UNIVERSIDAD DE LA REPÚBLICA - Facultad de Ingeniería

Examen de PROBABILIDAD Y ESTADÍSTICA - 15 de Febrero de 2003

APELLIDOS	NOMBRE	CEDULA DE IDENTIDAD

Nro. DE EXAMEN

Total de puntos: 100. *(se necesitan 50 puntos para aprobar)*
Duración: 4 horas

Problema 1 (Total: 50 puntos)

Se sabe que la opinión de una población se divide en dos opciones (A y B). No hay individuos indecisos. Sin embargo, al consultar un individuo al azar por su opción de preferencia, éste puede no responder a la pregunta. Sea p la probabilidad de que un individuo prefiera la opción A y sean q_1 y q_2 las probabilidades de que un individuo consultado, conteste a la pregunta dado que su opción de preferencia es la A o la B respectivamente. Suponemos además que aquellos que responden no mienten.

1. Calcular las siguientes probabilidades:

   (a) Probabilidad de que un individuo al ser consultado, efectivamente conteste.

   (b) Probabilidad de que un individuo conteste efectivamente la opción A.

   (c) Probabilidad de que la respuesta sea la opción A, dado que el individuo contestó alguna opción.

2. En una población suficientemente grande, se consulta a n individuos. Sea X la variable aleatoria "número de respuestas A en la muestra". Hallar $\mathbf{P}(X = k)$ con $0 \leq k \leq n$, $\mathbf{E}(X)$ y $\mathbf{Var}(X)$.

3. Se propone usar como estimador de p a la proporción de respuestas por la opción A en la muestra. ¿Esto es correcto? Justifique.

4. Supongamos ahora que $n = 100$ y se computan 45 respuestas por la opción A en la muestra. Hallar un Intervalo de Confianza al 95% para la proporción de respuestas A.

5. La decisión de quién será el futuro técnico de la Selección Uruguaya se definirá en un plebiscito entre dos candidatos (de iniciales JRC y OWT). Una empresa encuestadora desea anticipar el resultado realizando una encuesta sobre 100 individuos. Suponemos como antes que no hay indecisos, pero existe la posibilidad de no contestar a la pregunta. En base a resultados históricos, la empresa supone que el porcentaje de personas partidarias del candidato JRC que contestan es del 80%. Se obtienen 45 respuestas a favor del candidato JRC. ¿Hay evidencia suficiente que permita afirmar que el candidato ganador es OWT?

6. ¿Arriesgaría usted un ganador si el porcentaje real de personas partidarias de JRC que contestan es del 70%? En caso afirmativo, explique en qué basa su decisión.

Problema 2 (Total: 50 puntos)

Un fabricante de fusibles asegura que, con una sobrecarga del 20%, los tiempos de vida de sus fusibles (desde que se conectan con sobrecarga hasta que se funden) se distribuyen uniformemente entre 10 minutos y 15 minutos.

Para probar esta afirmación una muestra de 8 fusibles fue sometida a una sobrecarga del 20%. Los tiempos en que tardaron en fundirse dichos fusibles fueron los siguientes:

13.34	10.69	13.37	11.14	13.87	13.75	10.76	12.63

1. Estudiar la aleatoriedad de la muestra.
2. ¿Es razonable suponer que la distribución de la muestra es $\mathbf{U}[10, 15]$?. Justifique su respuesta.
3. Si la distribución de los tiempos de vida fuese $\mathbf{U}[10, 15]$ el tiempo "medio" de vida sería de 12.5 minutos ¿Los datos experimentales confirman la afirmación anterior? Justifique su respuesta.

