

Grados en Informática
Métodos Estadísticos Examen Junio 2013

- **Tiempo: 2 horas 45 minutos.**
- Dejar DNI encima de la mesa. **Apagar y guardar el MÓVIL.**

APELLIDOS, NOMBRE:

DNI:

Grupo:

Titulación:

1. **Solo Parcial 1:** La cantidad de miel en Kg. producida por una colmena (X), depende del número (N) de las mismas situadas en un entorno de 500 m.

Para valores de $N \leq 8$, se sospecha que la relación es de la forma: $X = 10 + aN^2$

(a) Ajustar dicha función a los datos:

N	0	1	2	3	3	4
X	10.01	9.93	9.81	9.58	9.56	9.22

- (b) ¿Es bueno este ajuste? Justificar con el parámetro apropiado.

- (c) Según el ajuste del apartado (a), ¿Para qué valores de N la producción de miel no llegará al 85% de la que se obtendría sin colmenas alrededor?

(1+0.75+0.5=2.25 Puntos)

2. En una granja se dispone de 3 tipos de gallinas. La probabilidad de que una del tipo A ponga un huevo con 2 yemas es del 0.002, mientras que las del tipo B lo ponen con probabilidad 0.005 y las del tipo C nunca ponen huevos con 2 yemas. Además se sabe que todas ponen 1 huevo diario y el 70% de las gallinas son del tipo A y el 20% del B. Hallar:

- (a) Proporción de huevos con 2 yemas en la granja.

- (b) Si un huevo tiene 2 yemas. ¿Qué probabilidad hay de que lo haya puesto una del tipo A?

- (c) Hallar un intervalo de confianza al nivel del 99.2% para el número de huevos de 2 yemas puestos por gallinas tipo A, si la granja dispone de 5400 gallinas de este tipo.

(0.5+0.5+0.5=1.5 Puntos)

3. La variable aleatoria ξ mide el tiempo hasta que se produce un fallo en el sistema cuando éste se deteriora con el uso y tiene por función de distribución: $F(x) = 1 - e^{-(2x)^{1.1}}$

- (a) Hallar $P(\xi > 1.5)$.

- (b) Hallar la probabilidad de que de 100 variables aleatorias independientes siguiendo esa distribución, 2 ó más sean mayores que 1.5.

- (c) Hallar la mediana de la distribución ξ .

(0.25+0.5+0.25=1 Punto)

4. Un artículo consta de 4 componentes iguales. Tras 100 horas de uso el número de componentes que debe ser sustituido η se supone (si son independientes) que debe seguir una distribución binomial con $N = 4$, es decir: $B(4, p)$.

- (a) Ajustar una binomial $B(4, p)$ a los datos históricos:

x_i	0	1	2	3	4
o_i	320	431	183	61	5

- (b) Analizar la bondad del ajuste realizado.

NOTA: Se han analizado 1000 veces 4 componentes.

(0.5+0.75= 1.25 Puntos)

5. En un experimento agrícola se están estudiando nuevas técnicas de cultivo de una determinada variedad de tomate “Cherry”. Mediante el cultivo “clásico” se obtenía una producción media de 9 Tm por parcela, con una varianza de 30 Tm.

Para dicho estudio se utiliza la técnica A, desarrollada en 10 parcelas y la B en 11 parcelas. Los resultados fueron:

A	12.6	9.2	6.4	9.8	15.3	14.0	10.1	13.6	9.8	13.4	
B	15.1	10.3	12.2	14.1	10.5	6.1	6.2	10.1	12.2	9.3	10.6

Se pide (al 95%):

- ¿Podemos aceptar que la varianza de la técnica B es la misma que la del cultivo clásico?
 - Contrastar si la producción media mediante la técnica B ha aumentado con respecto a la clásica.
- ¿Podemos aceptar que las dispersiones mediante las técnicas A y B son iguales?
 - Contrastar que la producción media de la técnica A es mayor que la de B.

