**Grado en Ingeniería Informática**

# Estadística

# PRIMER PARCIAL

31 de marzo de 2014

| Apellidos, nombre: | |
|---|---|
| Grupo: | Firma: |

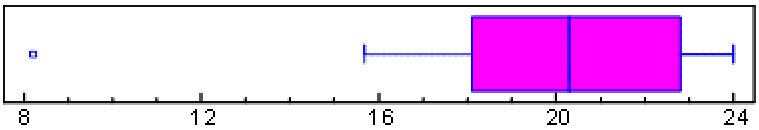
## Instrucciones

1. Rellenar la información de cabecera del examen.
2. Responder a cada pregunta en la hoja correspondiente.
3. Justificar todas las respuestas.
4. No se permiten anotaciones personales en el formulario.
5. No se permite tener teléfonos móviles encima de la mesa. Sobre la mesa sólo se permite el DNI, calculadora, útiles de escritura, las tablas y el formulario.
6. No desgrapar las hojas.
7. Todas las preguntas puntúan lo mismo (sobre 10).
8. Se debe firmar en las hojas que hay en la mesa del profesor al entregar el examen. Esta firma es el justificante de la entrega del mismo.
9. Tiempo disponible: **2 horas**

**1.** Un cierto programa informático realiza búsquedas en una base de datos de tamaño considerable. Con el fin de mejorar la eficiencia del programa, se realiza una modificación y se ensaya 13 veces, obteniéndose los siguientes valores de tiempo (en milisegundos):

{ 8,2 ; 15,7 ; 17,3 ; 18,1 ; 18,5 ; 19,2 ; 20,3 ; 21,1 ; 21,6 ; 22,8 ; 23,2 ; 24 ; 24 }

Para explorar la muestra se representan los datos mediante el siguiente gráfico:




**a)** ¿Cuál es la población objeto de estudio? ¿Cuáles son los individuos de dicha población? *(1,5 puntos)*

**b)** Calcular el primer cuartil y el rango. *(2 puntos)*

**c)** Indica cómo se denomina este tipo de gráficos. *(1 punto)*

**d)** A la vista del gráfico anterior y teniendo en cuenta que el coeficiente de asimetría estandarizado vale -2,276, ¿puede afirmarse que los tiempos de búsqueda en la base de datos siguen una distribución asimétrica negativa? *(2 puntos)*

**e)** Obtener, aproximadamente, un parámetro de posición que sea representativo de los valores de la muestra. *(1,5 puntos)*

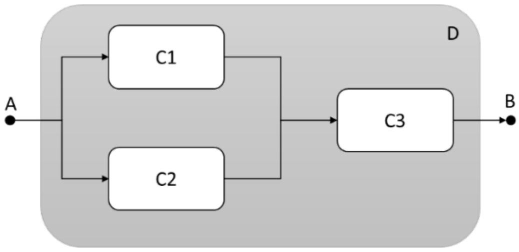
**f)** ¿Qué método emplearías para estudiar si los datos proceden de una distribución Normal? ¿Qué ventajas presenta dicho método frente al gráfico anterior? *(2 puntos)*

**2.** Una industria fabrica componentes electrónicos (C) cuya vida media es de 4 años. Asumiendo que la duración de este tipo de componentes se puede modelar según la ley exponencial, se pide:

**a)** Calcula el porcentaje de los componentes fabricados que tendrán una duración superior a los 5 años (define la variable aleatoria utilizada, su distribución y sus parámetros). *(2,5 puntos)*

**b)** Sabiendo que una componente lleva funcionando un año, calcula la probabilidad de que ésta siga funcionando correctamente otros 5 años más (es decir, después de 6 años en total). *(2,5 puntos)*

**c)** Calcula la fiabilidad a los 5 años del dispositivo **D** que se muestra en la figura, sabiendo que C1, C2 y C3 son componentes electrónicos del tipo descrito en el enunciado y de funcionamiento independiente. Define y describe las variables aleatorias utilizadas, así como lo que se entiende por dicha fiabilidad. *(5 puntos)*




**3.** Un fabricante de placas base desea garantizar que la proporción de placas defectuosas en los lotes que produce es inferior al 1%. El fabricante establece un plan de muestreo que consiste en seleccionar N placas y aceptar el lote si hay menos de 2 defectuosas.

¿Cuánto debe valer N como mínimo para que la probabilidad de rechazar un lote que no cumpla el requisito deseado sea superior al 99%? *(10 puntos)*

**4.** Se sabe que el tiempo de compilación de un determinado tipo de programas fluctúa uniformemente entre 20 y 30 segundos. Se pide indicar la variable aleatoria y contestar a las siguientes preguntas:

**a)** ¿Cuál es la probabilidad de que el programa se compile antes de 25 segundos? *(2 puntos)*

**b)** Si se compilan 10 programas secuencialmente, ¿cuál es la probabilidad de que el tiempo total de ejecución sea inferior a 4,5 minutos? *(4 puntos)*

**c)** Si el tiempo total de ejecución se expresa en minutos, ¿cuánto valdría la varianza? ¿Qué unidades tendría? *(4 puntos)*

**SOLUCIÓN:**

**1a)** La población es el conjunto de búsquedas que pueden realizarse con el programa modificado. Los individuos corresponden a cada una de las posibles búsquedas que pueden realizarse sobre la base de datos con el programa.

