BIOESTADÍSTICA
EXAMEN 14 DE FEBRERO DE 2007

DATOS DEL ESTUDIANTE

Nombre	Cédula

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

Problema 1 (26 puntos)

Se estudia la cantidad de visitas a los museos, a lo largo de un año, en la población universitaria. Esta población queda dividida en dos categorías: En la Categoría A, que constituyen el 30 % del estudiantado, la cantidad de visitas anuales a los museos se modela con una variable de Poisson de parámetro $\lambda_A = 2{,}7$. En la Categoría B, que constituyen el 70 % del estudiantado, la cantidad de visitas anuales se modela con una variable de Poisson de parámetro $\lambda_B = 0{,}6$.

a) **(8 puntos)** Si se seleccionan al azar 3 estudiantes de la Categoría A, ¿cuál es la probabilidad de que ninguno de ellos haya hecho una visita a los museos a lo largo del último año?

b) **(8 puntos)** Calcule la probabilidad de que un estudiante de la categoría A realice al menos una visita anual a los museos. Calcule esa misma probabilidad para un estudiante de la categoría B.

c) **(10 puntos)** Suponga que se selecciona un estudiante al azar. Calcule la probabilidad de que ese estudiante haya hecho al menos una visita anual a los museos.

Problema 2 (22 puntos)

El tiempo de reacción muscular a un determinado tipo de impulso eléctrico se modela mediante una variable aleatoria continua X, cuya densidad de probabilidad es

$$f_X(x) = \begin{cases} \frac{2}{a^2}(a-x) & \text{si } 0 \leq x \leq a \\ 0 & \text{en caso contrario,} \end{cases} \tag{1}$$

donde $a > 0$.

a) **(8 puntos)** Calcule $\mathbf{E}(X)$ en función del parámetro a.

b) **(6 puntos)** Calcule $\mathbf{Var}(X)$ en función del parámetro a.

c) **(8 puntos)** Suponga que $X_1, X_2, \ldots, X_n$ son variables aleatorias i.i.d. con densidad definida en (1), tales que $\overline{X_n} = 2{,}4$. Estime el parámetro a.

Problema 3 (20 puntos)

Nota: En cada una de las partes de este problema considere $\alpha = 0{,}10$.

Un indicador del grado de polución atmosférica está dado por el número medio de microgramos de partículas en suspensión por metro cúbico de aire; se considera que el nivel de contaminación es alto si esa cantidad supera los 60 microgramos. Se mide la calidad del aire en un período de 8 días y se obtiene la siguiente muestra:

67.5	55.1	58.1	64.4	60.8	57.6	60.5	66.2

Se asume que los datos corresponden a una distribución normal de valor esperado μ.

a) **(8 puntos)** Construya un intervalo de confianza exacto para μ.

b) **(12 puntos)** Realice una prueba para decidir entre las siguientes hipótesis:

$$\begin{cases} H_0: & \mu \leq 60 \\ H_1: & \mu > 60. \end{cases}$$

Problema 4 (32 puntos)

Nota: En las pruebas de hipótesis utilice el siguiente criterio de decisión: se acepta la hipótesis nula si el p-valor es superior a $0{,}10$.

La siguiente muestra corresponde a los tiempos, medidos en horas, que demoran en recuperarse un grupo de personas de un ataque de pánico, luego de suministrarles una dosis de 1 mg de Alprazolam:

2.25	2.39	0.06	2.74	2.31	1.96	0.76	2.14	1.56	2.69

a) **(8 puntos)** Realice dos pruebas de hipótesis para decidir si es razonable suponer que los datos son independientes e idénticamente distribuidos.

b) **(12 puntos)** Implemente la prueba de ajuste de Kolmogorov-Smirnov para decidir si es razonable suponer que la muestra ajusta a una distribución uniforme, $\mathcal{U}[a, b]$, de parámetros $a = 0$ y $b = 3$.

Se considera ahora una nueva muestra, independiente de la anterior, correspondiente a los tiempos de recuperación de un grupo de pacientes cuando se les aplica una nueva droga que está en etapa de experimentación:

2.85	3.62	3.45	1.51	3.01	1.90	2.40	3.55	2.90	3.42

c) **(12 puntos)** Implemente la prueba de comparación de muestras de Kolmogorov-Smirnov para decidir si es razonable suponer que las dos muestras tienen la misma distribución.

