Solución

Grado en Ingeniería Informática

Estadística

PRIMER PARCIAL

13 de mayo de 2013

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. Justificar si las siguientes afirmaciones son o no correctas:

a) La desviación típica no es un parámetro útil cuando el objetivo del análisis es estudiar si unos datos se distribuyen normalmente. **(2,5 puntos)**

CORRECTO. La desviación típica es un parámetro de dispersión, no de forma.

Para analizar la normalidad de unos datos se podrían usar otros parámetros como los coeficientes de asimetría y curtosis.

b) La variable discreta (con dos valores “0” y “1”) que refleja la situación de paro (valor “1”) o no paro (valor “0”) en el sector informático está definida sobre el conjunto de todos los trabajadores parados en dicho sector. **(2,5 puntos)**

INCORRECTO. Población = {todos los trabajadores del sector informático}

Sobre cada individuo se estudia si está en paro(1) o no (0).

c) La resistencia a la rotura de los ratones fabricados por dos empresas (A y B) de material informático constituye una variable aleatoria bidimensional. **(2,5 puntos)**

INCORRECTO. Son 2 variables aleatorias unidimensionales:

XA = {Resistencia a la rotura ratones fabricados por la empresa A}

XB = {Resistencia a la rotura ratones fabricados por la empresa B}

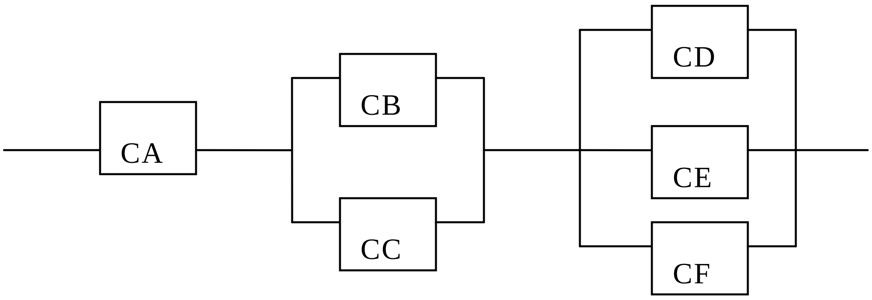
Cada variable está definida sobre una población, según la empresa.

d) El tercer cuartil calculado en una muestra es una característica de dispersión de los datos de dicha muestra. **(2,5 puntos)**

INCORRECTO. Los cuartiles son características de POSICIÓN, no de dispersión.

El Recorrido o Intervalo Intercuartílico (C3-C1) sí sería una característica de dispersión.

2. Un dispositivo está formado por seis componentes idénticos (CA, CB, CC, CD, CE, CF) montados como aparece en la siguiente figura:



Se sabe que el tiempo de funcionamiento de cada componente hasta el fallo sigue una distribución exponencial de mediana 650 horas.

a) Sabiendo que una componente lleva funcionando 300 horas, ¿cuál es la probabilidad de que siga funcionando correctamente más de 400 horas en total?. **(5 puntos)**

v.a. T= {Tiempo de funcionamiento de una componente}~ Exp(α)

Si mediana = 650 h → $P(T > 650) = 0{,}5$; $e^{-650\alpha} = 0{,}5 \rightarrow -650\,\alpha = \ln(0{,}5)$;

$\boldsymbol{\alpha} = \ln(0{,}5)/-650 = \mathbf{0{,}0011}$

La distribución EXPONENCIAL NO TIENE MEMORIA (400 = 300 + 100)

$\underline{P(X>400/X>300)} = P(X > 100) = e^{-100\times 0{,}0011} = \underline{0{,}899}$

b) Asumiendo que las 6 componentes del dispositivo funcionan independientemente, calcular la fiabilidad del dispositivo a las 800 horas y definir todas las variables y sucesos empleados. **(5 puntos)**

SUCESOS

C_i = La componente C_i funciona correctamente a las 800 h (i=A, B,…,F)

DIS = El dispositivo funciona correctamente a las 800 h

$DIS = CA \cap (CB \cup CC) \cap (CD \cup CE \cup CF)$

VARIABLES

T_i = {Tiempo de funcionamiento (h) de la componente C_i } i=A, B,…,F

T_{DIS} = {Tiempo de funcionamiento (h) del dispositivo}

$T_i \sim Exp(\alpha = 0{,}0011)$ Como en el apartado a)

FIABILIDAD DISPOSITIVO A LAS 800 h → $P(T_{DIS} > 800) = P(DIS)$

$P(DIS) = P(CA \cap (CB \cup CC) \cap (CD \cup CE \cup CF)) = P(CA) \times P(CB \cup CC) \times P(CD \cup CE \cup CF)$

- $P(CB \cup CC) = 1 - P(\overline{CB \cup CC}) = 1 - P(\overline{CB} \cap \overline{CC}) = 1 - P(\overline{CB}) \text{ x } P(\overline{CC})$
- $P(CD \cup CE \cup CF) = 1 - P(\overline{CD \cup CE \cup CF}) = 1 - P(\overline{CD} \cap \overline{CE} \cap \overline{CF}) = 1 - P(\overline{CD}) \text{ x } P(\overline{CE}) \text{ x } P(\overline{CF})$