(0.75+0.75=1.5 Puntos)

===== Entregar en folio aparte =====

Indicar, tan solo, las órdenes necesarias (MATLAB o lenguaje equivalente) para resolverlos, pero sin usar calculadora ni tablas.

6. MATLAB Solo Parcial 1: Dada la tabla bidimensional:

- Ajustar la recta $Y/X: y = mx + n$, la $X/Y: x = py + q$, y una parábola $Y/X: y = a + bx + cx^2$ a los datos de la tabla:

$X \backslash Y$	0	1	2	3	4
$(-\infty, 0]$	0	0	35	35	0
$(0, 10]$	0	35	25	15	30
$(10, 20]$	11	20	5	0	0
$(20, 40]$	22	12	14	0	0

- Hallar el coeficiente de determinación cada uno de los ajustes realizados.

(0.5+0.25=0.75 Puntos)

7. MATLAB: Las variables ξ_i son todas variables aleatorias independientes que siguen una exponencial de media 50.

Estimar, mediante el método de Montecarlo, con 10000 iteraciones, las probabilidades (dar el intervalo de confianza al 95% de cada una):

- $P(2\xi_1 > 100)$
- $P(\xi_1 + \xi_2 > 100)$
- $P(\xi_1 + \xi_2 + \xi_3 > 150)$

(0.25+0.25+0.25=0.75 Puntos)

8. MATLAB: Un estudiante está interesado en averiguar el camino más rápido para ir de la Facultad a su casa. Quiere distinguir entre 2 recorridos posibles, para ello durante 4 semanas realiza 20 recorridos, 10 por cada uno. Además tiene el cuidado de hacer el mismo número (2) por cada uno de los días: Lunes, Martes, etc. Obtiene los siguientes resultados (en minutos).

Recorrido 1:	24	20	18	16	30	24	36	18	16	26
Recorrido 2:	26	20	16	16	36	22	38	22	18	30

- Contrastar al nivel $\alpha = 3\%$ que el primer recorrido es mejor, pues disminuye el tiempo medio del trayecto.
- Contrastar que al mismo nivel que la media del primer recorrido es 21 minutos.

(0.5+0.5=1 Punto)

Soluciones:

Problema 1:

a: El ajuste $X = 10 + aN^2$ se realiza con: $X - 10 = aN^2$ luego $X' = X - 10 = aN^2$ y queremos expresar X' en la base $B = \{N^2\}$, por lo que la única ecuación es:

$$\langle X', N^2 \rangle = a \langle N^2, N^2 \rangle \Rightarrow \sum_i X'_i N_i^2 = a \sum_i N_i^4 \Rightarrow -21.05 = 435a \Rightarrow a \approx -0.0484$$

y el ajuste pedido es: $\mathbf{X}^* = \mathbf{10} - \mathbf{0.0484N}^2$

Los cálculos se realizan en la tabla siguiente, donde: $X' = X - 10$, $X^* = 10 - 0.0484N^2$

N	X	X'	$X_i N_i^2$	N_i^4	X^*	$X - X^*$	$(X - X^*)^2$	X^2	N^2	$X_i N_i$
0	10.01	0.01	0.00	0	10.0000	0.0100	0.000100	100.2001	0	0
1	9.93	-0.07	-0.07	1	9.9516	-0.0216	0.000466	98.6049	1	9.93
2	9.81	-0.19	-0.76	16	9.8064	0.0036	0.000013	96.2361	4	19.62
3	9.58	-0.42	-3.78	81	9.5644	0.0156	0.000243	91.7764	9	28.74
3	9.56	-0.44	-3.96	81	9.5644	-0.0044	0.000019	91.3936	9	28.68
4	9.22	-0.78	-12.48	256	9.2256	0.0056	0.000031	85.0084	16	36.88
13	58.11		-21.05	435		-0.0024	0.000872	563.2195	39	123.85

b: $\bar{X} = \frac{58.11}{6} \approx 9.685$, $\sigma_x^2 = \frac{563.22}{6} - 9.685^2 \approx 0.0707$

