UNIVERSIDAD DE SAN PEDRO SULA

4. Regresión Lineal Simple

Introducción

Una vez conociendo las medidas que se utilizan para expresar la fuerza y la dirección de la relación lineal entre dos variables, se tienen elementos base para incursionar en el tema de los pronósticos. Toda empresa necesita hacer pronósticos, ya sea para producir o para vender.

Análisis de regresión

Una de las técnicas más utilizadas es la que recurre a la ecuación de regresión que se base en una recta lineal en la cual se estima el valor de Y basándose en un valor X. Esta técnica solo pronostica tendencias lineales; si una empresa no tiene un comportamiento lineal, los resultados que se obtienen no son adecuados.

> "ECUACIÓN DE REGRESIÓN: Ecuación que expresa la relación lineal entre dos variables." (Lind |Marchal |Wathen, 2008, p.470).

Principios de mínimos cuadrados

Es la ecuación que se extrae de la interpolación de los puntos de un diagrama de dispersión, la posición de cada uno de los puntos y su respectiva variación permite trazar una recta que se denomina Tendencia y con la cual es posible establecer pronósticos y se elimina el juicio subjetivo.

> "PRINCIPIO DE MÍNIMOS CUADRADOS: Determina una ecuación de regresión al minimizar la suma de los cuadrados de las distancias verticales entre los valores reales de "Y" y los valores pronosticados de "Y"." (Lind |Marchal |Wathen, 2008, p.471).

Desde el punto de vista de la matemática, una ecuación lineal tiene la forma de Y=a + bX; en donde "b" se conoce como la "pendiente de la ecuación" y "a" es el intercepto en el eje Y. La fórmula general de una ecuación de regresión lineal se denota como:

ECUACIÓN DE REGRESIÓN	$\hat{Y} = a + bX$

Donde:

$\hat{Y}$: Valor de pronóstico
X : Variable independiente
a : Intersección en Y
b : Pendiente de la ecuación de regresión

Tanto el intercepto como la pendiente son calculados a través de las siguientes fórmulas:

USAP
UNIVERSIDAD DE SAN PEDRO SULA

PENDIENTE DE LA RECTA DE REGRESIÓN	$b = r\left(\frac{s_Y}{s_X}\right)$

Donde:

b	:	Valor de pronóstico
r	:	Coeficiente de correlación
s_Y	:	Desviación estándar de la variable dependiente
s_X	:	Desviación estándar de la variable independiente

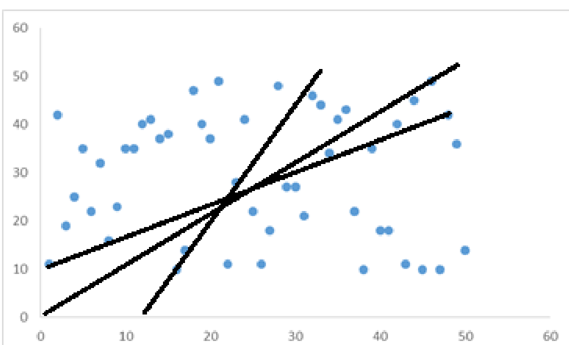
INTERCEPTO CON EL EJE "Y"	$a = \bar{Y} - b\bar{X}$

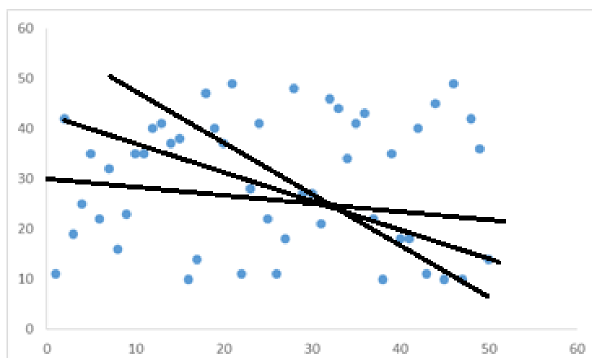
Donde:

a	:	Intercepto con el eje "Y"
$\bar{Y}$	:	Media aritmética de la variable dependiente
$\bar{X}$	:	Media aritmética de la variable independiente
s_X	:	Desviación estándar X

