

Matemática IV – 2024

TP N°2 – Regresión Lineal

1) Suponga que $(x_1, y_1), \dots, (x_n, y_n)$ son pares observados generados por los siguientes modelos y deduzca los estimadores de mínimos cuadrados de β_1 y β_0 .

a) $Y = \beta_1 x + \varepsilon$

b) $Y = \beta_1(ax + c) + \beta_0 + \varepsilon$

2) Una cadena de supermercados financia un estudio sobre los gastos mensuales en alimentos, de familias de 4 miembros. La investigación se limitó a familias con ingresos netos entre \$688.000 y \$820.000, con lo cual se obtuvo la siguiente recta de estimación $\hat{y} = 0,85x - 18.000$.

$y = \text{gastos}$; $x = \text{Ingresos}$

a) Estime los gastos en alimentos en un mes, para una familia de 4 miembros con un ingreso de \$700.000

b) Uno de los directivos de la compañía se preocupa por el hecho de que la ecuación aparentemente indica que para una familia que tiene un ingreso de \$12.000 no gastaría nada en alimentos ¿Cuál sería su respuesta?

3) La empresa META quiere pronosticar el precio de sus acciones en función de los días en el periodo del 03/09/23 al 30/08/24, pero durante las fechas del 02/02/24 al 24/04/24 implementaron una serie de actualizaciones en sus distintas plataformas que dispararon el precio de sus acciones y querían saber en que porcentaje afectaron dichas actualizaciones al ajuste y a la linealidad.

Utilizando los datos proporcionados en el archivo “META” haga los cálculos necesarios y responda

Sugerencia: Realice dos analisis diferentes y para una de ellas desestimar los datos del periodo de actualización.

4) Los siguientes datos corresponden a los tiempos relativos en segundos que tardaron en ejecutarse seis programas elegidos al azar en el entorno Windows y en DOS :

	Programas					
Windows	2,5	7,1	5	8,5	7	8,1
DOS	2,3	7,1	4	8	6,6	5

a) Realizar el grafico de dispersión de los puntos

b) Si un programa tarda 6 segundos en ejecutarse en Windows, ¿Cuánto tardara en ejecutarse en DOS?

c) Se estima que los tiempos de Windows mejoraran reduciéndose en un 10% en los próximos años, estime la recta de regresión considerando esta mejora. Suponga que los tiempos DOS no se modifican.

5) En la tabla siguiente, se muestran la variable y , rendimiento de un sistema informático, respecto a la variable x , numero de buffer:

x	5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
y	9.6	20.1	29.9	39.1	50.0	9.6	19.4	29.7	40.3	49.9	10.7	21.3	30.7	41.8	51.2

A partir de la tabla anterior, se quiere ajustar la variable y como función de x .

a) Realizar el analisis de regresión de los datos (Estimación de la recta, Test de Hipótesis, Indicadores).

b) Comentar los resultados siguientes:

- Recta de regresión del rendimiento del sistema informático frente al número de buffers e interpretación de los coeficientes.
- Contraste de hipótesis sobre la pendiente de la recta.
- Coeficiente de determinación y correlación lineal.

6) Determine si las siguientes relaciones son posibles o no y justifique su respuesta:

a) $\hat{\sigma}^2 = 0,2$ $n = 102$ $R^2 = 0,8$ $S_{yy} = 100$

b) $\hat{y} = 7x + 4$ $\bar{x} = 10$ $\bar{y} = 64$ $r = -0,8$

c) $\hat{\beta}_0 = 10,073$ $\hat{\beta}_1 = -2,06$ $\bar{x} = 8,5$ $\bar{y} = 8,325$

7) Indique si las siguientes afirmaciones son correctas o no. Justifique su respuesta:

a) $SS_R = S_{yy} - \hat{\beta}_0 S_{xy}$

b) El error del intervalo de predicción es $\sqrt{n+1}$ veces mayor que el intervalo confianza para la respuesta media cuando $x^* = \bar{x}$ e igual $(1 - \alpha)$.

c) El coeficiente de determinación R^2 indica el grado de relación lineal que existe entre la variable independiente y dependiente.

d) El principio de mínimos cuadrados consiste en minimizar la suma de los residuos al cuadrado considerando la distancia perpendicular entre el valor observado y el estimado.

