



PROBLEMA 1

(2'5 puntos)

Una empresa de electrodomésticos localiza la venta de sus productos en 2 ciudades: Valencia y Madrid. La mitad de los productos que vende por año en Madrid son frigoríficos, al igual que también lo son la mitad de los que vende por año en Valencia. Además, se sabe que también vende lavadoras y lavavajillas, a razón de 2 lavadoras por cada lavavajillas en Madrid; siendo en Valencia de 2 lavavajillas por cada lavadora. La empresa no vende ningún otro producto. De los productos que vende en Madrid, se sabe que en un 80% de los frigoríficos, un 90% de las lavadoras y un 75% de los lavavajillas son de color blanco. Respecto a Valencia, el 30% de los frigoríficos, el 25% de las lavadoras y el 40% de los lavavajillas no son de color blanco.

Sepa que el 70% de sus ventas tienen lugar en Madrid.

- Calcule la probabilidad de que un producto elegido al azar vendido por esta empresa sea de color blanco.
- Suponga que se selecciona un producto de esta empresa al azar, de los vendidos en Madrid. Si resulta que es de color blanco, ¿cuál es la probabilidad de que sea una lavadora? Indique si el hecho de saber que fuera blanco, ha influido en la probabilidad de que fuese una lavadora, a la vista del resultado obtenido, justificando la respuesta.
- Si nos centramos en los productos de esta empresa que tienen destino Madrid ese año, y seleccionamos un producto al azar, ¿cuál es la probabilidad de que sea un frigorífico, si se sabe que no es blanco?
- Calcule la probabilidad de que un producto vendido por esta empresa, elegido al azar, sea blanco o su venta haya sido en Valencia.

SOLUCIÓN:

B= tener color blanco	a_1 =frigorífico-Madrid	a'_1 =frigorífico-Valencia
A_1 =destino Madrid	a_2 =lavadora-Madrid	a'_2 =lavadora-Valencia
A_2 =destino Valencia	a_3 =lavavajillas-Madrid	a'_3 =lavavajillas-Valencia

a) Aplicamos el teorema de Probabilidad Total:

$$\begin{aligned}
 P(B) &= P(A_1)P\left(\frac{B}{A_1}\right) + P(A_2)P\left(\frac{B}{A_2}\right) = \\
 &= P(A_1)P\left[P(a_1)P\left(\frac{B}{a_1}\right) + P(a_2)P\left(\frac{B}{a_2}\right) + P(a_3)P\left(\frac{B}{a_3}\right)\right] + \\
 &+ P(A_2)P\left[P(a'_1)P\left(\frac{B}{a'_1}\right) + P(a'_2)P\left(\frac{B}{a'_2}\right) + P(a'_3)P\left(\frac{B}{a'_3}\right)\right] = \\
 &= 0'7\left(\frac{1}{2}0'8 + \frac{1}{3}0'9 + \frac{1}{6}0'75\right) + 0'3\left(\frac{1}{2}0'7 + \frac{1}{6}0'75 + \frac{1}{3}0'6\right) = \boxed{0'78}
 \end{aligned}$$



APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

CURSO: 3º GRUPO: _____

PROBLEMA 1: Continuación

b) Aplicamos el teorema de Bayes:

$$P\left(\frac{a_2}{B_{MADRID}}\right) = \frac{P(a_2)P\left(\frac{B_{MADRID}}{a_2}\right)}{P(a_1)P\left(\frac{B_{MADRID}}{a_1}\right) + P(a_2)P\left(\frac{B_{MADRID}}{a_2}\right) + P(a_3)P\left(\frac{B_{MADRID}}{a_3}\right)} =$$
$$= \frac{\frac{1}{3} \cdot 0'9}{\frac{1}{2} \cdot 0'8 + \frac{1}{3} \cdot 0'9 + \frac{1}{6} \cdot 0'75} = \boxed{0'3636}$$

Existe dependencia, ya que la probabilidad condicionada es distinta a la probabilidad sin condicionar.

Además, la dependencia es favorable, ya que el hecho de saber que sale blanco en Madrid incrementa la probabilidad de que se trate de una lavadora.

$$c) \quad P\left(\frac{a_1}{B_{MADRID}^*}\right) = \frac{P(a_1)P\left(\frac{B_{MADRID}^*}{a_1}\right)}{P(a_1)P\left(\frac{B_{MADRID}^*}{a_1}\right) + P(a_2)P\left(\frac{B_{MADRID}^*}{a_2}\right) + P(a_3)P\left(\frac{B_{MADRID}^*}{a_3}\right)} =$$
$$= \frac{\frac{1}{2} \cdot 0'2}{\frac{1}{2} \cdot 0'2 + \frac{1}{3} \cdot 0'1 + \frac{1}{6} \cdot 0'25} = \boxed{0'5714}$$

$$d) \quad P(B \cup A_2) = P(B) + P(A_2) - P(B \cap A_2) = 0'78 + 0'3 - 0'2025 = \boxed{0'8775}$$

donde se necesita calcular que:

$$P(B \cap A_2) = P\left(\frac{B}{A_2}\right)P(A_2) = P\left(P\left(\frac{a_1}{\cdot}\right)P\left(\frac{B}{a_1}\right) + P\left(\frac{a_2}{\cdot}\right)P\left(\frac{B}{a_2}\right) + P\left(\frac{a_3}{\cdot}\right)P\left(\frac{B}{a_3}\right)\right)P(A_2) =$$
$$= \left(\frac{1}{2} \cdot 0'7 + \frac{1}{6} \cdot 0'75 + \frac{1}{3} \cdot 0'6\right)0'3 = 0'2025$$



PROBLEMA 2

(2'5 puntos)

Un fabricante de disquetes está interesado en analizar en sus discos la cantidad de bytes que quedan en sectores dañados tras darles formato. Ha tomado unas mediciones que indican que el 85% de los disquetes tienen menos de 10.000 bytes en sectores dañados. También se ha observado que el 90% de los disquetes tienen más de 4.000 bytes en sectores dañados. Suponga que la distribución del número de bytes existentes en sectores dañados puede aproximarse a una normal.

a) Calcule el número medio de bytes en sectores dañados, y su desviación típica.

Para el resto de apartados olvide los números anteriores y suponga que el número medio de bytes dañados es 7500 y la desviación típica es 2500.

b) Calcule la probabilidad de que la cantidad de bytes situados en sectores dañados se encuentre entre 5000 y 7000.

c) Calcule la probabilidad de que la diferencia entre el número de bytes que tenga en sectores dañados un disquete tomado al azar, y el número medio de bytes en sectores dañados en los disquetes de este fabricante, sea superior a 3000 (por exceso, o por defecto).

d) Calcule cuál es el número de bytes que debe haber en sectores dañados de un disquete, para el que el 60% de los disquetes tenga mayor número de bytes en sectores dañados que él.

e) Se le ha indicado que considere normal la distribución de la variable "número de bytes existentes en sectores dañados". Sin embargo esta variable sólo puede tomar valores naturales. Explique esta situación aparentemente contradictoria.

SOLUCIÓN:

Se define la variable X = "número de bytes situados en sectores dañados"

a) Resolvemos el sistema que se plantea:

$$P(X < 10000) = 0'85 \Rightarrow P(X \geq 10000) = 0'15 \Rightarrow$$

$$P\left(Z > \frac{10000 - \mu}{\sigma}\right) = 0'15 \Rightarrow \left\{ \begin{array}{l} 1'03 - 1'04 \\ Z_{0'15} - 1'04 \end{array} \quad \begin{array}{l} 0'1515 - 0'1492 \\ 0'15 - 0'1492 \end{array} \right\} \Rightarrow$$

$$\Rightarrow Z_{0'15} = 1'0365 \Rightarrow \boxed{1'0365 = \frac{10000 - \mu}{\sigma}}$$

$$P(X > 4000) = 0'9 \Rightarrow P(X \leq 4000) = 0'1 \Rightarrow P\left(Z \leq \frac{4000 - \mu}{\sigma}\right) = 0'1 \Rightarrow$$

$$P\left(Z \geq \frac{\mu - 4000}{\sigma}\right) = 0'1 \Rightarrow \left\{ \begin{array}{l} 1'28 - 1'29 \\ Z_{0'1} - 1'29 \end{array} \quad \begin{array}{l} 0'1003 - 0'0985 \\ 0'1 - 0'0985 \end{array} \right\} \Rightarrow$$

$$\Rightarrow Z_{0'1} = 1'2816 \Rightarrow \boxed{1'2816 = \frac{\mu - 4000}{\sigma}}$$



PROBLEMA 2: Continuación

Resolviendo las dos ecuaciones obtenidas para las dos incógnitas μ y σ :

$\mu = 7317'2$ $\sigma = 2588'3$

b) $P(5000 \leq X \leq 7000) = P\left(\frac{5000 - 7500}{2500} \leq Z \leq \frac{7000 - 7500}{2500}\right) =$

$P(-1 \leq Z \leq -0'2) = 0'262$

c) $P(|X - 7500| \geq 3000) = 1 - P(|X - 7500| \leq 3000) = 1 - P(4500 \leq Z \leq 10500) =$

$1 - P(-1'2 \leq Z \leq 1'2) = 0'2302$

d) $P(X \geq X_{0'6}) = 0'6$ $P\left(Z \geq \frac{X_{0'6} - 7500}{2500}\right) = 0'6$ $Z_{0'6} = -Z_{0'4}$

$\frac{X_{0'6} - 7500}{2500} = -0'25$ $X_{0'6} = 6875$

e) La variable aleatoria se trabaja mediante una normal, porque nos lo indica el enunciado, pero realmente es una v.a. discreta (los valores son naturales, no reales), que por trabajar con grandes números está siendo utilizada como una continua.

En este tipo de aproximaciones se comete poco error al calcular probabilidades de intervalos de valores. Sin embargo, al intentar calcular la probabilidad de un valor concreto de la variable, aparecería la incongruencia de la correspondencia de valores. Sería falso decir que la probabilidad de un valor concreto (6500, por ejemplo) fuera nula. La probabilidad de un valor de la variable discreta se deberá corresponder con la de un cierto intervalo de la continua, no siendo nula, sino muy baja.

En el curso se ha visto la aproximación de la v.a. binomial y de la v.a. Poisson, mediante normales, utilizando aproximaciones por continuidad. En este caso ha de existir una aproximación por continuidad igual o similar a las vistas.



PROBLEMA 3

(2'5 puntos)

Una variable aleatoria X , tiene por función de distribución:

$$F(x) = \begin{cases} 0 & x \leq 1 \\ ax + b & 1 < x \leq 2 \\ 3/5 & 2 < x \leq 4 \\ cx + d & 4 < x \leq 6 \\ 1 & 6 < x \end{cases}$$

- a) ¿Es X discreta?, ¿continua?, ¿puede ser discreta o continua?. Razone suficientemente las respuestas
- b) Si fuese continua calcule:
 - b1) a, b, c y d
 - b2) La función de densidad, comprobando después de calculada si realmente es función de densidad o no.
 - b3) Media y varianza de X
 - b4) $P(X \geq 3)$; $P[(X \geq 3.5) \cup (1.5 \leq X \leq 5)]$ y $P[(X \leq 5) \cap (X \geq 1.5)]$

Nota: Haga todos los cálculos en función de a, b, c y d y sustituya finalmente sus valores numéricos

SOLUCIÓN

- a) Se sabe que $F(x)$ es función de distribución de una v.a. X .

Si los valores de a, b, c y d fueran tales que $F(x)$ fuese una función continua, entonces X sería una v.a. continua. Esto es posible, como veremos después en el apartado b.1.

Para que X fuese discreta, $F(x)$ debería ser una función en escalera, pero entonces, por la forma en que está definida, vemos que, en todos los puntos de salto, $F(x)$ no es continua por la derecha, con lo que no sería función de distribución.

Cabe observar que sí se puede conseguir que $F(x)$ sea una función escalonada: en el intervalo $1 < x \leq 2$ debería ser $a=0$ y b un valor cualquiera del intervalo $0 \leq b \leq 3/5$; y, análogamente, en el intervalo $4 < x \leq 6$ bastaría tomar $c=0$ y $3/5 \leq d \leq 1$. Sin embargo, en ninguno de estos casos la función sería continua por la derecha en los puntos de salto, por lo que nunca sería una v.a. discreta.

Por tanto, X tiene que ser continua.



APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

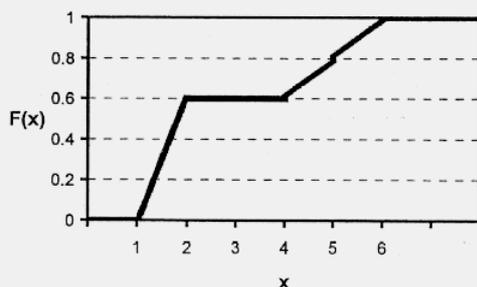
CURSO: 3º GRUPO: _____

PROBLEMA 3: Continuación A

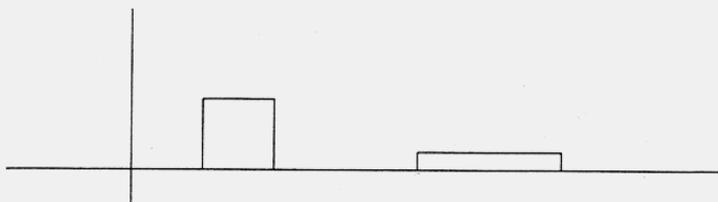
b.1) Al ser X continua, $F(x)$ es una función continua, luego

$$\left. \begin{array}{l} \text{para } x=1 \quad F(1)=0=a+b \\ x=2 \quad F(2)=2a+b=\frac{3}{5} \end{array} \right\} \Rightarrow a=\frac{3}{5} \Rightarrow b=-\frac{3}{5}$$

$$\left. \begin{array}{l} \text{para } x=4 \quad F(4)=\frac{3}{5}=4c+d \\ x=6 \quad F(6)=6c+d=1 \end{array} \right\} \Rightarrow 2c=\frac{2}{5} \Rightarrow c=\frac{1}{5} \Rightarrow d=-\frac{1}{5}$$



$$\text{b.2) } f(x) = \frac{dF(x)}{dx} = \begin{cases} a = \frac{3}{5} & 1 \leq x \leq 2 \\ c = \frac{1}{5} & 4 \leq x \leq 6 \\ 0 & \text{resto} \end{cases}$$



En la gráfica que representa esta función, se ve que el área que encierra con el eje X es:

$$\text{Área} = 1 \cdot \frac{3}{5} + 2 \cdot \frac{1}{5} = 1, \text{ luego sí es función de densidad}$$

$$\text{También: } \int_{-\infty}^{\infty} f(x) dx = \int_1^2 a dx + \int_4^6 c dx = a + 2c = \frac{3}{5} + 2 \cdot \frac{1}{5} = 1$$



PROBLEMA 3: Continuación B

b.3)