El procedimiento usual para determinar la ecuación de regresión cuando se tiene una muestra es el siguiente:

1. Calcular la media aritmética de cada variable de la muestra
2. Calcular las variaciones simples y cuadradas
3. Calcular la desviación estándar de cada variable de la muestra
4. Calcular el valor de la pendiente (b)
5. Calcular el valor del intercepto (a)
6. Construir la ecuación con los valores resultantes
7. Trazar la recta sobre el diagrama de dispersión





Ejemplo 4.1

1. En una empresa multinacional se estudió la relación entre las ventas reportadas y los gastos generados por publicidad. La información del último cuatrimestre, en millones de dólares, se detalla a continuación:

Mes	Gastos en publicidad	Ganacias por ventas
Julio	2	7
Agosto	1	3
Septiembre	3	8
Octubre	4	10

UNIVERSIDAD DE SAN PEDRO SULA

a. Determinar la ecuación de regresión
b. Estimar las ventas cuando se gastan 3 millones en publicidad.

Desarrollo

a. Media aritmética de ambas variables

Mes	Gastos en publicidad (X)	Ganancias por ventas (Y)
Julio	2	7
Agosto	1	3
Septiembre	3	8
Octubre	4	10
∑	10	28
$\bar{X}, \bar{Y}$	**2.5**	**7.0**

b. Variación simple y cuadrada

Mes	Gastos en publicidad (X)	Ganancias por ventas (Y)	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X}) * (Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
Julio	2	7	-0.5	0	-	0.25	0
Agosto	1	3	-1.5	-4	6.0	2.25	16
Septiembre	3	8	0.5	1	0.5	0.25	1
Octubre	4	10	1.5	3	4.5	2.25	9
∑	10	28			**11.0**	**5.0**	**26.0**

c. Desviación estándar

$$s_X = \sqrt{\frac{5.0}{4-1}} = 1.29$$

$$s_Y = \sqrt{\frac{26.0}{4-1}} = 2.94$$

d. Coeficiente de correlación

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_X s_Y} = \frac{11}{(4-1)(1.29)(2.94)} = 0.967$$

e. Pendiente de la ecuación de regresión

$$b = r\left(\frac{s_Y}{s_X}\right) = 0.967\left(\frac{2.94}{1.29}\right) = 2.2$$

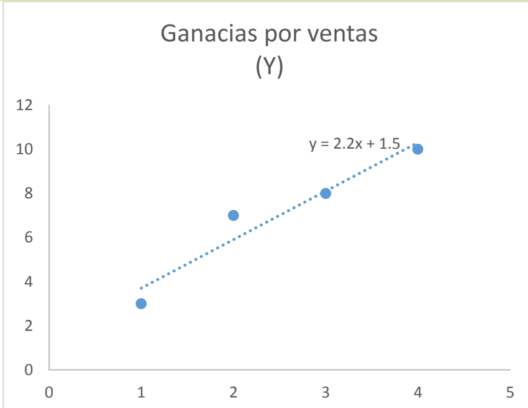
f. Intercepto de la ecuación de regresión

$$a = \bar{Y} - b\bar{X} = 7 - (2.2)(2.5) = 1.5$$

g. Ecuación de regresión

$$\hat{Y} = 1.5 + 2.2X$$

h. Gráfica de ecuación de regresión



i. Pronóstico o estimación para X=3 millones de dólares en publicidad

$\boldsymbol{X = 3} \qquad \boldsymbol{\hat{Y} = 1.5 + 2.2(3) = 8.4}$

2. La empresa Sara determinó que existe una fuerte relación entre las llamadas realizadas por 6 empleados durante el mes de abril y las ventas de unidades de aire acondicionado facturadas.