Las componentes son iguales →

$P(CA) = P(CB) = P(CC) = P(CD) = P(CE) = P(CF) = \mathbf{P(C_i)}$

$\mathbf{P(C_i) = P(T_i > 800)} = e^{-800 \times 0{,}0011} = \mathbf{0{,}415} \rightarrow \mathbf{P(\overline{C_i}) = 1 - P(Ci) = 1 - 0{,}415 = 0{,}582}$

$$\boxed{\underline{P(DIS)} = P(CA) \text{ x } \left(1 - P\left(\overline{C_i}\right)^2\right) \text{x} \left(1 - P\left(\overline{C_i}\right)^3\right) = 0{,}415 \text{ x } (1 - \mathbf{0{,}582}^2) \text{ x } (1 - \mathbf{0{,}582}^3) \approx \underline{0{,}22}}$$

3. La empresa ELECTONYS utiliza tarjetas de red AA comercializadas por Chisco en los equipos que monta. Esta empresa quiere garantizar que el porcentaje de tarjetas defectuosas en cada lote que compra sea inferior al 10%. Para comprobar este requisito, de cada lote de tarjetas que compra toma N tarjetas seleccionadas al azar, aceptando dicho lote sólo si el número de tarjetas defectuosas es inferior a 3. ¿Cuánto debe valer como mínimo N para que la probabilidad de admitir una partida con un 10% o más de tarjetas defectuosas no supere el 5%?

v.a. X= {Número de tarjetas de red defectuosas en un lote}~ B(N, p)

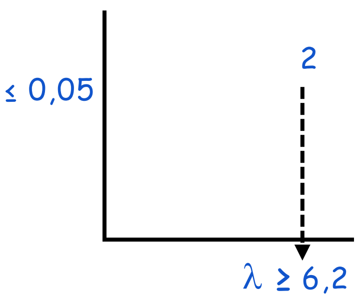
Se quiere **garantizar** $p < 0{,}1$

Aceptar partida o lote **si** $X < 3 \leftrightarrow \mathbf{X \leq 2}$

$P(\text{Aceptar} / p \geq 0{,}1) = P(X \leq 2) \leq 0{,}05$

B(N, p) → Como p es pequeño (<0,1) suponiendo y N grande → aproximación → $X \sim Ps(\lambda)$, donde $\lambda = Np = N \times 0{,}1$

Mirando en el Ábaco



$\lambda = N \times 0{,}1 \geq 6{,}2 \rightarrow N \geq 62{,}5 \rightarrow$ <u>N = 62 tarjetas</u>

4. El tiempo empleado por los visitantes que usan los ordenadores de búsqueda instalados en una biblioteca se distribuye como una variable normal de media 200 segundos y desviación típica 30 segundos.

a) ¿Qué probabilidad hay de que un visitante que emplea los ordenadores de búsqueda los utilice entre 170 y 200 segundos? **(5 puntos)**

v.a. X= {Tiempo de búsqueda de un visitante} ~ $N(m=200, \sigma=30)$

<u>**$P(170 \le X \le 200)$**</u> $=P(X \le 200) - P(X \le 170)$ = [tipificando] =

$= P(Z \le (200-200)/30) - P(Z \le (170-200)/30) = P(Z \le 0) - P(Z \le -1) =$

$P(Z < 0) - P(Z \ge 1)$ = [Tabla] = $1 - 0{,}5 - 0{,}1587 =$ <u>**0,34**</u>

b) ¿Cuál es la probabilidad de que la suma del tiempo empleado por 3 visitantes que usan los ordenadores de búsqueda supere los 590 segundos? **(5 puntos)**

v.a. Y= {Tiempo de búsqueda de 3 visitantes}

$Y = X_1 + X_2 + X_3$

X_i = {Tiempo de búsqueda de un visitante i} ~ $N(m_i = 200, \sigma_i = 30)$

- La suma de variables normales independientes es otra v.a. que sigue una distribución Normal→ Y ~ Normal
- La media de una suma de variables es la suma de sus medias
  $m_Y = m_1 + m_2 + m_3 = 3 \times 200 = 600$ s
- La varianza de una suma de variables independientes es la suma de sus varianzas
  $\sigma^2_Y = \sigma^2_1 + \sigma^2_2 + \sigma^2_3 = 3 \times (30)^2 = 2700 \text{ s}^2 \rightarrow \sigma_Y = \sqrt{2700} = 51{,}96$

$Y \sim N(m_Y = 600, \sigma_Y = 51{,}96)$

<u>**$P(Y > 590)$**</u> = [tificando] = $P(Z > (590-600)/51{,}96) = P(Z > -0{,}19) =$

$= 1 - P(Z \ge 0{,}19)$= [Tabla] = $1 - 0{,}4247 =$ <u>**0,5753**</u>

5. En un estudio sobre la utilización de videojuegos según el sexo en universitarios españoles se recogieron datos de 112 alumnos, obteniéndose los siguientes resultados:

	Tiempo dedicado a videojuegos (Horas/Semana)				
Parámetros muestrales	**N**	$\bar{X}$	**S**	**Estándar Asimetría**	**Estándar Curtosis**
MUJER	51	3,28	0,78	-0,90	-0,73
HOMBRE	61	5,16	0,84	-0,53	-0,75

A la vista de estos resultados contesta las siguientes preguntas:

a) ¿Se puede admitir que el comportamiento respecto al tiempo medio de utilización de videojuegos es el mismo en los chicos que en las chicas (α=5%)? Utiliza el TEST correspondiente. **(3 puntos)**

H_0: $m_1 = m_2$ — **tiempo medio** de utilización de videojuegos es el **mismo** en los **chicos** (m_1) que en las **chicas** (m_2)

H_1: $m_1 \neq m_2$ — **tiempo medio** de utilización de videojuegos es el **distinto** en los **chicos** (m_1) que en las **chicas** (m_2)

$$S_{\bar{X}_1-\bar{X}_2} = S\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} = 0,81\sqrt{\frac{1}{51}+\frac{1}{61}} \approx 0,15$$

$$S = \sqrt{\frac{S_1^2(N_1-1)+S_2^2(N_2-1)}{(N_1-1)+(N_2-1)}} == \sqrt{\frac{0,6084x50+0,7056x60}{110}} = 0,81$$

$$t_{(N_1-1)+(N_2-1)}^{\alpha/2} = t_{110}^{\frac{\alpha}{2}=0,025} \approx t_{120}^{\frac{\alpha}{2}=0,025} = 1,98 \quad \text{[Tabla]}$$

$$\left[\left|t_{Calculada}\right| = 12,53\right] > \left[t_{120}^{\frac{\alpha}{2}=0,025} = 1,98\right] \Rightarrow \text{Se Rechaza } H_0$$

$$t_{Calculada} = \frac{\bar{X}_1-\bar{X}_2}{S_{\bar{X}_1-\bar{X}_2}} = \frac{3,28-5,16}{0,15} \approx -12,53$$

No se puede admitir que el comportamiento respecto al tiempo medio de utilización de videojuegos es el mismo en los chicos que en las chicas (α=5%)

b) Indica si puede admitirse la hipótesis de igualdad de varianzas (α=5%) **(3 puntos)**

H_0: $\sigma^2_1 = \sigma^2_2$ — **Varianza del tiempo** de utilización de videojuegos es la **misma** en los **chicos** (σ^2_1) que en las **chicas** (σ^2_2)

H_1: $\sigma^2_1 \neq \sigma^2_2$ — **Varianza del tiempo** de utilización de videojuegos es

distinta en los **chicos** ($\sigma^2{}_1$) que en las **chicas** ($\sigma^2{}_2$)

$$IC^{(1-\alpha)\%}_{\sigma_1^2/\sigma_2^2} = \left[\frac{S_1^2}{S_2^2 f_2}, \frac{S_1^2}{S_2^2 f_1}\right] = \left[\frac{0,6048}{0,7056 \times 1,7}, \frac{0,6048}{0,7056 \times 0,58}\right] = [0,5 \,, 1,47]$$

$$f_1 / P(F_{50,60} \le f_1) = \frac{\alpha}{2} = 0,025 \to f_1 = 0,58$$

$$f_2 / P(F_{50,60} \ge f_2) = \frac{\alpha}{2} = 0,025 \to f_2 = 1,7$$

Como el "1" está dentro del intervalo $[0,5 \,, 1,47] \to$ Aceptamos H_0

Se puede admitir que las varianzas son iguales.

c) Justifica si, tras el estudio, se puede llegar a la siguiente conclusión: "Los universitarios varones dedican a los videojuegos un promedio de 5 horas a la semana (nivel de confianza 90%)" **(3 puntos)**

$$\left.\begin{array}{l} H_0: m = 5 \\ H_1: m \neq 5 \end{array}\right\}$$

$$IC^{90\%}_{m} = \bar{X}_2 \pm \frac{S_2}{\sqrt{N_2}} t^{\alpha/2}_{(N_2-1)} = 5,16 \pm \frac{0,84}{\sqrt{61}} 1,671 \approx [4,5 \,, 5,34] \qquad t^{\alpha/2}_{(N_2-1)} = t^{\frac{\alpha}{2}=0,05}_{60} = 1,671 \quad [\text{Tabla}]$$

Como $5 \subset [4,5 \,, 5,34] \to$ Aceptamos H_0

Sí se puede llegar a la conclusión de que "Los universitarios varones dedican a los videojuegos un promedio de 5 horas a la semana (nivel de confianza 90%)

d) ¿Qué hipótesis básicas deben asumirse en las técnicas de inferencia utilizadas? Justifica su cumplimiento en este caso. **(1 punto)**

Igualdad de varianzas: como se ha visto en el apartado b)

Normalidad en los datos: el tiempo de utilización de videojuegos se distribuye normalmente tanto en la población de las chicas como en la de los chicos. Se observa que los coeficientes de asimetría y curtosis estandarizados están dentro del intervalo [-2, 2], como corresponde a datos normales.

Independencia: asumimos que las dos muestras son aleatorias simples e independientes.