Grado en Ingeniería Informática

Estadística

PRIMER PARCIAL

2 de abril de 2012

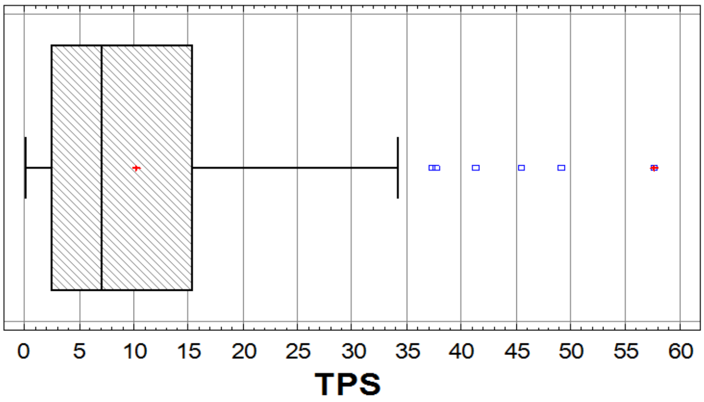
Apellidos, nombre	
Grupo, Firma	

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. Sobre la mesa sólo se permite el DNI, calculadora, útiles de escritura, las tablas y el formulario.
5. No desgrapar las hojas.
6. Todas las preguntas puntúan lo mismo (sobre 10).
7. 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.
8. Tiempo disponible: 2 horas

1. En el estudio del rendimiento de un sistema operativo se ha registrado durante una semana el tiempo de proceso (expresado en ms) de un determinado servicio (TPS).

El gráfico que se presenta a continuación muestra los resultados obtenidos:



Se pide:

a) ¿Cuál es la variable implicada en el análisis? ¿De qué tipo es? **(2 puntos)**

v.a TPS={tiempo de proceso de un servicio S}

Es una v.a. cuantitativa continua

b) ¿Cómo se llama este tipo de gráfico? ¿Para qué tipos de variables es aconsejable su utilización? **(2 puntos)**

Este tipo de gráfico se denomina "Box & Whisker" o "Caja y Bigotes" y puede usarse para representar cualquier tipo de variable de naturaleza cuantitativa.

A partir de la información proporcionada por el gráfico anterior:

c) ¿Qué se puede decir sobre la distribución de la variable analizada? **(2 puntos)**

Se trata de una distribución asimétrica positiva, ya que la mitad izquierda de la caja y el bigote izquierdo son mucho más pequeños que la mitad derecha de la caja y el bigote correspondiente. También encontramos muchos datos extremos por la derecha.

d) ¿Qué parámetro nos aportaría información fiable sobre la posición de los datos analizados?. Justifica tu respuesta y calcula, si es posible, su valor aproximado. **(2 puntos)**

Como estamos estudiando una distribución asimétrica positiva, el parámetro de posición más adecuado sería la mediana, cuyo valor aproximado es 7 ms.

e) ¿Qué parámetro nos aportaría información fiable sobre la dispersión de los datos analizados?. Justifica tu respuesta y calcula, si es posible, su valor aproximado. **(2 puntos)**

Como estamos estudiando una distribución asimétrica positiva, el parámetro de dispersión más adecuado sería el Intervalo Intercuartílico, cuyo valor aproximado es 13 ms (15,5 -2,5).

2. En un grupo de 30 alumnos de Informática cada uno tiene un Tablet PC, y están distribuidos de la siguiente manera: la mitad son de la marca Sungsam, 10 son CERA y el resto PH.

Se sabe que el 30% de los Tablets CERA se han comprado en una tienda online, al igual que el 20% de los Sungsam y el 40% de los PH.