   Determinar la ecuación de regresión y estimar cuántas unidades de aire acondicionado se deberían vender cuando un vendedor hace 20 llamadas.

AGENTE	LLAMADAS	UNIDADES VENDIDAS
Tomás García	20	30
José Girón	40	60
Gregorio Figueroa	30	60
Carlos Ramírez	10	40
Miguel Godoy	20	50
Marcos Reyes	20	30

Desarrollo

a. Calcular la media aritmética de cada variable

AGENTE	LLAMADAS	UNIDADES VENDIDAS
Tomás García	20	30
José Girón	40	60
Gregorio Figueroa	30	60
Carlos Ramírez	10	40
Miguel Godoy	20	50
Marcos Reyes	20	30

$$\boldsymbol{\bar{X} = \frac{140}{6} = 23.0}$$

$$\boldsymbol{\bar{Y} = \frac{270}{6} = 45.0}$$

b. Calcular las variaciones simples y cuadradas

AGENTE	LLAMADAS	UNIDADES VENDIDAS	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X}) * (Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
Tomás García	20	30	$(20 - 23)^2 =$ -3	$(20 - 45)^2 =$ -15	45	9	225
José Girón	40	60	$(40 - 23)^2 =$ 17	$(60 - 45)^2 =$ 15	255	289	225
Gregorio Figueroa	30	60	$(30 - 23)^2 =$ 7	$(60 - 45)^2 =$ 15	105	49	225
Carlos Ramírez	10	40	$(10 - 23)^2 =$ -13	$(40 - 45)^2 =$ -5	65	169	25
Miguel Godoy	20	50	$(20 - 23)^2 =$ -3	$(50 - 45)^2 =$ 5	-15	9	25
Marcos Reyes	20	30	$(20 - 23)^2 =$ -3	$(30 - 45)^2 =$ -15	45	9	225
Σ	140	270			500	534	950

USAP UNIVERSIDAD DE SAN PEDRO SULA

c. Calcular la desviación estándar de cada variable

$$s_X^2 = \frac{534}{6-1} = 106.7$$

$$s_Y^2 = \frac{950}{6-1} = 190$$

$$s_X = \sqrt{106.7} = 10.3$$

$$s_Y = \sqrt{190} = 13.8$$

d. Calcular el coeficiente de correlación

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{(n-1)s_X s_Y}$$

$$r = \frac{500}{(6-1)(10.3)(13.8)} = 0.702$$

e. Calcular la pendiente de la ecuación de regresión

$$b = r\left(\frac{s_Y}{s_X}\right) = 0.702\left(\frac{13.8}{10.3}\right) = 0.702(1.33) = 0.93$$

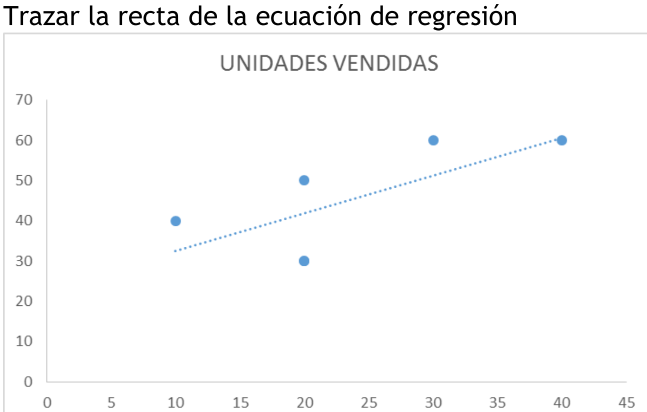
f. Calcular el intercepto de la ecuación

$$a = \bar{Y} - b\bar{X} = 45.0 - (0.93)(23) = 23.61$$

g. Construir la ecuación

$$\hat{Y} = 23.61 + 0.93X$$

h. Trazar la recta de la ecuación de regresión



Error estándar de estimación

Por lo general, no todos los puntos del diagrama de dispersión están ubicados sobre la recta que resulta de la ecuación de regresión lineal; en el trazo de toda ecuación de regresión siempre se presenta un error por las múltiples variaciones que se presentan; a este error se le llama "Error estándar de estimación".