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_1^2 axdx + \int_4^6 cxdx = a \left[\frac{x^2}{2} \right]_1^2 + c \left[\frac{x^2}{2} \right]_4^6 = a \left(2 - \frac{1}{2} \right) + c(18 - 8) = \frac{3}{2}a + 10c$$
$$= \frac{9}{10} + \frac{10}{5} = 2.9$$

$$Var[X] = E[X^2] - E^2[X]$$

$$E[X^2] = \int_1^2 ax^2dx + \int_4^6 cx^2dx = a \left[\frac{x^3}{3} \right]_1^2 + c \left[\frac{x^3}{3} \right]_4^6 = a \left(\frac{8}{3} - \frac{1}{3} \right) + c \left(\frac{216}{3} - \frac{64}{3} \right) = \frac{7}{3}a + \frac{152}{3}c$$
$$= \frac{173}{15} = 11.53$$

$$Var[X] = \frac{173}{15} - 2.9^2 = 3.123$$

b.4)

$$P(X \geq 3) = 1 - P(X \leq 3) = 1 - F(3) = 1 - \frac{3}{5} = \frac{2}{5}$$

$$P[(X \geq 3.5) \cup (1.5 \leq X \leq 5)] = P(X \geq 1.5) = 1 - P(X \leq 1.5) = 1 - F(1.5) = 1 - (a \cdot 1.5 + b)$$
$$= 1 - \left(\frac{3}{5} \cdot 1.5 - \frac{3}{5} \right) = 0.7$$

También:

$$P[(X \geq 3.5) \cup (1.5 \leq X \leq 5)] = P(X \geq 3.5) + P(1.5 \leq X \leq 5) - P[(X \geq 3.5) \cap (1.5 \leq X \leq 5)]$$
$$= 1 - F(3.5) + F(5) - F(1.5) - P(3.5 \leq X \leq 5) = 1 - F(3.5) + F(5) - F(1.5) - F(5) + F(3.5)$$
$$= 1 - F(1.5) = 0.7$$

$$P[(X \leq 5) \cap (X \geq 1.5)] = P(1.5 \leq X \leq 5) = F(5) - F(1.5) = 5c + d - 1.5a - b = 0.5$$



UNIVERSIDAD PONTIFICIA DE SALAMANCA EN MADRID

Departamento de Electrónica y Comunicaciones 2002'40

ESTADÍSTICA: 16 de septiembre de 2002 (Tarde)

APELLIDOS Y NOMBRE: _____
Nº DE EXPEDIENTE: _____ CURSO: 3º GRUPO: _____

PROBLEMA 4

(2'5 puntos)

Se ha desarrollado una nueva técnica de fabricación de un componente de una antena de radiofrecuencia, mediante la cual se afirma que se obtienen componentes que en término medio son capaces de soportar frecuencias máximas superiores a 7GHz.

Se sabe que la distribución de las frecuencias máximas soportadas por los componentes así fabricados tienen una desviación típica de 400MHz.

Se quiere comprobar si es cierto que en media al menos soportan 7GHz, para lo cual el departamento de control de calidad ha desarrollado un contraste de hipótesis para la medida de 100 componentes con una región crítica de $\bar{x} < 6900\text{MHz}$.

Se pide:

- Calcule la probabilidad de que siendo realmente cierta la afirmación de que la frecuencia máxima de trabajo es superior a 7GHz, sin embargo se concluya que esto no se cumple.
- Calcule la probabilidad de aceptar como cierto que se superan dichos 7GHz en media, si realmente sólo se superan los 6'94 GHz.
- ¿Cuál es el valor de la frecuencia máxima que en media debería tener la población de componentes para que la potencia del contraste con dicha población de componentes fuera de 89'25%?
- ¿Para qué valores la potencia aumentará: para los superiores al pedido en el apartado c), o para los inferiores?

SOLUCIÓN:

Se ha realizado un contraste unilateral (porque se pretende detectar que se superen o no los 7GHz, y porque la región crítica es $\bar{x} < \text{valor}$) para el valor medio de una población, con muestra suficientemente grande como para aproximar mediante el Teorema Central del Límite ($n=100$) y desviación típica conocida ($\sigma = 400\text{ MHz}$)

$$H_0 : \mu = 7.000\text{ MHz} \quad \text{Reg. Aceptación : } \bar{x} \geq 6.900\text{ MHz}$$

$$H_A : \mu < 7.000\text{ MHz} \quad \text{Reg. Crítica : } \bar{x} < 6.900\text{ MHz}$$

$$\begin{aligned} \text{a) } \alpha &= P(\varepsilon_{\text{tipol}}) = P(\text{rechazo } H_0 | H_0 \text{ cierto}) = P(\bar{x} < 6900 | \mu = 7000) = P\left(z < \frac{6900 - 7000}{400/\sqrt{100}}\right) = \\ &= P(z < -2'5) = P(z > 2'5) = 0'00621 \end{aligned}$$



PROBLEMA 4: Continuación

b) Se pide la probabilidad de error tipo II, para $\mu_a = 6940$

$$\beta|_{\mu_a} = P(\varepsilon_{T\text{ip}oII})_{\mu_a} = P(\text{aceptar } H_0 | \mu = \mu_a) = P(\bar{x} \geq 6900 | \mu = \mu_a) = P\left(z \geq \frac{6900 - \mu_a}{400/\sqrt{100}}\right)$$

$$\Rightarrow \beta|_{\mu_a} = P\left(z \geq \frac{6900 - \mu_a}{400/\sqrt{100}}\right) \Rightarrow$$

$$\beta|_{\mu=6940} = P(\varepsilon_{T\text{ip}oII})_{\mu=6940} = P\left(z \geq \frac{6900 - 6940}{400/\sqrt{100}}\right) = P(z \geq -1) = 1 - P(z > 1) = 1 - 0'1587$$

$$\Rightarrow \beta|_{\mu=6940} = 0'8413$$

c) La potencia es 0'8925, cuando la probabilidad de error tipo II es de $1 - 0'8925 = 0'1075$.

Volviendo a los cálculos del apartado anterior:

$$\beta|_{\mu_a} = P\left(z \geq \frac{6900 - \mu_a}{400/\sqrt{100}}\right) = 0'1075$$

Buscando este valor en las tablas, vemos que $P(z > 1'24) = 0'1075$

luego

$$\frac{6900 - \mu_a}{400/\sqrt{100}} = 1'24 \Rightarrow \mu_a = 6900 - 1'24 \frac{400}{\sqrt{100}} = 6850'4 \text{ MHz}$$

d) La potencia aumenta para valores del parámetro más alejados del de la hipótesis nula, ya que el contraste es capaz de distinguir con más facilidad la diferencia. En este caso, por tanto, para valores inferiores al calculado: $\mu < 6850'4 \text{ MHz}$

Esta argumentación teórica se puede comprobar numéricamente por comparación de los valores calculados en los apartados b) y c).

De igual manera, a partir de la expresión obtenida en b) $\beta|_{\mu_a} = P\left(z \geq \frac{6900 - \mu_a}{400/\sqrt{100}}\right)$, y por

observación de la tabla de valores de la distribución $N(0,1)$, puede observarse que β es menor cuanto mayor sea el numerador, y por tanto, cuanto más pequeño sea el valor de μ_a .



PROBLEMA 1

(2'5 puntos)

Se tienen 5 fichas, cada una marcada con una cifra distinta del 1 al 5. Se realiza con ellas un experimento, consistente en elegir al azar tres fichas y colocarlas en fila, en el mismo orden de su elección, de izquierda a derecha. Calcule las probabilidades de los siguientes sucesos:

- a) Que aparezca el número "123"
- b) Que aparezca un número compuesto por las cifras "1", "2", y "3"
- c) Que aparezca un número que no contenga la cifra "3"
- d) Que aparezca un número que contenga al menos una de las cifras "2" ó "3"
- e) Que aparezca un número compuesto por cifras sucesivas.
- f) Que aparezca un número compuesto por cifras sucesivas, ordenadas ascendentemente de izquierda a derecha.
- g) Que aparezca un número par

SOLUCIÓN:

a) $P_A = \frac{1}{5} \cdot \frac{1}{4} \cdot \frac{1}{3} = \frac{1}{60}$ o lo que es lo mismo: $P_A = \frac{1}{V_{5,3}} = \frac{1}{5 \cdot 4 \cdot 3} = \frac{1}{60}$ *favorable*

b) $P_B = \frac{1}{60} \cdot P_3 = \frac{3!}{60} = \frac{1}{10}$ o lo que es lo mismo: $P_B = \frac{1}{C_{5,3}} = \frac{3! \cdot 2!}{5!} = \frac{1}{10}$ *probable favorable*

o por hipergeométrica para varios sucesos: $P_B = \frac{\binom{1}{1} \binom{2}{1} \binom{3}{1} \binom{4,5}{0}}{\binom{5}{3}} = \frac{3! \cdot 2!}{5!} = \frac{1}{10}$

c) $P_C = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} = \frac{2}{5} = \frac{C_{4,3}}{C_{5,3}}$

d) $D =$ Que contenga el "2" y/ó el "3" $\bar{D} =$ Que no contenga el "2" ni el "3"

$P_D = 1 - P_{\bar{D}} = 1 - \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = 1 - \frac{1}{10} = \frac{9}{10}$
= 1 - C_{2,2} / C_{5,3}

e) $P_E = P_{1,2,3} + P_{2,3,4} + P_{3,4,5} = 3 \cdot \left(\frac{1}{60} \cdot P_3 \right) = \frac{3 \cdot 3 \cdot 2}{60} = \frac{3}{10}$ *C_{5,3}*

f) $P_F = P_{"123"} + P_{"234"} + P_{"345"} = 3 \cdot \frac{1}{60} = \frac{3}{60} = \frac{1}{20}$

g) $G =$ Que salga número par = Que el último número sea el "2" ó el "4"

$P_G = \frac{2}{5}$



PROBLEMA 2

(2'5 puntos)

En la fabricación de aleación de estaño-plomo para la soldadura de componentes electrónicos se quiere estudiar la relación existente entre la proporción de plomo en la aleación y la cantidad de aleación producida por día.

Nombraremos por X a la proporción de plomo en la aleación, y por Y a la producción diaria de aleación (en Tm). Se sabe que la densidad conjunta de las variables aleatorias X e Y es

$$f(x,y) = \begin{cases} k y (1+3x^2) & \forall 0 < x < 1; 0 < y < 2 \\ 0 & \text{resto} \end{cases}$$

Se pide:

- a) Calcule el valor de k
- b) Obtenga las funciones de densidad marginales
- c) ¿Son independientes las variables X e Y ?
- d) Calcule la esperanza de las variables X , Y y de su producto $G = X \cdot Y$
- e) Calcule σ_x , σ_y y σ_{xy}
- f) Obtenga la recta de regresión de Y sobre X
- g) ¿Cuánto vale el coeficiente de correlación lineal entre las variables X e Y ?

Nota: Se recomienda que trabaje los apartados sin sustituir el valor de "k" obtenido en el apartado a), realizando esta sustitución al final de cada apartado.

SOLUCIÓN:

a) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1 \Rightarrow 1 = k \int_0^1 (1+3x^2) dx \int_0^2 y dy = k [x+x^3]_0^1 \left[\frac{y^2}{2} \right]_0^2 =$
 $= k \cdot [1+1] \cdot [2] = 4k = 1 \Rightarrow \boxed{k = \frac{1}{4}}$

b) $f(x) = \int_{-\infty}^{\infty} f(x,y) dy$
 $\forall 0 < x < 1 \quad f(x) = \int_0^2 k y (1+3x^2) dy = k(1+3x^2) \left[\frac{y^2}{2} \right]_0^2 = 2k(1+3x^2)$
 $f(x) = \begin{cases} 2k(1+3x^2) & \forall 0 < x < 1 \\ 0 & \text{resto} \end{cases} \quad f(x) = \begin{cases} \frac{1+3x^2}{2} & \forall 0 < x < 1 \\ 0 & \text{resto} \end{cases}$

$f(y) = \int_{-\infty}^{\infty} f(x,y) dx$
 $\forall 0 < y < 2 \quad f(y) = \int_0^1 k y (1+3x^2) dx = k y [x+x^3]_0^1 = 2ky$
 $f(y) = \begin{cases} 2ky & \forall 0 < y < 2 \\ 0 & \text{resto} \end{cases} \quad f(y) = \begin{cases} \frac{y}{2} & \forall 0 < y < 2 \\ 0 & \text{resto} \end{cases}$



PROBLEMA 2: Continuación

c) $f(x) \cdot f(y) = \begin{cases} 4k^2 y(1+3x^2) & \forall 0 < x < 1; 0 < y < 2 \\ 0 & \text{resto} \end{cases}$

Dado que $k = \frac{1}{4} \Rightarrow f(x, y) = f(x) \cdot f(y) \quad \forall (x, y) \Rightarrow X$ e Y son independientes

d)
$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^1 x \cdot 2k(1+3x^2) dx = 2k \int_0^1 (x+3x^3) dx = 2k \left[\frac{x^2}{2} + \frac{3x^4}{4} \right]_0^1 = \\ &= 2k \left[\frac{1}{2} + \frac{3}{4} \right] = \frac{5k}{2} = \frac{5}{8} \end{aligned}$$

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f(y) dy = \int_0^1 y \cdot 2ky dy = 2k \left[\frac{y^3}{3} \right]_0^1 = \frac{16k}{3} = \frac{4}{3}$$

Dado que X e Y son independientes, tendremos $E(X \cdot Y) = E(X) \cdot E(Y)$. Así:

$$E(X \cdot Y) = E(X) \cdot E(Y) = \frac{5k}{2} \cdot \frac{16k}{3} = \frac{40k^2}{3} = \frac{5}{6}$$

Podemos comprobarlo de manera inmediata, sin más que plantear:

$$E(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y) dx dy = \int_0^1 2kx(1+3x^2) dx \cdot \int_0^2 2ky^2 dy$$

que, efectivamente, es $E(X) \cdot E(Y)$

e) $\sigma_X^2 = E(X^2) - E^2(X)$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = 2k \int_0^1 (x^2 + 3x^4) dx = 2k \left[\frac{x^3}{3} + \frac{3x^5}{5} \right]_0^1 = \frac{28k}{15} = \frac{7}{15}$$

$$\sigma_X^2 = \frac{28k}{15} - \left(\frac{5k}{2} \right)^2 = \frac{7}{15} - \left(\frac{5}{8} \right)^2 = \frac{64 \cdot 7 - 15 \cdot 25}{64 \cdot 15} = \frac{73}{960}$$

$$\Rightarrow \sigma_X = \sqrt{\frac{73}{960}} \cong \sqrt{0'076} \cong 0'27576$$

$\sigma_Y^2 = E(Y^2) - E^2(Y)$

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 \cdot f(y) dy = 2k \int_0^2 y^3 dy = 2k \left[\frac{y^4}{4} \right]_0^2 = 8k = 2$$

$$\sigma_Y^2 = 8k - \left(\frac{16k}{3} \right)^2 = 2 - \left(\frac{4}{3} \right)^2 = 2 - \frac{16}{9} = \frac{2}{9} \Rightarrow \sigma_Y = \frac{\sqrt{2}}{3} \cong 0'4714$$

Dado que X e Y son independientes, tendremos $E(X \cdot Y) = E(X) \cdot E(Y)$. Así $\sigma_{XY} = 0$

f) Dado que X e Y son independientes, tendremos $\hat{Y} = E(Y) = \frac{16k}{3} = \frac{4}{3}$

g) Dado que X e Y son independientes, tendremos $\rho_{XY} = 0$



PROBLEMA 3

(2'5 puntos)

Una empresa vende repuestos de maquinaria industrial. El 10% de las compras que se realizan son de un gasto inferior a 1.200€ cada una. Por otro lado, el 60% de las compras son de un gasto superior a 1.800€ cada una. Se sabe que el gasto por compra sigue una distribución normal.

a) Calcule el gasto medio por compra y la desviación típica.