Examen de PROBABILIDAD Y ESTADISTICA - Febrero 2003
Soluciones

Solución problema 1

1. Definimos los sucesos:

$$\begin{aligned} A &= \text{El individuo prefiere la opción A.} \\ B &= \text{El individuo prefiere la opción B.} \\ C &= \text{El individuo contesta efectivamente la pregunta.} \end{aligned}$$

   Conocemos entonces de la letra:

$$\begin{aligned} \mathbf{P}(A) &= p \\ \mathbf{P}(B) &= 1-p \text{ (ya que no hay indecisos)} \\ \mathbf{P}(C \mid A) &= q_1 \\ \mathbf{P}(C \mid B) &= q_2 \end{aligned}$$

   (a) Se desea hallar $\mathbf{P}(C)$. Escribiendo $C = (C \cap A) \cup (C \cap B)$ donde la unión es disjunta podemos calcular:

$$\begin{aligned} \mathbf{P}(C) &= \mathbf{P}(C \cap A) + \mathbf{P}(C \cap B) = \\ &= \mathbf{P}(C \mid A)\,\mathbf{P}(A) + \mathbf{P}(C \mid B)\,\mathbf{P}(B) = \\ &= q_1 p + q_2(1-p) \end{aligned}$$

   (b) Se desea hallar $\mathbf{P}(C \cap A) = \mathbf{P}(C \mid A)\,\mathbf{P}(A) = q_1 p$.

   (c) Se desea hallar $\mathbf{P}(A \mid C)$. Utilizando la definición y las partes anteriores:

$$\mathbf{P}(A \mid C) = \frac{\mathbf{P}(C \cap A)}{\mathbf{P}(C)} = \frac{q_1 p}{q_1 p + q_2(1-p)}$$

2. Si X es el número de respuestas A en la muestra, se tiene un muestreo sin reposición donde la proporción de respuestas A la obtenemos de $\mathbf{P}(C \cap A)$ (ya que el individuo debe contestar *y además* preferir la opción A). Sin embargo, como la población es grande, aproximamos la distribución de X por un muestreo con reposición. Es decir, la variable X es tal que:

$$X \sim Bin(n, \mathbf{P}(C \cap A)) \Rightarrow X \sim Bin(n, q_1 p)$$

   De lo anterior se sigue que:

$$\begin{aligned} \mathbf{P}(X = k) &= C_k^n (q_1 p)^k (1 - q_1 p)^{n-k} \\ \mathbf{E}(X) &= n q_1 p \\ \mathbf{Var}(X) &= n q_1 p (1 - q_1 p) \end{aligned}$$

3. Definamos:

$$\hat{p}_{A,n} = \frac{\sharp\{\text{resp. A en la muestra}\}}{n} = \frac{X}{n}$$

Como $X \sim Bin\,(n, q_1 p) \Rightarrow X = \sum_{i=1}^{n} X_i$ con X_i $iid \sim Ber\,(q_1 p)$. Entonces:

$$\hat{p}_{A,n} = \frac{X}{n} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}_n \xrightarrow[n]{c.s.} \mathbf{E}X_1 = q_1 p$$

donde hemos usado la LFGN. Por lo tanto, no es correcto en general utilizar $\hat{p}_{A,n}$ como estimador de p ya que $\hat{p}_{A,n} \to p$ cuando $n \to \infty$ solo si $q_1 = 1$, es decir, todos aquellos individuos cuya opción de preferencia es A contestan efectivamente la pregunta.

4. Se tiene ahora $n = 100$ y $\sharp\,\{\text{resp. A en la muestra}\} = X = 45$. Por lo tanto:

$$\hat{p}_{A,100} = \frac{45}{100} = 0.45$$

Tomando $p_A = q_1 p$ (proporción de respuestas A) y utilizando el intervalo de confianza para proporciones dado por:

$$I_\alpha = \left[\hat{p}_{A,n} - \frac{z_{\alpha/2}\sqrt{\hat{p}_{A,n}\,(1-\hat{p}_{A,n})}}{\sqrt{n}}, \hat{p}_{A,n} + \frac{z_{\alpha/2}\sqrt{\hat{p}_{A,n}\,(1-\hat{p}_{A,n})}}{\sqrt{n}}\right]$$

donde los valores en nuestro caso son:

$$\begin{aligned} \hat{p}_{A,100} &= 0.45 \\ n &= 100 \Rightarrow \sqrt{n} = 10 \\ \alpha &= 0.05 \Rightarrow z_{\alpha/2} = 1.96 \end{aligned}$$

se obtiene:

$$I_{0.05} = [0.45 - 0.0975, 0.45 + 0.0975] = [0.3525, 0.5475]$$

que constituye un Intervalo de Confianza al 95% para la proporción de respuestas A ($q_1 p$).

