Grados en Informática
Métodos Estadísticos Examen Septiembre 2015

- **Tiempo: 2 horas 30 minutos**.
- Dejar DNI encima de la mesa. **Apagar y guardar el MÓVIL**.

APELLIDOS, NOMBRE:
DNI: Grupo: Titulación:

---

1. Se desea encontrar el mejor recorrido para ir a casa entre 2 posibles A y B. Para ello se mide el tiempo transcurrido por cada uno de ellos en minutos, obteniéndose los valores:

A	29.5	31.3	28.5	29.1	30.9	
B	29.98	29.3	28.42	34.5	32.5	33.3

Supuesto que el tiempo por cada uno de ellos son distribuciones normales:

(a) Dar intervalos de confianza al 95% para la media μ_A y para la diferencia de medias $\mu_A - \mu_B$.
(b) Contrastar la igualdad o desigualdad de varianzas (σ_A^2 y σ_B^2).
(c) ¿Podemos asegurar con el 95% de confianza que el recorrido A es más rápido que el B?

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

2. La proporción de azufre (en ppm) en un acero inoxidable sigue una variable aleatoria S de función de densidad:

$$f_S(x) = \begin{cases} \frac{x-1}{8} & 1 \leq x \leq 3 \\ \frac{9-x}{24} & 3 < x \leq 9 \end{cases}$$

(a) Calcular la media.
(b) Calcular la mediana.

(0.6+0.9=1.5 Puntos)

3. El tiempo de fabricación de una pieza t (expresado en minutos), consta de 3 partes bien diferenciadas ($t = t_1 + t_2 + t_3$).

El tiempo t_i que se tarda en cada proceso sigue respectivamente:

$p(t_1 = 0) = 0.4$, $p(t_1 = 1) = 0.6$ (Preparación del material)

$t_2 \longrightarrow N(2, 0.4)$ (Fabricación propiamente dicha)

$t_3 \longrightarrow N(3, 0.3)$ (Empaquetado)

Hallar:

(a) Probabilidad de que $t_2 + t_3 > 5.5$. ($P(t_2 + t_3 > 5.5)$)
(b) Probabilidad de que $t > 5.5$. ($P(t > 5.5)$)
(c) Probabilidad de que si ha tardado más de 5.5 ($t > 5.5$), haya necesitado preparación $t_1 = 1$
(d) Probabilidad de que de 50 unidades fabricadas, elegidas al azar, 25 ó más de ellas no hayan necesitado preparación ($t_1 = 0$).

En caso necesario, debe realizarse la corrección de continuidad. *(0.4+0.4+0.4+0.4=1.6 Pts.)*

4. Dada la tabla de frecuencias absolutas bidimensional:

$Y\backslash X$	$[-7,-3]$	$(-3,3]$	$(3,7]$	$(7,\infty)$
1	15	10	0	0
2	0	5	5	0
3	0	0	0	15

Se pide:

(a) Representar el histograma de frecuencias absolutas de la variable marginal de X.
(b) Hallar el rango intercuartílico de la marginal de X.
(c) Ajustar una función de la forma $X = a + b\,\mathrm{Ln(Y)}$
(d) Hallar el coeficiente de determinación del ajuste propuesto.

(0.5+0.5+0.7+0.7=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. Dada la tabla bidimensional:

   (a) Ajustar la función: $y = a + bx^2 + cx^4$ a los datos de la tabla y hallar su razón de determinación.

$Y \backslash X$	-2	-1	0	1	2	3
$(-\infty, 0]$	0	215	135	65	0	0
$(0, 5]$	84	17	5	0	0	0
$(5, 10]$	10	986	115	32	0	0
$(10, 20]$	0	0	140	220	145	25

   (b) Hallar media y varianza de $Y/_{X \geq 0}$

(0.3+0.3=0.6 Puntos)

6. El número diario de artículos desechados por el control de calidad establecido previo a la puesta a la venta N_1, sigue una distribución de Poisson de parámetro 7.77. Por otra parte, de los artículos diarios puestos en venta, son devueltos en número N_2 que sigue una Poisson de parámetro 2.13. Estimar mediante simulación con 100000 iteraciones:

   Hallar:

   (a) Media y varianza de artículos rechazados $N = N_1 + N_2$.