Para los apartados b) y c) suponga que el gasto medio por compra es de 2000€ y la desviación típica es de 600€.

b) Calcule la probabilidad de que el gasto por compra difiera del gasto medio por compra en más de 400€.

c) Calcule para qué gasto por compra, el 30% de las compras tienen menor gasto que él.

Se tiene una variable aleatoria X , de la que se conoce que sigue una distribución Ji-Dos de Pearson, con 61 grados de libertad.

d) Calcule la probabilidad de que la variable tome un valor inferior a 30.

Nota: En a) deberá interpolar, mientras que en los demás se podrá coger el valor más próximo.

SOLUCIÓN:

a) Se toma X como variable "gasto":

$$P(X < 1.200) = 0'1 \quad P\left(Z < \frac{1.200 - \mu}{\sigma}\right) = 0'1 \quad P\left(Z > \frac{\mu - 1.200}{\sigma}\right) = 0'1$$

Interpolando:

$$\begin{array}{l} 1'28 - 1'29 \quad 0'1003 - 0'0985 \\ Z_{0'1} - 1'29 \quad 0'1 - 0'0985 \end{array} \quad Z_{0'1} = 1'2816 \quad \boxed{\frac{\mu - 1.200}{\sigma} = 1'2816}$$

$$P(X > 1.800) = 0'6 \quad P\left(Z > \frac{1.800 - \mu}{\sigma}\right) = 0'6 \quad P\left(Z > \frac{\mu - 1.800}{\sigma}\right) = 0'4$$

Interpolando:

$$\begin{array}{l} 0'25 - 0'26 \quad 0'4013 - 0'3974 \\ Z_{0'4} - 0'26 \quad 0'4 - 0'3974 \end{array} \quad Z_{0'4} = 0'2533 \quad \boxed{\frac{\mu - 1.800}{\sigma} = 0'2533}$$

Resolviendo las dos ecuaciones obtenidas para las dos incógnitas μ y σ :

$$\mu = 1947'78$$

$$\sigma = 583'48$$



PROBLEMA 3: Continuación

b) La probabilidad pedida es:

$$1 - P(2000 - 400 < X < 2000 + 400) = 1 - P(1600 < X < 2400) =$$

$$1 - P\left(\frac{1600 - 2000}{600} < Z < \frac{2400 - 2000}{600}\right) = 1 - P(-0'666 < Z < 0'666) = \boxed{0'5092}$$

c) $P(X \leq X_{0'7}) = 0'3$ $P\left(Z \leq \frac{X_{0'7} - 2000}{600}\right) = 0'3$ $P\left(Z \geq \frac{2000 - X_{0'7}}{600}\right) = 0'3$

$$\frac{2000 - X_{0'7}}{600} = 0'52$$

$$\boxed{X_{0'7} = 1688}$$

d) $X = \chi_{61}^2$

Como $n = 61 > 30$, se realiza la aproximación de Ji-Dos de Pearson a una Normal,

definiendo una nueva variable aleatoria: $Y = \sqrt{2X} \Rightarrow X = \frac{Y^2}{2}$, con distribución:

$$N(\sqrt{2n-1}, 1) = N(11, 1)$$

La probabilidad que se pide es:

$$P(X < 30) = 1 - P(X \geq 30) = 1 - P\left(\frac{Y^2}{2} \geq 30\right) = 1 - P(Y \geq \sqrt{30 \cdot 2}) = 1 - P(Y \geq 7'745) =$$

$$1 - P\left(Z \geq \frac{7'745 - 11}{1}\right) = 1 - P(Z \geq -3'255) = 1 - (1 - P(Z \geq 3'255)) = \boxed{0'000687}$$



PROBLEMA 4

(2'5 puntos)

En el centro de llamadas de una entidad bancaria se quiere evaluar el número de llamadas entrantes en la franja horaria de 9h a 11h de la mañana en los días laborables. En función de ello se dimensiona la plantilla, teniendo en cuenta que: si se asignara más personal del necesario, ese recurso humano estaría siendo desperdiciado; mientras que si se asignara menos del necesario, se perderían llamadas al no ser atendidas, lo que se considera perjudicial para los intereses de la entidad bancaria.

La entidad bancaria toma decisiones sobre la dimensión de la plantilla asumiendo como cierta la premisa de que entre las 9h y las 11h de un día laborable, entran 245 llamadas en promedio. Se contabilizan las llamadas de 9h a 11h durante 27 días laborables, obteniendo en total 6750 llamadas. En esta muestra, la variable "número de llamadas obtenido en esa franja horaria, por día" tiene una cuasivarianza de 20.

Suponga que el número de llamadas entrantes en esa franja horaria en los días laborables sigue una distribución normal, y utilice un error tipo I del 5% para responder a las siguientes cuestiones:

- a) ¿Es correcta la premisa asumida? Justifique su respuesta por contraste de hipótesis
- b) Calcule el tamaño de la muestra que delimita la aceptación o el rechazo en dicho contraste, señalando si para valores superiores o inferiores al obtenido se rechaza o se acepta.
- c) Calcule el Error Tipo II, suponiendo que las llamadas que realmente entran un día laborable en promedio entre las 9 y las 11 es de 248.
- d) Si desconocemos el número medio de llamadas existente entre las 9h y las 11h en un día laborable, pero queremos estimarla, ¿con qué probabilidad podremos afirmar que no cometeremos un error superior a 2 llamadas en dicha estimación, si el número de días analizados se aumenta en 10?

SOLUCIÓN:

a) Se plantea el contraste bilateral:

$$\begin{aligned}
 H_0: \mu &= 245 & \sum_{i=1}^{27} X_i &= 6750 & \bar{X} &= \frac{1}{n} \sum_{i=1}^{27} X_i = \frac{6750}{27} = 250 \\
 H_a: \mu &\neq 245 & S_M^2 &= 20 & S_M &= 4'472
 \end{aligned}$$

Se utiliza el estadístico de contraste adecuado, teniendo en cuenta que la varianza poblacional es desconocida y el tamaño de la muestra es inferior a 30.

$$t = \frac{\bar{X} - \mu_0}{S_M / \sqrt{n}} = \frac{250 - 245}{4'472 / \sqrt{27}} = 5'8139 \qquad t_{\alpha/2; n-1} = t_{0'025; 26} = 2'056$$

Cae en la región de rechazo, ya que no cumple: $|t| \leq t_{\alpha/2; n-1}$, porque $5'8139 > 2'056$

Por tanto, con un Error Tipo I del 5% se puede afirmar que la premisa que está asumiendo no es correcta.



APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____ CURSO: 3º GRUPO: _____

PROBLEMA 4: Continuación

b) Teniendo en cuenta que la media muestral es superior a la media poblacional que sea asume en la hipótesis nula, la condición de aceptación con la que vamos a operar es:

$$t = \frac{\bar{X} - \mu_0}{S_M / \sqrt{n}} \leq t_{\alpha/2; n-1}$$

Para valores superiores a 30 de tamaño de muestra, teniendo en cuenta que estaríamos con una Z , siempre nos daría rechazo, al ser el valor de Z mayor que $Z_{0'025} = 1'96$, por tanto, operamos con la expresión anterior:

$$\frac{250 - 245}{4'472 / \sqrt{n}} \leq 2'056 \quad n \leq 3'3815$$

Luego, para $n \leq 3$ se acepta la hipótesis nula, y para $n > 4$ se rechaza.

c) $\beta = P\left(\frac{\text{Aceptar } H_0}{H_0 \text{ falsa}}\right)$

$$= P\left(\frac{\mu_0 - t_{\alpha/2; n-1} \frac{s}{\sqrt{n}} - \mu_a}{s / \sqrt{n}} \leq t_{n-1} \leq \frac{\mu_0 + t_{\alpha/2; n-1} \frac{s}{\sqrt{n}} - \mu_a}{s / \sqrt{n}}\right) =$$

$$= P\left(\frac{245 - 2'056 \frac{4'472}{\sqrt{27}} - 248}{4'472 / \sqrt{27}} \leq t_{26} \leq \frac{245 + 2'056 \frac{4'472}{\sqrt{27}} - 248}{4'472 / \sqrt{27}}\right) =$$

$$= P(-5'5443 \leq t_{26} \leq -1'2318) = P(t_{26} \geq 1'2318) - P(t_{26} \geq 5'5443) = 0'11 - 0 = \boxed{0'11}$$

d) El nuevo tamaño de muestra es: $n = 27 + 10 = 37$, es decir, mayor que 30, por lo que la expresión con la que vamos a operar es:

$$E = Z_{\alpha/2} \frac{S_M}{\sqrt{n}} \quad 2 = Z_{\alpha/2} \frac{4'472}{\sqrt{37}} \quad Z_{\alpha/2} = 2'72$$

Luego: $\frac{\alpha}{2} = 0'00326 \quad \alpha = 0'00652 \quad \boxed{1 - \alpha = 0'9934}$



UNIVERSIDAD PONTIFICIA DE SALAMANCA EN MADRID

Departamento de Electrónica y Comunicaciones 2002'31

ESTADÍSTICA: 3-junio-2002 (Tarde)

APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

CURSO: 3º

GRUPO: _____

PROBLEMA 1

(2'5 puntos)

Las capacidades de los condensadores de una serie en una cadena de fabricación siguen una distribución normal, con una media de 1'3nF (nano-Faradios), y una desviación típica de 0'04nF.

- 1) Escogido un condensador al azar, calcule la probabilidad de que su capacidad esté comprendida entre:
 - 1a) 1'28nF y 1'30nF
 - 1b) 1'30nF y 1'32nF
 - 1c) 1'31nF y 1'33nF
- 2) Calcule entre qué dos valores (simétricamente distribuidos en torno a la media) se encuentran las capacidades del 80% de los condensadores de esta serie.
- 3) Si se seleccionaran muestras de 16 condensadores:
 - 3a) ¿Cuál se esperaría que fueran la media y la desviación típica de las capacidades medias de cada muestra?
 - 3b) ¿Cuál sería la expresión de la función densidad de probabilidad de las medias de estas muestras.
- 4) Escogida una muestra de 16 condensadores al azar, calcule la probabilidad de que su capacidad media esté comprendida entre:
 - 4a) 1'28nF y 1'30nF
 - 4b) 1'30nF y 1'32nF
 - 4c) 1'31nF y 1'33nF
- 5) Calcule entre qué dos valores (simétricamente distribuidos en torno a la media) se encuentran las capacidades medias del 80% de las muestras de 16 condensadores que pudiéramos tomar sobre esta serie.
- 6) Compare y analice cada resultado de 4) frente a los de 1); y el resultado de 5) frente al de 2)
- 7) ¿Qué es más probable que ocurra:
 - a) un condensador de capacidad superior a 1'34nF;
 - b) una capacidad media superior a 1'32nF en una muestra de 4 condensadores;
 - c) o una capacidad media que supere los 1'31nF en una muestra de tamaño 16?.
 Explíquelo.

SOLUCIÓN:

La variable involucrada es X : Capacidad de un condensador de la serie (nF)

$$X : N(1'3, 0'04) \quad \Rightarrow \quad Z = \frac{X - 1'3}{0'04} : N(0, 1)$$

$$1a) \quad P(1'28 \leq x \leq 1'30) = P\left(\frac{1'28 - 1'30}{0'04} \leq z \leq \frac{1'30 - 1'30}{0'04}\right) = P(-0'5 \leq z \leq 0) = P(0 \leq z \leq 0'5) =$$

$$= P(z \geq 0) - P(z \geq 0'5) = 0'5 - 0'3085 = 0'1915$$

$$1b) \quad P(1'30 \leq x \leq 1'32) = P\left(\frac{1'30 - 1'30}{0'04} \leq z \leq \frac{1'32 - 1'30}{0'04}\right) = P(0 \leq z \leq 0'5) = 0'1915$$



APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

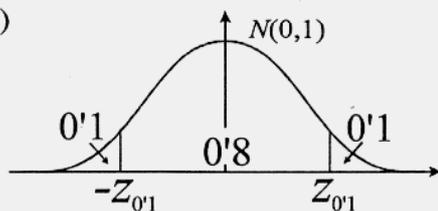
CURSO: 3º GRUPO: _____

PROBLEMA 1: Continuación A

$$1c) \quad P(1'31 \leq x \leq 1'32) = P\left(\frac{1'31-1'30}{0'04} \leq z \leq \frac{1'33-1'30}{0'04}\right) = P(0'25 \leq z \leq 0'75) =$$

$$= P(z \geq 0'25) - P(z \geq 0'75) = 0'4013 - 0'2266 = 0'1747$$

2)



$$Z = \frac{X-1'3}{0'04} \Rightarrow x = 1'3 + z \cdot 0'04$$

$$\left. \begin{array}{l} 1'28 \rightarrow 0'1003 \\ z_{0.1} \rightarrow 0'1 \\ 1'29 \rightarrow 0'0985 \end{array} \right\} \Rightarrow z_{0.1} = 1'281\bar{6}$$

$$0'8 = P(-z_{0.1} \leq z \leq z_{0.1}) = P(-1'281\bar{6} \leq z \leq 1'281\bar{6}) = P(1'3 - 1'281\bar{6} \cdot 0'04 \leq x \leq 1'3 + 1'281\bar{6} \cdot 0'04)$$

$$= P(1'3 - 0'0512\bar{6} \leq x \leq 1'3 + 0'0512\bar{6}) = P(1'2487\bar{3} \leq x \leq 1'3512\bar{6})$$

$$x \in [1'3 \pm 0'0512\bar{6}] \Rightarrow x \in [1'2487\bar{3}, 1'3512\bar{6}]$$

$$3) \quad n=16 \Rightarrow \bar{X}_{16} : N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(1'3, \frac{0'04}{\sqrt{16}}\right) = N(1'3, 0'01)$$

$$3a) \quad E[\bar{X}_{16}] = 1'3nF$$

$$3b) \quad f_{\bar{X}_{16}}(t) = \frac{1}{\sigma_{\bar{X}_{16}} \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{t - \mu_{\bar{X}_{16}}}{\sigma_{\bar{X}_{16}}}\right)^2} = \frac{1}{0'01 \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{t-1'3}{0'01}\right)^2}$$

$$4) \quad \bar{X}_{16} : N(1'3, 0'01) \Rightarrow Z = \frac{\bar{X}_{16} - \mu_{\bar{X}_{16}}}{\sigma_{\bar{X}_{16}}} = \frac{\bar{X}_{16} - \mu_X}{\sigma_X / \sqrt{n}} = \frac{\bar{X}_{16} - 1'3}{0'01} : N(0,1)$$

$$4a) \quad P(1'28 \leq \bar{x}_{16} \leq 1'30) = P\left(\frac{1'28-1'30}{0'01} \leq z \leq \frac{1'30-1'30}{0'01}\right) = P(-2 \leq z \leq 0) = P(0 \leq z \leq 2) =$$

$$= P(z \geq 0) - P(z \geq 2) = 0'5 - 0'0228 = 0'4772$$

$$4b) \quad P(1'30 \leq \bar{x}_{16} \leq 1'32) = P\left(\frac{1'30-1'30}{0'01} \leq z \leq \frac{1'32-1'30}{0'01}\right) = P(0 \leq z \leq 2) = 0'4772$$

