Guía de Ejercitación 8. Modelo de regresión lineal simple

Conjunto de n mediciones $\{(x_1, y_1);\ (x_2, y_2); \cdots;\ (x_n, y_n)\}$.

Modelo de Regresión Lineal Simple	$Y = \alpha X + \beta + \varepsilon$

X: variable aleatoria independiente o explicativa.
Y: variable aleatoria dependiente o explicada.
ε: variable aleatoria correspondiente a la perturbación; $E[\varepsilon] = 0, V(\varepsilon) = \sigma^2$.
α, β: parámetros del modelo.

$E[Y/X] = \hat{Y} = aX + b$	$a = \hat{\alpha}$	$b = \hat{\beta}$

Residuo i-ésimo $e_i = y_i - \hat{y}_i, i = 1, 2, \cdots, n$.

Por cuadrados mínimos:

$\hat{\alpha} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$	$\hat{\beta} = \bar{y} - \hat{\alpha}\bar{x}$	$\widehat{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}$

Fórmulas de cálculo

$S_{xy} = \left(\sum_{i=1}^{n} x_i y_i\right) - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$	$S_{xx} = \left(\sum_{i=1}^{n} x_i^2\right) - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$

Coeficiente de determinación r^2	$r^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Coeficiente de correlación muestral	$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}}$

Ejercicio 1. El artículo "Promising Quantitative Nondestructive Evaluation Techniques for Composite Materials" (*Materials Evaluation*, 1985, pp. 561-565) reporta sobre un estudio para investigar la forma como la propagación de una onda de esfuerzo ultrasónico que pasa por una sustancia depende de las propiedades de ésta. Los datos siguientes sobre resistencia a la fractura (x, como porcentaje de resistencia final a la tracción) y atenuación (y, en neper/cm, la disminución en amplitud de la onda de esfuerzo) en materiales compuestos de fibra de vidrio y poliéster reforzado, se obtuvieron de un gráfico presentado en ese artículo.

x	12	30	36	40	45	57	62	67	71	78	93	94	100	105
y	3,3	3,2	3,4	3,0	2,8	2,9	2,7	2,6	2,5	2,6	2,2	2,0	2,3	2,1

a) Representar el diagrama de dispersión de los datos.
b) Hallar la ecuación de la recta de regresión lineal para los datos muestreados.
c) Interpretar la pendiente de la recta en términos del problema.
d) Calcular el coeficiente de determinación. ¿Sugiere este valor que el modelo de regresión simple describe de manera eficiente la relación entre las dos variables? Explicar.
e) Estimar, a partir del modelo de regresión lineal calculado, la reducción en la amplitud de onda de esfuerzo para una resistencia a la fractura del 65%.

f) Si la resistencia a la fractura fuese el doble de la del punto anterior, ¿se considera razonable utilizar el modelo anterior? Justificar.

Rta. Ej 1. b) $Y = 3{,}6209 - 0{,}0147X$ **d)** $r^2 = 0{,}8974$. Sí.

Ejercicio 2. Es difícil determinar la resistencia al corte de puntos de soldaduras, mientras que es relativamente sencillo medir el diámetro de puntos de soldadura. Sería ventajoso si se pudiera predecir la resistencia al corte a partir de una medición del diámetro de soldadura. Los datos son:

Diámetro de soldadura (0,00001 pulgadas)	400	800	1250	1600	2000	2500	3100	3600	4000	4000
Resistencia al corte (psi)	370	780	1210	1560	1980	2450	3070	3550	3940	3950

a) Representar el diagrama de dispersión y, previo a ningún cálculo, discutir si es razonable aplicar el modelo de regresión lineal y, en caso afirmativo, visualizar a partir del gráfico si se espera que la pendiente de la recta predictora será unitaria y si tendrá ordenada al origen nula.
b) Hallar la ecuación de la recta de regresión y dibujarla sobre el diagrama de dispersión.
c) Estimar el valor esperado de resistencia al corte cuando el diámetro de soldadura es de 0,2300 pulgadas.
d) Hallar el coeficiente de determinación. Interpretar.

Rta. Ej 2. b) $Y = -22{,}2142 + 0{,}9928\, X$. **c)** $Y = 2261\, psi$. **d)** $r^2 = 0{,}9999$

Ejercicio 3. Un Jefe de Producto desea convencer a la Gerencia de que, por cada dólar que se invierte en publicidad, pueden esperar un aumento de 3 dólares al menos en las ventas. A tal efecto, se registraron los datos Publicidad y Ventas (ambos en miles de dólares) para los últimos de 5 meses.

Publicidad (miles de $u\$s$)	5	7,2	6,3	4,2	5,4	6,8
Ventas (miles de $u\$s$)	72	79	73	59	67	69

a) ¿Consideras que los datos son consistentes con la afirmación del Jefe de Producto? Justificar con un planteo y cálculos adecuados.
b) Interpretar la pendiente de la recta de regresión lineal asociada a los datos en términos del problema.
c) Teniendo en cuenta el concepto de extrapolación, ¿tiene sentido darle una interpretación a la ordenada al origen en este problema?
d) Estimar las ventas promedio para un mes en que se invirtieron en publicidad $u\$s\ 6000$.

