SOLUCIÓN AL EXAMEN DE ESTADÍSTICA II DE 1 DE FEBRERO DE 2012

PROBLEMA 1

Sea X = Número de hombres que entran en un comercio en una hora, X~P(3)

Sea Y = Número de mujeres que entran en un comercio en una hora, Y~P(4)

A) Sea Z = número de clientes(hombres y mujeres) en 4 horas y se obtiene como: $U=X_1+X_2+X_3+X_4+Y_1+Y_2+Y_3+Y_4$, la suma de los hombres y mujeres que entran en cada hora. Su distribución sigue siendo Poisson y su parámetro es la suma de parámetros: U~P(3+3+3+3+4+4+4+4)≡P(28). A continuación expresamos la probabilidad que nos piden, transformando el suceso en uno de mayor o igual, y buscamos en tablas

$$P\{U < 35\} = 1 - P\{U \geq 35\} = 1 - 0{,}1121 = 0{,}8879$$

El 88,79% de las mañanas tendremos que atender a menos de 35 clientes en el comercio.

B) Sea V = gasto en euros que realiza un cliente en el comercio, V~N(120, 30). A continuación expresamos la probabilidad que nos da el enunciado, llamando k a la cantidad que consumen el 10% de los clientes más gastadores.

$$P\{V > k\} = 0{,}10$$

Puesto que es una normal, tendremos que tipificar y calcular el percentil del 10% para despejar el valor de k.

$$P\{V > k\} = P\left\{Z > \frac{k-120}{30}\right\} = P\{Z > z_{0,10}\} = 0{,}10$$

Buscando en la tabla de la normal (Tabla 4) tenemos el percentil del 10%: $z_{0,10}$=1,2816. Igualamos y despejamos la cantidad mínima solicitada:

$$\frac{k-120}{30} = z_{0,10} = 1{,}2816 \Rightarrow k = 120 + 30 \cdot 1{,}2816 = 158{,}45€$$

El 10% de los clientes que más gastan en la tienda consumen 158,45 euros.

C) Sea W = precio rebajado de los productos, es decir, W=0,70V. Por lo tanto, es un cambio de escala (transformación lineal) de una distribución normal y su distribución será otra normal cuya media y varianza se obtienen así:

$$E[W] = E[0{,}7V] = 0{,}7E[V] = 0{,}7 \cdot 120 = 84€$$
$$Var[W] = Var[0{,}7V] = 0{,}7^2 Var[V] = 0{,}7^2 \cdot 30^2 = 441 \Rightarrow \sigma_W = 21€$$

A continuación se repite un experimento dicotómico 10 veces (para cada cliente) donde el éxito es gastar menos de 66,36€ y el fracaso gastar 66,36€ o más. Por lo tanto, sea Q=número de clientes que han gastado menos de 66,36€ y tendrá una distribución Binomial de parámetros n=10 (repeticiones) y p=P{W<66,36} que tenemos que calcular.

$$P\{W < 66{,}36\} = P\left\{Z < \frac{66{,}36 - 84}{21}\right\} = P\{Z < -0{,}84\} = P\{Z > 0{,}84\} = 0{,}2005$$

Dada la variable aleatoria Q~Bi(10, 0,20) buscamos en tabla la probabilidad solicitada:

$$P\{Q \geq 3\} = 0{,}3222$$

El 32,22% de las ocasiones que entren 10 clientes en el comercio al menos tres gastarán menos de 66,36€.

PROBLEMA 2

Sea X una variable dicotómica (1=cliente propenso a cambiar de compañía y 0=cliente fiel a la compañía). X~Be(p) con p=proporción de clientes propensos a cambiar de compañía de telefonía móvil.

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: $P\{X = X_i\} = p^{X_i}(1-p)^{1-X_i}$

   Tomamos el logaritmo: $\ln P\{X = X_i\} = X_i \ln p + (1-X_i)\ln(1-p)$

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

$$\frac{\partial \log P\{X = X_i\}}{\partial p} = \frac{X_i}{p} - \frac{(1-X_i)}{1-p} = \frac{X_i - p}{p(1-p)}$$

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

$$\frac{1}{p(1-p)}\sum_{i=1}^{n}(X_i - p) = \frac{1}{p(1-p)}\left[\sum_{i=1}^{n} X_i - np\right] = 0 \Rightarrow \left[\sum_{i=1}^{n} X_i - np\right] = 0 \Rightarrow$$

$$p = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{p}$$

   El estimador máximo verosímil de la proporción p de una Bernoulli es la proporción muestral.

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

$$P\left\{|\hat{p} - p| < k\sqrt{\frac{p(1-p)}{n}}\right\} = 1 - \frac{1}{k^2}$$

   Dado el nivel de confianza del 97% tenemos que $k^2 = \frac{1}{1-0{,}97} = 33{,}33$, la peor varianza posible, es decir, max{p(1-p)} se alcanza cuando p=0,5 y entonces vale 0,25 y el error máximo es del 3%, así pues:

$$n = \frac{k^2 p(1-p)}{e^2} = \frac{33{,}33 \cdot 0{,}25}{0{,}03^2} \cong 9259 \text{ clientes}$$

   Visto que el tamaño muestral será elevado entonces podemos considerar que el estimador proporción muestral tiene una distribución aproximadamente normal con media p y desviación típica $\sqrt{p(1-p)/n}$. Por lo tanto, podemos expresar las condiciones del enunciado a través de su distribución:

$$P\left\{\left|\hat{p}-p\right|<z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right\}=1-\alpha$$

El nivel de confianza 1-α=0,97, por lo que buscamos el percentil que deja el 0,015 de probabilidad al final de la normal estándar: $z_{0,015}=2{,}1701$

El tamaño muestral será,

$$n=\frac{z_{0,02}^{2}\,p(1-p)}{e^{2}}=\frac{2{,}1701^{2}\cdot 0{,}25}{0{,}03^{2}}\cong 1309 \text{ clientes}$$

Tendremos que realizar una encuesta a 1309 clientes de la compañía para estimar la proporción de ellos que son propensos a cambiar.

