Grados en Informática

Métodos Estadísticos   Examen Junio 2018

- **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, 2 y 6 marcados con (E.D.).

APELLIDOS, NOMBRE:
DNI: Grupo: Titulación:

---

1. **(E.D.)** En un estudio de mercado, las encuestas han obtenido los datos siguientes sobre el porcentaje de compradores (C), en función del precio del producto (P en euros):

C	85	50	34	26	21	11	6
P	6	10	15	20	25	50	100

   a) Ajustar una función de la forma $C = m + \frac{n}{P}$.
   b) Comparar con el ajuste de $C = aP + bP^2$. *(0.5+1=1.5 Puntos)*

2. **(E.D.)** El IPC y los precios y cantidades producidas de 2 productos han evolucionado desde al año 2000 de la forma indicada en la tabla:

año	2000	...	2008	...	2016
Precio A	80	...	84	...	87
Precio B	120	...	140	...	170
Cantidad A	10	...	10	...	11
Cantidad B	38	...	40	...	45
IPC	3.25	...	3.49	...	3.70

   Determinar los **índices de precios** de Laspeyres y Paashe del año 2016, respecto al año base 2000, sin deflaccionar los precios y deflaccionándolos. *(0.75 Puntos)*

3. El diámetro (en mm.) de la variedad de frambuesa "negra"se distribuye según una variable aleatoria $N(27.8, 0.5)$, si ha sido cultivada en condiciones de suelo y humedad óptimos (corrigiendolo si es necesario), mientras que sigue una $N(27.1, 0.4)$ si se cultiva sin corregir el suelo. Estas son clasificadas como de calidad "extra" si tienen un diámetro superior a 27.5 mm., hallar:

   a) ¿Qué proporción de cada tipo será clasificada como "extra"?
   b) En una cooperativa, el 20 % del terreno está corregido y este produce un 25 % más de fruto por Ha. De las frambuesas de calidad "extra"que llega a la cooperativa, ¿qué proporción proviene de terreno corregido?
   c) Un agricultor que ha corregido el terreno por primera vez y desea hacer comprobaciones sobre los valores $\mu = 27.8$ y $\sigma = 0.5$. Para ello toma una muestra de 300 frambuesas de forma aleatoria y mide la suma de sus diametros $\sum_i d_i = 8310$ y de sus cuadrados $\sum_i d_i^2 = 230202.82$. Contrastar con $\alpha = 0.05$:
      1) Si el valor de la desviación típica está de acuerdo con las medidas realizadas.
      2) Si la media es menor de 27.8. *(0.75+0.5+(0.5+0.5)=2.25 Puntos)*

4. Una empresa de transporte tiene comprobado que la probabilidad de que se pinche una rueda determinada a cualquiera de sus camiones (8 ruedas) determinado día es 1/2000. Hallar:

   a) Probabilidad de que determinado día se pinche alguna rueda a determinado camión.
   b) Probabilidad de que a determinado camión se le pinche alguna rueda, por primera vez, tras haber transcurrido 30 dias sin pinchazos, pero antes del día 45º. ($31 \leq \xi < 45$)
   c) Probabilidad de que se le pinche más de 1 rueda un día determinado a alguno de los camiones de la flota de la empresa. La empresa cuenta con una flota de 150 camiones. *(0.5+0.5+0.5=1.5 Puntos)*

5. Se desea encontrar el recorrido idóneo para la recogida de basura de un barrio. Para ello se mide el tiempo empleado empleado cuando se realiza en dos secuencias diferentes A y B durante determinados días. Encontrándose (tiempo en minutos):

   Secuencia A: $t_A = \{114, 123, 94, 116, 112, 101, 107, 112, 135, 129\}$

   Secuencia B: $t_B = \{83, 154, 117, 104, 116, 102, 103, 129\}$

   ¿Puede concluirse, al 1 % de significación, que la recogida con la secuencia B, se realiza en menos tiempo que con la A?

