UNIVERSIDAD DE MÁLAGA

Departamento de Matemática Aplicada

Grados en Informática, 10–09–2012

Convocatoria de Septiembre - Curso 2011/2012

Estadística

Apellidos: ..... **Nombre:** .....

DNI: ..... **Titulación:** .....

1. El tiempo que tarda una determinada máquina en cortar un acero (variable "y"), depende del contenido de impurezas (variable "x"), medido en %. La función que liga ambas variables es: $y = 1/(ax^2 + b)$. Se pide:

   a) Ajustar dicha función a los datos:

x	0	0	1	2	3
y	10	12	5	2	1

   b) Calcular la fiabilidad de dicho ajuste.

   c) ¿Resulta un ajuste mejor que el lineal? Justificar.

   (Puntuación = 1.25 + 0.5 + 0.5 = 2.25)

2. La función de densidad de una variable aleatoria X = tiempo de espera (en minutos) para acceder a Internet, es:

$$f(x) = \begin{cases} \frac{2x}{3} & 0 \leq x < 1 \\ \frac{3-x}{3} & 1 \leq x \leq 3 \\ 0 & \text{en el resto} \end{cases}$$

   Se pide:

   a) Hallar el tiempo medio de espera en el acceso a la red.

   b) Hallar la mediana.

   c) Si un ordenador conectó con la red antes del minuto y medio, ¿Qué es mas probable: que lo hiciera durante el primer minuto, o pasado éste?

   (Puntuación = 0.5+0.5+0.5=1.5)

3. Se ha realizado, mediante muestreo aleatorio simple, un estudio sobre los salarios abonados por ciertas empresas a sus altos directivos, hombres y mujeres, obteniéndose una muestra de 30 individuos de la que se conocen los siguientes datos, expresados en miles de euros:

Muestra de 10 hombres	$\sum x = 1710$	$\sum x^2 = 293220$
Muestra de 20 mujeres	$\sum x = 3800$	$\sum x^2 = 722228$

   Supuesto que los salarios siguen una ley Normal, se pide:

   a) ¿Puede afirmarse que el salario conjunto, (de hombres y mujeres), es superior a 179000 euros?

   b) Obtener un intervalo de confianza para el cociente de varianzas de los salarios de hombres y mujeres.

   c) A la vista de los datos del estudio, el presidente de los empresarios afirma que los hombres ganan menos que las mujeres. ¿Justifican los datos disponibles esta afirmación?

   (Puntuación = 0.75 + 0.75 + 0.75 = 2.25)

4. El número de averías observadas en una empresa durante un periodo de $\mathbf{10.000}$ días, viene expresado en la siguiente tabla:

$\boldsymbol{X}$	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$
$\boldsymbol{n}$	$\mathbf{6828}$	$\mathbf{2500}$	$\mathbf{550}$	$\mathbf{100}$	$\mathbf{11}$	$\mathbf{11}$

donde $\boldsymbol{X}$ es el número de averías y $\boldsymbol{n}$ es el número de días en que se obtuvo dicho el valor. Ajustar una distribución de Poisson y analizar la bondad de dicho ajuste al 95 %.

(Puntuación = 0.5 + 1=1.5)

5. **MATLAB:**

   a) Escribir el programa con las órdenes necesarias para resolver el problema 1.

   b) Cuando un componente $\boldsymbol{k}$ se avería, es sustituido por otro igual. Si $\boldsymbol{\xi_k} \rightsquigarrow \boldsymbol{E(2)}$ (exponencial de media 1/2) ($\boldsymbol{\xi_k}$ representa la duración del componente en años). Escribir un programa que estime, mediante el método de Montecarlo con 10.000 iteraciones, la probabilidad de que la suma de 5 de ellas (independientes) ($\boldsymbol{\xi = \sum_{k=1}^{5} \xi_k}$) sea mayor de 3.

   c) Se ha medido el tiempo que tarda (en días) en descargarse la batería de un móvil resultando:

Marca A	$\mathbf{7.4}$	$\mathbf{8.2}$	$\mathbf{6.4}$	$\mathbf{7.3}$	$\mathbf{9}$	$\mathbf{7.3}$	$\mathbf{7.5}$	$\mathbf{8}$
Marca B	$\mathbf{8}$	$\mathbf{7.2}$	$\mathbf{8.7}$	$\mathbf{5.3}$	$\mathbf{8.7}$	$\mathbf{8.3}$	$\mathbf{8.5}$	$\mathbf{7.5}$

   Escribir los programas que realicen (al 95 % de confianza) los contrastes necesarios para dare respuesta a las cuestiones:

   1) ¿Debe admitirse que ambas marcas tardan el mismo tiempo en descargarse?
   2) ¿Puede rechazarse la afirmación de que la marca A dura más de 7.8 días?

   (Puntuación = 0.75 + 0.75 + 1=2.5)

ES OBLIGATORIO ENTREGAR ESTA HOJA DEBIDAMENTE CUMPLIMENTADA

- Los alumnos que quieran conservar la nota del control, no contestarán a las preguntas 1 y 5-a.

Soluciones:

Problema 1:

a) Linealizamos la ecuación: $y = \frac{1}{ax^2+b} \Leftrightarrow \frac{1}{y} = ax^2 + b$.

Hacemos el cambio $Y = \frac{1}{y},\ X = x^2,\ A = b;\ B = a, \Rightarrow Y = BX + A$

x_i	y_i	$X_i = x_i^2$	$Y_i = \frac{1}{y_i}$	X_i^2	Y_i^2	X_iY_i	y_i^{est}	r_i	r_i^2	y_i^2	x_iy_i
0	10	0	0.1000	0	0.0100	0	10.5882	−0.5882	0.3460	100	0
0	12	0	0.0833	0	0.0069	0	10.5882	1.4118	1.9931	144	0
1	5	1	0.2000	1	0.0400	0.2	5.1220	−0.1220	0.0149	25	5
2	2	4	0.5000	16	0.2500	2	2.0096	−0.0096	0.0001	4	4
3	1	9	1.0000	81	1.0000	9	0.9984	0.0016	0.0000	1	3
6	30	14	1.8833	98	1.3069	11.2		0.6931	2.3541	274	12

De donde obtenemos: $\bar{X} = \frac{14}{5} = 2.8$, $\bar{Y} = \frac{1.8833}{5} = 0.3763$, $V(X) = \frac{98}{5} - 2.8^2 = 11.76$

$V(Y) = \frac{1.3069}{5} - 0.3763^2 = 0.1195$, $Cov(X,Y) = \frac{11.2}{5} - (2.8)(0.3763) = 1.1853$

$a = B = \frac{Cov(X,Y)}{V(X)} = \frac{1.1853}{11.76} = 0.1008$

$b = A = \bar{Y} - B\bar{X} = 0.3763 - 0.1008(2.8) = 0.0944$ y el ajuste pedido resulta ser:

$$\mathrm{y} = \frac{1}{0.1008\mathrm{x}^2 + 0.0944}$$

b) La fiabilidad, o bondad del ajuste, la estimamos mediante el coeficiente R^2. Para ello, calculamos los valores estimados por el ajuste y_i^{est}, los residuos $r_i = y_i - y_i^{est}$, la varianza residual y aplicamos $R^2 = 1 - \frac{V_r}{V_y}$.

$V_r = \frac{2.3541}{5} - \left(\frac{0.6931}{5}\right)^2 = 0.4516$

$V(y) = \frac{274}{5} - \left(\frac{30}{5}\right)^2 = 18.8$ y $R^2 = 1 - \frac{0.4516}{18.8} = 0.976$

c) Para la bondad del ajuste lineal calculamos r^2

$V(x) = \frac{14}{5} - \left(\frac{6}{5}\right)^2 = 1.36$

$cov(x,y) = \frac{12}{5} - (1.2)5 = -4.8 \Rightarrow r^2 = \frac{cov(x,y)^2}{V(x)V(y)} = \frac{(-4.8)^2}{1.36*18.8} = 0.9011$

Como $R^2 > r^2$ el mejor ajuste es el propuesto.

