Grados en Ingeniería Informática
Estadística
Examen **Convocatoria Septiembre 2014**

- A resolver en **2 horas**
- Dejar DNI encima de la mesa.
- **Apagar y guardar el MÓVIL**.

APELLIDOS, NOMBRE:
DNI: Especialidad: Grupo:

---

1. Un productor de mandarinas desea comprobar cómo influye el abonado (N) en *gr.* $\times$ árbol $\times$ mes, con el tamaño del fruto (T) obtenido (diametro en mm.). En el estudio se obtuvo la siguiente tabla de frecuencias:

$N \backslash T$	$[52, 58]$	$(58, 62]$	$(62, 66]$	$(66, 72]$
$[30, 38]$	4	2	1	0
$(38, 42]$	1	6	4	2
$(42, 52]$	0	2	4	4

Calcular:

(a) Calcular la moda y la mediana de la variable T=tamaño del fruto.
(b) Calcular la dispersión relativa de la variable $N/T \in (58, 62]$.
(c) Hallar el coeficiente de correlación lineal e interpretarlo.

(1+0.5+0.5= 2 Puntos)

2. El tiempo de atención a un cliente, (en minutos), sigue una variable aleatoria ξ de la que se ha estimado su función de densidad $f(x)$:

$$f(x) = \begin{cases} \frac{x}{50} & \text{Si } 0 \leq x < 5 \\ K - \frac{x}{25} & \text{Si } 5 \leq x < 10 \\ 0 & \text{En el resto} \end{cases}$$

Hallar:

(a) Calcular el valor de K, la media y la mediana de la variable ξ.
(b) Calcular la probabilidad de que el tiempo de espera sea, a lo sumo de 7 minutos.
(c) Calcular la probabilidad de que si han transcurrido más de 4 minutos, la duración sea menor o igual a 7 minutos.

(1.5+0.5+0.5= 2.5 Puntos)

3. Se realiza un estudio en una plantación de mangos de la Axarquía para ver cuál de 2 tipos de abonado produce mayor cantidad de frutos. Se toman 45 parcelas de igual superficie y se abonan 25 con la fórmula A=(15-15-15) y 20 de ellas con la B=(15-46-15). Se obtuvo una producción media por parcela de 57 Kg. y 62 Kg. respectivamente. Mientras que $\sum_{i \in A} x_i^2 = 81609$ y $\sum_{i \in B} x_i^2 = 77260$.

(a) Intervalo de confianza para la producción media de una parcela abonada con el tipo B, al nivel del 90%.
(b) Saber si puede ser aceptada la hipótesis de que la media por parcela abonado con el tipo A es superior o igual a 58 Kg. al nivel del 90%.
(c) Se puede afirmar, al 95%, que las parcelas con abono tipo B dan más producción que las abonadas con el tipo A.
(d) Se puede afirmar, al 95%, que las dispersiones varían según el tipo de abonado.

(0.75+0.75+0.75+0.75 = 3 Puntos)

4. **PRÁCTICAS (Sólo instrucciones)**

   Los datos de venta anuales son:

año	2006	2007	2008	2009	2010	2011	2012	2013
Uds.	1017	1043	1066	1100	1098	1110	1141	1150

   Se pide:

   (a) Ajustar la recta de tendencia.

   (b) Hallar el coeficiente de correlación lineal.

   (c) Estimar las ventas previstas por el modelo para los años 2014 y 2015.

   (0.4+0.3+0.3=1 Punto)

5. **PRÁCTICAS (Sólo instrucciones)**

   El tiempo que se tarda en llegar a TEATINOS en metro se compone de el tiempo de llegada del tren ξ_1 más el del recorrido ξ_2, donde ξ_1 sigue una uniforme en [0,15], mientras que ξ_2 sigue una normal de media 15 minutos y 3 de desviación. Hallar mediante simulación (10000 iteraciones) los apartados a y b:

   (a) La probabilidad de tardar más de 25 min.

   (b) La probabilidad de tardar más que un autobús que pasa en ese momento por la parada del metro y cuya duración del trayecto sigue una normal de media 26 min. con desviación de 5 min.

   (c) Queremos contrastar que la duración del trayecto del metro es en realidad superior a 15 minutos, (sin fiarnos tampoco de que la desviación sea 3) para ello medimos durante 14 días el tiempo del trayecto obteniendo los tiempos 14, 16, 18, 14, 17, 15, 15, 16, 14, 19, 17, 16, 16, 15. Hacer el contraste al nivel del 5%.

