Trimestre Enero-Marzo 2008
Departamento de Cómputo Científico y Estadística
Guía de ejercicios. Regresión Lineal Simple
Práctica N° 6

CONTENIDO
- Método de mínimos cuadrados.
- Propiedades de los estimadores de mínimos cuadrados.
- Inferencia respecto a los parámetros
- Predicción.
- Ajuste de otros modelos.

1. Se realiza un experimento para contar el número de manatíes observados en un área, desde un helicóptero y desde un avión. Después de contar manatíes por 10 días de estas dos maneras, se obtiene

Día	N° de manatíes	
	Desde avión	Desde helicóptero
1	24	30
2	31	30
3	32	33
4	39	38
5	47	58
6	47	58
7	35	48
8	76	75
9	95	85
10	85	55

Considerando y: número de manatíes contado desde el avión, x: número de manatíes contado desde el helicóptero

a) Ajuste una recta a estos datos e interprete los coeficientes.
b) Realice las siguientes pruebas t al nivel 95%

$$\begin{aligned} H_0 &: \beta_1 = 0 \\ H_a &: \beta_1 \neq 0 \end{aligned}$$

$$\begin{aligned} H_0 &: \beta_1 = 1 \\ H_a &: \beta_1 \neq 1 \end{aligned}$$

c) ¿Qué significa cada una de ellas en este caso?
d) ¿Cuál les parece la más adecuada?

e) ¿Qué significado tiene la constante?
f) ¿Le parecería mejor una regresión por el origen? Hágala
g) ¿Qué pendiente obtuvo?
h) ¿Qué se puede deducir? ¿Hay una ventaja significativa en contar desde el helicóptero?

2. Los datos siguientes muestran el número de derrames de petróleo en el mar y la cantidad perdida en ellos (en millones de toneladas métricas) para los años 1973-1985

Año	N° de derrames (x)	Cantidad de petróleo derramada (y)
1973	36	84,5
1974	48	67,1
1975	45	188,0
1976	29	204,2
1977	49	213,1
1978	35	260,5
1979	65	723,5
1980	32	135,6
1981	33	45,3
1982	9	1,7
1983	17	387,8
1984	15	24,2
1985	8	15,0

a) Realice una regresión lineal simple con constante $y = \beta_0 + \beta_1 x$ para estos datos. ¿Cómo se interpreta en este caso el valor que Ud. obtuvo para β_0? ¿Cree Ud. que tiene sentido un modelo con constante para estos datos?
b) Ajuste un modelo sin constante $y = \beta x$.
c) Examine ambos modelos y diga cuál elegiría Ud.
d) Interprete los estadísticos Fy R^2en ambos casos.

3. Se desea verficar si el personal de un laboratorio es capaz de detectar correctamente la cantidad de un cierto antibiótico presente en muestras de sangre. Se envían al laboratorio 13 muestras de las cuales se conoce la cantidad de antibiótico presente (x) y se pide al personal medir esta cantidad en cada una de ellas (y). Se obtienen los siguientes datos:

Cantidad presente (x) en $\mu g/ml$	Cantidad hallada (y) en $\mu g/ml$
0	0
5	4,5
5	5,0
5	4,8
10	8,9
10	8,9
10	8,9
20	17
20	18,2
20	15,4
40	32,6
40	36,1
40	31,5

a) Si el laboratorio pudiese detectar exactamente la cantidad de antibiótico presente en las muestras, tendríamos que las varibles x e y serían iguales y, por lo tanto, la recta de regresión debería ser aquella que tiene pendiente 1 y pasa por el origen. Grafique la nube de puntos y diga si esta parece ser la situación en este caso.

b) Ajuste a estos datos un modelo de regresión lineal simple. En base a lo obtenido diga si se cumple la situación planteada en la pregunta anterior.

c) Utilizando los resultados de la parte b) realice la prueba para la hipótesis nula $H_0 : \beta_1 = 1$, frente a la hipótesis $H_a : \beta_1 \neq 1$. ¿cuál es el valor p o nivel de significación de lo observado en este caso?