   (b) Probabilidad de que determinado dia sean rechazados más de 10 artículos. ($P(N > 10)$)

(0.35+0.35=0.7 Puntos)

7. Un estudio desea datar un documento. Para ello necesita conocer el contenido del isótopo carbono 14 (^{14}C) en el mismo. Extrae 32 muestras al azar y las manda a laboratorios diferentes, obteniendo los datos (en $10^{-12}\%$):

Lab_1	44	40	38	36	50	44	56	38	36	46	43	44	42	40	46	44
Lab_2	46	40	36	36	56	42	58	42	38	50	48	48	45	42	45	46

   (a) Contrastar al nivel α =3% que el primer laboratorio indica concentraciones inferiores al otro.

   (b) Contrastar, al mismo nivel, que la media del primer recorrido es 44 (expresado en $10^{-12}\%$).

(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: $I = [\bar{x}_A \pm t_{\frac{\alpha}{2},n-1}\frac{s_A}{\sqrt{n_A}}]$.

$\bar{x}_A = 29.86$, $V_A = 1.1424$, $s_A^2 = \frac{5}{4}V_A = 1.4280$, $s_A = 1.1950$ y $t_{0.0.25,4} = 2.776$, por lo que:

$I = [29.86 \pm 2.776\frac{1.195}{\sqrt{5}}] = [28.3765, 31.3435]$

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 varianzas son " desconocidas y distintas" o "desconocidas pero iguales", que debemos tratar como 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: $H_0 : \sigma_A^2 = \sigma_B^2 :$ $\qquad H_a : \sigma_A^2 \neq \sigma_B^2 :$

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

$\bar{x}_B = 31.3333$, $V_B = 4.9517$, $s_B^2 = \frac{6}{5}V_B = 5.942$ y $s_B = 2.4376$, luego $F_{exp} = \frac{s_A^2}{s_B^2} \approx 0.2403$

$\left[F_{1-\frac{\alpha}{2},n_A-1,n_B-1}, F_{\frac{\alpha}{2},n_A-1,n_B-1}\right] = [F_{0.975,4,5}, F_{0.025,4,5}] = \left[\frac{1}{9.364}, 7.388\right] = [0.1068, 7.388]$ y como $0.2403 \in [0.1068, 7.388]$, aceptaremos la hipótesis nula de igualdad de varianzas.

Continuación de 1-a2: Al ser las varianzas iguales: $I = \left[(\bar{x}_A - \bar{x}_B) \pm t_{\frac{\alpha}{2},n_A+n_B-2}s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}\right]$ donde $s_p^2 = \frac{(n_A-1)s_A^2+(n_B-1)s_B^2}{n_A+n_B-2} = \frac{4(0.0892)+5(0.4156)}{9} = 3.9358 \Rightarrow s_p = 1.9839$.

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

$I_{\mu_A-\mu_B} = \left[(29.86 - 31.3333) \pm 2.262(1.9839)\sqrt{\frac{1}{5} + \frac{1}{6}}\right] = [-4.1907,\ 1.2440]$

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

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

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

$E_{exp} = \frac{29.86-31.3333}{1.9839\sqrt{\frac{1}{5}+\frac{1}{6}}} = -1.2264 \not< -1.8331 = t_{0.05,9}$ por lo que aceptamos la hipótesis nula y no podremos asegurar que el recorrido A sea más rápido.

