Examen-Junio-2022-Estadistica.pdf

ApuntesDeConfi_

Metodos Estadisticos para la Computacion

1º Grado en Ingeniería Informática

Escuela Técnica Superior de Ingeniería Informática
Universidad de Málaga

Métodos Estadísticos para la Computación Junio 2022

APELLIDOS, NOMBRE: ____________________

GRADO: __________ GRUPO: ____ DNI: __________

1. (1 punto) Considera las variables estadísticas: X = rentabilidad, en tanto por ciento, de las acciones de Macdonalds e Y = tiempo, en años, que un accionista tarda en vender. Las rectas de regresión de X e Y son $6y + 5x = 7$ y $3y + 2x = 4$, donde la variable x representa a la variable estadística X e y, a Y.

    a) ¿Cuál es la recta de regresión de Y sobre X? ¿Y la de X sobre Y?

    b) Tengo pensado mantener las acciones de Macdonalds durante 3 años. ¿Qué predicción puedes hacer sobre la rentabilidad?

    c) Sabiendo que $\sigma_Y^2 = 5$, calcula la varianza de X, la covarianza y el coeficiente de correlación lineal.

2. (1 punto) La variable bidimensional (X,Y) toma los siguientes 4 valores $(1, 1)$, $(2, 4)$, $(3, 9)$, $(4, 16)$. Calcula el Error Cuadrático Medio (MSE) para estos dos modelos lineales y determina en base a eso cuál de los dos modelos es mejor:

$$y = -6 + 6 * x \qquad y = -5 + 5 * x$$

3. (1,5 puntos) Un laboratorio produce dos medicamentos A y B para la misma afección. En el primero de ellos se reportan efectos secundarios en el 3 % de los casos, mientras que esta tasa aumenta a un 10 % para el medicamento B. Si se estima que de cada tres pacientes en tratamiento, dos toman el medicamento A,

    a) (0,5 puntos) Calcula la probabilidad de que un paciente en tratamiento, elegido aleatoriamente, reporte efectos secundarios.

    b) (1 puntos) Si se eligen aleatoriamente 40 pacientes que toman el mismo medicamento, y nos aseguran que entre ellas hay menos de dos personas con efectos secundarios, ¿cuál es la probabilidad de que estuvieran tomando el medicamento A?.

4. (2 puntos) De cierta variable aleatoria se sabe que su mediana es 3/2, y que su función de densidad es la siguiente:

$$f(x) = \begin{cases} 0 & si \quad x \le 0 \\ x^2 & si \quad 0 < x \le 1 \\ a + bx & si \quad 1 < x \le 3 \\ c & si \quad 3 < x \end{cases}$$

    a) (1 punto) Determina los valores $a, b, c \in \mathbb{R}$. Determina su esperanza.

    b) (1 punto) Realizamos el experimento de elegir aleatoriamente un número real X siguiendo esta distribución de probabilidad. Consideramos la variable aleatoria N = *número de puntos obtenidos* según la siguiente regla: **3 puntos si $X \le 1$, 2 puntos si $1 < X \le 2$ y 1 punto si $X > 2$.** Calcula la función de distribución de N, determina su esperanza y su moda.

5. (1,5 puntos) Una empresa produce dos tipos de balones esféricos (A y B). El radio de los de tipo A se distribuye según una normal de media 20 cm y varianza 16 cm^2, y los de tipo B según una normal de media 16 cm y desviación típica 3 cm. El 70 % de los balones que produce la empresa son del tipo A y el resto de tipo B.

a) (0,5 puntos) Calcular la probabilidad de que el radio de un balón de tipo A mida $3cm$ más que el radio de un balón de tipo B.

b) (0,5 puntos) Calcular la probabilidad de que el radio de un balón escogido al azar de esta empresa mida menos de 22 cm

c) (0,5 puntos) Control de calidad escoge 10 balones de la empresa. ¿Cuál es la probabilidad de que al menos 2 de ellos midan más de 22 cm de radio?

