Grados en Informática
Métodos Estadísticos   Examen Junio 2016

- **Tiempo: 2 horas 30 minutos**.
- Dejar DNI encima de la mesa. **Apagar y guardar el MÓVIL**.
- El alumno puede optar por conservar la nota del Control o realizar los problemas 1 y 5 marcados con (E.D.).

APELLIDOS, NOMBRE:
DNI: Grupo: Titulación:

---

1. **(E.D.)** Se desea comprobar con precisión el voltaje V y la resistencia interna r de una batería. Se conecta a un circuito eléctrico, para el que la ley de Ohm da la fórmula: $V = (R + r)I$, y se determinan varias medidas de la resistencia ohmica R y de la intensidad de corriente I: (Trabajar con 4 decimales)

   a) Estimar los valores de V y r de la batería, ajustando dicha función (ley de Ohm) a los datos:

R	3.8	4.2	4.0	4.5	3.5
I	1.39	1.26	1.33	1.18	1.50

   b) Determinar la varianza residual de R en función de I.

   c) Hallar el sesgo de los datos R

   (1+0.5+0.75=2.25 Puntos)

2. El diámetro (en cm.) de una variedad de fruta sigue una variable aleatoria $N(7.8, 0.4)$, si ha sido cultivada en terreno volcánico, mientras que sigue una $N(7, 0.6)$ si lo ha sido en otro tipo de terreno. Estas son clasificadas como de calidad "extra" si tienen un diámetro superior a 8.5 cm., hallar:

   a) ¿Qué proporción de cada tipo será clasificada como "extra"?

   b) Si el 30 % de las frutas que llegan a un almacén lo hacen de terrenos volcánicos. ¿En qué proporción estarán las que han sido cultivadas en terreno volcánico, entre las de que han sido seleccionadas de calidad "extra".

   c) En la recepción del almacén hacemos una inspección consistente en tomar 10 frutas (al azar y todas cultivadas en el mismo campo) y calculamos la suma de sus diámetros. Si esta da superior a 74 cm. consideramos que el lote proviene de terreno volcánico, considerándolo como "no volcánico", en caso contrario (suma diámetros menor o igual a 74 cm.). ¿Cuál es la probabilidad de equivocarnos cuando el lote proviene de terreno volcánico?

   d) En las hipótesis del apartado anterior, ¿cuál es la probabilidad de equivocarnos si el lote proviene de terreno no volcánico?

   (0.5+1+0.4+0.35=2.25 Puntos)

3. Un estudio desea determinar la uniformidad del conocimiento de inglés según el Centro de estudio. Para ello, se pregunta por el significado de 4 palabras: Blizzard, throne, ligthning, hood, obteniéndose los resultados:

Centro	*A*	*B*	*C*	*D*
a_i	30	40	40	35
n_i	120	200	160	120

   donde a_i es el número de respuestas correctas y n_i es el de respuestas realizadas.

   Contrastar al 2 % si existen diferencias de conocimiento entre los Centros educativos *(1.25 Puntos)*

4. Se desea contrastar si un fertilizante empleado en el abonado del trigo, resulta mejorado mediante la adición de hierro (Fe). Para su estudio se disponen de 25 parcelas de igual tamaño, de las que se seleccionan 12 al azar a las que se las abona con el fertilizante modificado, mientras que las 13 restantes se les abona con el fertilizante sin modificar, obteniéndose las producciones siguientes en Tm. por parcela.

Sin Fe	6.0	5.7	5.4	5.4	6.0	5.7	6.4	6.5	6.3	6.0	5.9	6.3	5.8
Con Fe	5.9	6.6	6.1	6.5	7.0	6.6	6.1	6.1	6.0	6.3	6.2	5.9	–

   a) Contrasta al 5 % si la adición de hierro aumenta la producción.

   b) Contrasta al 5 % si la adición de hierro aumenta la producción en 0.5 Tm/parcela.