ERROR ESTÁNDAR DE ESTIMACIÓN: Medida de dispersión de los valores observados respecto de la recta de regresión." (Lind |Marchal |Wathen, 2008, p.478).

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Si el error estándar de la estimación es pequeño, los datos están relativamente cercanos a la recta de la ecuación de regresión lineal y significa que se predice Y con poco error; si es grande, los datos están muy dispersos de la recta de la ecuación de regresión lineal y no proporciona una estimación precisa de Y.

ERROR ESTÁNDAR DE LA ESTIMACIÓN	$s_{Y\cdot X} = \sqrt{\frac{\Sigma(Y-\hat{Y})^2}{n-2}}$

Ejemplo 4.2

1. En una empresa multinacional se estudió la relación entre las ventas reportadas y los gastos generados por publicidad. La información del último cuatrimestre, en millones de dólares, se detalla a continuación:

Mes	Gastos en publicidad (X)	Ganancias por ventas (Y)
Julio	2	7
Agosto	1	3
Septiembre	3	8
Octubre	4	10

Si la ecuación de regresión de la muestra es $\hat{Y} = \mathbf{1.5 + 2.2X}$, calcular el error estándar de la estimación

Desarrollo

a. Calcular el pronóstico

Mes	Gastos en publicidad (X)	Ganancias por ventas (Y)	VENTAS PRONOSTICADAS $\hat{Y}$
Julio	2	7	1.5 + 2.2 (2) = 5.9
Agosto	1	3	1.5 + 2.2 (1) = 3.7
Septiembre	3	8	1.5 + 2.2 (3) = 8.1
Octubre	4	10	1.5 + 2.2 (4) = 10.3

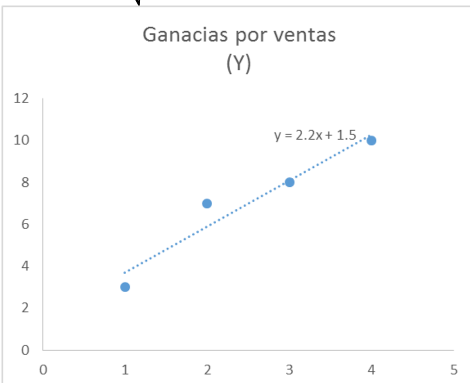
b. Calcular la variación cuadrada de Y con respecto al pronóstico.

Mes	Ganancias por ventas (Y)	$\hat{Y}$	$(Y-\hat{Y})^2$
Julio	7	5.9	$(7 - 5.9)^2 =$ 1.21
Agosto	3	3.7	$(3 - 3.7)^2 =$ 0.49
Septiembre	8	8.1	$(8 - 8.1)^2 =$ 0.01
Octubre	10	10.3	$(10 - 10.3)^2 =$ 0.09
		Σ	1.80

USAP
UNIVERSIDAD DE SAN PEDRO SULA

c. Calcular el error estándar de la estimación

$$s_{Y\cdot X} = \sqrt{\frac{1.8}{4-2}} = \sqrt{0.9} = 0.95$$



2. La gerente de ventas de Sara, determinó que la ecuación de regresión de mínimos cuadrados era $\hat{Y} = 23.61 + 0.93X$, donde Y es el número de unidades que se pronostican se venderán en el siguiente mes y base al número X de llamadas de parte de los vendedores. Determinar el error estándar de la estimación como una medida para observar qué tan bien se ajustan los valores a la recta de regresión.