C) Sea I = ventas diarias, en euros, de la compañía durante la promoción tiene una media de 1000€ y una desviación típica de 175€.

Los gastos diarios son 900€, por lo tanto, el beneficio diario en euros, B=I-900, será una variable aleatoria con media y varianza:

$$E[B]=E[I-900]=E[I]-900=100€$$
$$Var[B]=Var[I-900]=Var[I]=175^{2}\Rightarrow\sigma_{B}=175€$$

Si tenemos en cuenta los 64 días de la promoción, el beneficio total será $T=\sum_{i=1}^{64}B_i$, donde B_i es el beneficio del i-ésimo día. Es una suma 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 T se aproxima a una normal cuya media es la suma de medias y cuya varianza es la suma de varianzas.

$$E[T]=64E[B]=64\cdot 100=6400€$$
$$Var[T]=64Var[B]=64\cdot 175^{2}\Rightarrow\sigma_{T}=8\cdot 175=1400€$$

Conocida la distribución y sus parámetros, podemos calcular la probabilidad de que la promoción resulte satisfactoria:

$$P\{T>4000\}=P\left\{Z>\frac{4000-6400}{1400}\right\}=P\{Z>-1{,}71\}=1-P\{Z>1{,}71\}=1-0{,}0436=0{,}9564$$

El gerente estará satisfecho con una probabilidad del 95,64%.

PROBLEMA 3

Sea X una variable dicotómica (1=cliente paga con retrasos y 0=paga regularmente) X~Be(p). Hasta la fecha me indican que p=0,10 y queremos saber si va a aumentar el siguiente trimestre y toman una m.a.s. de tamaño n=1200 clientes.

A) El contraste de hipótesis se plantea como hipótesis nula la situación actual mientras que la alternativa será el cambio posible, es decir, el aumento de retrasos en los pagos.

$$H_0 : p = 0{,}10$$
$$H_1 : p > 0{,}10$$

El error de tipo I es rechazar la hipótesis nula cuando realmente es cierta. Por lo tanto, la empresa piensa que va a tener más dificultades en los pagos que los actuales y realmente esa situación no se produce.

El error de tipo II es no rechazar la hipótesis nula cuando realmente es cierta la alternativa. Por lo tanto, la empresa piensa que se mantiene la situación y realmente va a tener más dificultades y retrasos este trimestre.

B) La proporción muestral obtenida es de 0,115. Puesto que el tamaño muestral n=1200 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
$p = 0{,}10$	$p > 0{,}10$	$Z > z_\alpha$	$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$

El estadístico toma el siguiente valor:

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} = \frac{0{,}115 - 0{,}10}{\sqrt{0{,}10 \cdot 0{,}90}}\sqrt{1200} = 1{,}732$$

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 clientes que van a retrasar sus pagos aumentará este trimestre.

El p-valor será: p-valor= $P\{Z > 1{,}732\} = 0{,}0418$

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 clientes que retrasarán sus pagos aumentará este trimestre.

C) Tenemos dos muestras: los que retrasan pagos y los que pagan regularmente. La primera muestra tiene 11,5%*1200=138 clientes y la segunda muestra tiene 88,5%*1200=1062 clientes. Los clientes con retrasos tienen una compra media de 990€ y una cuasidesviación típica de 146€. Los clientes sin retrasos tienen una compra media de 1125€ y una cuasidesviación típica de 260€.

Tenemos dos variables aleatorias poblacionales independientes:

X=compra de los clientes que ha tenido retrasos en el pago.

Y=compra de los clientes que han pagado regularmente.

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

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

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: 1125-990=135. El error de estimación viene dado por:

$$1{,}96\sqrt{\frac{S_{1X}^2}{n} + \frac{S_{1Y}^2}{m}} = 1{,}96\sqrt{\frac{146^2}{138} + \frac{260^2}{1062}} = 1{,}96 \cdot 14{,}77 = 28{,}95$$

El intervalo de confianza para la diferencia de compras medias reales a un nivel de confianza del 95% es (106,05; 163,95). La diferencia de medias reales está entre 106,05 euros y 163,95 euros.

A la vista del intervalo podemos afirmar que la diferencia de compras medias es positiva, es decir, hay una diferencia clara entre las compras medias de los clientes que han tenido retrasos y los que no. La compra media de los clientes que pagan regularmente es significativamente mayor que la compra media de los clientes que retrasan sus pagos, al menos en 106,05 euros.

PROBLEMA 4

A) Los motivos de seleccionar una muestra es recopilar información sobre un fenómeno real cuando no tenemos acceso a toda la población porque es infinita, porque necesitamos conclusiones rápidas y económicas, porque el estudio necesario supone la destrucción del elemento o porque no conocemos todos los elementos de la población.

B) Un estimador debe ser eficiente para asegurarnos que la estimación puntual proporcionada sea lo más fiable posible. Puesto que un estimador eficiente va a ser insesgado y su varianza mínima entonces estamos seguro que nuestra estimación va a estar alrededor y lo más cerca posible del parámetro desconocido.

C) Un intervalo de confianza es un intervalo de extremos aleatorios que va a contener en su interior el parámetro desconocido con una probabilidad fijada a priori. Si está calculado al 95%, significará que el 95% de las muestras y, por tanto, el 95% de los intervalos construidos de esa forma contendrán a dicho parámetro en su interior.