UNIVERSIDAD NACIONAL

JOSÉ FAUSTINO SÁNCHEZ CARRIÓN

FACULTAD DE CIENCIAS E INGENIERÍA

INSTITUTO DE INVESTIGACIÓN

TEXTO UNIVERSITARIO

ANALISIS DE REGRESIÓN Y CORRELACIÓN LINEAL Y NO LINEAL

AUTORES

Mstro. CASTAÑEDA CARRIÓN Yolanda Marianela

Mstro. TREJO LÓPEZ Mirtha Sussan

Mo. VALVERDE FLORES Cosme Ulises

HUACHO

2011

OBJETIVOS

1. Elaborar un texto de Análisis de Regresión y Correlación Lineal y No Lineal a nivel universitario que sirva como material bibliográfico en el transcurso de la formación profesional del estudiante de la Universidad Nacional José Faustino Sánchez Carrión.

2. Contribuir a la buena formación básica profesional de los estudiantes de las diferentes Escuelas Profesionales de la Universidad.

JUSTIFICACIÓN

Actualmente en la Universidad la asignatura de Estadística se imparte en las diferentes carreras profesionales, debido a ello es necesario contar con un texto de fácil acceso para la preparación básica de los estudiantes que sirva de guía y consulta permitiendo comprender, identificar, analizar e interpretar la relación y la intensidad de las variables.

RESULTADOS

1. Satisfacer la necesidad de un material bibliográfico al alcance de los estudiantes de las diferentes carreras de la Universidad.
2. Proporcionar al estudiante un aprendizaje interactivo y constructivo generando oportunidades para una discusión creativa respetando los conocimientos previos.
3. Determinar la recta de regresión utilizando el método de los mínimos cuadrados.
4. Proporcionar los procedimientos y técnicas para expresar y medir la relación o afinidad entre las variables para n observaciones.

TEXTO UNIVERSITARIO

ANÁLISIS DE REGRESIÓN Y CORRELACIÓN LINEAL Y NO LINEAL

ÍNDICE GENERAL

AGRADECIMIENTO

Reiteramos nuestra gratitud y reconocimiento a nuestros estudiantes por su constante estímulo y paciencia para que logremos escribir un texto universitario que es el desafío más noble de la cátedra universitaria y difundir ideas es hacer la vida eterna por que los hombres pasan y sus obras quedan. Que mejor satisfacción en nuestra vida dejarles este libro como testimonio de nuestro compromiso y aporte al desarrollo técnico y científico.

PROLOGO

Este texto de Análisis de Regresión y Correlación Lineal y No Lineal tiene la finalidad de proporcionar temas que se imparten en el curso de Estadística Inferencial y está orientado a las aplicaciones de nuestro contexto social.

Estamos comprometidos a ayudar a los estudiantes para que se acerquen sin angustia a la Estadística. Esta orientación de la enseñanza – aprendizaje ha dado como resultado una gran cantidad de auxiliares efectivos para el aprendizaje. En cada capítulo se presentan problemas planteados para dar a los estudiantes la oportunidad de trabajar con problemas semejantes a los ejemplos desarrollados y que sirvan para reforzar la comprensión del material elaborado.

Después del análisis de cada concepto hay al menos un ejemplo y su solución. Al final de cada capítulo se incluye un breve resumen.

Al principio de cada capítulo se plantea un conjunto de objetivos, en ellos se indica lo que el estudiante será capaz de hacer al concluir el capítulo.

Los Autores

CAPÍTULO I

REGRESIÓN LINEAL

OBJETIVOS:

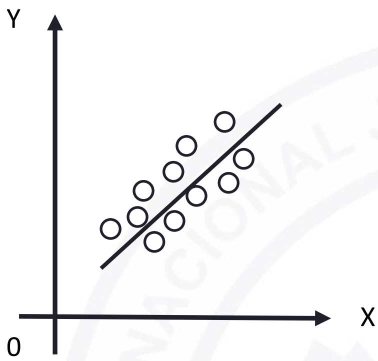
1. Trazar un diagrama de dispersión.
2. Analizar el objetivo de la regresión lineal
3. Determinar una ecuación que pueda utilizarse en pronósticos.
4. Calcular e interpretar el error estándar de estimación.
5. Determinar e interpretar los intervalos de predicción.
6. Identificar las fórmulas para los intervalos de predicción.

1.1. Diagrama de Esparcimiento

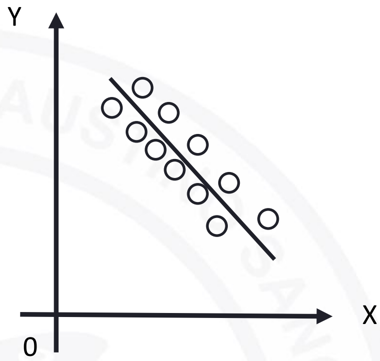
Es una ilustración gráfica que se utiliza en el análisis de regresión está constituido por una dispersión de puntos tal que cada punto representa un valor de la variable independiente (medido a lo largo del eje horizontal) y un valor asociado de la variable dependiente (medido a lo largo del eje vertical). Se le conoce con el nombre de nube de puntos o diagrama de dispersión.

La construcción del diagrama de esparcimiento constituye el primer paso para investigar la relación existente entre dos variables, la posición y forma de esta nube proporciona la idea del tipo de relación existente entre dos variables, de este modo se facilita la elección de la correspondiente función matemática.

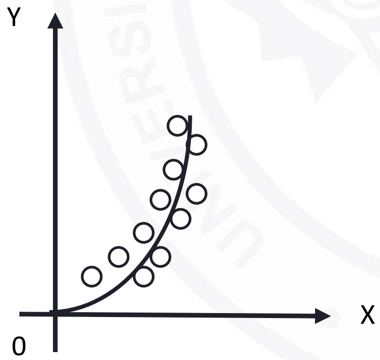
GRÁFICO Nº 1.1
DIAGRAMAS DE ESPARCIMIENTO
PARA DATOS BIDIMENSIONALES



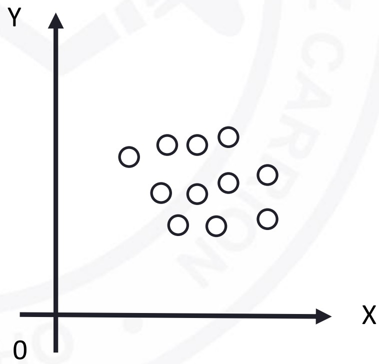
A: RELACIÓN LINEAL POSITIVA



B: RELACIÓN LINEAL NEGATIVA



C: RELACIÓN NO LINEAL



D: SIN RELACION DEFINIDA

Ahora bien graficada y visualizada la forma del diagrama de esparcimiento, interesa analizar y expresar el tipo de relación entre las variables. Para expresar esta relación se elige una función matemática que mejor represente o se ajuste al diagrama de esparcimiento.

1.2. Regresión

Es el método estadístico que investiga y define la relación funcional entre dos o más variables.

1.3. Función o Modelo de Regresión

Ecuación matemática que representa el modelo estadístico correspondiente.

1.4. Regresión Lineal

El propósito de la regresión lineal es estimar la relación que existe entre dos variables X y Y, se expresa: Y = f(x) "Y depende de X"

1.4.1. Variable Independiente.

Es la variable cuyo valor se supone conocido y se utiliza para predecir o explicar el valor de otra variable de interés, se simboliza por X. se le conoce con los nombres: VARIABLE EXPLICATORIA, VARIABLE PREDICTORIA. VARIABLE REGRESORA, VARIABLE CONTROL O ESTIMULO.

1.4.2. Variable Dependiente.

Es la variable cuyo valor se supone desconocido y que se explica o predice con ayuda de otra variable, se simboliza por Y, recibe los nombres: VARIABLE EXPLICATIVA, VARIABLE PREDECIDA, VARIABLE REGRESANDO O VARIABLE RESPUESTA.

1.4.3. Ajuste de una Función de Regresión

Ajustar una función de regresión significa buscar o definir la función que exprese con mayor precisión la relación entre variables. Gráficamente será aquella función que mejor se adecue a la nube de puntos.

El problema de ajuste de una función de regresión a un conjunto de n valores (X,Y), comprende tres pasos:

1. Graficar el diagrama de esparcimiento o nube de puntos.
2. Definir la forma de la función de regresión (recta, parábola, exponencial, etc.)
3. Determinar el valor numérico de los parámetros de la función elegida.

1.4.4. Método de Mínimos Cuadrados

Técnica empleada para llegar a la ecuación de regresión minimizando la suma de los cuadrados de las distancias verticales entre los valores verdaderos y los valores pronosticados Y.

1.5. Regresión Simple Lineal

1.5.1. Ecuación de Regresión

Expresión matemática que define la relación entre dos variables. Se le denomina también ecuación de estimación o ecuación de pronóstico

1.5.2. Línea Recta de Regresión

Consideremos la ecuación de la recta:

$$Y^* = a + bX$$

Donde:

Y^* : Valor pronosticado de la variable Y para un valor seleccionado de X

a : Ordenada de la intersección con el eje Y (o intersección Y). Es el valor estimado de Y cuando X = 0. Otra forma de decir esto es a: es el valor estimado de Y, en donde la recta de regresión cruza el eje Y cuando X es cero. Llamado también ordenada de origen.

b : Pendiente de la recta o sea cambio promedio en Y* por unidad de cambio (incremento o decremento) en la variable independiente X, se le conoce con el nombre de coeficiente angular.

X : Cualquier valor seleccionado para la variable independiente.

Para el cálculo de a y b usaremos el Método de Mínimos Cuadrados mediante las ecuaciones normales

$$a = \overline{Y} - b\overline{X}$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2},$$

Donde:

X	es un valor de la variable independiente
Y	es un valor de la variable dependiente
n	es el número de elementos de la muestra
$\overline{X}$	es la media de la variable independiente
$\overline{Y}$	es la media de la variable dependiente

1.5.3. Interpretación de "b"

Consideremos el signo de este coeficiente se tiene:

i) Si b es positivo (b>O), entonces existe una relación lineal positiva o directa es decir que ante incrementos en la variable independiente, corresponde incrementos en la variable dependiente.

ii) Si b es negativo (b<O), se tiene una relación lineal negativa o inversa es decir, que incrementos en la variable independiente origina decrementos o disminuciones en la variable dependiente.

1.5.4. Graficación de la Recta de Regresión

Dada la ecuación de regresión:

$$Y^{*} = a + bX$$

Para graficar una recta es suficiente definir dos puntos luego se grafican estos dos puntos en el plano y por ellos se traza una recta.

1.5.5. Error Estándar de Estimación

Llamado también desviación estándar de regresión. Mide la dispersión de los valores observados, con respecto a la recta de regresión. En general entre el valor Y, y el estimado Y* existe una diferencia o sesgo, que puede ser menor o mayor en la medida que los “n" puntos del diagrama de esparcimiento estén más o menos cerca de la línea de regresión.

$$S_e = S_{YX} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

Donde:

Y : valor real de Y

Y* : valor pronosticado de Y

n : número de elementos de la muestra

X : valores de la variable independiente

Y : valores de la variable dependiente

a : ordenada de origen

b : pendiente de la recta o coeficiente angular

1.5.6. Interpretación del Error Estándar de Estimación

Como se aplicaba en la desviación estándar, mientras más grande sea el error estándar de la estimación, mayor será la dispersión de los puntos alrededor de la línea de regresión. De manera inversa, si $S_e = 0$, esperamos que la ecuación de estimación sea un estimador perfecto de la variable dependiente. En este caso, todos los puntos de datos caerían directamente sobre la línea de regresión no habrá puntos dispersos alrededor.