4. Los auditores a menudo necesitan comparar el valor revisado (o actual) de un artículo del catálogo de inventario con el valor en los libros (o nominal). Si una compañía tiene su inventario y sus libros al día, debe existir una relación lineal muy estrecha entre los valores revisados y los valores nominales. Una muestra de diez artículos del catálogo de cierta compañía dio los datos que contiene la tabla siguiente acerca de los valores revisados y los nominales. Ajuste el modelo $y = \beta_0 + \beta_1 x + \varepsilon$ a estos datos. ¿Cuál es su estimación para el cambio que se espera en el valor revisado para un cambio de una unidad en el valor nominal? Si el valor nominal es $x = 100$, ¿qué utilizaría para estimar el valor revisado?

Artículo	Valor revisado (y)	Valor nominal (x)
1	9	10
2	14	12
3	7	9
4	29	27
5	45	47
6	109	112
7	40	36
8	238	241
9	60	59
10	170	167

5. Se realizó un experimento para observar el efecto del incremento de la temperatura en la eficacia de un antibiótico. Tres porciones de una onza de antibiótico se almacenaron durante periodos iguales en cada una de las siguientes temperaturas Fahrenheit: 30°, 50°, 70° y 90°. Las lecturas de la eficacia observada al final del periodo experimental se muestran en la siguiente tabla.

Lectura de eficacia (y)	38, 43, 29	32, 26, 33	19, 27, 33	14, 19, 21
Temperatura (x)	30°	50°	70°	90°

a) Encuentre la recta de mínimos cuadrados adecuada para estos datos.
b) Represente los puntos y trace una gráfica de la recta para verificar sus cálculos.
c) Calcule S^2.
d) Determine un intervalo de confianza de 95% para la potencia media de una onza de antibiótico almacenada a $65°F$.
e) Determine un intervalo de predicción de 95% para la efectividad de una porción de una onza del antibiótico almacenado a $65°F$. Compare este intervalo con el calculado en la parte d).

6. Las medianas de los precios de venta de casas nuevas parar una sola familia durante un periodo de ocho años se indican en la tabla siguiente. Sea y la mediana de los precios de venta y x el año (representado con números enteros, 1, 2, ..., 8), ajuste el modelo $y = \beta_0 + \beta_1 x + \varepsilon$. ¿Qué se puede concluir con los resultados?

Año	Mediana del precio de venta (x 1000)
1972 (1)	\$ 27.6
1973 (2)	\$ 32.6
1974 (3)	\$ 35.9
1975 (4)	\$ 39.3
1976 (5)	\$ 44.2
1977 (6)	\$ 48.8
1978 (7)	\$ 55.7
1979 (8)	\$ 62.9

a) Calcule el SSE y el $S^2 = \left(\frac{1}{n-2}\right) SSE$ para este ejercicio.

b) A veces es conveniente, desde el punto de vista del cálculo, contar con valores de x separados simétricamente y a la misma distancia de cero. Estos valores de x se pueden reescalar (o codificar) de forma conveniente sin pérdida de información en el análisis estadístico. Codifique los valores de x (originalmente en una escala de 1 a 8) mediante la fórmula:

$$x^* = \frac{x - 4.5}{0.5}$$

En seguida ajuste el modelo $y = \beta_0^* + \beta_1^* x^* + \varepsilon$. Calcule SSE. Compare el valor de SSE con el valor que obtuvo en la parte a).

7. Supongamos que propusimos el modelo

$$y_i = \beta_1 x_i + \varepsilon_i \qquad i = 1, ...n$$

donde ε_i representa variables aleatorias independientes e idénticamente distribuidas con $E(\varepsilon_i) = 0$. De esta manera, $\widehat{y}_i = \widehat{\beta}_1 x_i$ es el valor esperado de y cuando $x = x_i$ y $SSE = \sum_{i=1}^{n} \left[y_i - \widehat{\beta}_1 x_i\right]^2$. Determine el estimador de mínimos cuadrados de β_1.

8. Suponga que $Y_1, ..., Y_n$ son variables aleatorias normales independientes con $E(Y_i) = \beta_0 + \beta_1 x_i$ y $Var(Y_i) = \sigma^2$, para $i = 1, ..., n$. Demuestre que los estimadores de máxima verosimilitud de β_0 y β_1 son iguales a los estimadores de mínimos cuadrados. De acuerdo con estas mismas suposiciones determine el estimador de máxima verosimilitud de σ^2.