   (1.5 Puntos)

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

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

   a) Ajustar una curva de la forma $Y = bX + cX^3$ 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)

7. **(MATLAB)**

   a) Si $\xi_1 \rightsquigarrow N(3, 0, 5)$. Halla la probabilidad de que $\xi_1 \in (3, 4)$, sabiendo que supera a la media.

   b) Si $\xi_2 \rightsquigarrow P(4.3)$ (Poisson con $\lambda = 4.3$). Hallar la probabilidad de que salga un valor entre 2 y 5. $(2 \leq \xi_2 \leq 5)$.

   c) Hallar mediante simulación con 100000 iteraciones la probabilidad de que la suma de ξ_1 y ξ_2 sea mayor que 9. $(P(\xi_1 + \xi_2 > 9))$.

   (0.25+0.25+0.25=0.75 Puntos)

8. **(MATLAB)** Un mecanismo está formado por 3 componentes A, B y C, que funcionan de forma independiente. Los tiempos hasta que se produce una avería en un componente concreto, siguen exponenciales de medias 10, 6 y 7, respectivamente. El mecanismo funciona (F) si lo hacen conjuntamente A y B, o bien, A y C. $(F = A \wedge (B \vee C))$

   a) Calcular mediante simulación con 10000 iteraciones la probabilidad de que funcione tras 8 años.

   b) La media y varianza del tiempo de funcionamiento del mecanismo.

   (0.5+0.5=1 Punto)

SOLUCIONES:
Problema 1:
Hacemos el cambio $X = \frac{1}{P} \Rightarrow C = m + nX$

C_i	P_i	x_i	x_i^2	$C_i x_i$	$Cest_i$	res_i	res_i^2	C_i^2
85	6	0.1667	0.0278	14.1667	84.4252	0.5748	0.3304	7225
50	10	0.1000	0.0100	5.0000	50.9695	−0.9695	0.9399	2500
34	15	0.0667	0.0044	2.2667	34.2416	−0.2416	0.0584	1156
26	20	0.0500	0.0025	1.3000	25.8777	0.1223	0.0150	676
21	25	0.0400	0.0016	0.8400	20.8593	0.1407	0.0198	441
11	50	0.0200	0.0004	0.2200	10.8226	0.1774	0.0315	121
6	100	0.0100	0.0001	0.0600	5.8042	0.1958	0.0383	36
233		0.4533	0.0468	23.8533		0.0000	1.4332	12155

El sistema de ecuaciones normales queda:

$$\left.\begin{array}{rcll} 233 & = & 7m & +0.4533n \\ 23.8533 & = & 0.4533m & +0.0468n \end{array}\right\} \Rightarrow m = 0.7859,\ m = 0.7859,\ \ n = 501.8360 \Rightarrow \mathbf{C = 0.7859 + \frac{501.836}{P}}$$

La varianza de C es: $V_C = \frac{12155}{7} - \left(\frac{233}{7}\right)^2 \approx 628.4898$
La varianza residual vale: $V_r = \frac{1.4332}{7} - 0^2 \approx 0.2047$
Y por tanto: $R^2 = 1 - \frac{0.2047}{628.4898} \approx 0.9997$ indicando que es un muy buen ajuste.
1-b: Ahora se trata de representar el vector C en la base $\{P, P^2\}$, las ecuaciones quedan:

$$\left.\begin{array}{l} \sum C_i P_i = a \sum P_i^2 + b \sum P_i^3 \\ \sum C_i P_i^2 = a \sum P_i^3 + b \sum P_i^4 \end{array}\right\} \Rightarrow \left.\begin{array}{c} 3715 = 13886m + 1153216n \\ 126735 = 1153216m + 106862546n \end{array}\right\} \Rightarrow \begin{array}{c} a = 1.6290 \\ b = -0.0164 \end{array} \Rightarrow \mathbf{C = 1.6290P - 0.0164P^2}$$