AGENTE	LLAMADAS	UNIDADES VENDIDAS
Tomás García	20	30
José Girón	40	60
Gregorio Figueroa	30	60
Carlos Ramírez	10	40
Miguel Godoy	20	50
Marcos Reyes	20	30
Σ	140	270

Desarrollo

a. Ventas pronosticadas para la ecuación de regresión $\hat{Y} = 23.61 + 0.93X$

AGENTE	LLAMADAS (X)	UNIDADES VENDIDAS	VENTAS PRONOSTICADAS $\hat{Y}$	
Tomás García	20	30	23.61 + 0.92 (20) =	42.2
José Girón	40	60	23.61 + 0.92 (40) =	60.8
Gregorio Figueroa	30	60	23.61 + 0.92 (30) =	51.5
Carlos Ramírez	10	40	23.61 + 0.92 (10) =	32.9
Miguel Godoy	20	50	23.61 + 0.92 (20) =	42.2
Marcos Reyes	20	30	23.61 + 0.92 (20) =	42.2

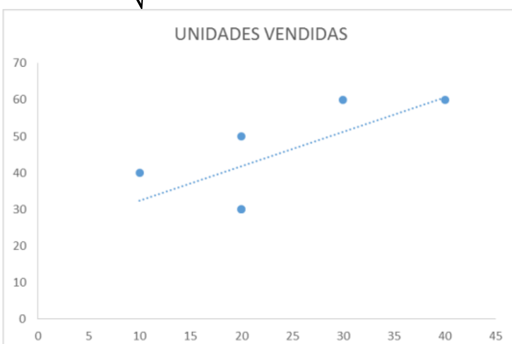
b. Calcular la variación cuadrada de Y con respecto al pronóstico.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

AGENTE	UNIDADES VENDIDAS (Y)	$\hat{Y}$	$(Y - \hat{Y})^2$	
Tomás García	30	42.2	$(30 - 42.2)^2 =$	149.08
José Girón	60	60.8	$(60 - 60.8)^2 =$	0.66
Gregorio Figueroa	60	51.5	$(60 - 51.5)^2 =$	72.08
Carlos Ramírez	40	32.9	$(40 - 32.9)^2 =$	50.27
Miguel Godoy	50	42.2	$(50 - 42.2)^2 =$	60.68
Marcos Reyes	30	42.2	$(30 - 42.2)^2 =$	149.08
				481.86

c. Error estándar de la estimación

$$s_{Y \cdot X} = \sqrt{\frac{481.86}{6-2}} = \sqrt{120.46} = 10.98$$



Intervalos de confianza e intervalos de predicción

Una de las principales utilidades del error estándar de estimación es la oportunidad de establecer intervalos de confianza cuando el tamaño de la muestra es grande y la dispersión de la recta de regresión se aproxima a una distribución normal.

En pronósticos se hace uso de dos tipos de intervalos; los de confianza y los de predicción y cada uno tiene sus propias características, entre las que destacan:

a. Intervalo de confianza: Reporta el valor medio de Y con respecto a la muestra.
b. Intervalo de predicción: Reporta un rango de valores de Y para un valor dado de X.

El intervalo de confianza se auxilia de la distribución t-Student para establecer el mejor intervalo de confianza. El método para calcular el valor de t es el siguiente:

a. Definir el nivel de confianza
b. Del tamaño de la muestra reducir 2 unidades (1 por cada variable)
c. Definir el número de las colas (2).
d. Buscar el resultado en la tabla

Ejemplo 4.3

1. Determinar el valor de t para una muestra con dos variables de tamaño 10 a dos colas al 95% de confiabilidad.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Desarrollo

Tamaño de la muestra:	n=10
Nivel de confianza:	95%
Total de colas	2
Grados de libertad:	gl=10-2 = 8
Nivel de significancia:	α=0.05

