

Grado en Ingeniería Informática

Estadística

PRIMER PARCIAL

30 de marzo de 2015

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. Responde a los siguientes apartados:

a) ¿Es posible que un conjunto de datos tenga mayor curtosis que otro y, sin embargo, tenga menor desviación típica? Justifica la respuesta.

(2 puntos)

b) Indica cuál es la población o poblaciones, la variable aleatoria (o variables aleatorias) implicada/s, la dimensión de la variable considerada y el tipo de cada una de ellas en los siguientes ejemplos:

1. Número de consultas atendidas y número de peticiones pendientes diariamente en cierta empresa informática. *(2 puntos)*

2. Volumen de ventas de una muestra de 100 empresas españolas del sector informático durante el año 2014. *(2 puntos)*

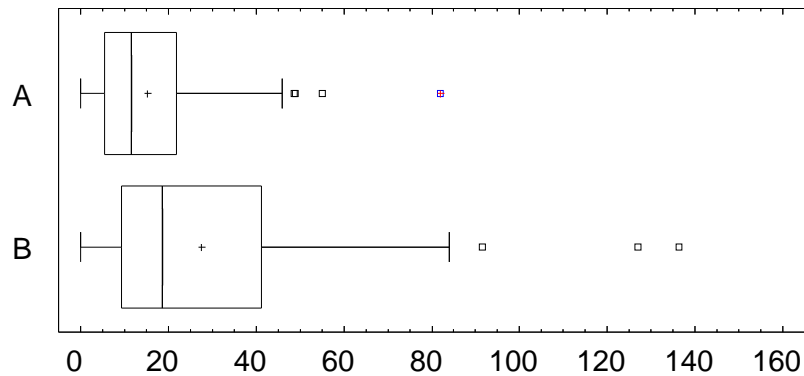
c) En una industria se mide un cierto parámetro de calidad que sigue una distribución Normal cuando el proceso está bajo control. Explica con detalle las distintas pautas no normales que podrían detectarse mediante un histograma de frecuencias de los datos, indicando cómo se pondrían de manifiesto dichas pautas atípicas en el gráfico del histograma. *(4 puntos)*

2. Una empresa que ofrece un determinado servicio por internet trabaja con 3 servidores distintos (A, B y C). El 25% de las solicitudes (accesos) de este servicio se ejecutan en el servidor A, el 35% en el servidor B y el resto en el C. Se tiene constatado que en el caso del servidor A se produce algún error en el 1% de los accesos, mientras que este porcentaje es del 2% en el caso del servidor B. Además, se sabe que si se elige al azar un acceso, la probabilidad de que tenga algún error y se haya ejecutado en el servidor C es del 1%.

a) Dada una solicitud al azar, ¿cuál es la probabilidad de que durante el acceso al servidor se produzca algún error? *(5 puntos)*

b) Sabiendo que un acceso ha sido correcto, calcula la probabilidad de que éste se haya hecho sobre el servidor A. *(5 puntos)*

3. Para la fabricación de cierto dispositivo electrónico se emplean componentes de dos tipos (A y B). Tras obtener experimentalmente el tiempo de funcionamiento hasta el fallo (en meses) de 100 componentes de cada tipo, se obtiene el siguiente gráfico box-whisker múltiple (gráfico superior obtenido con los tiempos de A y gráfico inferior con los de B). Se asume que los tiempos siguen una distribución exponencial, y que la media del tiempo de B es 27,5.



a) Si se monta un circuito con 4 componentes en serie, siendo dos de tipo A y dos de tipo B (es decir: -A-A-B-B-), calcular la probabilidad de que el circuito falle antes de un año. Se asume que los fallos son independientes del tipo de componente. (5 puntos)

b) ¿Cuántos componentes de tipo B deberían conectarse en paralelo como mínimo para obtener una fiabilidad superior al 95% tras un año de funcionamiento? (5 puntos)

4. El sistema de comprobación de errores a nivel de red de un sistema de comunicaciones tiene que garantizar el requisito de que el porcentaje de paquetes incorrectos que circulan por la red sea inferior al **uno** por mil. Para ello en cada sesión el sistema de comprobación de errores, a intervalos constantes de tiempo, selecciona al azar n paquetes y verifica si son o no correctos. Si el sistema de comprobación de errores encuentra 3 o más paquetes incorrectos, no acepta la sesión como íntegra impidiendo la comunicación temporalmente y desplegando el procedimiento de corrección.