   (0.5+0.5+0.5=1.5 Puntos)

Problema 1:
a:

I_i	T_i	n_i	N_i	a_i	h_i
$[52,58]$	55	5	5	6	$\frac{5}{6}$
$(58,62]$	60	10	15	4	2.5
$(62,66]$	64	9	24	4	2.25
$(66,72]$	69	6	30	6	1
		30			

Mediana: N=30 $\Rightarrow \frac{N}{2} = 15$ El intervalo mediano es el (58,62] y la mediana es 62, pues tiene 15 menores y 15 mayores. Por la fórmula:

$$M_e = L_i + \frac{\frac{N}{2} - N_{i-1}}{n_i} a_i = 58 + \frac{15-5}{10} 4 = 62$$

Moda: La mayor altura (h_i) del histograma se da en el intervalo (58,62] que será el intervalo modal.

$$Mo = L_{i-1} + \frac{\Delta_1}{\Delta_1 + \Delta_2} a_i = 58 + \frac{\frac{5}{3}}{\frac{5}{3} + 0.25} \approx 61.4783$$

b: Calculamos la condicional de N a que $T \in (58,62]$ y operamos en la tabla:

I_i	N_i	n_i	n_iN_i	$n_iN_i^2$
$[30,38]$	34	2	68	2312
$(38,42]$	40	6	240	9600
$(42,52]$	47	2	94	4418
		10	402	16330

Media=40.2, Varianza= $\frac{4418}{10} - 40.2^2 = 16.96$
Desv. Típica= $\sqrt{16.96} = 4.1183$, Coef. Var.= $\frac{4.1183}{40.2} = 0.1024$
c: $\bar{T} = \frac{55*5+60*10+64*9+69*6}{30} = 62.1667$
$m_2(T) = \frac{55^2*5+60^2*10+64^2*9+69^2*6}{30} = 3885.1667$
$V(T) = 3885.1667 - 62.1667^2 \approx 20.47 \Rightarrow \sigma_T \approx 4.524$
$\bar{N} = \frac{34*7+40*13+47*10}{30} = 40.9333$
$m_2(N) = \frac{34^2*7+40^2*13+47^2*10}{30} = 1699.4$
$V_N = 1699.4 - 40.9333^2 \approx 23.8622, \Rightarrow \sigma_N \approx 4.8849$
$M_{NT} = \frac{4*34*55+40*55+2*34*60+6*40*60+2*47*60+34*64+4*40*64+4*47*64+2*40*69+4*47*69}{30} \approx 2558$
y $Cov = 2301.13 - 62.1667 * 40.9333 = 13.3118$, $\mathbf{r} = \frac{\mathbf{13.3118}}{\mathbf{4.524*4.8849}} \approx \mathbf{0.6023}$
Luego existe una correlación directa (al crecer la cantidad de abono, aumenta el tamaño del fruto).

Problema 2:
a: Tenemos que elegir K que verifique $\int_0^\infty f(x)dx = 1 \Rightarrow \int_0^5 \frac{x}{50}dx + \int_5^{10} K - \frac{x}{25}dx = 1$
$\Rightarrow \left[\frac{x^2}{100}\right]_0^5 + \left[Kx - \frac{x^2}{50}\right]_5^{10} = \frac{25}{100} + 5K - 2 + 0.5 = 1, \Rightarrow K = 0.45 = \frac{9}{20}$

b: Media: $\bar{x} = \int_0^5 \frac{x^2}{50}dx + \int_5^{10} \frac{9x}{20} - \frac{x^2}{25}dx = \left[\frac{x^3}{150}\right]_0^5 + \left[\frac{9x^2}{40}\right]_5^{10} - \left[\frac{x^3}{75}\right]_5^{10} \frac{5^3}{150} + \frac{9(75)}{40} - \frac{1000-125}{75} = \frac{145}{24} \approx 6.0417$
Mediana: Se debe cumplir: $F(a) = 0.5$. Veamos en qué intervalo está. Si estuviese en [0,5], entonces $\int_0^5 f(x)dx \geq 0.5$, pero $\int_0^5 f(x)dx = 0.25$, luego deducimos que la mediana pertenece al intervalo [5,10].

$$\int_0^5 \frac{x}{50}dx + \int_5^a \frac{9}{20} - \frac{x}{25}dx = 0.5 \Rightarrow 0.25 + \frac{9}{20}(a-5) - \frac{a^2-25}{50} = 0.5 \Rightarrow a \approx 6.0961$$

La mediana vale $Me = \frac{45-\sqrt{425}}{4} \approx 6.0961$
b: $P(\xi \leq 7) = \int_0^5 \frac{x}{50}dx + \int_5^7 (\frac{9}{20} - \frac{x}{25})dx = 0.25 + \frac{18}{20} - \frac{49-25}{50} = 0.67$
c: $P(\xi \leq 7/\xi \geq 4) = \frac{P(4 \leq \xi \leq 7)}{P(\xi \geq 4)} = \frac{F(7)-F(4)}{1-F(4)} = \frac{0.67-0.16}{1-0.16} = 0.607$
pues $F(4) = \int_0^4 \frac{x}{50}dx = \left[\frac{x^2}{100}\right]_0^4 = 0.16$
NOTA: Los apartados b y c pueden hacerse también calculando previamente la función de distribución:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x^2}{100} & 0 \leq x \leq 5 \\ -\frac{x^2}{50} + \frac{9x}{20} - \frac{3}{2} & 5 \leq x \leq 10 \\ 1 & x \geq 10 \end{cases}$$

