

## Guía Tema IV

### Ejercicios regresión lineal simple.

1) La siguiente tabla proporciona cifras de las tasas de renuncias por cada 1000 empleados y las tasas de desempleo en cierta industria de los E. Unidos para el período 2000-2012

Año	Tasa Renuncias (Y) (# x cada mil empleados)	Tasa de desempleo (X) (% pob. en edad laboral)
2000	1.3	6.2
2001	1.2	7.8
2002	1.4	5.8
2003	1.4	5.7
2004	1.5	5.0
2005	1.9	4.0
2006	2.6	3.2
2007	2.3	3.6
2008	2.5	3.3
2009	2.7	3.3
2010	2.1	5.6
2011	1.8	6.8
2012	2.2	5.6

a) Dibuje el diagrama de dispersión.

b) Suponiendo que la tasa de renuncias Y está relacionada linealmente con la tasa de desempleo X a través del modelo  $Y = \beta_1 + \beta_2 X + U$ , halle la ecuación de regresión estimada.

c) Realice una d'cima para validar el modelo. ( $\alpha = 0,01$ )

d) Pronostique la Tasa de Renuncia para una Tasa de desempleo del 4,5%.

2) Los siguientes datos corresponden a los por cientos de alfabetización e infectados por el SIDA en 8 países y se desea establecer un modelo de regresión lineal que permita estimar la incidencia media de SIDA para una tasa dada de alfabetización.

Alfabetización	12	32	34	20	56	60	44	72
SIDA	5	4,1	3,8	4,2	3	2,2	3	1,5

a) Calcule la ecuación de regresión estimada

b) Estime la infección por SIDA media para un país con un 25 por ciento de alfabetización

c) Efectúe la D'cima F simple para validar el modelo ( $\alpha = 0,05$ ) y calcule el coeficiente de Determinación.

3) Dadas las variables "x" y "y" las que se suponen están relacionadas linealmente:

a) Halle la ecuación de regresión estimada.

b) Estime el valor de "y" para cuando  $x = 12$

c) Halle e interprete el valor de  $R^2$  y verifique, para un  $\alpha = 0.05$ , el supuesto de relación lineal entre las variables seleccionadas.

d) Diga cuál es la expresión de los estimadores mínimo cuadráticos de  $\beta_0$  y  $\beta_1$ , si el modelo fuera:  $y = \beta_0 + \beta_1 x^3 + u$

$X_i$	$Y_i$
1	2.5
2	3.0
3	3.4
4	4.0
5	4.5
6	5.1
7	6.0
8	6.4
9	7
10	7.5

4) Durante los primeros 9 meses del año 2020 una pequeña empresa productora de calzado, demostró una mejor producción (miles de pesos) como queda demostrado en los siguientes valores observados:

Se pide:

- Realice el diagrama de dispersión e interprete.
- Halle el  $R^2$  e interprete.
- Halle la ecuación de regresión estimada más adecuada para los datos obtenidos.
- Valide el modelo seleccionado a través de la Dócima F simple ( $\alpha = 0,05$ ).

Meses	Costo (miles) Y	Ing.(m) X
Enero	1	5
Febrero	3	6
Marzo	1	8
Abril	3	7
Mayo	2	10
Junio	4	11
Julio	3	12
Agosto	2	9
Septiembre	1	6

5) Con un conjunto de 100 observaciones tomadas sobre las variables: y = producción de acero(en TM) y x = consumo de combustible(en TM). Se hicieron dos ajustes: uno con el modelo  $y = \beta_0 + \beta_1 x + U$  y otro mediante el modelo  $y = \beta_0 + \beta_1 x^4 + U$  obteniéndose las siguientes ecuaciones de predicción:

$\hat{y} = 2 + 13x$ ;  $\hat{y} = 1.5 + 0.1x^4$ . Si para el primero de los ajustes, se obtuvieron los siguientes datos:

**Suma de cuadrados:** S.C.RESIDUAL = 200; S.C.TOTAL = 800

Y para el segundo:  $\sum(y_i - \hat{y})^2 = 50$   $\sum(\hat{y}_i - \bar{y})^2 = 750$

- Halle el coeficiente de determinación para ambos ajustes e interprete los mismos.
- Selecciones de los dos modelos, el que ofrece mejor ajuste.
- Construya una prueba **F** para el modelo seleccionado en (b) y comente sobre la validez del ajuste hecho ( $\alpha = 0.01$ )

6) Los siguientes datos representan las ganancias (en miles), por concepto de consumo, de cierto parque de atracciones turísticas, se supone que las mismas dependen en gran medida del número de ofertas recreativas que se ofrecen. Para ello se tomaron 12 observaciones.

Ofertas.	21	24	28	16	18	19	20	29	34	35	15	32
Ganancias.	345	356	389	213	255	310	300	398	400	395	120	400

- Dibuje el Diagrama de Dispersión.
- Determine a través del coeficiente de correlación lineal que modelo lineal es el más adecuado entre los siguientes:

$$y = \beta_1 + \beta_2 X \quad y = \beta_1 + \beta_2 \ln X \quad y = \beta_1 + \beta_2 \frac{1}{X}$$

- Estime los parámetros del modelo seleccionado.

## Algunos resultados para comprobar ejercicios 1-6:

### **Ej1:**

#### **Regresión Simple - Y vs. X**

Variable dependiente: Y (Tasa de renuncia) Variable independiente: X (Tasa de desempleo) Lineal:  $Y = a + b \cdot X$

#### **Coefficientes**

	<i>Mínimos Cuadrados</i>	<i>Estándar</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimado</i>	<i>Error</i>	<i>T</i>	<i>Valor-P</i>
Intercepto	3.36626	0.331084	10.1674	0.0000
Pendiente	-0.286212	0.062885	-4.55135	0.0008

### **Ej2:**

Variable dependiente: Y1 (SIDA) Variable independiente: X1 (Alfabetización) Lineal:  $Y = a + b \cdot X$

#### **Coefficientes**

	<i>Mínimos Cuadrados</i>	<i>Estándar</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimado</i>	<i>Error</i>	<i>T</i>	<i>Valor-P</i>
Intercepto	5.58939	0.229928	24.3093	0.0000
Pendiente	-0.0542881	0.00505062	-10.7488	0.0000

### **Ej3:**

valores de las sumas

$\sum x$	55
$\sum y$	49.4
$\sum x^2$	385
$\sum y^2$	271.28
$\sum xy$	319
n	10

Coefficiente de Determinación	
$R^2_{xy}$	0.9954

Ecuación de Regresión:

Pendiente	0.57333333
Intercepto	1.78666667

Dócima F Simple

FV	SC	gl	CM	F	F(Tabla)
Total	27.24				
Regresión	27.12	1	27.12	1730.98	5.32
Residuos	0.12	8	0.016		

#### **Regresión Simple - Y vs. $X^3$**

Coefficiente correlación	
$R_{x^3y}$	0.93503748

Coefficiente de Determinación		
$R^2_{x^3y}$	0.87429509	

Entre los valores de las sumas

$\sum x^3$	3025
$\sum y$	49.4
$\sum x^6$	1978405
$\sum y^2$	271.28
$\sum x^3y$	19976.2
n	10

Dócima F Simple

FV	SC	gl	CM	F	F(Tabla)
Total	27.244				
Regresión	23.8192956	1	23.8192956	55.6411124	5.31765506
Residuos	3.42470444	8	0.42808805		

Ecuación de Regresión:

Pendiente	0.00473291
Intercepto	3.50829596

#### Ej4:

valores de las sumas

$\sum x$	74
$\sum y$	20
$\sum x^2$	656
$\sum y^2$	54
$\sum xy$	176
n	9

Coefficiente de Determinación	
$R^2_{xy}$	0.29384916

Ecuación de Regresión:

Pendiente	0.24299065
Intercepto	0.22429907

#### Ej5:

N=100

x = consumo de combustible(en TM).

