

ESTADISTICA GENERAL

ASOCIAACION DE VARIABLES REGRESION LINEAL COEFICIENTE DE CORRELACION



Mg. Edward I. Terrones Gálvez



ESCENARIO

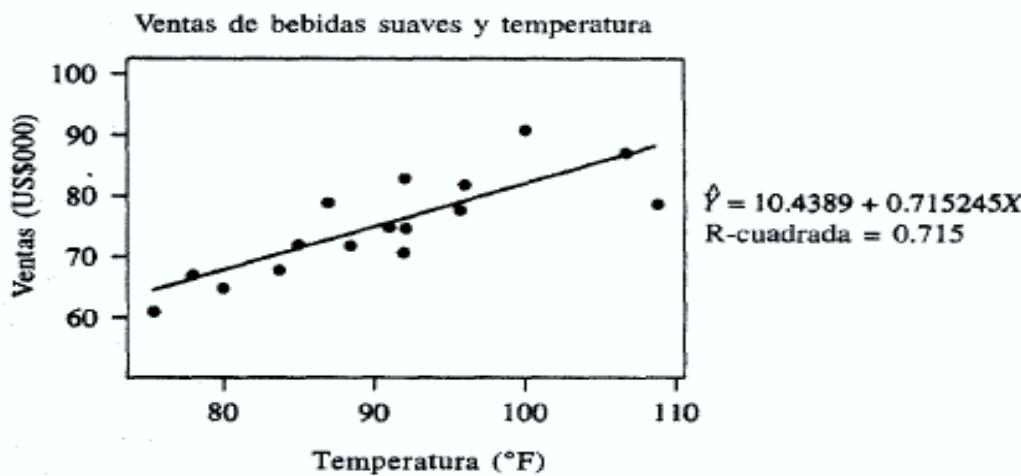
La competencia en la industria de las bebidas suaves siempre ha sido intensa. Recientemente, la lucha entre Coca-Cola y Pepsi-Cola se ha puesto álgida por incrementar sus participaciones respectivas de US\$27 mil millones en el mercado nacional de bebidas. Cada compañía ha ofrecido su propia marca de *flair* promocional en un esfuerzo continuo por reorganizar la mezcla en su mercadeo y promocionar su respectivo producto. Coca-Cola actualmente goza de un 21.7% de participación en el mercado, seguido de Pepsi al 18.9%.

Sin duda alguna los ejecutivos de mercadeo, los especialistas en gerencia y los estadísticos trabajan duro en ambas compañías intentando sobrepasar a sus contrapartes de mente competitiva. Hasta ahora se han puesto de acuerdo

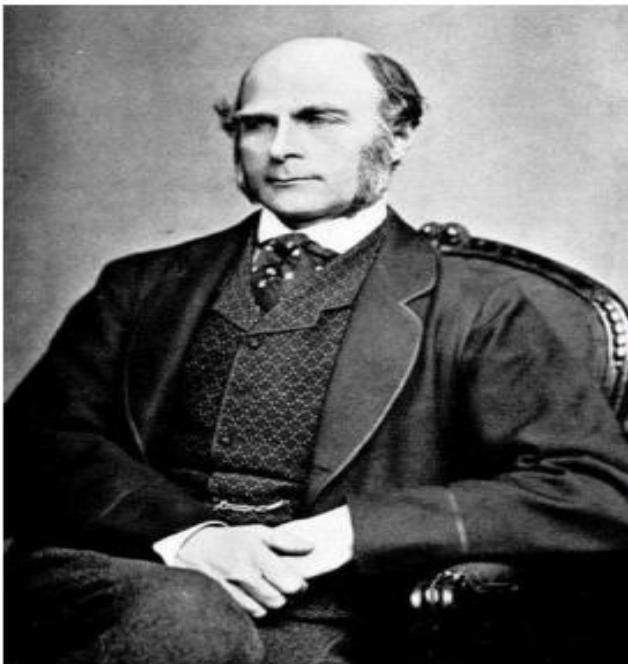
en muy poco, salvo en que las ventas parecen incrementar con las elevadas temperaturas del verano.

Predecir las tendencias en la participación de mercado es una tarea especialmente ardua y difícil. Muchos ejecutivos han echado a perder sus carreras en el frustrado intento de anticipar correctamente el comportamiento de los volubles consumidores.

La regresión y el análisis de correlación son las dos herramientas más poderosas y útiles que los analistas de todo tipo tienen a su disposición para escudriñar el interior del futuro sombrío. En este capítulo se analizarán estos procedimientos y se enseñará cómo ellos pueden orientar a los profesionales en negocios en su búsqueda de una carrera exitosa.



Un visionario: Sir Francis Galton



Antropólogo, geógrafo, explorador, inventor, meteorólogo, estadístico, psicólogo y eugenista británico, Francis Galton (1822-1911) es mundialmente reconocido por su afición a medir y cuantificar prácticamente todo. Cursó estudios en el King's College de la Universidad de Londres y en el Trinity College de Cambridge, aunque es notable que no tuvo cátedras universitarias y realizó la mayoría de sus investigaciones por su cuenta. Sus múltiples contribuciones recibieron reconocimiento formal cuando, a la edad de 87 años, se le concedió el título de Sir o Caballero del Reino.

De intereses muy variados, Galton contribuyó a diferentes áreas de la ciencia. A menudo sus investigaciones fueron continuadas dando lugar a nuevas disciplinas.

ASOCIACION ENTRE DOS VARIABLES

1. Regresión lineal simple

El término regresión fue originalmente utilizado por Francis Galton, para indicar ciertas relaciones en la teoría de la herencia biológica.

