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

- A resolver en **2 horas y media**
- **Apagar y guardar el MÓVIL y dejar el DNI sobre la mesa**.
- **El control realizado en Marzo puede sustituir a los problemas 1 y 6.**

APELLIDOS, NOMBRE:
DNI: Especialidad: Grupo:

---

1. Un estudio sobre el precio de los alquileres en TORRELANDIA, proporcionó las siguientes tablas:

Precios 2013 en euros	n_i
Menos de 200	13
[200,300]	21
(300,500]	25
Más de 500	11

Año	Precio medio	IPC (base 2008)
2009	288	105
2010	308	110
2011	324	115
2012	335	119

   (a) Calcular el precio medio y mediano del alquiler para el año 2013.
   (b) Calcular la serie de índices simples de los precios medios de alquiler con base 2010.
   (c) Calcular la serie de números índices simples, con base 2010, una vez deflaccionada la serie temporal con el IPC.
   (d) Calcular la recta de tendencia (datos 2009 a 2012) y ver si predice bien el valor observado en 2013.
   (e) Hallar el coeficiente de regresión lineal y la varianza residual.

((0.35+0.3)+0.3+0.3+0.5+0.5= 2.25 Puntos)

2. Una franquicia gestiona 3 empresas. Se ha observado que las ventas diarias de cada una de ellas, en euros (X_i), siguen respectivamente: $X_1 \to N(1000, 200)$, $X_2 \to N(1100, 200)$ y $X_3 \to N(900, 150)$.

   Por otra parte, el número de bajas diarias por enfermedad de cada empresa, (Y_i), siguen respectivamente distribuciones de Poisson de parámetros 1. ($Y_1 \to P(1), Y_2 \to P(1), Y_3 \to P(1)$).

   Se sabe que la distribución de Poisson es reproductiva, esto es:

   "Si $\xi_1 \to \mathbf{P}(\lambda_1)$ y $\xi_2 \to \mathbf{P}(\lambda_2)$, entonces $\xi = \xi_1 + \xi_2$ seguirá también una Poisson de parámetro $\lambda_1 + \lambda_2$". ($\xi \to P(\lambda_1 + \lambda_2)$).

   Todas las variables son independientes entre sí.

   (a) $P(X_1 > 1000)$.
   (b) $P(X_2 > X_1)$.
   (c) Probabilidad de que la venta total de las 3 empresas sea mayor de 3500 euros.
   (d) Probabilidad de que en un día el número total de bajas sea superior a 2.
   (e) Probabilidad de que en el día de hoy y de mañana no haya ninguna baja en la primera empresa.
   (f) Si se observan las ventas de 10 días en la primera empresa, determinar la probabilidad de que en alguno de ellos se haya vendido más de 1088 euros.

(0.3+0.3+0.3+0.3+0.3+0.5=2 Puntos)

3. La variable aleatoria ξ representa el tiempo transcurrido, en minutos, hasta que un motor alcanza la temperatura ideal de trabajo. Conocemos la función de densidad de ξ que viene dada por:

$$f(x) = \begin{cases} kx & x \in [0,1] \\ k & x \in (1,2] \\ 0 & x \notin [0,2] \end{cases}$$

   (a) Hallar k y la función de distribución de ξ.
   (b) Hallar la mediana de ξ e interpretarla.

(0.5+0.25= 0.75 Puntos)

4. Un estudio previo realizado 25 años antes determinó que en una región los grupos sanguíneos "0", "A", "B" y "AB", se encontraban en la relación 14:12:3:1. Se quiere contrastar si esa relación sigue siendo correcta. Para ello se toma una muestra al azar de 300 individuos observándose los valores $n(0) = 126$, $n(A) = 117$, $n(B) = 42$ y $n(AB) = 15$.

