Grados en Informática
Métodos Estadísticos Examen Junio 2015

- **Tiempo: 2 horas 30 minutos**.
- Dejar DNI encima de la mesa. **Apagar y guardar el MÓVIL**.
- Los problemas 4 y 5 están marcados con **(1P), indicando que pueden ser sustituidos por la nota del control. Contestar a alguno de ellos significa renunciar a la nota obtenida.**

APELLIDOS, NOMBRE:
DNI: Grupo: Titulación:

---

1. Se desea contrastar si el grosor del corcho en el Alentejo (Portugal) es mayor que en Extremadura. Para ello se han dividido ambas regiones en áreas de igual tamaño y se han obtenido los siguientes resultados para el grosor (en cm.):

Alentejo	4.25	4.70	4.00	4.15	4.60	
Extremadura	4.12	4.20	3.98	5.50	5.00	5.20

Supuesto que el grosor sigue una distribución normal:

(a) Dar intervalos de confianza al 95% para el grosor medio del corcho en el Alentejo y para la diferencia de medias $\mu_{Alent} - \mu_{Extr}$.

(b) Contrastar la igualdad o desigualdad de varianzas (σ^2_{Alent} y σ^2_{Extr}).

(c) ¿Podemos asegurar con el 95% de confianza que el grosor medio del corcho alentejano es menor que el del corcho extremeño?

((0.5+0.5)+0.75+0.75=2.5 Puntos)

2. El control de calidad de una cadena de montaje exige que se hagan inspecciones de 4 piezas consecutivas cada cierto tiempo. Al cabo de una año se observó que se habían realizado 1500 pruebas (de 4 piezas cada una). Sea $\mathcal{O}_i$ el número de veces que ha salido i piezas defectuosas, resultando $\mathcal{O}_0 = 981$, $\mathcal{O}_1 = 450$, $\mathcal{O}_2 = 60$, $\mathcal{O}_3 = 6$, $\mathcal{O}_4 = 3$.

(a) Ajustar una distribución binomial a la variable ξ (número de piezas defectuosas).

(b) Contrastar la bondad del ajuste realizado.

(0.35+0.9=1.25 Puntos)

3. Una panadería produce "pan campesino" y ha realizado un estudio de su demanda diaria D, resultando que sigue aproximadamente una Normal de media $\mu = 967.5$ y $\sigma = 30$ unidades. Decide producir K=1000 unidades. Hallar:

(a) Probabilidad de que en determinado día le falten unidades (las venda todas y le pidan más).

(b) Decide que incrementará el número de panes producidos si en dos o más días de la misma semana (6 días) le faltan panes. ¿Cuál es la probabilidad de que ocurra en una semana concreta?

(c) Probabilidad de que en un día que le han faltado panes, no llegue a vender todos los de otra hornada, o sea, no lleguen a venderse 50 panes más, (una hornada son 50 panes).

NOTA: En caso necesario, ebe realizarse la corrección de continuidad. *(0.5+0.6+0.75=1.85 Pts.)*

4. **(1P)** Dada la tabla de frecuencias absolutas:

Intervalo	$[0,6]$	$(6,10]$	$(10,20]$	$(20,30]$	$(30,\infty)$	$\sum$
n_i	1850	989	177	117	67	3200

Se pide:

(a) Representar el histograma de frecuencias absolutas.

(b) Hallar la moda.

(c) Encontrar un intervalo $[C_5, C_{95}]$ que abarca al 90% de la población. (C_k indica el centil k).

(d) Hallar la varianza.

(0.5+0.7+0.7+0.5=2.4 Puntos)

APELLIDOS, NOMBRE:

DNI: Grupo: Titulación:

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

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

5. **(1P)** Dada la tabla bidimensional:

   (a) Ajustar la parábola X/Y: $x = a + by + cy^2$ a los datos de la tabla y hallar su razón de determinación.

$X\backslash Y$	0	1	2	3	4
$(-\infty, 0]$	0	0	315	350	0
$(0, 5]$	0	105	75	45	90
$(5, 10]$	110	200	50	0	0
$(10, 20]$	222	102	140	0	0

   (b) Hallar media y varianza de $Y/_{X\leq 5}$

(0.3+0.3=0.6 Puntos)

6. La duración de un artículo ξ sigue una exponencial de media 8 años. Por otra parte el tiempo desde que se produce hasta su venta (puesta en servicio), sigue una normal ν de media $\mu = 1$ y desviación típica $\sigma = 0.2$. Ambas, ξ y ν, son independientes entre sí.

