Regresión Lineal Simple

Irene Patricia Valdez y Alfaro
Marco Antonio Gómez Ramírez

Junio de 2015

Regresión

El análisis de regresión es una técnica estadística para investigar la **relación funcional entre dos o más variables**, ajustando algún modelo matemático.

En la práctica profesional de la Ingeniería es frecuente que se requiera resolver problemas que implican un conjunto de variables, por ejemplo, en la fabricación de un artículo doméstico, en términos generales, para mejorar la calidad del producto, debemos conocer: El proceso de fabricación, la materia prima, la capacitación de los obreros, la actualización del equipo de producción, etc., estas podrían ser las variables que nos interesan para la producción del artículo, con ellas elaboraríamos un modelo matemático $y=\beta_0+ \beta_1 x_1+ \beta_2 x_2 + \beta_3 x_3+\ldots.. + \beta_n x_n$, que nos permitiría conocer la forma de producción actual y pronosticar el mejoramiento del producto al modificar alguna(s) de las variables, x_1, x_2, x_3.... x_4, en este caso son llamadas variables de regresión.

Regresión lineal simple

La regresión lineal simple utiliza una sola variable de regresión y el caso más sencillo es el modelo de línea recta.

Por ejemplo las horas de estudio con las calificaciones obtenidas por los alumnos, aquí podemos diferenciar la variable dependiente o respuesta Y (calificación) y la variable independiente o regresor X (horas de estudio).

Una forma sencilla de relacionarlas es mediante la **Recta de Regresión** $y = \beta_0 + \beta_1 x$, donde β_0 y β_1 son parámetros o coeficientes de regresión, que son estimados a partir de los datos reales, β_0 es la ordenada al origen de la recta de regresión y β_1 la pendiente correspondiente.

Para estimar los parámetros usamos el método de mínimos cuadrados, por lo que necesitamos elaborar un tabla con la información siguiente:

	Por ejemplo deseamos conocer la interrelación de las calificaciones con respecto a las horas de estudio de quince alumnos, los datos obtenidos fueron los siguientes:					
	X(horas)	Y(cal.)	X^2	Y^2	XY	
	8	10	64	100	80	
	6	6	36	36	36	
	2	4	4	16	8	
	2	3	4	9	6	
	4	5	16	25	20	
	8	8	64	64	64	
	5	6	25	36	30	
	9	10	81	100	90	
	5	5	25	25	25	
	9	9	81	81	81	
	7	8	49	64	56	
	7	7	49	49	49	
	6	7	36	49	42	
	3	5	9	25	15	
	7	9	49	81	63	
Sumas	88	102	592	760	665	

Formulario para la recta de regresión lineal simple, por el método de mínimos cuadrados

Modelo matemático $$Y = \beta_0 + \beta_1 X$$

β_0 y β_1 son constantes desconocidas (parámetros del modelo de regresión)

donde: $$\beta_0 = \bar{y} - \beta_1 \bar{x}$$ $$\beta_1 = \frac{S_{xy}}{S_{xx}}$$

Suma de cuadrados de X

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

Suma de cuadrados de Y

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

Suma de productos cruzados de X y Y

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) y_i = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}$$

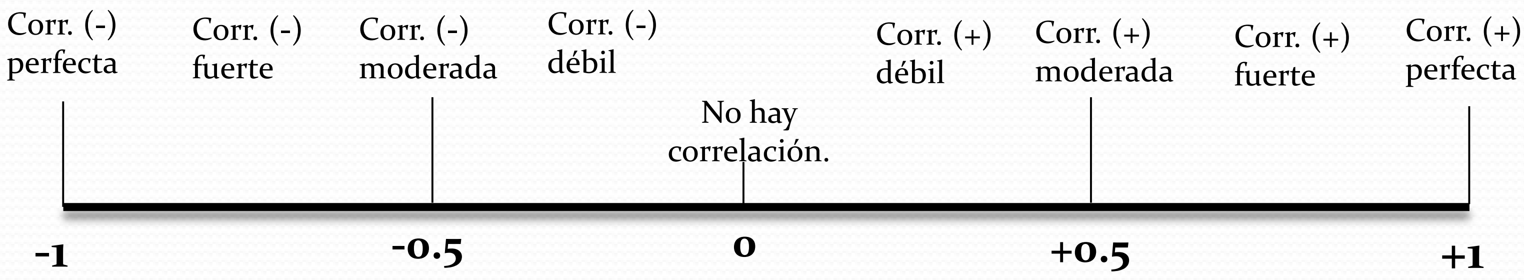
Coeficiente de correlación:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

Coeficiente de determinación:

$$R^2 = r^2$$

El coeficiente de correlación (r) mide la bondad del ajuste realizado, es decir, cuantifica la relación lineal entre las dos variables.



