

## BIOESTADÍSTICA EXAMEN FEBRERO 2006

### DATOS DEL ESTUDIANTE

Nombre	Cédula

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

### Problema 1 (38 puntos)

Un gen con cuatro alelos (denotados  $A$ ,  $a^{ch}$ ,  $a^h$  y  $a$ ) determina la coloración de la piel de los conejos:

- El alelo  $A$  produce una coloración negra.
- El alelo  $a^{ch}$ , cuando es homocigota produce conejos de color gris oscuro llamados *chinchilla*, y cuando es heterocigota con alelos menores en la jerarquía de dominancia produce conejos de color gris claro.
- El alelo  $a^h$  produce conejos blancos con patas negras llamados *Himalaya*.
- El alelo  $a$  produce conejos albinos.

La jerarquía de dominancia es  $A > a^{ch} > a^h > a$  y los diferentes fenotipos están resumidos en la siguiente tabla:

Fenotipos	Posibles genotipos
Negro	$AA, Aa^{ch}, Aa^h, Aa$
Chinchilla	$a^{ch}a^{ch}$
Gris claro	$a^{ch}a^h, a^{ch}a$
Himalaya	$a^ha^h, a^ha$
Albino	$aa$

Suponga que se cruza un macho  $Aa$  con una hembra  $a^{ch}a$  (se asume en lo que sigue que cada individuo aporta un alelo con igual probabilidad y de manera independiente).

- a) (6 puntos) Determine los posibles fenotipos de las crías y sus probabilidades.
- b) (8 puntos) Si una de las crías es de color negro, ¿cuál es la probabilidad de que su genotipo sea  $Aa$ ?
- c) (10 puntos) Si la pareja tiene cuatro crías, ¿cuál es la probabilidad de que dos o más de ellas sean albinas?

- d) **(14 puntos)** Suponga ahora que se cruzan dos de las crías seleccionadas aleatoriamente, ¿cuál es la probabilidad de que ellas tengan una cría chinchilla?

**Sugerencia:** Construya un diagrama con todos los posibles genotipos.

### Problema 2 (18 puntos)

El contenido bacteriano sedimentario del agua de lluvia se modela mediante una variable aleatoria continua  $X$ , cuya densidad de probabilidad es

$$f_X(x) = \begin{cases} \frac{3}{\theta^3}x^2 & \text{si } 0 \leq x \leq \theta \\ 0 & \text{en caso contrario,} \end{cases} \quad (1)$$

donde  $\theta > 0$ .

- a) **(8 puntos)** Calcule  $\mathbf{E}(X)$  en función de  $\theta$ .
- b) **(10 puntos)** Suponga que  $X_1, X_2, \dots, X_n$  son variables aleatorias i.i.d. con densidad definida en (1) tales que  $\overline{X}_n = 3.5$ . Estime el parámetro  $\theta$ .

### Problema 3 (44 + 5 puntos)

La siguiente muestra corresponde a un análisis de glicemia practicado en el mismo día a doce pacientes:

0.90	1.60	1.29	1.64	1.54	0.99	0.81	1.18	1.10	0.95	1.35	1.65
------	------	------	------	------	------	------	------	------	------	------	------

**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

- a) **(8 puntos)** Realice una prueba de hipótesis para decidir si es razonable suponer que los datos son independientes e idénticamente distribuidos.
- b) **(10 puntos)** Realice una prueba de hipótesis para decidir si es razonable suponer que los datos corresponden a una distribución normal.

**Ayuda:** Acá y en lo que sigue pueden ser de utilidad los siguientes valores obtenidos a partir de la muestra:  $\sum_{i=1}^{12} X_i = 15$  y  $\sum_{i=1}^{12} (X_i)^2 = 19.78$ .

- c) **(26 puntos)** Asumiendo que los datos corresponden a una distribución normal de valor esperado  $\mu$ ,
- Construya un intervalo de confianza exacto para  $\mu$  al nivel  $\alpha = 0.10$ .
  - Realice una prueba de hipótesis para decidir entre las siguientes hipótesis:

$$\begin{cases} H_0 : \mu \leq 1.2 \\ H_1 : \mu > 1.2 \end{cases}$$

- d) **Parte extra: (5 puntos)** Sabiendo que el verdadero valor esperado es  $\mu = 1.3$ , calcule la probabilidad de decidir  $H_0$ .

## SOLUCIÓN

### Problema 1

a) Los posibles genotipos de las crías son

$Aa^{ch}$	$Aa$	$aa^{ch}$	$aa$
-----------	------	-----------	------

y cada uno de ellos tiene probabilidad  $\frac{1}{4}$ . De manera que los posibles fenotipos son: Negro, Gris claro y Albino; y sus probabilidades:

$$\mathbf{P}(\text{Negro}) = \frac{1}{2}, \quad \mathbf{P}(\text{Gris claro}) = \frac{1}{4} \quad \text{y} \quad \mathbf{P}(\text{Albino}) = \frac{1}{4}.$$

b) Si una cría es de color negro entonces su genotipo es  $Aa^{ch}$  ó  $Aa$ , de manera que la probabilidad de que su genotipo sea  $Aa$  es  $\frac{1}{2}$ .

