

Estadística Gerencial

Doc. Juan Morales Romero

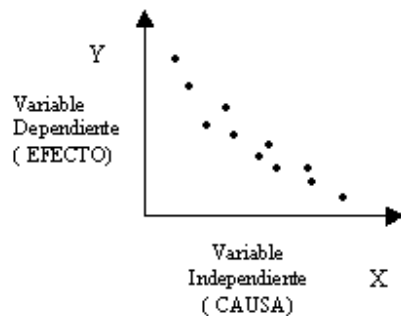
Análisis de Correlación y Regresión Lineal

ANALISIS DE CORRELACION

Conjunto de técnicas estadísticas empleadas para medir la intensidad de la asociación entre dos variables

DIAGRAMA DE DISPERSION

Gráfica que representa la relación entre dos variables denominada nube de puntos



VARIABLE DEPENDIENTE

Llamada también variable efecto o variable que se pretende explicar (variable explicada). La evolución de la variable dependiente (efecto) se explica por las fluctuaciones de la variable causa.

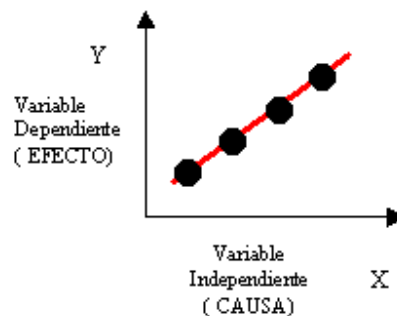
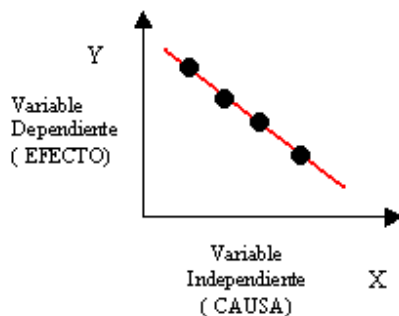
VARIABLE INDEPENDIENTE

Llamada también variable causa o variable explicadora (variable explicativa).

COEFICIENTE DE CORRELACION

Originado por el investigador Karl Pearson en 1900. Describe la intensidad de la relacion entre dos variables. Puede tomar cualquier valor de -1 a +1

- 1.0 = Perfectamente relacionados en sentido lineal negativo.
- +1.0 = Perfectamente relacionados en sentido lineal positivo
- 0 = No existe relacion alguna entre dos variables
- 0.9 o +0.9 = Correlación lineal muy intensa



CORRELACION NEGATIVA PERFECTA

CORRELACION POSITIVA PERFECTA

Estadística Gerencial

Doc. Juan Morales Romero

FORMULA DEL COEFICIENTE DE CORRELACION

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

Donde :

n = Numero de observaciones

SX = Suma de los valores de la variable X

SY = Suma de los valores de la variable Y

(SX²) = Suma de los valores de X elevados al cuadrado

(SX)² = Cuadrado de la suma de los valores de X

(SY²) = Suma de los valores de Y elevados al cuadrado

(SY)² = Cuadrado de la suma de los valores de Y

SX Y = Suma de los productos de X e Y

FORMA GENERAL DE LA ECUACION DE REGRESION LINEAL SIMPLE

$$Y = a + bX$$

Y = Valor pronosticado de la variable Y (variable dependiente)

X = Valor de la variable x (variable independiente)

a = Es la ordenada de la intersección en el eje Y. Es decir Y cuando X= 0

b = Es la pendiente de la recta es decir el cambio promedio en Y por unidad de cambio en la variable X.