Consiste en determinar una relación funcional (recta de regresión) entre ellas, con el fin de que se pueda predecir el valor de una variable en base a la otra.

$\hat{y} = b_0 + b_1 x = a + bx$: Regresión de “y” en “x” donde:

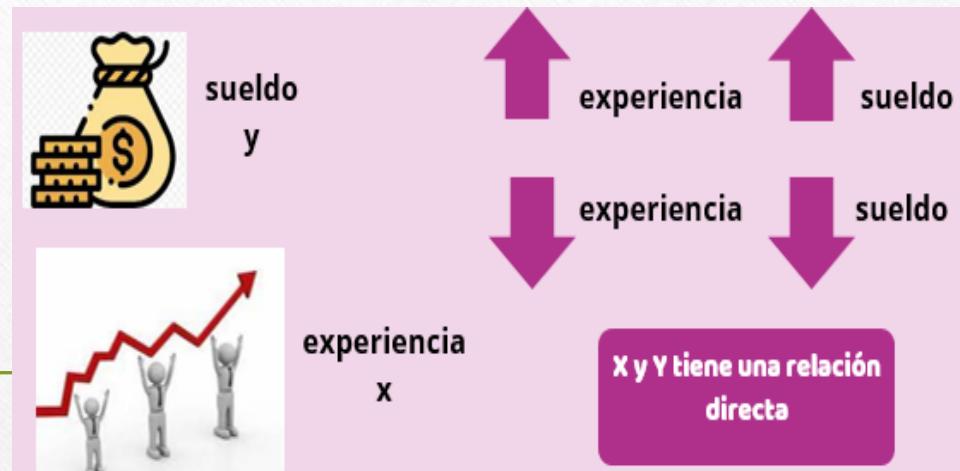
\hat{y} : Variable dependiente; variable que se desea explicar o predecir, también se le denomina regresando o variable respuesta.

x : Variable independiente; es la variable que supuestamente se conoce, también se le denomina predictora o explicativa.

b_0 ,a : Intercepto, ordenada en el origen o coeficiente de posición.

b_1 ,b : Pendiente, coeficiente angular o coeficiente de regresión; nos determina el ángulo de inclinación de la recta

Por ej: Deseamos predecir el sueldo mensual(en soles) en función a la experiencia de un trabajador



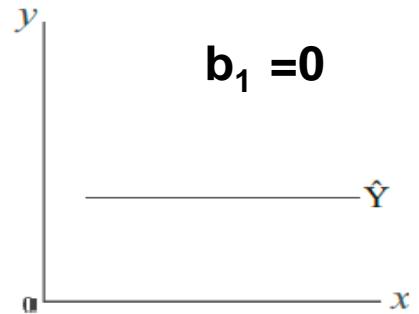
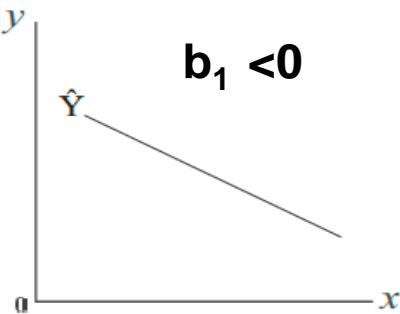
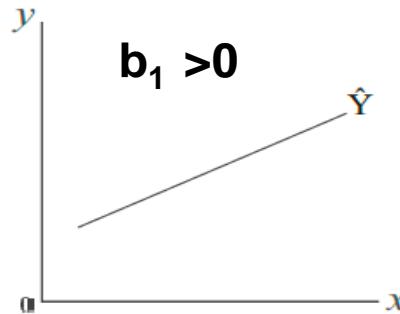
Deseamos predecir el tiempo de espera en una entidad bancaria
En función del numero de Trabajadores que contratataron



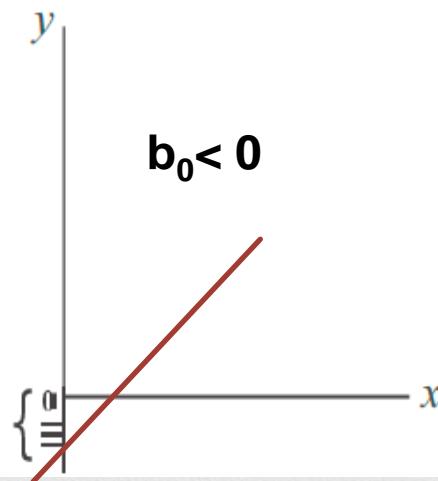
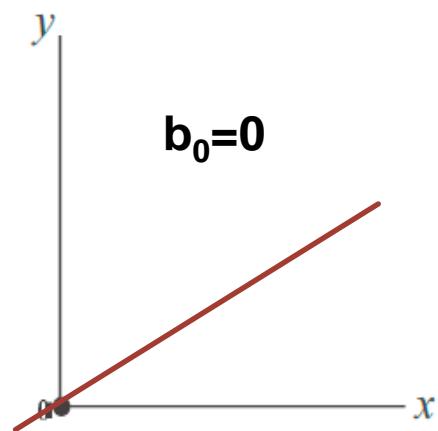
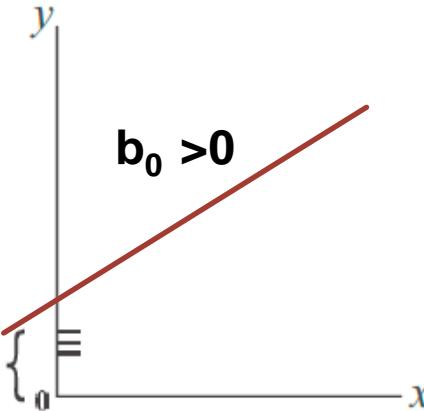
Ejemplos de relaciones entre dos variables

- ¿Existe alguna relación entre la cantidad que gasta una empresa en publicidad y sus ventas mensuales?
- ¿Existe relación entre la reducción de precios y las ventas?
- ¿Existe relación entre las horas trabajadas y la edad en los trabajadores de una empresa?
- ¿Existe relación entre el numero de horas que estudian los alumnos para un examen y la calificación que obtiene?

