Estadística II
Grado en Matemáticas, UAM, 2019-2020

Examen final, convocatoria ordinaria, 20-1-2020

Ejercicio 1. Un vector $\mathbb{X} = (X_1, X_2)$ se distribuye, en dos poblaciones π_0 y π_1, como se indica a continuación:

- En π_0, las variables X_1 y X_2 son dos normales estándar independientes.
- En π_1, el vector $\mathbb{X}$ se distribuye uniformemente en el rectángulo centrado en el origen cuyos lados vertical y horizontal miden 2 y $e\pi$, respectivamente.

Ponemos que las probabilidades a priori de cada población son iguales. Identifica (y dibuja con precisión) las regiones óptimas R_0 y R_1 de clasificación en π_0 y π_1, respectivamente.

Ejercicio 2. Una cierta característica en una población tiene cuatro niveles, que designamos por C_1, C_2, C_3 y C_4.

En una muestra de 100 individuos de la población se han encontrado 8, 28, 10 y 54, respectivamente, de cada nivel.

Se desea contrastar un modelo teórico que asigna probabilidades $1/4 - p_1$, $1/4 + p_1$, $1/4 - p_2$ y $1/4 + p_2$, respectivamente, a cada categoría. Aquí $0 \leq p_1, p_2 \leq 1/4$.

Escribe las conclusiones que obtienes (sobre el modelo y la muestra) *argumentando con el p-valor* de la muestra.

Ejercicio 3. Suponemos que una variable respuesta Y depende linealmente de una única variable regresora X. La muestra va a ser de tamaño n, del tipo $(x_1, y_1), \ldots, (x_n, y_n)$. Proponemos, pues, el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \text{para } i = 1, \ldots, n,$$

donde las ε_i son variables aleatorias idénticas e independientes, cada una de las cuales se distribuye con la siguiente función de densidad:

$$f(x) = \begin{cases} x + 1 & \text{si } x \in [-1, 0], \\ 1 - x & \text{si } x \in (0, 1]. \end{cases}$$

Los parámetros del modelo son $\beta_0, \beta_1 \in \mathbb{R}$.

Como estimador de β_0 elegimos el habitual

$$\widehat{\beta}_0 = \overline{Y} - \frac{\overline{x}}{V_x} \operatorname{cov}_{x,Y}.$$

de mínimos cuadrados.

a) Comprueba si $\widehat{\beta}_0$ es un estimador insesgado de β_0 para muestras (x_i, Y_i) de tamaño n.

b) Calcula la varianza de $\widehat{\beta}_0$ para muestras (x_i, Y_i) de tamaño n.

Ejercicio 4. El vector $\mathbb{X} = (X_1, X_2)^\intercal$ sigue una normal bidimensional con vector de medias $\mu = (1, -\sqrt{2}/2)^\intercal$ y matriz de covarianzas $V = 3 \cdot I_{2\times 2}$. Considera la matriz

$$B = \begin{pmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{pmatrix}.$$

Determina cómo se distribuye la variable aleatoria

$$Z = \frac{1}{3}\, \mathbb{X}^\intercal \cdot B \cdot \mathbb{X}.$$

Ejercicio 5. Se plantea un modelo de regresión lineal múltiple con dos variables regresoras, X_1 y X_2, para muestras de tamaño n:

$$Y_i = \beta_0 + \beta_1\, x_{i,1} + \beta_2\, x_{i,2} + \varepsilon_i, \quad \text{para } i = 1, \ldots, n,$$

donde las ε_i son variables normales independientes, de media 0 y varianza σ^2.

Se dispone de la siguiente muestra de tamaño $n = 4$:

X_1	X_2	Y
1	2	3
2	1	2
−1	1	0
1	3	2

Llamando X a la matriz de diseño, se han calculado los siguientes productos matriciales:

$$(X^\intercal \cdot X)^{-1} = \frac{1}{50}\begin{pmatrix} 69 & -3 & -31 \\ -3 & 11 & -3 \\ -31 & -3 & 19 \end{pmatrix}, \quad (X^\intercal \cdot X)^{-1} \cdot X^\intercal = \frac{1}{50}\begin{pmatrix} 4 & 32 & 41 & -27 \\ 2 & 16 & -17 & -1 \\ 4 & -18 & -9 & 23 \end{pmatrix}$$

Además, se ha obtenido que $R^2 = 0.730526$.

a) Halla un intervalo de confianza al 95 % para el parámetro β_1.

b) Se dispone de una nueva observación $(1, 1)$. Se pide calcular cuántas veces es mayor

- la longitud del intervalo de confianza, al 95 %, para predecir el *valor* de la variable respuesta que correspondería a esa observación,
- que la longitud del intervalo de confianza, al 95 %, para predecir el *valor medio* de la variable respuesta que correspondería a esa observación.

1. Percentiles de la t de Student

La siguiente tabla contiene valores de percentiles $t_{\{\nu;\alpha\}}$, redondeados a tres cifras decimales, donde los grados de libertad ν etiquetan las columnas, y los valores de α (en porcentajes) etiquetan filas:

	1	2	3	4
20 %	1.376	1.061	0.978	0.941
10 %	3.078	1.886	1.638	1.533
5 %	6.314	2.920	2.353	2.132
4 %	7.916	3.320	2.605	2.333
3 %	10.579	3.896	2.951	2.601
2.50 %	12.706	4.303	3.182	2.776
2.00 %	15.895	4.849	3.482	2.999
1.50 %	21.205	5.643	3.896	3.298
1.00 %	31.821	6.965	4.541	3.747
0.50 %	63.657	9.925	5.841	4.604
0.10 %	318.309	22.327	10.215	7.173
0.05 %	636.619	31.599	12.924	8.610
0.001 %	31830.989	223.603	47.928	23.332

1. Percentiles de la χ^2

La siguiente tabla contiene valores de percentiles $\chi^2_{\{\nu;\alpha\}}$, redondeados a tres cifras decimales, donde los grados de libertad ν etiquetan las columnas, y los valores de α (en porcentajes) etiquetan filas:

	1	2	3	4
20 %	1.642	3.219	4.642	5.989
10 %	2.706	4.605	6.251	7.779
5 %	3.841	5.991	7.815	9.488
4 %	4.218	6.438	8.311	10.026
3 %	4.709	7.013	8.947	10.712
2.50 %	5.024	7.378	9.348	11.143
2.00 %	5.412	7.824	9.837	11.668
1.50 %	5.916	8.399	10.465	12.339
1.00 %	6.635	9.210	11.345	13.277
0.50 %	7.879	10.597	12.838	14.860
0.10 %	10.828	13.816	16.266	18.467
0.05 %	12.116	15.202	17.730	19.997
0.001 %	19.511	23.026	25.902	28.473