FORMULA DE (a)

$$a = (\sum Y / n) - b(\sum X / n)$$

FORMULA DE (b)

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

FORMA GENERAL DE LA ECUACION DE REGRESION MULTIPLE

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

Y = Valor pronosticado de la variable Y (variable dependiente)

X₁ = Valor de la variable x 1 (variable independiente 1)

X₂ = Valor de la variable x 2 (variable independiente 2)

b₁ = Es el cambio en Y por unidad de cambio en la variable X1 manteniendo X2 constante

b₂ = Es el cambio en Y por unidad de cambio en la variable X2 manteniendo X1 constante

- En la ecuación de regresión lineal múltiple la variable efecto (dependiente) se encuentra explicada por múltiples causas (muchas variables independientes) o muchas variables explicativas.

Estadística Gerencial

Doc. Juan Morales Romero

ECUACION DE TENDENCIA

$$Y = a + b T$$

Calculo de una ecuación de regresión teniendo como variable independiente , causa o explicativa al tiempo. La ecuación de tiempo permite predecir o pronosticar por periodos cortos siempre que el coeficiente de determinación indique un alto grado de explicatividad.

Y = Valor pronosticado de la variable Y (variable dependiente)

T = Valor de la variable T (variable independiente)

a = Es la ordenada de la intersección en el eje Y. Es decir Y cuando T= 0

b = Es la pendiente de la recta es decir el cambio promedio en Y por unidad de cambio en la variable T

FORMULA DE (a)

$$a = (\sum Y / n) - b(\sum T / n)$$

FORMULA DE (b)

$$b = \frac{n(\sum TY) - (\sum T)(\sum Y)}{n(\sum T^2) - (\sum T)^2}$$

ECUACION DE TENDENCIA PARA REALIZAR PRONOSTICO (PROSPECCION) Y RETROSPECCION (PASADO)

El primer paso es establecer las variables en este caso las variables son dos : La Variable dependiente (efecto) la que se pronosticara) y la Variable independiente (causa) que es el tiempo a partir del cual se explicaría la evolución de la variable efecto .

EJEMPLO

En la tabla se presenta numero de los delitos de Estados Unidos en miles de personas de 1989 a 1995 se pide proyectar el numero de delitos para 1996 (prospección) y 1988 (retrospección).

AÑO	1989	1990	1991	1992	1993	1994	1995
NUMERO	660	671.4	688.0	695.5	717.1	759.2	807

Por ser una ecuación de tendencia la variable independiente- causa o explicativa es el tiempo y la variable dependiente - efecto o explicada es el numero de delitos. Codificando la variable tiempo asignamos cero para el primer año 1989 (t = 0), 1990 (t=1),1991 (t=2),1992 (t=3),1993 (t=4), 1994 (t=5) , 1995 (t=6) y **1997 (t=7) y 1988 (t = -1)**

CODIFICACION :

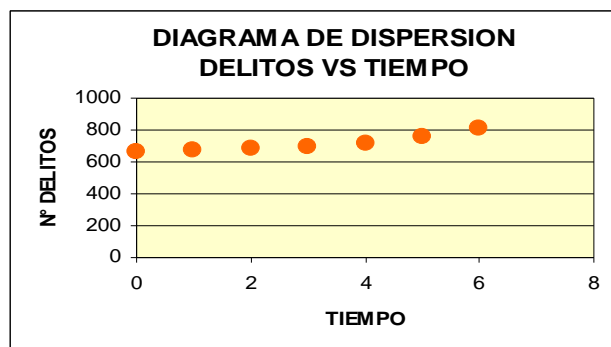
T	0	1	2	3	4	5	6
NUMERO	660	671.4	688.0	695.5	717.1	759.2	807

Estadística Gerencial

Doc. Juan Morales Romero

Construyendo diagrama de dispersión :

Uniendo intersecciones de la variable dependiente (delitos) con variable independiente (tiempo) es decir pares (X,Y)



Ahora calculamos los coeficientes necesarios para estimar la ecuación de regresión :