Representación del coeficiente angular o pendiente



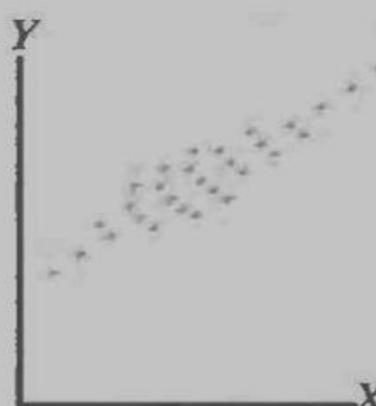
Ubicación del coeficiente de posición



Método de mínimos cuadrados: La recta de mejor ajuste

$$b_1 = \frac{\text{Cov}(x,y) = S_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y}) / (n-1)}{\sum(x_i - \bar{x})^2 / (n-1)}}{\frac{S_x^2}{S_x^2}} = b_1 = \frac{\sum x*y - \frac{(\sum x)*(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$
$$b_0 = \bar{y} - b_1 \bar{x}$$

Diagrama de dispersión o nube de puntos: Es la grafica de los $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, n valores de la variable bidimensional (X, Y) , observados en una muestra, en el sistema cartesiano.



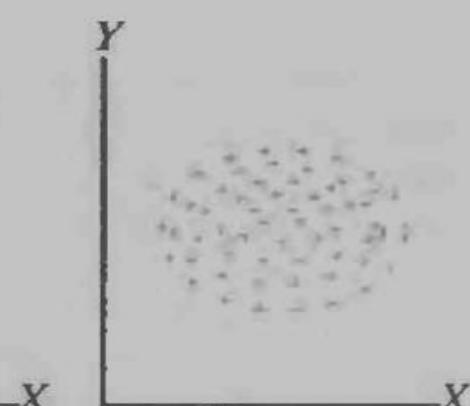
a) lineal positiva



b) lineal negativa



c) no lineal



d) ninguna relación

2. Análisis de correlación: Coeficiente de correlación de Pearson

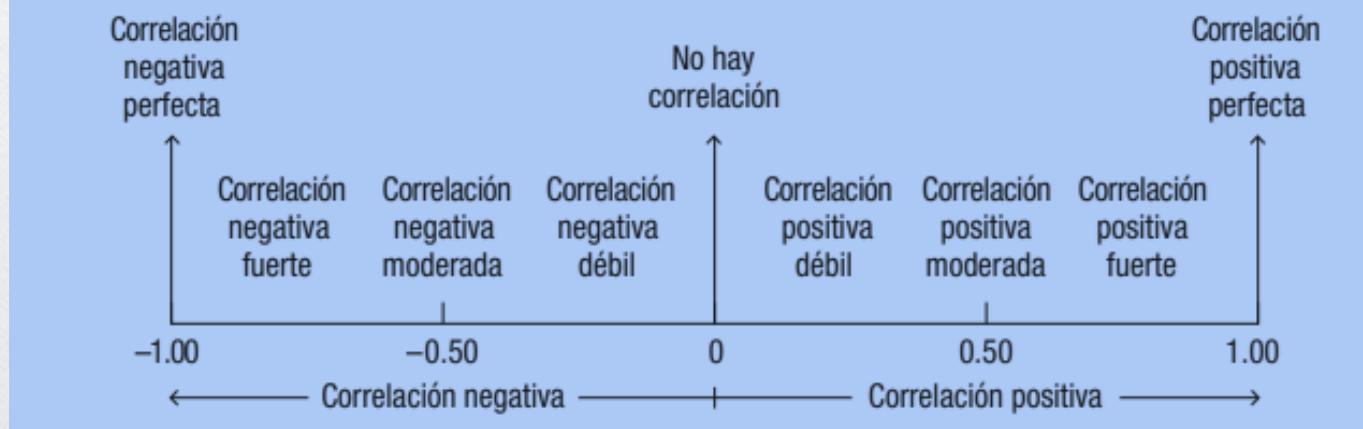
Fue propuesto por Carl Pearson y determina el grado de relación que puede haber entre dos variables.

$$r = \frac{\text{Cov}(x,y)}{\frac{S_x S_y}{S_x S_y}} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1) * S_x S_y}$$