   ¿Puede admitirse que las anteriores proporciones siguen siendo válidas? *(1 Punto)*

5. Un estudio sobre el tamaño de las memorias empleadas en los teléfonos móviles (en GB) dió en los estudiantes de Informática los siguientes resultados: $N_I = 11$, $\sum I_{i=1}^{11} = 143$, $\sum_{i=1}^{11} I_i^2 = 1889$, mientras que para los de Medicina: $N_M = 9$ dando $\sum_{i=1}^{9} M_i = 108$ y $\sum_{i=1}^{9} M_i^2 = 1328$.

   Contrastar al 5% que los estudiantes de Informática usan móviles con más memoria que los de Medicina.

   (1.5 Puntos)

PRÁCTICAS Indicar solo las órdenes necesarias en MATLAB para resolverlo

=== Entregar en folio aparte ===

6. (a) Ajustar una función de la forma $Y = a + b\cos(X)$ a los datos de la tabla:

$X \backslash Y$	$[0, 8]$	$(8, 12]$	$(12, 16]$	$(16, 24]$
0	0	0	50	75
$\frac{\pi}{4}$	0	71	51	12
$\frac{\pi}{2}$	36	41	0	0
$\frac{3\pi}{4}$	0	49	44	0
π	0	0	25	18

   (b) Hallar el coeficiente de determinación del ajuste realizado.

   (c) Hallar media, varianza y sesgo de $Y/(X \geq \frac{3\pi}{4})$.

   (0.25+0.25+0.25=0.75 Puntos)

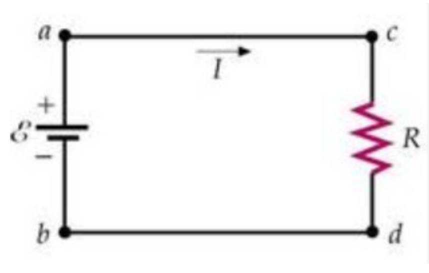
7. El circuito de la figura consta de una pila $\mathcal{E}$ y una resistencia R. Por márgenes de fabricación, el valor de $\mathcal{E}$ en Voltios sigue una distribución normal de media 1.5 y desviación 0.2, teniendo la pila una resistencia interna r que sigue, en ohmios, también una normal de media 0.3 y desviación típica 0.01. Por otra parte, el valor, en ohmios, de la resistencia R sigue una N(1.2,0.02).

   Se sabe que, por la Ley de Ohm, el valor de la intensidad (I), en amperios, es $I = \frac{\mathcal{E}}{R+r}$.

   Estimar por simulación con 10000 iteraciones:

   (a) La media y la varianza de I.

   (b) La probabilidad de que la intensidad sea mayor que 1.1 amperios.



   (0.5+0.5=1 Puntos)

8. Se quiere contrastar si las resistencias R del problema anterior tienen realmente una media $\mu = 1.2$ y una desviación $\sigma = 0.02$. Para ello se toma una muestra aleatoria de ellas, obteniéndose los valores:

   1.23, 1.21, 1.17, 1.11, 1.3, 1.28, 1.19, 1.18, 1.31, 1.12, 1.19, 1.24, 1.22

   (a) ¿Puede considerarse que la desviación es 0.02?

   (b) ¿Puede considerarse que la media es 1.2?

   (0.4+0.35=0.75 Puntos)

SOLUCIONES:

Ejercicio 1:

a: En la tabla pondremos las frecuencias absolutas acumuladas N_i, las marcas de clase x_i y los productos $x_i n_i$:

Precios 2013 en euros	n_i	N_i	x_i	$x_i n_i$
Menos de 200	13	13	150	1950
[200,300]	21	34	250	5250
(300,500]	25	59	400	10000
Más de 500	11	70	600	6600
	70			23800

obteniéndose la media $\bar{x} = \frac{\sum_i x_i n_i}{N} = \frac{23800}{70} = 340$.