Variable Independiente X=	TIEMPO				
Variable Dependiente Y=	DELITOS				
AÑO	T	Y	T2	Y2	TY
1989	0	660	0	435600	0
1990	1	671.4	1	450778	671.4
1991	2	688	4	473344	1376
1992	3	695.5	9	483720	2086.5
1993	4	717.1	16	514232	2868.4
1994	5	759.2	25	576385	3796
1995	6	807	36	651249	4842
	21	4998.2	91	3585308	15640.3
	S(T)	S(Y)	S(T2)	S(Y2)	S(TY)
	441	24982003			
	(S(T))2	(S(Y))2			
N° de observaciones	7				

Coeficiente de Correlación y Determinación :

r =	0.951
R2	0.905

Reemplazando datos en las formulas

$$a = (\sum Y / n) - b(\sum T / n)$$

$$b = \frac{n(\sum TY) - (\sum T)(\sum Y)}{n(\sum T^2) - (\sum T)^2}$$

$$a = 644.846$$

$$b = 23.061$$

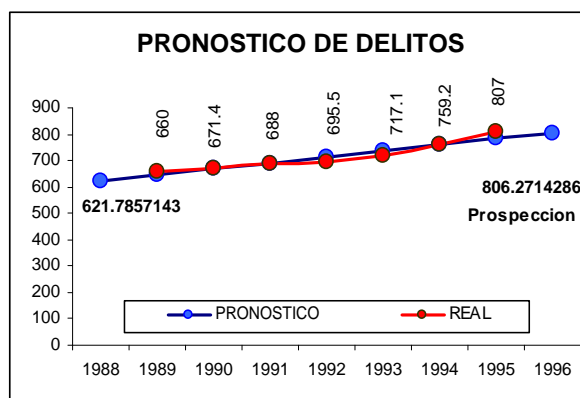
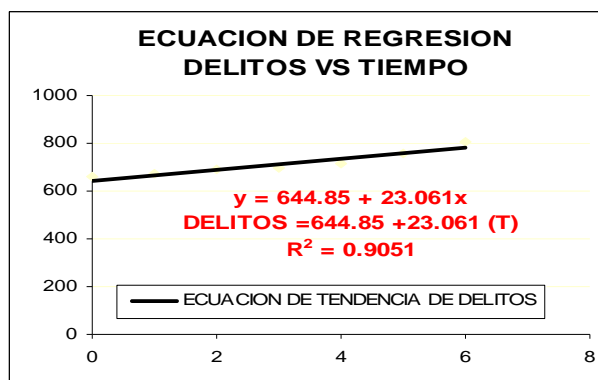
Entonces la ecuación de regresión es :

$$Y = a + b T$$

$$\text{DELITOS} = 644.846 + 23.061 * (T)$$

Estadística Gerencial

Doc. Juan Morales Romero



Para calcular el numero de delitos ocurridos en 1988 utilizamos ($t = -1$) dado que el primer periodo 1989 se ha codificado como ($t = 0$) Reemplazando ($t = -1$) en la Ecuación de Regresión para estimar el numero de delitos cometidos un año antes 1988.

$$\text{DELITOS} = 644.846 + 23.061 * (T)$$

- 1988 ($t = -1$)
 $\text{DELITOS} = 644.846 + 23.061 * (-1) = 621.78$
- 1989 ($t = 0$)
 $\text{DELITOS} = 644.846 + 23.061 * (-0) = 644.846$
- 1990 ($t = 1$)
 $\text{DELITOS} = 644.846 + 23.061 * (1) = 667.907$

Y asi sucesivamentehasta

- 1996 ($t = 7$)
 $\text{DELITOS} = 644.846 + 23.061 * (7) = 806.27$

FINALMENTE ELABORANDO UNA TABLA

CODIFICACION TIEMPO	AÑO	PRONOSTICO DELITOS = 644.85 + 23.061 (T)	REAL HISTORICO
-1	1988	621.7857143	
0	1989	644.8464286	660
1	1990	667.9071429	671.4
2	1991	690.9678571	688
3	1992	714.0285714	695.5
4	1993	737.0892857	717.1
5	1994	760.15	759.2
6	1995	783.2107143	807
7	1996	806.2714286	

