BIOESTADÍSTICA
EXAMEN 18 DE AGOSTO DE 2006

DATOS DEL ESTUDIANTE

Nombre	Cédula

- **La duración del examen es 3 horas.**
- **El puntaje mínimo para aprobar es 50 puntos.**

Problema 1 (38 puntos)

Una caja contiene lamparitas: Una proporción p de las lamparitas son de la marca A y una proporción $(1-p)$ de la marca B. Se sabe que la duración, medida en horas, de una lamparita de la marca A tiene distribución exponencial de parámetro $\lambda_A = 0,002$ y que la duración, medida en horas, de una lamparita de la marca B tiene distribución exponencial de parámetro $\lambda_B = 0,001$.

a) **(6 puntos)** Calcule la probabilidad de que una lamparita de la marca A tenga una duración comprendida entre 100 y 200 horas.

b) **(6 puntos)** Calcule la probabilidad de que la duración de una lamparita de la marca A sea superior a 200 horas, y la probabilidad de que la duración de una lamparita de la marca B sea superior a 200 horas.

c) **(12 puntos)** Suponga que se extrae al azar una lamparita de la caja. Calcule, en función de p, la probabilidad de que su duración sea superior a 200 horas.

d) **(14 puntos)** Estime el parámetro p, sabiendo que en una muestra de diez lamparitas extraídas de la caja, siete de esas lamparitas tuvieron una duración superior a 200 horas.

Problema 2 (24 puntos)

Los siguientes datos corresponden a los niveles de calcio en la sangre (medidos en $mmol/l$) para un grupo de 10 individuos:

4,23	12,55	5,59	6,34	9,20	6,18	5,71	3,50	6,88	1,99

a) **(12 puntos)** Asumiendo que los datos tienen distribución normal con media μ y varianza desconocida, construya un intervalo de confianza exacto al nivel 5 % para μ.

b) **(12 puntos)** Suponiendo ahora que los niveles de calcio en la sangre tienen distribución normal con media μ y desviación estándar $\sigma = 3$, calcule el mínimo tamaño que debería tener una muestra para que el intervalo de confianza exacto al nivel 5 % para μ tenga longitud 2.

Problema 3 (38 puntos)

La siguiente muestra (X) registra los volúmenes de aire expulsado en un segundo (medido en *litros*), correspondientes a doce pacientes con enfermedades coronarias:

3,167	1,935	3,725	3,887	2,454	4,790	4,789	3,562	3,927	3,774	3,413	4,325

Nota: En las pruebas de hipótesis utilice el siguiente criterio de decisión: se acepta la hipótesis nula si el p-valor es superior a $0,05$.

a) **(10 puntos)** Realice dos pruebas de hipótesis para estudiar si es razonable suponer que los datos son independientes e idénticamente distribuidos.

b) **(12 puntos)** Asumiendo que los datos tienen distribución normal con media μ y desviación estándar σ (que suponemos desconocida), implemente la siguiente prueba de hipótesis:

$$\begin{cases} H_0: & \mu = 3,5 \\ H_1: & \mu > 3,5. \end{cases}$$

Considere ahora una nueva muestra (Y), que también registra los volúmenes de aire expulsado en un segundo (medido en *litros*), pero que corresponde a doce pacientes que no tienen enfermedades coronarias:

3,475	4,801	4,556	3,783	2,653	3,202	4,811	3,318	3,780	3,915	4,920	3,890

c) **(16 puntos)** Asumiendo que la muestra Y es i.i.d. y que además es independiente de la muestra X, implemente la prueba de comparación de Kolmogorov-Smirnov para decidir si las muestras X e Y tienen la misma distribución.

SOLUCIÓN

Problema 1

En lo que sigue, X es una variable exponencial de parámetro $\lambda_A = 0,002$ e Y es una variable exponencial de parámetro $\lambda_A = 0,001$.

a) $\mathbf{P}(100 \leq X \leq 200) = \mathbf{P}(X \geq 100) - \mathbf{P}(X \geq 200) = e^{-\lambda_A 100} - e^{-\lambda_A 200} = 0,1484.$

b) $\mathbf{P}(X \geq 200) = e^{-\lambda_A 200} = 0,6703$ y $\mathbf{P}(Y \geq 200) = e^{-\lambda_B 200} = 0,8187.$

c) En notación obvia:

$$\begin{aligned}\mathbf{P}(\text{duración superior a } 200 \text{ horas}) &= \mathbf{P}(\text{duración superior a } 200 \text{ horas}|\text{ Marca } A)\mathbf{P}(\text{ Marca } A)\\ &+ \mathbf{P}(\text{duración superior a } 200 \text{ horas}|\text{ Marca } B)\mathbf{P}(\text{ Marca } B)\\ &= 0,6703\,p + 0,8187\,(1-p).\end{aligned}$$

d) A partir de la muestra podemos estimar $\mathbf{P}(\text{duración superior a } 200 \text{ horas}) \simeq 0,7$. Con esto, usando el resultado de la parte anterior, estimamos p por el valor $\hat{p}$ que cumple:

$$0,7 = 0,6703\,\hat{p} + 0,8187\,(1-\hat{p}),$$

de donde resulta $\hat{p} = 0,8$.

Problema 2

a) A partir de la muestra se obtiene: $n = 10$, $\overline{X}_n = 6,217$, $s_n = 2,9685$. Por otra parte, con $\alpha = 0,05$ resulta $t_{\frac{\alpha}{2}}(n-1) = 2,2622$. Por lo tanto:

$$I = \left[\overline{X}_n - \frac{1}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\, s_n\,,\, \overline{X}_n + \frac{1}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\, s_n\right] = [4,0934\ ;\ 8,3406].$$

b) Para que el intervalo de confianza tenga longitud 2 se debe cumplir: $\frac{1}{\sqrt{n}} z_{\frac{\alpha}{2}}\sigma = 1$, de manera que podemos estimar n por:

$$n \simeq (1,96 \times 3)^2 = 34,57$$

Por lo tanto tomamos $n = 35$.

Problema 3

a) **Test de Rachas:** El número de rachas es $R = 8$ y el p-valor: $\alpha^* = 0,5547$.

Test de correlación de rangos de Spearman: $\sum_{i=1}^{12} (R(X_i) - i)^2 = 162$, de manera que el estadístico es $r_s = 0{,}4336$ y $\alpha^* > 0{,}05$.

En ambos casos es razonable suponer que la muestra es aleatoria.

b) A partir de la muestra se obtiene $\overline{X}_n = 3,6457$ y $s_n = 0,8461$. Con esto:

$$\sqrt{12}\frac{(\overline{X}_n - 3,5)}{s_n} = 0,5964.$$

Como el percentil 0,95 de la distribución t con 11 grados de libertad es 1,7959, aceptamos H_0.

c) El estadístico del test es $D = 0,25$. Por lo tanto $mnD = 36$ y de la tabla se obtiene $\alpha^* > 0,200$, de manera que aceptamos H_0.