Contestar razonadamente a las siguientes preguntas:

a) ¿Cuál es la probabilidad de que si escogemos un Tablet cualquiera, no se haya comprado en una tienda online?. **(5 puntos)**

Sean los sucesos:

S = El Tablet es de la marca Sunsam

C = El Tablet es de la marca CERA

P = El Tablet es de la marca PH

O = El Tablet se ha comprado en una tienda online

$P(S) = 1/2 \quad P(C) = 1/3 \quad P(P) = 1/6$

$P(O / S) = 0{,}2 \quad P(O / C) = 0{,}3 \quad P(O / P) = 0{,}4$

$P(O) = P(O / S) * P(S) + P(O / C) * P(C) + P(O / P) * P(P) = 0{,}267$

Así, $1 - P(O) = 1 - 0{,}267 = 0{,}733$

b) Si escogemos un Tablet que se ha comprado en una tienda online, ¿cuál es la probabilidad de que sea Sungsam? **(5 puntos)**

$P(S / O) = [P(O / S) * P(S)] / P(O) = 0{,}3745$

3. Una empresa dedicada a la fabricación de piezas para aviones está interesada en el estudio de la calidad en producción de una de las piezas que fabrica. La calidad de este tipo de piezas viene determinada por un parámetro de longitud de una de sus partes.
Para realizar el estudio de calidad la empresa ha decidido realizar para dicha pieza un control estadístico de calidad, midiendo, para ello, dicha característica clave (longitud).
Se fija un límite de tolerancia mínimo y máximo para la característica a medir y se conoce que, en condiciones normales de fabricación, el 99% de las piezas fabricadas están dentro de las tolerancias admitiéndose, por tanto, como correctas.

Se pide:

a) Si se extrae al azar una muestra de 5 piezas, ¿cuál es la probabilidad de obtener 2 defectuosas? **(3 puntos)**

$$X = \{Nº\ de\ piezas\ defectuosas\ en\ una\ muestra\ de\ 5\ piezas\}$$

$$p = 0.01$$

$$n = 5$$

$$X \sim Binomial(n = 5, p = 0.01)$$

$$P(X = 2) = \binom{5}{2} 0{,}01^2 0{,}99^3 = 0{,}000970299$$

b) Si se extrae una muestra al azar 50 piezas, calcular la probabilidad de encontrar al menos 3 piezas defectuosas (utilizar la aproximación a la distribución de Poisson). **(3,5 puntos)**

$$X = \{Nº\ de\ piezas\ defectuosas\ en\ una\ muestra\ de\ 50\ piezas\}$$

$$p = 0.01$$

$$n = 50$$

$$X \sim Binomial(n = 50, p = 0.01)$$

$$N \uparrow\uparrow, p \downarrow\downarrow$$

$$X \approx Poisson(\lambda = np = 50 \cdot 0{,}01 = 0{,}5)$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0{,}9856123 = 0{,}01438768$$

c) Si tomamos al azar una muestra de 10.000 piezas, calcular aproximadamente la probabilidad de encontrar menos de 120 defectuosas. **(3,5 puntos)**

$$X = \{Nº\ de\ piezas\ defectuosas\ en\ una\ muestra\ de\ 10000\ piezas\}$$

$$p = 0.01$$

$$n = 10000$$

$$X \sim Binomial(n = 10000, p = 0.01)$$
$$\sigma^2(X) = np(1-p) = 100 \cdot 0{,}99 = 99 \gg 9$$
$X \approx V \sim N\left(m = np = 100, \sigma = \sqrt{99}\right)$ por el teorema central del límite
$$P(X < 120) = P(X \leq 119) \approx P(V < 119{,}5) = P\left(Z < \frac{119{,}5 - 100}{\sqrt{99}}\right)$$
$$= P(Z < 1{,}96) = 1 - P(Z > 1{,}96) = 1 - 0{,}0250 = 0{,}975$$

4. En una red de ordenadores corporativa, el tiempo entre accesos de los usuarios (procesos de petición de acceso o *login*) varía con una distribución exponencial de mediana 1,04 minutos.

a) Determina completamente la distribución que sigue la variable aleatoria en estudio mediante su(s) parámetro(s) fundamental(es). **(4 puntos)**

Despejando el valor de α:

$P(T > t_{mediana}) = 0.5 \rightarrow \alpha = \ln(0.5)/-1.04 = 0.667$

$T \sim Exp(\alpha = 0.667) \rightarrow E(T) = 1/\alpha = 1.5$ minutos ; $\sigma_T = 1/\alpha = 1.5$ minutos

b) Calcula la probabilidad de que se produzca una petición de acceso en los 30 segundos siguientes a la puesta en servicio de la red. **(3 puntos)**

Trabajando en minutos: $P(T < 0.5)$?