6. (1,5 puntos) Se han medido la altura de 50 chimpancés de una especie A y se ha obtenido una media de $100cm$ y una varianza de $98cm^2$ y la altura de 55 chimpancés de una especie B y se ha obtenido una media de $103cm$ y una varianza de $118.8cm^2$.

a) (1 punto) Hallar un intervalo de confianza al 95 % para la diferencia de medias entre las dos alturas, bajo la hipótesis de normalidad de los datos.

b) (0,5 puntos) A la vista de lo anterior, ¿podríamos afirmar con esa confianza que la media de altura de la especie B es mayor que la media de altura de la especie A?

7. (1,5 puntos) Una empresa de estudios de mercado lleva a cabo un sondeo en Madrid y Barcelona. Se pregunta a residentes de ambas cidudades por su marca de smartphone favorito, entre 3 conocidas marcas: Huawei, Samsung y Xiaomi. Este es el resultado:

$Ciudad\backslash Marca$	Huawei	Samsung	Xiaomi
Madrid	150	210	110
Barcelona	200	200	130

Proporciona una de estas dos informaciones (**sólo una de ellas, la que prefieras**):

a) Contrasta con un nivel de significación del $2,5\,\%$ si la elección de marca preferida de smartphone está relacionada con la ciudad en la que residen los encuestados.

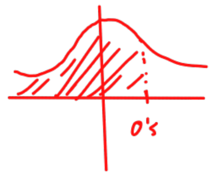
b) Proporciona el p-value de testear si la elección de marca preferida de smartphone está o no relacionada con la ciudad en la que residen los encuestados. (Es decir, la probabilidad de que haya ocurrido lo que ha ocurrido si no están relacionadas).

DATOS ÚTILES:

Función de distribución de una normal tipificada (media 0, desviación típica 1)

colas a la izquierda →

0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2
0.5793	0.6554	0.7257	0.7881	0.8413	0.8849	0.9192	0.9452	0.9641	0.9772



Función cuantil de una normal tipificada

0.9	0.925	0.95	0.975
1.2816	1.4395	1.6449	1.9600

Función de distribución de una χ^2

	grados de libertad 1	2	3	4	5	6
4	0.9545	0.8647	0.7385	0.5940	0.4506	0.3233
4.5	0.9661	0.8946	0.7877	0.6575	0.5201	0.3907
5	0.9747	0.9179	0.8282	0.7127	0.5841	0.4562
5.5	0.9810	0.9361	0.8614	0.7603	0.6421	0.5185
6	0.9857	0.9502	0.8884	0.8009	0.6938	0.5768

Función cuantil de una χ^2

cuantil	grados de libertad 1	2	3	4	5	6
0.025	0.0010	0.0506	0.2158	0.4844	0.8312	1.2373
0.05	0.0039	0.1026	0.3518	0.7107	1.1455	1.6354
0.95	3.8415	5.9915	7.8147	9.4877	11.0705	12.5916
0.975	5.0239	7.3778	9.3484	11.1433	12.8325	14.4494


FNAC.ES
FNAC.ES
QUE PUEDAS RESERVAR UN DISCO ANTES DE SU LANZAMIENTO
FNAC.ES
TIENE SU PUNTO ES
DESCÚBRELO Y COMPRA EN FNAC.ES
fnac
TE FALTA FNAC | FNAC.ES

5. (1,5 puntos) Una empresa produce dos tipos de balones esféricos (A y B). El radio de los de tipo A se distribuye según una normal de media 20 cm y varianza 16 cm^2, y los de tipo B según una normal de media 16 cm y desviación típica 3 cm. El 70 % de los balones que produce la empresa son del tipo A y el resto de tipo B.

a) (0,5 puntos) Calcular la probabilidad de que el radio de un balón de tipo A mida $3cm$ más que el radio de un balón de tipo B.

b) (0,5 puntos) Calcular la probabilidad de que el radio de un balón escogido al azar de esta empresa mida menos de 22 cm

c) (0,5 puntos) Control de calidad escoge 10 balones de la empresa. ¿Cuál es la probabilidad de que al menos 2 de ellos midan más de 22 cm de radio?