8) En un departamento de informática, un grupo de investigación dedicado al estudio de las comunicaciones por la red desea conocer la relación entre el tiempo de transmisión de un fichero y la información útil del mismo. Para ello se han hecho algunos experimentos en los que se enviaban paquetes de distintas longitudes (bytes) de información útil y se median los tiempos (en milisegundos) que tardaban desde el momento en que se enviaban hasta que llegaban al servidor. Los resultados del experimento se resumen en los siguientes estadísticos:

$$S_{xx} = 47.990 \quad \bar{x} = 194 \quad \hat{\beta}_0 = 27,3275$$

$$\sum x_i^2 = 424.350 \quad \sum x_i y_i = 183.760 \quad \sum y_i^2 = 81.715$$

Se pide estudiar la relación entre las variables tiempo (y) y longitud (x) de los ficheros. Para ello, se pide:

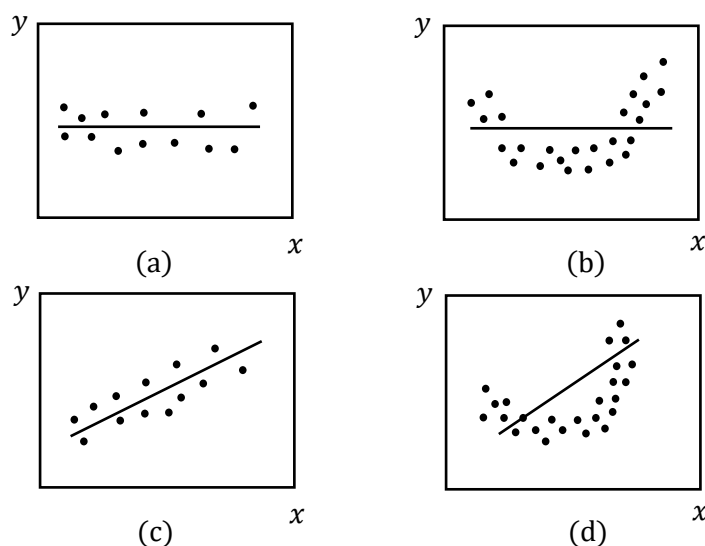
- Obtener la recta de regresión del tiempo en función de la longitud de los ficheros. Interpretar los resultados obtenidos.
- Indicar el valor que toma el coeficiente de determinación y correlación lineal. Interpretar los resultados.
- Estudiar la significación del modelo.
- Obtener el intervalo de confianza, al 95%, para la pendiente de la recta,
- ¿Cuál será el tiempo de transmisión para un fichero que tiene una longitud 250 bytes?

9) De un análisis de regresión realizada sobre un Dataset, el cual consiste en un pequeño relevamiento del tiempo que demandan las llamadas a servicio técnico de una empresa (x) y la cantidad de unidades de hardware reparadas (y), se sabe que el $IC(\beta_0) = (-0,4348 ; -0,4248)$, que la estimación de la pendiente es 12 veces el error que se comete al estimar la verdadera ordenada al origen con $\hat{\beta}_0$ y que la proporción de variación total observada no explicada por el modelo de regresión lineal es tan solo del 2%.

A partir de los datos proporcionados determinar:

- El error que se comete al estimar la verdadera ordenada al origen con $\hat{\beta}_0$
- La recta de regresión estimada
- La bondad del ajuste

10) Observando los siguientes gráficos de regresión y considerando las hipótesis $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ Indique para cada una, si se acepta o no H_0 y la implicancia de esta.



11) La autoridad aeronáutica argentina realizó un estudio de operaciones de aerolíneas, en 18 compañías, que reveló que la relación entre el número de pilotos empleados y el número de aviones en servicio tenía una pendiente de 4.3. Estudios anteriores indicaban que la pendiente de esta relación era 4.0. Si se calculó que la desviación estándar de la de pendiente de regresión es 0.17, ¿hay razones para creer, a un nivel de significancia de 0.05, que la pendiente verdadera ha cambiado?