   (1.25+0.5=1.75 Puntos)

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

5. **(E.D.) (MATLAB)** Dada la tabla bidimensional:

   a) Ajustar una parábola de la forma $Y = a + bX + cX^2$ a los datos de la tabla:

x_i	0	1	1	-1	-1	2	-2
y_i	0	0	1	-1	0	2	2
n_i	2	5	4	2	6	4	5

   b) Hallar el coeficiente de determinación del ajuste realizado.

(0.5+0.25=0.75 Puntos)

6. **(MATLAB)** Se ha estimado que el número de colonias de hormigas por Dm^2 sigue un Poisson de media 1.84

   a) Hallar la probabilidad de que haya 5 ó más en 1 Dm^2.

   b) Hallar la probabilidad de que de 100 parcelas de Dm^2 elegidas al azar, al menos 8 tengan 5 ó más colonias.

   c) Estimar, mediante el método de Montecarlo con 10000 iteraciones, la probabilidad de que elegidas dos parcelas al azar tengan el mismo número de colonias.

(0.25+0.25+0.25=0.75 Puntos)

7. **(MATLAB)** Con el enunciado del problema 4 pero mediante MATLAB:

   a) Contrastar al 5 % si la media del fertilizante sin añadir Fe es 6.

   b) Contrastar al 5 % si la media del fertilizante con adición de Fe es mayor que 6.

   c) Contrastar al 5 % si la adición de hierro aumenta la producción.

   d) Contrastar al 5 % si la adición de hierro aumenta la producción en 0.5 Tm/parcela.

(0.25+0.25+0.25+0.25=1 Punto)

SOLUCIONES:

Problema 1:

La expresión $V = (R + r)I$ la podemos poner como: $\frac{V}{I} = R + r \Rightarrow R = \frac{V}{I} - r$ y llamando $Y = R$, $V = b$, $X = \frac{1}{I}$, $a = -r$ tenemos $Y = a + bX$

$Y_i = R_i$	I_i	$X_i = 1/I_i$	$X_i * Y_i$	X_i^2	R_i^{est}	res_i	res_i^2	R_i^2	R_i^3
3.8	1.3900	0.7194	2.7338	0.5176	3.7997	0.0003	0.0000	14.44	54.872
4.2	1.2600	0.7937	3.3333	0.6299	4.2083	-0.0083	0.0001	17.64	74.088
4.0	1.3300	0.7519	3.0075	0.5653	3.9783	0.0217	0.0005	16.00	64.000
4.5	1.1800	0.8475	3.8136	0.7182	4.5045	-0.0045	0.0000	20.25	91.125
3.5	1.5000	0.6667	2.3333	0.4444	3.5092	-0.0092	0.0001	12.25	42.875
20.0		3.7791	15.2216	2.8754		0.0000	0.0007	80.58	326.960

La ecuaciones normales de la recta quedan:

$$\left.\begin{array}{r} 20 = 5a + 3.7791b \\ 15.2216 = 3.7791a + 2.8754b \end{array}\right\} \Rightarrow \left.\begin{array}{r} a = -0.1608 \\ b = 5.5051 \end{array}\right\} \Rightarrow Y = -0.1608 + 5.5051X$$

y por tanto: $V \approx 5.0551$ y $r \approx 0.1608$

1-b: El ajuste será $R = \frac{V}{I} - r$ que produce $R_i^{est} = \frac{5.0551}{I_i} - 0.1608$ y $res_i = R_i - R_i^{est}$

Calculamos la varianza residual: $\mathbf{V_r} = \frac{\mathbf{0.0007}}{\mathbf{5}} - \mathbf{0^2} \approx \mathbf{0.00014}$ y aunque no se pide $R^2 = 1 - \frac{0.00014}{0.116} = 0.9988$ pues la varianza de R es:$V_R = \frac{80.58}{5} - (20/5)^2 \approx 0.116$

1-c:

Calculamos los momentos ordinarios $m_1 = \frac{20}{5} = 4$, $m_2 = \frac{80.58}{5} = 16.116$ y $m_3 = \frac{326.96}{5} = 65.3920$ y calculamos $\mu_3 = m_3 - 3m_2m_1 + 2m_1^3 = 65.392 - 3(16.116)4 - 2 * 4^3 = 0$ y el coeficiente de sesgo: $\gamma_1 = \frac{\mu_3}{\sigma^3} = 0$ indicando que los datos de R están centrados en la media.