Problema 2:

a) $\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{0} 0dx + \int_0^1 x\frac{2x}{3}dx + \int_1^3 x\frac{3-x}{3}dx + \int_3^{\infty} 0dx = 0 + \left[\frac{2x^3}{9}\right]_0^1 + \left[\frac{x^2}{2}\right]_1^3 - \left[\frac{x^3}{9}\right]_1^3 =$
$= \frac{2}{9} + \left(\frac{9}{2} - \frac{1}{2}\right) + \left(3 - \frac{1}{9}\right) = \frac{4}{3}$

b) Debemos calcular la función de distribución y resolver $F(x) = 0.5$. $F(x) = \int_{-\infty}^{x} f(x)dx \Rightarrow$

$$F(x) = \begin{cases} 0 & x \le 0 \\ \frac{x^2}{3} & 0 < x \le 1 \\ \frac{-x^2+6x-3}{6} & 1 < x \le 3 \\ 1 & x > 3 \end{cases}$$

NOTA: El valor obtenido en el intervalo [1,3] se debe a: $F(x) = \int_{-\infty}^{x} f(x)dx = \int_0^1 \frac{2x}{3}dx + \int_1^x \frac{3-x}{3}dx =$
$\left[\frac{x^2}{3}\right]_0^1 + \left[x - \frac{x^2}{6}\right]_1^x = \frac{1}{3} + (x-1) - \frac{x^2}{6} + \frac{1}{6} = (-x^2 + 6x - 3)/6$

Debemos resolver $F(x) = 0.5$ y sabemos que F(x) es una función creciente. $F(1) = 1/3 < 0.5$ luego la solución está en el intervalo [1,3]: $\frac{-x^2+6x-3}{6} = 0.5 \Rightarrow -x^2 + 6x - 6 = 0 \Rightarrow x = 3 \pm \sqrt{3} \Rightarrow$ $\mathrm{Me} = 3 - \sqrt{3} \approx 1.2679$, ya que la otra raíz está fuera del intervalo [1,3].

c) Nos están pidiendo que comparemos $P(\xi \leq 1/\xi < 1.5)$ y $P(\xi > 1/\xi < 1.5)$:

$P(\xi \leq 1/\xi < 1.5) = \frac{P(\xi \leq 1)}{P(\xi < 1.5)} = \frac{F(1)}{F(1.5)} = \frac{1/3}{5/8} = \frac{8}{15}$

$P(\xi > 1/\xi < 1.5) = \frac{P(1 < \xi \leq 1.5)}{P(\xi < 1.5)} = \frac{F(1.5) - F(1)}{F(1.5)} = \frac{5/8 - 1/3}{5/8} = \frac{7}{15}$

Luego es más probable que conecte durante el primer minuto.

Problema 3:

a) En este caso se trata de hacer un contraste unilateral de la media μ de los salarios (trabajaremos en miles de euros).

Hipótesis nula: $H_0: \ \mu \leq 179$, Hipótesis alternativa: $H_a: \ \mu > 179$

En este caso la muestra tiene 30 elementos (hombres y mujeres) luego se trata de muestra pequeña, con varianza desconocida.

