Grados en Ingeniería Informática

Métodos Estadísticos para la Computación
Examen 23 Septiembre 2013

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

APELLIDOS, NOMBRE:
DNI: Grupo: Titulación:

---

1. La energía (E), en julios, necesaria para regresar un brazo robot a su posición de reposo depende del ángulo x en el que se encuentre (en rad.). Se quiere aproximar una función de la forma $E = a\text{sen}(x) + b\cos(x)$. Para ello se determinan unos puntos que son expresados en la tabla:

x	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π
E	0	1	2	3	0

   Se pide:

   (a) Realizar el ajuste.
   (b) Hallar el coeficiente de determinación.
   (c) Hallar el coeficiente de correlación lineal. Comparar ambos ajustes.

   (1+0.5+0.5=2 Puntos)

2. Se sabe que el tiempo que tarda un frigorífico hasta que se produce su primera avería sigue una distribución exponencial.

   Un servicio técnico observa que el 23.87% de los electrodomésticos hacen uso de la garantía en los 36 primeros meses.

   (a) Hallar la media y el parámetro λ de la exponencial.
   (b) El comercio ofrece la posibilidad de contratar por 15€, una garantía adicional hasta los 72 meses (36 meses adicionales). Si sabemos que el precio medio de una reparación cuesta 100€, ¿Qué resulta más conveniente, contratar la garantía adicional o no contratarla? Justificar la respuesta.

   (0.75+0.75=1.5 Puntos)

3. En una campaña agrícola, cada platanera produce una piña de plátanos excepto si es atacada por el insecto COMEPLAT. Sabemos que número de plataneras atacadas por el insecto COMEPLAT en una plantación A, sigue una distribución normal de media 55 y varianza 54.879, mientras que en otra plantación B, también sigue una normal, pero de media 30 y varianza 29.982, calcular:

   (a) Probabilidad de que sean atacadas más de 50 plataneras en la plantación A y más de 50 en la B. Es decir, se pide que **ocurran ambos hechos en la misma campaña**.
   (b) Probabilidad de que sean atacadas más de 100 plataneras, **contando ambas plantaciones**.
   (c) Probabilidad de que sean atacadas más plataneras en la plantación B que en la A.
   (d) Probabilidad de que si han sido atacadas exactamente 100 plataneras en total, hayan sido atacadas exactamente 60 en la plantación A.

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

4. Se quiere analizar la preferencia alimenticia de una especie de primates. Uno de los experimentos consiste en ofrecerles una banana (B) o un mango (M), pudiendo solo tomar uno de ellos. Se hacen 150 experimentos de ese tipo. Determinar al 95% de confianza:

   (a) Si obtenemos nB=90 (nB="número de veces que opta por la banana"), ¿podemos concluir que la banana es preferida al mango?.
   (b) Si obtenemos nB=90, ¿podemos rechazar que ambas frutas son igualmente preferidas?.
   (c) ¿Qué número mínimo de observaciones debe realizarse (tamaño de la muestra) para que una proporción observada $\hat{p}(B) = 0.51$, nos indique (resulte significativa) que la banana es preferida al mango.

   (0.75+0.5+0.75=2 Puntos)

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

5. **Indicar las órdenes necesarias en MATLAB para: (Sin realizar cálculos)** Una tienda desea como le afecta a su tienda la llegada de cruceros al puerto. Así anota de forma diferenciada la facturación en 10^3€, realizada durante los días en que existe un crucero en el puerto y en los que no existe crucero, obteniendo los resultados:

Con crucero	1.34	1.72	1.09	2.11	1.45	1.88	1.71	1.67	1.44	1.97	–
Sin crucero	1.12	1.45	1.98	1.87	1.23	1.45	1.38	1.45	1.67	1.49	1.70

Contrastar al 98% de confianza:

(a) Que las diferencias observadas no son significativas, es decir, la llegada de cruceros no afecta a la facturación.

(b) Que la facturación los días con crucero es mayor que en los que no los hay.

(c) Que la varianza de la facturación los días con crucero es mayor que los sin crucero.

(0.5+0.5+0.5=1.5 Puntos)

6. **Indicar las órdenes necesarias en MATLAB para: (Sin realizar cálculos ni aproximaciones).**

Una compañía considera que en una provincia tiene 62000 posibles clientes. Como campaña de introducción ofrece un servicio gratuito durante los 6 primeros meses a 2000 de ellos, estimando que el 12% de ellos continuarán recibiéndolo una vez pasado el periodo gratuito. Para los 60000 restantes, estima que la probabilidad de que lo contrate es del 0.3%.