   Hallar:

   (a) Un vendedor que compra el producto recién fabricado al 30% de su valor de venta. ¿Cuánto tiempo T_0 estima que debe transcurrir hasta que venda el 30% del producto y comience a producir ganancias? $P(\nu \leq T_0) = 0.3$

   (b) (Mediante simulación con 100000 iteraciones). Estimar la proporción de artículos que necesitarán reparación antes de su garantía de 2 años. ($P(\xi \leq 2)$)

   (c) (Mediante simulación con 100000 iteraciones). Hallar la media y varianza del tiempo trascurrido T desde la fabricación del artículo hasta que se avería: $T = \nu + \xi$.

(0.25+0.25+0.2=0.7 Puntos)

7. Un investigador está interesado en averiguar el algoritmo más rápido para resolver una tarea. Quiere distinguir entre 2 posibles Alg_1 y Alg_2. Obtiene los siguientes resultados (en segundos).

Alg_1	44	40	38	36	50	44	56	38	36	46	43
Alg_2	46	40	36	36	56	42	58	42	38	50	$-$

   (a) Contrastar al nivel α =2% que el primer recorrido es mejor, pues disminuye el tiempo medio.

   (b) Contrastar, al mismo nivel, que la media del primer recorrido es 41 minutos.

(0.35+0.35=0.7 Puntos)

SOLUCIONES:

Problema 1:

a1: Se trata de un intervalo de confianza para la media, desviación típica desconocida, muestra pequeña: $I = [\bar{x}_A \pm t_{\frac{0.05}{2},4} \frac{s_A}{\sqrt{n_A}}]$.

$\bar{x}_A = 4.34$, $V_A = 0.0714$, $s_A^2 = \frac{5}{4} V_A = 0.0892$ y $s_A = 0.2987$, por lo que:

$I = [4.34 \pm 2.776 \frac{0.2987}{\sqrt{5}}] = [3.9691, 4.7109]$

a2: Se trata de un intervalo de confianza para la diferencia de medias, desviaciones desconocidas y muestra pequeña. Para calcularlo, debemos antes averiguar si las variaciones son desconocidas y distintas o iguales, que se trata de un contraste que se pregunta como apartado 1-b.

En resumen, que vamos a realizar el apartado 1-b y luego haremos el 1-a2.

1-b: $H_0 : \sigma_A^2 = \sigma_E^2 :$ $\qquad$ $H_0 : \sigma_A^2 \neq \sigma_E^2 :$

En las tablas encontramos como región crítica: $\frac{s_A^2}{s_E^2} \notin \left[F_{1-\frac{\alpha}{2}, n_A-1, n_E-1}, F_{\frac{\alpha}{2}, n_A-1, n_E-1}\right]$

$\bar{x}_E = 4.6667$, $V_E = 0.3464$, $s_E^2 = \frac{6}{5} V_E = 0.4156$ y $s_B = 0.6447$, luego $F_{exp} = \frac{s_A^2}{s_E^2} = 0.2147$

$\left[F_{1-\frac{\alpha}{2}, n_A-1, n_E-1}, F_{\frac{\alpha}{2}, n_A-1, n_E-1}\right] = \left[\frac{1}{9.364}, 7.388\right] = [0.1068, 7.388]$ y como $0.2147 \in [0.1068, 7.388]$, aceptaremos la hipótesis nula de igualdad de varianzas.

Continuación de 1-a2: Al ser las varianzas iguales: $I = \left[(\bar{x}_A - \bar{x}_E) \pm t_{\frac{\alpha}{2}, n_A+n_E-2} s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_E}}\right]$ donde $s_p^2 = \frac{(n_A-1)s_A^2 + (n_B-1)s_E^2}{n_A+n_E-2} = \frac{4(0.0892)+5(0.4156)}{9} = 0.2706 \Rightarrow s_p = 0.5202$.

En las tablas $t_{0.025,9} = 2.262$ y el intervalo pedido será:

$I_{\mu_A - \mu_E} = \left[(4.34 - 4.6667) \pm 2.262(0.5202) \sqrt{\frac{1}{5} + \frac{1}{6}}\right] = [-1.0392, 0.3859]$

1-c: Se trata de un contraste unilateral para la diferencia de medias, muestras pequeñas varianzas desconocidas pero iguales:

$H_0 : \mu_A \geq \mu_E, \qquad H_a : \mu_A < \mu_E$

En las tablas la región crítica es aquella que $E_{exp} = \frac{\bar{x}_A - \bar{x}_E}{s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_E}}} < -t_{\alpha, n_A+n_E-2}$

$E_{exp} = \frac{4.34-4.6667}{0.5202\sqrt{\frac{1}{5}+\frac{1}{6}}} = -1.0371 \not< -1.8331 = t_{0.05,9}$ por lo que aceptamos la hipótesis nula y no podremos asegurar que el grosor del alentejano sea menor.