$X_A \rightsquigarrow N(20,4)$ $\qquad X_B \rightsquigarrow N(16,3)$

$P(A) = 0'7$

a.) $P(X_A - X_B > 3) = P(D > 3) = P(Z > \frac{3-4}{5}) = P(Z < 0'2) = 0'5793$

$D \rightsquigarrow N(20-16, \sqrt{4^2+3^2}) = N(4,5)$

la probabilidad de que A mida 3cm más que B es del 57'93 %

b.) $P(X < 22) = P(X_A < 22)\, P(A) + P(X_B < 22)\, P(B) =$

$= P(Z < \frac{22-20}{4})\, P(A) + P(Z < \frac{22-16}{3})\, P(B) = P(Z < 0'5)\, P(A) + P(Z < 2)\, P(B) =$

$= 0'6905 \cdot 0'7 + 0'9772 \cdot 0'3 = 0'7765.$

la probabilidad que que el radio de un balón cogido al azar sea menor de 22 cm es del 77'65 %

c.) $Y \rightsquigarrow B(10, P(X > 22)) = B(10, 0'2235)$

$P(X > 22) = 1 - P(X < 22) = 1 - 0'7765 = 0'2235$

$P(Y \geq 2) = 1 - P(Y=0) - P(Y=1) = 1 - \binom{10}{0} 0'2235^0 (1-0'2235)^{10} - \binom{10}{1} 0'2235^1 (1-0'2235)^9 =$

$= 0'691$

la probabilidad de que haya más de dos balones que midan más de 22 es del 77'65 %

6. (1,5 puntos) Se han medido la altura de 50 chimpancés de una especie A y se ha obtenido una media de $100cm$ y una varianza de $98cm^2$ y la altura de 55 chimpancés de una especie B y se ha obtenido una media de $103cm$ y una varianza de $118.8cm^2$.

   a) (1 punto) Hallar un intervalo de confianza al 95 % para la diferencia de medias entre las dos alturas, bajo la hipótesis de normalidad de los datos.

   b) (0,5 puntos) A la vista de lo anterior, ¿podríamos afirmar con esa confianza que la media de altura de la especie B es mayor que la media de altura de la especie A?

$n_A = 50 \quad \bar{X}_A = 100 \quad \bar{\sigma}_A^2 = 98 \quad \bar{S}_A^2 = 100$

$n_B = 55 \quad \bar{X}_B = 103 \quad \bar{\sigma}_B^2 = 118'8 \quad \bar{S}_B^2 = 121$

$$I_{1-\alpha} = \left[(\bar{X}_B - \bar{X}_A) \pm Z_{\alpha/2} \cdot \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \right]$$

$0'95 = 1-\alpha \; ; \quad \alpha = 0'05 \qquad Z_{0'025} = 1'96$

$$I_{0'95} = \left[(103-100) \pm 1'96 \cdot \sqrt{\frac{100}{50} + \frac{121}{55}} \right] = \left[3 \pm 4'017 \right] =$$

a.) $\boxed{I_{0'95} = \left[-1'017 \; , \; 7'017 \right]}$

b.) No podríamos afirmarlo ya que el intervalo tiene un tramo en zona negativa lo que implica que X_B puede ser menor X_A.

7. (1,5 puntos) Una empresa de estudios de mercado lleva a cabo un sondeo en Madrid y Barcelona. Se pregunta a residentes de ambas ciudades por su marca de smartphone favorito, entre 3 conocidas marcas: Huawei, Samsung y Xiaomi. Este es el resultado:

*Ciudad**Marca*	Huawei	Samsung	Xiaomi	
Madrid	150	210	110	470
Barcelona	200	200	130	530
	350	410	240	1000

Proporciona una de estas dos informaciones (**sólo una de ellas, la que prefieras**):

a) Contrasta con un nivel de significación del $2,5\,\%$ si la elección de marca preferida de smartphone está relacionada con la ciudad en la que residen los encuestados.

b) Proporciona el p-value de testear si la elección de marca preferida de smartphone está o no relacionada con la ciudad en la que residen los encuestados. (Es decir, la probabilidad de que haya ocurrido lo que ha ocurrido si no están relacionadas).

