

Práctica

Regresión lineal simple

- 1) Se utiliza regresión lineal para analizar los datos de un estudio donde se investigó la relación que existe entre la temperatura de la superficie de una carretera (x) y la deformación del pavimento (y). El resumen de cantidades es el siguiente:

$$n = 20, \sum y_i = 12.75, \sum y_i^2 = 8.86, \sum x_i = 1478, \sum x_i^2 = 143215.8, \sum x_i y_i = 1083.67$$

- Calcular las estimaciones de mínimos cuadrados de la pendiente y la ordenada al origen. Hacer un gráfico de la recta de regresión, y estimar σ^2 .
- Utilice la ecuación de la recta ajustada para predecir la deformación del pavimento observada cuando la temperatura de la superficie sea 85°F.

- 2) Se tiene la siguiente información sobre la relación entre una medida de la corrosión del hierro (Y) y la concentración de NaPO₄ (X, en ppm)

x	2.50	5.03	7.60	11.60	13.00	19.60	26.20	33.00	40.00	50.00	55.00
y	7.68	6.95	6.30	5.75	5.01	1.43	0.93	0.72	0.68	0.65	0.56

- Construya un gráfico de dispersión de los datos. ¿Parece ser razonable el modelo de regresión lineal?
 - Calcule la ecuación de la recta de regresión estimada, utilícela para pronosticar el valor de la rapidez de corrosión que se observaría para una concentración de 33 ppm, y calcule el residuo correspondiente.
 - Estime la desviación estándar de observaciones alrededor de la recta de regresión verdadera.
 - ¿Se puede concluir que el modelo de regresión lineal simple especifica una relación útil entre las dos variables?. Establezca y pruebe las hipótesis adecuadas al nivel de significación de 0.05.
- 3) En pruebas diseñadas para medir el efecto de cierto aditivo en el tiempo de secado de pintura se obtuvieron los siguientes datos

Concentración de aditivo (%)	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8
Tiempo de secado (horas)	8.7	8.8	8.3	8.7	8.1	8.0	8.1	7.7	7.5	7.2

Se tiene el siguiente resumen estadístico:

$$\bar{x} = 4.9 \quad \bar{y} = 8.11 \quad S_{xx} = 3.29967 \quad S_{xy} = -2.75 \quad S_{yy} = 2.58894$$

- Realizar un gráfico de dispersión y estimar la recta de regresión lineal
- Estimar la varianza
- Pronostique el tiempo de secado para una concentración de 4.4%
- ¿Puede utilizarse la recta de mínimos cuadrados para pronosticar el tiempo de secado respecto a una concentración de 7%?

e) ¿Para qué concentración pronosticaría un tiempo de secado de 8.2 horas?

- 4) Un químico está calibrando un espectrómetro que se utilizará para medir la concentración de monóxido de carbono en muestras atmosféricas. Para comprobar la calibración, se miden muestras de concentración conocida.

Las concentraciones verdaderas (x) y las medidas (y) están dadas en la tabla siguiente:

x (ppm)	0	10	20	30	40	50	60	70	80	90	100
Y (ppm)	1	11	21	28	37	48	56	68	75	86	96

Para comprobar la calibración se ajusta un modelo lineal $y = \beta_0 + \beta_1 x + \varepsilon$. Idealmente, el valor de β_0 debe ser 0 y el valor de β_1 debe ser 1.

- Calcule los estimadores de mínimos cuadrados $\hat{\beta}_0$ y $\hat{\beta}_1$
- ¿Se puede rechazar la hipótesis nula $H_0: \beta_0 = 0$? Utilice $\alpha = 0.05$
- ¿Se puede rechazar la hipótesis nula $H_0: \beta_1 = 1$? Utilice $\alpha = 0.05$
- ¿Los datos proporcionan suficiente evidencia para concluir que la máquina está fuera de calibración?