Resultado: t = 2.306

2. Con un 90% de confiabilidad, calcular el valor de t para una muestra de variables de tamaño 8 a 2 colas,

Desarrollo

Tamaño de la muestra:	n=8
Nivel de confianza:	90%
Total de colas	2
Grados de libertad:	gl=8-2 = 6
Nivel de significancia:	α=0.10
Resultado:	t=1.943

Intervalo de Confianza

Para determinar el intervalo de confianza del valor medio de Y para una X dada, la fórmula es la siguiente:

INTERVALO DE CONFIANZA PAR LA MEDIA DE Y, DADA X	$\hat{Y} \pm t * s_{Y \cdot X} \sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\sum(X-\bar{X})^2}}$

Intervalo de Predicción

Para determinar el intervalo de predicción de un valor particular de Y para una X dada, la fórmula se transforma de la siguiente manera:

INTERVALO DE CONFIANZA PAR LA MEDIA DE Y, DADA X	$\hat{Y} \pm t * s_{Y \cdot X} \sqrt{1 + \frac{1}{n} + \frac{(X-\bar{X})^2}{\sum(X-\bar{X})^2}}$

El intervalo de confianza es determinante para el total de la muestra, mientras que el intervalo de predicción es solamente para uno de los miembros de la muestra.

Ejemplo 4.4

1. En una agencia de publicidad, los pronósticos de las ganancias por ventas (Y) están relacionados con los gastos en publicidad (X) en base a la ecuación $\hat{Y}$= 1.5+2.2X, con 8.49 de error estándar en la estimación. En una muestra de 10 cuentas el gasto promedio en publicidad fue de 7 millones de Lempiras y una variación aleatoria de 5 millones. Con una confiabilidad del 95% determinar el intervalo de confianza para cuando la empresa gaste 5 millones de lempiras. Predecir cuál sería la predicción de las ventas de Juan Pérez si se le asignaran 5 millones de lempiras para publicidad.

Desarrollo

a. Datos conocidos

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Variable independiente (X): Gastos en publicidad
Variable dependiente (Y): Ganancias en ventas

Ecuación de regresión: $\hat{Y} = 1.5 + 2.2X$
Error estándar de estimación: $s_{Y \cdot X} = 2.49$
Tamaño de la muestra: $n = 10$
Media aritmética (X): 7
Variación aleatoria ($\sum(X - \bar{X})^2$): 5
Nivel de confianza: 95%
Valor a pronosticar: X=5

b. Determinar el valor de t en la distribución t-Student
Nivel de confianza: 95%
Nivel de significancia: $\alpha = 0.05$
Tamaño de la muestra: $n = 10$
Grados de libertad: $gl = 10 - 2 = 8$
Valor crítico: t= 2.306

c. Calcular el pronóstico de las ganancias para X=5.
X=5
$$\hat{Y} = 1.5 + 2.2(5)$$
$$\hat{Y} = 1.5 + 11.0$$
$$\hat{Y} = 12.5$$

d. Calcular el intervalo de confianza

$$IC_{95\%} = \hat{Y} \pm t * s_{Y \cdot X}\sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2}}$$

$$IC_{95\%} = 12.5 \pm (2.306)(2.49)\sqrt{\frac{1}{10} + \frac{(5 - 7)^2}{5}}$$

$$IC_{95\%} = 12.5 \pm (5.742)\sqrt{0.1 + \frac{4}{5}}$$

$$IC_{95\%} = 12.5 \pm (5.742)\sqrt{0.1 + 0.8}$$
$$IC_{95\%} = 12.5 \pm (5.742)\sqrt{0.9}$$
$$IC_{95\%} = 12.5 \pm (5.742)(0.9487)$$
$$IC_{95\%} = 12.5 \pm (5.45)$$

$$IC_{95\%} = \begin{cases} 12.5 - 5.45 \\ 12.5 + 5.45 \end{cases}$$

$$IC_{95\%} = \begin{cases} 7.05 \\ 17.95 \end{cases}$$

Con un 95% de confianza se puede prever una ganancia entre 7 y 16 millones.