1.5.7. Intervalos de Predicción

Los estadísticos aplican los intervalos de predicción basados en la distribución normal sólo a grandes muestras (n≥30)

Suponiendo que los puntos observados estén normalmente distribuidos alrededor de la línea de regresión, se tendrá los intervalos siguientes:

$Y^* \pm S_e$ se encuentra el 68,3% de los valores reales de Y

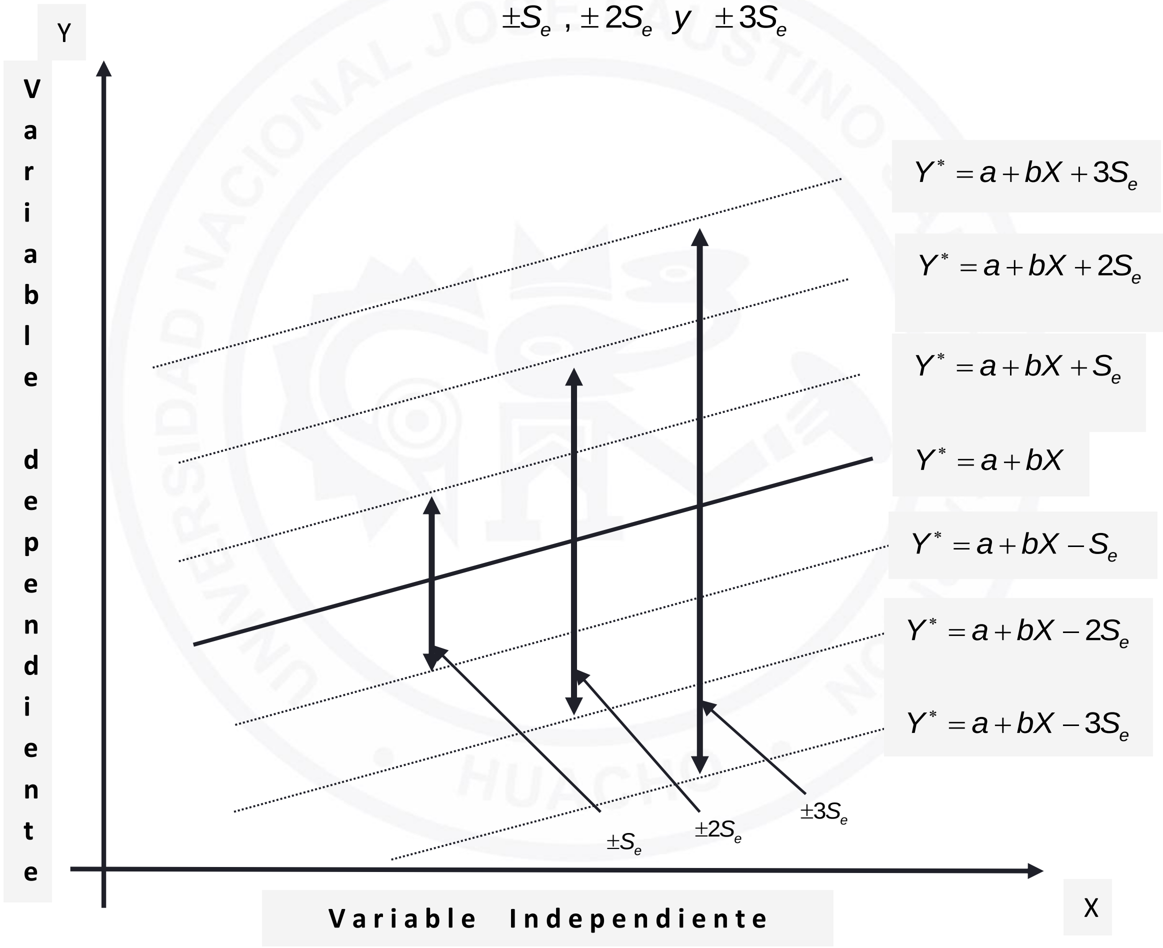
$Y^* \pm 2S_e$ se encuentra el 95,5% de los valores reales de Y

$Y^* \pm 3S_e$ se encuentra el 99,7% de los valores reales de Y

GRÁFICO Nº1.2

LIMITES ALREDEDOR DE LA LÍNEA DE REGRESIÓN

$\pm S_e$, $\pm 2S_e$ y $\pm 3S_e$



Cuando las muestras son pequeñas (n<30) utilizamos la distribución t-student, los intervalos serán:

$$Y^* \pm t_{(1-\alpha, n-2)} S_e$$

EJERCICIOS RESUELTOS

1.- El administrador de EMAPA HUACHO S.A. desea establecer la relación entre el consumo mensual domiciliario y el tamaño de la familia en la urbanización Los Jardines en octubre de 2011, se conocen los datos siguientes:

Tamaño de familia	Galones de agua utilizados
9	1050
2	650
7	850
9	1000
4	550
11	1050
6	900
3	650
3	600
2	600
2	700
3	500
5	550
10	1000
6	800
7	900
6	700
4	500
3	550
5	800
8	950
8	1100
6	750
7	800
5	750
4	600
7	950
11	1000
12	1050
9	1100
12	1000
12	1100
10	1050
9	800

a) Construir el diagrama de esparcimiento.
b) Determine la ecuación de regresión. Interprételo.
c) Graficar la línea de regresión
d) Calcular el error estándar de estimación. Interprétalo
e) Establecer los intervalos de predicción estimando el consumo de agua de las familias que tienen un tamaño de 4 personas con un nivel de confianza del 95.5%. Interprétalo.

SOLUCIÓN:

a) Definamos primeramente la variable dependiente como el consumo de galones de agua utilizando (Y) y la variable independiente como el tamaño de la familia (X);

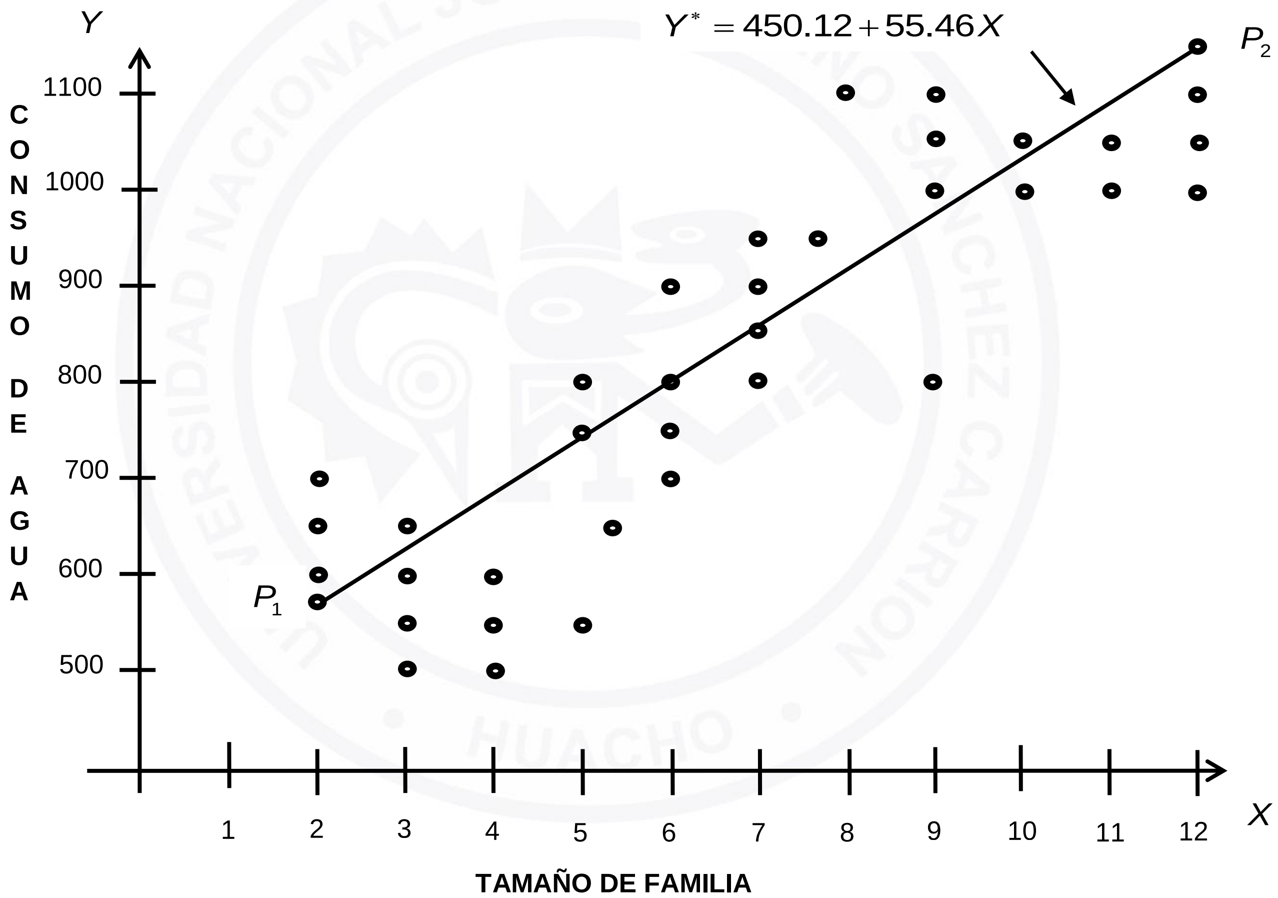
Luego:

Consumo de galones de agua =f (tamaño de la familia)

Y = f (X)

GRÁFICO Nº1.3

DIAGRAMA DE ESPARCIMIENTO Y RECTA ESTIMADA DE REGRESIÓN DEL CONSUMO DE AGUA Y EL TAMAÑO DE FAMILIA. URBANIZACIÓN LOS JARDINES. OCTUBRE 2011. HUACHO



b)

CUADRO Nº1.1

ECUACIÓN DE REGRESIÓN Y ERROR ESTÁNDAR DE ESTIMULACIÓN DEL TAMAÑO DE FAMILIA Y CONSUMO DE AGUA EN LA URBANIZACIÓN LOS JARDINES. OCTUBRE 2011. HUACHO.

X	Y	XY	X^2	Y^2
9	1050	9450	81	1102500
2	650	1300	4	422500
7	850	5950	49	722500
9	1000	9000	81	1000000
4	550	2200	16	302500
11	1050	11550	121	1102500
6	900	5400	36	810000
3	650	1950	9	422500
3	600	1800	9	360000
2	600	1200	4	360000
2	700	1400	4	490000
3	500	1500	9	250000
5	550	2750	25	302500
10	1000	10000	100	1000000
6	800	4800	36	640000
7	900	6300	49	810000
6	700	4200	36	490000
4	500	2000	16	250000
3	550	1650	9	302500
5	800	4000	25	640000
8	950	7600	64	902500
8	1100	8800	64	121000
6	750	4500	36	562500
7	800	5600	49	640000
5	750	3750	25	562500
4	600	2400	16	360000
7	950	6650	49	902500
11	1000	11000	121	1000000
12	1050	12600	144	1102500
9	1100	9900	81	1210000
12	1000	12000	144	1000000
12	1100	13200	144	1210000
10	1050	10500	100	1102500
9	800	7200	81	640000
227	27900	204100	1837	24185000
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$

La recta de regresión o ecuación de regresión:

$$Y^* = a + bX$$

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - \left(\sum X\right)^2}$$

Reemplazando los valores del Cuadro Nº 1.1 en b se tiene:

$$b = \frac{34(204100) - (227)(27900)}{34(1837) - \left(227\right)^2}$$

$$b = 55.46$$

Encontramos el valor de a:

$$a = \frac{27900}{34} - 55.46\left(\frac{227}{34}\right)$$

$$a = 450.12$$

La ecuación de regresión:

$$Y^* = 450.12 + 55.46X$$

Interpretación:

$$Y^* = 450.12 + 55.46X$$

Iniciando con un consumo base de 450, 12 galones de agua cada incremento unitario en el tamaño de la familia está asociado con un aumento promedio de 55,46 galones de agua.

a = 450,12

Nos indica que es el consumo de agua cuando el tamaño de familia es cero.

b = 55,46

Nos indica en general que el consumo de agua (Y) aumenta a medida que se incremente el tamaño de familia (X) esto es que existe una relación lineal directa entre el tamaño de familia (X) y el consumo de agua (Y) de la urbanización los Jardines. Huacho.

c) Definimos dos puntos para graficar la recta de regresión y luego se grafican en el mismo plano del diagrama de esparcimiento.

$$X = 2 \qquad \Rightarrow \qquad Y_2^* = 450.12 + 55.46(2) = 561.04$$

$$\therefore P_1(2\,;561.04)$$

$$X = 12 \qquad \Rightarrow \qquad Y_{12}^* = 450.12 + 55.46(12) = 1115.64$$

$$\therefore P_2(12\,;1115.64)$$

d) Error estándar de estimación

$$S_e = S_{YX} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}}$$

$$S_e = S_{YX} = \sqrt{\frac{24185000 - 450.12(27900) - 55.46(204100)}{32}}$$

$$S_e = S_{YX} = 97.9901143$$

$$S_e = S_{YX} \cong 98$$

Interpretación:

Como el error estándar de estimación es pequeño, indica que las dos variables están relacionadas muy cerca esto nos indica que el coeficiente de correlación tiende a ser grande ya que los puntos de diagrama de dispersión parecen cercanos a la línea de regresión.

e) $n = 34, \quad X = 4.1 \quad 1-\alpha = 95.5\% \quad S_e = 98$

Como n es grande $n = 34 \geq 30$ el intervalo de predicción:

$$Y^* \pm 2S_e = 95.5\%$$

El intervalo de predicción propuesto es para X = 4 encontramos Y^*

$$Y^* = 450.12 + 55.46(4) = 450.12 + 221.84 = 671.96$$

$$L_1 = 671.96 - 2(98)$$

$$L_1 = 476$$

$$L_2 = 671.96 + 2(98)$$

$$L_2 = 868$$

El intervalo de predicción será:

$$P\ [476 \leq Y^* \leq 868] = 95.5\%$$

Interpretación:

El 95,5% de las casas habitacionales en la urbanización Los Jardines que tienen un tamaño de familia de 4 personas tendrían un consumo mensual entre 476 y 868 galones de agua.