c) De las cuatro crías, denotamos con  $X$  la cantidad de las que son albinas. Está claro que  $X$  tiene distribución binomial de parámetros  $n = 4$  y  $p = \frac{1}{4}$ . Por lo tanto:

$$\mathbf{P}(X \geq 2) = 1 - \mathbf{P}(X = 0) - \mathbf{P}(X = 1) = 1 - (0.75)^4 - 4(0.25)(0.75)^3 = 0.2617$$

d) El espacio muestral para los genotipos de las crías de la segunda generación se obtiene estudiando todas las combinaciones posibles para los genotipos de la parte a):

$AA$	$Aa^{ch}$	$AA$	$Aa$	$Aa$	$Aa^{ch}$	$Aa$	$Aa$
$a^{ch}A$	$a^{ch}a^{ch}$	$a^{ch}A$	$a^{ch}a$	$a^{ch}a$	$a^{ch}a^{ch}$	$a^{ch}a$	$a^{ch}a$
$AA$	$Aa^{ch}$	$AA$	$Aa$	$Aa$	$Aa^{ch}$	$Aa$	$Aa$
$aA$	$aa^{ch}$	$aA$	$aa$	$aa$	$aa^{ch}$	$aa$	$aa$
$aA$	$aa^{ch}$	$aA$	$aa$	$aa$	$aa^{ch}$	$aa$	$aa$
$a^{ch}A$	$a^{ch}a^{ch}$	$a^{ch}A$	$a^{ch}a$	$a^{ch}a$	$a^{ch}a^{ch}$	$a^{ch}a$	$a^{ch}a$
$aA$	$aa^{ch}$	$aA$	$aa$	$aa$	$aa^{ch}$	$aa$	$aa$
$aA$	$aa^{ch}$	$aA$	$aa$	$aa$	$aa^{ch}$	$aa$	$aa$

El fenotipo Chinchilla corresponde al genotipo  $a^{ch}a^{ch}$ , por lo tanto:

$$\mathbf{P}(\text{Chinchilla}) = \frac{4}{64} = 0.0625$$

### Problema 2

a) De la definición de valor esperado:

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_0^{\theta} \frac{3}{\theta} x^3 dx = \frac{3}{4}\theta$$

b) Sabemos que el valor esperado se estima por el promedio  $\overline{X_n}$ , de manera que el parámetro  $\theta$  se puede estimar a partir de la igualdad anterior como

$$\hat{\theta} = \frac{4}{3}\overline{X_n} = 4.67$$

### Problema 3

a) **Test de Rachas:** El número de rachas es  $R = 7$  y el  $p$ -valor:  $\alpha^* = 0.4453$ .

**Test de correlación de rangos de Spearman:**  $\sum_{i=1}^{12} (R(X_i) - i)^2 = 264$ , de manera que el estadístico es  $r_s = 0.0769$  y  $\alpha^* > 0.10$ .

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

b) A partir de la muestra se obtienen los valores

$$\overline{X_n} = 1.25 \quad \text{y} \quad s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X_n})^2} = 0.306$$

**Test de D'Agostino para normalidad:** El estadístico es

$$D = \frac{1}{n^2 s_n} \sum_{i=1}^n \left( i - \frac{n+1}{2} \right) X_i^* = 0.273,$$

de manera que  $\alpha^* > 0.20$ . Es razonable suponer entonces que los datos corresponden a una distribución normal.

**Nota:** Otros tests que podrían haberse usado son el test de normalidad de Shapiro-Wilk y el test de ajuste de Kolmogorov-Smirnov (este último bastante más complicado de implementar).

c) i) Asumimos que los datos corresponden a una distribución normal con desviación estándar desconocida (que estimamos por  $s_n$ ), de manera que un intervalo de confianza exacto para  $\mu$  al nivel  $\alpha$  es:

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

De la tabla de Student se obtiene  $t_{0.05}(11) = 1.80$  y por lo tanto el intervalo buscado es

$$I = [1.091, 1.409]$$

- ii) Usamos como estadístico el promedio  $\overline{X}_n$ : si esta cantidad es suficientemente grande decidimos  $H_1$ ; en caso contrario decidimos  $H_0$ .

Asumiendo que  $H_0$  es verdadera con  $\mu = 1.2$  (que es el valor del borde), la región crítica al nivel  $\alpha$  es

$$\mathcal{R}_\alpha = \left\{ \overline{X}_n \geq 1.2 + \frac{s_n}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right\}$$

De la tabla de Student:  $t_{0.10}(11) = 1.363$ , de manera que la región crítica es

$$\mathcal{R}_{0.10} = \{ \overline{X}_n \geq 1.32 \}$$

Como el valor observado es  $\overline{X}_n = 1.25$ , que no pertenece a  $\mathcal{R}_{0.10}$ , decidimos  $H_0$ .

- d) **Parte extra:** En la parte anterior calculamos la región crítica; ahora queremos calcular cuál es la probabilidad de no caer en esa región crítica (que es la condición para decidir  $H_0$ ) sabiendo que  $\mu = 1.3$ . Esto es:

$$\mathbf{P}(\overline{X}_n \leq 1.32) = \mathbf{P}\left( \sqrt{n} \frac{(\overline{X}_n - 1.3)}{s_n} \leq \sqrt{n} \frac{(1.32 - 1.3)}{s_n} \right) = \mathbf{P}(t(n-1) \leq 0.226) \simeq 0.60$$