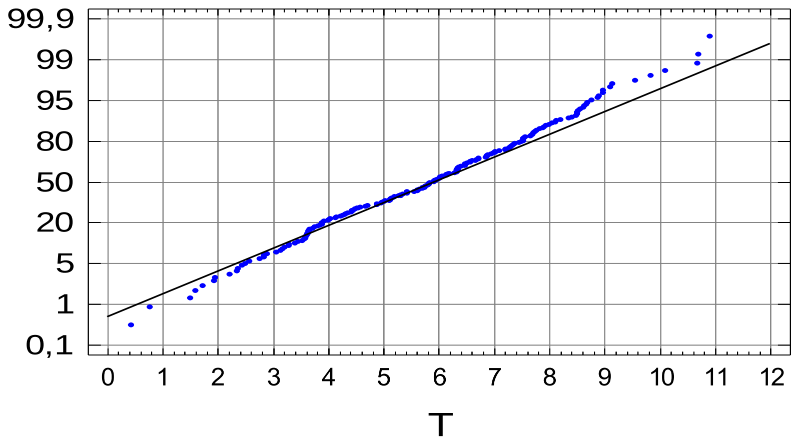
$P(T < 0.5) = 1 - e^{-\alpha * t} = 1 - e^{-0.667*0.5} = 0.2835 \equiv 28.35\%$

c) Suponiendo que la red lleva funcionando 3 minutos y no se ha producido ninguna petición de *login*, ¿cuál es la probabilidad de que se produzca una petición de *login* en los próximos 30 segundos?. **(3 puntos)**

Por la propiedad de falta de memoria se puede establecer directamente que:

$P(X < 3.5 \mid X > 3) = P(X < 0.5) =$ (calculada anteriormente) $= 0.2835$

5. Con el fin de comprobar el funcionamiento de una unidad de almacenamiento masivo se han medido los tiempos de acceso en milisegundos a los datos del mismo. El tiempo de acceso a los datos fluctúa aleatoriamente en función de la situación de éstos dentro del dispositivo. Los resultados se han representado en el siguiente gráfico:



a) ¿Cómo se denomina a este tipo de gráficos? **(1 punto)**

Papel Probabilístico Normal

b) A partir de la información proporcionada por el gráfico estima aproximadamente, si es posible, la media y la desviación típica del tiempo de acceso a un dato en la unidad. **(4 puntos)**

m=6 (abcisa del 50%)

σ=2 (la abcisa de 84,15% es m+σ, que resulta 8 sobre el gráfico. Por tanto σ=2)

c) ¿Qué porcentaje de las veces el tiempo total de lectura de un vector de 300 datos superará los 1750 milisegundos? **(5 puntos)**

T={Tiempo de acceso a 1 dato} ~ N(m=6, σ=2)

Y={Tiempo de acceso 300 datos}

$Y=X_1 + X_2 + X_3 + \ldots + X_{300}$ Y es combinación lineal de variables normales e independientes, por lo que se distribuirá normalmente. Teniendo en cuanta las propiedades de la media y la varianza, se tiene:

Y ~ N(m=300x6, σ=Raíz (300x4))$\Rightarrow$Y ~ N(m=1800, σ=34,64)

P(Y>1750) = P(N(0,1)>(1750-1800)/34,64) = P(N(0,1)>-1,44) =

1 - P(N(0,1)>1,44) = 1 - 0,0749 = 0,9251 → **92,51%**