Tarea Examen - Probabilidades y Estadística - Primavera 2009

Profesor: Fernando Lema
Auxiliares: Abelino Jiménez - Benjamín Palacios

Pregunta 1.

a) Sean X, Y variables aleatorias independientes, ambas $N(0,1)$. Calcule usando el Teorema de Cambio de Variable la función de densidad de la variablea aleatoria $\frac{X}{Y}$.

b) En una fábrica se seleccionan diariamente motores y se inspeccionan hasta encontrar el primer motor defectuoso. Se llama X la variable el número de motores inspeccionados hasta encontrar uno defectuoso:
i) Justifique que X sigue una distribución geométrica de parámetro p (la probabilidad de fallar).
ii) Sea una muestra aleatoria de n días. Encuentre el estimador de máxima verosimilitud del parámetro p.
iii) Se muestra en la tabla adjunta la información del número de motores inspeccionados en un registro de $n = 100$ días. Estime la probabilidad de que un día cualquiera se deban inspeccionar más de dos motores (estricto) para encontrar uno defectuoso.

Número de motores inspeccionados	1	2	3	4	6	Total
Número de días	8	10	15	25	42	100

Pregunta 2.

a) Sea una muestra aleatoria simple $\{x_1, ..., x_n\}$de distribución uniforme sobre $[0, \theta]$, con $\theta > 0$.
i) Describa un test estadístico para verificar si se puede admitir que la distribución es uniforme.
ii) Sea ahora la hipótesis nula de Ho: $\theta = 1$ contra la hipótesis alternativa H1: $\theta = a$, dode $a > 1$ dado. Muestre que el estadístico del test es $Max\{x_i\}$ y deduzca la región crítica de nivel de significación igual a α.
iii) Dé la expresión de la función de potencia del test Ho: $\theta = 1$ contra H1: $\theta = a$.
iv) Dé el error de tipo II en función de a.
v) Dé el mínimo valor n tal que los errores de tipo I y II sean inferiores a 1%.

b) Sea p_1 la proporción de mujeres entre 18 y 30 años que no está inscrita en los registros electorales para votar en Diciembre en las próximas elecciones. Sea p_2 la proporción de hombre entre 18 y 30 años que tampoco está inscrito. Sean $\hat{p}_1$ y $\hat{p}_2$ los estimadors habituales de p_1 y p_2 basados en encuestas de tamaños n y m respectivamente (suponga ambos números grandes). Determine un procedimiento para estudiar

$$H_0 : \; p_1 = p_2$$
$$H_1 : \; p_1 \neq p_2$$

Pregunta 3.

a) Un laboratorio requiere determinar la concentración de ADN de una bacteria en un extracto dado. Considerando que se comenten errores de medición, se replica 20 veces la medición. En la tabla adjunta se presenta los datos obtenidos de manera aleatoria e independiente. Se sabe que las mediciones se ajustan a una distribución $N(\mu, \; \sigma^2)$, donde μ y σ son desconocidos y μ es el valor real de la concentración de ADN.

Replica	Concentración	Replica	Concentración	Replica	Concentración	Replica	Concentración
1	0.503	2	0.249	3	1.047	4	0.670
5	0.882	6	0.858	7	0.805	8	0.549
9	0.563	10	0.706	11	0.503	12	0.536
13	1.168	14	0.813	15	0.496	16	0.790
17	0.518	18	0.463	19	0.735	20	1.038

i) Encuentre el intervalo de confianza al 95% para μ a partir de los datos de la tabla.
ii) Para el laboratorio es más arriesgado sobreestimar el valor real de μ de la concentración. Encuentre un intervalo $[a.b]$ de modo que la concentración de ADN pertenezca a dicho intervalo con un nivel de confianza de 95% y tal que $P(\mu < a) = \frac{1}{4}P(\mu > b)$.
iii) Para demotrar al laboratoio la baja calidad de las mediciones, se solicita a usted que encuentre un intervalo al cual pertenezca σ^2, el valor real de la varianza de la medición, con un nivel de confianza de 90%

b) En el año 2000, en las salas cunas de la Región Metropolitana se obtuvieron datos sobre 622 niñas y 694 niños recién nacidos. Supongamos que X e Y son las variables aleatorias que representan los perímetros del cráneo de las niñas y de los niños que se distribuyen según las distribuciones normales $N(\mu_1, \sigma_1^2)$ y $N(\mu_2, \sigma_2^2)$ respectivamente. Se observó en las muestras que las medias de cada grupo son $\overline{x} = 34,643\,cm$, $\overline{y} = 35,062\,cm$ y las varianzas de $S_1^2 = 2,18\,cm^2$, $S_2^2 = 2,16\,cm^2$.

i) Se sabe que en la totalidad del país, los datos indican que el perímetro del cráneo de los recién nacidos de sexo masculino es en promedio 35cm con una desviación estándar de 1,35cm. ¿Los recién nacidos de sexo masculino de la Región Metropolitana poseen un perímetro de cráneo significativamente diferentes a la totalidad del país?

ii) Queremos ahora comparar los perímetros del cráneo de los recién nacidos niños y niñas de la Región Metropolitana. ¿Podemos admitir que $\sigma_1^2 = \sigma_2^2$? Compare los perímetros del cráneo de los recién nacidos niños y niñas de la Región Metropolitana.