2.- Se desea analizar la relación entre edad y el tiempo efectivo de servicios, se considera una muestra de 15 trabajadores de la Empresa Textil ¨La Alameda¨ S.A. Trujillo 2011, obteniéndose los resultados siguientes:

Edad	48	40	30	39	46	42	27	36	34	46	32	42	40	32	27
Tiempo de servicio	24	18	9	14	22	22	4	13	10	20	12	18	16	8	6

a) Construir el diagrama de esparcimiento

b) Determinar la ecuación de regresión.

c) Graficar la línea de regresión

d) Calcular el error estándar de estimación. Interprételo

e) Establecer los intervalos que tienen 35 años de edad con un nivel de confianza del 99%

SOLUCIÓN:

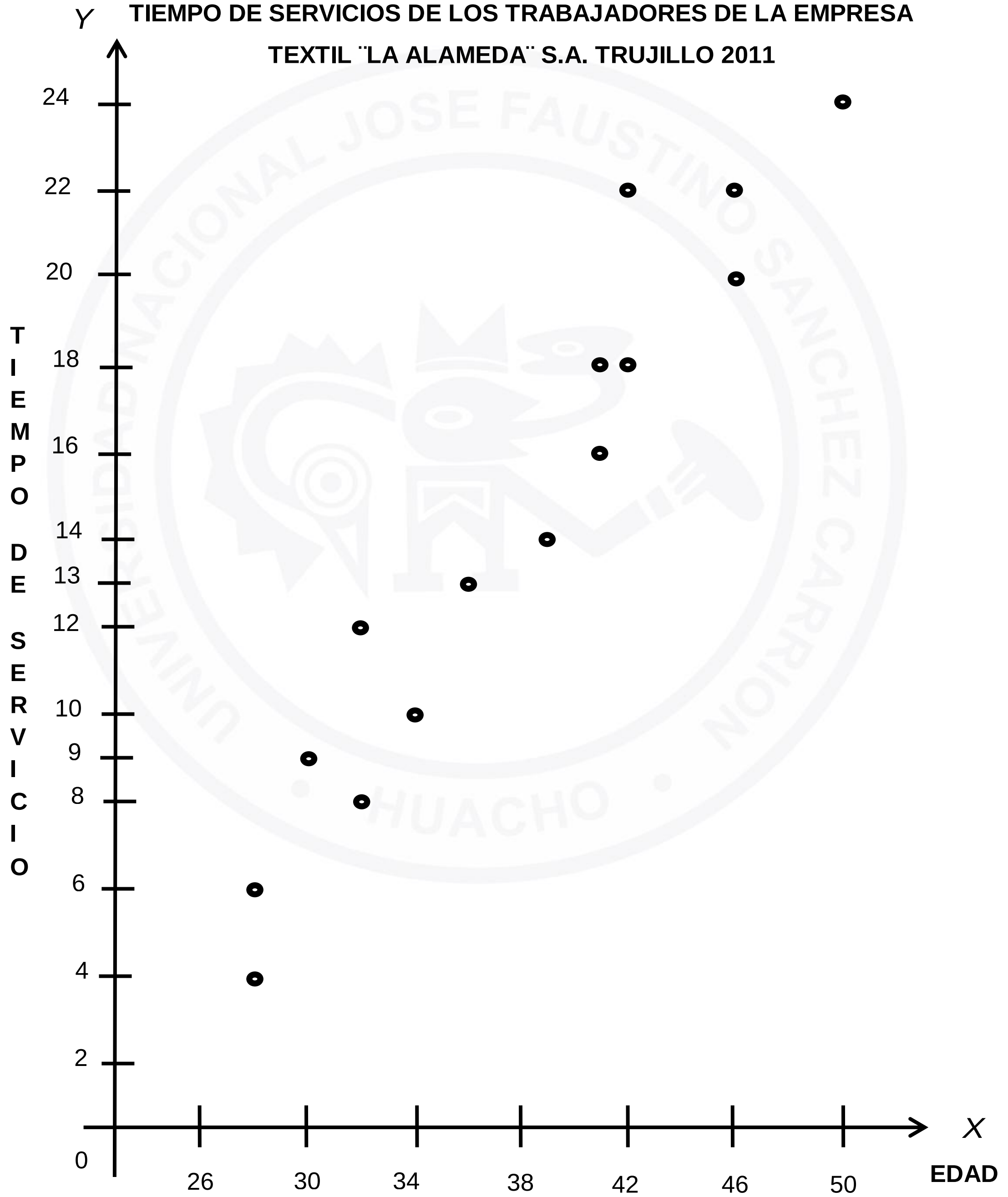
a) Primeramente definimos cual será la variable dependiente (Y) y cual la variable independiente (X) Luego:

Tiempo de servicio =f (edad)

Y = f (x)

GRÁFICO Nº1.4

DIAGRAMA DE ESPARCIMIENTO – RECTA ESTIMADA DE REGRESIÓN DE LA EDAD Y TIEMPO DE SERVICIOS DE LOS TRABAJADORES DE LA EMPRESA TEXTIL ¨LA ALAMEDA¨ S.A. TRUJILLO 2011



b)

CUADRO Nº 1.2

ECUACIÓN DE REGRESIÓN Y ERROR ESTÁNDAR DE ESTIMACIÓN DE LA EDAD Y TIEMPO DE SERVICIO DE LOS TRABAJADORES DE LA EMPRESA TEXTIL ¨LA ALAMEDA¨ S.A. TRUJILLO 2011

X	Y	XY	X^2	Y^2
48	24	1152	2304	576
40	18	720	1600	324
30	9	270	900	81
39	14	546	1521	196
46	22	1012	2116	484
42	22	924	1764	484
27	4	108	729	16
36	13	468	1296	169
34	10	340	1156	100
46	20	920	2116	400
32	12	384	1024	144
42	18	756	1764	324
40	16	640	1600	256
32	8	256	1024	64
27	6	162	729	36
561	216	8658	21643	3654
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$

Previamente reemplazamos los valores del cuadro Nº1.2 en b tenemos:

$$b = \frac{15(8658) - (561)(216)}{15(21643) - (561)^2}$$

$$b = 0.876$$

Encontramos el valor de a:

$$a = \frac{216}{15} - 0.876\left(\frac{561}{15}\right)$$

$$a = -18.362$$

La ecuación de regresión.

$$Y^* = -18.362 + 0.876X$$

c) Para graficar la recta de regresión:

$X = 30 \qquad \Rightarrow \qquad Y^* = -18.362 + 0.876(30) = 7.9$

$\therefore P_1(30\ ;\ 7.9)$

$X = 48 \qquad \Rightarrow \qquad Y^* = -18.362 + 0.876(48) = 23.7$

$\therefore P_2(48\ ;\ 23.7)$

d) Error Estándar de Estimación

$$S_e = S_{YX} = \sqrt{\frac{3654-(18.362)(216)-(0.876)(8658)}{13}}$$

$$S_e = S_{YX} = 1.66$$

Interpretación:

Como el error estándar de estimación (S_e) es pequeño, indica que las dos variables están relacionadas muy cerca esto nos indica que el coeficiente de correlación tiende a se grande ya que los puntos sobre el diagrama de dispersión parecen cercanos a la línea de regresión.

e) $n = 15, \quad X = 35 \quad 1-\alpha = 0.99 \quad S_e = 1.66$

Como n es pequeño $n = 15 < 30$ el intervalo de predicción será:

$$Y^* \pm t_{(1-\alpha, n-2)} S_e$$

Encontraremos el valor de Y^* cuando X = 35

$Y^* = -18.362 + 0.876(35) = -18.362 + 0.876(35) = 12.298$

$Y^* = 12.3$

$L_1 = 12.3 - t_{(0.99;13)}(1.66)$

$L_1 = 12.3 - 2.6501(1.66)$

$L_1 = 8$

$L_s = 12.3 - t_{(0.99;13)}(1.66)$

$L_s = 12.3 + 2{,}6501(1.66)$

$L_s = 17$

El intervalo de predicción será:

$P[8 \leq Y^* \leq 17] = 0.99$

Interpretación:

El 99% de los trabajadores de 35 años de edad tendría un tiempo de servicio comprendido entre 8 y 17 años

EJERCICIOS PROPUESTOS

1.- Se desea investigar la estructura de los sueldos de los docentes y saber si existe relación con los años de experiencia.

Años de Experiencia	9	7	5	16	15	18	20	13	10	8	4	12
Sueldo (miles)	1.2	0.8	0.7	2.5	2.3	2.6	3.2	1.9	1.7	1.5	0.7	1.8

a) Construir el diagrama de esparcimiento
b) Determinar e Interpretar la ecuación de regresión
c) Graficar la línea de regresión
d) Calcular e interpretar el error estándar de estimación.
e) Establecer e interpretar los intervalos de predicción.

2.- Un profesor desea establecer si existe la relación entre el tiempo (minutos) para hacer la tarea y el tiempo gastado para aprender (en minutos) de 11 estudiantes.

Tiempo de hacer la tarea	40	35	20	38	17	26	22	12	12	5	28
Tiempo gastado en aprender	30	30	40	40	50	50	60	60	70	70	60

a) Construir el diagrama de esparcimiento
b) Determinar e Interpretar la ecuación de regresión
c) Graficar la línea de regresión
d) Calcular e interpretar el error estándar de estimación.
e) Establecer e interpretar los intervalos de predicción.

3.- El gerente de un taller desea saber si existe relación entre las horas de trabajo y la producción (unidades producidas).

Horas	80	79	83	84	78	60	82	85	79	84	80	62
Producción	300	302	315	330	300	250	300	340	315	330	310	240

a) Construir el diagrama de esparcimiento
b) Determinar e Interpretar la ecuación de regresión
c) Graficar la línea de regresión
d) Calcular e interpretar el error estándar de estimación.
e) Establecer e interpretar los intervalos de predicción.

4.- Una compañía de seguros considera que el número de vehículos que circulan por una determinada autopista a más de 120 km/h puede ser en función del número de accidentes que ocurren en ella durante 5 días se obtuvo:

Accidentes	5	7	2	1	9
Nº Vehículos	15	18	10	8	20

a) Construir el diagrama de esparcimiento
b) Determinar e Interpretar la ecuación de regresión
c) Graficar la línea de regresión
d) Calcular e interpretar el error estándar de estimación.
e) Establecer e interpretar los intervalos de predicción.

5.- Un centro comercial desea saber que en función de la distancia en km, a la que se encuentra un núcleo de la población, acuden los clientes.

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

a) Construir el diagrama de esparcimiento
b) Determinar e Interpretar la ecuación de regresión
c) Graficar la línea de regresión
d) Calcular e interpretar el error estándar de estimación.
e) Establecer e interpretar los intervalos de predicción.

RESUMEN

Una investigación estadística con frecuencia implica el examen de la relación entre dos conjuntos de variables. El análisis de regresión se ocupa en parte del desarrollo de una expresión matemática para tal relación. La forma general de la ecuación de regresión es $Y^* = a + bX$ en donde Y^* es el valor pronosticado de Y, dado un valor específico de X.

La recta de "mejor ajuste" se puede determinar utilizando la técnica de mínimos cuadrados. Minimizar la suma de los cuadrados de las desviaciones entre los valores Y reales y los valores Y^* pronosticados sobre la recta regresión.

El error estándar de estimación S_{XY} mide la exactitud del pronóstico. También se utiliza para desarrollar un intervalo de coeficiente correspondiente.

CAPÍTULO II

CORRELACIÓN LINEAL

OBJETIVOS:

1. Calcular y explicar el coeficiente de correlación lineal.
2. Interpretar el significado del coeficiente de correlación rectilínea.
3. Calcular y explicar el uso de los coeficientes de determinación y no determinación.

2.1. Correlación

Expresa el grado de asociación o afinidad entre las variables consideradas, también explica el grado de la bondad del ajuste de las líneas de regresión

2.2. Análisis de Correlación

Grupo de técnicas estadísticas empleado para medir la intensidad de la relación entre dos variables. El principal objetivo del análisis de correlación consiste en determinar que tan intensa es la relación entre dos variables.

2.3. Correlación Simple

Aquella que trata de analizar la relación entre dos variables

2.4. Correlación Rectilínea o Lineal

Cuando la función de regresión es una recta.

2.5. Coeficiente de Correlación Rectilínea

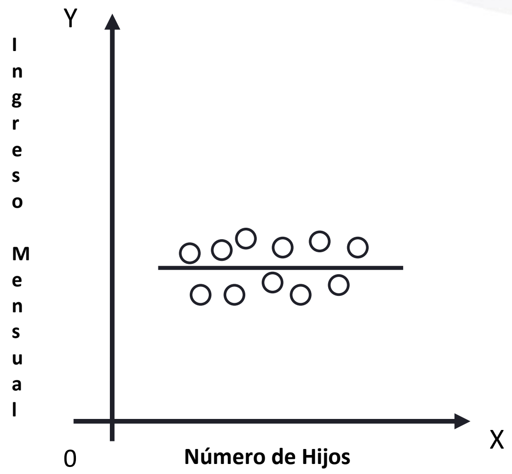
Es el estadígrafo que mide o expresa el grado de afinidad o asociación entre dos variables, cuando ellas están relacionadas mediante una línea recta.