$\sigma_r^2 = \frac{0.000872}{6} - \left(\frac{-0.0024}{6}\right)^2 \approx 0.000145 \Rightarrow \mathbf{R}^2 = \mathbf{1} - \frac{0.000145}{0.0707} = \mathbf{0.9979}$

El valor de R^2 justifica que el ajuste es muy bueno pues justifica el 99.79% de las variaciones observadas.

c: Para $N=0 \Rightarrow X^* = 10$ el 85% de ese valor es 8.5, por lo que resolvemos $10 - 0.0484N^2 < 8.5, \Rightarrow 0.0484N^2 > 1.5 \Rightarrow \mathbf{N} > \mathbf{5.56}$

Luego con 6 alrededor ya no se alcanza el 85% del que tendría estando sola.

Problema 2:

a: $P(2/A) = 0.002$, $P(2/B) = 0.005$, $P(2/C) = 0$, $P(A) = 0.7$, $P(B) = 0.2$ y $P(C) = 1 - 0.7 - 0.2 = 0.1$

Por el Teorema de la Probabilidad Total:

$\mathbf{P(2) = P(A) * P(2/A) + P(B) * P(2/B) + P(C) * P(2/C) = 0.7(0.002) + 0.2(0.005) + 0.1(0) = 0.0024}$

b: Se pide $P(A/2)$. Pero $P(A/2) = \frac{P(A \cap 2)}{P(2)} = \frac{P(A) * P(2/A)}{P(2)} = \frac{0.7(0.002)}{0.0024} = \mathbf{0.5833}$

c: Según las tablas para la proporción, un intervalo de confianza para p es:

$$I_p = \left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \right] = \left[0.002 \pm 2.65 \sqrt{\frac{0.002(0.998)}{5400}} \right]$$

Si la proporción $p = \frac{n(2)}{n}$ está en ese intervalo con probabilidad del 99.2%, el número de huevos de 2 yemas $n(2)$ estará en $I_{n(2)} = [(5400)0.002 \pm 2.65(5400)\sqrt{\frac{0.002(0.998)}{5400}}] = [\mathbf{2.1}, \mathbf{19.5}]$

En este caso hemos obtenido como $z_{\frac{\alpha}{2}} = z_{0.004}$ buscando en las tablas de la normal 0.004 (en el centro) obtenemos el valor 2.65 (en el borde).

En general el intervalo para n puede deducirse del de p ($p = n/N$), y resulta: $I_n = [n\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{n\hat{p}\hat{q}}]$

Problema 3:

a: $P(\xi > 1.5) = 1 - F(1.5) = 1 - \left(1 - e^{-(2(1.5))^{1.1}}\right) = e^{-3^{1.1}} = e^{-3.3484} = \mathbf{0.0351}$

b: La variable aleatoria $\eta = \{\text{Numero de mayores de 1.5}\}$ seguirá una binomial de parámetros $N=100$ y $p=0.0351$.

Como $N > 30$ y $Np = 3.51 < 5$ puede aproximarse por una distribución de Poisson de parámetro $\lambda = 3.51$

Nos piden $P(\eta \geq 2) = 1 - P(\eta = 0) - P(\eta = 1) = 1 - \left[e^{-3.51} \frac{(3.51)^0}{0!}\right] - \left[e^{-3.51} \frac{(3.51)^1}{1!}\right] = 1 - e^{-3.51}(1 + 3.51) \approx \mathbf{0.8652}$

c: la mediana debe verificar $F(x) = 0.5 \Rightarrow 1 - e^{-(2x)^{1.1}} = 0.5 \Rightarrow 0.5 = \exp(-(2x)^{1.1}) \Rightarrow -Ln(0.5) = (2x)^{1.1} \Rightarrow [-Ln(0.5)]^{1/1.1} = (2x) \Rightarrow x = \frac{[-Ln(0.5)]^{1/1.1}}{2} = \mathbf{0.358316}$