Estimar mediante el método de Montecarlo, con 10.000 iteraciones.

(a) La media y varianza del número de clientes que contratan el servicio (ϕ):

$$\phi = \phi_1 + \phi_2 \rightarrow B(2000, 0.12) + B(60000, 0.003)$$

(b) Si la empresa necesita, al menos, 300 clientes para ser viable. Dar un intervalo de confianza al 95% para la probabilidad de que ϕ sea mayor o igual a 300. ($P(\phi \geq 300)$)

(0.5+0.5=1 Punto)

SOLUCIONES:

Problema 1:

a: Queremos expresar E en función de la base $\mathcal{B} = \{\text{sen(x)}, \cos(\text{x})\}$, por lo que las ecuaciones normales son:

$$\left.\begin{array}{ll} \langle E, \text{senx}\rangle & = a\,\langle \text{sen(x)}, \text{sen(x)}\rangle \quad +b\,\langle \cos(x), \text{sen(x)}\rangle \\ \langle E, \cos(x)\rangle & = a\,\langle \text{sen(x)}, \cos(\text{x})\rangle \quad +b\,\langle \cos(x), \cos(x)\rangle \end{array}\right\} \Rightarrow$$

$$\left.\begin{array}{lll} \sum_i E_i\text{sen}(\text{x}_\text{i}) & = a\sum_i \text{sen}^2(\text{x}_\text{i}) & +b\sum_i \text{sen}(\text{x}_\text{i})\cos(\text{x}_\text{i}) \\ \sum_i E_i\cos(x_i) & = a\sum_i \text{sen}(\text{x}_\text{i})\cos(\text{x}_\text{i}) & +b\sum_i \cos^2(x_i) \end{array}\right\} \Rightarrow \left.\begin{array}{ll} 4.8284 & = 2a + 0b \\ -1.4142 & = 0a + 3b \end{array}\right\} \Rightarrow \begin{array}{c} \mathbf{a = 2.4142} \\ \mathbf{b = -0.4714} \end{array}$$

Luego el ajuste pedido es $\mathbf{E = 2.4142\text{sen}(x) - 0.4714\cos(x)}$.

x_i	0	$\frac{\pi}{4}$	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	π	$\frac{5\pi}{2}$
E_i	0	1	2	3	0	6
$\text{sen}(\text{x}_\text{i})$	0	$\frac{\sqrt{2}}{2}$	1	$\frac{\sqrt{2}}{2}$	0	
$\cos(x_i)$	1	$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$	-1	
$E_i\text{sen}(\text{x}_\text{i})$	0	$\frac{\sqrt{2}}{2}$	2	$\frac{3\sqrt{2}}{2}$	0	$2+2\sqrt{2} \approx 4.8284$
$E_i\cos(x_i)$	0	$\frac{\sqrt{2}}{2}$	0	$-\frac{3\sqrt{2}}{2}$	0	$-\sqrt{2} \approx -1.4142$
$\text{sen}(\text{x}_\text{i})\cos(\text{x}_\text{i})$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	0
$\text{sen}^2(\text{x}_\text{i})$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	0	2
$\cos^2(x_i)$	1	$\frac{1}{2}$	0	$\frac{1}{2}$	1	3
E_i^*	-0.4714	1.3738	2.4142	2.0404	0.4714	
$r_i = E_i - E_i^*$	0.4714	-0.3738	-0.4142	0.9596	-0.4714	0.1716
r_i^2	0.2222	0.1397	0.1716	0.9208	0.2222	1.6765
E_i^2	0	1	4	9	0	14
x_i^2	0	0.6169	2.4674	5.5517	9.8696	18.5055
$E_i x_i$	0	$\frac{\pi}{4}$	π	$\frac{9\pi}{4}$	0	$\frac{7\pi}{2} \approx 10.9956$

b:

$$Vr = \sigma_r^2 = \frac{\sum_i r_i^2}{N} - \left(\frac{\sum_i r_i}{N}\right)^2 = \frac{1.6765}{5} - \left(\frac{0.1716}{5}\right)^2 \approx \mathbf{0.3341}$$