5. Identificando a JRC con la opción A y a OWT con la opción B de las partes anteriores, el dato de la letra es que:

$$\mathbf{P}\,(C \mid A) = 0.8 = q_1$$

Como $n = 100$ y la cantidad de respuestas por el candidato JRC es 45 podemos utilizar el intervalo de confianza para $q_1 p$ de la parte anterior, es decir que con una confianza del 95%:

$$0.3525 \leq q_1 p \leq 0.5475$$

y si suponemos $q_1 = 0.8$ (exactamente) entonces dividimos la desigualdad anterior entre 0.8 para obtener que con una confianza del 95%:

$$0.4406 \leq p \leq 0.6844$$

por lo que no podemos afirmar que $p < 0.5$ (lo que daría el triunfo al candidato OWT) ya que no hay evidencia suficiente.

6. En este caso, repetimos el razonamiento de antes pero utilizando $q_1 = 0.7$ para obtener que con una confianza del 95%:

$$0.5036 \leq p \leq 0.7821$$

lo que nos permitiría afirmar con una confianza del 95% que $p > 0.5$ y por lo tanto el ganador será JRC.

Solución problema 2

En todos las pruebas de hipótesis se acepta $\mathbf{H}_0$ si el p-valor $\alpha^* \geq 0,1$

(a) TEST DE ALEATORIEDAD

Test de rachas de subidas y bajadas

13.34	
10.69	0
13.37	1
11.14	0
13.87	1
13.75	0
10.76	0
12.63	1

Estadístico de prueba
$\mathbf{R} =$ "número de rachas" $= 6$

Como $\mathbf{R} = 6 > \frac{2(8)-1}{3} = 5$, planteamos

$\mathbf{H}_0$	$\mathbf{H}_1$
$X_1, ..., X_n$ son i.i.d	"hay muchas rachas"

Usando la Tabla 9 tenemos que:

n	$\mathbf{R}$	P
$\downarrow$	$\downarrow$	$\downarrow$
8	$\vdots$	$\vdots$
	6	$0.3250 = \alpha^*$

Conclusión: No se rechaza $\mathbf{H}_0$.

Test de correlación de rangos de Spearman

Datos	13.34	10.69	13.37	11.14	13.87	13.75	10.76	12.63
índice	1	2	3	4	5	6	7	8
rango	5	1	6	3	8	7	2	4

Estadístico de prueba
$\mathbf{R}_s = 1 - 6\dfrac{\sum\limits_{i=1}^{i=n} \left(R\left(X_i\right) - i\right)^2}{n\left(n^2-1\right)} = 1 - 6\dfrac{78}{8\left((8)^2 - 1\right)} = 0.07\,142\,9$

Como el estadístico de Spearman dió positivo planteamos:

$\mathbf{H}_0$	$\mathbf{H}_1$
$X_1, ..., X_n$ son i.i.d	hay tendencia creciente

Usando la Tabla tenemos que:

n	$\mathbf{R}$	P
$\downarrow$	$\downarrow$	$\downarrow$
$\vdots$	$\vdots$	$\vdots$
8	$\downarrow$	$\vdots$
	0.071	$0.441 = \alpha^*$

Tenemos que

$$0.1 \leq \alpha^* \text{ (p-valor)}$$

Conclusión: No se rechaza $\mathbf{H}_0$.