Rta. Ej 3. **a)** Sí. **b)** Por cada mil dólares invertidos en publicidad se estima un incremento promedio de 4615 dólares en las ventas. **c)** No tiene sentido, el modelo se construye con datos entre 4200 y 7200 de publicidad, nada garantiza que la linealidad se va a mantener hasta publicidad = 0. **d)** 40680 dólares en promedio.

Ejercicio 4. Una firma de servicios de electricidad debe prever constantemente la demanda de electricidad semanal para administrar adecuadamente los inventarios de petróleo utilizados. Se supone que la demanda de electricidad puede variar directamente con la temperatura. Se analizaron 12 semanas del año pasado obteniendo los siguientes resultados:

Predicción de temperatura media hecha por el servicio meteorológico (Grados Celsius)	Consumo de petróleo (Miles de litros)
23	101
25	105
24	102
27	111
29	112
21	105
20	98
19	96
22	100
29	113
28	110
32	122

a) Determinar los estimadores de cuadrados mínimos de los parámetros del modelo de regresión lineal simple. Dar la expresión de la ecuación correspondiente.

b) Interpretar los coeficientes de la ecuación en términos del problema.

c) Estimar el consumo promedio de petróleo para una semana en la que la predicción del servicio meteorológico para la temperatura media fue de 26 Grados Celsius.

d) Estimar el coeficiente de correlación lineal. Interpretar dicho valor.

Rta. Ej 4. a) $Y = 62{,}828 + 1{,}742\, X$. **c)** Consumo promedio de $108137{,}9$ litros de petróleo. **d)** $r = 0{,}94847$.

Ejercicio 5. En una gran carpintería se registraron, para una pieza determinada, los costos (en horas hombre: hh) en función de la superficie de las piezas en m^2.

Superficie (m^2)	Costo (hh)	Superficie (m^2)	Costo (hh)	Superficie (m^2)	Costo (hh)	Superficie (m^2)	Costo (hh)
0,2	2	0,1	2	1,5	3	1,1	2
1,4	2,58	1,3	2,52	1	2,51	3,1	3,95
2,6	3,21	1	2,59	2	2,91	0,9	2,47
0,7	2,03	2,5	3,58	2,7	3,71	2,1	3,12
1,3	2,36	0,4	2,29	2,2	3,13	1,3	2,45
1,7	2,69	0,3	1,88	0,7	2,33	1,1	2,49
0,6	2,48	2,6	3,48	1,7	2,93	2,6	3,27
2	3,23	1,6	2,97	2,4	3,46	1,1	2,77
0,4	2,24	2,9	3,87	0,8	2,57	1,8	2,96
2,8	3,72	1	2,44	0,3	1,83	2,3	3,37

a) Realizar el gráfico de dispersión

b) Estimar la recta de regresión. Interpretar la pendiente de la recta en términos del problema.

c) ¿Dirías que existe un costo fijo de producción? Justificar. Si la respuesta es afirmativa, ¿cuál es su estimación?

d) Recibes un pedido de presupuesto por 150 piezas: 50 de $0{,}8\, m^2$, 80 de $1{,}2\, m^2$ y el resto de $2{,}3\, m^2$. Decides presupuestar un monto equivalente al costo de producción mas un 20% en concepto de margen de utilidad. ¿Cuál debería ser el importe de la cotización?

Ejercicio 6. El Gerente de Logística desea construir un modelo para presupuestar el costo anual de mantenimiento de los auto-elevadores. En base a su experiencia cree que la variable de mayor relevancia es la antigüedad del equipo. Un analista de la gerencia recopiló la siguiente información para 7 equipos.

Antigüedad (años)	4,5	1	1	5	0,5	4	6
Costo Anual ($u\$s$)	619	549	466	1194	163	723	1345

a) Determinar los estimadores de cuadrados mínimos de los parámetros del modelo de regresión lineal simple. Dar la expresión de la ecuación correspondiente.
b) Interpretar los coeficientes de la ecuación en términos del problema.
c) Estimar los coeficientes de correlación y determinación e interpretar.
d) La gerencia decide reemplazar los dos equipos más antiguos (5 y 6 años respectivamente) por dos equipos nuevos. ¿Cuál será la reducción al cabo de un año en costo de mantenimiento en estos dos equipos?

Ejercicio 7. Cohetes a propulsión. El motor de un cohete se fabrica al unir dos tipos de propulsores, uno de encendido y un impulsor. Se piensa que la resistencia al esfuerzo cortante de la junta, Y, es una función lineal de la edad del propulsor (tiempo de vida) cuando se arma, X (en semanas). Se pudo hacer las siguientes 20 mediciones.