Problema 4:

a: Lo primero será ajustar una binomial. En total se han analizado $1000*4=4000$ componentes de los que se han sustituido: $Num = 320*0 + 431*1 + 183*2 + 61*3 + 5*4 = 1000$ y por tanto $p = \frac{1000}{4000} = 0.25$

Luego la binomial ajustada será: $\mathbf{B(4, 0.25)}$

b: Los cálculos debemos realizarlos en forma de tabla:

x_i	0	1	2	3	4	\sum_i
O_i	320	431	183	61	5	1000
p_i	0.3164	0.4219	0.2109	0.0469	0.0039	
$1000p_i$	316.4	421.9	210.9	46.9	3.9	

Como una de las frecuencias esperadas $e_i = 1000p_i < 5$ debemos juntar clases.

x_i	0	1	2	3 y 4	\sum_i
O_i	320	431	183	66	1000
p_i	0.3164	0.4219	0.2109	0.0508	
$1000p_i$	316.4	421.9	210.9	50.8	
$\frac{O_i^2}{e_i}$	323.64	440.30	158.79	85.75	1008.48

Calculamos el $\chi_{exp}^2 = \sum_i \frac{O_i^2}{e_i} - n \approx 1008.48 - 1000 = 8.48$

Que se compara con $\chi_{teo}^2 = \chi_{0.05, 4-1-1}^2 = 5.9915$ y como $\chi_{exp}^2 = 8.48 > 5.9915 = \chi_{teo}^2$, deducimos que **no es un buen ajuste**.

Problema 5:

Debemos calcular a partir de los datos:

$$n_A = 10, \bar{x}_A = 11.42, s_A^2 = \frac{10}{9} 6.8896 = 7.655$$

$$n_B = 11, \bar{x}_B = 10.609, s_B^2 = \frac{11}{10} 7.208 = 7.929, s_B = 2.815$$

$$\mathbf{a-i: } H_0 : \sigma_B^2 = 30, H_a : \sigma_B^2 \neq 30$$

$$E = \frac{(n-1)s_B^2}{\sigma^2} = \frac{(11-1)7.929}{30} = 2.643 \notin [3.247, 20.483] \Rightarrow \mathbf{Rechazo } H_0$$

$$\mathbf{a-ii: } H_0 : \mu_B \leq 9, H_a : \mu > 9$$

Se trata de un contraste de la media (unilateral) para varianza desconocida y muestra pequeña.

$$E = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} > t_{\alpha, n-1}, \text{ que en nuestro caso vale:}$$

$E = \frac{10.609 - 9}{\frac{2.815}{\sqrt{11}}} = 1.895$, mientras que $t_{0.05, 10} = 1.812$ luego estamos en la región crítica y rechazamos H_0 , **la media ha aumentado**.

$$\mathbf{b-i: } H_0 : \sigma_A^2 = \sigma_B^2, H_a : \sigma_A^2 \neq \sigma_B^2$$

$$E = \frac{s_A^2}{s_B^2} = \frac{7.655}{7.929} \in [F_{0.975, 9, 10}, F_{0.025, 9, 10}] = [\frac{1}{3.964}, 3.779] \Rightarrow \mathbf{Las varianzas son iguales.}$$

$$\text{Recordar que } F_{0.975, 9, 10} = \frac{1}{F_{0.025, 10, 9}}$$

$$\mathbf{b-ii: } H_0 : \mu_A \leq \mu_B, H_a : \mu_A > \mu_B$$

$$\text{Veamos si } E = \frac{\bar{x}_A - \bar{x}_B}{s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} > t_{0.05, 11+10-2}$$

$$s_p = \sqrt{\frac{(10-1)7.655 + (11-1)7.929}{11+10-2}} = \sqrt{\frac{148.185}{19}} = 2.7924$$

$$E = \frac{11.42 - 10.609}{2.7924 \sqrt{\frac{1}{11} + \frac{1}{10}}} = \frac{0.811}{1.2204} = 0.664, \text{ mientras que } t_{0.05, 19} = 1.729$$

Como $E < t_{0.05, 19}$ acepto la hipótesis nula (H_0) y concluyo con que **no hay evidencia de que la producción media de la variedad A sea mayor que la de B**.