Doc. Juan Morales Romero

CASO APLICATIVO DE PRONOSTICO CON INDICES DE CRIMINALIDAD EN EL PERU

I. POBLACIÓN PENAL

8.26 POBLACIÓN RECLUIDA EN ESTABLECIMIENTOS PENITENCIARIOS, SEGÚN DEPARTAMENTOS, 1991 - 2002

Departamento	1998	1999	2000	2001	2002
Total	26 059	27 400	27 734	26 989	28 135
Amazonas	330	399	459	473	467
Ancash	774	817	836	784	783
Apurimac	169	182	185	174	210

Fuente: Instituto Nacional Penitenciario - Oficina de Estadística.

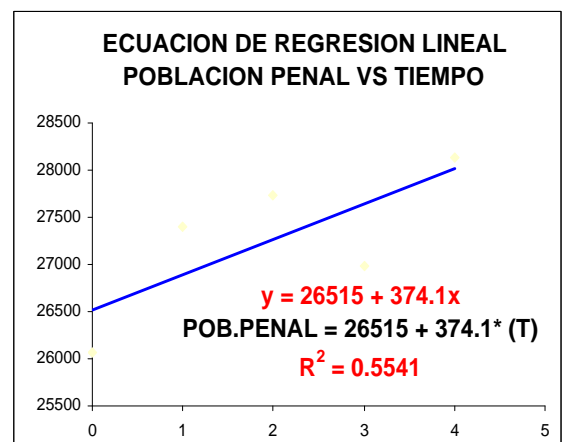
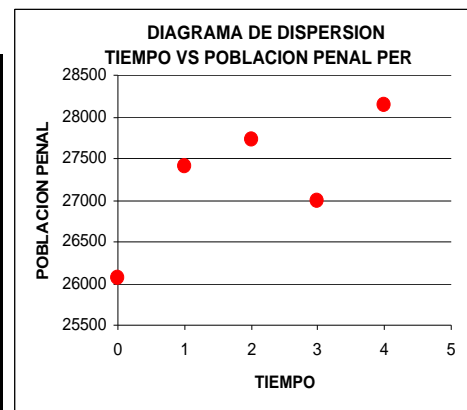
Con la siguiente información recopilada del Compendio Estadístico del INEI 2004 se le pide estimar utilizando una ecuación de tendencia la población reducida en los centros penitenciarios del año 1997 (retrospección) y la población penal de los años 2003, 2004, 2005, 2006 y 2007 (prospección o proyección)

Variable Independiente X=	TIEMPO					
	POBLACION PENAL					
	T		Y	T2	Y2	TY
	1998	0	26059	0	679071481	0
	1999	1	27400	1	750760000	27400
	2000	2	27734	4	769174756	55468
	2001	3	26989	9	728406121	80967
	2002	4	28135	16	791578225	112540
		10	136317	30	3718990583	276375
	S(X)	S(Y)	S(X2)	S(Y2)	S(XY)	
	100	18582324489				
	(S(X))2	(S(Y))2				
N° de observaciones	5					
	18705		25128.0978			