Para la mediana buscamos la clase que la contiene $N/2 = 35$, hasta 300 hay 34 datos, luego estará en el intervalo [300,500], aplicando la fórmula:

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} a_i = 300 + \frac{35 - 34}{25} 200 = 308$$

b: Los calculamos en la tabla: Indices simples $= \frac{\text{Precio medio del año}}{\text{Precio medio año 2010}}$.

c: Deflaccionamos dividiendo los precios medios por el IPC y multiplicando por 110 (valor del IPC en 2010)

Año	Precio medio	IPC (base 2008)	Ind. simples	Defl.(2010)	Ind. Defl.
2009	288	105	0.9351	301.7143	0.9796
2010	308	110	1.0000	308.0000	1.0000
2011	324	115	1.0519	309.9130	1.0062
2012	335	119	1.0877	309.6639	1.0054

d: Para calcular la recta de tendencia cambiamos la variable "Año" por T=2*(Año-2010.5):

Año	P_i	T_i	Ind. simples	Deflacc.
2009	288	-3	-864	9
2010	308	-1	-308	1
2011	324	1	324	1
2012	335	3	1005	9
$\sum$	1255	0	157	20

Resolvemos las ecuaciones normales:

$$\left.\begin{array}{ll} \sum_i P_i & = aN + b\sum_i T_i \\ \sum_i P_i T_i & = a\sum_i T_i + b\sum_i T_i^2 \end{array}\right\} \Rightarrow \left.\begin{array}{ll} 1255 & = 4a + 0b \\ 157 & = 0a + 20b \end{array}\right\} \Rightarrow a = 313.75, \quad b = 7.85$$

La recta de tendencia es: $P = 313.75 + 7.85T_i = 313.75 + 7.85 * 2(\text{Año} - 2010.5) = -31251.1 + 7.85(\text{Año})$

El valor que predice para el año 2013, $T_i = 2(2013 - 2010.5) = 5$ es: $P_i = 313.75 + 7.85 * 5 = 353$, mientras el valor real, obtenido en el apartado "a", es 340.

Por tanto el error (residuo) sería r=353-340=13, que parece elevado y podría deberse a un cambio en la tendencia, o a que el ajuste lineal no fuese bueno. Esto último vamos a comprobarlo en el siguiente apartado.

e: Calculariamos:

$$r = \frac{cov(P, A)}{\sigma_A \sigma_P} = \frac{19.6250}{1.118 \cdot 17.6971} \approx 0.9919$$

$V_A = 1.2500 \Rightarrow \sigma_A = 1.1180$, $V_P \approx 313.1875 \Rightarrow \sigma_P \approx 17.6971$ y $cov(P, A) = 19.6250$

$$V_r = (1 - r^2)V_P = 313.1875(1 - 0.9919^2) = 5.0750$$

NOTA: En la teoría se demuestra que podríamos haber calculado r y V_r usando la variable T en lugar de la A (más engorrosa), obteniendo los mismos valores de r y V_r, aunque la $cov(P, T)$ y $\sigma(T)$ sean ambas el doble de las que resultan usando A.

Problema 2:

a: $P(X_1 > 1000) = 0.5$, ya que $X_1 \to N(1000, 200)$ es la probabilidad de ser mayor que la media.

b: $P((X_2 > X_1) = P(X_2 - X_1) > 0)$ y $X_2 - X_1 \to N\left(1100 - 1000, \sqrt{200^2 + 200^2}\right) = N(100, 200\sqrt{2}) \Rightarrow P(z > \frac{0-100}{200\sqrt{2}}) = P(z > -0.3535534) = 1 - P(z > 0.3535534) \approx 0.63816$

c: $P(V_T > 3500)$ donde $V_T = X_1 + X_2 + X_3 \to N(3000, \sqrt{200^2 + 200^2 + 150^2} = N(3000, 320.1562)$.