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

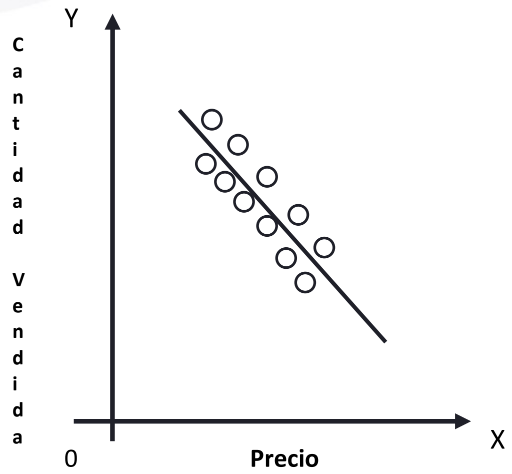
GRÁFICO Nº 2.1

DIAGRAMAS DE DISPERSIÓN QUE MUESTRAN CORRELACIÓN CERO, NEGATIVA Y POSITIVA

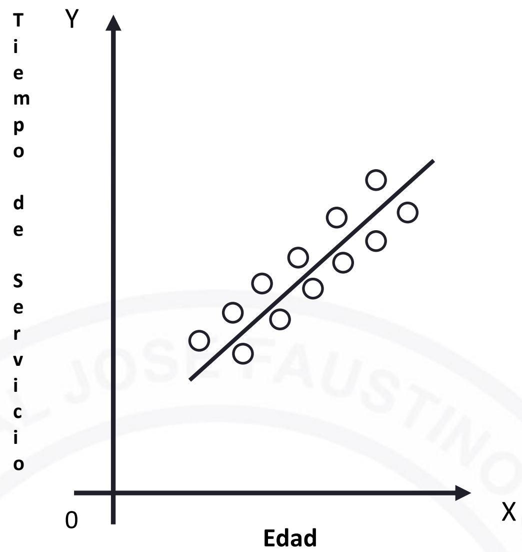
Correlación Cero, r=0
(X y Y no están relacionadas linealmente)



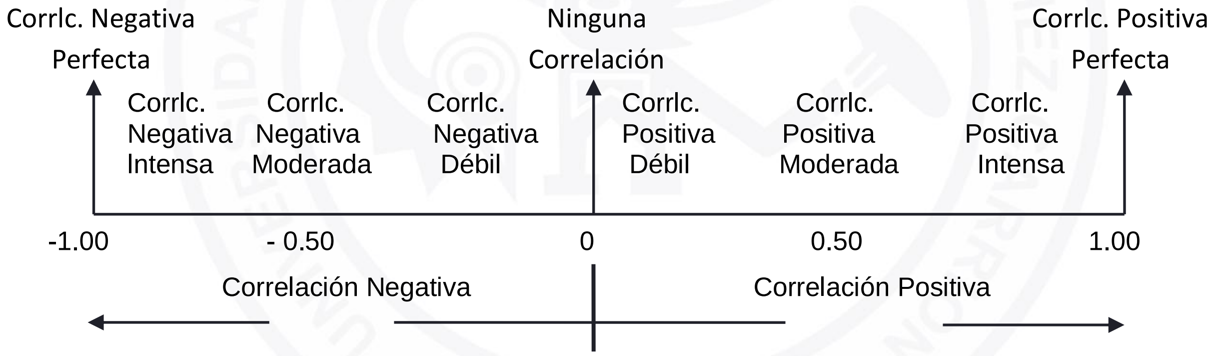
Correlación Negativa
(X y Y tienen relación lineal inversa)



Correlación Positiva
(X y Y tienen relación lineal directa)



El esquema que sigue representa adecuadamente la intensidad y la dirección del coeficiente de correlación.



2.6 Propiedades de r

La propiedad fundamental del coeficiente de correlación “r” es:

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

a) Si r > 0 entonces existe correlación directa positiva

b) Si r < 0 se trata de una correlación inversa negativa

c) Si r= 1 los datos forman una línea recta, en el caso de la correlación rectilínea.

d) Si r = +1 hay una correlación perfecta positiva

e) Si r = -1 hay una correlación perfecta negativa

f) Si r = 0 los datos son incorrelacionados

El signo de "r" es el mismo que el signo "b" coeficiente angular o pendiente de la recta de la ecuación de regresión $Y^* = a + bX$

2.7. Interpretación Clásica

a) $0 < r < \pm 0{,}20$ existe correlación no significativa.
b) $\pm 0{,}20 < r < \pm 0{,}40$ existe una correlación baja.
c) $\pm 0{,}40 \leq r < \pm 0{,}70$ existe una significativa correlación.
d) $\pm 0{,}70 \leq r < \pm 1{,}00$ existe alto grado de asociación.

2.8. Coeficiente de Determinación (r^2)

Proporción de la variación total en la variable dependiente Y que se explica por, o se debe a, la variación en la variable independiente X. Se calcula al elevar al cuadrado el coeficiente de correlación.

2.9. Coeficiente de No Determinación

Proporción de la variable total en Y que no es explicada por la variación en X. Se calcula por medio de $1 - r^2$.

Los coeficientes de determinación y no determinación solo pueden ser positivos (porque al elevar al cuadrado una r negativa da como resultado un número positivo). Los coeficientes pueden tomar cualquier valor entre 0 y 1,00 inclusive.

El coeficiente de determinación siempre es menor que el coeficiente de correlación.
Algunos estadígrafos preferirán utilizar el coeficiente de determinación como una medida más conservadora considerando que el coeficiente de correlación puede exagerar la relación entre los dos conjuntos de variables.

EJERCICIOS RESUELTOS

1.- Una gasolinera desea saber si existe relación entre la recaudación (miles de soles) durante las últimas 7 semanas, así como el número de clientes (en miles) que acudieron durante esos periodos.

Recaudación	1,5	10	8	3	5	15	2
Nº de Clientes	3	6	5	3,5	4	8	3,2

a) Calcular el coeficiente de correlación rectilínea.
b) Calcular e interpretar el coeficiente de determinación.
c) Calcular e interpretar el coeficiente de no determinación.

SOLUCIÓN:

a)

CUADRO Nº 2.1
COEFICIENTE DE CORRELACIÓN RECTILÍNEA, COEFICIENTE DE DETERMINACIÓN Y NO DETERMINACIÓN DE LA RECAUDACIÓN Y EL NÚMERO DE CLIENTES DE UNA GASOLINERA

X	Y	XY	X^2	Y^2
3	1,5	4,5	9	2,25
6	10	60	36	100
5	8	40	25	64
3,5	3	10,5	12,25	9
4	5	20	16	25
8	15	120	64	225
3,2	2	6,4	10,24	4
32,7	44,5	261,4	172,49	429,25
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$

Coeficiente de correlación rectilínea del Cuadro Nº 2.1

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

$$r = \frac{1829,8 - 1455,15}{\sqrt{[1207,43 - 1069,29][3004,75 - 1980,25]}}$$

$$r = 0.99$$

b) Calcular e interpretar el coeficiente de determinación.

$$r^2 = (0.99)^2 = 0.9801$$

$$r^2 = 0.98$$

Interpretación:

El 98% de la variación total de la recaudación semanal se explica por, o se debe a la variación en el número de clientes.

c) Calcular e interpretar el coeficiente de no determinación.

$$1 - r^2 = 1 - 0.98 = 0.02$$

Interpretación:

El 2% de la variación total de la recaudación semanal no se debe a la variación en el número de clientes.

2.- El director de una institución educativa desea establecer si existe relación entre los pesos en libras y las estaturas en pulgadas de una muestra de niños de cinco años de edad:

Estatura	38	39	40	41	42	43	44	45	46	47
Peso	34	35	36	38	39	41	44	46	47	49

a) Determinar el coeficiente de correlación rectilínea.
b) Determinar e interpretar el coeficiente de determinación.
c) Determinar e interpretar el coeficiente de no determinación.

SOLUCIÓN:

a) Coeficiente de correlación rectilínea del Cuadro Nº 2.2

$$r = \frac{10(17527) - (409)(425)}{\sqrt{\left[10(16985) - (409)^2\right]\left[10(18145) - (425)^2\right]}}$$

$$r = 0.99$$

CUADRO Nº 2.2

COEFICIENTE DE CORRELACIÓN RECTILÍNEA, COEFICIENTE DE DETERMINACIÓN Y NO DETERMINACIÓN DE LOS PESOS Y ESTATURA DE NIÑOS DE CINCO AÑOS DE EDAD DE UNA INSTITUCIÓN EDUCATIVA

X	Y	XY	X^2	Y^2
34	38	1292	1156	1444
35	39	1365	1225	1521
36	40	1440	1296	1600
38	41	1558	1444	1681
39	42	1638	1521	1764
41	43	1763	1681	1849
44	44	1936	1936	1936
46	45	2070	2116	2025
47	46	2162	2209	2116
49	47	2303	2401	2209
409	425	17527	16985	18145
$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$

b) Determinar e interpretar el coeficiente de determinación.

$$r^2 = (0.99)^2 = 0.9801$$

$$r^2 = 0.98$$

Interpretación:

El 98% de la variación total de la estatura de los niños se explica por, o se debe a la variación en el peso.

c) Determinar e interpretar el coeficiente de no determinación.

$$1 - r^2 = 1 - 0.98 = 0.02$$

Interpretación:

El 2% de la variación total de la estatura de los niños no se debe a la variación en el peso.

EJERCICIOS PROPUESTOS

1.- Se desea investigar la estructura de los sueldos de los docentes y saber si existe relación con los años de experiencia.

Años de Experiencia	9	7	5	16	15	18	20	13	10	8	4	12
Sueldo (miles)	1.2	0.8	0.7	2.5	2.3	2.6	3.2	1.9	1.7	1.5	0.7	1.8

a) Hallar e interpretar el coeficiente de correlación.
b) Hallar e interpretar el coeficiente de determinación.
c) Hallar e interpretar el coeficiente de no determinación.

2.- Un gerente desea saber si existe relación entre las horas de trabajo en un taller y el número de unidades producidas.

Horas	80	79	83	84	78	60	82	85	79	84	80	62
Producción	300	302	315	330	300	250	300	340	315	330	310	240

a) Hallar e interpretar el coeficiente de correlación
b) Hallar e interpretar el coeficiente de terminación y no determinación.

3.- Un profesor desea establecer si existe la relación entre el tiempo (minutos) para hacer la tarea y el tiempo gastado para aprender (en minutos) de 11 estudiantes.

Tiempo de hacer la tarea	40	35	20	38	17	26	22	12	12	5	28
Tiempo gastado en aprender	30	30	40	40	50	50	60	60	70	70	60

a) Hallar e interpretar el coeficiente de correlación.
b) Hallar e interpretar el coeficiente de determinación y no determinación.

4.- Una compañía de seguros considera que el número de vehículos que circulan por una determinada autopista a más de 120 km/h puede ser en función del número de accidentes que ocurren en ella durante 5 días se obtuvo:

Accidentes	5	7	2	1	9
Nº Vehículos	15	18	10	8	20

a) Hallar e interpretar el coeficiente de correlación.
b) Hallar e interpretar el coeficiente de determinación y no determinación.

RESUMEN

El coeficiente de correlación r puede tomar cualquier valor entre -1.00 y +1.00 coeficiente cercanos a -1 y +1 indican que existe una correlación intensa entre las dos variables de interés. Un coeficiente cercano a cero indica correlación débil, y uno de cero significa que no existe correlación. El signo negativo antes de r indica que existe una relación inversa, lo cual significa que conforme X aumenta Y disminuye. Una correlación positiva significa que si X aumenta Y se incrementa. El signo de r no tiene que ver con la intensidad; r = - 0.31 y r = + 0.31 denotan igual intensidad pero ambos indican relaciones débiles.

Otra dos medidas de relación son el coeficiente de determinación y el coeficiente de no determinación el primero se determina al elevar r al cuadrado y se define como la proporción de la variación en Y explicada por medio de X. el coeficiente de no determinación se obtiene por medio de 1- r^2 y es la proporción de la variación en Y no explicada en X

CAPÍTULO III

REGRESIÓN NO LINEAL

OBJETIVOS:

1. Trazar un diagrama de esparcimiento.
2. Describir la relación mediante una función no lineal entre las variables.
3. Calcular e interpretar el error estándar de estimación.
4. Graficar la curva de regresión no lineal.

3.1. Parábola de Segundo Grado

La curva o función de regresión parabólica se construye a partir la ecuación polinomial de segundo grado:

$$Y^{*} = a + bX + cX^{2}$$

Que tiene tres parámetros o coeficientes desconocidos a, b, c. Para determinar el valor de estos tres parámetros se requiere de tres ecuaciones. Precisamente aplicando el Método de Mínimos Cuadrados se obtiene las tres ecuaciones normales siguientes:

1) $\sum Y = na + b\sum X + c\sum X^{2}$

2) $\sum XY = a\sum X + b\sum X^{2} + c\sum X^{3}$

3) $\sum X^{2}Y = a\sum X^{2} + b\sum X^{3} + c\sum X^{4}$

3.1.1. Deducción de las Ecuaciones Normales

Por el método de los mínimos cuadrados se parte de la expresión:

$Min\ \phi = \sum\left(Y - Y^{*}\right)^{2}$ donde reemplazando

$Y^{*} = a + bX + cX^{2}$ se tiene:

$$Min\ \phi = \sum\left(Y - a - bX - cX^{2}\right)^{2}$$

Derivando parcialmente con respecto a los parámetros a, b, c e igualando a cero resulta:

$$\frac{\partial\phi}{\partial a} = 2\sum\left(Y - a - bX - cX^{2}\right)\left(-1\right) = 0$$

$$\frac{\partial\phi}{\partial b} = 2\sum\left(Y - a - bX - cX^{2}\right)\left(-X\right) = 0$$

$$\frac{\partial\phi}{\partial c} = 2\sum\left(Y - a - bX - cX^{2}\right)\left(-X^{2}\right) = 0$$

Desarrollando sumatorias y trasponiendo términos, se obtiene las tres ecuaciones normales anotadas anteriormente en este sistema de tres ecuaciones con tres incógnitas (a,b,c) se reemplaza el valor de las sumatorias que se obtiene de los datos y luego se resuelve.