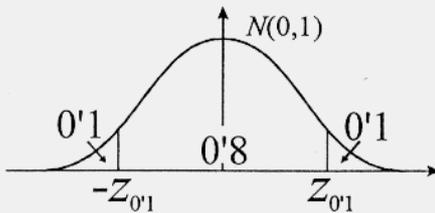


PROBLEMA 1: Continuación B

$$4c) \boxed{P(1'31 \leq \bar{x}_{16} \leq 1'32)} = P\left(\frac{1'31 - 1'30}{0'01} \leq z \leq \frac{1'33 - 1'30}{0'01}\right) = P(1 \leq z \leq 3) =$$

$$= P(z \geq 1) - P(z \geq 3) = 0'1587 - 0'00135 = \boxed{0'15735}$$

5) De la misma forma que en 2), pero ahora con



$$Z = \frac{\bar{X}_{16} - 1'3}{0'01} : N(0,1) \Rightarrow x = 1'3 + z \cdot 0'01$$

$$z_{0'1} = 1'2816 \Rightarrow \bar{x}_{16} \in 1'3 \pm 1'2816 \cdot 0'01$$

Así: $\boxed{\bar{x}_{16} \in [1'3 \pm 0'012816]} \Rightarrow \boxed{\bar{x}_{16} \in [1'287183, 1'312816]}$

$$6) \begin{cases} P_{1a} = 0'1915; P_{4a} = 0'4772 \\ P_{1b} = 0'1915; P_{4b} = 0'4772 \end{cases}$$

La distribución de \bar{X}_{16} es más concentrada en torno a $\mu = 1'30$, luego un intervalo que comience en este valor recoge más probabilidad en la distribución de \bar{X}_{16} , que en la de X .

$$P_{1c} = 0'1747; P_{4c} = 0'15735$$

Por el mismo motivo, si el intervalo se aleja de la media recoge menos probabilidad el de distribución más concentrada.

$$\text{ii) } x \in [1'3 \pm 0'05126] \quad \text{iii) } \bar{x}_{16} \in [1'3 \pm 0'012816]$$

De nuevo por el mismo motivo, para acumular la misma probabilidad en un intervalo centrado en la media, este intervalo resulta menor en la distribución más concentrada.

$$7) \begin{array}{l} X : N(1'3, 0'04) \Rightarrow P(x \geq 1'34) = \left(z \geq \frac{0'04}{0'04}\right) = (z \geq 1) \\ \bar{X}_4 : N\left(1'3, \frac{0'04}{\sqrt{4}}\right) = N(1'3, 0'02) \Rightarrow P(\bar{x}_4 \geq 1'32) = \left(z \geq \frac{0'02}{0'02}\right) = (z \geq 1) \\ \bar{X}_{16} : N\left(1'3, \frac{0'04}{\sqrt{16}}\right) = N(1'3, 0'01) \Rightarrow P(\bar{x}_{16} \geq 1'31) = \left(z \geq \frac{0'01}{0'01}\right) = (z \geq 1) \end{array}$$

Son equiprobables. Cada una es más concentrada que la anterior, y en todas se calcula la probabilidad de superar la media más una desviación típica ($z > 1$).



PROBLEMA 2

(2'5 puntos)

Se tiene la función de densidad de probabilidad conjunta:

$$f(x, y) = \begin{cases} \frac{kxy}{3} & \text{en } R_{xy} \\ 0 & \text{en el resto} \end{cases}$$

donde R_{xy} es la región delimitada por: $y \geq \frac{x}{2}$, $y \leq x$, $x \geq 0$, $y \geq 0$, $x + y \leq 2$

- a) Calcule el valor de k.
- b) Obtenga las funciones de densidad marginales.
- c) ¿Son independientes las variables x e y? A partir del resultado de independencia o no que haya obtenido, ¿puede deducir alguna conclusión respecto al valor del coeficiente de correlación?
- d) Calcule $P(x \geq 1)$.
- e) Calcule el valor de la función de distribución en el punto (1,1).
- f) Calcule $P(x \leq 1; y \geq 0.5)$.

Nota: No es necesario sustituir el valor de k en el apartado b).

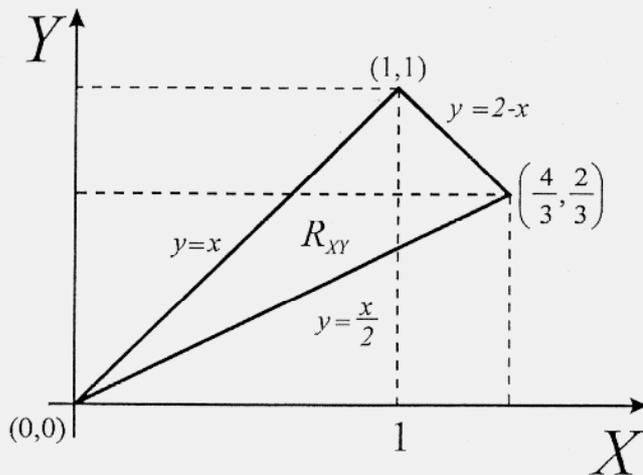
SOLUCIÓN:

a) Calculamos los dos puntos de corte:

$$\left. \begin{matrix} x + y = 2 \\ y = x \end{matrix} \right\} \Rightarrow x = 1; y = 1 \Rightarrow (1, 1)$$

$$\left. \begin{matrix} x + y = 2 \\ y = \frac{x}{2} \end{matrix} \right\} \Rightarrow x = \frac{4}{3}; y = \frac{2}{3} \Rightarrow \left(\frac{4}{3}, \frac{2}{3} \right)$$

$$\int_0^1 \int_{\frac{x}{2}}^x \frac{kxy}{3} dy dx + \int_1^{4/3} \int_{\frac{x}{2}}^{2-x} \frac{kxy}{3} dy dx = \frac{k}{6} \left[\frac{3}{2^4} + \frac{173}{3^4 2^4} \right] = \frac{13}{3^5} k \cong 0.053497942k = 1$$



$$\Rightarrow k = \frac{3^5}{13} \cong 18.6923$$



PROBLEMA 2: Continuación

b) $\int_{\frac{x}{2}}^x \frac{kxy}{3} dy = \frac{kx^3}{8}$ $\int_{\frac{x}{2}}^{2-x} \frac{kxy}{3} dy = \frac{kx(2-x)^2}{6} - \frac{kx^3}{24}$

$$f(x) = \begin{cases} \frac{kx^3}{8} & 0 \leq x \leq 1 \\ \frac{kx(2-x)^2}{6} - \frac{kx^3}{24} & 1 < x \leq \frac{4}{3} \\ 0 & \text{resto} \end{cases}$$

Asimismo:

$\int_y^{2-y} \frac{kxy}{3} dx = \frac{ky^3}{2}$ $\int_y^{2-y} \frac{kxy}{3} dx = \frac{ky(2-y)^2}{6} - ky^3$

$$f(y) = \begin{cases} \frac{ky^3}{2} & 0 \leq y \leq \frac{2}{3} \\ \frac{ky(2-y)^2}{6} - ky^3 & \frac{2}{3} < y \leq 1 \\ 0 & \text{resto} \end{cases}$$

c) Probando con un punto, como por ejemplo el (1,1):

$f(x,y) \neq f(x)f(y)$ ya que: $k \neq \frac{k}{8}$

Por tanto, no son independientes.

Esto implica que hay algún tipo de dependencia, luego existe la posibilidad de que pudiera ser lineal, por lo que no podemos concluir un valor de antemano; habría que calcularlo con su correspondiente expresión.

d) $P(x \geq 1) = \int_1^{4/3} \int_{\frac{x}{2}}^{2-x} \frac{kxy}{3} dy dx \cong 0'4089$

e) $F(1,1) = 1 - P(x \geq 1) \cong 1 - 0'4089 = 0'5911$

f) $\int_{0.5}^1 \int_{0.5}^x \frac{kxy}{3} dy dx = 0'407$



PROBLEMA 3

(2'5 puntos)

- a) Se sabe que una determinada característica de una población tiene distribución $N(\mu,3)$. Al realizar un muestreo aleatorio simple de tamaño n , se obtiene, para la media, μ , de dicha característica, un intervalo de confianza cuya semilongitud vale a , con un nivel de significación del 5%:
 - a1) ¿Cuál debería ser el tamaño muestral para obtener una estimación el doble de precisa, manteniendo la misma confianza?
 - a2) ¿Con qué confianza podemos conseguir un intervalo de la mitad de longitud que el inicial, con el mismo tamaño muestral, n ?
- b) El número de clientes que acude por hora a una oficina bancaria sigue un proceso de Poisson. Contados los clientes que acuden en 150 horas elegidas al azar, se obtiene que han sido 1350.
 - b1) Obtenga un intervalo de confianza para el verdadero valor medio de clientes por hora, con un nivel de significación del 90%
 - b1') Obtenga un intervalo de confianza para el verdadero valor medio de clientes por hora, con un nivel de confianza del 90%
 - b2) ¿Cuál sería el nivel de confianza de un intervalo cuyo extremo inferior fuera 9, si la muestra hubiera sido de la mitad de tamaño que la anterior y la media muestral hubiera sido un 10% mayor que la anterior?

NP

SOLUCIÓN:

a1) Estimamos la media de una población normal de varianza conocida. El intervalo de confianza simétrico en este caso es $I = \left[\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$, de semilongitud $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = a$.

Manteniendo la misma confianza (y mismo nivel de significación), una estimación doble de precisa será la que obtenga un intervalo de confianza de longitud mitad que el anterior, y, por tanto, de semilongitud mitad que la anterior. Luego $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n'}} = \frac{a}{2}$.

Dividiendo ambas expresiones resulta $\frac{\sqrt{n'}}{\sqrt{n}} = 2 \Rightarrow n' = 4n$

es decir, el tamaño muestral debe ser 4 veces mayor, y esto independientemente de cuál sea el nivel de significación, siempre que sea el mismo en los dos casos.

a2) Ahora mantenemos el tamaño muestral y queremos conseguir también el doble de precisión, es decir, un intervalo de confianza de la mitad de longitud. Esto lo conseguiremos con un nivel de confianza inferior.

$$\left. \begin{aligned} z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} &= a \\ z_{\alpha'/2} \cdot \frac{\sigma}{\sqrt{n}} &= \frac{a}{2} \end{aligned} \right\} \Rightarrow \frac{z_{\alpha/2}}{z_{\alpha'/2}} = 2 \Rightarrow z_{\alpha'/2} = \frac{1}{2} z_{\alpha/2}$$



PROBLEMA 3: Continuación

como

$$\alpha = 0'05 \Rightarrow \frac{\alpha}{2} = 0'025 \Rightarrow z_{\alpha/2} = 1'96 \Rightarrow z_{\alpha'/2} = 0'98 \Rightarrow \frac{\alpha'}{2} = 0'1635$$

$$\Rightarrow \alpha' = 0'327 \Rightarrow 1 - \alpha' = 0'673$$

- b1) Un nivel de significación del 90% significa que la probabilidad de que el intervalo sea erróneo es del 90%. Esto no tiene ningún sentido, y así hay que indicarlo; pero sí se puede calcular. La mejor respuesta sería indicar su falta de sentido, y calcularlo:

$$I = \left[\bar{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{\bar{x}}{n}} \right], \text{ donde } \bar{x} = \frac{1350}{150} = 9 = \hat{\lambda}$$

$$\text{y } 1 - \alpha = 0'1 \Rightarrow \alpha = 0'9 \Rightarrow \frac{\alpha}{2} = 0'45 \Rightarrow z_{\alpha/2} \cong 0'125,$$

$$\text{luego } I = \left[9 \pm 0'125 \sqrt{\frac{9}{150}} \right] = [9 \pm 0'03062] = [8'9694, 9'0306]$$

- b1') Queremos estimar la media de una distribución de Poisson. Como el tamaño muestral es grande (150), sabemos que en este caso un intervalo de confianza de la media es

$$I = \left[\bar{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{\bar{x}}{n}} \right], \text{ donde } \bar{x} = \frac{1350}{150} = 9 = \hat{\lambda}$$

$$\text{y } 1 - \alpha = 0'9 \Rightarrow \alpha = 0'1 \Rightarrow \frac{\alpha}{2} = 0'05 \Rightarrow z_{\alpha/2} = 1'645,$$

$$\text{luego } I = \left[9 \pm 1'645 \sqrt{\frac{9}{150}} \right] = [9 \pm 0'403] = [8'597, 9'403]$$

- b2) El nuevo tamaño muestral es $n' = \frac{n}{2} = \frac{150}{2} = 75$ y la nueva media muestral

$$\bar{x}' = \bar{x} + 0'1\bar{x} = 9 + 0'1 \cdot 9 = 9'9,$$

con lo que el nuevo intervalo de confianza sería $I' = \left[\bar{x}' \pm z_{\alpha'/2} \sqrt{\frac{\bar{x}'}{n'}} \right]$

y como su extremo inferior es 9 :

$$\bar{x}' - z_{\alpha'/2} \cdot \sqrt{\frac{\bar{x}'}{n'}} = 9 \Rightarrow z_{\alpha'/2} \cdot \sqrt{\frac{\bar{x}'}{n'}} = 0'9 \Rightarrow z_{\alpha'/2} \cdot \sqrt{\frac{9'9}{75}} = 0'9 \Rightarrow z_{\alpha'/2} = 2'4772$$

$$\Rightarrow \frac{\alpha'}{2} = 0'0066 \Rightarrow \alpha' = 0'0132 \Rightarrow 1 - \alpha' = 0'9868$$



PROBLEMA 4

(2'5 puntos)

Una empresa de asesoramiento informático quiere estimar si el número medio de peticiones de asesoramiento que cabe esperar que tengan cada día es 5 o en su defecto mayor que 5. De estudios anteriores se considera aceptable que la varianza poblacional del número de peticiones por día es 4. El Director Técnico de la empresa (que aprobó con buena nota la Estadística en la UPS) dice que si observando 100 días, elegidos al azar, el número medio de peticiones por día es mayor o igual a 5'35 rechazará la hipótesis de que el número medio es 5.