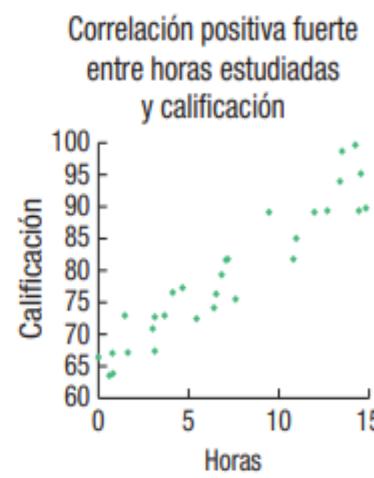
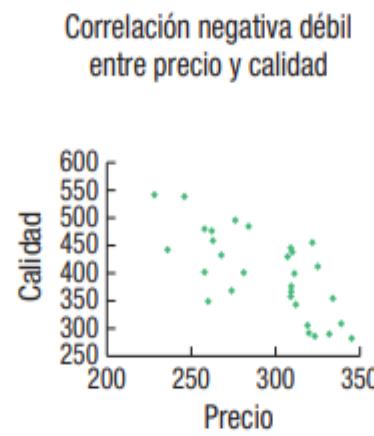
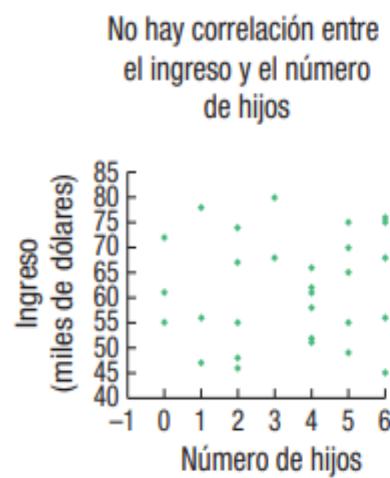
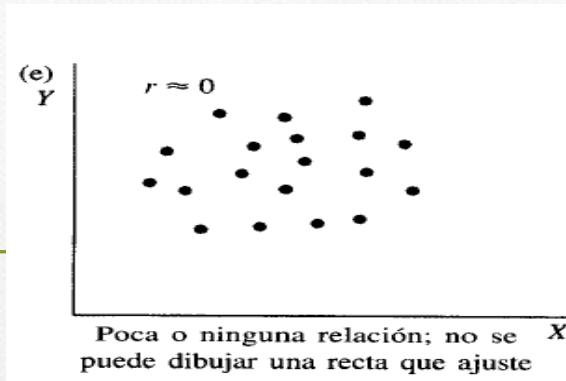
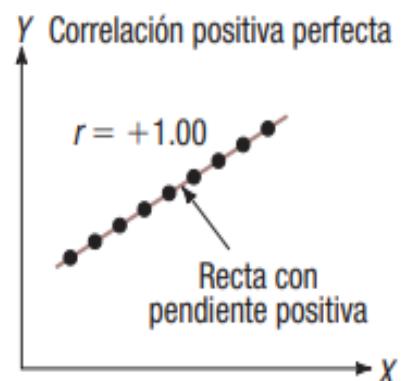
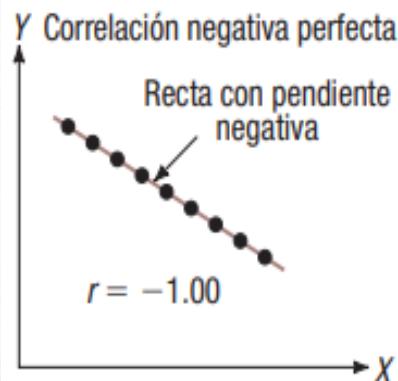
Donde:

$$-1 \leq r \leq 1$$

En el siguiente diagrama se resumen la fuerza y la dirección del coeficiente de correlación.



Diagramas de dispersión con diferentes tipos de correlaciones



Coeficiente de determinación: Mide la proporción de la varianza que queda explicada por la ecuación de regresión; o mide el porcentaje de variabilidad en Y que puede explicarse a través del conocimiento de la variable independiente.

$$r^2 = r * r \quad \text{donde: } 0 \leq r^2 \leq 1$$

Cuanto mayor sea el valor de r^2 , mejor será el ajuste y mas útil la recta de regresión como instrumento de medición. Ej:

$r^2 = 0.90$: indica que de 100 pares de puntos, 90 están en la recta de regresión y 10 fuera de la recta de regresión.

Ejercicio resuelto 1

Armand's Pizza Parlors es una cadena de restaurantes de comida italiana. Sus mejores ubicaciones son las que se encuentran cerca de los campus de las universidades. Los gerentes creen que las ventas trimestrales de estos restaurantes (que se denotan por y) están directamente relacionadas con el tamaño de la población estudiantil (que se denota x); es decir, en los restaurantes que están cerca de campus que tienen una población estudiantil grande se generan más ventas que en los restaurantes situados cerca de campus con una población estudiantil pequeña. Empleando el análisis de regresión, se puede obtener una ecuación que muestre cuál es la relación entre la variable dependiente y y la variable dependiente x .

POBLACIÓN DE ESTUDIANTES Y VENTAS TRIMESTRALES
EN 10 RESTAURANTES ARMAND'S PIZZA PARLORS

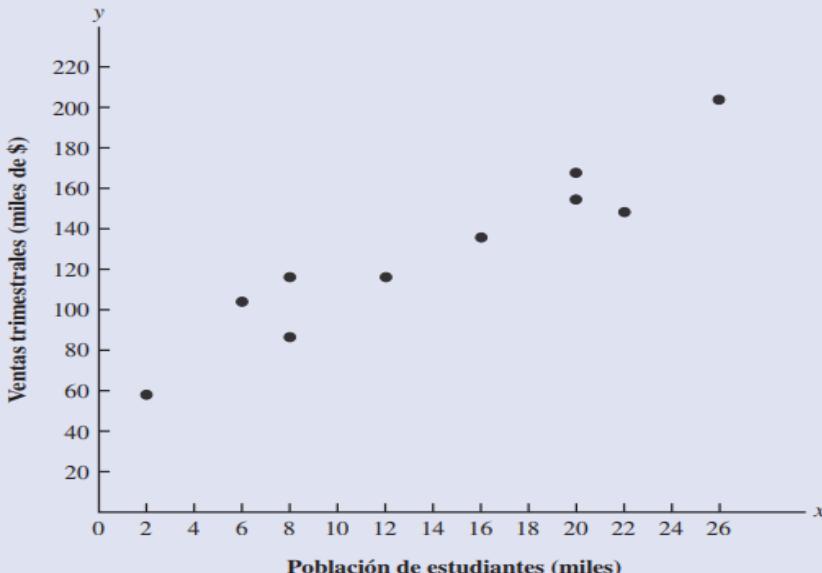
Restaurante <i>i</i>	Población de estudiantes (miles) <i>x_i</i>	Ventas trimestrales (miles de \$) <i>y_i</i>
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137
7	20	157
8	20	169
9	22	149
10	26	202