3.1.2. Error Estándar de Estimación

Para determinar el error estándar de la Parábola se parte de la definición:

$$S_{yx} = \sqrt{\frac{\sum\left(Y - Y^{*}\right)^{2}}{n}}\ .$$

Donde se sustituye $Y^* = a + bX + cX^2$ para obtener:

$$S_{yx} = \sqrt{\frac{\sum (Y - a - bX - cX^2)^2}{n}}$$

Desarrollando el cuadrado, factorizando y simplificando términos resulta:

$$S_{yx} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY - c\sum X^2 Y}{n}}$$

Que finalmente es la formula para calcular el error estándar de estimación.

Aquí $S_{yx}^2 = S_e^2$ constituye la Varianza Residual.

EJERCICIO RESUELTO

En las primeras columnas del cuadro Nº 3.1.están el volumen mensual de venta (Y) en millones de soles y los años de experiencia en ventas (X) de 10 vendedores profesionales de una fabrica productora de alimentos.

a) Construir el diagrama de esparcimiento.
b) Determinar la Curva de Regresión Parabólica.
c) Calcular el Error Estándar de Estimación.
d) Graficar la curva de regresión no lineal obtenida.

SOLUCIÓN:

a) El diagrama de esparcimiento de los puntos (X,Y) se muestra en el Grafico Nº 3.1.

b) Para la curva de regresión parabólica.

$$Y^* = a + bX + cX^2$$

Las ecuaciones normales son:

$$\sum Y = na + b\sum X + c\sum X^2$$

$$\sum XY = a\sum X + b\sum X^2 + c\sum X^3$$

$$\sum X^2 Y = a\sum X^2 + b\sum X^3 + c\sum X^4$$

Sustituyendo el valor de las sumatorias calculadas en el cuadro Nº3.1. resulta:

$$64 = 10a + 48b + 278c$$

$$361 = 48a + 278b + 1764c$$

$$2297 = 278a + 1764b + 11798c$$

Resolviendo el sistema se obtiene:

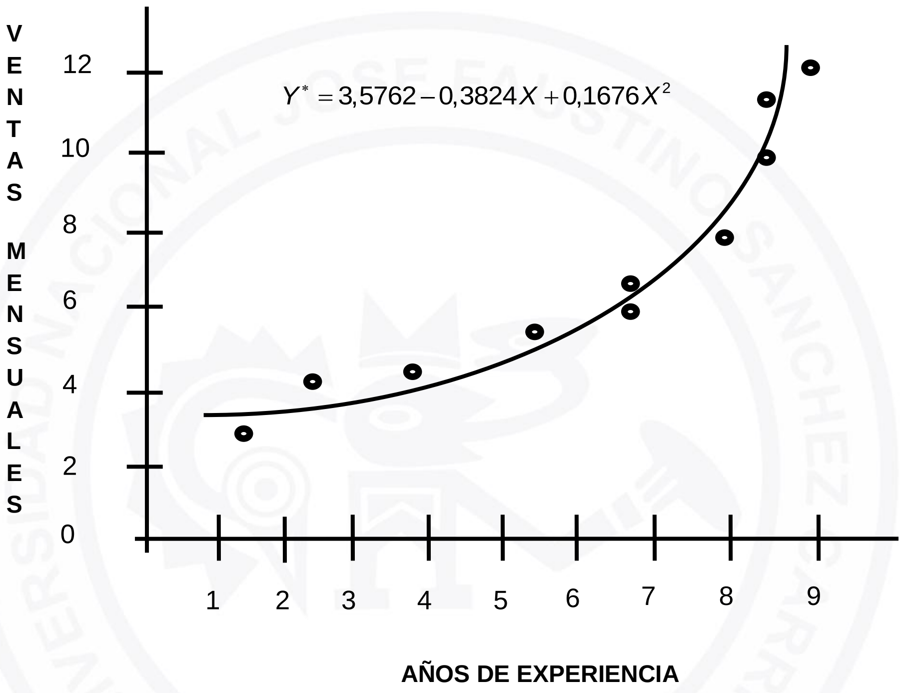
$$a = 3.5762 \qquad b = -0.3824 \qquad c = 0.1676$$

Luego la ecuación de regresión parabólica es:

$$Y^* = 3.5762 - 0.3824X + 0.1676X^2$$

GRÁFICO Nº3.1

DIAGRAMA DE ESPARCIMIENTO Y CURVA DE REGRESIÓN PARABÓLICA



c) En la fórmula del error estándar de estimación.

$$S_{yx} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY - c\sum X^2Y}{n}}$$

Se remplaza las sumatorias y parámetros:

$$S_{yx} = \sqrt{\frac{478 - (3.5762)(64) - (-0.3824)(361) - (0.1676)(2297)}{10}}$$

$$S_{yx} = \sqrt{0.2192} = 0.4682$$

CUADRO Nº3.1

DETERMINACIÓN DE LA CURVA DE REGRESIÓN PARABÓLICA DE LAS VENTAS MENSUALES (Y) Y EXPERIENCIA (X) EN VENTAS DE 10 VENDEDORES PROFESIONALES

Y	X	XY	X^2	X^3	X^4	X^2 Y	Y^2
5	4	20	16	64	256	80	25
6	5	30	25	125	625	150	36
4	2	8	4	8	16	16	16
5	5	25	25	125	625	125	25
7	6	42	36	216	1296	252	49
10	7	70	49	343	2401	490	100
3	1	3	1	1	1	3	9
11	8	88	64	512	4096	704	121
4	3	12	9	27	81	36	16
9	7	63	49	343	2401	441	81
64	48	361	278	1764	11798	2297	478
$\sum Y$	$\sum X$	$\sum XY$	$\sum X^2$	$\sum X^3$	$\sum X^4$	$\sum X^2Y$	$\sum Y^2$

d) Para graficar la ecuación:

$$Y^* = 3.5762 - 0.3824X + 0.1676X^2$$

Tenemos que determinar algunos puntos aislados asignando valores a X, si tenemos:

Si:

$X = 0$	$Y = 3.6$	$P_1(0;3,6)$
$X = 1$	$Y = 3.4$	$P_2(1;3,4)$
$X = 5$	$Y = 5.8$	$P_3(5;5,8)$
$X = 8$	$Y = 11.2$	$P_4(8;11,2)$

Estos 4 puntos se ubican en el plano y por ellos aproximadamente a mano alzada construimos la parábola como se observa en el grafico Nº 3.1, la parábola pasa a lo largo del diagrama de esparcimiento. Los valores observados Y están ubicados alrededor de la parábola, en tanto que los valores teóricos o estimados Y^* están ubicados en el lugar geométrico de la parábola $Y^* = 3,5762 - 0,3824X + 0,1676X^2$

3.2. La Función Potencial

La curva de regresión potencial se construye a partir de la función potencial cuya ecuación es:

$$Y^* = bX^a$$

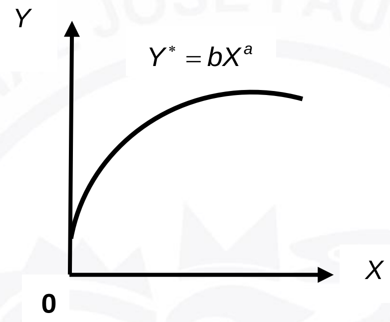
que tiene dos parámetros desconocidos a,b. en este caso se trata de ajustar una curva potencial a la nube de puntos (X,Y).

Recordemos que por una nube de puntos (X,Y) pueden pasar muchas funciones potenciales, de esta familia de curvas se elige la que mejor se ajusta a los valores de (X,Y), es decir que las diferencias o residuo ($Y-Y^*$) sean mínimos. Tal como se ha establecido, el método de los mínimos cuadrados permite determinar la mejor curva.

Como hay dos parámetros (a,b) se necesita dos ecuaciones normales.

Para facilitar la determinación de las dos ecuaciones normales conviene expresar la función original en términos logarítmicos.

Si $Y^* = bX^a$ entonces $Log\ Y^* = \log b + a.\log X$ ahora los parámetros son: log b; a.



3.2.1. Deducción de las Ecuaciones Normales

Aplicando el método de los mínimos cuadrados, la expresión por minimizar es:

$$Min\ \phi = \sum (\log Y - \log Y^*)^2$$

Sustituyendo $\log Y^*$: $$= \sum (\log Y - \log b - a.\log X)^2 .$$

Derivando parcialmente, respecto a los parámetros log b; a,

$$\frac{\partial \phi}{\partial (\log b)} = 2\sum (\log Y - \log b - a.\log X)(-1) = 0$$

$$= \sum (\log Y - \log b - a.\log X) = 0$$

$$\frac{\partial \phi}{\partial a} = 2\sum (\log Y - \log b - a.\log X)(-\log X) = 0$$

$$= \sum (\log X \log Y - \log b.\log X - a.\log^2 X) = 0$$

Resultando las ecuaciones normales siguientes:

$$\sum \log Y = n \log b + a \sum \log X$$
$$\sum \log X \log Y = \log b \sum \log X = a \sum (\log X)^2$$

En este par de ecuaciones reemplazar el valor de las sumatorias obtenidas en el respectivo cuadro de trabajo a partir de los datos bidimensionales (X,Y), luego se resuelve simultáneamente el sistema de ecuaciones para obtener el valor de los dos par a y log b.

EJERCICIO RESUELTO

El ingreso y consumo promedio mensual (en miles de soles) de una muestra de 12 familias de distintos estratos sociales, fue el siguiente:

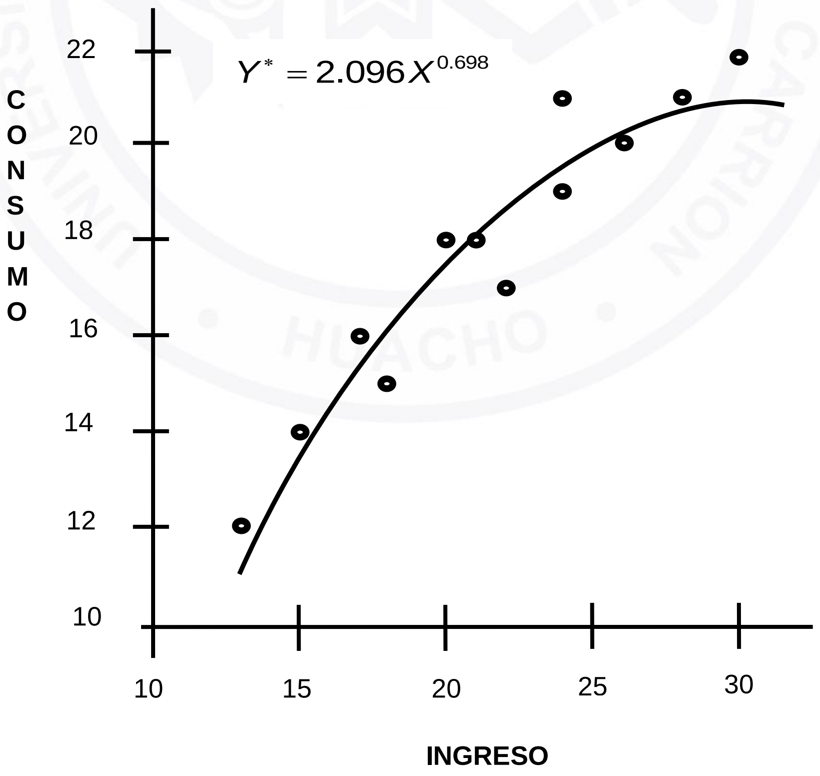
Ingreso :	13	15	17	18	20	21	22	24	24	26	28	30
Consumo:	12	14	16	15	18	18	17	19	21	20	21	22

a) Construir la nube de puntos o diagrama de esparcimiento.
b) Ajustar los datos a una curva de regresión potencial
c) Graficar la ecuación potencial.
d) Estimar el valor del consumo de una familia con un ingreso mensual de 25 000 soles.

SOLUCIÓN:

a) Los puntos (X,Y) se grafican en un plano rectangular y resulta el diagrama de esparcimiento o nube de puntos del Grafico Nº3.2. como se observa la nube de puntos sugiere una curva.

GRAFICO Nº3.2
DIAGRAMA DE ESPARCIMIENTO Y CURVA POTENCIAL DEL INGRESO (X) Y CONSUMO (Y) DE UNA MUESTRA DE 12 FAMILIAS. 2010



b) Trabajamos con la curva de regresión potencial en su forma logarítmica:

$$Log\ Y^{*} = \log b + a.\log X$$

Cuyas Ecuaciones Normales son:

$$\sum \log Y = n \log b + a \sum \log X$$
$$\sum \log X \log Y = \log b \sum \log X = a \sum (\log X)^2$$

Sustituyendo valores de las sumatorias calculadas en el Cuadro Nº3.2 se obtiene:

$$14.9130 = 12 \log b + 15.8410\ a$$
$$19.7794 = 15.8410 \log b + 21.0450\, a$$

De donde

$a = 0.698 \qquad \log b = 0.3213 \qquad o \qquad b = 2.096$

Luego:

$$Log\ Y^{*} = 0.3213 + 0.698 X$$

O también:

$$Y^{*} = 2.096 X^{0.698}$$

c) Para graficar la ecuación:

$$Y^{*} = 2.096 X^{0.698}$$

El método más sencillo es determinar algunos puntos aislados (X,Y) para el efecto se asigna valores a X de donde resulta el valor de Y.

