

# Matemática IV- 2023

## TP2 - Regresión Lineal Simple y Múltiple

1. Suponga que un investigador cuenta con datos sobre el espacio de memoria RAM dedicado a guardar información de ciertos códigos ( $x$ ) y los tamaños de ingreso ( $y$ ) de esos códigos. Es posible que el investigador desee adoptar un modelo para calcular la recta de regresión verdadera que pase a través de  $(0, 0)$ . El modelo apropiado es  $Y = \beta_1 x + \varepsilon$ . Suponga que  $(x_1, y_1), \dots, (x_n, y_n)$  son pares observados generados con este modelo y deduzca el estimador de mínimos cuadrados de  $\beta_1$ .
2. 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 3 segundos en ejecutarse en Windows, ¿Cuánto tardara en ejecutarse en DOS?
  - (c) Se estima que los tiempos de Windows mejoraran reduciéndose en un 5% en los próximos años, estime la recta de regresión considerando esta mejora.
3. Una compañía de transporte analiza una muestra de 10 embarques por camión y registra la distancia en kilómetros, además el tiempo de entrega de mercancía medido por días, obteniéndose la siguiente información:

$$\sum x_i = 7620 \quad \sum y_i = 28,5 \quad \sum x_i y_i = 26370 \quad \sum x_i^2 = 7104300 \quad \sum y_i^2 = 99,75$$

- (a) La compañía informa que si el kilometraje es de 1500, el tiempo para la entrega es mayor de tres dias. ¿Está usted de acuerdo?
  - (b) ¿Qué porcentaje de información es explicada por la recta de regresión?
  - (c) Dar una medida de la relación lineal entre las variables involucradas
4. De un dataset, el cual consiste en un pequeño relevamiento del tiempo que demandan las llamadas a servicio técnico de una empresa y la cantidad de unidades de hardware reparadas accidentalmente se perdieron los datos observados, quedando los residuos de estos y la estimación para los minutos 87 y 120 de llamadas que son 5.260 y 7.418 unidades respectivamente.

$e_i$	-0,074	0,533	-0,775	0,244	-0,410	-0,260	0,151
$e_i$	0,086	0,301	0,647	-0,315	-0,053	0,358	-0,427

- (a) Determinar la recta de regresión estimada para pronosticar los hardware reparados a partir de los minutos de llamada.
- (b) Estimar la variabilidad de los datos observados alrededor de la recta de regresión y dar una interpretación del resultado.
5. Suponga que se está estudiando el efecto de cierta fricción constante y uniforme sobre laminas metálicas del mismo grosor. Se ha efectuado 8 observaciones cuyos resultados son:

$$x = \text{Tiempo de fricción (hs)} \quad y = \text{Grosor de la lamina (mm)}$$

Se pide mediante el cálculo de la ecuación de regresión:

x	3	5	6	8	9	10	12	15
y	9,5	9,0	8,8	8,5	8,2	8,0	7,6	7,0

- (a) ¿Cuál fue el grosor original de las láminas?
- (b) ¿Cuántas horas tardara la fricción en destruir una lámina?
- (c) Tiene algún sentido pronosticar el grosor de una lámina después de 70 horas de fricción?
6. Un virus informático atacó el disco duro (HDD) de una computadora y a partir de una investigación realizada por expertos sobre el tema se adoptó el siguiente modelo

$$Y = 10\beta_1 x + \beta_0 + \varepsilon$$

para determinar el porcentaje daño producido por el virus en función de los días y se obtuvieron los siguientes estadísticos:

$$S_{xx} = 27768,357 \quad \bar{x} = 97,2143 \quad \hat{\beta}_0 = -0,4298$$

$$\sum x_i y_i = 9885 \quad \sum x_i^2 = 160077 \quad \sum y_i^2 = 27768,357$$

- (a) ¿Qué miden la variable independiente y dependiente?
- (b) Determinar el tamaño de la muestra utilizado para obtener los estadísticos. (aproxime el valor encontrado al entero más próximo)
- (c) Estime la Recta de regresión.
- (d) ¿En cuantos días se pronostica un daño mayor al 90%?
- (e) La variación de las observaciones respecto a la recta de regresión estimada es ¿cercana o lejana? Justifique.

7. 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:

- (a) El error que se comete al estimar la verdadera ordenada al origen con  $\beta_0$ .
- (b) La recta de regresión estimada.
- (c) La bondad del ajuste.

8. 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 que la pendiente verdadera ha cambiado?

9. El siguiente intervalo de confianza para la respuesta media de 95%, para un Toyota Hilux de 2175 kg, se obtuvo de una serie de datos sobre el número de kilómetros para ciertos automóviles, en  $\frac{km}{litro}$  y su peso en kg:

$$\left[ 6,5 - 2,306\sqrt{2,214\left(\frac{1}{10} + \frac{(2175 - 3395)^2}{7283652,5}\right)} ; 6,5 + 2,306\sqrt{2,214\left(\frac{1}{10} + \frac{(2175 - 3395)^2}{7283652,5}\right)} \right]$$

- (a) ¿Cuántos kilogramos más puede pesar el Toyota Hilux para que el ancho del intervalo sea mínimo?
- (b) En condición del inciso anterior ¿Puede recorrer más  $8,5 \frac{km}{litro}$  en promedio?

10. 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

- (a) ¿Hay suficiente evidencia para indicar que la frescura está linealmente relacionada con el tiempo de almacenaje?
- (b) Estime mediante un intervalo de 98% el descenso de frescura de las rosas por cada día que pasa.
- (c) Estime mediante un intervalo de 98% la frescura de las rosas cuando no han sido almacenada ni empacada.

- (d) Estime la medición de frescura para un tiempo de almacenaje de 14 días con un intervalo de confianza de 95%.
11. 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

- (a) 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.
- (b) ¿Cuál de las dos estimaciones anteriores se ajusta mejor a los datos?
- (c) Encuentre un intervalo de confianza 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?
- (d) 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 la vida del producto basada en el uso diario con las baterías.
- (e) Mejora la estimación utilizando los dos tipos de batería juntas que por separado. Explique.
12. 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.

13. 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?