El coeficiente de determinación (R^2) proporciona un indicador sobre la calidad del modelo, es decir, si lo podemos usar para predecir futuros resultados, por ejemplo, un coeficiente R^2=0.905, significa que el 90.5% de la variabilidad es explicada por el modelo, en otras palabras, el modelo es confiable en un 90.5% para la predicción de futuras resultados.

Cálculos del ejercicio

n=	15		$\beta_0 =$	0.8794
$\bar{X} =$	5.8667		$\beta_1 =$	1.6408
$\bar{Y} =$	6.8		r=	0.9392
$SS_{XX}=$	75.7333		$R^2=$	0.8821
$SS_{YY}=$	66.4			
$SS_{XY}=$	66.6		$Y = 0.8794 - 1.6408X$	

Gráfica

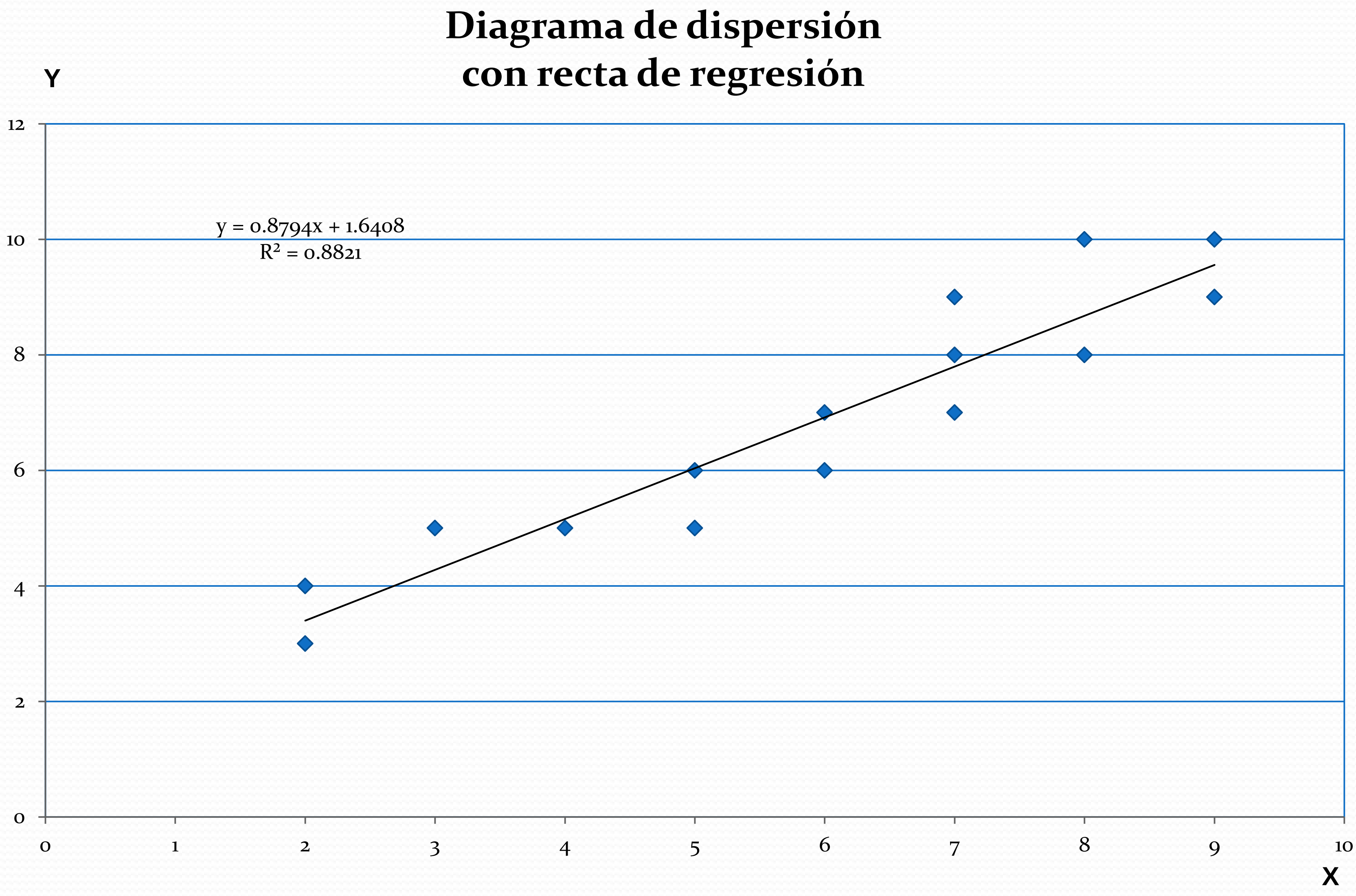


Diagrama de dispersión
con recta de regresión
Y
y = 0.8794x + 1.6408
R² = 0.8821
0
2
4
6
8
10
12
0
1
2
3
4
5
6
7
8
9
10
X