- a) Determinar la relación entre las variables usando la grafica adecuada
- b) Determinar la ecuación de la recta
- c) Estimar las ventas trimestrales de un restaurante cuando la población sea de 16000 estudiantes
- d) Determinar el grado de relación entre las variables
- e) Determinar el % de varianza de las ventas que es explicada por el tamaño de población

SOL

a)

DIAGRAMA DE DISPERSIÓN EN EL QUE SE MUESTRA LA POBLACIÓN DE ESTUDIANTES Y LAS VENTAS TRIMESTRALES DE ARMAND'S PIZZA PARLORS



$$b) Y = a + bX = b_0 + b_1 X$$

Restaurante i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	2	58	-12	-72	864	144
2	6	105	-8	-25	200	64
3	8	88	-6	-42	252	36
4	8	118	-6	-12	72	36
5	12	117	-2	-13	26	4
6	16	137	2	7	14	4
7	20	157	6	27	162	36
8	20	169	6	39	234	36
9	22	149	8	19	152	64
10	26	202	12	72	864	144
Totales	140	1300			2840	568
	Σx_i	Σy_i			$\Sigma(x_i - \bar{x})(y_i - \bar{y})$	$\Sigma(x_i - \bar{x})^2$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{140}{10} = 14$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1300}{10} = 130$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

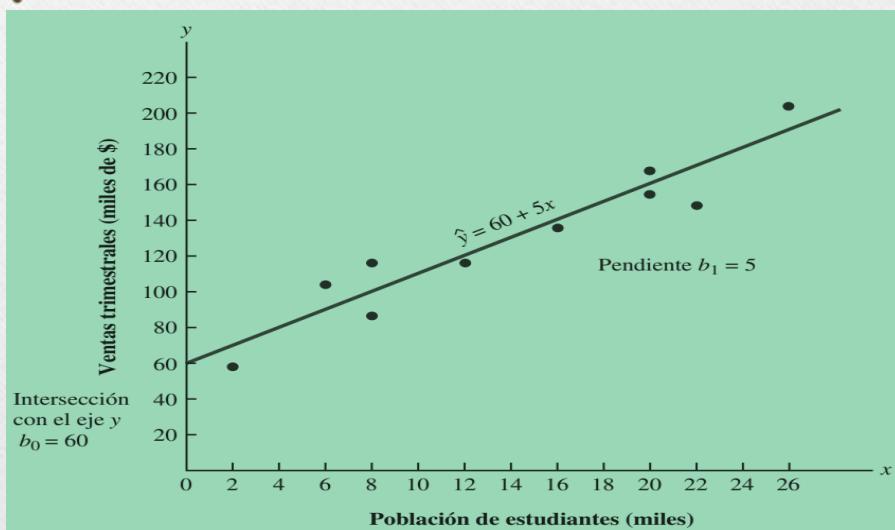
$$\begin{aligned} b_1 &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ &= \frac{2840}{568} \\ &= 5 \end{aligned}$$

La intersección con el eje y (b_0) se calcula como sigue.

$$\begin{aligned} b_0 &= \bar{y} - b_1 \bar{x} \\ &= 130 - 5(14) \\ &= 60 \end{aligned}$$

Por lo tanto, la ecuación de regresión estimada es

$$\hat{y} = 60 + 5x$$



c) Estimar las ventas trimestrales de un restaurante cuando la población sea de 16000 estudiantes

Se pide: Estimar “y”, cuando $x=16 \Rightarrow \hat{y} = 60 + 5(16) = 140$

d) Determinar el grado de relación entre las variables

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1) * S_x S_y} = \frac{2840}{9 * (7.944) * (41.806)} \Rightarrow r = 0.95016$$

Por lo tanto existe una relación alta positiva entre el tamaño de población estudiantil y las ventas trimestrales

e) Determinar el % de varianza de las ventas que es explicada por el tamaño de población

$$r^2 = 0.95016 * 0.95016 \Rightarrow r^2 = 0.90280 = 90.280\%$$

El 90.280% de la variabilidad en las ventas se explica por la relación lineal que existe entre el tamaño de la población de estudiantes y las ventas

Ejercicio resuelto 2

North American Copier Sales vende copiadoras a empresas de todos los tamaños en Estados Unidos y Canadá. Hace poco ascendieron a la señora Marcy Bancer al puesto de gerente nacional de ventas. Los representantes de todo el país asistirán a la siguiente junta de ventas, y ella desea destacar la importancia de hacer una última llamada de ventas adicional cada día; por lo tanto, decide reunir información sobre la relación entre el número de llamadas de ventas y el de copiadoras vendidas, así que selecciona una muestra aleatoria de 15 representantes y determina el número de llamadas que hicieron el mes anterior y las copiadoras que vendieron. La información muestral se reporta en la tabla 13.1. ¿Qué observaciones cabe hacer respecto de la relación entre el número de llamadas de ventas y la cantidad de copiadoras vendidas? Elabore un diagrama de dispersión para representar la información.