y = producción de acero(en TM)

Modelo I:  $y = \beta_0 + \beta_1 x + U$  --  $\hat{y} = 2 + 13x$  -- S.C.RESIDUAL = 200; S.C.TOTAL = 800

Modelo II:  $y = \beta_0 + \beta_1 x^4 + U$  --  $\hat{y} = 1.5 + 0.1x^4$  --  $\sum (y_i - \hat{y})^2 = 50$   $\sum (\hat{y}_i - \bar{y})^2 = 750$

a)

$$R^2_{MI} = \frac{SC\ Reg}{SCT} = \frac{600}{800} = 0.75 \quad SCReg = SCT - SCRes = 800 - 200 = 600$$

$$R^2_{MII} = \frac{SC\ Reg}{SCT} = \frac{750}{800} = 0.9375 \quad SCT = SCReg + SCRes = 750 + 50 = 800$$

b) El modelo de mejor ajuste es el 2.

c) Prueba F

FV	SC	gl	CM	F	F(Tabla)
Total	800				
Regresión	750	1	750	1470.59	6.9
Residuos	50	98	0.51		

$\alpha =$	0.01
------------	------

Decisión: Rechazo  $H_0$

#### Ej6:

Coefficiente correlación	Modelo y-x
r	0.86450484
$\sum x$	291
$\sum y$	3881
$\sum x^2$	7613
$\sum y^2$	1342405
$\sum xy$	100136
n	12

Coefficiente correlación	Modelo y-lnx
r	0.9097839

$\sum (1/x)$	0.53606975
$\sum y$	3881
$\sum (1/x)^2$	0.02583339
$\sum y^2$	1342405
$\sum (1/x)y$	161.224276
n	12

Coefficiente correlación	Modelo y-1/x
r	-0.9473078

$\sum \ln x$	37.780703
$\sum y$	3881
$\sum (\ln x)^2$	119.917621
$\sum y^2$	1342405
$\sum \ln xy$	12483.4277
n	12

**Ejercicios sobre regresión lineal simple y múltiple. (De los ejercicios 1, 2 y 3 se brindan al final un conjunto de información adicional, que los ayudará a resolver estos problemas)**

- 1- La gerencia de Hop Scotch Airlines, la aereolínea transportadora más pequeña del mundo, considera que existe una relación directa entre los gastos publicitarios y el número de pasajeros que escogen viajar por Hop Scotch. Para determinar si esta relación existe, y si es así cual podría ser la naturaleza exacta se recolectaron los valores mensuales por gastos de publicidad y número de pasajeros para los n=15 meses más recientes. Los datos aparecen a continuación representando a los pasajeros con la variables Y, ya que se asume que dependen de la publicidad.

Obs. (Mes)	Publicidad (X-Miles)	Pasajeros (Y-Miles)
1	10	15
2	12	17
3	8	13
4	17	23
5	10	16
6	15	21
7	10	14
8	14	20
9	19	24
10	10	17
11	11	16
12	13	18
13	16	23
14	10	15
15	12	16

- a) Realice la validación del modelo ( $\alpha=0,05$  y  $\alpha=0,01$ ) y su estimación.  
b) Pronostique el número de pasajero que se espera para un gasto publicitario de 15500 y 12750 respectivamente.

- 2- Un director administrativo está intentando desarrollar un sistema diseñado para identificar qué atributos personales son esenciales para avanzar gerencialmente. A quince empleados que han sido ascendidos recientemente se les practica una serie de pruebas para determinar ( $X_1$ ) sus habilidades comunicativas; ( $X_2$ ) la habilidad para relacionarse con otros y ( $X_3$ ) la habilidad para tomar decisiones. A la clasificación de cada empleado (Y) se le hace regresión sobre estas tres variables. Los datos originales son los siguientes:

Y	$X_1$	$X_2$	$X_3$	$E_i$
80	50	72	18	9.8594
75	51	74	18	2.1899
84	42	79	22	3.5386
62	42	71	17	-5.0361
92	59	85	25	-0.5293
75	45	73	17	5.0004
63	48	75	16	-9.2960
69	39	73	19	-1.4517
68	40	71	20	-0.7434
87	55	80	30	-1.9707
92	48	83	33	-1.7266
82	45	80	20	1.1705
74	45	75	18	0.8108
80	61	75	20	3.1243
62	59	70	15	-4.9400

- a) Evalúe la significación del modelo completo.  
b) Desarrolle los métodos paso a paso y teniendo en cuenta los coeficientes de determinación y compare si por ambos se obtiene el mismo resultado.

- 3- Grafique los siguientes datos. Compare un modelo lineal con la forma cuadrática. Valide el modelo seleccionado y con el mismo pronostique Y para  $X=22$ .

Y	731	730	815	1291	1408	2170	2312	2768	2877	2929	7641
X	18	18	19	24	25	31	32	35	36	36	48

4- Una empresa necesita estimar la función de demanda de su producto principal y para ello recogió una serie de datos anuales en el período 2000-2020. Los datos recogidos corresponden al consumo del producto en decenas de miles de toneladas (Y), precio del producto por Kg ( $X_1$ ) y precio de un producto alternativo por Kg ( $X_2$ ).

- Diga cuál de las variables independientes posee una mayor correlación con Y.
- Realice una dística F simple para la variable seleccionada en el inciso anterior e interprete.
- A partir del resultado obtenido en el inciso anterior realice una prueba que permita analizar si en un modelo de regresión debe incluirse las dos variables e interprete.
- Estime el modelo de regresión que considere más adecuado. ¿Qué porcentaje de la variabilidad de Y no viene explicada por el modelo de regresión seleccionado?
- Estime el consumo esperado del producto para un año donde el precio del producto por Kg sea igual a 120 \$ y el precio del producto alternativo sea de 130 \$.

Datos Adicionales:

$$\begin{aligned} \sum Y &= 952,4 & \sum Y^2 &= 44529,92 & \sum X_1 &= 2942,59 & \sum X_1^2 &= 465757,43 \\ \sum X_2 &= 2622 & \sum X_2^2 &= 336606,7 & \sum YX_1 &= 131063,06 & \sum YX_2 &= 121101,98 \end{aligned}$$

$$SC_{Error X_1} = 1229,35 \quad SC_{Regresión X_1 X_2} = 704,165 \quad R^2_{yx_2} = 0,388$$

$$\text{Para las dos variables independientes: } b_1 = 21,62 \quad b_2 = -0,0595 \quad b_3 = 0,257$$

### Datos adicionales:

#### Ejercicio 1: (Datos y valores)

X	Y	X <sup>2</sup>	Y <sup>2</sup>	XY	Yestimada
10	15	100	225	150	15.1994
12	17	144	289	204	17.3621
8	13	64	169	104	13.0368
17	23	289	529	391	22.7686
10	16	100	256	160	15.1994
15	21	225	441	315	20.6060
10	14	100	196	140	15.1994
14	20	196	400	280	19.5247
19	24	361	576	456	24.9313
10	17	100	289	170	15.1994
11	16	121	256	176	16.2807
13	18	169	324	234	18.4434
16	23	256	529	368	21.6873
10	15	100	225	150	15.1994
12	16	144	256	192	17.3621

Sumas

$\sum x$	187
$\sum y$	268
$\sum x^2$	2469
$\sum y^2$	4960
$\sum xy$	3490
n	15

Dística F Simple

FV	SC
Total	171.733333
Regresión	161.044079
Residuos	10.6892546

Coficiente correlación	
$R_{xy}$	0.9684
Coficiente de Determinación	
$R^2_{xy}$	0.9378

Ejercicio 2. Datos y valores

<i>Parámetro</i>	<i>Estimación</i>
CONSTANTE	-40.0258
X1	0.147018
X2	1.26122
X3	0.66711

**Análisis de Varianza**

<i>Fuente</i>	<i>Suma de Cuadrados</i>
Modelo	1146.03
Residuo	297.305
Total (Corr.)	1443.33

R-cuadrada = 79.4015 porciento

R-cuadrado (ajustado para g.l.) = 73.7837 porciento

Ejercicio 3 (Datos).