Problema 2:

a:

$$E(S) = \int_1^3 x\frac{x-1}{8}dx + \int_3^9 x\frac{9-x}{24}dx = \frac{1}{8}\left[\frac{x^3}{3} - \frac{x^2}{2}\right]_1^3 + \frac{1}{24}\left[\frac{9x^2}{2} - \frac{x^3}{3}\right]_3^9 = \frac{7}{12} + \frac{15}{4} = \frac{13}{3} \approx 4.3333$$

b: Debemos resolver $F(x) = 0.5$

Si la mediana estuviese en el primer intervalo ($1 \leq x \leq 3$):

$$F(x) = \int_1^x \frac{x-1}{8}dx = 0.5 \Rightarrow \frac{1}{8}\left[\frac{x^2}{2} - x\right]_1^x = \frac{x^2}{16} - \frac{x}{8} + \frac{1}{16} = 0.5 \Rightarrow x^2 - 2x - 7 = 0$$

que tiene por raíces $x_1 \approx 3.8284$ y $x_2 \approx -1.8284$, que no sirven pues no están en [1,3].

Por tanto la mediana tiene que estar en [3,9]:

$$F(x) = \int_1^3 \frac{x-1}{8}dx + \int_3^x \frac{9-x}{24}dx = 0.5 \Rightarrow \frac{1}{4} + \frac{1}{24}\left[9x - \frac{x^2}{2}\right]_3^x = 0.5 \Rightarrow \frac{1}{4} + \frac{1}{24}\left(9x - \frac{x^2}{2} - 27 + \frac{9}{2}\right) = 0.5$$

$$\Rightarrow x^2 - 18x + 57 = 0 \Rightarrow x_1 \approx 13.8990 \text{ (no sirve) y } x_2 \approx 4.1010 \text{ que es la mediana.}$$

Problema 3:

3-a: Nos piden $P(t_2 + t_3 > 5.5)$, pero $\xi = t_2 + t_3$ por ser suma de normales independientes seguirá una normal: $\xi \to N(2 + 3, \sqrt{0.4^2 + 0.3^2}) = N(5, 0.5)$

$P(\xi > 5.5) = P(z > \frac{5.5-5}{0.5}) = P(\xi > 1) \approx 0.1587$

3-b: Nos piden $P(t > 5.5) = P(t_1 + t_2 + t_3 > 5.5)$, para calcular esa probabilidad debemos distinguir 2 casos, que no se necesite preparación ($t_1 = 0$) y que si se necesite ($t_1 = 1$).

Asi, $P(t_1+t_2+t_3 > 5.5) = P(t_1 = 0)P(t_1+t_2+t_3 > 5.5)+P(t_1 = 1)P(t_1+t_2+t_3 > 5.5) = 0.4P(0+t_2+t_3 > 5.5) + 0.6P(1 + t_2 + t_3 > 5.5) = 0.4(0.1587) + 0.6(0.8413) = 0.5683$, pues $P(1 + t_2 + t_3 > 5.5) = P(t_2 + t_3 > 4.5) = P(z > \frac{4.5-5}{0.5}) = P(z > -1) = 1 - P(z \leq -1) = 1 - P(z \geq 1) = 1 - 0.1587 = 0.8413$.

3-c: Nos piden $P(t_1 = 1/t > 5.5) = \{\text{Teorema de Bayes}\} = \frac{P(t_1=1 \wedge t>5.5)}{P(t>5.5)} = \frac{0.6(0.8413)}{0.5683} = 0.8882$

3-d: El número X de las que no necesitan preparación seguirá una distribución binomial: $X \to B(50, 0.4)$ al ser $N = 50 > 30$ y $Np = 50(0.4) = 20 > 5$ $Nq = 30 > 5$ se aproximará por una normal X^* de media $\mu = Np = 20$ y desviación $\sigma = \sqrt{Npq} = \sqrt{50(0.4)(0.6)} = \sqrt{12} = 3.4641$, es decir $N \to N(20, 3.4641)$.