Número de llamadas de ventas y copiadoras vendidas

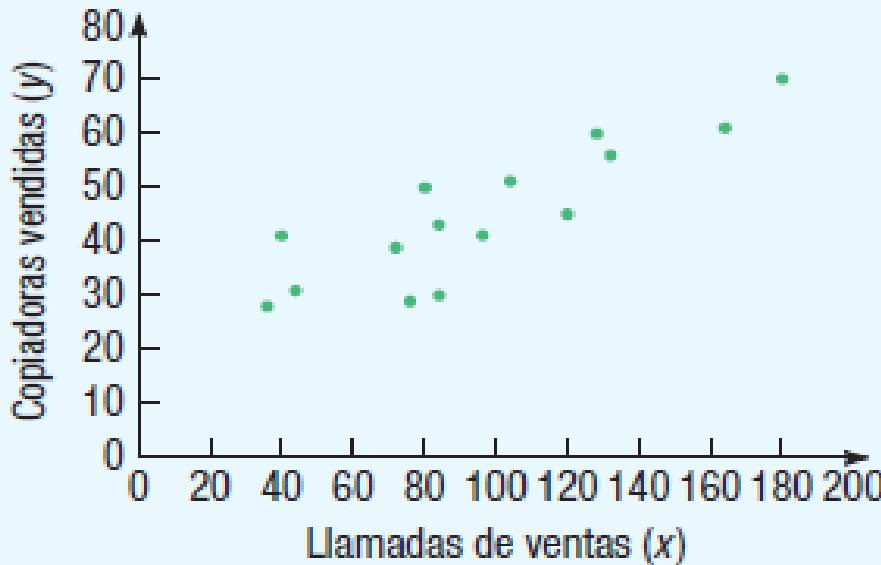
Representantes de ventas	Número de llamadas de ventas	Número de copiadoras vendidas
Brian Virost	96	41
Carlos Ramírez	40	41
Carol Saia	104	51
Greg Fish	128	60
Jeff Hall	164	61
Mark Reynolds	76	29
Meryl Rumsey	72	39
Mike Kiel	80	50
Ray Snarsky	36	28
Rich Niles	84	43
Ron Broderick	180	70
Sal Spina	132	56
Soni Jones	120	45
Susan Welch	44	31
Tom Keller	84	30

- a) Determinar la relación entre las variables usando la grafica adecuada
- b) Determinar la ecuación de la recta
- c) Estimar la cantidad de copiadoras vendidas si se realizan 100 llamadas de ventas
- d) Determinar el grado de relación entre las variables
- e) Determinar el % de varianza de la cantidad de copiadoras vendidas que es explicada por el numero de llamadas.

SOL

a)

Diagrama de dispersión que representa las llamadas de ventas y las copiadoras vendidas



$$(x_i - \bar{x})^2$$

b) $Y = a + bX = b_0 + b_1X$

	X	Y						
Representantes	Nro de llamadas	Nro de copiadoras	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	
Brian Virost	96	41	0	-4	0	0	16	
Carlos Ramírez	40	41	-56	-4	224	3136	16	
Carol Saia	104	51	8	6	48	64	36	
Greg Fish	128	60	32	15	480	1024	225	
Jeff Hall	164	61	68	16	1088	4624	256	
Mark Reynolds	76	29	-20	-16	320	400	256	
Meryl Rumsey	72	39	-24	-6	144	576	36	
Mike Kiel	80	50	-16	5	-80	256	25	
Ray Snarsky	36	28	-60	-17	1020	3600	289	
Rich Niles	84	43	-12	-2	24	144	4	
Ron Broderick	180	70	84	25	2100	7056	625	
Sal Spina	132	56	36	11	396	1296	121	
Soni Jones	120	45	24	0	0	576	0	
Susan Welch	44	31	-52	-14	728	2704	196	
Tom Keller	84	30	-12	-15	180	144	225	
Total	1440	675			6672	25600	2326	

$$\bar{x} = \frac{\sum x_i}{n} = \frac{1440}{15} = 96$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{675}{15} = 45$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = 6672 / 25600$$

$$b_1 = 0.260625$$

$$b_0 = 45 - (0.260625) * 96$$

$$b_0 = 19.98$$

Por lo tanto

$$\hat{y} = b_0 + b_1 x \Rightarrow \hat{y} = 19.98 + 0.260625 X$$

c) Estimar la cantidad de copiadoras vendidas si se realizan 100 llamadas de ventas

Se pide: Estimar “y”, cuando $x=100 \Rightarrow \hat{y} = 19.98 + 0.260625(100) = 46.0425$

d) Determinar el grado de relación entre las variables

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1) * S_x S_y} = \frac{6672}{14 * (42.762) * (12.89)} \Rightarrow r = 0.8646$$

Por lo tanto existe una relación alta positiva entre el numero de llamadas de ventas y la cantidad de copiadoras vendidas

e) Determinar el % de varianza de la cantidad de copiadoras vendidas que es explicada por el numero de llamadas

$$r^2 = 0.8646 * 0.8646 \Rightarrow r^2 = 0.7476 = 74.76\%$$

El 74.76% de la variabilidad en las ventas de copiadoras se explica por la relación lineal que existe entre el numero de llamadas y las ventas de las copiadoras