C_i	P_i	P_i^2	P_i^3	P_i^4	$C_i P_i$	$C_i P_i^2$	$Cest_i$	res_i	res_i^2
85	6	36	216	1296	510	3060	9.1838	75.8162	5748.1
50	10	100	1000	10000	500	5000	14.6505	35.3495	1249.6
34	15	225	3375	50625	510	7650	20.7463	13.2537	175.7
26	20	400	8000	160000	520	10400	26.0224	−0.0224	0.0
21	25	625	15625	390625	525	13125	30.4788	−9.4788	89.8
11	50	2500	125000	6250000	550	27500	40.4659	−29.4659	868.2
6	100	10000	1000000	100000000	600	60000	−1.0352	7.0352	49.5
		13886	1153216	106862546	3715	126735		92.4875	8180.9

No nos indican el parámetro que debemos usar para comparar las bondades de los ajustes, así podemos elegir entre la varianza residual, coeficiente de determinación y SSE.

El más fácil de calcular es $SSE = \sum_i e_i^2$ (suma de los cuadrados de los errores o residuos). Para el primer caso, SSE=1.4332, mientras que para la parábola: SSE=8180.9, indicando que es un ajuste mucho peor.

A la misma conclusión se llega si comparamos las varianzas residuales: $V_{r1} = 0.2047$, mientras que $V_{r2} = 994.134$ (es mejor el primer ajuste por tener varianza residual menor).

Problema 2
Sin deflaccionar:
$L_{2016} = \frac{10(87)+38(170)}{10(80)+38(120)} = 1.3675$
$P_{2016} = \frac{11(87)+45(170)}{11(80)+45(120)} = 1.3705$
Deflaccionando los precios:
$P(A)'_{2016} = P(A)_{2016}\frac{3.25}{3.70} = 76.4189$ y $P(B)'_{2016} = P(B)_{2016}\frac{3.25}{3.70} = 149.3243$
$L'_{2016} = \frac{10(76.4189)+38(149.3243)}{10(80)+38(120)} = 1.2012$
$P'_{2016} = \frac{11(76.4189)+45(149.3243)}{11(80)+45(120)} = 1.2039$
NOTA: Resulta que los deflaccionados, son los sin deflaccionar multiplicados por: $k = \frac{3.25}{3.70} = 0.8784$

Problema 3:
3-a: Llamemos respectivamente ξ_1 y ξ_2 a las variables aleatorias que siguen el diámetro de las frambuesas según el terreno esté o no corregido. $\xi_1 \rightsquigarrow N(27.8, 0.5)$ y $\xi_2 \rightsquigarrow N(27.1, 0.4)$. (A=Terreno corregido, B=Terreno no corregido, E=extra.)
P(E/A) $=P(\xi_1 > 27.5) = P\left(\frac{27.5-27.8}{0.5}\right) = P(z > -0.6) = 1 - P(z > 0.6) = \mathbf{0.7257}$
P(E/B) $=P(\xi_2 > 27.5) = P\left(\frac{27.5-27.1}{0.4}\right) = P(z > 1) = \mathbf{0.1587}$
3-b: Si no produjese más el terreno acondicionado, las cajas estarían en las proporciones 20 y 80, pero al producir un 25 % más, el 20 % del terreno producirá como el 20(1.25)=25 %. Así pues, de cada 105 unidades de producto 25/105 serán de terreno acondicionado y 80/105 del no acondicionado.

Lo que nos piden es $\mathbf{P(A/E)} = \frac{P(E\wedge B)}{P(E)} = \frac{\frac{25}{105}(0.7257)}{0.2937} = \mathbf{0.5884}$

Pues por el teorema de la probabilidad total: $P(E) = \frac{25}{105}(0.7257) + \frac{80}{105}(0.1587) = 0.2937$. (La proporción de "extra"será del 29.37 %.)

3-c1: Será un contraste de la varianza: $H_0 : \sigma^2 = 0.5^2 = 0.25$ y $H_a : \sigma^2 \neq 0.25$, $\alpha = 0.05$ y $n = 300$.