$$V_E = \sigma_E^2 = \frac{\sum_i E_i^2}{N} - \left(\frac{\sum_i E_i}{N}\right)^2 = \frac{14}{5} - \left(\frac{6}{5}\right)^2 \approx \mathbf{1.36}$$

$$\mathbf{R^2 = 1 - \frac{V_r}{V_E} = 1 - \frac{0.3341}{1.36} \approx 0.7543}$$

c:

$$\bar{x} = \frac{5\pi}{10} \approx 1.5708, \qquad \bar{y} = \frac{6}{5} = 1.2$$

$$V_x = \frac{\sum_i x_i^2}{N} - \left(\frac{\sum_i x_i}{N}\right)^2 = \frac{18.5055}{5} - \left(\frac{5\pi}{10}\right)^2 \approx \mathbf{1.2337}$$

$$cov = \frac{\sum_i x_i E_i}{N} - \bar{x}\bar{y} = \frac{10.9956}{5} - (1.5708)(1.2) \approx \mathbf{0.3142}$$

$$r = \frac{cov}{\sigma_x \sigma_E} = \frac{0.3142}{\sqrt{(1.2337)(1.36)}} \approx \mathbf{0.2425} \Rightarrow \mathbf{r^2} \approx \mathbf{0.0588}$$

Resulta mucho mejor el ajuste trigonométrico que el lineal, pues explica el 75.43% de las variaciones observadas, mientras que el lineal solo explica el 5.88%.

Problema 2:

a:

Nos dicen que $P(x \leq 36) = 0.2387 \Rightarrow F(36) = 0.2387 = 1 - e^{-\lambda 36} \Rightarrow e^{-36\lambda} = 0.7613 \Rightarrow$

$$\lambda = \frac{-\text{Ln}(0.7613)}{36} \approx \mathbf{7.57577\,10^{-3}} \qquad \Rightarrow \qquad \mu = \frac{1}{\lambda} = \mathbf{132}$$

b: La probabilidad de tener alguna avería en el periodo de ampliación de la cobertura es:
$P(36 < x \leq 72) = F(72) - F(36) = \left[1 - e^{-72\lambda}\right] - 0.2387 \approx 0.1817$
El costo esperado para una avería en ese periodo será al menos de 100*0.1817=18.17€. Por tanto, **pagar 15€ y ampliar la cobertura es la mejor opción.**

Problema 3:
a: Nos dicen que: $A \to N(55, \sqrt{55.879}) = N(55, 7.4752)$, mientras que $B \to N(30, \sqrt{29.982}) = N(55, 5.4756)$.
Y nos piden $P((A > 50) \cap (B > 50))$.

$$P((A > 50) \cap (B > 50)) = P(A > 50)P(B > 50) = P\left(z > \frac{50-55}{7.4752}\right) P\left(z > \frac{50-30}{5.4756}\right) =$$

$$= P(z > -0.6688)P(z > 3.653) \approx 0.748(0.00013) \approx \mathbf{0.000097}$$

b: Ahora nos piden $P(A + B > 100)$.
Sabemos que por ser normales independientes A y B, $A+B \to N(55+30, \sqrt{54.879 + 29.982}) = N(85, 9.212)$
$P(A + B > 100) = P\left(z > \frac{100-85}{9.212}\right) = P(z > 1.6283) \approx \mathbf{0.0517}$

c: Ahora nos piden $P(B - A > 0)$ y sabemos que por ser normales independientes:
$B - A \to N(\mu_B - \mu_A, \sqrt{s_B^2 + s_A^2}) = N(-15, 9.212)$
$P(B - A > 0) = P\left(z > \frac{0-(-15)}{9.212}\right) = P(z > 1.6283) \approx \mathbf{0.0517}$

d: Nos piden $P(A = 60/(A + B = 100))$.
$P(A = 60/(A + B = 100)) = \frac{P((A=60)\cap(A+B=100))}{P(A+B=100)} = \frac{P((A=60)\cap(B=40))}{P(A+B=100)} = \frac{P(A=60)P(B=40)}{P(A+B=100)} =$ (por la corrección de continuidad) $= \frac{P(59.5<A<60.5)P(39.5<B<40.5))}{P(99.5<A+B<100.5)} \approx \frac{(0.0429)(0.0138)}{0.0115} = \mathbf{0.0514}$
$P(59.5 < A < 60.5) = P\left(\frac{59.5-50}{7.4752} < z < \frac{60.5-50}{7.4752}\right) = P(0.6074 < z < 0.7424) \approx 0.0429$
Ya que:
$P(39.5 < A < 40.5) = P\left(\frac{39.5-30}{5.4756} < z < \frac{40.5-30}{5.4756}\right) = P(1.7350 < z < 1.9176) \approx 0.0138$
$P(99.5 < A + B < 100.5) = P\left(\frac{99.5-85}{9.212} < z < \frac{100.5-85}{9.212}\right) = P(1.5740 < z < 1.6826) \approx 0.0115$