- a) ¿Cuál es la probabilidad de que la decisión de rechazar la hipótesis fuese errónea?
- b) ¿Cuál sería el criterio de decisión para $\alpha = 0'01$?
- c) Si $\alpha = 0'05$, ¿cuál sería la probabilidad de aceptar que el número medio es 5, cuando realmente fuese 5'5? ¿y cuando realmente fuese 6?

SOLUCIÓN

Se quiere contrastar las hipótesis:

$$H_0 : \mu = \mu_0 = 5$$

$$H_a : \mu > 5$$

- a) La probabilidad de que la decisión de rechazar la hipótesis planteada (la hipótesis nula) fuese errónea es la probabilidad de rechazar la hipótesis nula cuando es cierta, es decir, el error de tipo I, α : $P(\text{rechazar } H_0 / H_0 \text{ cierta}) = \alpha$

Dado que el tamaño muestral es grande y la varianza poblacional conocida, bajo la hipótesis nula, el estadístico \bar{x} tiene distribución normal $\bar{x} : N\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$

y por tanto el estadístico de contraste $\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} : N(0,1)$.

La región crítica de este contraste es $R.C. : \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha \Rightarrow \bar{x} - \mu_0 > z_\alpha \frac{\sigma}{\sqrt{n}}$

y, según el enunciado, se rechazará si $\bar{x} \geq 5'35$, es decir, si $\bar{x} - \mu_0 \geq 0'35$;

$$\text{luego } z_\alpha \frac{\sigma}{\sqrt{n}} = 0'35 \Rightarrow z_\alpha = 0'35 \cdot \frac{\sqrt{n}}{\sigma} = 0'35 \cdot \frac{\sqrt{100}}{2} = 1'75 \Rightarrow \alpha = \mathbf{0'04}$$



PROBLEMA 4: Continuación

b) La región crítica es $\frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_\alpha$ o bien $\bar{x} - \mu_0 > z_\alpha \frac{\sigma}{\sqrt{n}}$;

luego $\bar{x} > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$

como $\alpha = 0.01 \Rightarrow z_\alpha = 2.32$ y el criterio de decisión será:

Rechazamos H_0 si $\bar{x} > 5'464$

Aceptamos H_0 si $\bar{x} \leq 5'464$

c) Nos están pidiendo la probabilidad de error de tipo II, es decir, de aceptar la hipótesis nula cuando es falsa

$$c1) P(\text{aceptar } H_0 | \mu = 5.5) = \beta = P\left(\bar{x} < 5 + \frac{1}{5} z_\alpha \mid \mu = 5.5\right) = P\left(\frac{\bar{x} - 5.5}{1/5} < \frac{5 + \frac{z_\alpha}{5} - 5.5}{1/5}\right) =$$

$$= P(Z < z_\alpha - 5 \cdot 0.5)$$

$\alpha = 0.05 \Rightarrow z_\alpha = 1.645$

$\Rightarrow \beta = P(Z < 1.645 - 2.5) = P(Z < -0.855) = P(Z > 0.855) = \underline{\underline{0'1963}}$

$$c2) P(\text{aceptar } H_0 | \mu = 6) = \beta = P\left(\bar{x} < 5 + \frac{1}{5} z_\alpha \mid \mu = 6\right) = P\left(\frac{\bar{x} - 6}{1/5} < \frac{5 + \frac{z_\alpha}{5} - 6}{1/5}\right) =$$

$$= P(Z < z_\alpha - 5 \cdot 1) = P(Z < 1.645 - 5) = P(Z < -3.355) =$$

$$= P(Z > 3.355) = \underline{\underline{0'0004027}}$$

Vemos que la probabilidad de cometer error de tipo II, es decir de aceptar que el número medio es 5, cuando realmente es 5.5 es de casi el 20%, pero cuando realmente es 6 es de sólo el 0.04%!



PROBLEMA 1

(2 puntos)

En una fábrica se consulta a los empleados sobre unas nuevas medidas de seguridad a tomar. Un 65% de los empleados del turno de noche apoyan estas medidas; el 40% de las mujeres empleadas las apoyan.

Además, sabemos que el 50% de la plantilla trabaja en turno de día y el otro 50% en turno de noche. En este turno de noche, el 20% de los empleados son mujeres, como también son mujeres el 30% de la plantilla total.

Se pide:

- a) Probabilidad de que un empleado elegido al azar sea una mujer que apoya las medidas.
- b) Probabilidad de que un empleado elegido al azar sea una mujer y/o un trabajador del turno de noche.
- c) ¿Es independiente el sexo de los trabajadores de si trabajan o no en turno de noche?
- d) Si el 50% de los empleados varones apoyan el plan, calcule la probabilidad de que un empleado (varón o mujer) elegido al azar no trabaje en el turno de noche, y no apoye las nuevas medidas de seguridad.

SOLUCIÓN:

Definimos los sucesos:

A - Apoyar las medidas N - Ser de turno de noche M - Ser mujer

Datos: $P(A|N) = 0.65$ $P(A|M) = 0.4$ $P(N) = 0.5$ $P(M|N) = 0.2$ $P(M) = 0.3$

a) $P(M \cap A) = P(M) \cdot P(A|M) = 0.3 \cdot 0.4 = 0.12$

b) $P(M \cup N) = P(M) + P(N) - P(M \cap N)$

$P(M \cup N) = P(M) + P(N) - P(N) \cdot P(M|N) = 0.3 + 0.5 - 0.5 \cdot 0.2 = 0.7$

c) $\left\{ \begin{matrix} P(M) = 0.3 \\ P(M|N) = 0.2 \end{matrix} \right\} \Rightarrow P(M|N) \neq P(M)$ No son independientes

d) $P(\bar{N} \cap \bar{A}) = P(\overline{N \cup A}) = 1 - P(N \cup A)$

↑
 $P(N \cup A) = P(N) + P(A) - P(N \cap A)$

↑
 $P(N \cap A) = P(N) \cdot P(A|N)$

↑
 $P(A) = P(M) \cdot P(A|M) + P(\bar{M}) \cdot P(A|\bar{M})$

$P(A|\bar{M}) = 0.5 \Rightarrow$

Así

$P(N \cup A) = P(N) + P(M) \cdot P(A|M) + P(\bar{M}) \cdot P(A|\bar{M}) - P(N) \cdot P(A|N)$

$P(\bar{N} \cap \bar{A}) = 1 - P(N \cup A) = 1 - P(N) - P(M) \cdot P(A|M) - P(\bar{M}) \cdot P(A|\bar{M}) + P(N) \cdot P(A|N) =$
 $= 1 - 0.5 - 0.3 \cdot 0.4 - 0.7 \cdot 0.5 + 0.5 \cdot 0.65 = 0.355$



PROBLEMA 2

(2'5 puntos)

Dada la variable aleatoria bidimensional (X, Y), con función densidad conjunta

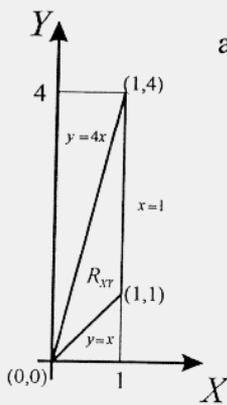
f(x, y) = { 8xy / 15 for (x,y) in R_XY, 0 otherwise

donde R_XY es el recinto plano comprendido en el interior de la poligonal cerrada formada por los puntos (0,0), (1,1) y (1,4). Se pide:

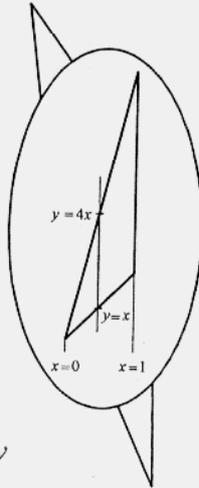
- a) Compruebe que es función de densidad conjunta.
b) Calcule la función densidad de probabilidad de la variable aleatoria unidimensional X.
c) Calcule la función de distribución de la variable aleatoria unidimensional X.
d) Calcule la función de densidad de la variable Y condicionada al valor de X.
e) Calcule la esperanza matemática de la variable X+2Y

Nota: En cada función que se le pide, debe expresar tras los cálculos, en un resumen final, el valor de la función para todos los puntos de su dominio.

SOLUCIÓN:



a) Integral from x=0 to 1 of integral from y=x to y=4x of (8xy/15) dy dx = 1



=> Integral from -infinity to infinity of integral from -infinity to infinity of f(x,y) dx dy = 1

luego es f.d.p

b) f(x) = integral from -infinity to infinity of f(x,y) dy

forall 0 <= x <= 1

f(x) = 8/15 integral from y=x to y=4x of xy dy = 4x^3

Resumen:

f(x) = { 4x^3 for 0 <= x <= 1, 0 otherwise



APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

CURSO: 3º GRUPO: _____

PROBLEMA 2: Continuación

c) $F_X(x) = \int_{t=-\infty}^{t=x} f_X(t) dt$

$\forall x \leq 0 \Rightarrow F(x) = 0$

$\forall 0 \leq x \leq 1 \Rightarrow F(x) = \int_{t=0}^{t=x} 4t^3 dt = 4 \left[\frac{t^4}{4} \right]_{t=0}^{t=x} = x^4$

$\forall x \geq 1 \Rightarrow F(x) = 1$

Resumen:

$$F(x) = \begin{cases} 0 & \forall x \leq 0 \\ x^4 & \forall 0 \leq x \leq 1 \\ 1 & \forall x \geq 1 \end{cases}$$

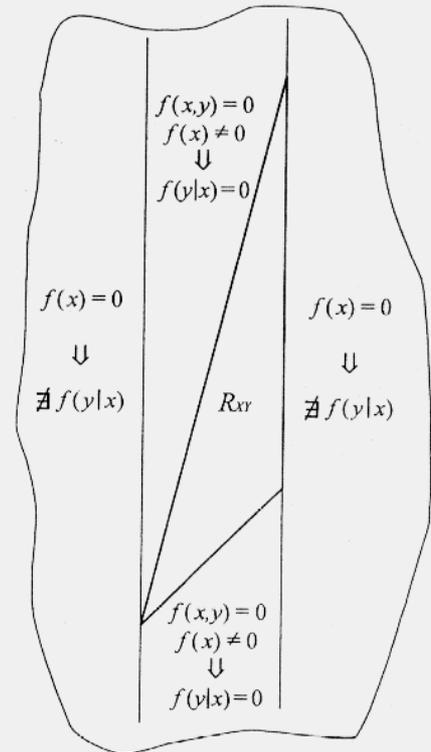
d) $f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \quad \forall f_X(x) \neq 0$

Definimos las diferentes zonas:

$\forall (x,y) \in R_{XY} \Rightarrow f(y|x) = \frac{8xy}{4x^3} = \frac{2y}{15x^2}$

Resumen:

$$f(y|x) = \begin{cases} \frac{2y}{15x^2} & \forall (x,y) \in R_{XY} \\ \emptyset & \forall x < 0, \forall x > 1 \\ 0 & \text{resto} \end{cases}$$





PROBLEMA 2: Continuación B

e) Podemos calcular la esperanza desde la función densidad de probabilidad conjunta; o calcularla como combinación lineal de las marginales, que además, podrían calcularse cada una de forma bidimensional (a través de la fdp conjunta), o unidimensional (a través de las densidades marginales).

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) \cdot dx \cdot dy \quad (\text{desde la fdpc, bidimensional})$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx \quad (\text{desde la fdp unidimensional calculada en b)}$$

$$E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot f(x, y) \cdot dx \cdot dy \quad (\text{desde la fdpc, bidimensional})$$

$$E(X + 2Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + 2y) \cdot f(x, y) \cdot dx \cdot dy \quad (\text{desde la fdpc, bidimensional})$$

Así,

$$\begin{aligned} E(X) &= \int_{x=0}^{x=1} \int_{y=x}^{y=4x} \frac{8x^2 y}{15} \cdot dx \cdot dy = \int_{x=0}^{x=1} \frac{8x^2}{15} \left[\frac{y^2}{2} \right]_{y=x}^{y=4x} \cdot dx = \int_0^1 \frac{8x^2}{15} \cdot \frac{16x^2 - x^2}{2} \cdot dx = \\ &= \int_0^1 4x^4 \cdot dx = 4 \cdot \left[\frac{x^5}{5} \right]_0^1 = \frac{4}{5} \end{aligned}$$

$$\begin{aligned} E(Y) &= \int_{x=0}^{x=1} \int_{y=x}^{y=4x} \frac{8xy^2}{15} \cdot dx \cdot dy = \int_{x=0}^{x=1} \frac{8x}{15} \left[\frac{y^3}{3} \right]_{y=x}^{y=4x} \cdot dx = \int_0^1 \frac{8x}{15} \cdot \frac{64x^3 - x^3}{3} \cdot dx = \\ &= \int_0^1 \frac{8x}{15} \cdot 21x^3 \cdot dx = \int_0^1 \frac{168}{15} x^4 \cdot dx = \frac{168}{15} \cdot \left[\frac{x^5}{5} \right]_0^1 = \frac{56}{5} \end{aligned}$$

$$E(X + 2Y) = E(X) + 2 \cdot E(Y) = \frac{4}{5} + 2 \cdot \frac{56}{5} = \frac{20 + 112}{5} = \frac{132}{5}$$



UNIVERSIDAD PONTIFICIA DE SALAMANCA EN MADRID

Departamento de Electrónica y Comunicaciones 2002/18

ESTADÍSTICA: 3-junio-2002 (Mañana)

APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

CURSO: 3º GRUPO: _____

PROBLEMA 3

(3 puntos)

Se quiere realizar un estudio sobre ciertas costumbres de los jubilados en una gran ciudad. Para ello se necesitan datos sobre sus tiempos medios de desplazamiento como peatón por la calle por semana (bien sea paseando, de compras, etc.); y sobre la proporción de jubilados que durante al menos tres días por semana se encuentran a pie en la calle entre las 5 y 5:30 de la tarde.

Damos por hecho que la distribución de estas variables es independiente de la semana elegida, por lo que se tomará una muestra suficiente, y se estudiará a todos los individuos de dicha muestra a lo largo de una semana completa.

Se le pide inicialmente, sin tener ningún dato de la población, que calcule 1) y 2):

- 1) ¿Suponiendo^(*) que la desviación típica de la permanencia en la calle por semana de los jubilados fuese de 4 horas, qué tamaño de muestra se necesitaría para lograr un nivel de confianza del 95% de estar en el valor correcto de la permanencia media a pie por semana, con un margen de ± 1 hora de error?

(*) Nota: la suposición indicada se utilizará exclusivamente en este apartado.

- 2) ¿Qué tamaño de muestra se necesitaría para obtener una confianza del 95% de estar dentro de ± 0.015 de la proporción verdadera (en tanto por uno) de personas jubiladas que se encuentran en la calle como peatones entre las 5 y las 5:30 de la tarde, al menos durante tres días por semana?