$\bar{x} = \frac{\sum x}{30} = \frac{1710+3800}{30} \approx 183.6667$ y su varianza vale: $V(x) = \frac{\sum x^2}{30} - \bar{x}^2 = \frac{293220+722228}{30} - 183.6667^2 = 114.81$, la cuasivarianza o varianza muestral vale $s^2 = \frac{30}{29} V(x) = 118.7689 \Rightarrow s = 10.8981$

Cuando no se indica el nivel de significación se toma $\alpha = 0.05$

El estadístico de contraste $\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = 2.3454 > t_{\alpha, n-1} = t_{0.05,29} = 1.699$ por lo que se rechaza la hipótesis nula y **la media de los salarios es superior a 179000**.

b) Se pide un intervalo de confianza para el cociente de varianzas:

Para hombres: n=10, $\bar{x_H} = \frac{1710}{10} = 171$, $s_H^2 = \frac{10}{9}\left[\frac{293220}{10} - 171^2\right] = 90 \Rightarrow s = \sqrt{90}$, $t_{0.025,9} = 2.262$

Para mujeres: n=20, $\bar{x_M} = \frac{3800}{20} = 190$, $s_M^2 = \frac{20}{19}\left[\frac{722228}{20} - 190^2\right] = 12 \Rightarrow s = \sqrt{12}$, $t_{0.025,19} = 2.093$

Calculamos $F_{0.025;9,19}$ y $F_{0.975;9,19}$ mirando en las tablas:

$F_{0.025;9,19} = 2.880$, mientras que $F_{0.975;9,19} = \frac{1}{F_{0.025;19,9}} = \frac{1}{3.7} = 0.2703$

Para el cálculo de $F_{0.025;19,9}$ hemos interpolado entre $F(0.025, 15, 9) = 3.769$ y $F(0.025, 24, 9) = 3.614$

$F_{0.025;19,9} = 3.769 + \frac{19-15}{24-15}(3.614 - 3.769) \approx 3.700$

Luego el intervalo pedido resulta:

$$I = \left[\frac{s_1^2/s_2^2}{F_{\frac{\alpha}{2};n_1-1,n_2-1}}, \frac{s_1^2/s_2^2}{F_{1-\frac{\alpha}{2};n_1-1,n_2-1}}\right] = \left[\frac{90/12}{2.88}, \frac{90/12}{0.2703}\right] = [2.6042, 27.7469]$$

c) Se trata de un contraste unilateral de diferencias de medias de poblaciones normales, muestras pequeñas y varianzas desconocidas.

Para hacer el contraste pedido, tendremos que distinguir entre varianzas iguales o diferentes, por lo que usualmente tendríamos que hacer un contraste previo, pero que en este caso podemos evitar ya que el intervalo de confianza obtenido (al mismo nivel de significación) no contiene al 1 (lo que significa varianzas desiguales). Así pues, haremos un contraste unilateral de la diferencia de medias de poblaciones normales, tamaño pequeño y varianzas diferentes.

$\mathbf{H_0}: \mu_H \geq \mu_M$, $\mathbf{H_a}: \mu_H < \mu_M$

El estadístico de contraste es: $E = \frac{\bar{x_H} - \bar{x_M}}{\sqrt{\frac{s_H^2}{n_H} + \frac{s_M^2}{n_M}}} = \frac{171-190}{\sqrt{\frac{90}{10} + \frac{12}{20}}} = -6.1322$, que se compara con $-t_{\alpha,f}$

Hallamos f: $A = \frac{s_H^2}{n_H} = 810$, $B = \frac{s_M^2}{n_M} = 7.2$, $f = \frac{(A+B)^2}{\frac{A^2}{11} + \frac{B^2}{21}} - 2 = 9.196 \approx 9$.

Se busca el valor de $t_{0.05,9}$ en las tablas, obteniéndose $t_{0.05,9} = 1.833$

Como el valor del estadístico $E = -6.1322$ es menor que $-t_{\alpha,f} = -1.833$ se rechaza la hipótesis nula y aceptamos la alternativa, es decir, **se acepta que los hombres ganan menos que las mujeres.**

Problema 4:

a) Se ajusta una Poisson con parámetro $\boldsymbol{\lambda}$, donde $\boldsymbol{\lambda}$ es la media de los datos.