Problema 2:

Llamamos d_1 a la v.a cuando proviene de terreno "volcánico (V)": $d_1 \rightarrow N(7.8, 0.4)$ y d_2 cuando proviene de terreno "no volcanico (NV)", entonces $d_2 \rightarrow N(7, 0.6)$. Nos piden: (Llamamos "Extra (E)")

$\mathbf{P(E/V) = P(d1 > 8.5)} = P\left(z > \frac{8.5-7.8}{0.4}\right) = P(z > 1.75)=$ **0.0401**

$\mathbf{P(E/NV) = P(d2 > 8.5)} = P\left(z > \frac{8.5-7}{0.6}\right) = P(z > 2.5)=$ **0.0062**

2-b: $\mathbf{P(V/E)} = \frac{P(V \wedge E)}{P(E)} = \frac{P(V)P(E/V)}{P(V)P(E/V)+P(NV)P(E/NV)} = \frac{0.3(0.0401)}{0.3(0.0401)+0.7(0.0062)} \approx$ **0.7349**

2-c: Si provienen de terreno volcánico, $SD1 = \sum_i d_i^1 \rightarrow N(78, 0.4\sqrt{10}) = N(78, 1.2649)$ y nos piden la probabilidad de equivocarnos, lo que ocurre si la suma es menor o igual a 74:

$\mathbf{P}(\text{equivocarnos}/\mathbf{V}) = P(SD1 \leq 74) = P\left(z \leq \frac{74-78}{1.2649}\right) = P(z \leq -3.1623) \approx$ **0.00078**

2-d: Si provienen de terreno no volcánico, $SD2 = \sum_i d_i^2 \rightarrow N(70, 0.6\sqrt{10}) = N(70, 1.8974)$ y nos piden la probabilidad de equivocarnos, lo que ocurre si la suma es mayor de 74:

$\mathbf{P}(\text{equivocarnos}/\mathbf{NV}) = P(SD2 > 74) = P\left(z \leq \frac{74-70}{1.8974}\right) = P(z > 2.1082) \approx$ **0.0175**

Problema 3:

Se trata de un contraste no paramétrico de homogeneidad entre muestras donde a_i son las frecuencias observadas O_i.

Fijamos las Hipótesis: H_0 : Las muestras son homogéneas. H_a : las muestras no son homogéneas.

Calculamos la proporción $p = \sum O_i \sum n_i = \frac{145}{600} = 0.2417$ y calculamos las frecuencias esperadas $e_i = n_i p$ con la hipótesis nula.

Centro	A	B	C	D
O_i	30	40	40	35
n_i	120	200	160	120
$e_i = n_i p$	29.0000	48.3333	38.6667	29.0000
$(O_i - e_i)$	1.0000	-8.3333	1.3333	6.0000
$(O_i - e_i)^2$	1.0000	69.4444	1.7778	36.0000
$\frac{(O_i-e_i)^2}{n_i}$	0.0083	0.3472	0.0111	0.3000

resultando que $\sum_i \frac{(O_i-e_i)^2}{n_i} = 0.0083 + 0.3472 + 0.0111 + 0.3 = 0.6667$ y el estadístico experimental:

$E = \frac{1}{p(1-p)} \sum_i \frac{(O_i-e_i)^2}{n_i} = \frac{10.6667}{0.2417(0.7583)} = 3.6379$, mientras que el teórico es $\chi^2_{0.02,3} = 9.837$ y como:

$E < 9.837 = \chi^2_{0.02,3}$ admitimos la hipótesis nula de homogeneidad.