**1b)** El primer cuartil es 18,1 (extremo izquierdo de la caja), que corresponde al cuarto valor (de menor a mayor).
Por definición, rango = máximo - mínimo = 24 - 8,2 = 15,8

**1c)** Se denomina diagrama box-whisker (diagrama de caja - bigotes).

**1d)** No se puede afirmar. El bigote derecho es más corto que el izquierdo, lo que sugiere una distribución asimétrica negativa si la mediana está desplazada hacia la derecha. Pero no es el caso (mediana=20,3), por lo que no hay suficiente evidencia para afirmar que la distribución sea asimétrica. Asumiendo que la distribución es normal, el valor 8,2 está muy separado del extremo del bigote, lo que sugiere que es un dato anómalo que conviene descartar. El coeficiente de asimetría está muy condicionado por el valor anómalo, de modo que no es un valor útil en este caso para discutir la asimetría de la población.

**1e)** Dado que existe un dato anómalo según lo discutido en el apartado anterior, la mediana (20,3) es un parámetro de posición más representativo que la media (19,54) en este caso. Si se descarta el dato anómalo, la media es 20,48 que también sería un parámetro de posición adecuado.

**1f)** Se pueden representar los valores sobre un papel probabilístico normal, ya que es más útil para estudiar la normalidad que el box-whisker en el caso de pocos datos y para diagnosticar la posible existencia de datos anómalos.

**2a)** v.a. **T** = {Duración (años) de una componente C} ~ Exp($\alpha$)
$m=1/\alpha \rightarrow \alpha = 1/m = ¼$ ; $P(T>5) = e^{-5/4} = 0{,}287 \rightarrow$ **<u>28,7%</u>**

**2b) P(T>6/T>1)** = P(T > 6-1) = P(T > 5) = $e^{-5/4}$ =**<u>0,287</u>**.
La distribución exponencial NO tiene memoria.

**2c) variables aleatorias**:
T1={Duración (años) de la componente C1}
T2={Duración (años) de la componente C2}
T3={Duración (años) de la componente C3}
TD={Duración (años) del dispositivo D}
T1, T2 y T3 siguen una distribución Exponencial de parámetro $\alpha$ = ¼

La **fiabilidad a los 5 años** se define como la **P(TD>5)**

- **$P(TD>5) = P[((T1>5)\cup(T2>5))\cap(T3>5)]$** $= P[(T1>5)\cup(T2>5)]\ P(T3>5) =$
$= [P(T1>5) + P(T2>5) - P((T1>5)\cap(T2>5))]\ P(T3>5) =$
$= [P(T1>5) + P(T2>5) - P(T1>5)\ P(T2>5)]\ P(T3>5)$

- P(T1>5) = P(T2>5) = P(T3>5) = 0,287 (apartado a)

- **<u>P(TD>5)</u>** = [P(T1>5) + P(T2>5) – P(T1>5) P(T2>5)] P(T3>5) = (0,287 + 0,287 - $0{,}287^2$) 0,287 = **<u>0,141</u>**

**3a)** X: Nº de placas base defectuosas en la muestra de tamaño N
X se distribuye con el modelo Binomial de parámetros N y $p\geq0{,}01$ para los lotes rechazables.
Se aproxima como Poisson con $\lambda=N\cdot0{,}01$ en el caso más difícil de detectar.
$P(X>1)>0{,}99$ por lo que $P(X\leq1)\leq0{,}01$ mirando el ábaco $\lambda\geq6{,}5$ de donde $N\geq6{,}5/0{,}01=650$

**4a)** X: tiempo de compilación de un programa en segundos, X ~ Un (20, 30)
P(X < 25) = 0,5 por ser la mediana (valor medio entre 20 y 30).

**4b)** E(X) = 25, $Var(X) = (b-a)^2/12 = 10^2/12 = 8{,}33$
Y= tiempo de compilación de 10 programas en segundos:
$Y = X_1 + ... + X_{10}$ ; $Y \sim N\ (m=10\cdot25,\ \sigma^2=10\cdot8{,}33)$ ; $Y \sim N\ (250, \sqrt{83{,}3})$
4,5 minutos = 270 segundos
$P(Y< 270) = P(Z < (270-250)/\sqrt{83{,}3}) = P(Z < 2{,}19) = 1 - P(Z > 2{,}19) =$
$= 1 - 0{,}0143 =$ **0,9857**

**4c)** Z= tiempo de compilación de 10 programas en minutos
$Z = Y / 60 = (1/60)\cdot Y$
$Var (Z) = (1/60)^2\cdot var(Y) = 83{,}3/60^2 =$ **0,02315**
Las unidades de la varianza son cuadráticas, por lo que serían minutos$^2$.