SOLUCIÓN AL EXAMEN DE ESTADÍSTICA II DE 28 DE JUNIO DE 2012

PROBLEMA 1

Sea X=Ventas diarias en euros de un comercio, X~N(450,96).

A) El beneficio diario es la diferencia entre las ventas y el gasto: B=X-200. Esta variable es un cambio de origen (transformación lineal de una normal) y tendrá otra distribución normal cuya media y varianza son:

$$E[B] = E[X - 200] = E[X] - 200 = 250€$$
$$V[B] = V[X - 200] = V[X] = 96^2 \Rightarrow \sigma_B = 96€$$

La probabilidad solicitada se obtiene tipìficando, puesto que es una normal, y buscando en tablas:

$$P\{B > 300\} = P\left\{Z > \frac{300 - 250}{96}\right\} = P\{Z > 0{,}52\} = 0{,}3015$$

El 30,15% de los días el comercio tendrá un beneficio superior a 300 euros

B) Repetimos el experimento dicotómico (éxito=beneficio mayor de 300 y fracaso=beneficio inferior a 300) 48 días laborales. La v.a. U=días con beneficio superior a 300 euros es una binomial de parámetros n=48 y p=0,3015; U~Bi(48,0,3015). Como no está en tablas, podemos aproximar a una normal cuya media es np=48·0,3015=14,47 y varianza np(1-p)=10,11.

Puesto que es una normal, tendremos que tipificar y calcular la probabilidad de que al menos 15 días tengamos un beneficio superior a 300, recordando la corrección por continuidad:

$$P\{U \geq 14{,}5\} = P\left\{Z > \frac{14{,}5 - 14{,}47}{3{,}18}\right\} = P\{Z > 0{,}01\} = 0{,}4960$$

El 49,60% de los bimestres tendremos al menos 15 días con beneficio superior a 300€.

C) Sea W=tiempo de entrega, en días, de un pedido, W~ε(1/10). Nos piden la probabilidad $P\{W > 10 | W > 5\}$. Recordando la propiedad de falta de memoria de la exponencial:

$$P\{W > 10 | W > 5\} = P\{W > 10 - 5\} = P\{W > 5\}$$

A continuación se calcula la probabilidad utilizando la función de densidad puesto que no tenemos esta distribución tabulada:

$$P\{W>5\}=\int_5^{\infty}\frac{1}{10}e^{-\frac{1}{10}x}dx=\left[-e^{-\frac{1}{10}x}\right]_5^{\infty}=0-\left(-e^{-\frac{5}{10}}\right)=0{,}6065$$

El 60,65% de las ocasiones que hayamos esperado más de cinco días un pedido, tardará al menos otros cinco días en llegar.

D) Hacemos 50 pedidos y tendremos 50 tiempos de entrega $W_1, W_2, \ldots, W_{50}$ y nos piden la probabilidad de que el tiempo medio de esos 50 tiempos sea mayor de 12 días. El tiempo medio de entrega de los 50 pedidos será: $\overline{W}=\sum_{i=1}^{50}W_i$

Es una media de variables aleatorias independientes, idénticamente distribuidas, con media y desviación típica finitas y el número de sumandos es elevado; por lo tanto, utilizando el Teorema Central del Límite la distribución de $\overline{W}=\sum_{i=1}^{50}W_i$ se aproxima a una normal cuya media es la media de la variables y cuya varianza es la varianza de la variable entre el número de sumandos.

$$E\left[\overline{W}\right]=E\left[W\right]=10$$

$$V[\overline{W}]=\frac{V[W]}{n}=\frac{10^2}{50}\Rightarrow \sigma_{\overline{W}}=1{,}41$$

Conocida la distribución y sus parámetros, podemos calcular la probabilidad de que el tiempo medio sea mayor de 12 días:

$$P\{\overline{W}>12\}=P\left\{Z>\frac{12-10}{1{,}41}\right\}=P\{Z>1{,}41\}=0{,}0793$$

Tiene una probabilidad del 7,93%.

