# SOLUCIÓN AL EXAMEN DE ESTADÍSTICA II DE 1 DE FEBRERO DE 2013

## PROBLEMA 1

Sea X = Tiempo de fabricación de un producto. $\sigma = \sqrt{39} = 6{,}24 \Rightarrow X \sim N(102,\ 6{,}24)$.

A) Me piden P{X<90, por lo tanto, tipifico y busco en tablas.

$$P\{X < 90\} = P\left\{Z < \frac{90-102}{6{,}24}\right\} = P\{Z < -1{,}92\} = P\{Z > 1{,}92\} = 0{,}0274$$

El 2,74% de los productos se fabrican en menos de 90 minutos.

B) El coste por minuto es de 2,54 euros y el producto se vende por 265 euros. Por lo tanto, la variable B=beneficio de un producto=265-2,54X.

Dará beneficios si B>0, por lo tanto, 265-2,54X>0 y despejando:

$$-2{,}54X > -265 \Rightarrow X < \frac{-265}{-2{,}54} = 104{,}33$$

En un producto tendremos beneficios si cuesta menos de 104,33 minutos fabricarlo. La probabilidad de que suceda es:

$$P\{X < 104{,}33\} = P\left\{Z < \frac{104{,}33-102}{6{,}24}\right\} = P\{Z < 0{,}37\} = 1 - P\{Z > 0{,}37\}$$
$$= 1 - 0{,}3557 = 0{,}6443$$

El 64,43% de los productos fabricados producen beneficio a la empresa..

C) Para cada solicitud tenemos un experimento dicotómico (tiene o no los requisitos para ser contratado) y se repite 20 veces de forma independiente. Por lo tanto, Y=nº solicitudes con los requisitos adecuados sigue una distribución binomial de parámetros n=20 y p=0,30. Y~Bi(20; 0,30).

Nos piden que cubramos al menos las 5 plazas, por lo tanto, P{Y≥5}. Se busca en tablas y se obtiene la probabilidad solicitada:

$$P\{Y \geq 5\} = 0{,}7624$$

**PROBLEMA 2**

Sea X= número de clientes atendidos en un día y sabemos que X~P($\lambda$).

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: $P\{X = X_i\} = e^{-\lambda}\dfrac{\lambda^{X_i}}{X_i!}$

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

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

$$\frac{\partial \ln P\{X = X_i\}}{\partial\lambda} = -1 + \frac{X_i}{\lambda} = \frac{X_i - \lambda}{\lambda}$$

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

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

$$\lambda = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

El estimador máximo verosímil del número medio de clientes atendidos en un día ($\lambda$) es la media muestral.

B) Suponiendo que el tamaño muestral necesario va a ser elevado, pensamos que el estimador media muestral se aproxima a un distribución normal: $\overline{X} \sim N\left(\lambda,\ \sqrt{\dfrac{\lambda}{n}}\right)$.

Por lo tanto, expresamos la condición para obtener el tamaño muestral:

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

Despejando de la ecuación, tenemos que:

$$n = \frac{z_{\alpha/2}^2\lambda}{e^2}$$

El nivel de confianza 1-$\alpha$=0,99, por lo que buscamos el percentil que deja el 0,005 de probabilidad al final de la normal estándar: $z_{0,005} = 2{,}5758$

El tamaño muestral, teniendo en cuenta que $\lambda$ como máximo es 50, será,

$$n = \frac{z_{0,05}^2\lambda}{e^2} = \frac{2{,}5758^2 \cdot 50}{2^2} = 82{,}93 \cong 83\,\text{días}$$

Tendremos que realizar un estudio de 83 días y contar el número de clientes atendidos en esos días.

C) Sea Y=tiempo de atención a un cliente. Se ha tomado una m.a.s. de n=200 clientes y se ha obtenido $\overline{Y} = 32{,}5$ minutos y $S_{1Y} = 7{,}8$ minutos.

El intervalo de confianza para la media $\mu$=E[Y], dado que el tamaño muestral es elevado porque n=200, viene dado por:

$$IC(\mu) = \left[ \overline{Y} - z_{\alpha/2} \frac{S_{1Y}}{\sqrt{n}}, \quad \overline{Y} + z_{\alpha/2} \frac{S_{1Y}}{\sqrt{n}} \right]$$

El nivel de confianza 1-$\alpha$=0,98, por lo que buscamos el percentil que deja el 0,01 de probabilidad al final de la normal estándar: $z_{0,01} = 2{,}3263$ y el intervalo se calculará:

$$IC(\mu) = \left[ 32{,}5 - 2{,}3263 \frac{7{,}8}{\sqrt{200}}, \quad 32{,}5 + 2{,}3263 \frac{7{,}8}{\sqrt{200}} \right] = [31{,}22; 33{,}78]$$

El tiempo medio de atención a un cliente es una cantidad comprendida entre 31,22 y 33,78 minutos.