SOLUCIÓN

Problema 1

a) Sea $X \sim \mathcal{P}(\lambda_A)$. Hay que calcular:

$$[\mathbf{P}(X = 0)]^3 = (e^{-\lambda_A})^3 = 3{,}03 \times 10^{-4}.$$

b) Para $X \sim \mathcal{P}(\lambda_A)$:

$$\mathbf{P}(X \geq 1) = 1 - e^{-\lambda_A} = 0{,}933.$$

Para $Y \sim \mathcal{P}(\lambda_B)$:

$$\mathbf{P}(Y \geq 1) = 1 - e^{-\lambda_B} = 0{,}451.$$

c)

$$[1 - e^{-\lambda_A}]\, 0{,}30 + [1 - e^{-\lambda_B}]\, 0{,}70 = 0{,}5956.$$

Problema 2

a) $\mathbf{E}(X) = \frac{2}{a^2} \int_0^a x(a - x)dx = \frac{a}{3}$.

b) Se cumple: $\mathbf{E}(X^2) = \frac{2}{a^2} \int_0^a x^2(a - x)dx = \frac{a^2}{6}$. De manera que:

$$\mathbf{Var(X)} = \mathbf{E}(X^2) - \mathbf{E}^2(X) = \frac{a^2}{18}.$$

c) De la primera parte se obtiene $a = 3\mathbf{E}(X)$; así que, usando el hecho que $\overline{X_n}$ es un estimador de $\mathbf{E}(X)$, tenemos:

$$\hat{a} = 3\overline{X_n} = 7{,}2.$$

Problema 3

De la muestra se obtiene $\overline{X_n} = 61{,}275$ y $s_n = 4{,}396$.

a)

$$I = \left[\overline{X_n} - \frac{s_n}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \, , \, \overline{X_n} + \frac{s_n}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right].$$

De la tabla de Student se obtiene $t_{0,05}(7) = 1{,}895$ y por lo tanto el intervalo queda

$$I = [58{,}33 \, , \, 64{,}22].$$

b) La región crítica es

$$\mathcal{R}_\alpha = \left\{ \overline{X_n} \geq 60 + \frac{s_n}{\sqrt{n}} t_\alpha(n-1) \right\}.$$

De la tabla de Student: $t_{0,10}(7) = 1{,}415$, de manera que la región crítica queda

$$\mathcal{R}_\alpha = \{ \overline{X_n} \geq 62{,}199 \}.$$

Como el valor observado ($\overline{X_n} = 61{,}275$) no pertenece a la región crítica decidimos H_0.

Problema 4

a) **Test de Rachas:**

El número de rachas es $R = 7$ y el p-valor: $\alpha^* = 0{,}454$.

Test de correlación de rangos de Spearman:

$\sum_{i=1}^{10} (R(X_i) - i)^2 = 180$, de manera que el estadístico es $r_s = -0{,}09$ y $\alpha^* = 0{,}406$.

En ambos casos es razonable suponer que los datos son i.i.d.

b) Sabemos que si $X \sim \mathcal{U}[0,3]$ entonces:

$$F_o(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{3} & \text{si } 0 \leq x \leq 3 \\ 1 & \text{si } 3 < x. \end{cases}$$

Para implementar el test de ajuste de Kolmogorov-Smirnov construimos la siguiente tabla:

X_i^*	$F_o(X_i^*)$	$\frac{i-1}{n}$	$\lvert F_o(X_i^*) - \frac{i-1}{n}\rvert$	$\lvert F_o(X_i^*) - \frac{i}{n}\rvert$
0.06	0.02	0	0.02	0.08
0.76	0.253	0.1	0.153	0.053
1.56	0.52	0.2	0.32	0.22
1.96	0.653	0.3	**0.353**	0.253
2.14	0.713	0.4	0.313	0.213
2.25	0.75	0.5	0.25	0.15
2.31	0.77	0.6	0.17	0.07
2.39	0.796	0.7	0.096	0.004
2.69	0.896	0.8	0.096	0.096
2.74	0.913	0.9	0.013	0.087
		1		

De manera que el estadístico es $D = 0{,}353$ y entonces $\alpha^* > 0{,}10$. Por lo tanto aceptamos H_0.

c) Es sencillo ver que la máxima diferencia entre las funciones de distribución empíricas es $D = 0{,}70$. De manera que el estadístico del test de comparación de Kolmogorov-Smirnov es $mnD = 70$ y el p-valor $\alpha^* = 0{,}05$. Por lo tanto decidimos H_1.