Valores de las sumas

$\sum x$	322
$\sum y$	25672
$\sum x^2$	10316
$\sum y^2$	98337750
$\sum xy$	921885
n	11

Ecuación de Regresión:

Pendiente	191.416565
Intercepto	-3269.46671

Entre los valores de las sumas

$\sum x^2$	10316
$\sum y$	25672
$\sum x^4$	13203044
$\sum y^2$	98337750
$\sum x^2 y$	35364293
n	11

Otros estadígrafos

$\sum (e_i - e_{i-1})^2$	2412312.7
$\sum (e_i)^2$	2308729.19
d(dur-Wat)	1.04486603
$\chi^2$	3.6000

5- En el pasado muchos economistas han estudiado los patrones de gasto de los consumidores en la economía. Un estudio famoso realizado por Milton Friedman concluye que el consumo es una función de *ingreso permanente*, que se define como el nivel de ingreso promedio que el consumidor espera recibir en el futuro. La teoría de *hábito-persistencia* de T.M. Brown argumenta que el consumo es moldeado por el ingreso pico más reciente de un consumidor, el ingreso más alto recibido en el pasado inmediato. Para combinar estas dos teorías, un economista recolectó datos sobre el *consumo* ( $C$ ), *ingreso permanente* ( $IPE$ ), e *ingreso pico* ( $IPI$ ) entre 20 consumidores seleccionados al azar en cierta región. A partir del análisis se presentaron los siguientes posibles modelos (*todos los análisis están en miles de pesos*):

Modelo I:  $\hat{C} = -2.8612 + 0.9527(IPE)$

Modelo II:  $\hat{C} = -4.2082 + 0.8769(IPI)$

Modelo III:  $\hat{C} = -3.008 + 0.8965(IPE) + 0.054(IPI)$

Modelo I:  $\sum C = 442$   $\sum C^2 = 10510$   $R^2 = 0.973$  Modelo II:  $SCE = 48.14$

Modelo III:  $SC Reg = 721.894$

- Diga el modelo que considera más adecuado a través del método paso a paso. En caso de excluir alguna variable, explique en su opinión, qué consideraría que pueda ocurrir.
- A partir del modelo seleccionado estime el consumo esperado para un consumidor cuyo ingreso permanente sea de \$30000 y el ingreso pico de \$ 40000.
- Si dos consumidores recibieron el mismo ingreso pico, pero uno ingresa dos mil pesos menos que el otro, ¿en cuánto difiere el consumo esperado entre ellos?

Nota: Para todas las pruebas utilice  $\alpha=0.01$

6- Un director administrativo está intentando desarrollar un sistema diseñado para identificar qué atributos personales son esenciales para avanzar gerencialmente. A quince empleados que han sido ascendidos recientemente se les practica una serie de pruebas para determinar: *sus habilidades comunicativas* ( $X_1$ ); *la habilidad para relacionarse con otros* ( $X_2$ ) y *la habilidad para tomar decisiones* ( $X_3$ ). A la clasificación de cada empleado ( $Y$ ) se le quiere hacer regresión sobre estas tres variables.

Modelo	Datos	Modelo	Datos
$\hat{Y} = 55.37 + 0.43X_1$	$R^2 = 0.0918$	$\hat{Y} = -70.12 + 0.13X_1 + 1.85X_2$	$SCE = 359.27$
$\hat{Y} = -67.6 + 1.9X_2$	$r_{yx_2} = 0.862$	$\hat{Y} = 32.69 + 0.24X_1 + 1.55X_3$	$SCE = 466.7$
$\hat{Y} = 43.18 + 1.6X_3$	$SC Reg = 936.74$	$\hat{Y} = -37.4 + 1.32X_2 + 0.67X_3$	$SCE = 308.6$