Nos piden $P(X \geq 25) = \{$ Por la corrección de continuidad $\} = P(X^* > 24.5) = P\left(z > \frac{24.5-20}{3.4641}\right) = P(z > 1.299) \approx \mathbf{0.097}$ que hemos calculado mediante interpolación en las tablas de la normal:

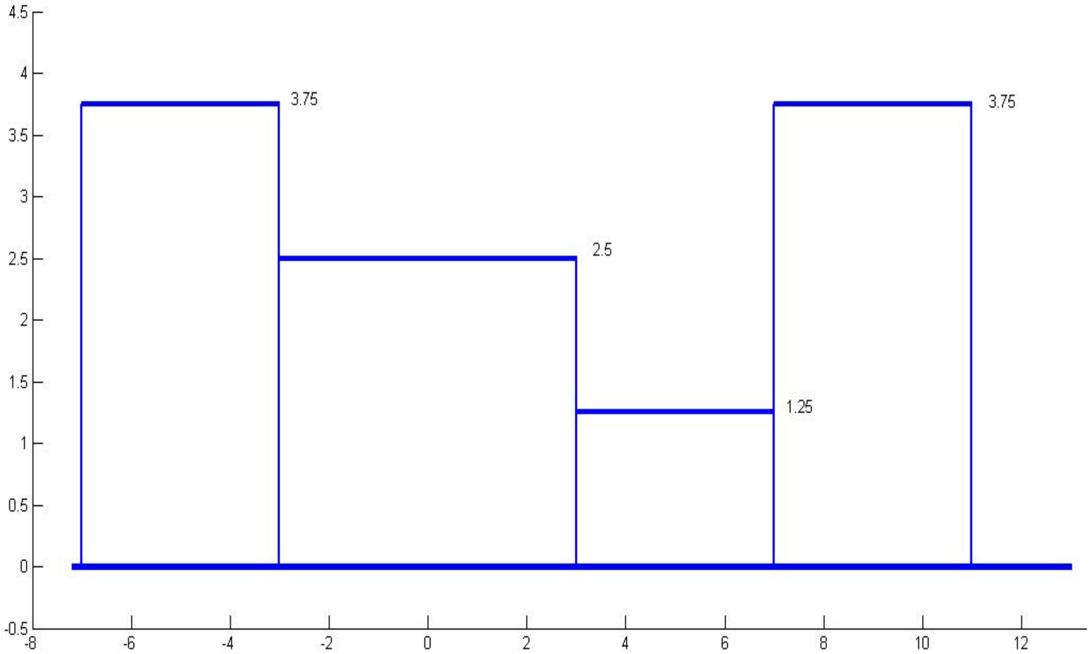
En las tablas obtenemos: $P(z > 1.29) = 0.0985$ y $P(z > 1.30) = 0.0968$, por lo que $P(z > 1.299) = 0.0985 + 0.9(0.0968 - 0.0985) \approx 0.0970$

Problema 4:

La variable marginal de X es:

Int.	n_i	a_i	$h_i = \frac{n_i}{a_1}$	N_i
$[-7, -3]$	15	4	3.75	15
$(-3, 3]$	15	6	2.5	30
$(3, 7]$	5	4	1.25	35
$(7, \infty)$	15	4	3.75	50

El histograma queda:



4-b: $R_{IC} = Q_3 - Q_1$ por lo que calcularemos el cuartil 1 y 3:
Para Q_1 Nc=50(0.25)=12.5 que está en el primer intervalo, así: $Q_1 = -7 + \frac{12.5-0}{15}4 = -7 + \frac{10}{3} = -\frac{11}{3}$
Para Q_3 Nc=50(0.75)=37.5 que está en el último intervalo, así: $Q_3 = 7 + \frac{37.5-35}{15}4 = 7 + \frac{10}{15} = \frac{23}{3}$
y el rango intercuartílico pedido vale: $R_{IC} = \frac{23}{3} - \left(-\frac{11}{3}\right) = \frac{34}{3}$