Si

$X = 10$ resulta $Y = 10.5$ $\qquad P_1(10;10.5)$

$X = 15$ resulta $Y = 13.9$ $\qquad P_2(15;13.9)$

$X = 20$ resulta $Y = 17.0$ $\qquad P_3(20;17.0)$

$X = 25$ resulta $Y = 19.8$ $\qquad P_4(25;19.8)$

$X = 30$ resulta $Y = 22.5$ $\qquad P_5(30;22.5)$

Estos 5 puntos se grafican en el mismo plano de la nube de puntos (Grafico Nº 3.2) luego por estos puntos trazar a “mano alzada” la curva potencial obtenida. Se observa que los puntos del diagrama de esparcimiento están muy cerca a la curva; entonces puede afirmarse que el error estándar de estimación es muy pequeño, es decir que la curva obtenida se ajusta muy bien a los datos bidimensionales Consumo (Y) e Ingreso (X).

d) Para un ingreso de 25 000 soles mensuales, significa reemplazar en la función potencial X = 25; luego:

$$Y^* = 2.096(25)^{0.698} \qquad Y^* = 19.82 \text{ miles de soles}$$

Aproximadamente una familia que tiene un ingreso de 25 000 soles gastará o consume alrededor de 19 820 soles mensuales.

CUADRO Nº3.2

DETERMINACIÓN DE LA CURVA DE REGRESIÓN POTENCIAL DEL CONSUMO (Y) PROMEDIO MENSUAL E INGRESO (X) DE UNA MUESTRA DE 12 FAMILIAS. 2010

X	Y	Log X	Log Y	Log X. Log Y	(Log X)^2	(Log Y)^2
13	12	1.1139	1.0792	1.2021	1.2409	1.1646
15	14	1.1761	1.1461	1.3480	1.3832	1.3136
17	16	1.2304	1.2041	1.4816	1.5140	1.4499
18	15	1.2553	1.1761	1.4763	1.5757	1.3832
20	18	1.3010	1.2553	1.6331	1.6927	1.5757
21	18	1.3222	1.2553	1.6597	1.7483	1.5757
22	17	1.3424	1.2304	1.6518	1.8021	1.5140
24	19	1.3802	1.2788	1.7650	1.9050	1.6352
24	21	1.3802	1.3222	1.8249	1.9050	1.7483
26	20	1.4150	1.3010	1.8409	2.0021	1.6927
28	21	1.4472	1.3222	1.9135	2.0943	1.7483
30	22	1.4771	1.3424	1.9829	2.1819	1.8021
258	213	15.8411	14.9132	19.7799	21.0451	18.6033
$\sum X$	$\sum Y$	$\sum Log\ X$	$\sum Log\ Y$	$\sum Log\ X.Log\ Y$	$\sum (Log\ X)^2$	$\sum (Log\ Y)^2$

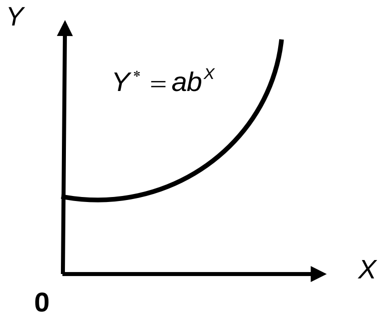
3.3. La Función Exponencial

La curva de regresión exponencial se determina a partir de la función exponencial de la forma:

$$Y^* = ab^X$$

Que también tiene dos parámetros (a,b).

Esta función se utiliza cuando interesa calcular tasas de incrementos considerando todos los puntos observados durante el periodo, aquí se supone que existe un crecimiento no lineal de tipo geométrico.



El problema es ajustar líneas a nubes de puntos (X,Y) de comportamiento no lineal, en esta oportunidad interesa determinar la función exponencial que mejor se ajuste a la nube de puntos (X,Y).

De la misma manera que la función potencial, transformar la función exponencial original en forma logarítmica. Entonces aplicando el operador logaritmo a la expresión $Y^* = ab^X$ se transforma en $\log Y^* = \log a + X \log b$ en donde los parámetros son: log a, log b.

3.3.1. Deducción de las Ecuaciones Normales

Para determinar la mejor curva exponencial de un conjunto de puntos se aplica el método de mínimos cuadrados, que significa minimizar la expresión:

$$Min\ \phi = \sum(\log Y - \log Y^*)^2$$

Sustituyendo $\log Y^*$ se tiene:

$$Min\ \phi = \sum(\log Y - \log a - X \log b)^2$$

Derivando parcialmente respecto a los parámetros log a, log b se obtienen las ecuaciones normales para el caso exponencial.

$$\frac{\partial \phi}{\partial(\log a)} = 2\sum(\log Y - \log a - X \log b)(-1) = 0$$

$$\frac{\partial \phi}{\partial \log b} = 2\sum(\log Y - \log a - X \log b)(-X) = 0$$

Efectuando las operaciones indicadas en las derivadas, resultan las siguientes ecuaciones normales:

$$\sum \log Y = n \log a + \log b \sum X$$

$$\sum X \log Y = \log a \sum X + \log b \sum X^2$$

Donde las incógnitas son (log a, log b), cuyos valores se obtienen remplazando las sumatorias.

EJERCICIO RESUELTO

Para estimar la función de Costo Total (C medida en millones de soles) con respecto a la producción total (Q medida en miles de unidades), un fabricante ha obtenido el siguiente conjunto de datos muestrales:

Producción (Q):	10	20	30	40	50	60	70	80
Costo Total (C):	30	36	40	48	50	54	66	68

Con estos datos:

a) Construir el diagrama de esparcimiento.
b) Determinar la función de Costo Total, a través de una curva de regresión exponencial.
c) Graficar la curva de regresión o función de costo total.
d) Estimar el costo total, si se produce 42,000 unidades.

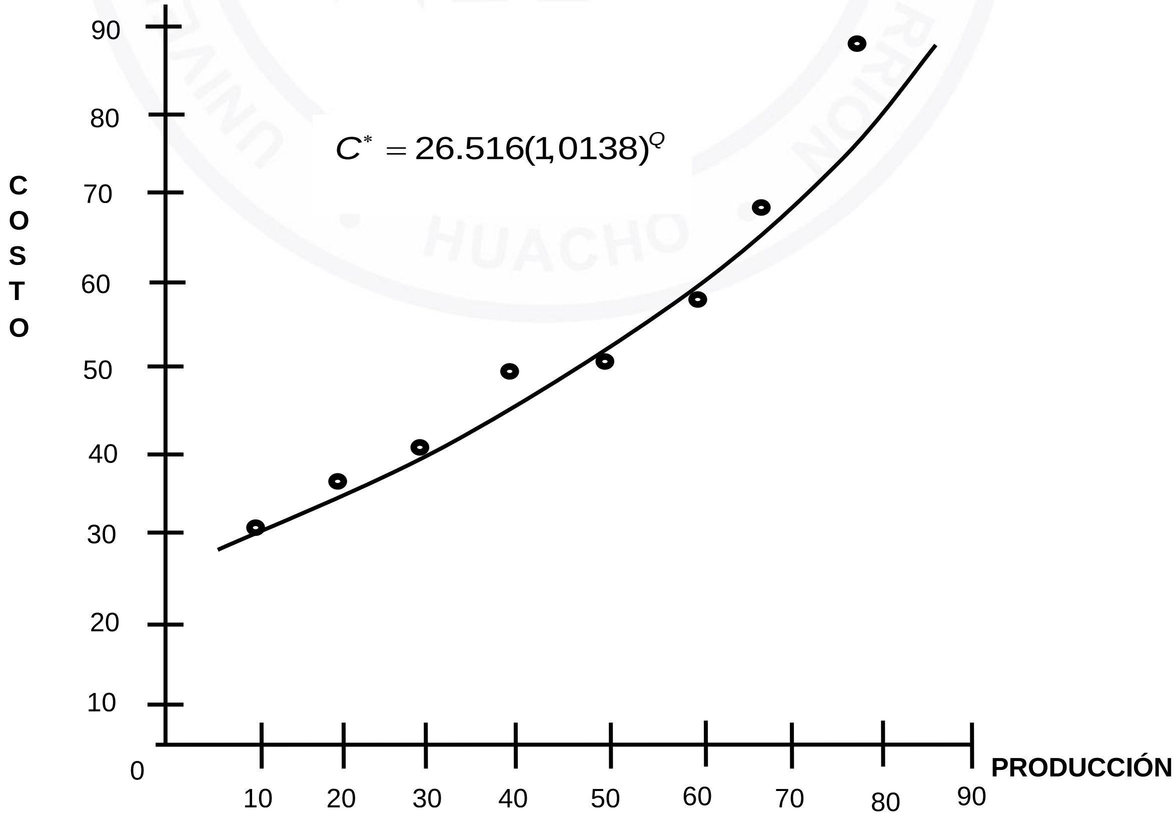
SOLUCIÓN:

En este caso es claro que: Costo Total = f (producción) o sea C= f (Q) donde los puntos bidimensionales serian (Q,C), equivalente a (X,Y).

a) Los puntos (Q,C) constituyen el diagrama de esparcimiento o nube de puntos, que se representa en el Gráfico Nº 3.3.

GRAFICO Nº3.3

DIAGRAMA DE ESPARCIMIENTO Y CURVA DE EXPONENCIAL DE PRODUCCIÓN (Q) Y COSTO TOTAL (C)



b) La curva de regresión exponencial (función costo total) es:

$$Y^* = ab^X \qquad \log Y^* = \log a + X \log b$$

$$C^* = ab^Q \qquad \log C^* = \log a + Q \log b$$

Luego las Ecuaciones Normales son:

$$\sum \log Y = n \log a + \log b \sum X$$
$$\sum Q \log C = \log a \sum Q + \log b \sum Q^2$$

El cuadro de trabajo para obtener el valor de los parámetros log a, log b, es el siguiente:

CUADRO Nº3.3

FUNCIÓN DE REGRESIÓN EXPONENCIAL DE LA PRODUCCIÓN Y COSTO TOTAL

Q_1	C_1	Log C	Q^2	Q log C	$(\log C)^2$
10	30	1.4771	100	14.77	2.1819
20	36	1.5563	400	31.13	2.4221
30	40	1.6021	900	48.06	2.5666
40	48	1.6812	1600	67.25	2.8266
50	50	1.6990	2500	84.95	2.8865
60	54	1.7324	3600	103.94	3.0012
70	66	1.8195	4900	127.37	3.3107
80	88	1.9445	6400	155.56	3.7810
360	412	13.5121	20400	633.03	22.9766
$\sum Q_i$	$\sum c_i$	$\sum \log C$	$\sum Q^2$	$\sum Q \log C$	$\sum (\log C)^2$

Reemplazando los valores en las ecuaciones normales

$$13.512 = 8 \log a + 360 \log b$$
$$633 = 360 \log a + 204000 \log b$$

Resolviendo se obtiene:

$$\log b = 0.0059 \qquad b = \text{antlog } 0.0059 = 1.0138$$
$$\log a = 1.4235 \qquad a = \text{antlog } 1.4235 = 26.5155$$

Luego la función costo estimado:

$$C^* = 26.516(1.0138)^Q$$

c) Para graficar la función exponencial o la función de costo total:

$$C^* = 26.516(1.0138)^Q$$

Se calcula algunos puntos aislados (Q, C) a partir de la ecuación dando valores a Q. donde:

Si:

$Q = 20$ entonces $C = 34.9$

$Q = 40$ entonces $C = 45.9$

$Q = 60$ entonces $C = 60.3$

$Q = 80$ entonces $C = 79.4$

$Q = 90$ entonces $C = 91.0$

Los cinco puntos se grafican en el mismo plano del diagrama de esparcimiento (Gráfico Nº 3.3), y por estos puntos trazar a "mano alzada" la curva exponencial correspondiente. Como es de esperar los datos observados (Q, C) están ubicados muy cerca al lugar geométrico de la curva que reafirma el elevado valor de r = 0.9770.

d) El costo total, para una producción de 42000 unidades significa que Q=42, luego reemplazamos en:

$C^* = 26.516(1.0138)^Q$ se tiene

$C^* = 26.516(1.0138)^{42} = 47.151$ millones de soles

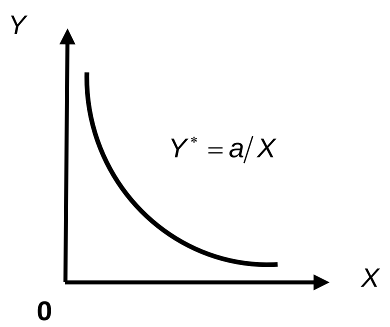
$C^* = 47152000$ soles

3.4. La Hipérbola Equilátera

En un caso especial de la función potencial. Esta curva se usa con frecuencia para el ajuste de curvas de demanda. Es una función asintótica con los ejes coordenados. La forma más simple de su ecuación es:

$$Y^* = \frac{a}{X}$$

Es una función con un sólo parámetro "a".