Problema 6:

a:

```
x=[-5 -5 5 5 5 5 15 15 15 30 30 30];
y=[ 2 3 1 2 3 4 0 1 2 0 1 2];
n=[35 35 35 25 15 30 11 20 5 22 12 14];
disp('Recta Y/X')
N=sum(n);
A=[sum(n.*x) N; sum(n.*x.^2) sum(n.*x)]
B=[sum(n.*y); sum(n.*y.*x)]
sol1=A\B, m=sol(1), n=sol(2)
% La solución es y=mx+n
```

```
disp('Recta X/Y')
A2=[sum(n.*y) N; sum(n.*y.^2) sum(n.*y)]
B2=[sum(n.*x); sum(n.*x.*y)]
sol2=A2\B2, p=sol2(1), q=sol2(2)
% La solución es x=py+q
```

```
disp('Parabola Y/X')
```

```
A3=[sum(n.*x.^2) sum(n.*x) N;
     sum(n.*x.^3) sum(n.*x.^2) sum(n.*x);
     sum(n.*x.^4) sum(n.*x.^3) sum(n.*x.^2)]
B3=[sum(n.*y); sum(n.*y.*x); sum(n.*y.*x.^2)]
```

```

sol3=A3\B3, c=sol3(1), b=sol3(2), a=sol3(3)

disp('Apartado b')

disp('Recta Y/X')
yest=m*x+n
res=y-yest
Vra=sum(n.*res.^2)/N-(sum(n.*res)/N)^2 % Podemos evitar la resta pues sum(n.*res)=0
Vy=sum(n.*y.^2)/N-(sum(n.*y)/N)^2
R2a=1-Vra/Vy

disp('Recta X/Y')
xest=p*y+q
resb=x-xest
Vrb=sum(n.*resb.^2)/N-(sum(n.*resb)/N)^2 % Podemos evitar la resta pues sum(n.*resb)=0
Vx=sum(n.*x.^2)/N-(sum(n.*x)/N)^2
R2a=1-Vrb/Vx

disp('Parabola Y/X')
yestc=a+b*x+c*x.^2
resc=y-yestc
Vrc=sum(n.*resc.^2)/N-(sum(n.*resc)/N)^2
R2c=1-Vrc/Vy

```

Problema 7:

```

disp('7-a')
NIT=10000
x=exprnd(50,NIT,1);
c=(2*x>100);
pa=sum(c)/NIT
qa=1-pa
Ip=[pa-1.96*sqrt(pa*qa/NIT),pa+1.96*sqrt(pa*qa/NIT)]

disp('b')
xb=exprnd(50,NIT,1)+exprnd(50,NIT,1);
cb=(xb>100);
pb=sum(cb)/NIT
qb=1-pb
Ipb=[pb-1.96*sqrt(pb*qb/NIT),pb+1.96*sqrt(pb*qb/NIT)]

disp('c')
xc=exprnd(50,NIT,1)+exprnd(50,NIT,1)+exprnd(50,NIT,1);
cc=(xc>150);
pc=sum(cc)/NIT
qc=1-pc
Ipc=[pc-1.96*sqrt(pc*qc/NIT),pc+1.96*sqrt(pc*qc/NIT)]

```

Problema 8:

```

R1=[24 20 18 16 30 24 36 18 16 26]
R2=[26 20 16 16 36 22 38 22 18 30]
[Ha,Pa]=ttest2(R1,R2,0.03,'left')
[Hb,Pb]=ttest(R1,21,0.03,'both')

```