12) Un horticultor inventó una escala para medir la frescura de rosas que fueron empacadas y almacenadas durante periodos variables antes de trasplantarlas. La medición y de frescura y el tiempo x en días que la rosa está empacada y almacenada antes de trasplantarla, se dan a continuación.

x	5	5	10	10	15	15	20	20	25	25
y	15,3	16,8	13,6	13,8	9,8	8,7	5,5	4,7	1,8	1,0

- ¿Hay suficiente evidencia para indicar que la frescura está linealmente relacionada con el tiempo de almacenaje?
- Estime mediante un intervalo de 98% el descenso de frescura de las rosas por cada día que pasa.
- Estime mediante un intervalo de 98% la frescura de las rosas cuando no han sido almacenada ni empacada.
- Estime la medición de frescura media para un tiempo de almacenaje de 14 días con un intervalo de confianza de 95%.

13) Un fabricante de teléfonos celulares está probando dos tipos de baterías para ver cuánto duran con una utilización normal. La siguiente tabla contiene los datos provisionales:

Horas de uso diario	2	1,5	1	0,5
Vida aproximada (meses) Litio	3,1	4,2	5,1	6,3
Vida aproximada (meses) Alcalina	1,3	1,6	1,8	2,2

- Desarrolle dos ecuaciones de estimación lineales, una para pronosticar la vida del producto basada en el uso diario con las baterías de litio y otra para las baterías alcalinas.
- ¿Cuál de las dos estimaciones anteriores se ajusta mejor a los datos?
- Encuentre un intervalo para la estimación del 90% para la vida (en meses) con 1,25 horas de uso diario, para cada tipo de batería. ¿Puede la compañía asegurar algo respecto a qué batería proporciona la vida más larga según estos números?
- El fabricante considera realizar una batería compuesta por los dos tipos de batería y pide para ello que se estime la ecuación lineal para pronosticar las horas de uso diario basada en el vida aproximada (en meses).
- Mejora la estimación utilizando los dos tipos de batería juntas que por separado. Explique.

14) Una Empresa de desarrollo de software le pide relacionar sus Ventas en función del número de pedidos de los tipos de software que desarrolla (Sistemas, Educativos y Automatizaciones Empresariales), para atender 10 proyectos en el presente año. En la Tabla representa Y (Ventas miles de S/.) e X (Nº pedidos de sistemas), W (Nº de pedidos de Aplicaciones Educativas) y Z (Nº de pedidos de Automatizaciones empresariales).

y	440	455	470	510	506	480	460	500	490	450
x	50	40	35	45	51	55	53	48	38	44
w	105	140	110	130	125	115	100	103	118	98
z	75	68	70	64	67	72	70	73	69	74

a) Mediante un software a elección estime la ecuación de regresión múltiple para cumplir con el requerimiento de la empresa.

b) La empresa quiere tener indicadores para asegurarse que la ecuación estimada se ajusta bien a los datos y si la relación lineal es la más correcta. ¿Cuáles recomendaría? Calcule las mismas y comente.

15) En la Facultad de Sistemas Informáticos se quiere entender los factores de aprendizaje de los alumnos que cursan la asignatura de PHP, para lo cual se escoge al azar una muestra de 15 alumnos y ellos registran notas promedios en las asignaturas correlativas de Algoritmos, Base de Datos y Programación como se muestran en el siguiente cuadro.

PHP	Algoritmos	Base de Datos	Programación
13	15	15	13
13	14	13	12
13	16	13	14
15	20	14	16
16	18	18	17
15	16	17	15
12	13	15	11
13	16	14	15
13	15	14	13
13	14	13	10
11	12	12	10
14	16	11	14
15	17	16	15
15	19	14	16
15	13	15	10

a) Construir un modelo para determinar la dependencia que exista de aprendizaje reflejada en las notas de la asignatura de PHP, conociendo las notas de las asignaturas Algoritmos, Base de Datos y Programación.

b) Si más el 80% del aprendizaje del Curso de PHP no puede ser explicado mediante las notas obtenidas por las asignaturas de Algoritmos, Base de Datos y Programación, se destinarán más recursos a estas asignaturas para obtener mejores resultados. ¿Cuál es seria su respuesta?