En conclusión, **la muestra es homogénea.**

Problema 4:

Nos piden que hagamos un contraste de diferencias de medias:

$H_0 : \mu_{Fe} \leq \mu_S$, $\mu_{Fe} > \mu_S$ con muestras pequeñas y varianzas desconocidas pero ¿serán iguales o distintas? pues los cálculos son diferentes según sea el caso.

Para ver cuál usamos debemos hacer previamente un contraste de igualdad de varianzas:

H_0 : Las varianzas son iguales $\sigma_S^2 = \sigma_{Fe}^2$, contra que H_a : Las varianzas son diferentes: $\sigma_S^2 \neq \sigma_{Fe}^2$

A partir de los datos dados obtenemos: $n_S = 13$, $n_{Fe} = 12$, $\bar{x}_S = \frac{\sum x_i^S}{13} = 5.9538$, $\bar{x}_{Fe} = \frac{\sum x_i^{Fe}}{12} = 6.2750$, $s_S = 0.3550$, $s_{Fe} = 0.3361$

Así: $E = \frac{s_S^2}{s_{Fe}^2} = \frac{0.3550^2}{0.3361^2} = \frac{0.1260}{0.1130} \approx 1.1157$

El intervalo es $I = [F_{1-\frac{\alpha}{2},n_S-1,n_{Fe}-1}, F_{\frac{\alpha}{2},n_S-1,n_{Fe}-1}] = [F_{0.975,12,11}, F_{0.025,12,11}] = [0.3007, 3.43]$

ya que $F_{0.975,12,11} = \frac{1}{F_{0.025,11,12}} \approx \frac{1}{3.3255} = 0.3007$ donde $F_{0.025,11,12} \approx \frac{F_{0.025,10,12}+F_{0.025,12,12}}{2} = \frac{3.374+3.277}{2} \approx 3.3255$ (pues interpolamos entre los 2 valores más cercanos de las tablas) .

Como resulta que $E \in [0.3007, 0.34]$ nos quedamos con la **hipótesis nula de igualdad de varianzas**.

NOTA: Alternativamente, podría haberse comprobado que: $E = \frac{s_{Fe}^2}{s_S^2} \in [F_{0.975,11,12}, F_{0.025,11,12}]$, llegándose a la misma conclusión.

Ahora continuamos con el de la diferencia de medias:

Calculamos s_d mediante la fórmula $s_d^2 = \frac{(n_S-1)s_S^2+(n_{Fe}-1)s_{Fe}^2}{n_S+n_{Fe}-2} = \frac{12(0.3550)^2+11(0.3561)^2}{23} = 0.1198$ por lo que $sd = \sqrt{0.1198} \approx 0.3461$

$E = \frac{\bar{x}_{Fe}-\bar{x_S}}{sd\sqrt{\frac{1}{n_S}+\frac{1}{n_{Fe}}}} = \frac{6.2750-5.9538}{0.3461\sqrt{\frac{1}{13}+\frac{1}{12}}} \approx 2.3181$ que se compara con $t_{0.05,23} = 1.714$ y como es mayor se rechaza la hipótesis nula por lo que $\mu_{Fe} > \mu_S$.

4-b:

Se estudia si la diferencia de medias entre la variable aleatoria Fe y $S + 0.5$ son iguales o diferentes. El contraste a realizar es:

$H_0 : \mu_{Fe} = \mu_S + 0.5$, $H_a : \mu_{Fe} \neq \mu_S + 0.5$ con muestras pequeñas y varianzas iguales (sigue sirviendo el contraste de igualdad de varianzas anterior).