4-c: Formamos la tabla de cálculos:

x_i	y_i	n_i	$Y_i = Ln(y_i)$	n_iY_i	$n_iY_i^2$	n_ix_i	$n_ix_iY_i$	X_i^{est}	r_i	n_ir_i	$n_ir_i^2$	$n_ix_i^2$
-5	1	15	0	0	0	-75	0	-3.2317	-1.7683	-26.5251	46.9054	375
0	1	10	0	0	0	0	0	-3.2317	3.2317	32.3166	104.4363	0
0	2	5	0.6931	3.4657	2.4023	0	0	4.0692	-4.0692	-20.3461	82.7926	0
5	2	5	0.6931	3.4657	2.4023	25	17.3287	4.0692	0.9308	4.6539	4.3318	125
9	3	15	1.0986	16.4792	18.1042	135	148.3127	8.3400	0.6600	9.9007	6.5349	1215
		50		23.4107	22.9088	85	165.6413			0	245.0009	1715

Resolviendo el sistema de ecuaciones normales:

$$\left.\begin{array}{r} 50a + 23.4107b = 85 \\ 23.4107a + 22.9088b = 165.6413 \end{array}\right\} \Rightarrow a = -3.2317, \quad b = 10.5329 \Rightarrow \mathbf{X} = \mathbf{-3.2317} + \mathbf{10.5329}\mathrm{Ln(Y)}$$

4-d:

Hallamos los $X_i^{est} = -3.2317 + 10.5329\mathrm{Ln(y_i)}$ y los residuos $r_i = X_i - X_i^{est}$
La varianza residual será: $V_r = \frac{245.0009}{50} - \left(\frac{0}{50}\right)^2 = 4.9$, mientras que $V_x = \frac{1715}{50} - \left(\frac{85}{50}\right)^2 = 31.41$
El coeficiente de determinación vale: $\mathbf{R^2 = 1 - \frac{4.9}{31.41} = 0.844}$

Problema 5:

```
clc,clear all,format compact
disp('Problema 5')
y=[-2.5*ones(1,3) 2.5*ones(1,3) 7.5*ones(1,4) 15*ones(1,4)]
x=[-1:1 -2:0 -2:1 0:3]
n=[215 135 65 84 17 5 10 986 115 32 140 220 145 25]
N=sum(n);
A=[N sum(n.*x.^2) sum(n.*x.^4)
   sum(n.*x.^2) sum(n.*x.^4) sum(n.*x.^6)
   sum(n.*x.^4) sum(n.*x.^6) sum(n.*x.^8)]
B=[sum(n.*y);sum(n.*y.*x.^2); sum(n.*y.*x.^4)]
sol=A\B; a=sol(1), b=sol(2),c=sol(3)
yest=a+b*x.^2+c*x.^4;
r=y-yest;
Vr=sum(n.*r.^2)/N-(sum(n.*r)/N)^2;
Vy=sum(n.*y.^2)/N-(sum(n.*y)/N)^2;
R2=1-Vr/Vy
disp('5-b')
yy=[-2.5 2.5 7.5 15]
nn=[135+65 5 115+32 140+220+145+25]
NN=sum(nn)
med=sum(nn.*yy)/NN
V=sum(nn.*yy.^2)/NN-med^2
```

Problema 6:

```
clear all
disp('Problema 6')
Nit=1e5;
n1=poissrnd(7.77,Nit,1);
n2=poissrnd(2.13,Nit,1);
n=n1+n2;
med=mean(n)
V=var(n), % Cuasivarianza (mejor estimador del valor correcto)
disp('6-b')
c=(n>10);
P=sum(c)/Nit %Estimador puntual
Q=1-P;
I=[P-1.96*sqrt(P*Q/Nit), P+1.96*sqrt(P*Q/Nit)]
```

Problema 7:

```
clear all
disp('Problema 7')
L1=[44 40 38 36 50 44 56 38 36 46 43 44 42 40 46 44]
L2=[46 40 36 36 56 42 58 42 38 50 48 48 45 42 45 46]
alfa=0.03;
[Ha,Pa]=ttest2(L1,L2,alfa,'left')
disp('7-b')
[Hb,Pb]=ttest(L1,44,alfa,'both')
```