3.4.1. Deducción de las Ecuaciones Normales

Como tiene un sólo parámetro será necesario disponer de una ecuación normal. Ahora la expresión que minimiza es:

$$Min\ \phi = \sum (Y - Y^*)^2 = \sum \left(Y - \frac{a}{X} \right)^2$$

Derivando respecto al parámetro "a" se tiene:

$$\frac{\partial \phi}{\partial a} = 2\sum \left(Y - \frac{a}{X} \right)\left(-\frac{1}{X} \right) = 0$$

Luego:

$$\sum \left(\frac{Y}{X} \right) = \sum \left(\frac{a}{X^2} \right)$$

Esta ecuación normal permite obtener el valor de "a".

EJERCICIO RESUELTO

Considerar las cantidades vendidas (miles de unidades) de un artículo según la variación de su precio (en soles), observados en un periodo de 9 meses, cuyos datos se indican en el Cuadro Nº3.4

a) Construir el diagrama de esparcimiento
b) Determinar la curva de regresión hiperbólica equilátera.
c) Graficar la curva de regresión obtenida.

SOLUCIÓN:

a) El diagrama de esparcimiento se aprecia en el Grafico Nº 3.4. donde hay nueve puntos (X,Y).

b) La ecuación de la hipérbola equilátera de nuestro análisis y la correspondiente ecuación normal son:

$$Y^* = \frac{a}{X} \quad \text{o también} \quad Y^* = aX^{-1}$$

$$\sum \left(\frac{Y}{X} \right) = \sum \left(\frac{a}{X^2} \right)$$

Reemplazando valores

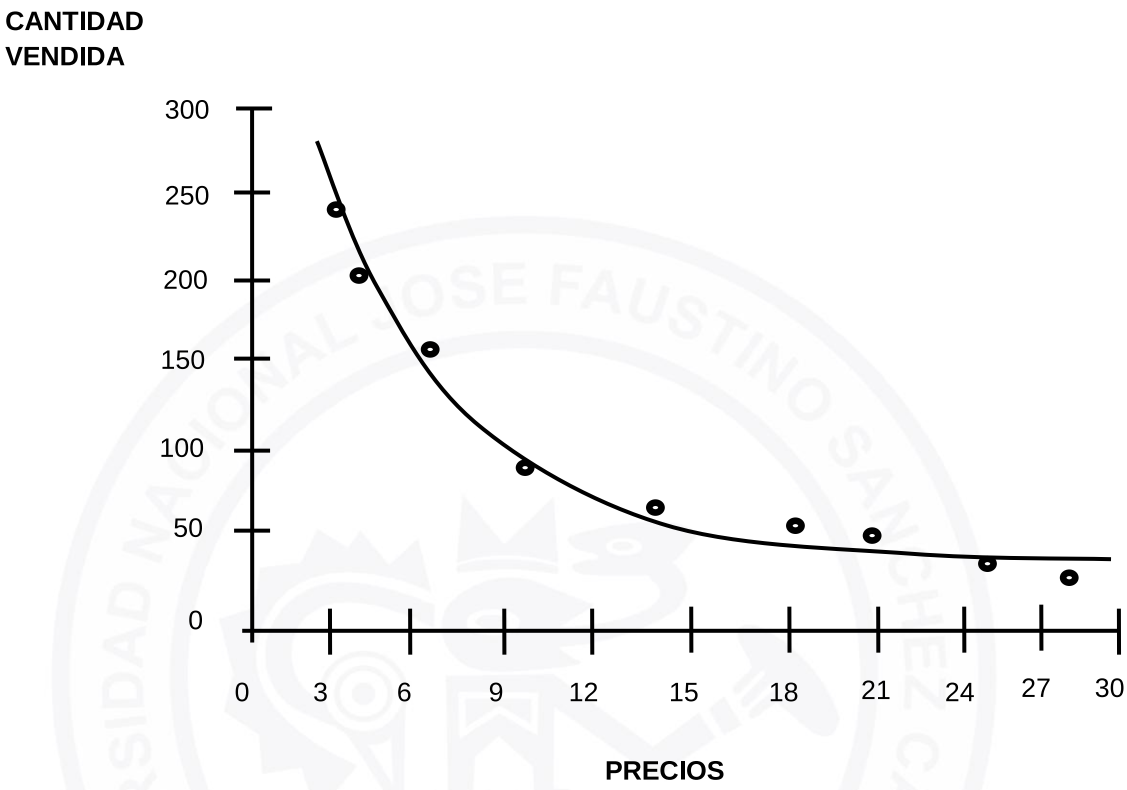
$$143.88 = 0.1414a$$
$$a = 1017.5$$

Luego la ecuación es:

$$Y^* = \frac{1017.5}{X}$$

GRAFICO Nº3.4

DIAGRAMA DE ESPARCIMIENTO Y CURVA DE REGRESIÓN HIPÉRBOLA EQUILÁTERA PARA PRECIOS (X) Y CANTIDADES VENDIDAS (Y) DE UN ARTÍCULO



c) Nuevamente, para graficar la hipérbola equilátera será necesario determinar algunos puntos, como:

Si:

$X = 4 \qquad Y = 254$

$X = 10 \qquad Y = 102$

$X = 16 \qquad Y = 63$

$X = 26 \qquad Y = 39$

Los puntos se grafican en el plano rectangular elegido y por ellos se traza a mano alzada la curva exponencial correspondiente, resultando la curva del Gráfico Nº 3.4.

CUADRO Nº3.4

PRECIOS (X) Y CANTIDADES (Y) TRANSADAS EN EL MERCADO DE UN ARTÍCULO EN 9 MESES

X	Y	$\frac{Y}{X}$	$\frac{1}{X}$	$\frac{1}{X^2}$	Y^2
4	240	60.00	0.250	0.0625	57600
5	200	40.00	0.200	0.0400	40000
8	150	18.75	0.125	0.0156	22500
10	100	10.00	0.100	0.0100	10000
14	80	5.71	0.071	0.0051	6400
18	70	3.89	0.056	0.0031	4900
21	60	2.86	0.048	0.0023	3600
25	40	1.60	0.040	0.0016	1600
28	30	1.07	0.036	0.0013	900
133	970	143.88	0.925	0.1415	147500
$\sum X$	$\sum Y$	$\sum\left(\frac{Y}{X}\right)$	$\sum\left(\frac{1}{X^2}\right)$	$\sum\left(\frac{1}{X^2}\right)$	$\sum Y^2$

EJERCICIOS PROPUESTOS

1. Proyectar la demanda en función del ingreso, teniendo como demanda histórica lo siguiente:

INGRESO	1.2	1.8	3.1	4.9	7.1	8.6
DEMANDA	4.5	5.9	7	7.8	6.8	4.5

a) Construir el diagrama de esparcimiento.
b) Determinar la Curva de Regresión Parabólica.
c) Calcular e interpretar el Error Estándar de Estimación.

2. Proyectar la Oferta de cierto producto tomando en cuenta los datos obtenidos en el estudio de mercado:

AÑOS	TIEMPO	OFERTA
1989	1	100 000
1990	2	120 000
1991	3	140 000
1992	4	110 000
1993	5	170 000
1994	6	150 000
1995	7	180 000
1996	8	200 000
1997	9	210 000
1998	10	200 000

De los datos de la tabla establecer:

a) Construir el diagrama de esparcimiento.
b) Determinar la Curva de Regresión Exponencial.
c) Calcular e interpretar el Error Estándar de Estimación.

3. Un analista de mercado revela la producción de cierto producto y el costo total del mismo en la siguiente tabla:

PRODUCCIÓN	COSTO TOTAL
3	42
6	58
9	62
12	75
15	82
20	93
25	109

De los datos de la tabla establecer:

a) Construir el diagrama de esparcimiento.
b) Determinar la Curva de Regresión Exponencial.
c) Calcular e interpretar el Error Estándar de Estimación.

4. La demanda de un producto sigue la siguiente ecuación:

$$Q = e^{(a-b/p)}$$

Si se tiene los siguientes datos:

AÑO	OFERTA	PRECIO
1990	10	1
1991	11	2
1992	13	3
1993	14	4
1994	17	5

a) Construir el diagrama de esparcimiento.
b) Determinar la Curva de Regresión Hiperbólica Equilátera.
c) Calcular e interpretar el Error Estándar de Estimación.

RESUMEN

En este capitulo nos hemos ocupado del análisis de regresión no lineal, primeramente identificando la variable dependiente y la variable independiente de un conjunto de n observaciones trazando el diagrama de dispersión, con la facilidad de observar el comportamiento de la mejor función no lineal que represente ese conjunto de datos y finalmente se ha determinado el error estándar de estimación correspondiente.

CAPÍTULO IV

CORRELACIÓN NO LINEAL

OBJETIVOS:

1. Calcular el coeficiente de correlación parabólico.
2. Calcular el coeficiente de correlación potencial.
3. Calcular el coeficiente de correlación exponencial.
4. Calcular el coeficiente de correlación hiperbólica.

4.1. Coeficiente de Correlación Parabólica

Para deducir la fórmula del coeficiente de correlación parabólica se considera la definición de coeficiente de correlación, dado por:

$$r = \sqrt{\frac{S^2_{Y^*}}{S^2_Y}} \quad ó \quad r^2 = \frac{S^2_{Y^*}}{S^2_Y}$$

De donde:

$$r^2 = \frac{\sum Y^{*2} - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2}$$

Sustituyendo Y^* se tiene:

$$r^2 = \frac{\sum\left(a + bX + cX^2\right)^2 - n\bar{Y}}{\sum Y^2 - n\bar{Y}^2}$$

Ejecutando operaciones, factorización y sustituyendo expresiones por ecuaciones normales, se obtiene finalmente:

$$r^2 = \frac{a\sum Y + b\sum XY + c\sum X^2 Y - nY}{\sum Y^2 - n\bar{Y}^2}$$

Cuya raíz cuadrada (r) constituye la formula del coeficiente de correlación parabólica.
También se puede aplicar la fórmula:

$$r^2 = 1 - \frac{S^2_{YX}}{S^2_Y}$$

Donde S^2_{YX} (varianza residual) y S^2_Y (varianza de Y) se pueden calcular fácilmente por las formulas ya conocidas.

EJERCICIO RESUELTO

En las primeras columnas del cuadro Nº 3.1 están el volumen mensual de venta (Y) en millones de soles y los años de experiencia en ventas (X) de 10 vendedores profesionales de una fábrica productora de alimentos, calcular el coeficiente de Correlación Parabólica.

SOLUCION

Para calcular e interpretar el coeficiente de correlación parabólica podemos usar:

$$r^2 = 1 - \frac{S^2_{YX}}{S^2_Y}$$

$$r^2 = 1 - \frac{0.2192}{6.84}$$
$$r^2 = 0.9680$$
$$r = 0.9838$$

Sabiendo que:

$$S_Y^2 = \frac{\sum Y^2}{n} - \left(\frac{\sum Y}{n}\right)^2 = \frac{478}{10} - \left(\frac{64}{10}\right)^2 = 6.84$$

También se puede aplicar la fórmula:

$$r^2 = \frac{a\sum Y + b\sum XY + c\sum X^2Y - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}$$

Reemplazando los valores:

$$r^2 = \frac{(3.5762)(64) + (-0.3824)(361) + (0.1676)(2297) - 10(6.4)^2}{478 - 10(6.4)^2}$$

$$r^2 = 0.9679$$
$$r = 0.9838$$

El valor de r = 0.9838, indicaría que la correlación es muy significativa, es decir que la curva parabólica expresa adecuadamente la relación entre las variables consideradas.

4.2. Coeficiente de Correlación Potencial

La determinación del coeficiente de correlación potencial, se realiza a partir de la definición general, dada por la expresión:

$$r = \sqrt{\frac{\sum Y^{*2} - n\overline{Y}^2}{\sum Y^2 - n\overline{Y}^2}}$$

Elevando al cuadrado y sustituyendo la función en términos de logaritmos, resulta

$$r^2 = \frac{\sum(\log Y^*)^2 - n\left(\overline{\log Y}\right)^2}{\sum(\log Y)^2 - n\left(\overline{\log Y}\right)^2}$$

Donde $\overline{\log Y} = \frac{\sum \log Y}{n}$. Ahora reemplazamos $\log Y^*$ resulta:

$$r^2 = \frac{\sum(\log b + a\log X)^2 - n\left(\overline{\log Y}\right)^2}{\sum(\log Y)^2 - n\left(\overline{\log Y}\right)^2}$$

Desarrollando el binomio $(\log b + a \log X)^2$, factorizando y simplificando se obtiene:

$$r^2 = \frac{a\sum \log X \, \log Y + \log b \sum \log Y - n\left(\overline{\log Y}\right)^2}{\sum (\log Y)^2 - n\left(\overline{\log Y}\right)^2}$$

Donde la raíz cuadrada (r) constituye la fórmula del coeficiente de correlación potencial, cuyo valor debe satisfacer la propiedad fundamental: $-1 \le r \le +1$

EJERCICIO RESUELTO

El ingreso y consumo promedio mensual (en miles de soles) de una muestra de 12 familias de distintos estratos sociales, fue el siguiente:

Ingreso :	13	15	17	18	20	21	22	24	24	26	28	30
Consumo:	12	14	16	15	18	18	17	19	21	20	21	22

Calcular el coeficiente de correlación potencial del Cuadro Nº4.2.