El estadístico experimental es: $E = \frac{|\bar{x}_{Fe}-(\bar{x_S}+0.5)|}{sd\sqrt{\frac{1}{n_S}+\frac{1}{n_{Fe}}}} = \frac{|6.2750-(5.9538+0.5)|}{0.3461\sqrt{\frac{1}{13}+\frac{1}{12}}} \approx 1.2909$ que se compara con $t_{0.025,23} = 2.069$ y como es menor, se acepta la hipótesis nula, por lo que, **es aceptable que la media ha aumentado en 0.5 Tm. con la adición de hierro**.

Problema 5:

Se trata de expresar y en la base: $B = \{1, x, x^2\}$

```
clc,clear,disp('Problema 5')
x=[0 1 1 -1 -1 2 -2]
y=[0 0 1 -1  0 2  2]
n=[2 5 4  2  6 4  5]
N=sum(n);
A=[    N          sum(n.*x)   sum(n.*x.^2);
   sum(n.*x)    sum(n.*x.^2) sum(n.*x.^3);
   sum(n.*x.^2) sum(n.*x.^3) sum(n.*x.^4)]
B=[sum(n.*y); sum(n.*y.*x); sum(n.*y.*x.^2)]
sol=A\B, a=sol(1), b=sol(2), c=sol(3)
% La parabola ajustada es y=a+bX+cX^2
my=sum(n.*y)/N, Vy=sum(n.*y.^2)/N-my^2
yest=a+b*x+c*x.^2; res=y-yest
mr=sum(n.*res)/N, Vr=sum(n.*res.^2)/N-mr^2
R2=1-Vr/Vy
```

Problema 6:

Se pide $P_a = P(N \geq 5)$ con $N \to P(1.84)$ que se obtiene mediante:

$P_a = 1 - P(N = 0) - P(N = 1) - P(N = 2) - P(N = 3) - P(N = 4)$, pero también valdría emplear la función de distribución de la Poisson:

$P_a = 1 - F_N(x)$ con $4 \leq x < 5$.

```
Pa=1-sum(poisspdf(0:4,1.84))
Pa_2=1-poisscdf(4.5,1,84) %Tambien sirve esto
```

6-b: Ahora ξ sigue una Binomial con n=100 y p el valor obtenido en el apartado anterior.

$P(\xi \geq 8) = 1 - \sum_{k=0}^{k=7} P(\xi = k) = 1 - F_\xi(7.5)$

```
Pb=1-sum(binopdf(0:7,100,Pa))
Pb_2=1-binocdf(7.5,100,Pa) %Tambien sirve esto
```

6-c:

```
NIT=10000;                % Numero iteraciones indicadas
x1=poissrnd(1.84,1,NIT); % Genera un vector aleatorio 1x10000 segun una Poisson(1.84)
x2=poissrnd(1.84,1,NIT); % Genera otro vector aleatorio 1x10000 segun una Poisson(1.84)
c=(x1==x2);               % Genera un vector con 10000 componentes.
                          % La componente k-esima es 1 si x1(k)=x2(k) y 0 en caso contrario.
p=sum(c)/NIT              % sum(c) contiene el número de veces que se ha dado la igualdad,
                          % al dividirlo por NIT es la frecuencia relativa.
q=1-p;
Ip=[p-1.96*sqrt(p*q/NIT),p+1.96*sqrt(p*q/NIT)] % Intervalo de confianza al 95% para p.
                                               % Aunque no se pide Ip
```

Problema 7:

```
S=[6.0 5.7 5.4 5.4 6.0 5.7 6.4 6.5 6.3 6.0 5.9 6.3 5.8]
Fe=[5.9 6.6 6.1 6.5 7.0 6.6 6.1 6.1 6.0 6.3 6.2 5.9]
[Ha,Pa]=ttest(S,6,0.05,'both')
[Hb,Pb]=ttest(Fe,6,0.05,'right')
[Hc,Pc]=ttest2(Fe,S,0.05,'right')
[Hd,Pd]=ttest2(Fe,S+0.5,0.05,'both')
```