Medición nº	1	2	3	4	5	6	7	8	9
Tiempo de vida	15,5	23,75	8,00	17,00	5,00	19,00	24,00	2,50	7,50
Resistencia (psi)	2158,70	1678,15	2316,00	2061,30	2307,30	1808,30	1784,70	2575,00	2357,90

10	11	12	13	14	15	16	17	18	19	20
11,00	13,00	3,75	25,00	9,75	22,00	18,00	6,00	12,50	2,00	21,50
2277,70	2165,20	2399,55	1779,80	2336,75	1765,30	2053,50	2414,20	2200,50	2654,20	1753,70

a) Construir un diagrama de dispersión de los datos. ¿Parece plausible un modelo de regresión lineal? Corroborarlo con algún cálculo apropiado.
b) Encontrar las estimaciones de cuadrados mínimos para la pendiente y la ordenada al origen del modelo de regresión lineal.
c) Estimar la resistencia al esfuerzo promedio de un motor armado y construido con un propulsor de 20 semanas.

Rta. Ej 7. a) $r^2 = 0{,}9363$. **b)** $Y = 2638{,}53 - 37{,}199\ X$. **c)** $1894{,}55\ psi$

EJERCICIOS CON ALGUNA CONSIDERACIÓN TEÓRICA ADICIONAL

Ejercicio 8. La tabla muestra valores observados de dos variables aleatorias X y Y.

X	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

a) Hallar la ecuación de la recta de regresión muestral de Y sobre X y también la ecuación de la recta de regresión muestral de X sobre Y. Comparar ambos resultados. ¿Se puede despejar una de la otra? Justificar.
b) Calcular el coeficiente de correlación muestral para ambos modelos de regresión. ¿Coinciden?
c) ¿Cuál es el ángulo entre ambas rectas de regresión?

Rta. Ej 8. **a)** $Y = 0{,}5454 + 0{,}6363\ X$; $X = -0{,}5 + 1{,}5\ Y$. **d)** $\theta = 23{,}84°$

Ejercicio 9. La luz arranca electrones. Cuando sobre una superficie metálica incide radiación electromagnética en cierto intervalo de frecuencias se pone en evidencia la presencia de electrones libres provenientes del metal. A este fenómeno se lo llama **efecto fotoeléctrico**. Mediante un campo magnético se puede frenar esos electrones emitidos y de esa manera tener una medida de la energía cinética máxima de ellos. Asociando a la radiación incidente de frecuencia f, fotones de energía hf con h la constante de Planck se puede plantear la ecuación de Einstein para el efecto fotoeléctrico:

$$e \cdot V = h \cdot f - \Phi$$

donde e es la carga eléctrica del electrón ($1{,}6 \times 10^{-19} C$), V el potencial de frenado, y Φ es el llamado trabajo de extracción fotoeléctrico (mínima energía necesaria para extraer un electrón del metal).

A partir de un experimento controlado, se tienen los datos referentes al potencial de frenado en función de la frecuencia para cierto metal, expuestos en la siguiente tabla.

f	0,82	0,74	0,69	0,61	0,55	0,52
V	1,45	1,18	0,90	0,65	0,32	0,24

f: frecuencia en 10^{15} $Hertz$ $(seg^{-1})10^{15}$ Hertz V: potencial de frenado en volt

Usando el método de mínimos cuadrados determinar:

a) la relación h/e y con ella h (la constante universal por excelencia de la mecánica cuántica);
b) el trabajo de extracción fotoeléctrico del metal;
c) el coeficiente de correlación muestral.

Rta. Ej 9. a) $h/e = 4{,}119$; $h = 6{,}5904 \times 10^{-19}$ (en unidades adecuadas).
b) $\Phi = 3{,}0532 \times 10^{-19}$. **c)** $r = 0{,}9971$.

Ejercicio 10. Desintegración radiactiva. Se tomaron medidas de la actividad A (número de desintegraciones por unidad de tiempo) de una muestra radiactiva (PA 234) con un contador Geiger-Müller correspondiente a distintos instantes de tiempo. Se sabe que la relación entre la actividad y el tiempo es $A = A_0 e^{-\lambda t}$ donde A_0 es la actividad inicial y λ es una constante propia del elemento radiactivo que se desintegra (λ se denomina constante de desintegración).

Los datos disponibles están reunidos en la siguiente tabla.

t	75	90	105	120	135	150	165	180	195	210	225	240	255
A	9,00	7,50	7,00	5,50	5,00	4,50	3,50	3,25	2,75	2,25	2,10	1,85	1,50

A en miles de desintegraciones por minuto, y t en segundos.

Estimar por cuadrados mínimos:

a) las constantes A_0 y λ; y,
b) $T = \ln 2 / \lambda$, llamado período de desintegración, que es el tiempo en que el número de desintegraciones se reduce a la mitad que, al igual que λ, caracteriza a cada especie radiactiva.
c) Representar gráficamente la actividad en función del tiempo, superponiendo los datos medidos y la función interpoladora utilizada en los ítems anteriores.

Ayuda: considerar el logaritmo de la actividad en función del tiempo, expresión linealizada de la ley de dependencia, $\ln A = \ln A_0 - \lambda t$.

Rta. Ej 10. a) $A_0 = 18{,}331$; $\lambda = 0{,}009687$. **b)** $T = 70\ seg.$