Se ha tomado una muestra de 100 personas jubiladas en esta ciudad, tomando un registro del tiempo de actividad como peatón a lo largo de una semana completa, obteniendo los siguientes resultados:

- tiempo medio como peatón en la semana completa: 15'3h
- cuasivarianza del tiempo de permanencia como peatón en la semana: 14'44h².
- 27 personas, de las 100 estuvieron en la calle entre las 5 y 5:30 de la tarde durante al menos tres días de la semana.

- 3) Obtenga un intervalo de confianza del 95% para el tiempo medio de permanencia como peatón de los jubilados por semana, en esta ciudad.

- 4) Obtenga un intervalo de confianza del 95% para la proporción de jubilados que al menos durante tres días por semana se encuentran en la calle entre las 5 y las 5:30 de la tarde.

SOLUCIÓN:

Definimos las variables en uso:

X : Tiempo de desplazamiento como peatón de un jubilado de esta ciudad.

p : Proporción de jubilados de esta ciudad que se encuentran en la calle entre las 5 y 5:30pm durante al menos tres días de la semana.

$$1) \sigma = 4h; \quad \varepsilon_{MP} = 1h; \quad 1 - \alpha = 0.95 \Rightarrow z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$$

$$z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \varepsilon_{MP} \Rightarrow n \geq \left(z_{\frac{\alpha}{2}} \frac{\sigma}{\varepsilon_{MP}} \right)^2 = \left(1.96 \cdot \frac{4}{1} \right)^2 = 7.84^2 = 61.4656 \Rightarrow \boxed{n \geq 62}$$



ESTADÍSTICA: 3-junio-2002 (Mañana)

APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

CURSO: 3º GRUPO: _____

PROBLEMA 3: Continuación

2) $\varepsilon_{MP} = 0'015$; $1 - \alpha = 0'95 \Rightarrow z_{\frac{\alpha}{2}} = z_{0'025} = 1'96$;

diseño en caso peor: $p \cdot (1 - p) = 0'25$

$$z_{\frac{\alpha}{2}} \sqrt{\frac{p \cdot (1 - p)}{n}} \leq \varepsilon_{MP} \Rightarrow n \geq \left(\frac{z_{\frac{\alpha}{2}}}{\varepsilon_{MP}} \right)^2 p \cdot (1 - p) = \left(\frac{1'96}{0'015} \right)^2 \cdot 0'25 = 4268'4 \Rightarrow \boxed{n \geq 4269}$$

Nota: $npq \geq 4269 \cdot 0'25 = 1067'25 > 9$.

Cuando se tomen muestras se comprobará con las proporciones que se midan.

3) $n = 100 > 30 \Rightarrow \boxed{IC_{0'95}(\mu) = \bar{x} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 15'3 \pm 1'96 \sqrt{\frac{14'44}{100}} = 15'3 \pm 0'7448}$

$\boxed{IC_{0'95}(\mu) = [14'5552, 16'0448]}$

Nota: Sería correcto utilizar $t_{99, 0'025} = 1'9842$.

4) En las expresiones donde necesitemos p , tomaremos $\hat{p} = \frac{27}{100}$

$$npq = 100 \cdot \frac{27}{100} \cdot \left(1 - \frac{27}{100} \right) = 19'72 > 9$$

$$\Rightarrow \boxed{IC_{0'95}(p) = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p \cdot (1 - p)}{n}} = 0'27 \pm 1'96 \sqrt{\frac{0'27 \cdot (1 - 0'27)}{100}} = 0'27 \pm 0'087016}$$

$\boxed{IC_{0'95}(p) = [0'182984, 0'357016]} \approx [18'30\%, 35'70\%]$



PROBLEMA 4

(2'5 puntos)

Una compañía de aviación quiere tener información sobre el número de pasajeros medio que tendrá en vuelos correspondientes a ciertos tramos horarios. En función de las condiciones de los contratos de catering y de adquisición de combustible, el número medio de pasajeros que la compañía considera adecuado es de 136. Por ello se diseñó hace varios meses una política de precios, que consiguiera este objetivo. Para evaluar el éxito o no de esta política de la compañía, se ha tomado nota del número de pasajeros de los últimos 22 vuelos correspondientes a dichos tramos horarios, y se han observado que suman un total de 2950 pasajeros. Para esos vuelos, se obtuvo una varianza en el número de pasajeros, de 200.

- a) Para un Error Tipo I del 5%, mediante un contraste de hipótesis, compruebe si la política de precios ha tenido éxito.
b) Suponga que desconociera cuál es la suma total de pasajeros de los 22 vuelos. Con qué probabilidad se podría afirmar, mediante un contraste de hipótesis, que la política de precios ha fracasado, cuando en realidad no es así, suponiendo que en la región de aceptación la media muestral no se aleja en más de 5 pasajeros, respecto al objetivo.
c) Calcule la potencia del contraste para una alternativa de 140 pasajeros, pero suponiendo que el número de vuelos analizados fuera 40, con una suma total de pasajeros de 2950 y varianza para esos 40 vuelos de 200. Suponga el Error Tipo I del 5%. Razone sobre el resultado.

SOLUCIÓN:

a) alpha = 0'05 n = 22

sigma_M^2 = 200 s_M^2 = n/(n-1) * sigma_M^2 = 22/21 * 200 = 209'5 s_M = 14'47

sum_{i=1}^{12} X_i = 2950 X_bar = sum_{i=1}^{12} X_i / n = 2950 / 22 = 134

Se plantea el contraste bilateral, y se utiliza el estadístico de contraste adecuado, teniendo en cuenta que la varianza poblacional es desconocida y el tamaño de la muestra es inferior a 30.

H_0: mu = 136

H_a: mu != 136

t = (X_bar - mu_0) / (s_M / sqrt(n)) = (134 - 136) / (14'47 / sqrt(22)) = 0'648

t_{alpha/2; n-1} = t_{0'025; 21} = 2'08



PROBLEMA 4: Continuación

Cae en la región de aceptación, ya que :

$$|t| \leq t_{\alpha/2; n-1} \quad 0'648 < 2'08$$

Por tanto, con un Error Tipo I del 5% se puede afirmar que la política de precios ha tenido éxito.

b) Lo que se pide es: $\alpha = P(\text{Rechazar } H_0 / H_0 \text{ cierta})$

La región de aceptación es $\bar{X} \leq 136 + 5 = 141$ ó $\bar{X} \geq 136 - 5 = 131$

$$\mu_0 + t_{\alpha/2; n-1} \frac{S_M}{\sqrt{n}} = 141 \quad t_{\alpha/2; n-1} \frac{S_M}{\sqrt{n}} = 5 \quad t_{\alpha/2; 21} = \frac{5\sqrt{22}}{14'47} = 1'62$$

Con el valor más cercano en tablas, se obtiene: $\frac{\alpha}{2} = 0'05$ $\alpha = 0'1$

c) Potencia del contraste = $1 - \beta$

$$\beta = P(\text{Aceptar } H_0 / H_a \text{ cierta}) = P\left(\frac{\mu_0 - z_{\alpha/2} \frac{s}{\sqrt{n}} - \mu_a}{s/\sqrt{n}} \leq Z \leq \frac{\mu_0 + z_{\alpha/2} \frac{s}{\sqrt{n}} - \mu_a}{s/\sqrt{n}}\right) =$$

$$= P\left(\frac{136 - 1'96 \frac{14'47}{\sqrt{40}} - 140}{14'47/\sqrt{40}} \leq z \leq \frac{136 + 1'96 \frac{14'47}{\sqrt{40}} - 140}{14'47/\sqrt{40}}\right) = P(-8'484 \leq z \leq 0'209) =$$

$$= 1 - P(z \geq 0'209) - P(z \geq 8'484) = 1 - 0'4168 - 0 = 0'5832$$

$$\text{Potencia} = 1 - \beta = 1 - 0'5832 = 0'4168 \quad \boxed{41'68\%}$$

La potencia del contraste es baja, pero eso aisladamente no significa nada.

Es necesario saber qué debe hacer el contraste ante una población cuya media diste 4 de la hipótesis nula ($140 - 136 = 4$).

Si esa diferencia de 4 es poco relevante, la potencia obtenida puede ser suficiente; pero si esa diferencia debe ser detectada y rechazada, la potencia obtenida resulta demasiado baja, y será necesario aumentar el número de vuelos analizados.



UNIVERSIDAD PONTIFICIA DE SALAMANCA EN MADRID

Departamento de Electrónica y Comunicaciones 2002/14

ESTADÍSTICA:

11-Feb-2002 (Tarde)

APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____

CURSO: 3º GRUPO: _____

PROBLEMA 1

(2'5 puntos)

En un determinado juego, en el que se trata de conseguir la máxima puntuación, la probabilidad de superar en una partida los 30 puntos es 0'6, mientras que la de quedarse por debajo de 20 puntos es 0'2. Se sabe que la variable aleatoria "puntuación obtenida en una partida" sigue una distribución normal. Además, sabemos que el resultado de cada partida es independiente del de las anteriores.

- Obtenga los valores de los parámetros de la distribución de probabilidad de dicha variable aleatoria.
- Calcule la probabilidad de que la puntuación obtenida en una partida diste de la media en más de 5 puntos.
- Suponga que se realiza un torneo en el que hay que participar en 4 partidas. Se consigue premio si en al menos 3 de las 4 partidas se logra una puntuación de al menos 40 puntos. ¿cuál es la probabilidad de conseguir premio?
- ¿Cuál sería dicha probabilidad si en el torneo se tuviera que participar en 50 partidas, y se obtuviera premio logrando 40 puntos o más, en al menos 35 de ellas?

SOLUCIÓN:

$$a) P(x > 30) = 0'6 \quad P\left(z > \frac{30 - \mu}{\sigma}\right) = 0'6 \quad P\left(z < \frac{30 - \mu}{\sigma}\right) = 0'4 \quad P\left(z > -\frac{30 - \mu}{\sigma}\right) = 0'4$$

$$0'25 - 0'26 \quad 0'4013 - 0'3974$$

$$Z_{0'4} - 0'26 \quad 0'4 - 0'3974$$

$$Z_{0'4} = 0'2533$$

$$P(x < 20) = 0'2 \quad P\left(z < \frac{20 - \mu}{\sigma}\right) = 0'2 \quad P\left(z > -\frac{20 - \mu}{\sigma}\right) = 0'2$$

$$0'84 - 0'85 \quad 0'2005 - 0'1977$$

$$Z_{0'2} - 0'85 \quad 0'2 - 0'1977$$

$$Z_{0'2} = 0'842$$

$$-\frac{30 - \mu}{\sigma} = 0'2533$$

$$-\frac{20 - \mu}{\sigma} = 0'842$$

$$\mu = 34'302$$

$$\sigma = 16'986$$

$$b) P(|x - 34'302| \geq 5) = 1 - P(|x - 34'302| < 5) = 1 - P(29'302 < x < 39'302) =$$

$$1 - P\left(\frac{29'302 - 34'302}{16'986} < z < \frac{39'302 - 34'302}{16'986}\right) = 1 - P(-0'294 < z < 0'294) = 2P(z > 0'294) = 0'7718$$



PROBLEMA 1: Continuación

c) En cada partida:

$$P(x \geq 40) = P\left(z \geq \frac{40 - 34'302}{16'986}\right) = P(z \geq 0'335) = 0'3707$$

En el torneo:

h_4 = "número de veces que se consiguen al menos 40 puntos, en 4 partidas"

Hay independencia entre cada juego, luego lo modelizamos como una binomial:

$$P(h_4 \geq 3) = \binom{4}{3} (0'3707)^3 (1 - 0'3707)^1 + \binom{4}{4} (0'3707)^4 (1 - 0'3707)^0 = 0'1265$$

d) En cada partida, la probabilidad de lograr 40 ó más puntos es la ya calculada en c):

$$P(x \geq 40) = P\left(z \geq \frac{40 - 34'302}{16'986}\right) = P(z \geq 0'335) = 0'3707$$

En el torneo:

h_{50} = "número de veces que se consiguen al menos 40 puntos, en 50 partidas"

Hay independencia entre cada juego, luego lo modelizamos como una binomial:

$$P(h_{50} = k) = \binom{50}{k} (0'3707)^k (1 - 0'3707)^{50-k} \Rightarrow P(h_{50} \geq 35) = \sum_{k=35}^{50} \binom{50}{k} (0'3707)^k (1 - 0'3707)^{50-k}$$

Pero dado que $npq = 11'664 > 9$, y por el teorema de Moivre-Laplace, se podrá realizar una aproximación de la variable binomial " h_{50} " a la normal " Y ", con distribución:

$$N(np, \sqrt{npq}) = N(18'535, 3'415),$$

que por la corrección de continuidad queda:

$$P(h_{50} \geq 35) \cong P(y > 34'5) = P\left(z > \frac{34'5 - 18'535}{3'415}\right) = P(z > 4'6749) = 0'0000013$$



PROBLEMA 2

(2'5 puntos)

Para controlar las especificaciones de un medicamento de nuevo diseño, se ha seleccionado una muestra de 20 píldoras de dicho medicamento. Se está analizando la cantidad media de miligramos de paracetamol por píldora; y para ello se observa el peso total de esta sustancia contenido en el conjunto de las 20 píldoras de la muestra, resultando ser de 9800 mg.

- a) Realizando un contraste de hipótesis con un Error Tipo I del 5%, indique a qué conclusiones llega sobre la especificación de 500 mg por píldora, teniendo en cuenta que de los datos de la muestra se sabe que: $\sum_{i=1}^{20} (x_i - \bar{x})^2 = 610$
- b) ¿Qué respondería al apartado anterior si le informan que la varianza poblacional es 50?
- c) Teniendo en cuenta que la varianza poblacional es la del apartado anterior, obtenga la potencia del contraste para una hipótesis alternativa de 490 mg. A la vista del resultado obtenido, ¿qué diría sobre el tamaño de la muestra?

SOLUCIÓN

a) $S_M^2 = \frac{1}{n-1} \sum_{i=1}^{20} (x_i - \bar{x})^2 = \frac{610}{19} = 32.1$ $S = 5.66$ $\bar{x} = \frac{9800}{20} = 490$

El contraste es bilateral:

$H_0: \mu = 500$

$H_a: \mu \neq 500$

Teniendo en cuenta que la varianza poblacional es desconocida y que el tamaño de la muestra es inferior a 30 se utiliza el estadístico de contraste es t:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{490 - 500}{\frac{5.66}{\sqrt{20}}} = -7.9$$

$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$ $t_{\alpha/2; n-1} = t_{0.025; 19} = 2.093$

La región de aceptación viene dada por:

$|t| \leq t_{\alpha/2; n-1} \Rightarrow \frac{|\bar{x} - \mu_0|}{\frac{s}{\sqrt{n}}} \leq t_{\alpha/2; n-1}$

Luego, no cumple la condición de la región de aceptación, es decir, cae fuera de dicha región: $|-7.9| > 2.093$.

Por tanto, se rechaza la hipótesis nula con un nivel de significación de 0.05.

Con una muestra de tamaño 20, con una probabilidad de error no superior al 5%, se puede afirmar que la cantidad media de miligramos de paracetamol por píldora no es 500 mg.