Problema 2:

El número de piezas defectuosas en cada inspección de 4 seguirá una binomial de n=4, a la que le estimamos p, a partir de los datos, como la proporción de defectuosas de las 6000=1500*4 piezas analizadas. $p = \frac{0(981)+1(450)+2(60)+3(6)+4(3)}{6000} = 0.1$, luego $\xi \to B(4, 0.1)$

$p_0 = P(\xi = 0) = \begin{pmatrix} 4 \\ 0 \end{pmatrix} 0.1^0 0.9^4 \approx 0.6561, \quad p_1 = P(\xi = 1) = \begin{pmatrix} 4 \\ 1 \end{pmatrix} 0.1^1 0.9^3 \approx 0.2916$

$p_2 = P(\xi = 2) \approx 0.0486$, $p_3 = P(\xi = 3) \approx 0.0036$ y $p_4 = P(\xi = 4) \approx 0.0001$.

N.defect.	0	1	2	3	4	
Observadas	981	450	60	6	3	
p_i	0.6561	0.2916	0.0486	0.0036	0.0001	
$e_i = 1500p_i$	984.15	437.40	72.90	5.40	0.15	

Se detecta que la frecuencia esperada $e_4 < 5$ y por tanto debemos juntar la clase 3 y la 4:

N.defect.	0	1	2	$\{3,4\}$	$\sum_i$
O_i	981	450	60	9	1500
p_i	0.6561	0.2916	0.0486	0.0037	
$e_i = 1500p_i$	984.15	437.40	72.90	5.55	
O_i^2	962361	202500	3600	81	
$\frac{O_i^2}{e_i}$	977.8601	462.9630	49.3827	14.5946	1504.8004

$\chi^2_{exp} = \sum_i \frac{O_i^2}{e_i} = 1504.8 - 1500 = 4.8004, \qquad \chi^2_{teo} = \chi^2_{0.05,2} = 5.9915$ y como es mayor que la experimental, aceptaremos la hipótesis nula de que los datos se ajustan bien a una binomial.

Los grados de libertad (gdl) son 2 pues, existen k=4 clases, m=1 (se estima 1 parámetro), así gdl=k-m-1=2.

Problema 3:

3-a: $P(\xi^* > 1000) = P(\xi > 1000.5) = P\left(z > \frac{1000.5-967.5}{30}\right) = P(z > 1.1) = 0.1357$

3-b: El número de días (N) en que ocurre (faltan panes) a la semana seguirá una binomial con n=6 y p=0.1357 (q=1-p=0.8643). Se pide:

$P(N \geq 2) = 1 - p(N = 0) - p(N = 1) = 1 - (0.8643)^6 - 6(0.1357)(0.8643)^5 = 1 - 0.4170 - 0.3927 \approx 0.1904$

3-c $P(\xi^* < 1050/\xi^* > 1000) = P(\xi \leq 1049.5/\xi > 1000.5) = \frac{P(1000.5<\xi\leq1049.5)}{P(\xi>1000.5)} = \frac{0.13253}{0.1357} = 0.9767$

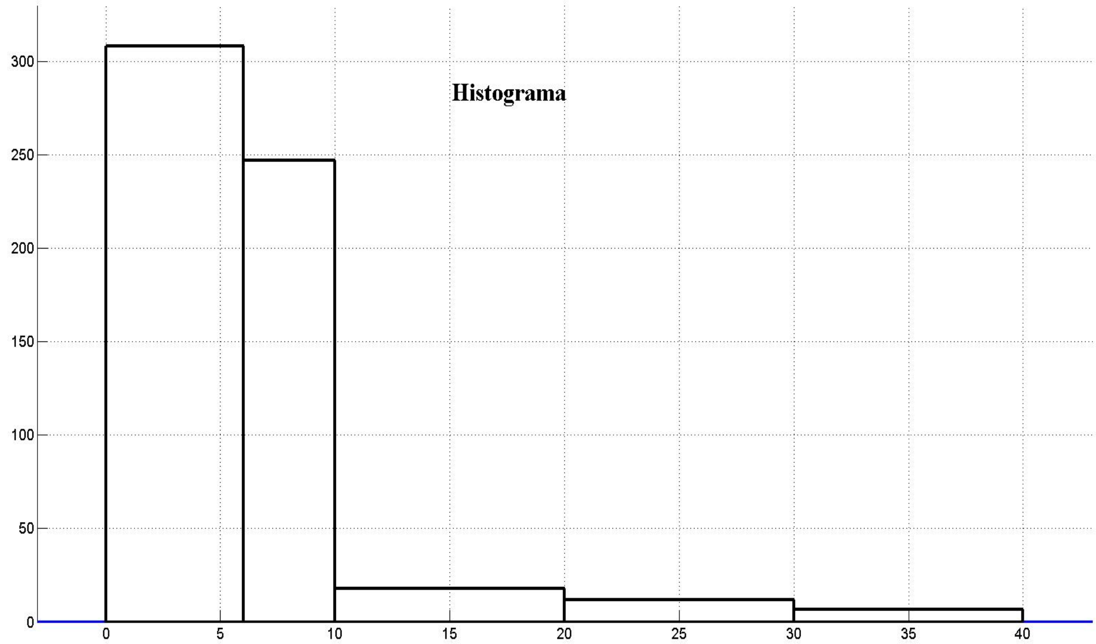
pues $P(1000.5 < \xi \leq 1049.5) = P(\frac{1000.5-967.5}{30} < z \leq \frac{1049.5-967.5}{30}) = P(1.1 < z \leq 2.7333) = 0.1357 - 0.003167 \approx 0.132533$