Así, $P(V_T > 3500) = P(z > \frac{3500-3000}{320.1562}) = P(z > 1.5617) \approx 0.0592$.

d: El número total de bajas será $Y = Y_1 + Y_2 + Y_3$, haciendo uso de la propiedad de ser reproductiva de la Poisson, Y seguirá una Poisson de parámetro (media y desviación) $\lambda = 1 + 1 + 1 = 3$.

$$P(Y > 2) = P(Y = 3) + Y = 4) + \ldots = 1 - P(Y = 0) - P(Y = 1) - P(Y = 2) = 1 - e^{-3}(1 + \frac{3}{1} + \frac{3^2}{2}) \approx 0.57681$$

.

NOTA: Podrían haberse usado las tablas de la Poisson, para con $\lambda = 3$ calcular $P(Y = 0)$, $P(Y = 1)$ y $P(Y = 2)$.

e: El número de bajas en la 1ª empresa sigue $P(1)$, la probabilidad de que no haya bajas un día (hoy) es $P(Y_1 = 0) = e^{-1} \approx 0.36788$

La probabilidad pedida es que no haya bajas ni hoy ni mañana, es decir:

$P(\text{no bajas}) = P(\text{no bajas hoy})P(\text{no bajas mañana}) = (0.36788)^2 \approx 0.135335$

f: Se pide $P(\text{alguno mayor de } 1088) = 1 - P(\text{ninguno mayor de } 1088) = 1 - 0.67^{10} \approx 0.9818$, ya que:

$P(Y_1 > 1088) = P(z > \frac{1088-1000}{200}) = P(z > 0.44) = 0.33 \Rightarrow P(Y_1 \leq 1088) = 1 - 0.33 = 0.67$

Problema 3:

a: Debe verificarse que $1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^1 kxdx + \int_1^2 kdx = \left[k\frac{x^2}{2}\right]_0^1 + [kx]_1^2 = \frac{k}{2} + (2k - k) = \frac{3k}{2} = 1, \Rightarrow$ $k = \frac{2}{3}$.

La función de distribución la vamos a calcular en 4 zonas:

Si $x < 0$, $F(x) = P(\xi \leq x) = 0$ y si $x > 2$, $F(x) = P(\xi \leq x) = 1$ son fáciles.

Si $0 \leq x < 1, \Rightarrow F(x) = P(\xi \leq x) = \int_{-\infty}^{0} 0dx + \int_0^x \frac{2}{3}xdx = \frac{x^2}{3}$

Si $1 \leq x < 2, \Rightarrow F(x) = P(\xi \leq x) = \int_{-\infty}^{0} 0dx + \int_0^1 \frac{2}{3}xdx + \int_1^x \frac{2}{3}dx = 0 + \frac{1}{3} + \frac{2(x-1)}{3} = \frac{2x-1}{3}$ Así pues:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{3} & 0 \leq x \leq 1 \\ \frac{2x-1}{3} & 1 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

b: Buscamos el valor x, tal que $F(x) = 0.5$, vemos que no está en [0,1] pues $F(1) = \frac{1}{3}$ y F(x) es no decreciente. Tiene que estar en [1,2] pues $F(1) = \frac{1}{3}$ y $F(2) = 1$. Por tanto, resolvemos: $\frac{2x-1}{3} = 0.5 \Rightarrow 2x-1 = 1.5 \Rightarrow x = 1.25$ y por tanto la mediana vale $Me = 1.25$.

Interpretaciones de la mediana existen varias:

1) Si nos fijamos en la función de distribución, es su corte con la recta y=0.5, indicando que el 50% de la población es menor que ese valor y el otro 50% mayor.

2) Si nos fijamos en la función de densidad, indica que el área del intervalo [0,1] y el trocito del [1,1.25] vale 0.5, mientras que el área restante [1.25, 2] vale también 0.5.

Problema 4:

La relación 14:12:3:1 debe entenderse, como que es 14 veces más frecuente que un individuo sea del grupo "0", que del grupo "AB', 12 veces del "A", ...