Determinar el valor mínimo del número de paquetes a comprobar, si se desea que la probabilidad de aceptar como íntegra una sesión que no satisfaga el requisito exigido, sea menor o igual al 5 por mil. *(10 puntos)*

5. Una empresa fabrica cierto tipo de componente electrónico para su uso en equipos informáticos. Se controla un parámetro de calidad el cual sigue una distribución Normal de media 90. Se considera que el componente es defectuoso cuando este parámetro es superior a 95.

a) La empresa ha constatado que produce un 2% de componentes defectuosos. Calcular la desviación típica del parámetro de calidad. *(5 puntos)*

b) Si selecciona al azar una muestra de 2000 componentes, ¿cuánto vale aproximadamente la probabilidad de que haya menos de 60 defectuosos?

(5 puntos)

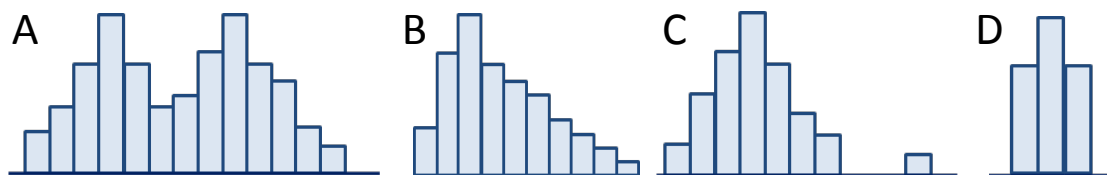
SOLUCIÓN

1a) Sí. El coeficiente de curtosis (CC) aporta una información muy distinta y absolutamente independiente de la dispersión. El CC indica la posible anomalía: $CC > 0$, leptocurtosis (tomando 0 como valor de referencia) o truncamiento de datos ($CC < 0$, planicurtosis), mientras que la dispersión de los datos (medida, por ejemplo, con la desviación típica) puede ser baja o elevada sin que existan datos anómalos ni truncados.

1b1) Población: Todos los días de la semana. Sobre cada día se registra una variable bidimensional (nº de consultas atendidas, nº de peticiones pendientes). Ambas componentes de la variable bidimensional son de naturaleza cuantitativa y discretas.

1b2) Población: Todas las empresas españolas del sector informático. Las 100 empresas a las que hace referencia el enunciado constituyen una muestra de dicha población. De cada empresa se registra la variable unidimensional volumen de ventas durante 2014 (no se especifica en el enunciado en qué unidades viene medida). Puesto que, en general, el volumen de ventas hace referencia al importe de los productos vendidos, se considera una variable unidimensional de naturaleza cuantitativa y continua.

1c) Un histograma construido a partir de un conjunto de datos que supuestamente siguen una distribución normal podría poner de manifiesto distintos tipos de problemas:



- Fig. A: Mezcla de poblaciones (en relación a lo que se plantea en el enunciado, podríamos estar representando una característica de calidad de dos tipos de productos diferentes).

- Fig. B: No normalidad de los datos. De esta manera un histograma como el representado en la figura B indicaría asimetría positiva.

- Fig. C: Existencia de datos marcadamente bajos o altos (en el ejemplo presentado, algún producto fabricado con un valor “anormalmente” alto para la característica de calidad que se está estudiando).

- Fig. D: Datos truncados. Se están despreciando valores en la muestra de datos recogidos. En relación a lo planteado en el enunciado, podrían haberse eliminado aquellos productos cuya característica de calidad presentara valores no comprendidos en cierto intervalo.

- Otras pautas no normales podrían ser: histograma con forma planicúrtica, frecuencia anómala de algún valor (altura de alguna barra del histograma claramente distinta de lo esperado) debido a la manipulación de los valores (posible redondeo), etc.