x_i	n_i	$x_i n_i$	f_i^{teor}	n_i^{teor}
0	**6828**	**0**	**0.6703**	**6703**
1	**2500**	**2500**	**0.2681**	**2681**
2	**550**	**1100**	**0.0536**	**536**
3	**100**	**300**	**0.0072**	**72**
4	**11**	**44**	**0.0007**	**7**
5	**11**	**55**	**0.0001**	**1**
	10000	**3999**		**10000**

Luego $\bar{x} = \frac{3999}{10000} = \mathbf{0.3999}$ y se ajusta la $\boldsymbol{P(0.3999)}$

b) Vamos a aproximar $\boldsymbol{P(0.3999)}$ mediante $\boldsymbol{P(0.4)}$ pues no tenemos las tablas para $\boldsymbol{\lambda = 0.3999}$

Calculamos las frecuencias teóricas (las relativas en las tablas de la Poisson y las absolutas multiplicando por n=10000), pero surge el problema de que ninguna frecuencia teórica debe ser menor de 5, y la última lo es, por lo que debemos juntar las 2 últimas clases.

x_i	n_i	f_i^{teor}	n_i^{teor}	e_i	e_i^2/n_i
0	**6828**	**0.6703**	**6703**	**125**	**2.3310**
1	**2500**	**0.2681**	**2681**	**−181**	**12.2197**
2	**550**	**0.0536**	**536**	**14**	**0.3657**
3	**100**	**0.0072**	**72**	**28**	**10.8889**
4 ó más	**22**	**0.0008**	**8**	**14**	**24.5**
	10000		**10000**		**50.3053**

Luego hemos calculado $\boldsymbol{E = \sum_i \frac{(n_i - n_i^{teor})^2}{n_i^{teor}} = 50.3053}$ y lo comparamos con el valor $\boldsymbol{\chi^2_{0.05,3} = 7.815}$ y como el estadístico $\boldsymbol{E > 7.815}$ rechazamos la hipótesis nula y **los datos no provienen de una distribución de Poisson.**

NOTAS:

- El número de grados de libertad es k-1-p=3 donde k=5 (número de clases) y p=1 (número de parámetros estimados desde la muestra (lambda)).
- Hemos comprobado que $\boldsymbol{\sum_i n_i^{teor} = 10000 = n}$ pues si (por redondeos) fuese menor que n, deberíamos sumarse la diferencia a la clase "4 ó más".

Problema 5 MATLAB

a)

```
x=[0 0 1 2 3], y=[10 12 5 2 1], N=5
disp('a) Ajuste pedido')
Y=1./y, X=x.^2, medX=sum(X)/N, medY=sum(Y)/N
varX=sum(X.^2)/N-medX^2, varY=sum(Y.^2)/N-medY^2, covXY=sum(X.*Y)/N-medX*medY
a=covXY/varX, b=medY-a*medX
disp('b) Coef. correlacion lineal')
medx=sum(x)/N, medy=sum(y)/N, varx=sum(x.^2)/N-medx^2, vary=sum(y.^2)/N-medy^2
cov=sum(x.*y)/N-medx*medy, sx=sqrt(varx), sy=sqrt(vary),r=cov/(sx*sy), r2=r^2
disp('b) Fiabilidad ajuste pedido'}
yest=1./(a*x.^2+b),res=y-yest,Vres=sum(res.^2)/N-(sum(res)/N)^2, R2=1-Vres/vary
```

b)

```
cont=0;n=10000;
for k=1:n
xi=sum(exprnd(0.5,1,5));
if xi>3,cont=cont+1;end
end
p=cont/n
q=1-p
I=[p-1.96*sqrt(p*q/n),p+1.96*sqrt(p*q/n)]
```

c)

```
alfa=0.05
A=[7.4 8.2 6.4 7.3 9 7.3 7.5 8];
B=8 7.2 8.7 5.3 8.7 8.3 8.5 7.5];
[Ha,pa]=ttest2(A,B,alfa,'both')
[Hb,pb]=ttest(A,7.8,0.05,'right')
```