- 5) En un experimento para investigar la relación entre el diámetro de un clavo (x) y su fuerza retirada final (y), se colocaron clavos de forma anular enhebrados en madera de abeto de Douglas, y después se midieron sus fuerzas de retirada en N/mm. Se obtuvieron los resultados siguientes para 10 diámetros diferentes (en mm):

x	2.52	2.87	3.05	3.43	3.68	3.76	3.76	4.50	4.50	5.26
y	54.74	59.01	72.92	50.85	54.99	60.56	69.08	77.03	69.97	90.70

- Calcule la recta de mínimos cuadrados para predecir la fuerza a partir del diámetro
 - Determine el intervalo de confianza de nivel 0.95 para la media de la fuerza de retirada de clavos de 4 mm de diámetro
 - Determine el intervalo de predicción de nivel 0.95 para la fuerza de retirada de clavos de 4 mm de diámetro
 - ¿Puede concluir que la media de la fuerza de retirada de clavos de 4 mm de diámetro es 60 N/mm con un nivel de significancia de 0.05?
- 6) Un comerciante realizó un estudio para determinar la relación que hay entre los gastos de la publicidad semanal y las ventas. Registró los datos siguientes:
- Costos de publicidad (en \$):* 40, 20, 25, 20, 30, 50, 40, 20, 50, 40, 25, 50.
- Ventas (en \$):* 385, 400, 395, 365, 475, 440, 490, 420, 560, 525, 480, 510.
- Haga un gráfico de dispersión
 - Encuentre la recta de regresión estimada para pronosticar las ventas semanales, a partir de los gastos de publicidad.
 - Estime las ventas semanales cuando los costos de la publicidad sean 35\$. ¿Es válido estimar las ventas semanales cuando los costos de la publicidad sean 75\$?
 - Pruebe la hipótesis de que $\beta_1 = 6$ contra la alternativa de que $\beta_1 < 6$, utilice $\alpha = 0.025$.
 - Construya un intervalo de confianza de 95% para la media de las ventas semanales cuando se gastan 45\$ en publicidad.
 - Construya un intervalo de predicción de 95% para la media de las ventas semanales cuando se gastan 45\$ en publicidad.
 - ¿Qué proporción de la variabilidad total en las ventas está explicada por el costo en publicidad?
- 7) Considere los siguientes datos sobre el número de millas para ciertos automóviles, en millas por galón (mpg) y su peso en libras (wt)

Modelo	GMC	Geo	Honda	Hyundai	Infiniti	Isuzu	Jeep	Land	Lexus	Linclon
Wt (x)	4520	2065	2440	2290	3195	3480	4090	4535	3390	3930
Mpg (y)	15	29	31	28	23	21	15	13	22	18

- Estime la recta de regresión lineal.
 - Estime las millas para un vehículo que pesa 4000 libras.
 - Suponga que los ingenieros de Honda afirman que, en promedio, el Civic (o cualquier otro modelo de vehículo que pese 2440 libras) recorre mas de 30 mpg. Con base en los resultados del análisis de regresión, ¿es esta afirmación creíble?, ¿por qué?
Sugerencia: calcule el intervalo de confianza para la media mpg cuando el peso es de 2440 libras, con $\alpha = 0.05$
 - Los ingenieros de diseño para el Lexus ES300 tienen por objetivo lograr 18 mpg como ideal para dicho modelo (o cualquier otro que pese 3390 libras), aunque se espera que haya cierta variación. ¿Es probable que sea realista ese objetivo?. Comente al respecto.
 - ¿Qué proporción de la variabilidad total en el millaje está explicada por el peso del motor?
- 8) Los valores siguientes son 26 lecturas sobre la congestión del transito y la concentración de monóxido de carbono efectuadas en un sitio de muestreo para determinar la calidad del aire de cierta ciudad. Enunciando las suposiciones necesarias , resolver los siguientes incisos :
- Prepare un diagrama de dispersión.
 - Calcule el coeficiente de correlación de la muestra.

congestion de transito		congestion de transito	
(automoviles por hora), X	CO (ppm), Y	(automoviles por hora), X	CO (ppm), Y
100	8.8	375	13.2
110	9.0	400	14.5
125	9.5	425	14.7
150	10.0	450	14.9
175	10.5	460	15.1
190	10.5	475	15.5
200	10.5	500	16.0
225	10.6	525	16.3
250	11.0	550	16.8
275	12.1	575	17.3
300	12.1	595	18.0
325	12.5	600	18.4