	NUMERADOR	DENOMINADOR
r =	0.744	
R2	0.554	

Ecuación de la recta de regresión				
Y=	a	(+)	b	X
Y=	26515.2	(+)	374.1	X

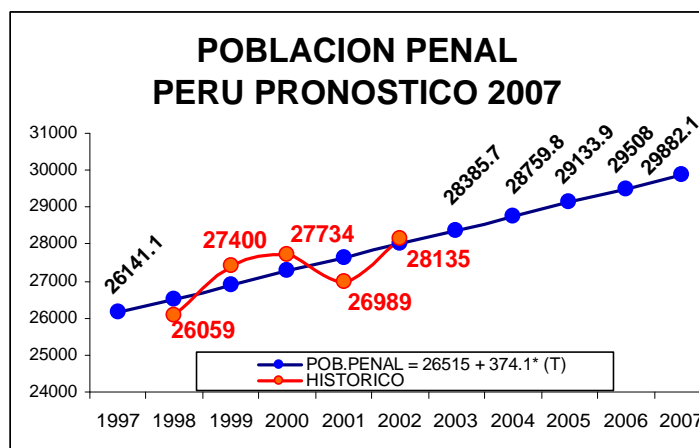
SI X=	
PRONOSTICO	



Estadística Gerencial

Doc. Juan Morales Romero

AÑO CODIFICADO	AÑO	PRONOSTICO $POB.PENAL = 26515 + 374.1 * (T)$	REAL HISTORICO
-1	1997	26141.1	
0	1998	26515.2	26059
1	1999	26889.3	27400
2	2000	27263.4	27734
3	2001	27637.5	26989
4	2002	28011.6	28135
5	2003	28385.7	
6	2004	28759.8	
7	2005	29133.9	
8	2006	29508	
9	2007	29882.1	



CALCULANDO COEFICIENTE DE CORRELACION CON EXCEL

1. Ingresar a Microsoft Excell
2. Digita tu data

AÑO	POBLACION PENAL
X	Y
1998	26059
1999	27400
2000	27734
2001	26989
2002	28135

3. Elige el Menú **Herramientas - Análisis de datos - Coeficiente de Correlación**

TIEMPO

POBLACION PENAL

T	Y
0	26059
1	27400
2	27734
3	26989
4	28135
10	136317

Coeficiente de correlación

Entrada

Rango de entrada:

Agrupado por: ☒ Columnas ☐ Filas

☐ Rótulos en la primera fila

Opciones de salida

☐ Rango de salida:

☒ En una hoja nueva:

☐ En un libro nuevo

Aceptar Cancelar Ayuda

Luego click en aceptar

	Columna 1	
Columna 1	1	
Columna 2	0.74438583	1

Doc. Juan Morales Romero

CALCULANDO LA ECUACION DE REGRESION CON EXCELL

1. Ingresar a Microsoft Excel
2. Digita tu data

AÑO	POBLACION PENAL
X	Y
1998	26059
1999	27400
2000	27734
2001	26989
2003	28135

- ### 3. Elige el Menú **Herramientas - Análisis de datos - Regresión**

B	C	D
Independiente X=	TIEMPO	
Dependiente Y=	POBLACION PENAL	
	T	Y
	0	26059
	1	27400
	2	27734
	3	26989
	4	28135
	10	136317
	S(X)	S(Y)
	100	18582324489
	(S(X))2	(S(Y))2
oservaciones	5	
	18705	25128.09782
	NUMERADOR	DENOMINADOR

Regresión

Entrada

Rango Y de entrada:

Rango X de entrada:

☐ Rótulos ☐ Constante igual a cero

☐ Nivel de confianza %

Opciones de salida

☐ Rango de salida:

☒ En una hoja nueva:

☐ En un libro nuevo

Residuales

☐ Residuos ☐ Gráfico de residuales

☐ Residuos estándares ☒ Curva de regresión ajustada

Probabilidad normal

☐ Gráfico de probabilidad normal

Aceptar

Cancelar

Ayuda

Luego click en botón aceptar para obtener resumen de estadísticas de la ecuación de regresión

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.744385832
Coefficiente de determinación R^2	0.554110267
R^2 ajustado	0.405480356
Error típico	612.692718
Observaciones	5
ANÁLISIS DE VARIANZA	
	<i>Grados de libertad</i>
Regresión	1
Residuos	3
Total	4
Coeficientes	
Intercepción	26515.2
Variable X 1	374.1

This document was created with Win2PDF available at <http://www.win2pdf.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.
This page will not be added after purchasing Win2PDF.