9. Ajuste una recta a los cinco datos siguientes. Obtenga las estimaciones de β_0 y β_1. Trace una gráfica de los puntos y represente la recta para verificar los cálculos.

y	x
3	-2
2	-1
1	0
1	1
0.5	2

a) ¿Presentan los datos suficiente evidencia para indicar que la pendiente β_1 difiere de cero? Haga la prueba al nivel de significancia de 5%.

b) Encuentre un intervalo de confianza de 95% para β_1.

10. La siguiente tabla proporciona los datos de la pesca de anchoas peruanas (en millones de toneladas métricas) y los precios de la carne de pescado (en dólares por tonelada) en los años de 1965 a 1978

Variable	1965	1966	1967	1968	1969	1970	1971
Precio de la carne (y)	190	160	134	129	172	197	167
Pesca de anchoas (x)	7.23	8.53	9.82	10.26	8.96	12.27	10.28
Variable	1972	1973	1974	1975	1976	1977	1978
Precio de la carne (y)	239	542	372	245	376	454	410
Pesca de anchoas (x)	4.45	1.78	4.0	3.3	4.3	0.8	0.5

a) Obtenga la recta de mínimos cuadrados adecuada para estos datos.

b) ¿Presentan los datos evidencia suficiente para indicar que el tamaño de la cantidad de anchoas capturadas x contribuye a la información que permite predecir los precios y de la harina del pescado? ¿Qué concluiría con un nivel de significancia de $\alpha = 0.10$?

11. Se llevó a cabo un estudio para determinar cómo afecta la privación del sueño la habilidad de los individuos para resolver problemas sencillos. La cantidad de horas sin dormir variaba entre 8, 12, 16, 20 y 24. Diez individuos participaron en el estudio, dos por cada nivel de privación de sueño. Después del periodo de privación de sueño se asignó a cada individuo un conjunto de problemas sencillos.en los que había que sumar y se registró el número de errores. La siguiente tabla contiene los resultados obtenidos.

Número de errores (y)	8,6	6,10	8,14	14,12	16,12
Número de horas sin dormir (x)	8	12	16	20	24

a) Obtenga la recta de mínimos cuadrados adecuada para estos datos.

b) ¿Presentan los datos evidencia suficiente para indicar que el número de errores se relaciona linealmente con el número de horas sin dormir?

c) ¿Qué concluiría con un nivel de significancia de $\alpha = 0.05$?

d) ¿Esperaría usted una relación lineal entre y y x si variara x en un margen más amplio, digamos, de $x = 4$ a $x = 48$?

e) Obtenga un intervalo de confianza de 95% para la pendiente. Dé una interpretación práctica para esta estimación por intervalo.

12. La siguiente tabla contiene la lista del número de casos de tuberculosis (por cada 100.000 habitantes) en el estado de Florida durante la década que va de 1967 a 1976. ¿Hay evidencia suficiente evidencia para afirmar que la tasa de tuberculosis decrece en tal periodo? Utilice $\alpha = 0.05$. (se puede codificar años de la manera que se considere conveniente)

Años	Casos
1967	26.3
1968	26.1
1969	24.7
1970	22.8
1971	22.1
1972	20.4
1973	19.0
1974	17.7
1975	19.3
1976	17.5

13. En el modelo de regresión lineal simple $y = \beta_0 + \beta_1 x + \varepsilon$, con $E(\varepsilon) = 0$ y $Var(\varepsilon) = \sigma^2$, utilice la expresión $Var\left(a_0\widehat{\beta}_0 + a_1\widehat{\beta}_1\right)$ para demostrar que

$$Var\left(\widehat{\beta}_0 + \widehat{\beta}_1 x^*\right) = \left[\frac{1}{n} + \frac{(x^* - \overline{x})^2}{S_{xx}}\right]\sigma^2$$

¿Para qué valor de x^* el intervalo de confianza de $E(Y)$ alcanza su mínima amplitud?

14. Consulte el ejercicio 10. Determine el intervalo de confianza de 90% para el precio medio por tonelada de alimento de pescado si la pesca de anchoas es de 5 millones de toneladas métricas.