Problema 3:

a: Tenemos los datos $n_A = 25$, $\bar{x_A} = 57$, $s_A^2 = \frac{25}{24}\left(\frac{81609}{25} - 57^2\right) = 16$

$n_B = 20$, $\bar{x_B} = 62$, $s_B^2 = \frac{20}{19}\left(\frac{77260}{20} - 62^2\right) = 20$

$I = \left[\bar{x} \pm t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}\right] = \left[62 \pm t_{0.05,19}\frac{\sqrt{20}}{\sqrt{20}}\right] = \left[62 \pm 1.729\frac{\sqrt{20}}{\sqrt{20}}\right] = [60.271, 63.729]$

b: Se trata de un contraste unilateral de la media con muestra pequeña.

$H_0 : \mu \geq 58$ contra $H_a : \mu < 58$

La región crítica es: $E = \frac{\bar{x}-58}{\frac{s}{\sqrt{n}}} < -t_{\alpha,n-1} \Rightarrow E = \frac{57-58}{\frac{4}{\sqrt{25}}} = -1.25 \not< -t_{0.1,24} = -1.318$, luego no estamos en la región crítica y se acepta la hipótesis nula. **Los diámetros de los frutos son mayores o iguales de 58 mm.**

c: Se trata de un contraste unilateral de la diferencia de medias muestra grande ($n_A + N_B = 45 > 30$) y varianzas desconocidas.

$H_0 : \mu_A \geq \mu_B$ contra que $H_a : \mu_A < \mu_B$

La región crítica es $\frac{\bar{X_A}-\bar{X_B}}{\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}} < -z_\alpha$ con $\alpha = 0.05$

$E = \frac{57-62}{\sqrt{\frac{16}{25}+\frac{20}{20}}} = -3.9 < -z_{0.05} - 1.645$, luego se rechaza la hipótesis nula y por tanto **La media de las parcelas abonadas con B dan mayor producción que las abonadas con el tipo A.**

d: Contraste bilateral de cociente de varianzas.

$H_0 : \sigma_A^2 = \sigma_B^2$, contra $H_a : \sigma_A^2 \neq \sigma_B^2$ y se rechaza H_0 (región crítica) si $\frac{s_A^2}{s_B^2} \notin \left[F_{1-\frac{\alpha}{2},n_A-1,n_B-1}, F_{\frac{\alpha}{2},n_A-1,n_B-1}\right]$,

$\Rightarrow \frac{16}{20} = 0.8 \in [F_{0.975,24,19}, F_{0.025,24,19}] = [\frac{1}{F_{0.025,19,24}}, 2.4523] = [0.3937, 2.4523]$ luego se acepta H_0 y **las dispersiones no varían.**

$F_{0.975,24,19} = \frac{1}{F_{0.025,19,24}} = \frac{1}{2.5401} \approx 0.3937$

$F_{0.025,19,24}$ no viene en las tablas pero si está $F_{0.025,15,24} = 2.437$ y $F_{0.025,24,24} = 2.669$. El valor de $F_{0.025,24,19}$ se obtiene interpolando y resulta ser: $F_{0.025,19,24} = 2.437 + \frac{4}{9}(2.669 - 2.437) = 2.5401$

Problema 4: Prácticas

```
disp('a:')
x=2006:2013
y=[1017 1043 1066 1100 1098 1110 1141 1150]
p=polyfit(x,y,1), a=p(2), b=p(1)
% Podría haberse calculado con:
% N=8, A=[N sum(x);sum(x) sum(x.^2)], B=[sum(y);sum(x.*y)], sol=A\B, a=sol(1), b=sol(2)
disp('b:')
r=corr(x',y')
% Podría haberse calculado con:
% N=8,mx=sum(x)/N,my=sum(y)/N,covar=sum(x.*y)/N-mx*my
% Vx=sum(x.^2)/N-mx^2,Vy=sum(y.^2)/N-my^2,r=covar/sqrt(Vx*Vy)
disp('c:')
v14=polyval(p,2014)
v15=polyval(p,2015)
% O también con:
% v14=a+b*2014, v15=a+b*2015
```

Problema 5: Prácticas

```
NIT=10000
ji1=unifrnd(0,15,NIT,1);
ji2=normrnd(15,3,NIT,1):
ji3=normrnd(26,5,NIT,1); % Para el apartado b
disp('a:')
ca=(ji1+ji2>25); p=sum(ca)/NIT
disp('b:')
cb=(ji1+ji2>ji3); p=sum(cb)/NIT
disp(c:)
x=[14 16 18 14 17 15 15 16 14 19 17 16 16 15]
N=length(x);
[H,p]=ttest(x,15,0.05,'right')
disp('Si H=1 indica que tarda mas de 15 min.')
```