Problema 4:
a:
Nos dan los datos $n = 150$, $nB = 90$ y $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$.
Y nos piden que contrastemos si $p(B) > 0.5$, por tanto se trata de un contraste unilateral de la proporción.
$H_0 : p(B) \leq 0.5$ contra $H_a : p(B) > 0.5$.
En nuestro caso $\hat{p} = \frac{90}{150} = 0.6$
La región crítica es:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} > z_\alpha, \text{ que sustituyendo: } \frac{0.6 - 0.5}{\sqrt{\frac{0.6(0.4)}{150}}} = \frac{0.1}{0.04} = 2.5$$

que en efecto es mayor que $z_{0.05} = 1.645$ y se rechaza la hipótesis nula, concluyendo que $p(B) > 0.5$ y, por tanto, **prefiere la banana.**

b: Aunque parece lo mismo es un contraste diferente, pues ahora es de dos colas.
$H_0 : p(B) = 0.5$ contra $H_a : p(B) \neq 0.5$.
La región crítica es:

$$\frac{|\hat{p} - p_0|}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} > z_{\frac{\alpha}{2}}, \text{ que sustituyendo: } \frac{|0.6 - 0.5|}{\sqrt{\frac{0.6(0.4)}{150}}} = \frac{0.1}{0.04} = 2.5$$

que en efecto es mayor que $z_{0.025} = 1.645$ y se rechaza la hipótesis nula, concluyendo que $p(B) > 0.5$ y, por tanto, **no son igualmente preferidas.** Se podría redactar lo mismo diciendo que **la elección no es aleatoria.**

c: Se trata del mismo contraste que en el apartado (a), pero ahora se busca n para que para resulte significativo $\hat{p} = 0.51$, es decir, el valor caiga en la región crítica. Se debe verificar:

$$\frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{0.51 - 0.5}{\sqrt{\frac{0.51(0.49)}{n}}} > z_\alpha = 1.645 \Rightarrow \left(\frac{0.01}{1.645}\right)^2 > \frac{0.51(0.49)}{n} \Rightarrow n > \frac{0.2499}{3.6955\,10^{-5}} \approx 28164.75$$

Luego necesitaríamos una muestra de tamaño igual o superior a 28165.

Problema 5 (MATLAB):
Comentarios: Al 98% de confianza, sigifica que $\alpha = 0.02$.
El apartado (a) se trata de un contraste de diferencias de medias de 2 colas (por defecto en MATLAB).
El apartado (b) es un contraste de diferencias de medias de 1 cola (media(C)¿media(S)).
El apartado (c) es un contraste de cociente de varianzas de 1 cola (var(C)¿var(S), o bien, var(C)/var(S)¿1).

```
format compact,clc
disp('Apartado a')
C=[1.34 1.72 1.09 2.11 1.45 1.88 1.71 1.67 1.44 1.97]
S=[1.12 1.45 1.98 1.87 1.23 1.45 1.38 1.45 1.67 1.49 1.70]
alfa=0.02
[Ha,Pa]=ttest2(C,S,alfa)
disp('Apartado b')
[Hb,Pb]=ttest2(C,S,alfa,'right')
disp('Apartado c')
[Hc,Pc]=vartest2(C,S,alfa,'right')
```

Problema 5:

```
clc,format compact
disp('Apartado a')
NIT=10000;
fi1=binornd(2000,0.12,NIT,1);
fi2=binornd(60000,0.003,NIT,1);
fi=fi1+fi2;
media=sum(fi)/NIT
varianza=sum(fi.^2)/NIT-media^2
% La media y varianza podrian calcularse con
% media=mean(fi), varianza=var(fi,1)
disp('Apartado b')
C=(fi>=300); % Condicion: En cada fila C=1 si fi>=300, en otro caso vale 0
p=sum(C)/NIT
I=[p-1.96*sqrt(p*(1-p)/NIT),p-1.96*sqrt(p*(1-p)/NIT)]
```