SOLUCIÒN:

El coeficiente de correlación potencial se obtiene sustituyendo valores en la fórmula:

$$r^2 = \frac{a\sum \log X \, \log Y + \log b \sum \log Y - n\left(\overline{\log Y}\right)^2}{\sum (\log Y)^2 - n\left(\overline{\log Y}\right)^2}$$

$$r^2 = \frac{(0.698)(19.7794) + (0.3213)(14.913) - 12(1.2427)^2}{18.6030 - 12(1.2427)^2}$$

$$r^2 = 0.9166$$

$$r = 0.9574$$

El valor de r es bastante cercano a +1, lo cual indica que la curva potencial se ajuste muy bien a los datos (X, Y) del ejemplo, y por tanto es un buen modelo para estimar el consumo familiar en función de sus ingresos, en la cuidad de Libertad, siempre que no hayan cambiado sustantivamente los patrones de consumo.

4.3. Coeficiente de Correlación Exponencial

La fórmula del coeficiente de correlación exponencial, se deduce de la definición, donde es suficiente reemplazar $\log Y^*$ resultando que:

$$r^2=\frac{\sum\left(\log Y^*-\overline{\log Y}\right)^2}{\sum\left(\log Y-\overline{\log Y}\right)^2}=\frac{\sum\left(\log Y^*\right)^2-n\left(\overline{\log Y}\right)^2}{\sum\left(\log Y\right)^2-n\left(\overline{\log Y}\right)^2}$$

$$r^2=\frac{\sum\left(\log a+X\log b\right)^2-n\left(\overline{\log Y}\right)^2}{\sum\left(\log Y\right)^2-n\left(\overline{\log Y}\right)^2}$$

Desarrollando el binomio, factorizando y simplificando:

$$r^2=\frac{\log b\sum X\log Y+\log a\sum\log Y-n\left(\overline{\log Y}\right)^2}{\sum\left(\log Y\right)^2-n\left(\overline{\log Y}\right)^2}$$

De esta expresión, la raíz cuadrada (r) es el valor del coeficiente de correlación exponencial.

EJERCICIO RESUELTO

Para estimar la función de Costo Total (C medida en millones de soles) con respecto a la producción total (Q medida en miles de unidades), un fabricante ha obtenido el siguiente conjunto de datos muestrales:

Producción (Q):	10	20	30	40	50	60	70	80
Costo Total (C):	30	36	40	48	50	54	66	68

Con estos datos calcular el coeficiente de correlación exponencial del Cuadro Nº3.3.

SOLUCIÓN:

El coeficiente de correlación exponencial, de la función: $C^*=ab^Q$ es:

$$r^2=\frac{\log b\sum Q\log C+\log a\sum\log C-n\left(\overline{\log C}\right)^2}{\sum\left(\log C\right)^2-n\left(\overline{\log C}\right)^2}$$

Reemplazando valores

$$r^2=\frac{(0.0059)(633)+(1.4235)(13.52)-8(1.689)^2}{22.9760-22.8217}$$

Luego: $r=0.9770$

El valor de (r) indica que existe una alta correlación o afinidad entre las variables C = costo total, Q= producción.

4.4. Coeficiente de Correlación Hiperbólica

El coeficiente de correlación de la hipérbola equilátera se obtiene de la expresión:

$$r^2 = \frac{\sum Y^{*2} - nY^2}{\sum Y^2 - nY^2}$$

$$r^2 = \frac{\sum \left(\frac{a}{X}\right)^2 - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2}$$

$$r^2 = \frac{a\sum \left(\frac{Y}{X}\right) - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2}$$

Donde la raíz cuadrada (r) constituye el valor del coeficiente de correlación hiperbólica

EJERCICIO RESUELTO

Considerar las cantidades vendidas (miles de unidades) de un artículo según la variación de su precio (en soles), observados en un periodo de 9 meses, cuyos datos se indican en el cuadro Nº3.4.
Calcular el coeficiente de correlación de los datos.

SOLUCIÓN:

Para obtener el valor del Coeficiente de Correlación se reemplaza valores en:

$$r^2 = \frac{a\sum \left(\frac{Y}{X}\right) - n\bar{Y}^2}{\sum Y^2 - n\bar{Y}^2} = \frac{(1017.5)(143.88) - 9(107.78)^2}{147.500 - 9(107.78)^2}$$

$r^2 = 0.9743$ entonces $r = 0.9871$

Como r es numérico alto, la curva de regresión obtenida se puede utilizar como un modelo de regresión no lineal para explicar el comportamiento de la variable dependiente (Y) dado los valores (X).

EJERCICIOS PROPUESTOS

1.- Proyectar la Oferta de cierto producto tomando en cuenta los datos obtenidos en el estudio de mercado de los cuales calcular e interpretar el coeficiente de correlación parabólica.

AÑOS	TIEMPO	OFERTA
1989	1	100 000
1990	2	120 000
1991	3	140 000
1992	4	110 000
1993	5	170 000
1994	6	150 000
1995	7	180 000
1996	8	200 000
1997	9	210 000
1998	10	200 000

2.- La demanda de un producto sigue la siguiente ecuación: $Q = e^{(a-b/p)}$ si se tiene los siguientes datos:

AÑO	OFERTA	PRECIO
1990	10	1
1991	11	2
1992	13	3
1993	14	4
1994	17	5

Calcular e interpretar el Coeficiente de Correlación para la función de producción .

3.- Proyectar la demanda en función del ingreso, teniendo como demanda histórica lo siguiente:

INGRESO	1.2	1.8	3.1	4.9	7.1	8.6
DEMANDA	4.5	5.9	7	7.8	6.8	4.5

Calcular e interpretar el Coeficiente de Correlación.

4.- Un analista de mercado revela la producción de cierto producto y el costo total del mismo en la siguiente tabla adjunta, calcular e interpretar el coeficiente de correlación.

PRODUCCIÓN	COSTO TOTAL
3	42
6	58
9	62
12	75
15	82
20	93
25	109

RESUMEN

Este capitulo trata de determinar a cada curva o función no lineal el coeficiente de correlación no lineal teniendo en consideración la propiedad fundamental: $-1 \leq r \leq +1$
Debe tenerse cuidado al interpretar los resultados de una correlación. El coeficiente de correlación r mide el grado de relación o afinidad entre variables, que ayuda a explicar el comportamiento de la variable dependiente dado algunos valores para la variable independiente.
Cuando r está cercano a -1 y +1 nos indica que la correlación es intensa, es decir que la curva expresa adecuadamente la relación entre las variables, lo cual es un buen modelo para realizar la estimación.

GLOSARIO

1. **Estimación,** conjunto de técnicas que permiten dar un valor aproximado de un parámetro de una población a partir de los datos proporcionados por una muestra.

2. **Grados de libertad,** son el número de variables aleatorias independientes de la muestra. Es el número de observaciones independientes en una muestra de los datos que están disponibles para estimar un parámetro de la población.

3. **Intervalo,** conjunto de números reales comprendidos entre a y b que se llaman extremos o límites del intervalo.

4. **Intervalo de confianza o predicción,** es un rango de valores calculado en una muestra en el cual se encuentra el verdadero valor del parámetro, con una probabilidad determinada.

5. **Limite superior del intervalo,** es el valor máximo que tiene el intervalo de confianza o de predicción.

6. **Limite inferior del intervalo,** es el valor mínimo que tiene le intervalo de confianza o de predicción.

7. **Nivel de confianza,** es la probabilidad de que el parámetro a estimas se encuentre en el intervalo de confianza. Los valores mas usados son: 90%, 95% y 99%

8. **Nivel de significación,** es la probabilidad de rechazar erróneamente la hipótesis nula.

9. **Variable,** característica observable, susceptible de adoptar distintos valores o ser expresados en varias categorías.

LISTA DE ABREVIATURAS

r : Coeficiente de Correlación.

r^2 : Coeficiente de determinación.

$1-r^2$: Coeficiente de no determinación.

t : Distribución t de student.

S_e : Error estándar de estimación.

$\bar{X}$: Media muestral de X

$\bar{Y}$: Media muestral de Y

$1-\alpha$: Nivel de confianza.

α : Nivel de significancia.

EPÍLOGO

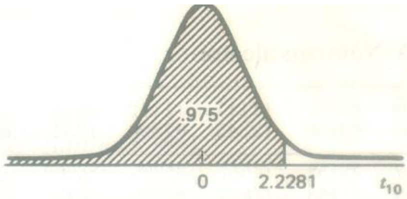
El presente trabajo de investigación texto universitario Análisis de Regresión y Correlación Lineal y no Lineal, tiene como finalidad poner al alcance de estudiantes de las diversas carreras profesionales de nuestra Universidad, un material bibliográfico, útil y práctico, que será utilizado en el transcurso de su formación profesional.

El contenido ha sido sistematizado con el propósito que estudiantes y lectores logren una mayor comprensión y aplicación de la Estadística, utilizando un vocabulario sencillo y preciso, reforzado con ejemplos y problemas prácticos facilitando el proceso de enseñanza aprendizaje.

El trabajo esta estructurado en cuatro capítulos. El primer capitulo, trata del análisis de regresión simple, el segundo capítulo se aborda el análisis de correlación simple, el tercer capítulo se refiere al análisis de regresión no lineal y el cuarto capítulo al análisis de correlación no lineal.

APÉNDICE

Tabla A. Percentil de la distribución t de estudent.



$P(t_{10} \leq 2.2281) = .975$

d. f.	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	3.078	6.3138	12.706	31.821	63.657
2	1.886	2.9200	4.3027	6.965	9.9248
3	1.638	2.3534	3.1825	4.541	5.8409
4	1.533	2.1318	2.7764	3.747	4.6041
5	1.476	2.0150	2.5706	3.365	4.0321
6	1.440	1.9432	2.4469	3.143	3.7074
7	1.415	1.8946	2.3646	2.998	3.4995
8	1.397	1.8595	2.3060	2.896	3.3554
9	1.383	1.8331	2.2622	2.821	3.2498
10	1.372	1.8125	2.2281	2.764	3.1693
11	1.363	1.7959	2.2010	2.718	3.1058
12	1.356	1.7823	2.1788	2.681	3.0545
13	1.350	1.7709	2.1604	2.650	3.0123
14	1.345	1.7613	2.1448	2.624	2.9768
15	1.341	1.7530	2.1315	2.602	2.9467
16	1.337	1.7459	2.1199	2.583	2.9208
17	1.333	1.7396	2.1098	2.567	2.8982
18	1.330	1.7341	2.1009	2.552	2.8784
19	1.328	1.7291	2.0930	2.539	2.8609
20	1.325	1.7247	2.0860	2.528	2.8453
21	1.323	1.7207	2.0796	2.518	2.8314
22	1.321	1.7171	2.0739	2.508	2.8188
23	1.319	1.7139	2.0687	2.500	2.8073
24	1.318	1.7109	2.0639	2.492	2.7969
25	1.316	1.7081	2.0595	2.485	2.7874
26	1.315	1.7056	2.0555	2.479	2.7787
27	1.314	1.7033	2.0518	2.473	2.7707
28	1.313	1.7011	2.0484	2.467	2.7633
29	1.311	1.6991	2.0452	2.462	2.7564
30	1.310	1.6973	2.0423	2.457	2.7500
35	1.3062	1.6896	2.0301	2.438	2.7239
40	1.3031	1.6839	2.0211	2.423	2.7045
45	1.3007	1.6794	2.0141	2.412	2.6896
50	1.2987	1.6759	2.0086	2.403	2.6778
60	1.2959	1.6707	2.0003	2.390	2.6603
70	1.2938	1.6669	1.9945	2.381	2.6480
80	1.2922	1.6641	1.9901	2.374	2.6388
90	1.2910	1.6620	1.9867	2.368	2.6316
100	1.2901	1.6602	1.9840	2.364	2.6260
120	1.2887	1.6577	1.9799	2.358	2.6175
140	1.2876	1.6558	1.9771	2.353	2.6114
160	1.2869	1.6545	1.9749	2.350	2.6070
180	1.2863	1.6534	1.9733	2.347	2.6035
200	1.2858	1.6525	1.9719	2.345	2.6006
∞	1.282	1.645	1.96	2.326	2.576

BIBLIOGRAFÍA

1. AVILA ACOSTA, ROBERTO. Estadística Elemental Lima: Estudios y Ediciones R. A. 2da Edición, 2001

2. CÓRDOVA ZAMORA, MANUEL. Estadística Inferencial Lima: Editorial MOSHERA S.R.L. 2da Edición, 2005.

3. ELORZA, HAROLDO. Estadística para Ciencias Sociales del Comportamiento y de la Salud. Lima: CENGASE 3ª Edición, 2008.

4. HEINZ, KOHLER. Estadística para Negocios y Economía. México: Compañía Editorial Continental S.A. 3era Edición, 2002.

5. LEVIN, RUBIN. Estadística para Administradores. México: Prentice Hall Hispanoamericana S.A. 3era Edición, 2007.

6. MASON LIND, MARCHAL. Estadística para Administración y Economía México: Alfa omega Grupo Editor S.A. 1ª Edición, 2007.

7. WEINER, Richard. Estadística. México: Compañía Editorial Continental S.A. 1ª Edición, 2007