Como la suma es 30 las frecuencias relativas son 14/30, 12/30 , 3/30 y 1/30.

Construimos la tabla:

Grupos	$"0"$	$"A"$	$"B"$	$"AB"$	$\sum$
O_i	126	117	42	15	300
p_i	$\frac{14}{30}$	$\frac{12}{30}$	$\frac{3}{30}$	$\frac{1}{30}$	1
$e_i = 300p_i$	140	120	30	10	300
$O_i - e_i$	-14	-3	12	5	
$(O_i - e_i)^2$	196	9	144	25	
$\frac{(O_i-e_i)^2}{e_i}$	1.40	0.075	4.800	2.500	8.775

El valor del estadístico $E = \sum_i \frac{(O_i-e_i)^2}{e_i} = 8.775$ se compara con el teórico al nivel $\alpha = 0.05$ en la χ^2 con 3 grados de libertad.

Número de clases=4 (todas con $e_i > 5$), Parámetros estimados al ajustar =0, g.d.l.=4-1-0=3.

Mirando en las tablas obtenemos $\chi^2_{3,0.05} = 7.815 < E$, por lo que se rechaza la hipótesis nula y se llega a que **la relación ha cambiado.**

Problema 5:

Se trata de un contraste de diferencias de medias, muestras pequeñas $N_I + N_M = 20 < 30$.

$\mathbf{H_0} : \mu_\mathbf{I} \leq \mu_\mathbf{M}$

$\mathbf{H_a} : \mu_\mathbf{I} > \mu_\mathbf{M}$

pero nos encontramos con el problema de no saber si debemos usar varianzas iguales o distintas, por lo que previamente debemos realizar un contraste de igualdad de varianzas:

$\mathbf{H_0} : \sigma_{\mathbf{I}}^{\mathbf{2}} = \sigma_{\mathbf{M}}^{\mathbf{2}}$
$\mathbf{H_0} : \sigma_{\mathbf{I}}^{\mathbf{2}} \neq \sigma_{\mathbf{M}}^{\mathbf{2}}$

La región crítica es: $\dfrac{s_I^2}{s_M^2} \notin \left[F_{1-\frac{\alpha}{2};n_I-1;n_M-1}, F_{\frac{\alpha}{2};n_I-1;n_M-1}\right] = [F_{0.975,10,8}, F_{0.025,10,8}] = [0.2594, 4.295]$

Ya que mirando en tablas de la F de Fisher-Snedecor, con $\alpha = 0.025$, $n_1 = 10$ y $n_2 = 8$, obtenemos: $F_{0.025;10,8} = 4.295$

Mientras que para $F_{0.975;10,8}$ debemos usar la propiedad de que $F_{0.975;10,8} = \frac{1}{F_{0.025;8;10}} = \frac{1}{3.855} = 0.2594$

$\bar{I} = \frac{143}{11} = 13$, $s_I^2 = (\frac{1889}{11} - 13^2) \cdot \frac{11}{10} = 3$, $\bar{M} = \frac{108}{9} = 12$, $s_M^2 = (\frac{1328}{9} - 12^2) \cdot \frac{9}{8} = 4$

Luego $\frac{s_I^2}{s_M^2} = \frac{3}{4} = 0.75 \in [0.2594, 4.295]$ y **aceptaremos la hipótesis nula de igualdad de varianzas**.