PROBLEMA 2

Sea X una variable aletaroria que mide la duración de una llamada telefónica, X~ε(1/λ)

A) El criterio de máxima verosimilitud se basa en maximizar la función de verosimilitud. Para ello planteamos el siguiente desarrollo:

Función de densidad de un dato muestral: $f(X = X_i) = \frac{1}{\lambda} e^{-\frac{1}{\lambda} X_i}$

Tomamos el logaritmo: $\ln f(X = X_i) = -\ln \lambda - \frac{1}{\lambda} X_i$

Derivamos respecto del parámetro que queremos estimar para buscar el máximo:

$$\frac{\partial \ln f(X = X_i)}{\partial \mu} = -\frac{1}{\lambda} + \frac{1}{\lambda^2} X_i = \frac{X_i - \lambda}{\lambda^2}$$

Hacemos la suma para todos los elementos de la muestra e igualamos a cero, despejando el parámetro:

$$\frac{1}{\lambda^2} \sum_{i=1}^{n} (X_i - \lambda) = \frac{1}{\lambda^2} \left[\sum_{i=1}^{n} X_i - n\lambda \right] = 0 \Rightarrow \left[\sum_{i=1}^{n} X_i - n\lambda \right] = 0 \Rightarrow$$
$$\lambda = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X} = \hat{\lambda}$$

El estimador máximo verosímil de la duración media de una llamada (la media λ de una exponencial) es la media muestral.

B) No conocemos la distribución de la proporción muestral, así que tendremos que utilizar la desigualdad de Chebychev para calcular el tamaño muestral.

$$P\left\{ \left| \overline{X} - \mu \right| < k \frac{\lambda}{\sqrt{n}} \right\} = 1 - \frac{1}{k^2}$$

Dado el nivel de confianza del 96% tenemos que $k^2 = \frac{1}{1 - 0{,}96} = 25$, la peor varianza posible, es decir, max{λ^2} se alcanza cuando λ=6 y entonces vale 36 y el error máximo es de 1,5 minutos, así pues:

$$n = \frac{k^2 \lambda^2}{e^2} = \frac{25 \cdot 36}{1{,}5^2} \cong 400 \text{ llamadas}$$

Visto que el tamaño muestral será elevado entonces podemos considerar que el estimador media muestral tiene una distribución aproximadamente normal con media λ y desviación típica $\lambda / \sqrt{n}$. Por lo tanto, podemos expresar las condiciones del enunciado a través de su distribución:

$$P\left\{\left|\overline{X}-\mu\right|<z_{\alpha/2}\frac{\lambda}{\sqrt{n}}\right\}=1-\alpha$$

El nivel de confianza 1-α=0,96, por lo que buscamos el percentil que deja el 0,02 de probabilidad al final de la normal estándar: $z_{0,02}=2,0537$

El tamaño muestral será,

$$n=\frac{z_{0,02}^{2}\lambda^{2}}{e^{2}}=\frac{2,0537^{2}\cdot 36}{1,5^{2}}\cong 68\,\text{llamadas}$$

Tendremos que realizar un estudio sobre 68 llamadas de la compañía para estimar la duración media.

C) Hacen una encuesta a 800 clientes y calculan que $\hat{\lambda}=\overline{X}=6,8$ y $S_1=3,2$ minutos.

El problema es estimar la media de una v.a. cualquiera con desviación típica desconocida y tamaño muestral elevado. La deducción teórica del intervalo está hecha en teoría, así que expresamos el intervalo de confianza para un nivel de confianza (1-α)100%:

$$\left(\overline{X}\pm z_{\alpha/2}\frac{S_1}{\sqrt{n}}\right)$$

A continuación calculamos el percentil correspondiente y sustituimos en el intervalo para obtener los extremos.

$$1-\alpha=0,95\Rightarrow\frac{\alpha}{2}=0,025\Rightarrow z_{0,025}=1,96$$

$$\left(\overline{X}\pm z_{\alpha/2}\frac{S_1}{\sqrt{n}}\right)\equiv\left(6,8\pm 1,96\cdot\frac{3,2}{\sqrt{800}}\right)\equiv(6,58,\,7,02)$$

La duración media de una llamada telefónica es un valor entre 6,58 minutos y 7,02 minutos.

PROBLEMA 3

Sea X una variable dicotómica (1=estudiante que asiste y 0=estudiante que no asiste) X~Be(p). Hasta la fecha me indican que p=0,60 y queremos saber si va a aumentar el siguiente trimestre y toman una m.a.s. de tamaño n=160 clientes.

A) El contraste de hipótesis se plantea como hipótesis nula la situación actual mientras que la alternativa será el cambio posible, es decir, la disminución de estudiantes en clase.

$$H_0 : p = 0{,}60$$
$$H_1 : p < 0{,}60$$

El error de tipo I es rechazar la hipótesis nula cuando realmente es cierta. Por lo tanto, el profesor piensa en una baja asistencia cuando realmente acudirán un 60%.

El error de tipo II es no rechazar la hipótesis nula cuando realmente es cierta la alternativa. Por lo tanto, el profesor piensa mantener en clase a un número considerable de estudiantes y realmente va a tener una baja asistencia.