PROBLEMA 2: Continuación

b) $\sigma^2 = 50 \Rightarrow \sigma = 7.07$

Se realiza el anterior contraste, pero dado que ahora se conoce la varianza poblacional, el estadístico de contraste es Z:

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{490 - 500}{7.07 / \sqrt{20}} = -6.325$$

La región de aceptación viene dada por:

$$|Z| \leq Z_{\alpha/2} \Rightarrow \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \leq Z_{\alpha/2}$$

$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \quad Z_{0.025} = 1.96$

Teniendo en cuenta que: $|-6.325| > 1.96$, se llega a la misma conclusión que en el apartado anterior, es decir, de rechazo de la hipótesis nula.

c)

$H_0: \mu = 500$

$H_a: \mu = 490$

Potencia del contraste = $1 - \beta$

$$\beta = P(\text{Aceptar } H_0 | H_a \text{ cierta}) = P\left(\frac{\mu_0 - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq z \leq \frac{\mu_0 + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}{\sigma / \sqrt{n}}\right) =$$

$$= P\left(\frac{\mu_0 - \mu_a}{\sigma / \sqrt{n}} - Z_{\alpha/2} \leq z \leq \frac{\mu_0 - \mu_a}{\sigma / \sqrt{n}} + Z_{\alpha/2}\right) = P\left(\frac{500 - 490}{7.07 / \sqrt{20}} - 1.96 \leq z \leq \frac{500 - 490}{7.07 / \sqrt{20}} + 1.96\right) =$$

$$= P(4.369 \leq Z \leq 8.289) = P(Z \geq 4.369) - P(Z \geq 8.289) \approx 0.00000854$$

Potencia del contraste = $1 - 0.00000854 = \boxed{0.999992}$ 99.9992%

La potencia del contraste es muy elevada, por lo que no sería necesario aumentar el tamaño de la muestra.



UNIVERSIDAD PONTIFICIA DE SALAMANCA EN MADRID
Departamento de Electrónica y Comunicaciones

ESTADÍSTICA: 11-Feb-2002 (Tarde)

APELLIDOS Y NOMBRE: _____
 Nº DE EXPEDIENTE: _____ CURSO: 3º GRUPO: _____

PROBLEMA 3

(2'5 puntos)

Dada la variable aleatoria bidimensional (X, Y) , cuya función densidad conjunta queda definida por

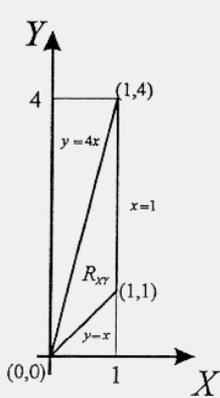
$$f(x, y) = \begin{cases} kxy & \forall (x, y) \in R_{XY} \\ 0 & \text{resto} \end{cases}$$

donde R_{XY} es el recinto plano comprendido en el interior de la poligonal formada por los puntos $(0,0)$, $(1,1)$ y $(1,4)$.

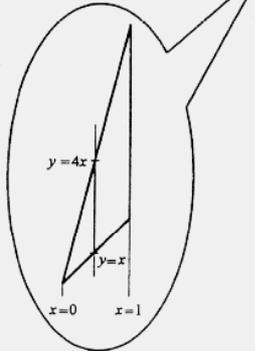
Se pide:

- Obtenga el valor de k .
- La función densidad de probabilidad de la variable aleatoria unidimensional Y .
- La función de distribución de la variable aleatoria unidimensional Y .
- La función de densidad de la variable X condicionada al valor $Y=y$.
- Esperanza matemática de la variable X y de la variable Y

SOLUCIÓN:

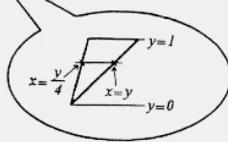
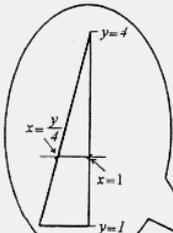


$$\begin{aligned} \text{a) } 1 &= \int_{x=0}^{x=1} \int_{y=x}^{y=4x} kxy \cdot dy \cdot dx = \int_{x=0}^{x=1} kx \left[\frac{y^2}{2} \right]_{y=x}^{y=4x} dx = \int_0^1 kx \left(\frac{16x^2 - x^2}{2} \right) dx = \\ &= \frac{15}{2} k \left[\frac{x^4}{4} \right]_0^1 = \frac{15}{8} k = 1 \Rightarrow \boxed{k = \frac{8}{15}} \end{aligned}$$



$$\text{b) } f(y) = \int_{-\infty}^{\infty} f(x, y) \cdot dx$$

$$\forall 0 \leq y \leq 1 \quad f(y) = \int_{x=\frac{y}{4}}^{x=y} kxy \cdot dx = yk \cdot \left[\frac{x^2}{2} \right]_{x=\frac{y}{4}}^{x=y} = \frac{ky \cdot \left(y^2 - \frac{y^2}{16} \right)}{2} = \frac{15}{32} ky^3 = \frac{y^3}{4}$$



$$\forall 1 \leq y \leq 4 \quad f(y) = \int_{x=\frac{y}{4}}^{x=1} kxy \cdot dx = yk \cdot \left[\frac{x^2}{2} \right]_{x=\frac{y}{4}}^{x=1} = \frac{ky \cdot \left(1 - \frac{y^2}{16} \right)}{2} = k \cdot \left(\frac{y}{2} - \frac{y^3}{32} \right) = \left(\frac{4y}{15} - \frac{y^3}{60} \right)$$



PROBLEMA 3: Continuación A

Resumen:

$$f(y) = \begin{cases} \frac{15}{32}ky^3 & \forall 0 \leq y \leq 1 \\ k \cdot \left(\frac{y}{2} - \frac{y^3}{32} \right) & \forall 1 \leq y \leq 4 \\ 0 & \text{resto} \end{cases} = \begin{cases} \frac{y^3}{4} & \forall 0 \leq y \leq 1 \\ \left(\frac{4y}{15} - \frac{y^3}{60} \right) & \forall 1 \leq y \leq 4 \\ 0 & \text{resto} \end{cases}$$

c) $\forall y \leq 0 \Rightarrow F(y) = 0$

$\forall 0 \leq y \leq 1 \Rightarrow F(y) = \int_{t=0}^{t=y} \frac{15}{32}(kt^3) \cdot dt = \frac{15}{32}k \cdot \left[\frac{t^4}{4} \right]_{t=0}^{t=y} = \frac{15}{128}ky^4 = \frac{y^4}{16}$

$\forall 1 \leq y \leq 4 \Rightarrow F(y) = \int_{t=0}^{t=1} \frac{15}{32}(kt^3) \cdot dt + \int_{t=1}^{t=y} k \left(\frac{t}{2} - \frac{t^3}{32} \right) dt = k \cdot \left\{ \frac{15}{32} \left[\frac{t^4}{4} \right]_{t=0}^{t=1} + \left[\frac{t^2}{4} - \frac{t^4}{128} \right]_{t=1}^{t=y} \right\}$

$$= \frac{k(-y^4 + 32y^2 - 16)}{128} = \frac{(-y^4 + 32y^2 - 16)}{240}$$

$\forall y \geq 4 \Rightarrow F(y) = 1$

Resumen:

$$F(y) = \begin{cases} 0 & \forall y \leq 0 \\ \frac{15}{128}ky^4 & \forall 0 \leq y \leq 1 \\ \frac{k}{128}(-y^4 + 32y^2 - 16) & \forall 1 \leq y \leq 4 \\ 1 & \forall y \geq 4 \end{cases} = \begin{cases} 0 & \forall y \leq 0 \\ \frac{y^4}{16} & \forall 0 \leq y \leq 1 \\ \frac{1}{240}(-y^4 + 32y^2 - 16) & \forall 1 \leq y \leq 4 \\ 1 & \forall y \geq 4 \end{cases}$$

d) $f(x|y) = \frac{f(x,y)}{f(y)} \quad \forall f(y) \neq 0$

Definimos las diferentes zonas:

$\forall (x,y) \in R_1 \equiv \forall y \leq 1 \text{ en } R_{XY} \equiv \begin{cases} y \leq 1 \\ y < 4x \\ y > x \end{cases} \Rightarrow f(x|y) = \frac{kxy}{\frac{15}{32}(ky^3)} = \frac{32x}{15y^2}$

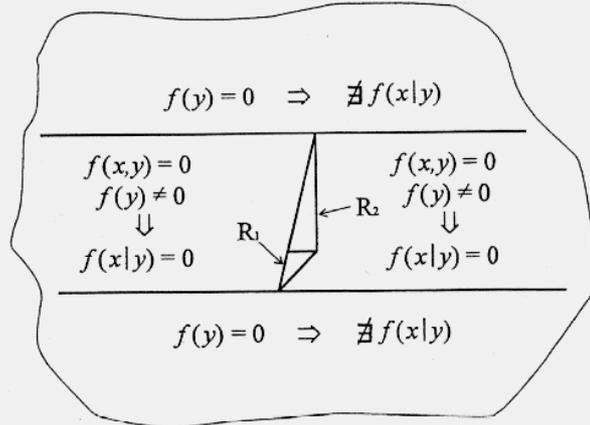
$\forall (x,y) \in R_2 \equiv \forall y \geq 1 \text{ en } R_{XY} \equiv \begin{cases} y \geq 1 \\ y < 4x \\ x < 1 \end{cases} \Rightarrow f(x|y) = \frac{kxy}{k \left(\frac{y}{2} - \frac{y^3}{32} \right)} = \frac{32x}{16 - y^2}$



PROBLEMA 3: Continuación B

Resumen:

$$f(x|y) = \begin{cases} \frac{32x}{15y^2} & \forall (x,y) \in R_1 \\ \frac{32x}{16-y^2} & \forall (x,y) \in R_2 \\ \exists & \forall y \leq 0 \cup y \geq 4 \\ 0 & \text{resto} \end{cases}$$



e) Podemos calcular las esperanzas marginales desde la función densidad de probabilidad conjunta (bidimensional), o a partir de las densidades marginales (unidimensionales). Dado que tenemos calculada la f.d.p. marginal para la Y (apartado b), pero no para la X, podemos, por ejemplo, calcular:

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x,y) \cdot dx \cdot dy \quad (\text{desde la fdpc, bidimensional})$$

$$E(Y) = \int_{-\infty}^{\infty} y \cdot f(y) \cdot dy \quad (\text{desde la fdp, unidimensional})$$

Así,

$$E(X) = \int_{x=0}^{x=1} \int_{y=x}^{y=4x} kx^2 y \cdot dx \cdot dy = \int_{x=0}^{x=1} kx^2 \left[\frac{y^2}{2} \right]_{y=x}^{y=4x} \cdot dx = \int_0^1 kx^2 \frac{16x^2 - x^2}{2} \cdot dx = \int_0^1 \frac{15}{2} kx^4 \cdot dx =$$

$$= \frac{15}{2} k \cdot \left[\frac{x^5}{5} \right]_0^1 = \frac{3}{2} k = \frac{4}{5}$$

$$E(Y) = \int_{y=0}^{y=1} y \frac{15}{32} (ky^3) \cdot dy + \int_{y=1}^{y=4} yk \left(\frac{y}{2} - \frac{y^3}{32} \right) \cdot dy = \int_{y=0}^{y=1} \frac{15}{32} ky^4 \cdot dy + \int_{y=1}^{y=4} k \left(\frac{y^2}{2} - \frac{y^4}{32} \right) \cdot dy =$$

$$= \frac{15k}{32} \cdot \left[\frac{y^5}{5} \right]_0^1 + \frac{k}{2} \cdot \left[\frac{y^3}{3} \right]_1^4 - \frac{k}{32} \cdot \left[\frac{y^5}{5} \right]_1^4 = \frac{3}{32} k + \frac{63}{6} k - \frac{1023}{160} k = \frac{2016}{480} k = \frac{21}{5} k = \frac{56}{25}$$



PROBLEMA 4

(2'5 puntos)

De una población fueron seleccionados aleatoriamente 300 individuos, de los que 93 eran fumadores.

- Estime la proporción de fumadores de la población, haciendo uso del intervalo de confianza del 95%.
- ¿Cuántas personas deberíamos seleccionar en una muestra para que, con un probabilidad de 0'978, el valor absoluto de la diferencia entre las proporciones (en tanto por uno) de fumadores existentes en la población y en la muestra fuera inferior a 0'1?

SOLUCIÓN:

a) $1 - \alpha = 0'95 \Rightarrow \frac{\alpha}{2} = 0'025 \Rightarrow z_{\alpha/2} = z_{0'025} = 1'96$

La estimación puntual que obtenemos de la muestra es $\hat{p} = \frac{93}{300} = 0'31$

Al ser muestra grande ($n = 300 > 50$), podemos tomar el estadístico $z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1-p)}{n}}}$: $N(0,1)$

Así, el intervalo de confianza solicitado es $I = \hat{p} \pm z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n}}$

Dado que no conocemos p (es, de hecho, lo que estamos estimando), lo aproximaremos por su estimación puntual. De modo que:

$$I \cong \hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} = 0'31 \pm 1'96 \sqrt{\frac{0'31 \cdot 0'69}{300}} = 0'31 \pm 0'0523 \Rightarrow I = [0'2576, 0'3623]$$

b) $1 - \alpha = 0'978 \Rightarrow \frac{\alpha}{2} = 0'011 \Rightarrow z_{\alpha/2} = z_{0'011} = 2'29$

$$\varepsilon_{\text{máx permitido}} \geq z_{\alpha/2} \sqrt{\frac{p \cdot (1-p)}{n}}$$

Dado que no conocemos p , lo aproximaremos por su estimación puntual. De modo que:

$$0'1 \geq 2'29 \sqrt{\frac{0'31 \cdot 0'69}{n}} \Rightarrow n \geq 2'29^2 \frac{0'31 \cdot 0'69}{0'1^2} = 112'17 \Rightarrow n \geq 113$$

Nota: Dado que tenemos una muestra de la población, no es necesario tomar el caso peor: $p \cdot (1-p) \cong 0'25$, y por tanto, no lo hemos hecho. La aproximación realizada: $p \cdot (1-p) = \hat{p} \cdot (1-\hat{p})$, es más realista (y por tanto siempre más optimista que el caso peor), si bien no diferirán mucho ambas soluciones, al ser $\hat{p} \cdot (1-\hat{p}) = 0'31 \cdot 0'69 = 0'2139$ cercano a 0'25. En cualquier caso, con este presupuesto, tendríamos:

a) $I = 0'31 \pm 1'96 \sqrt{\frac{0'25}{300}} \cong 0'31 \pm 0'0566 \Rightarrow I = [0'2534, 0'3666]$

b) $n \geq 2'29^2 \frac{0'25}{0'1^2} = 131'025 \Rightarrow n \geq 132$



PROBLEMA 1

(2 puntos)

Una llamada telefónica comienza en un instante aleatorio en el intervalo (0,T).

