

Practica III

1.

Hertzman *et al.* (A-10) condujeron un estudio para identificar las determinantes que provocan niveles elevados de plomo en la sangre de niños de edad preescolar; compare la situación actual con información anterior, determine la tendencia histórica en la contaminación ambiental por plomo en una comunidad canadiense, y encuentre las bases para identificar las precauciones adecuadas y medidas de protección contra futura exposición al plomo. Los individuos estudiados eran niños con edades entre dos y cinco años, inclusive, que vivían en una comunidad canadiense, un lugar donde existe una de las más grandes fundidoras de plomo y zinc de América del Norte. Los individuos fueron repartidos en dos grupos: 1) casos o grupo de niños con niveles de plomo en la sangre de $18 \mu\text{g/ml}$ o más, y 2) controles o grupo de niños con niveles de plomo en la sangre de $10 \mu\text{g/dl}$ o menos. Los niveles de plomo fueron evaluados en muestras de agua para beber, pinturas, polvo doméstico, vegetales cultivados en casa, y tierra. Entre los análisis aplicados por los investigadores estaba el análisis de regresión múltiple logístico, cuyas variables independientes eran sexo, edad y los logaritmos de los niveles de plomo en las muestras del ambiente (covariancias). Encontraron que el nivel de plomo en la tierra era el factor de riesgo mayor para los niveles altos de plomo en la sangre. El análisis produjo una relación de 14.25 para los grados de probabilidad, los cuales se pueden interpretar como “por cada diez veces que se incrementa el nivel de plomo en la tierra, aumenta 14.25 veces la proporción relativa de casos con respecto a los grupos de control”. La siguiente tabla

muestra los niveles de plomo en la tierra para el grupo 1 (codificado como 1) y el grupo 2 (codificado como 0). Utilice la regresión logística para analizar estos datos. Obtenga la razón de grados de probabilidad y compárela contra la que obtuvieron los autores. Realice las pruebas con nivel de significación de .05 y calcule el valor de p .

Categoría del individuo	Nivel de plomo en el suelo (ppm)	Categoría del individuo	Nivel de plomo en el suelo (ppm)	Categoría del individuo	Nivel de plomo en el suelo (ppm)
1 = caso	1290	0	197	1	852
0 = control	90	1	916	0	137
1	894	1	755	0	137
0	193	1	59	0	125
1	1410	1	1720	1	562
1	410	1	574	0	325
1	1594	1	403	1	1317
0	321	1	61	1	2125
0	40	1	1290	1	2635
0	96	1	1409	1	2635
0	260	1	880	0	544
0	433	1	40	1	731
0	260	1	40	1	815
0	227	0	68	0	328
0	337	0	777	1	1455
1	867	1	1975	0	977
1	1694	1	1237	1	624
0	302	0	133	1	392
1	2860	0	269	1	427
1	2860	0	357	1	1000
1	4320	0	315	1	1009
1	859	0	315	0	1010
0	119	0	255	1	3053
1	115	0	422	0	1220
0	192	0	400	0	46
1	1345	0	400	0	181
0	55	0	229	0	87
0	55	1	229	0	131
1	606	0	768	0	131
1	1660	0	886	1	1890
0	82	1	58	1	221
0	1470	0	508	1	221
1	600	1	811	0	79
1	2120	1	527	1	1570
1	569	1	1753	1	909
0	105	0	57	1	1720
1	503	0	769	1	308

(Continúa)

Categoría del individuo	Nivel de plomo en el suelo (ppm)	Categoría del individuo	Nivel de plomo en el suelo (ppm)	Categoría del individuo	Nivel de plomo en el suelo (ppm)
0	161	0	677	1	97
0	161	1	677	0	200
1	1670	1	424	0	1135
0	132	1	2230	0	320
1	974	0	421	1	5255
1	3795	1	628	0	176
0	548	1	1406	0	176
1	622	1	378	0	100
0	788	1	812		
1	2130	1	812		

FUENTE: Utilizada con autorización de Shona Kelly.

- ¿Cuál es la variable dependiente?
- ¿Cuáles son las variables independientes?
- ¿Cuáles son las hipótesis nula y alternativa adecuadas?
- ¿Cuál de las hipótesis nulas es rechazada y por qué?
- ¿Cuál es el objetivo más relevante, la predicción o la estimación? ¿Son ambos igual de importantes? Explique su respuesta.
- ¿Cuál es la población muestreada?
- ¿Cuál es la población objetivo?
- ¿Cuáles variables están relacionadas con otras variables? ¿La relación es directa o inversa?
- Describa la ecuación de regresión mediante números adecuados para la estimación de parámetros.
- Asigne valores numéricos para cualquier otra estadística accesible.
- Especifique para cada variable si es cuantitativa o cualitativa.
- Explique el significado de cualquier estadística para la que se dan los valores numéricos.

2.