2a) Sucesos:A: el acceso se realiza sobre el servidor A $\rightarrow P(A) = 0,25$ B: el acceso se realiza sobre el servidor B $\rightarrow P(B) = 0,35$ C: el acceso se realiza sobre el servidor C $\rightarrow P(C) = 1 - P(A) - P(B) = 0,4$

E: el acceso es erróneo (se produce algún error)

 $P(E/A) = 0,01$; $P(E/B) = 0,02$; $P(E \cap C) = 0,01$

Según el Teorema de la Probabilidad Total:

$$P(E) = P(E \cap A) + P(E \cap B) + P(E \cap C) = P(A) \cdot P(E/A) + P(B) \cdot P(E/B) + P(E \cap C) = 0,25 \cdot 0,01 + 0,35 \cdot 0,02 + 0,01 = \mathbf{0,0195}$$

2b) Según el Teorema de Bayes:

$$P(A/\bar{E}) = \frac{P(A \cap \bar{E})}{P(\bar{E})} = \frac{P(A) \cdot P(\bar{E}/A)}{1 - P(E)} = \frac{0,25 \cdot (1 - 0,01)}{1 - 0,0195} = \mathbf{0,2524}$$

3a) Suceso A: el componente de tipo A funciona más de un año.

Suceso B: el componente de tipo B funciona más de un año

Suceso C: el circuito en serie -A-A-B-B- funciona más de un año.

$$T_B \approx \exp(\alpha = 1/27.5) ; P(T_B > 12) = e^{-\alpha \cdot t} = e^{-12/27.5} = 0.646$$

El valor medio de los datos de A se representa como una pequeña cruz dentro de la caja, el cual vale aproximadamente 15 de modo que $\alpha = 1/15$.

$$P(T_A > 12) = e^{-\alpha \cdot t} = e^{-12/15} = 0.449$$

$$P(C) = P(A_1 \cap A_2 \cap B_1 \cap B_2) = P(A_1) \cdot P(A_2) \cdot P(B_1) \cdot P(B_2) = 0.449^2 \cdot 0.646^2 = 0.084$$

La probabilidad de que falle antes de un año será: $1 - 0.084 = \mathbf{0.916}$ **3b) Suceso F:** el componente de tipo B falla antes de un año.Suceso D: el circuito con n componentes de tipo B en paralelo falla antes de un año.Fiabilidad (probabilidad de que el circuito funcione más de un año) $> 95\%$;

$$P(D) < 0.05$$

$$P(F) = P(T_B < 12) = 1 - e^{-\alpha \cdot t} = 1 - e^{-12/27.5} = 0.354$$

$$P(D) = P(F_1 \cap \dots \cap F_n) = P(F_1) \cdot \dots \cdot P(F_n) = [P(F)]^n < 0.05 ; 0.354^n < 0.05$$

$$n > (\log 0.05) / \log (0.354) ; n > 2.88 ; \mathbf{n_{min} = 3}$$

4) v.a X = nº de paquetes incorrectos $\sim B(n, p < 0,001)$ Si $X \leq 2$ se acepta la sesión como íntegra; Si $X \geq 3$ se rechaza la sesión.¿n? / $P(\text{aceptar sesión íntegra} / p \leq 0,001) = P(X \leq 2 / p \leq 0,001) \leq 0,005$

$$P(X \leq 2 / p \leq 0,001) = P[B(n, p \leq 0,001) \leq 2] = P[Ps(\lambda = n \cdot 0,001) \leq 2] \leq \mathbf{0,005}$$

Mirando el Ábaco de Poisson se tiene que $\lambda \geq 9,25$

$$n \cdot 0,001 \geq 9,25 \rightarrow n \geq 9,25/0,001 \rightarrow \mathbf{n \geq 9250}$$

5a) $P[N(90, \sigma) > 95] = 0.02$; $P[N(0; 1) > (95 - 90) / \sigma] = 0.02$; $(95 - 90) / \sigma = 2.05$;
 $\sigma = 2.44$

5b) Y= n° componentes defectuosos en muestra de 2000 componentes

Y es B (n=2000, p=0,02). Dado que $n \cdot p \cdot (1-p) = 39,2 > 9$ se puede aproximar con la distribución normal de media $n \cdot p = 40$ y varianza 39,2 ($\sigma = 6,26$)

$P(Y < 60) \approx P[N(40; 6,26) < 59,5] = P[N(0, 1) < 3,11] = 1 - 0,00094 = \mathbf{0,99906}$