Es decir, de aquellos días en que pasa de 1000, el 97.67% de las veces no llegaría a vender la siguiente hornada.

Problema 4: Formamos la tabla:

Int	x_i	n_i	a_i	h_i	N_i	$n_i x_i$	$n_i x_i^2$
$[0,6]$	3	1850	6	308.33	1850	5550	16650
$(6,10]$	8	989	4	247.25	2839	7912	63296
$(10,20]$	15	177	10	17.70	3016	2655	39825
$(20,30]$	25	117	10	11.70	3133	2925	73125
$(30,\infty)$	35	67	10	6.70	3200	2345	82075
$\sum_i$		3200				21387	274971

4-a: Las alturas h_i del histograma se han calculado como $h_i = \frac{n_i}{a_i}$ (a_i amplitudes de intervalo).



4-b: El intervalo modal es el de mayor altura en el histograma y en este caso el [0,6].
$\Delta_1 = h_1 - 0 = 308.33$, $\Delta_2 = h_1 - h_2 = 308.33 - 247.25 = 61.08$ y el calculo de la moda da:
$Mo = 0 + \frac{308.33}{308.33+61.08} 6 \approx 5.0079$

4-c:
n1 =3200*0.05=160, n2 =3200*0.95=3040 las frecuencias acumuladas del primer intervalo ya rebasan el valor n1 (160¡1850) por lo que $c_5 = 0 + \frac{160-0}{1850} 6 \approx 0.5189$.

El primero que rebasa el valor n2=3040 es el (20,30], así: $c_{95} = 20 + \frac{3040-3016}{117} 10 \approx 22.0513$ y el intervalo pedido que abarca al 90% de la población es: I=[0.5189, 22.0513].

4-d:
Media=$\frac{21387}{3200} = 6.6834375$, **Varianza:**=$\frac{274971}{3200} - 6.6834375^2 = 41.2601$

Problema 5:

```
clear,clc,format compact
x=[-2.5 -2.5 2.5 2.5 2.5 2.5 7.5 7.5 7.5 15 15 15]
y=[2 3 1 2 3 4 1 2 3 1 2 3]
n=[315 350 105 75 45 90 110 200 50 222 102 140]
N=sum(n)
A=[N sum(n.*y) sum(n.*y.^2)
  sum(n.*y) sum(n.*y.^2) sum(n.*y.^3)
  sum(n.*y.^2) sum(n.*y.^3) sum(n.*y.^4)]
B=[sum(n.*x); sum(n.*x.*y); sum(n.*x.*y.^2)]
sol=A\B
a=sol(1),b=sol(2), c=sol(3)
```

```
disp('La solucion es x=a+by+cy^2')
xest=a+b*y+c*y.^2
res=x-xest
Vr=sum(n.*res.^2)/N-(sum(n.*res)/N)^2
Vx=sum(n.*x.^2)/N-(sum(n.*x)/N)^2
R2=1-Vr/Vx
disp('1-b')
yy=[1 2 3 4]
nn=[105 380 395 90]
NN=sum(nn)
med=sum(nn.*yy)/NN
varyy=sum(nn.*yy.^2)/NN-med^2
```

Problema 6:

```
T0=norminv(0.3,1,0.2)
disp('b:')
NIT=100000;
XI=exprnd(8,NIT,1);
C=(XI<=2);PROP=sum(C)/NIT
disp('El valor real, no estimado, puede calcularse:')
Valor_real=expcdf(2,8)
disp('c:')
NU=normrnd(1,0.2,NIT,1);
T=XI+NU;MED=mean(T),VARIANZA=var(T)
% La cuasivarianza es el mejor estimador de la varianza de T
% que es lo que queremos, asi pues, var(T) --> cuasiv.
```

Problema 7

```
A1=[44 40 38 36 50 44 56 38 36 46 43]
A2=[46 40 36 36 56 42 58 42 38 50]
alfa=0.02;
[Ha,Pa]=ttest2(A1,A2,alfa,'left')
[Hb,Pb]=ttest(A1,41,alfa,'both')
```