En la Tabla se recogen las siguientes variables medidas sobre 30 olmos hembra.

	nombre	unidades	breve descripción
X_1	Longitud	mm	mayor medida de la corteza
X_2	Diámetro	mm	perpendicular a la longitud
X_3	Altura	mm	con madera dentro de la corteza
X_4	Peso total	g	todo el olmo
X_5	Peso desvainado	g	peso de la madera
X_6	Peso de las vísceras	g	peso de la tripa (después de sangrar)
X_7	Peso de la corteza	g	después de ser secado

X1	X2	X3	X4	X5	X6	X7
0.53	0.42	0.135	0.677	0.2565	0.1415	0.21
0.53	0.415	0.15	0.7775	0.237	0.1415	0.33
0.545	0.425	0.125	0.768	0.294	0.1495	0.26
0.55	0.44	0.15	0.8945	0.3145	0.151	0.32
0.525	0.38	0.14	0.6065	0.194	0.1475	0.21
0.535	0.405	0.145	0.6845	0.2725	0.171	0.205
0.47	0.355	0.1	0.4755	0.1675	0.0805	0.185
0.44	0.34	0.1	0.451	0.188	0.087	0.13
0.565	0.44	0.155	0.9395	0.4275	0.214	0.27
0.55	0.415	0.135	0.7635	0.318	0.21	0.2
0.615	0.48	0.165	1.1615	0.513	0.301	0.305
0.56	0.44	0.14	0.9285	0.3825	0.188	0.3
0.58	0.45	0.185	0.9955	0.3945	0.272	0.285
0.68	0.56	0.165	1.639	0.6055	0.2805	0.46
0.68	0.55	0.175	1.798	0.815	0.3925	0.455
0.705	0.55	0.2	1.7095	0.633	0.4115	0.49
0.54	0.475	0.155	1.217	0.5305	0.3075	0.34
0.45	0.355	0.105	0.5225	0.237	0.1165	0.145
0.575	0.445	0.135	0.883	0.381	0.2035	0.26
0.45	0.335	0.105	0.425	0.1865	0.091	0.115
0.55	0.425	0.135	0.8515	0.362	0.196	0.27
0.46	0.375	0.12	0.4605	0.1775	0.11	0.15
0.525	0.425	0.16	0.8355	0.3545	0.2135	0.245
0.47	0.36	0.12	0.4775	0.2105	0.1055	0.15
0.5	0.4	0.14	0.6615	0.2565	0.1755	0.22
0.505	0.4	0.125	0.583	0.246	0.13	0.175
0.53	0.41	0.13	0.6965	0.302	0.1935	0.2
0.565	0.44	0.16	0.915	0.354	0.1935	0.32
0.595	0.495	0.185	1.285	0.416	0.224	0.485
0.475	0.39	0.12	0.5305	0.2135	0.1155	0.17

- Realizar el acp completo con gráficos de cargas.
- Cuales variables presentan más similitud.
- Cuantos componentes explican la mayor variabilidad de los datos.
- Que grupos de individuos presentan más similitud.
- Realizar un análisis de regresión lineal del componente 1 y sus respectivas variables de carga.

3.

Gases contaminantes

En la Tabla siguiente se presentan las 10 primeras observaciones de cinco variables de niveles de gases contaminantes (CO: X_3 , NO: X_4 , NO₂: X_5 , O₃: X_6 , y HC: X_7) y dos variables relacionadas (Intensidad del viento: X_1 , y Radiación solar: X_2).

X_1	X_2	X_3	X_4	X_5	X_6	X_7
8	98	7	2	12	8	2
7	107	4	3	9	5	3
7	103	4	3	5	6	3
10	88	5	2	8	15	4
6	91	4	2	8	10	3
8	90	5	2	12	12	4
9	84	7	4	12	15	5
5	72	6	4	21	14	4
7	82	5	1	11	11	3
8	64	5	2	13	9	4

- a. Calcular la covarianza.
- b. Calcular la matriz de correlación.
- c. Calcular las distancias euclidianas y mahalanobis.

4.

Una muestra formada por 500 estudiantes de primaria fueron clasificados de manera cruzada de acuerdo con el estado de nutrición y desempeño escolar de la siguiente manera:

Estado de nutrición y desempeño escolar de 500 estudiantes de primaria

Desempeño escolar	Estado de nutrición		Total
	Deficiente	Bueno	
Deficiente	105	15	120
Satisfactorio	80	300	380
Total	185	315	500

El ejercicio consiste en utilizar el análisis de regresión logística para calcular los coeficientes de regresión y estimar la razón de los grados de probabilidad, así como redactar la interpretación de los resultados obtenidos.

5. En Colombia existen 3 operadores principales de telefonía móvil como lo son; Tigo, claro y movistar. Los porcentajes actuales que tiene cada operador en el mercado actual son para Tigo 0.4 para Claro 0.25 y para movistar 0.35. Se tiene la siguiente información un usuario actualmente de Tigo tiene una probabilidad de permanecer en Tigo de 0.60, de pasar a Claro 0.2 y de pasarse a movistar de 0.2; si en la actualidad el usuario es cliente de Claro tiene una probabilidad de mantenerse en Claro del 0.5 de que esta persona se cambie a Tigo 0.3 y que se pase a movistar de 0.2; si el usuario es cliente en la actualidad de movistar la probabilidad que permanezca en movistar es de 0.4 de que se cambie a Tigo de 0.3 y a Claro de 0.3.
 - a. Se quiere conocer cuál es la probabilidad de permanencia en la primera y segunda corrida.
 - b. Realizar el grafico

