal y correlación

$ publicidad en millones	**Demanda millones**
3	2
4	4
6	4
8	5
9	7
11	8
14	9

REGRESIÓN LINEAL MÚLTIPLE

Cuando se usa más de una variable independiente para predecir los valores de una variable dependiente, el proceso se llama análisis de regresión múltiple, incluye el uso de ecuaciones lineales.

$$y_i = a + b_1 x_1 + \ ... + b_n x_n + \varepsilon_i$$

Se asume que los errores ε_i tienen las características siguientes:

- Tienen media cero y varianza común σ^2.
- Son estadísticamente independientes.
- Están distribuidos en forma normal.

REGRESIÓN LINEAL MÚLTIPLE: Ejemplo

Superficie (x1)	Oficinas (x2)	Entradas (x3)	Antigüedad (x4)	Valor (y)
2310	2	2	20	142
2333	2	2	12	144
2356	3	1.5	33	151
2379	3	2	43	150
2402	2	3	53	139
2425	4	2	23	169
2448	2	1.5	99	126
2471	2	2	34	142.9
2494	3	3	23	163
2517	4	4	55	169
2540	2	3	22	149

REGRESIÓN LINEAL MÚLTIPLE: Ejemplo

Ejemplo resuelto con Excel de Office 2013

	Coeficientes	Error típico
Intercepción	52.3178305	12.23736
Superficie (x1)	0.02764139	0.00542937
Oficinas (x2)	12.5297682	0.40006684
Entradas (x3)	2.55321066	0.53066915
Antigüedad (x4)	-0.2342372	0.01326801

REGRESIÓN LINEAL MÚLTIPLE: Ejercicio

Se ha recogido datos sobre usuarios de cierto servicio en la que además se realizan actividades tanto para niños como para adolescentes y adultos, y existe interés en analizar cuáles son las variables que determinan el nivel de satisfacción de sus usuarios; las variables recogidas son: afición a la lectura, al cine, a la música, número de hijos, renta… y, por supuesto, nivel de satisfacción.

REGRESIÓN LINEAL MÚLTIPLE: Ejercicio

Aficion_lectura	Num_hijos	Aficion_cine	Aficion_musica	renta_mens	Nivel_estudios	Aficion_TV	Satisfaccion
4	0	3	5	1200	4	4	4
3	0	3	4	1500	5	4	3
5	1	4	1	1800	3	5	5
2	2	1	3	1000	2	2	3
4	1	5	3	1300	3	4	4
3	1	3	4	1900	1	4	3
5	3	4	5	1300	4	5	5
3	0	2	3	1200	4	4	3
3	1	4	1	1600	2	5	4
1	3	2	1	1400	2	1	2
4	0	5	4	1700	3	4	4
5	0	5	5	2500	4	5	5
5	2	4	4	1100	5	3	5
5	2	5	3	1400	3	4	5
2	1	1	4	1800	4	3	3
4	2	5	4	2000	4	5	5
3	3	2	4	1500	4	3	3
1	1	2	3	1000	2	2	2
2	1	2	2	1300	3	3	3
1	0	2	5	1600	4	4	2
5	1	4	4	1800	3	4	4
2	2	3	3	1200	4	4	4
4	1	5	5	1700	2	5	4
4	1	4	3	1500	5	4	4
5	2	4	5	1100	5	5	5

REGRESIÓN LINEAL MÚLTIPLE: Ejercicio

Resolver utilizando algún software.

BIBLIOGRAFIA

1. Anderson, D. R., Sweeney, D. J. y Williams, T. A. Estadística para Administración y Economía. Décima edición. Cengage Editores.

México. 2008.

2. Fernández. A. C. Manual de estadística descriptiva aplicada al sector turístico. Ed. Síntesis. España. 2001.

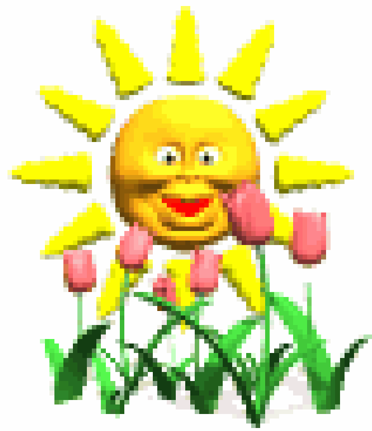
3. Ferran M. SPSS para Windows. Análisis Estadístico. Ed Mc Graw Hill. México 2001

4. Infante, S. G. y Zárate de L. G. Métodos Estadísticos. Ed. Trillas. México. 2000.

5. Levine, D. M., Krehbiel, T. C. y Berenson, M. L. Estadística para administración. Cuarta edición. Pearson. México. 2006

6. Lind Douglas A.,Marchal William G.,Wathen Samuel A. . Estadística aplicada a los negocios la economía. Décimo Tercera edición .Mc
Graw Hill 2008.

7. Riquelme P. Tablas y Gráficos en investigaciones. 2004.



FIN DE LA PRESENTACION