B) La proporción muestral obtenida es de 85/160=0,5312. Puesto que el tamaño muestral n=160 es elevado tenemos el siguiente estadístico de contraste y su región crítica según las hipótesis del apartado anterior..

Hipótesis nula	Hipótesis alternativa	Región crítica	Estadístico
$p = 0{,}60$	$p < 0{,}60$	$Z < -z_\alpha$	$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$

El estadístico toma el siguiente valor:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0{,}5312 - 0{,}60}{\sqrt{0{,}60 \cdot 0{,}40}}\sqrt{160} = -1{,}78$$

Para un nivel de significación del 5%, el percentil es $z_\alpha = z_{0,05} = 1{,}6449$

Es evidente que estamos en la región crítica y, por lo tanto, tenemos evidencia para rechazar la hipótesis nula. Creemos que la proporción de estudiantes que van a clase disminuirá en el segundo semestre.

El p-valor será: p-valor= $P\{Z < -1{,}78\} = 0{,}0375$

La probabilidad de observar un valor más extremo que el dado por nuestra muestra es muy pequeña, por lo tanto, tenemos una evidencia razonable en contra de la hipótesis nula. Por lo que rechazaríamos dicha hipótesis nula y creeríamos que la proporción de estudiantes asistentes en clase será menor en el segundo semestre.

C) Tenemos dos muestras: los que asisten a clase y los que no asisten. La primera muestra tiene 85 estudiantes y la segunda muestra tiene 75 estudiantes. Los estudiantes que van a clase tienen una nota media de 6,3 y una cuasidesviación típica de 2,5. Los estudiantes que no asisten tienen una nota media de 5,4 y una cuasidesviación típica de 3,3.

Tenemos dos variables aleatorias poblacionales independientes:

X=nota de un estudiante que asiste a clase.

Y=nota de un estudiante que no asiste a clase.

El intervalo de confianza para la diferencia de medias, teniendo en cuenta que los tamaños muestrales son elevados, es el siguiente:

$$\left(\overline{X}-\overline{Y}-z_{\alpha/2}\sqrt{\frac{S_{1X}^2}{n}+\frac{S_{1Y}^2}{m}},\ \overline{X}-\overline{Y}+z_{\alpha/2}\sqrt{\frac{S_{1X}^2}{n}+\frac{S_{1Y}^2}{m}}\right)$$

Para un nivel de confianza del 95%, el percentil de la normal es: $z_{\alpha/2}=z_{0,025}=1{,}96$. La diferencia entre las estimaciones de la media es: 6,3-5,4=0,9 puntos. El error de estimación viene dado por:

$$1{,}96\sqrt{\frac{S_{1X}^2}{n}+\frac{S_{1Y}^2}{m}}=1{,}96\sqrt{\frac{2{,}4^2}{85}+\frac{3{,}2^2}{75}}=1{,}96\cdot 0{,}45=0{,}89$$

El intervalo de confianza para la diferencia de notas medias reales a un nivel de confianza del 95% es (0,01; 1,79).

La diferencia de notas medias reales está entre 0,01 puntos y 1,79 puntos.

A la vista del intervalo podemos afirmar que la diferencia de notas medias es distinta de cero, es decir, hay una diferencia entre las notas medias de los estudiantes que asisten o no asisten a clase. En concreto, la diferencia es positiva y es claro que los estudiantes que asisten a clase tienen una nota media superior a los que no van a clase.

PROBLEMA 4

A) El estadístico muestral es la forma de evaluar y comparar nuestros resultados empíricos con el modelo teórico propuesto, tanto cuando construimos intervalos de confianza como contrastes de hipótesis. Por lo tanto, tenemos que conocer su distribución para saber qué decisiones y con qué seguridad las tomamos en el problema.

B) La precisión de un intervalo de confianza depende del nivel de confianza y del tamaño muestral. Por lo tanto, tenemos que disminuir el nivel de confianza o aumentar el tamaño muestral. Preferimos, si es posible, aumentar el tamaño muestral porque de esa forma tendremos más precisión con la misma probabilidad de cubrimiento.

C) La potencia de un contraste es la probabilidad de rechazar la hipótesis nula cuando es cierto un valor concreto del parámetro. Por lo tanto, si el valor del parámetro pertenece a la hipótesis alternativa entonces la potencia será el complementario de la probabilidad del error de tipo II, es decir, la probabilidad de tomar una decisión correcta al rechazar la nula. Es evidente que esta probabilidad debe ser lo más alta posible, es decir, un valor cercano a 1.