(b) TEST DE AJUSTE

Como la distribución está perfectamente determinada (es uniforme con parámetros conocidos) podemos aplicar el **Test de Kolmogorov- Smirnov** (No podemos aplicar un Test χ^2, pues es para n "grande")

$\mathbf{H}_0$	$\mathbf{H}_1$
$F = \mathbf{U}[10, 15]$	$F \neq \mathbf{U}[10, 15]$

Estadístico de prueba
$\mathbf{E} = \sup\limits_{t \in \mathbb{R}} \{\lvert F_n(t) - F_0(t) \rvert\}$

siendo

$$\begin{aligned} F_n(t) &= \frac{\text{cantidad de valores muestrales de } X \leq t}{n} \\ F_0\ (t) &= \frac{t-10}{5} \text{ (distribución uniforme } \mathbf{U}[10, 15]\text{)} \end{aligned}$$

En nuestro problema tenemos:

Muestra ordenada	$\mathbf{F_0}$	$\mathbf{F_n}$		$\lvert \mathbf{F_n} - \mathbf{F_0} \rvert$	
10.69	0.138	0.125	0	0.013	0.138
10.76	0.152	0.25	0.125	0.098	0.027
11.14	0.228	0.375	0.25	0.147	0.022
12.63	0.526	0.5	0.375	0.026	0.151
13.34	0.668	0.625	0.5	0.043	0.168
13.37	0.674	0.75	0.625	0.076	0.049
13.75	0.750	0.875	0.75	0.125	0
13.87	0.774	1	0.875	0.226	0.101

El estadístico de prueba es

$$D_n = 0.226$$

usando la Tabla 6

n	**0.2**	**0.1**
$\downarrow$	$\downarrow$	$\downarrow$
8	$\vdots$	$\vdots$
	0.358	0.410

Por lo tanto

$$\alpha^* >> 0.2$$

Conclusión: No se rechaza H_0

(c) TEST SOBRE LA MEDIANA

No podemos aplicar Test paramétricos para la media $\mu = 12.5$ pues tenemos pocos datos (no es aplicable el Teorema Central del Límite).

Pero si podemos aplicar Test no paramétricos para la mediana $m = 12.5$ (pues si la distribución es uniforme media y mediana coinciden y valen 12.5)

Test de signos

X_i	13.34	10.69	13.37	11.14	13.87	13.75	10.76	12.63
U_i	1	0	1	0	1	1	0	1

Estadístico de prueba
$\mathbf{U} = \sum_{i=1}^{i=n} U_i = 5$

Como $\mathbf{U} > \mathbf{E}(\mathbf{U}) = \frac{n}{2} = 4$, planteamos

$\mathbf{H}_0$	$\mathbf{H}_1$
$m = 12.5$	$m > 12.5$

usando la Tabla 5

n	**P**	**U**
$\downarrow$	$\downarrow$	$\downarrow$
8	$\vdots$	$\vdots$
	$0.3633 = \alpha^*$	5

Conclusión: No se rechaza H_0

Test rangos signados de Wilcoxon.

X_i	13.34	10.69	13.37	11.14	13.87	13.75	10.76	12.63
U_i	1	0	1	0	1	1	0	1
$\lvert X_i - 12.5\rvert$	0.84	1.81	0.87	1.36	1.37	1.25	1.74	0,13
$R^+(X_i)$	2	8	3	5	6	4	7	1

obtenemos nuestro estadístico

$$\mathbf{W}^+ = \sum_{j=1}^{n} U_j R^+(X_j) = 16$$

como

$$\mathbf{W}^+ = 16 < \mathbf{E}\left(\mathbf{W}^+\right) = \frac{n(n+1)}{4} = \frac{8(9)}{4} = 18$$

planteamos

$\mathbf{H}_0$	$\mathbf{H}_1$
$m = 12.5$	$m < 12.5$

Usando la tabla

n	$\mathbf{W}^+$	P
$\downarrow$	$\downarrow$	$\downarrow$
8	$\vdots$	$\vdots$
	16	$0.422 = \alpha^*$

Conclusión: No se rechaza $\mathbf{H}_0$