Trabajo Práctico 9. Modelo de regresión lineal simple

Conjunto de n mediciones $\{(x_1,y_1); (x_1,y_1); \ldots; (x_n,y_n)\}$

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

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

Fórmulas de cálculo

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

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 para los n pares

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

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 muestrados.
c) 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.

Ejercicio 2. Se determinaron los valores del módulo de elasticidad (MoE, la relación del esfuerzo −fuerza por unidad de área− a la deformación unitaria −deformación por unidad de longitud− en GPa) y la resistencia a la flexión (medida de la capacidad de resistir a fallas por flexión, en MPa) para una muestra de vigas de concreto de cierto tipo, resultando los siguientes datos (obtenidos de un gráfico del artículo "Effects of Aggregates and Microfilters on the Flexural Properties of Concrete", *Magazine of Concrete Research*, 1997, pp. 81-98).

MoE	29,8	33,2	33,7	35,3	35,5	36,1	36,2	36,3	37,5
Resistencia	5,9	7,2	7,3	6,3	8,1	6,8	7,0	7,6	6,8

MoE	37,7	38,7	38,8	39,6	41,0	42,8	42,8	43,5	45,6
Resistencia	6,5	7,0	6,3	7,9	9,0	8,2	8,7	7,8	9,7

MoE	46,0	46,9	48,0	49,3	51,7	62,6	69,8	79,5	78,0
Resistencia	7,4	7,7	9,7	7,8	8,7	11,6	11,3	11,8	10,7

a) Trazar el diagrama de dispersión. ¿Sugiere este gráfico que el valor de la resistencia está determinado total y únicamente por el valor del módulo de la elasticidad?
b) Hallar la ecuación de la recta de regresión lineal para los datos muestrados.
c) ¿Es de confiar el uso de la recta anterior obtenida con el método de los cuadrados mínimos para estimar la resistencia cuando el módulo de la elasticidad fuera 100? Explicar la respuesta.
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.

Ejercicio 3. 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 que 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) Estimar el coeficiente de determinación.

Ejercicio 4. 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

Hallar:
a) la ecuación de la recta de regresión muestral de Y sobre X;
b) la ecuación de la recta de regresión muestral de X sobre Y;
c) el coeficiente de correlación muestral;
d) el ángulo entre ambas rectas de regresión.

Ejercicio 5. 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 meteorlógico (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) Estimar el consumo promedio de petróleo para una semana en la que la predicción del servicio meteorológico para la temperatura media fue de 26º.
c) Estimar el coeficiente de correlación lineal.

Ejercicio 6. Un ingeniero ha sido asignado a la tarea de desarrollar un modelo de predicción de pico de carga en una planta nuclear. Para ello, inicialmente consideró la relación entre "demanda de carga" del sistema interconectado y la "temperatura máxima" registrada en el día, basándose en el supuesto de que las altas temperaturas provocan picos de mayor potencia demandada. Registró valores correspondientes a 6 días tomados al azar. Los datos registrados se muestran en la tabla.

Día	1	2	3	4	5	6
Temperatura máxima (ºC)	33,3	28,9	35,0	38,9	31,1	36,1
Pico de demanda (Megawatts)	207	139	211	273	156	244

a) Estimar el coeficiente de correlación lineal entre ambas variables. Formular un comentario sobre la linealidad entre la temperatura máxima y la demanda de carga.
b) Hacer una predicción de la carga pico demandada en un día con temperatura máxima de 36°C.

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.

Ejercicio 8. 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:

$$eV = h\,f - \Phi$$

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 (1/seg) *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.

Ejercicio 9. 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 $A = A_0 \exp(-\lambda t)$ donde A_0 es la actividad inicial y λ es una constante propia del elemento radiactivo que se desintegra (L 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 L, 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, $\ln A = \ln A_0 - \lambda t$.