Se define la V.A. X: "instante de comienzo de la llamada", siendo: $P(t_1 \leq X \leq t_2) = \frac{t_2 - t_1}{T}$

con $t_1, t_2 \in (0, T)$. Calcule:

- a) Función de distribución de dicha variable.
- b) Su función densidad de probabilidad.

SOLUCIÓN:

a) $F(x) = \int_{-\infty}^x f(t) dt$

Dado que la V.A. está definida en (0,T), deberemos distinguir tres zonas de la variable X:

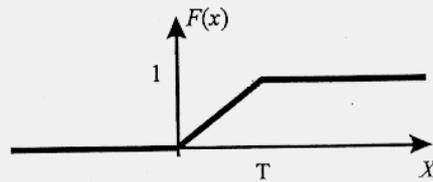
$x < 0 \Rightarrow F(x) = \int_{-\infty}^x 0 \cdot dt = 0$

$0 \leq x \leq T \Rightarrow F(x) = P(X \leq x) = P(-\infty < X \leq x) = P(-\infty < X \leq 0) + P(0 < X \leq x) = 0 + \frac{x-0}{T} = \frac{x}{T}$

$x > T \Rightarrow F(x) = P(X \leq T) = 1$

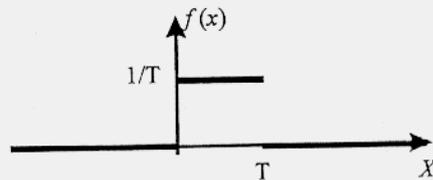
Por tanto:

$$F(x) = \begin{cases} 0 & \forall x \leq 0 \\ \frac{x}{T} & \forall 0 \leq x \leq T \\ 1 & \forall x \geq T \end{cases}$$



b) $f(x) = \frac{dF(x)}{dx}$

$$\Rightarrow f(x) = \begin{cases} 0 & \forall x < 0 \\ \frac{1}{T} & \forall 0 \leq x \leq T \\ 0 & \forall x > T \end{cases}$$



Nota: Este problema también podría haberse resuelto si se hubiera apreciado en el enunciado que $P(t_1 \leq X \leq t_2) = \frac{t_2 - t_1}{T}$ define una variable aleatoria uniforme continua.



PROBLEMA 2

(2'5 puntos)

Se han medido 100 láminas de acero en una planta de fabrica, obteniendo de estas medidas un valor medio de 0'256mm, y un valor $s^2 = 0'01$.

Se pide:

- a) Verifique el buen funcionamiento de esta fábrica, con un nivel de significación del 5%, comprobando la veracidad de la afirmación del director de la misma, que asegura que su producción tiene un grosor medio de 0'253mm.
- b) Calcule cuál es la potencia de la prueba que Vd. ha realizado en a), para una producción con grosor medio de 0'21mm.

SOLUCIÓN:

- a) Se trata de un contraste bilateral para la media poblacional de una normal de la que desconocemos su desviación típica.

El contraste se plantea: .

Hipótesis nula, $H_0 : \mu = 0'253mm (= \mu_0)$

Hipótesis alternativa, $H_a : \mu \neq 0'253mm$

Al ser grande la muestra ($n = 100 > 50$), el estadístico de contraste $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ será $N(0,1)$.

Los puntos críticos serán:

$$z_{C1} = -z_{\alpha/2} = \frac{\bar{x}_{C1} - \mu_0}{s/\sqrt{n}} ; z_{C2} = z_{\alpha/2} = \frac{\bar{x}_{C2} - \mu_0}{s/\sqrt{n}}$$

siendo $z_{\alpha/2} = z_{0'025} = 1'96$

$$\left. \begin{matrix} z_{C1} = -1'96 \\ z_{C2} = 1'96 \end{matrix} \right\} \Rightarrow \begin{cases} \text{Región de aceptación } |z_r| \leq 1'96 \\ \text{Región de aceptación } |z_r| > 1'96 \end{cases}$$

En nuestro caso:

$$z_r = \frac{\bar{x}_r - \mu_0}{s/\sqrt{n}} = \frac{0'256 - 0'253}{\sqrt{0'01}/\sqrt{100}} = \frac{0'003}{10^{-2}} = 0'3 \Rightarrow |z_r| = 0'3 \leq 1'96 \Rightarrow$$

\Rightarrow se acepta la hipótesis H_0

\Rightarrow La afirmación del director de la fábrica se considerará cierta, con significación del 5%



PROBLEMA 2: Continuación

b) $Pot = 1 - \beta$

$$\beta(\mu = \mu_a) = [\text{Prob. error Tipo II, para } \mu = \mu_a] = P[\text{Aceptar } H_0 | \mu = \mu_a] = P(\bar{x}_{C1} \leq \bar{x} \leq \bar{x}_{C2} | \mu = \mu_a) =$$

$$= P\left(\frac{\bar{x}_{C1} - \mu_a}{s/\sqrt{n}} \leq z \leq \frac{\bar{x}_{C2} - \mu_a}{s/\sqrt{n}}\right)$$

Recordando del apartado anterior,

$$\left. \begin{aligned} z_{C1} = -z_{\alpha/2} &= \frac{\bar{x}_{C1} - \mu_0}{s/\sqrt{n}} \Rightarrow \bar{x}_{C1} = \mu_0 - z_{\alpha/2} \frac{s}{\sqrt{n}} \\ z_{C2} = z_{\alpha/2} &= \frac{\bar{x}_{C2} - \mu_0}{s/\sqrt{n}} \Rightarrow \bar{x}_{C2} = \mu_0 + z_{\alpha/2} \frac{s}{\sqrt{n}} \end{aligned} \right\} \Rightarrow \begin{cases} \bar{x}_{C1} = 0'253 - 1'96 \sqrt{\frac{0'01}{100}} = 0'2334 \\ \bar{x}_{C2} = 0'253 + 1'96 \sqrt{\frac{0'01}{100}} = 0'2726 \end{cases}$$

$$\Rightarrow \beta(\mu = \mu_a) = P\left(\frac{\mu_0 - \mu_a}{s/\sqrt{n}} - z_{\alpha/2} \leq z \leq \frac{\mu_0 - \mu_a}{s/\sqrt{n}} + z_{\alpha/2}\right)$$

Para nuestro caso particular:

$$\Rightarrow \beta(\mu = 0'21) = P\left(\frac{0'253 - 0'21}{10^{-2}} - 1'96 \leq z \leq \frac{0'253 - 0'21}{10^{-2}} + 1'96\right) =$$

$$= P(4'3 - 1'96 \leq z \leq 4'3 + 1'96) = P(2'34 \leq z \leq 6'26) =$$

$$= \alpha(2'34) - \alpha(6'26) = 0'00964 - \varepsilon$$

(donde $0'0^9149 < \varepsilon < 0'0^{10}777$)

$$\Rightarrow \beta(\mu = 0'21) \cong 0'00964$$

$$\Rightarrow \boxed{Pot = 1 - \beta(\mu = 0'21) \cong 1 - 0'00964 = 0'99036 \cong 99\%}$$



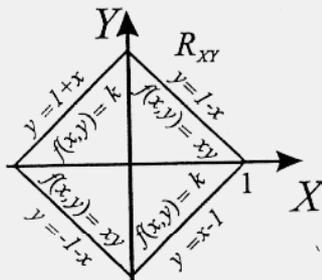
PROBLEMA 3

(3 puntos)

Se define la V.A. bidimensional (X,Y) , mediante su f.d.p.c., que toma valores distintos de cero en el cuadrado delimitado por los puntos: $(1,0)$, $(0,1)$, $(-1,0)$ y $(0,-1)$, de tal forma que en las zonas del cuadrado correspondientes al segundo y cuarto cuadrantes, la f.d.p.c. es uniforme de valor K ; mientras que en el primer y tercer cuadrantes, la f.d.p.c. toma la expresión: xy .

- Calcule el valor de K .
- Obtenga las funciones de densidad marginales.
- Compruebe la independencia o no de sus variables.
- Calcule la covarianza entre las variables.
- Obtenga la recta de regresión de la variable Y sobre la X .
- Mediante la recta que minimice el error cuadrático medio, ¿qué estimación daría como valor de la variable Y , para un valor de $X = 0.8$?

SOLUCIÓN:



Por simetría del recinto y de la f.d.p.c. respecto de la recta $y = x$, tenemos:

$$\alpha_{ab} = \alpha_{ba}$$

$$\mu_{ab} = \mu_{ba}$$

$$a) \int_0^1 \int_0^{1-x} xy \, dy \, dx + \int_{-1}^0 \int_0^{1+x} k \, dy \, dx + \int_{-1}^0 \int_{-x-1}^0 xy \, dy \, dx + \int_0^1 \int_{x-1}^0 k \, dy \, dx = \frac{1}{12} + 2k - k = 1 \quad \boxed{k = \frac{11}{12}}$$

b) Función de densidad marginal de la variable x :

$$\int_{-x-1}^0 xy \, dy + \int_0^{1+x} \frac{11}{12} \, dy = \frac{-x^3 - x}{2} - x^2 + kx + k \quad \int_{x-1}^0 \frac{11}{12} \, dy + \int_0^{1-x} xy \, dy = \frac{x^3 + x}{2} - x^2 - kx + k$$

$$f(x) = \begin{cases} \frac{-x^3 - x}{2} - x^2 + kx + k & -1 \leq x \leq 0 \\ \frac{x^3 + x}{2} - x^2 - kx + k & 0 < x \leq 1 \\ 0 & \text{resto} \end{cases}$$



UNIVERSIDAD PONTIFICIA DE SALAMANCA EN MADRID

Departamento de Electrónica y Comunicaciones 2002'02

ESTADÍSTICA: 11-Feb-2002 (Mañana)

APELLIDOS Y NOMBRE: _____

Nº DE EXPEDIENTE: _____ CURSO: 3º GRUPO: _____

PROBLEMA 3: Continuación

Análogamente para la variable y, se obtiene:

$$f(y) = \begin{cases} \frac{-y^3 - y}{2} - y^2 + ky + k & -1 \leq y \leq 0 \\ \frac{y^3 + y}{2} - y^2 - ky + k & 0 < y \leq 1 \\ 0 & \text{resto} \end{cases}$$

- c) No son independientes, ya que no se cumple la igualdad $f(x,y) = f(x)f(y) \forall (x,y)$
 Utilizando un contraejemplo podemos demostrarlo: evaluando en el punto (0'8, 0'6):

$$f(0'8,0'6) = 0; \quad f(0'8) \cdot f(0'6) = (\neq 0) \cdot (\neq 0) \neq 0 \quad \text{c.q.d.}$$

d) $\mu_{11} = \alpha_{11} - \alpha_{10}\alpha_{01}$

$$\alpha_{10} = \int_{-1}^0 x \left(\frac{-x^3 - x}{2} - x^2 + kx + k \right) dx + \int_0^1 x \left(\frac{x^3 + x}{2} - x^2 - kx + k \right) dx = 0$$

De igual forma se tiene que : $\alpha_{01} = 0$ (por simetría del recinto y de la fdpc)

$$\alpha_{11} = \int_0^1 \int_0^{1-x} xyxy \, dy \, dx + \int_{-1}^0 \int_0^{1+x} xy \frac{11}{12} \, dy \, dx + \int_{-1}^0 \int_{-x-1}^0 xyxy \, dy \, dx + \int_0^1 \int_{-1}^0 xy \frac{11}{12} \, dy \, dx = 1'2635$$

Por tanto la covarianza queda: $\mu_{11} = 1'2635$

- e) Teniendo en cuenta que la expresión de la recta de regresión es: $\hat{y} - \mu_y = \frac{\mu_{11}}{\mu_{20}} (x - \mu_x)$, nos falta la varianza de la variable x. Pero dado que : $\mu_{20} = \alpha_{20} - \alpha_{10}^2$, basta con calcular: α_{20}

$$\alpha_{20} = \int_{-1}^0 x^2 \left(\frac{-x^3 - x}{2} - x^2 + kx + k \right) dx + \int_0^1 x^2 \left(\frac{x^3 + x}{2} - x^2 - kx + k \right) dx = 0'1693$$

$$\mu_{20} = 0'1693 - 0 = 0'1693$$

Sustituyendo finalmente en la recta de regresión: $\hat{y} = \frac{1'2635}{0'1693} x$ $\hat{y} = 7'463x$

f) $\hat{y} = 7'463 \cdot 0'8 = \boxed{5'97}$



PROBLEMA 4

(2'5 puntos)

En una empresa que fabrica neumáticos para vehículos especiales, se obtiene una muestra de 22 unidades y se observa que el diámetro medio es de 1 metro. La variable X que mide dicho diámetro sigue una distribución normal. Sabiendo que la varianza muestral es 0'13.

- Obtenga un intervalo de confianza del 98% para la varianza del diámetro de los neumáticos fabricados.
- Obtenga un intervalo de confianza del 95% para el diámetro medio de estos neumáticos.
- Se quiere tener una probabilidad 0'99 de que el diámetro medio en la población no se aleje del diámetro medio obtenido en la muestra en más de 10 centímetros. ¿Cuántos neumáticos necesitaríamos en la muestra?

SOLUCIÓN

a) $1 - \alpha = 0'98 \Rightarrow \frac{\alpha}{2} = 0'01$; $1 - \frac{\alpha}{2} = 0'99$
 $\sigma_M^2 = 0'13 \Rightarrow S_M^2 = \frac{n}{n-1} \sigma_M^2 = \frac{22}{21} 0'13 = 0'136$

$$I = \left[\frac{(n-1)S_M^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)S_M^2}{\chi_{1-\alpha/2, n-1}^2} \right] = \left[\frac{21 \cdot 0'136}{\chi_{0'01, 21}^2}, \frac{21 \cdot 0'136}{\chi_{0'99, 21}^2} \right] = \boxed{(0'0733, 0'321)}$$

- b) Al ser la varianza de la población desconocida, y la muestra menor que 30, se obtiene el intervalo con la t-student:

$$1 - \alpha = 0'95 \Rightarrow \frac{\alpha}{2} = 0'025$$

$$I = \bar{x} \pm t_{\alpha/2, n-1} \sqrt{\frac{S_M^2}{n}} = 1 \pm t_{0'025, 21} \sqrt{\frac{0'136}{22}} = 1 \pm 0'1635 = \boxed{(0'8365, 1'1635)}$$

- c) Al ser un intervalo menor que en el apartado b) , y con una mayor probabilidad, necesitará de una muestra superior, que muy posiblemente supere 30. Por tanto:

$$1 - \alpha = 0'99 \Rightarrow \frac{\alpha}{2} = 0'005 \quad E = Z_{\alpha/2} \sqrt{\frac{S_M^2}{n}} \Rightarrow 0'1 = Z_{0'005} \sqrt{\frac{0'136}{n}}$$

Aproximando sin interpolar: $Z_{0'005} = 2'58$

Sustituyendo el valor obtenido en tablas, y despejando n: $n = \frac{0'136}{\left(\frac{0'1}{2'58}\right)^2} = 90'66 \quad \boxed{n \geq 91}$

(Al ser mayor de 30, ratifica la aproximación realizada)