La tablas proporcionan: $E = \frac{(n-1)s^2}{\sigma^2} = \frac{299(0.0529)}{0.5^2} = 63.28$

Pues $\bar{x} = \frac{8310}{300} = 27.7$ y $s^2 = \frac{300}{299}V = \frac{300}{299}\left(\frac{230202.82}{300} - \left(\frac{8310}{300}\right)^2\right) \approx 0.0529 \Rightarrow s \approx 0.23$.

Mientras que el intervalo de la región no crítica es (región de aceptación de H_0): $I = \left[\chi^2_{1-\frac{\alpha}{2},n-1}, \chi^2_{\frac{\alpha}{2},n-1}\right] = \left[\chi^2_{0.975,299}, \chi^2_{0.025,299}\right] \approx$

$\approx \left[\frac{(-1.96+\sqrt{597})^2}{2}, \frac{(1.96+\sqrt{597})^2}{2}\right] = [252.5310, 348.3106]$.

Por tanto, $E \notin I$ y rechazamos la hipótesis nula, aceptando que la varianza no es 0.5^2

3-c2: Ahora se trata de un contraste unilateral de la media: $H_0 : \mu \geq 27.8$ contra $H_a : \mu < 27.8$. Usamos varianza desconocida (pues en el contraste anterior salió rechazado $\sigma = 0.5$) y muestra grande.

En las tablas se observa que el estadístico de contraste es $E = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{27.7-27.8}{\frac{0.23}{\sqrt{300}}} = -7.53 < -1.645 = -z_{0.05}$. Por tanto, rechazamos la hipótesis nula y se concluye que el diámetro medio es menor de 27.8.

Problema 4:

4-a: El número de ruedas pinchadas por camión ψ sigue una Binomial de parámetros $n = 8$ y $p = 1/2000 = 0.0005$.

La probabilidad de que se pinche alguna rueda es: $P(\text{alguna}) = 1 - P(\text{ninguna}) = 1 - P(\psi = 0) = 1 - 0.9995^8 \approx 0.003993$

4-b: El tiempo ξ hasta que determinado camión pincha más de una rueda en el mismo día, seguirá una Geométrica o de Pascal con $p = 0.003993$, $q = 1 - p$ por tanto la probabilidad pedida es $P = \sum_{k=31}^{44} p(\xi = k) = q^{30}p + q^{31}p + q^{32}p + \ldots + q^{43}p = \frac{q^{44}p - q^{30}p}{p-1} \approx 1.93686(10)^{-4}$

4-c: La probabilidad de que se le pinche más de 1 rueda a un camión determinado es:

$P(\psi > 1) = P(\psi = 2) + P(\psi = 3) + \ldots = 1 - P(\psi = 0) - P(\psi = 1) = 1 - 0.9995^8 - 8(0.0005)(0.9995)^7 \approx 6.98(10)^{-6}$

Ahora el número de camiones al que se le pincha más de 1 rueda seguirá una Binomial ϕ de parámetros $n = 150$, $p = 6.98(10)^{-6}$, que se puede aproximar por una distribución de Poisson de $\lambda = 150(6.98)(10^{-6} = 0.001047$, así que:

$P(\text{alguno}) = 1 - P(\phi = 0) = 1 - e^{-0.001047} \approx 0.001046452086738$

NOTA: Si lo hubiésemos calculado directamente con la Binomial $P(\text{alguno}) = 1 - (0.99999302)^{150} \approx 0.001046455736960$

Problema 5

Se trata de un contraste unilateral de diferencias de medias: $H_0 : \mu_A \leq \mu_B$ contra $H_a : \mu_A > \mu_B$ con muestras pequeñas y varianzas desconocidas, por lo que debemos diferenciar si son iguales o diferentes previamente.