15. Suponga que el modelo $Y = \beta_0 + \beta_1 x + \varepsilon$ se ajusta a los n datos $(y_1, x_1), ..., (y_n, x_n)$. ¿En qué valor de x será mínima la longitud del intervalo de predicción de Y?

16. Utilice los datos y el modelo del ejercicio 7 para construir un intervalo de predicción de 95% para la mediana de los precios de venta en 1980.

17. Matis y Wehrly (1979) elaboraron la siguiente tabla de datos respecto a la proporción de peces de agua dulce que sobrevivieron a niveles fijos de contaminación térmica durante diversos periodos.

Proporción de sobrevivientes (y)	Tiempo con cambio de escala (x)
1.00	0.10
0.95	0.15
0.95	0.20
0.90	0.25
0.85	0.30
0.70	0.35
0.65	0.40
0.60	0.45
0.55	0.50
0.40	0.55

a) Ajuste el modelo lineal $Y = \beta_0 + \beta_1 x + \varepsilon$. Interprete los resultados.

b) Determine un intervalo de predicción de 95% para la proporción de sobrevivientes en el momento $x = 0.60$.

18. El oro radioactivo posee una afinidad hacia el tejido inflamado y se utiliza a veces como marcador para diagnosticar artritis. Los siguientes datos provienen de un experimento donde se investigó el tiempo y las concentraciones en las cuales el oro radioactivo se retiene en el plasma de una persona.

Se listan las concentraciones en plasma de oro radiactivo halladas en 10 muestras de sangre provenientes de 10 pacientes a quienes se les dio una dosis inicial de 50 mg. Las medidas se realizaron entre 1 y 7 días después de la inyección inicial y lo que se midió fue el porcentaje de la concentración inicial presente. Obteniéndose:

Días después de inyectado (x_i)	% de la concentración inicial (y_i)
1	94,5
2	71
1	86,4
2	80,5
2	81,4
3	67,4
5	49,3
6	46,8
6	42,3
7	36,6

a) Ajuste una curva $y = ae^{bx}$.
b) ¿Cuánto tiempo tarda en desaparecer el 90% de la dosis inyectada?

19. La edad a la cual un animal comienza a caminar y la edad en la cual comienza a jugar han sido estudiadas en diversas especies de mamíferos. La siguiente tabla muestra estos datos para 11 especies distintas (edades promedio en cada especie)

Especies	Comienzo de locomoción (x_i) en días	Comienzo de juego (y_i) en días
Homo Sapiens	360	90
Gorilla Gorilla	165	105
Felis Catus	21	21
Canis Familiaris	23	26
Rattus Norvegicus	11	14
Turdus Merula	18	28
Macaca Mulatta	18	21
Pan Troglodytes	150	105
Saimiri Sciurems	45	68
Cercocebus Alb.	45	75
Tamiasciureus Hud.	18	46

a) Ajuste una curva del tipo $y = ax^b$ a estos datos.
b) Grafique los datos y la curva que se ajustó.

20. Es bien sabido que los grandes depósitos de agua tienen un efecto regulador en la temperatura de las masas de tierra que los rodean. En una noche fría en Florida central se registraron las temperaturas en puntos equidistantes a lo largo de una línea recta en la dirección del viento de un gran lago. Los datos

que se obtuvieron aparecen en la siguiente tabla. Observe que las temperaturas disminuyen rápidamente y se nivelan a medida que se está más lejos del lago. El modelo sugerido para estos datos es $y = \alpha_0 e^{-\alpha_1 x}$

a) Convierta el modelo en un modelo lineal y estime los parámetros mediante el método de mínimos cuadrados.

b) Encuentre un intervalo de confianza de 90% para α_0. Interprete el resultado.

Sitio (x)	Temperatura (y)
1	37.00 $^\circ F$
2	36.25 $^\circ F$
3	35.41 $^\circ F$
4	34.92 $^\circ F$
5	34.52 $^\circ F$
6	34.45 $^\circ F$
7	34.40 $^\circ F$
8	34.00 $^\circ F$
9	33.62 $^\circ F$
10	33.90 $^\circ F$