Por tanto, volviendo al contraste de diferencia de medias:

$\mathbf{H_0} : \mu_{\mathbf{I}} \leq \mu_{\mathbf{M}}$ $\qquad$ $\mathbf{H_a} : \mu_{\mathbf{I}} > \mu_{\mathbf{M}}$, vemos que la región crítica es:

$$\frac{\bar{x_I} - \bar{x_M}}{s_p\sqrt{\frac{1}{n_I} + \frac{1}{n_M}}} > t_{\alpha;n_I+n_M-2} = t_{0.05,18} = 1.734 \qquad \text{con:} \qquad s_p^2 = \frac{(n_I-1)s_I^2 + (n_M-1)s_M^2}{n_I + n_M - 2}$$

$s_p^2 = \frac{10*3+8*4}{11+9-2} = 3.4444 \Rightarrow sp = \sqrt{3.4444} = 1.8559 \Rightarrow E = \frac{13-12}{1.8559\sqrt{\frac{1}{11}+\frac{1}{9}}} = 1.1988 < t_{0.05,18} = 1.734$ y por tanto aceptamos la hipótesis nula y **no podemos concluir de que los alumnos de informática tengan móviles con más memoria que los de medicina**.

Problema 6: Las ecuaciones del ajuste tras el cambio de variable $x = cos(X)$ son:

$$\left.\begin{array}{lll} \sum_i n_i Y_i & = aN & +b\sum_i n_i x_i \\ \sum_i n_i Y_i & = a\sum_i n_i x_i & +b\sum_i n_i x_i^2 \end{array}\right\}$$

El programa queda:

```
format compact,clc
X=[0 0 pi/4 pi/4 pi/4 pi/2 pi/2 3*pi/4 3*pi/4 pi pi]
Y=[14 20 10 14 20 4 10 10 14 14 20], n=[50 75 71 51 12 36 41 49 44 25 18]
x=cos(X), N=sum(n)
A=[N sum(n.*x);sum(n.*x) sum(n.*x.^2)],B=[sum(n.*Y);sum(n.*Y.*x)]
sol=A\B,a=sol(1),b=sol(2)
yest=a+b*cos(X)
r=Y-yest
Vr=var(r,1),Vy=var(Y,1),R2=1-Vr/Vy
yy=[10 14 20], nn=[49 44+25 18]
NN=sum(nn),m1=sum(nn.*yy)/NN, m2=sum(nn.*yy.^2)/NN,  m3=sum(nn.*yy.^3)/NN
media=m1,varianza=m2-m1^2,mu3=m3-3*m2*m1+2*m1^3,s=sqrt(varianza),sesgo=mu3/(s^3)
```

Problema 7:

```
NIT=10000;
E=normrnd(1.5,0.2,NIT,1);
r=normrnd(0.3,0.01,NIT,1);
R=normrnd(1.2,0.02,NIT,1);
I=E./(R+r);
media=mean(I)
varianza=var(I,1)
c=(I>1.1);p=sum(c)/NIT,q=1-p;
Ip=[p-1.96*sqrt(p*q/NIT),p+1.96*sqrt(p*q/NIT)]
```

Problema 8:

Contrastamos si la varianza es 0.02^2, esto es: [$\mathbf{H_0} : \boldsymbol{\sigma^2 = 0.0004}$] contra [$\mathbf{H_a} : \boldsymbol{\sigma^2 \neq 0.0004}$] con "vartest".

Después contrataremos si la media es 1.2, [$\mathbf{H_0} : \boldsymbol{\mu = 1.2}$] contra [$\mathbf{H_a} : \boldsymbol{\mu \neq 1.2}$], con "ttest".

```
x=[1.23, 1.21, 1.17, 1.11, 1.3, 1.28, 1.19, 1.18, 1.31, 1.12, 1.19, 1.24, 1.22]
[Ha,Pa,Ia,STa]=vartest(x,0.0004,0.05,'both')
[Hb,Pb,Ib,STb]=ttest(x,1.2,0.05,'both')
% [Hb,Pb,Ib,STb]=ztest(x,1.2,0.0004,0.05,'both')
```

NOTA: El resultado del primer contraste podría afectar a como hacer el segundo. Si se aceptase que la varianza fuese 0.0004, debería haberse usado "ztest", en lugar de "ttest".

En el examen se dijo que se considerara que: "En el test del apartado 'a' resulta rechazada la hipótesis nula".