**PROBLEMA 3**

Sea X=rentabilidad diaria de la inversión y hasta antes de la crisis sabíamos que $\mu$=E[X]=0,13%. Queremos saber si ha disminuido por culpa de la crisis.

A) El contraste de hipótesis se plantea como hipótesis nula la situación actual mientras que la alternativa será el cambio posible, es decir, la pérdida de rentabilidad diaria.

$$H_0 : \mu = 0{,}13$$
$$H_1 : \mu < 0{,}13$$

El error de tipo I es rechazar la hipótesis nula cuando realmente es cierta. Por lo tanto, la empresa piensa que va a disminuir su rentabilidad cuando realmente no es cierto. Va a buscar otro tipo de inversión cuando realmente no es necesario.

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 rentabilidad cuando realmente se está reduciendo, es decir, va a ingresar menos de lo que pensaba..

B) La media muestral obtenida en una muestra de 67 días es de 0,09% con una cuasidesviación típica de 0,115%. Puesto que el tamaño muestral n=67 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 |
|---|---|---|---|
| $\mu = 0{,}13$ | $\mu < 0{,}13$ | $Z < -z_\alpha$ | $Z = \dfrac{\overline{X} - \mu_0}{S_1}\sqrt{n}$ |

El estadístico toma el siguiente valor:

$$Z = \frac{\overline{X} - \mu_0}{S_1}\sqrt{n} = \frac{0{,}09 - 0{,}13}{0{,}125}\sqrt{67} = -2{,}62$$

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 $Z = -2{,}62 < -z_\alpha = -1{,}6449$ y, por lo tanto, tenemos evidencia para rechazar la hipótesis nula. Creemos que la rentabilidad media de nuestra inversión ha disminuido por culpa de la crisis.

El p-valor será: p-valor= $P\{Z < -2{,}62\} = 0{,}0044$

La probabilidad de observar un valor más extremo que el dado por nuestra muestra es muy pequeña del 0,44%, por lo tanto, tenemos una evidencia fuerte en contra de la hipótesis nula. Por lo que rechazaríamos dicha hipótesis nula y creeríamos que la rentabilidad diaria media ha disminuido con la crisis.

C) Tenemos dos muestras: los de la inversión anterior y la alternativa que proponen a la empresa. La primera muestra tiene 67 días y la segunda muestra tiene 107 días.

Tenemos dos variables aleatorias poblacionales independientes:

X=rentabilidad diaria de la inversión antigua.

Y=rentabilidad diaria de la inversión alternativa.

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

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

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: 0,13-0,09=0,04. El error de estimación viene dado por:

$$1{,}96\sqrt{\frac{S_{1X}^2}{n}+\frac{S_{1Y}^2}{m}}=1{,}96\sqrt{\frac{0{,}125^2}{67}+\frac{0{,}115^2}{107}}=1{,}96\cdot 0{,}0189=0{,}037$$

El intervalo de confianza para la diferencia de compras medias reales a un nivel de confianza del 95% es (0,003; 0,077). La diferencia de rentabilidades medias reales está entre el 0,003% y 0,077%.

A la vista del intervalo podemos afirmar que la diferencia de rentabilidades medias es positiva, es decir, hay una diferencia clara entre la rentabilidad media de la inversión nueva con respecto a la antigua. Por lo tanto, aconsejaríamos a la empresa que cambie su inversión a la alternativa nueva que le ofrecen.

**PROBLEMA 4**

A) El muestreo probabilístico se basa en la selección de elementos muestrales utilizando criterios probabilísticos mientras que el muestreo no probabilístico utiliza criterios de facilidad, accesibilidad, conveniencia, … Utilizamos el muestreo probabilístico porque nos permite conocer la probabilidad de cada muestra, con lo cual sabremos evaluar los resultados obtenidos y realizar inferencias para toda la población.

B) Un estimador T se dice insesgado con respecto a un parámetro $\theta$ si la media del estimador coincide con dicho parámetro desconocido, es decir, $E[T]=\theta$. Es aconsejable que suceda esta propiedad para asegurar que los posibles valores del estimador se distribuya alrededor del valor real que queremos aproximar. En caso contrario, podría ser que los posibles valores estuvieran mayoritariamente por debajo o por encima del verdadero valor del parámetro y entonces tendríamos problemas de infraestimación o sobreestimación.

C) Un contraste de hipótesis es una técnica estadística para tomar decisiones sobre una afirmación estadística realizada para la población. Si el nivel de significación es del 5% entonces estamos admitiendo que la probabilidad máxima de rechazar mi afirmación inicial sobre la población cuando realmente es cierta es del 5%, es decir, como mucho 5 de cada 100 veces que haga el contraste cambiaré de opinión inicial para equivocarme en la decisión.