e. Calcular el intervalo de predicción

$$IP_{95\%} = \hat{Y} \pm t * s_{Y \cdot X}\sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2}}$$

$$IP_{95\%} = 12.5 \pm (2.306)(2.49)\sqrt{1 + \frac{1}{10} + \frac{(5 - 7)^2}{5}}$$

$$IP_{95\%} = 12.5 \pm (5.742)\sqrt{1 + 0.1 + \frac{4}{5}}$$

$$IP_{95\%} = 12.5 \pm (5.742)\sqrt{1.9}$$
$$IP_{95\%} = 12.5 \pm (5.742)(1.3784)$$
$$IP_{95\%} = 12.5 \pm 7.91$$
$$IP_{95\%} = \begin{cases} 12.5 - 7.91 \\ 12.5 + 7.91 \end{cases}$$

$$IP_{95\%} = \begin{cases} 4.59 \\ 20.41 \end{cases}$$

Con un 95% de confianza se puede predecir que Juan Pérez puede obtener una ganancia entre 5 y 20 millones.

2. La gerente de ventas de Sara, sabe que la ecuación de regresión de mínimos cuadrados es $\hat{Y} = 23.61 + 0.93X$, con error estándar de la estimación de 10.98 y 23 llamadas en promedio al mes. Determinar un intervalo de confianza del 95% si la empresa realiza 25 llamadas y un intervalo de predicción del 95% para cuando el agente Carlos hace 25 llamadas , en base a la siguiente muestra:

AGENTE	LLAMADAS (X)	UNIDADES VENDIDAS (Y)
Tomás García	20	30
José Girón	40	60
Gregorio Figueroa	30	60
Carlos Ramírez	10	40
Miguel Godoy	20	50
Marcos Reyes	20	30

Desarrollo

a. Datos conocidos

Ecuación de regresión: $\hat{Y} = 23.61 + 0.93X$
Media aritmética de X $\bar{X} = 23$
Error estándar de estimación: $s_{Y \cdot X} = 10.98$
Tamaño de la muestra: $n = 6$

b. Calcular la sumatoria de la variación

AGENTE	LLAMADAS (X)	UNIDADES VENDIDAS (Y)	$(X - \bar{X})^2$
Tomás García	20	30	$(20 - 23)^2 =$ 9
José Girón	40	60	$(40 - 23)^2 =$ 289
Gregorio Figueroa	30	60	$(30 - 23)^2 =$ 49
Carlos Ramírez	10	40	$(10 - 23)^2 =$ 169
Miguel Godoy	20	50	$(20 - 23)^2 =$ 9
Marcos Reyes	20	30	$(20 - 23)^2 =$ 9
			∑ 534

c. Calcular el pronóstico para X=25

$$\hat{Y} = 23.61 + 0.93(25) = 23.61 + 23.25 = 46.9$$

d. Determinar el valor de t en la tabla de la Distribución t-Student

Tamaño de la muestra: n=6
Nivel de confianza: 95%

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Total de colas 2
Grados de libertad: gl=6 - 2 = 4
Nivel de significancia: α=0.05

	Intervalo de confianza, c					
	80%	90%	95%	98%	99%	99.9%
	Nivel de significancia para una prueba de una cola, α					
gl	0.100	0.050	0.025	0.010	0.005	0.0005
	Nivel de significancia para una prueba de dos colas, α					
	0.200	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869