H_0: son independientes

Busco la tabla e_i:

$$p_{ij} = \frac{O_{i\cdot} \quad O_{\cdot j}}{n \cdot n} \qquad e_{ij} = \frac{O_{i\cdot} \quad O_{\cdot j}}{n}$$

Marca \\ Ciudad	H	S	X	
M	164'5	192'7	112'8	470
B	185'5	217'3	127'2	430

Construyo el estimador:

$$\hat{\chi}^2 = \sum\sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}} = 5'47$$

$$\chi^2_{0'025,(3-1)(2-1)} = \chi^2_{0'025,2} = 7'378$$

Como $\hat{\chi}^2 < \chi^2_{0'025,2} \Rightarrow H_0$ es cierta: son independientes

Son independientes al 2'5 %

4. (2 puntos) De cierta variable aleatoria se sabe que su mediana es $3/2$, y que su función de densidad es la siguiente:

$$f(x) = \begin{cases} 0 & si \quad x \leq 0 \\ x^2 & si \quad 0 < x \leq 1 \\ a + bx & si \quad 1 < x \leq 3 \\ c & si \quad 3 < x \end{cases}$$

a) (1 punto) Determina los valores $a, b, c \in \mathbb{R}$. Determina su esperanza.

b) (1 punto) Realizamos el experimento de elegir aleatoriamente un número real X siguiendo esta distribución de probabilidad. Consideramos la variable aleatoria $N =$ *número de puntos obtenidos* según la siguiente regla: **3 puntos si** $X \leq 1$, **2 puntos si** $1 < X \leq 2$ **y 1 punto si** $X > 2$. Calcula la función de distribución de N, determina su esperanza y su moda.

a.) $\boxed{c=0}$ ya que si fuese una constante $\int_{-\infty}^{\infty} f(x)\,dx \neq 1$

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 x^2\,dx + \int_1^3 a+bx\,dx = \left.\frac{x^3}{3}\right]_0^1 + \left. ax + \frac{b}{2}x^2 \right]_1^3 =$$

$$= \frac{1}{3} + \left[\left(3a + \frac{9}{2}b\right) - \left(a + \frac{b}{2}\right)\right] = 2a + 4b + \frac{1}{3} = 1$$

$$\boxed{2a + 4b = 2/3}$$

$$0'5 = \int_{-\infty}^{Me} f(x)\,dx = \int_0^1 x^2\,dx + \int_1^{3/2} a+bx\,dx = \frac{1}{3} + \left[ax + \frac{b}{2}x^2\right]_1^{3/2} =$$

$$= \frac{1}{3} + \left(\frac{3}{2}a + \frac{9}{8}b\right) - \left(a + \frac{b}{2}\right) = \frac{1}{2}a + \frac{5}{8}b + \frac{1}{3} = 0'5$$

$$\boxed{1/2\, a + 5/8\, b = 1/6}$$

$$\boxed{a = \frac{1}{3} \qquad b=0 \qquad c=0}$$

$$f(x) = \begin{cases} 0 & si \quad x \leq 0 \\ x^2 & si \quad 0 < x \leq 1 \\ a + bx & si \quad 1 < x \leq 3 \\ c & si \quad 3 < x \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^1 x^3\,dx + \int_1^3 1/3\, x\,dx = \left. 1/4\, x^4 \right]_0^1 + \left. 1/6\, x^2 \right]_1^3 =$$

$$= \frac{1}{4} - 0 + \frac{3}{2} - \frac{1}{6} = \frac{19}{12}$$

$$\boxed{E(X) = \frac{19}{12}}$$

b.)

$$N = \begin{cases} 3 & x \leq 1 \\ 2 & 1 < x \leq 2 \\ 1 & 2 < x \end{cases} \qquad F(N) = \begin{cases} 0 & n < 1 \\ 1/3 & 1 \leq n < 2 \\ 1/3 + 1/3 & 2 \leq n < 3 \\ 1 & 3 \leq n \end{cases}$$

$$P(N=1) = P(x>2) = \int_2^{\infty} f(x)\,dx = \int_2^3 1/3\,dx = \frac{1}{3}x \Big]_2^3 = 1 - \frac{2}{3} = \frac{1}{3}$$

$$P(N=2) = P(1<x\leq 2) = \int_1^2 f(x)\,dx = \int_1^2 \frac{1}{3}\,dx = 1/3\,x \Big]_1^2 = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}$$

$$E(N) = 1 \cdot P(N=1) + 2 \cdot P(N=2) + 3 \cdot P(N=3) = 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = 2 \,//$$