Pregunta 4.

a) Un supermercado quiere ver si la espera de sus clientes se distribuye como una Uniforme(0,10), para así saber cómo plantear una nueva forma de publicidad. Para realizar el estudio, se consideró una muestra de clientes durante 3 horas, llegándose a la siguiente tabla:

5,4	5,1	1,5	3,6	4,2	1,4	1	7,8	4,9	3,4
7,4	5,7	2	8	5,9	9,4	7,4	2,9	0,8	4,8
4	2,4	6,5	8,4	3,4	2,6	4,3	7	2,4	1,4
4,6	4,8	6,5	3,2	4,6	3,7	4,8	8,6	3,2	1

Tabla tiempo de espera para 40 clientes

i) Realice un test de ajuste considerando un nivel de significación de $\alpha = 0,05$.

ii) ¿Se puede decir que el 40% de los clientes espera más de 5 minutos? (Realice un Test de Proporciones).

b) A una empresa de estudios de mercado se le encargó el estudio de la elección del nombre de una nueva marca de cigarrillos. Con este fin se realiza una encuesta cuyo resultado se presenta en la tabla de contingencia.

	Nombre de la Marca			
Percepción	Alazán	Corsario	Zodiaco	TOTAL
Viril	24	46	10	80
Ridículo	6	30	36	72
Sobrio	46	2	4	52
Vulgar	8	64	14	86
TOTAL	84	142	64	290

i) Analice si existe una relación entre la percepción del consumidor y el nombre de la marca. Use su resultado para elegir el nombre de la marca.

ii) Indique el nivel de confianza mínimo con el cual se puede decir que el que 50% considera todos los nombres propuestos vulgares.

Pregunta 5.

Se quiere estudiar la relación entre la Viscosidad del aceite (X), y el Volumen de desgaste del acero dulce (Y), para un determinado tipo de motor.

Por medio de una muestra, se obtuvieron los siguientes datos:

Y	240	181	193	155	172	110	113	75	94
X	1,6	9,4	15,5	20	22	35,5	43	40,5	33

i) Haga un gráfico a partir de los datos anteriores

ii) Realice la estimación de mínimos cuadrados a partir de la tabla.

iii) Estudie las Hipótesis

$$\begin{array}{ll} H_0: \ \beta_1 = 0 & \qquad H_0: \ \beta_2 = 0 \\ H_1: \ \beta_1 \neq 0 & \qquad H_1: \ \beta_2 \neq 0 \end{array}$$

y concluya.
iv) Construya un intervalo de confianza al 95% para β_1 y β_2.

Pregunta 6.
a) Partiendo de los datos $(x_1, y_1), ..., (x_n, y_n)$,
i) Demuestre que si $\alpha_1 = \beta_1 + \beta_2 x$ y $\alpha_2 = \beta_2$, entonces

$$\sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 + n(\alpha_1 - \overline{y})^2 + \alpha_2^2 \sum_{i=1}^{n} (x_i - \overline{x})^2 - 2\alpha_2 \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

ii) Usando la parte anterior, deduzca que $\overline{y}$ y

$$a = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}$$

son estimadores de mínimos cuadrados de α_1 y α_2 respectivamente.

b) Considere el modelo de regresión $\mathbb{E}(Y \mid X = x) = \beta x$, donde β es desconocido. Suponga que obtenemos los valores independientes $(x_1, y_1), ..., (x_n, y_n)$.
i) Determine el estimador de mínimos cuadrado de β.
ii) Demuestre que b, el estimador por mínimos cuadrado de β, es insesgado.
iii) Suponiendo que $var(Y \mid X = x) = \sigma^2$, demuestre que

$$Var(B \mid X_1 = x_1, ..., X_n = x_n) = \frac{\sigma^2}{\sum\limits_{i=1}^{n} x_i^2}$$

iv) Con las condiciones anteriores, demuestre que

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - b x_i)^2$$

es un estimador insesgado de σ^2.

NOTA: Esta tarea es de carácter obligatorio para los alumnos eximidos.

Fecha de Entrega: Hasta día del Examen.