$H_0 : \sigma_A^2 = \sigma_B^2$ contra $H_a : \sigma_A^2 \neq \sigma_B^2$

El estadístico de contraste es: $E = \frac{s_A^2}{s_B^2} = \frac{152.9}{451.71} \approx 0.33849$, mientras que el intervalo es:

$I = [F_{1-\frac{\alpha}{2},n_1-1,n_2-1}, F_{\frac{\alpha}{2},n_1-1,n_2-1}] = [F_{0.995,9,7}, F_{0.005,9,7}] = [0.1452, 8.5138]$, pues:

$F_{0.995,9,7} = \frac{1}{F_{0.005,7,9}} = \frac{1}{6.885} = 0.1452$.

Como E pertenece al intervalo nos quedaremos con la hipótesis nula y aceptamos igualdad de varianzas.

Continuamos con el contraste de la diferencia de medias usando varianzas desconocidas pero iguales, muestra pequeña, realizando cálculos:

$\bar{x}_A = 114.3$, $\bar{x}_B = 113.5$, $s_A^2 = 152.9$, $s_B^2 = 451.71$, $s_A = 12.365$, $s_B = 21.254$

El estadístico de contraste vale: $E = \frac{\bar{x}_A - \bar{x}_B}{s_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} = \frac{114.3 - 113.5}{16.6313\sqrt{\frac{1}{10} + \frac{1}{8}}} = 0.1001$, pues:

$s_p^2 = \frac{9s_A^2 + 7s_B^2}{16} = 283.6313 \Rightarrow s_p = 16.8414$

Se compara con el valor de una T-Student: $t_{\alpha,n_A+n_B-2} = t_{0.01,16} = 2.583$ Como $E = 0.1001$ no es mayor que 2.583, aceptamos la hipótesis nula, no concluyéndose que el tiempo medio del recorrido de B sea menor.

Problema 6 (MATLAB y ED)

```
clc,clear all,format compact
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=[sum(n.*x.^2) sum(n.*x.^4);sum(n.*x.^4) sum(n.*x.^6)]
B=[sum(n.*y.*x);sum(n.*y.*x.^3)]
sol=A\B
```

```
b=sol(1), c=sol(2)
disp('Estos son los coeficientes')
yest=b*x+c*x.^3
res=y-yest
mres=sum(n.*res)/N, m2res=sum(n.*res.^2)/N
Vres=m2res-mres^2
my=sum(n.*y)/N, m2y=sum(n.*y.^2)/N
Vy=m2y-my^2
R2=1-Vres/Vy
```

Problema 7: (MATLAB)

7-a: Nos piden: $P(\xi_1 \in (3,4)/\xi_1 > 3) = \frac{P((\xi_1 \in (3,4)) \wedge (\xi_1 > 3))}{P(\xi_1 > 3)} = \frac{F(4)-F(3)}{1-F(3)}$, por tanto:

```
Pa=(normcdf(4,3,0.5)-normcdf(3,3,0.5))(1-normcdf(3,3,0.5))
disp('Tambien:')
Pa=(normcdf(4,3,0.5)-0.5)/0.5
```

7-b:

```
Pb=sum(poisspdf(2:5,4.3))
```

7-c:

```
nit=100000
xi1=normrnd(3,0.5,1,nit);
xi2=poissrnd(4.3,1,nit);
c=(xi1+xi2>9);
p=sum(c)/nit,q=1-p;Ip=[p-1.96*sqrt(p*q/nit),p+1.96*sqrt(p*q/nit)]
```

Problema 8: (MATLAB)

```
clc,clear all
disp('8-a')
nit=10000
A=exprnd(10,1,nit);
B=exprnd(6,1,nit);
C=exprnd(7,1,nit);
cA=(A>8);cB=(B>8);cC=(c>8);
c=(cA&cB)|(cA&cC);
p8=sum(c)/nit,q8=1-p8;
Ip8=[p8-1.96*sqrt(p8*q8/nit),p8+1.96*sqrt(p8*q8/nit)]
disp('8-b')
TAB=min([A;B]);
TAC=min([A;C]);
T=max([TAB;TAC]);
m=mean(T)
V=var(T)
```