Resultado: t=2.776

e. Intervalo de confianza

$$IC_{95\%} = \hat{Y} \pm t * s_{Y \cdot X}\sqrt{\frac{1}{n} + \frac{(X-\bar{X})^2}{\sum(X-\bar{X})^2}}$$

$$IC_{95\%} = 46.9 \pm 2.776(10.98)\sqrt{\frac{1}{6} + \frac{(25-23)^2}{534}}$$

$$IC_{95\%} = 46.9 \pm 30.48\sqrt{0.167 + 0.0075}$$
$$IC_{95\%} = 46.9 \pm 30.48\sqrt{0.1745}$$
$$IC_{95\%} = 46.9 \pm 30.48(0.4177)$$
$$IC_{95\%} = 46.9 \pm 12.84$$

$$IC_{95\%} = \begin{cases} 46.9 - 12.73 \\ 46.9 + 12.73 \end{cases}$$

$$IC_{95\%} = \begin{cases} 34.17 \\ 59.63 \end{cases}$$

f. Intervalo de predicción

$$IP_{95\%} = \hat{Y} \pm t * s_{Y \cdot X}\sqrt{1 + \frac{1}{n} + \frac{(X-\bar{X})^2}{\sum(X-\bar{X})^2}}$$

$$IP_{95\%} = 46.9 \pm 2.776(10.98)\sqrt{1 + \frac{1}{6} + \frac{(25-23)^2}{534}}$$

$$IP_{95\%} = 46.9 \pm 30.48\sqrt{1 + 0.167 + 0.0075}$$
$$IP_{95\%} = 46.9 \pm 30.48\sqrt{1.1745}$$
$$IP_{95\%} = 46.9 \pm 30.48(1.0837)$$
$$IP_{95\%} = 46.9 \pm 33.03$$

$$IP_{95\%} = \begin{cases} 46.9 - 33.03 \\ 46.9 + 33.03 \end{cases}$$

$$IP_{95\%} = \begin{cases} 13.87 \\ 79.93 \end{cases}$$

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Ejercicio

Para cada uno de los ejercicios que se muestran a continuación realizar las siguientes actividades:

a. Determinar la ecuación de regresión lineal por el método de mínimos cuadrados
b. Calcular el error estándar de la estimación
c. Determinar el intervalo de confianza para el valor de X igual a 9
d. Determinar el intervalo de predicción para el caso que uno de los valores sea 9
e. Trazar la recta de la regresión lineal sobre el diagrama de dispersión

1. En una fábrica, la cantidad de artículos que se producen depende de la cantidad de empleados que se tengan contratados. Para la siguiente muestra que se seleccionó al azar, se obtuvieron los siguientes datos:

Empleados	4	5	7	6	8
Artículos	8	14	10	14	14

2. En una empresa que se dedica a hacer viajes a la ciudad de Tegucigalpa, los viajes que se realizan están relacionados con la cantidad de vehículos que se mantienen activos cada día. Una muestra de 8 días, reveló la cantidad de viajes que se habían realizado.

Vehículos	6	3	6	3	4	4	6	8
Viajes	14	15	7	12	13	11	9	7

3. La Empresa Nacional de Energía Eléctrica del país estudia la relación entre kilowatts-hora (miles) usados y el número de habitaciones en una residencia privada familiar. Una muestra aleatoria de 10 casas reveló lo siguiente:

Número de habitaciones	12	9	14	6	10	8
Kilowatts-hora (miles)	9	7	10	5	8	8

4. El señor James McWhinney, presidente de Daniel-James Financial Services, considera que hay una relación entre el número de contactos con sus clientes y la cantidad de ventas en dólares. Para documentar esta afirmación, el señor McWhinney reunió la siguiente información muestral. La columna X indica el número de contactos con sus clientes el mes anterior, y la columna Y muestra el valor de las ventas (miles de $) el mismo mes por cada cliente muestreado.

Número de contactos	14	12	20	16	46	23	48
Ventas (miles de dólares)	24	14	28	30	80	30	90

BIBLIOGRAFÍA

- Lind, D.A., Marchal, W.G., Wathen, S.A. (15). (2012). *Estadística Aplicada a los Negocios y la Economía*. México: McGraw-Hill

- David M. Levine, Timothy C. Krehbiel, Mark L. Berenson. 2006. *Estadística para Administración*. (4° edición). Naucalpan de Juárez, México.: Pearson Prentice Hall