Modelo Completo:

$\sum y = 1145$   $\sum y^2 = 88845$   $\hat{Y} = -39.6 + 0.14X_1 + 1.25X_2 + 0.68X_3$   $SC Reg = 1148.58$

- Demuestre que alguna de las variables es significativa con la d́cima correspondiente.
- Seleccione el modelo que considere más adecuado a partir de  $R^2$  y  $\bar{R}^2$ .
- A partir del modelo seleccionado, ¿qué clasificación se espera tenga un empleado que obtuvo 60 en habilidades comunicativas, 85 en habilidad para relacionarse con otros y 20 en habilidad para tomar decisiones?.
- Si dos empleados son evaluados con la misma puntuación en  $X_1$  y  $X_2$  pero tienen 5 puntos de diferencia en  $X_3$ , ¿en cuánto debe diferir la clasificación esperado entre ellos?

Nota: Para todas las pruebas utilice  $\alpha=0.05$

7- Se han aplicado los modelos lineal y log-lineal para relacionar la ganancia neta mensual Y (miles de pesos) de una empresa con el costo de materias primas por peso producido  $X_1$  (pesos) y la fuerza de trabajo empleada  $X_2$  (miles de horas de trabajo), obteniéndose los siguientes resultados en base a los datos de 14 meses.

Modelo	$R^2$	$\bar{R}^2$	F	Prob.
Lineal	81,45	78,08	24,152	0.00036
Log-lineal	87,50	85,23	38,5	0.00001

- Diga cuál de los dos modelos elegiría para representar la relación entre estas variables. Argumente.
- Valide el modelo seleccionado con la dícima correspondiente ( $\alpha = 0,05$ )

8- Cierta ministerio está evaluando distintos proyectos de inversión para los cuáles intenta determinar el *costo de inversión* a partir de los indicadores  $X_1$ : *costo de montaje* y  $X_2$ : *costo de construcción*, expresados en miles de pesos. Al recopilar los datos de estas variables en 30 proyectos anteriores de tipo similar a los que se consideran se obtuvieron los siguientes resultados al ajustar los distintos modelos posibles:

Modelo	$b_1$	$b_2$	$b_3$	SCT	SCE	SCReg
$Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + U$	9,88	0,35	0,27		7,32	
$Y = \beta_1 + \beta_2 X_1 + U$	10,2	0,24		57,73		33,2
$Y = \beta_1 + \beta_3 X_2 + U$	9,23		0,19			30,6

- Diga si podemos considerar que alguna de las variables independientes son significativas en el modelo. Use  $\alpha = 0,05$ .
- Calcule el Coeficiente de Determinación y el coeficiente de Determinación Ajustado para cada modelo y a partir de los mismos determine el modelo más adecuado.
- Determine el mejor modelo a través del método paso a paso. Compare con el modelo anterior.
- Estime el costo de inversión de un proyecto con un costo de montaje igual a 30 mil pesos y un costo de construcción de 20 mil.

10- Con base en información anual de los años 2001-2020 se obtuvieron los siguientes resultados de regresión:

$$\hat{Y}_t = -859.92 + 0.647X_{1t} - 23.195X_{2t} \quad R^2 = 0.9776 \quad SCT = 95728.46$$

$$\hat{Y}_t = -261.09 + 0.2452X_{1t} \quad R^2 = 0.9388$$

Donde Y= gasto de cierto país en bienes importados, miles de millones de dólares,  $X_1$  = ingreso personal disponible, miles de millones de dólares y  $X_2$ = variable de tendencia.

- Aplice una dícima F total al modelo múltiple y verifique que algunas de las variables independientes está relacionada con Y.
- Aplice el método paso a paso para seleccionar la mejor ecuación de regresión.
- A través de los coeficientes de Determinación y Determinación Ajustado decida el mejor modelo y compare si coincide con el resultado del inciso anterior.

Para todas las pruebas utilice  $\alpha = 0,05$ .