EJERCICIOS A DESARROLLAR

Ejercicio 1 :

Se realizó una investigación por muestreo a 10 empresas que producen un determinado producto, en cuanto al valor de la producción(miles de soles) y el costo del mismo(miles de soles). Los resultados fueron:

Nº	Producción=	Costo=
1	10	3
2	18	5
3	12	4
4	16	5
5	22	8
6	36	12
7	30	10
8	32	14
9	26	12
10	12	3

- a) Determinar gráficamente la relación entre las dos variables
- b) Determinar la ecuación de la recta
- c) Estimar los costos para una empresa que tiene en producción en 32(miles soles)
- d) Determinar el grado de relación entre las variables
- e) ¿Qué % de variabilidad de la variable Y es explicada por la variable X

Ejercicio 2:

Nº	Nº sucursales	Costos
1	3	2
2	5	3
3	3	5
4	2	4
5	4	6
6	1	2
7	5	5
8	2	1
9	6	3
10	3	5
11	5	3
12	1	2

Al observar el número de sucursales y los costos mensuales en comunicación telefónica con la casa central, en miles de soles, para 12 empresas se encontró:

- a) Determinar gráficamente la relación entre las dos variables
- b) Determinar la ecuación de la recta
- c) Estimar los costos mensuales para una empresa que tiene 7 sucursales
- d) Determinar el grado de relación entre las variables
- e) ¿Qué % de variabilidad de la variable Y es explicada por la variable X

Ejercicio 3:

Se ha observado, que para predecir la demanda (consumo) de combustible para calefacción, resulta ser más preciso el pronóstico a largo plazo de las temperaturas y el uso, de la relación temperatura - consumo, que el tratar de pronosticar directamente la demanda analizando las ventas de combustibles. Un distribuidor mantiene un registro de ventas mensuales de combustibles y de temperaturas máximas en esos meses. A continuación aparecen los datos de 8 de estos meses seleccionados al azar.

VOLUMEN VENTAS:	16,2	17,4	12,3	35,9	42,1	26,4	19,0	10,1
PROMEDIODE TEMPERATURAS MÁXIMAS (grados centígrados):	18,3	21,2	17,6	23,4	25,2	20,4	21,6	18,6

Se pide:

- Grafique los puntos (observaciones reales) y las rectas de \hat{Y} y \hat{X}
- ¿Qué porcentaje de la variación en las ventas mensuales se explica por la temperatura máxima promedio?

EJERCICIO 4:

El propietario de Maumee Ford-Mercury-Volvo desea estudiar la relación entre la antigüedad de un automóvil y su precio de venta. En la siguiente lista se observa una muestra aleatoria de 12 automóviles usados que vendió el concesionario durante el año anterior.

Automóvil	Antigüedad (años)	Precio de venta (miles de dólares)	Automóvil	Antigüedad (años)	Precio de venta (miles de dólares)
1	9	8.1	7	8	7.6
2	7	6.0	8	11	8.0
3	11	3.6	9	10	8.0
4	12	4.0	10	12	6.0
5	8	5.0	11	6	8.6
6	7	10.0	12	6	8.0

- Trace un diagrama de dispersión.
- Establezca el coeficiente de correlación.
- Interprete el coeficiente de correlación. ¿Le sorprende que sea negativo?

EJERCICIO 5:

Wageweb realiza estudios sobre datos salariales y presenta resúmenes de éstos en su sitio de la Red. Basándose en datos salariales desde el 1 de octubre de 2002 Wageweb publicó que el salario anual promedio de los vicepresidentes de ventas era \$142 111 con una gratificación anual promedio de \$15 432 (Wageweb.com, 13 de mayo de 2003). Suponga que los datos siguientes sean una muestra de salarios y bonos anuales de 10 vicepresidentes de ventas. Los datos se dan en miles de dólares.

Vicepresidente	Salario	Gratificación
1	135	12
2	115	14
3	146	16
4	167	19
5	165	22
6	176	24
7	98	7
8	136	17
9	163	18
10	119	11

- a. Trace un diagrama de dispersión con estos datos tomando como variable independiente los salarios.
- b. ¿Qué indica el diagrama de dispersión del inciso a) acerca de la relación entre salario y gratificación?
- c. Use el método de mínimos cuadrados para obtener la ecuación de regresión estimada.
- d. Dé una interpretación de la ecuación de regresión estimada.
- e. ¿Cuál será la gratificación de un vicepresidente que tenga un salario anual de \$120 000?

REGRESION LINEAL SIMPLE

Ejercicios 6: Se cuenta con las mediciones sobre la edad y la talla de 14 niños, y estamos interesados en determinar si existe algún tipo de relación entre la talla del niño y su edad.

niño	edad (meses)	talla (cm)
i	x_i	y_i
1	3	55
2	6	68
3	5	64
4	5	66
5	3	62
6	4	65
7	9	74
8	8	75
9	9	73
10	7	69
11	6	73
12	5	68
13	8	73
14	6	71

$$y = \text{talla} / x = \text{edad} / n = 14$$

$$\sum_{i=1}^{14} y_i = 956 \quad \bar{y} = 68,3 \quad s_y = 5,6$$

$$\sum_{i=1}^{14} x_i = 84 \quad \bar{x} = 6 \quad s_x = 2$$

$$\text{cov}(x, y) = 9,07 \quad r_{xy} = 0,88$$

$$\sum_{i=1}^{14} x_i y_i = 5863 \quad \sum_{i=1}^{14} x_i^2 = 556$$

Ejercicios 7: : Se ha efectuado un estudio donde se relacionan los puntajes de aptitud con la productividad en una industria. Después de 3 meses de entrenamiento del personal sus postulantes elegidos al azar obtuvieron los diez pares de puntajes y productividad que se indican a continuación. Determinar:

- a) Graficar la relación entre las dos variables.
- b) La ecuación de la recta
- c) ¿Cuál es la productividad esperada de un trabajador, cuyo puntaje de aptitud fue de 18.
- d) Calcular el puntaje de aptitud de un postulante cuando la productividad es de 36
- e) Calcula e interpreta el coeficiente de correlación.
- f) Que % de variabilidad de la productividad es explicada por los puntajes de aptitud

Puntaje de aptitud	Productividad
9	23
17	35
20	29
19	33
20	43
23	32
15	30
25	40
28	45
30	38

Ejercicios 8: los ingresos y los gastos mensuales en soles de una muestra de 100 familias, han dado las siguientes resultados:

$$\bar{x} = 210 , \bar{y} = 200 , S_x^2 = 5.76 , S_y^2 = 2.56 , \sum xy = 4200364$$

- a) Determinar la regresión de "y" en "x"
- b) Estime el gasto de una familia que tiene 250 soles de ingreso.

Ejercicios 9: El coeficiente de correlación entre dos variables X e Y es $r = 0.60$. Si $S_x = 1.50$, $S_y = 2$, $\bar{x} = 10$, $\bar{y} = 20$, hallar la recta de regresión

Ejercicios 10: En una tabla de frecuencias correspondientes a 20 parejas, se obtuvo la siguiente información:

$$\Sigma X = 174 ; \Sigma X^2 = 1626 ; \Sigma Y = 340 ; \Sigma Y^2 = 5904 ; \Sigma XY = 3062$$

- a) Hallar la recta de regresión lineal de Y en X
- b) Calcular e interpretar el coeficiente de Pearson
- c) Qué % de la variabilidad de Y es explicada por X?

Ejercicios 11: Al estudiar la relación entre la edad y el índice de masa corporal a partir de una muestra de una muestra de hombres se obtuvo:

$$S_X = 8.5, S_Y = 10, \bar{X} = 50, \bar{Y} = 120, r^2 = 0.90$$

- a) Determine la recta de regresión
- b) Estime el índice de masa corporal para un hombre de 40 años
- c) Calcular la edad cuando su índice de masa corporal se estima en 25.8
- d) Calcular e interpretar el coeficiente de correlación

Ejercicios 12: Al estudiar la regresión lineal entre los ingresos mensuales y el número de hijos por familia, se obtuvo la siguiente la información:

- $\bar{x} = 3 ; \bar{y} = 900 ; S_X = 0.05 x \sqrt{S_{XY}} , S_Y = 0.5$
- a) Determinar la ecuación de la recta de la regresión
 - b) Estimar los ingresos de las familias con 4 hijos
 - c) ¿A cuántos hijos por familia correspondería un ingreso estimado de 750 soles

Ejercicios13:

De una determinada empresa se conocen los sgtes datos, referidos al volumen de ventas(miles soles) y el gasto en publicidad (miles soles) de los últimos seis años:

Volumen ventas(miles soles)	Gastos publicidad(miles soles)
10	16
15	32
20	48
22	56
30	64
32	80

- a) Verificar que tipo de relación existe usando la gráfica apropiada
- b) Calcular la recta de regresión de Y en X
- c) ¿Qué volumen de ventas de la empresa se podría esperar en un año que se gaste de publicidad 60(miles soles)
- d) Determinar el grado de relación entre las variables
- d) Que % de variabilidad del volumen de ventas es explicada por los gastos de publicidad.

Ej 14 Se está estudiando la relación entre el número de años que una persona está afiliada al sindicato y el nivel de satisfacción con la actuación de dicho sindicato. Para ello se parte de los datos de 7 individuos tomados aleatoriamente de personas adscritas a partidos políticos, obteniéndose:

- a) Verificar que tipo de relación existe usando la gráfica apropiada
- b) Calcular la recta de regresión de Y en X
- c) Predecir el índice de satisfacción de una persona que lleva 11 años militando al sindicato
- d) Que % de variabilidad del nivel de satisfacción en el sindicato es explicada por los años que una persona está afiliada.

Satisfacción	Años
7	8
5	7
8	10
5	3
9	6
9	13
3	4

Ej 15

Un centro comercial sabe en función de la distancia, en kilómetros, a la que se sitúe de un núcleo de población, acuden los clientes, que figuran en la tabla:

Nº de Clientes (X)	Distancia (Y)
8	15
7	19
6	25
4	23
2	34
1	40

- 1 Calcular el coeficiente de correlación lineal.
- 2 Si el centro comercial se sitúa a 2 km, ¿cuántos clientes puede esperar?
- 3 Si desea recibir a 5 clientes, ¿a qué distancia del núcleo de población debe situarse?

Ej 16 Un gerente de ventas recolecto los datos siguientes sobre ventas anuales y años de experiencia

Vendedor	Años de experiencia	Ventas anuales (miles de \$)
1	1	80
2	3	97
3	4	92
4	4	102
5	6	103
6	8	111
7	10	119
8	10	123
9	11	117
10	13	136

- Elabore un diagrama de dispersión con estos datos, en el que la variable independiente sean los años de experiencia.
- Dé la ecuación de regresión estimada que puede emplearse para predecir las ventas anuales cuando se conocen los años de experiencia.
- Use la ecuación de regresión estimada para pronosticar las ventas anuales de un vendedor de 9 años de experiencia.

Gracias

