SOLUCIÓN AL EXAMEN DE ESTADÍSTICA II DE 20 DE ENERO DE 2014

PROBLEMA 1

Sea X = Tiempo de funcionamiento de una componente. $E[X]=1/\lambda=2300$ ⇨ $X \sim \varepsilon(\lambda=1/2300)$

A) Me piden P{X>2000|X>500}, por lo tanto, empleando la falta de memoria de la distribución exponencial es igual a la probabilidad de la diferencia de tiempos:

$$P\{X > 2000 | X > 500\} = P\{X > 1500\}$$

Como es una variable aleatoria continua sin tabular tendremos que calcularla utilizando la función de densidad:

$$P\{X > 1500\} = \int_{1500}^{\infty} \lambda e^{-\lambda x} dx = \int_{1500}^{\infty} \frac{1}{2300} e^{-\frac{x}{2300}} dx = \left[-e^{-\frac{x}{2300}} \right]_{1500}^{\infty} = e^{-\frac{1500}{2300}} = 0{,}5209$$

B) Tengo 10 componentes y cada una es independiente del resto. Por lo tanto, tengo 10 repeticiones independientes de un mismo experimento dicotómico (funciona más de 660 horas o menos). Calculo la probabilidad de que una componente funcione más de 660 horas:

$$P\{X > 660\} = \int_{660}^{\infty} \lambda e^{-\lambda x} dx = \int_{660}^{\infty} \frac{1}{2300} e^{-\frac{x}{2300}} dx = \left[-e^{-\frac{x}{2300}} \right]_{660}^{\infty} = e^{-\frac{660}{2300}} = 0{,}7505$$

Sea Y=número de componentes que funcionan más de 660 horas de las 10 que tiene instaladas una máquina. Y~Bi(n=10, p=0,7505).

Si la máquina opera más de 660 horas es que las 10 componentes funcionan más de 660 horas, es decir, la variable Y toma el valor 10:

$$P\{Y = 10\} = \binom{n}{k} p^k (1-p)^{n-k} = \binom{10}{10} (0{,}7505)^{10} (0{,}2495)^0 = (0{,}7505)^{10} = 0{,}0567$$

El 5,67% de las máquinas operarán más de 660 horas sin ninguna avería.

C) Sea U=número de roturas de una máquina en un mes Y~P(2).

Tenemos 50 máquinas, por lo tanto, V=número de roturas de todas las máquinas en un mes será otra Poisson porque hay independencia entre las máquinas. V~P(2·50)=P(100). Dado que el intervalo de estudio se duplica (2 meses), entonces el parámetro será proporcional, es decir, el doble.

La variable W=número de roturas de las 50 máquinas en 2 meses sigue una distribución Poisson de media 200 y, como el parámetro es muy elevado, se aproxima a una distribución normal: $W \approx N(\mu=200, \sigma=\sqrt{200})=N(200, 14{,}14)$

La probabilidad solicitada es: $P\{W>230\}=P\{W\geq 231\}$ y, al realizar la corrección por continuidad, tenemos $P\{W\geq 230{,}5\}$. Al ser una distribución normal, basta tipificar y buscar en tablas convenientemente:

$$P\{W \geq 230.5\} = P\left\{Z \geq \frac{230{,}5-200}{14{,}14}\right\} = P\{Z \geq 2{,}16\} = 1 - P\{Z < 2{,}16\}$$
$$= 1 - 0{,}9846 = 0{,}0154$$

PROBLEMA 2

A) Tenemos una variable aleatoria Bernoulli X~Be(p) donde p representa la proporción de contactos que estarían dispuestos a contratar a la nueva empresa. La población consta de N=480 contactos y nos preguntan sobre el tamaño muestral necesario para estimar dicho valor de p con un error máximo del e=3% y una confianza del 1-α=90%. El estimador de p es la proporción muestral $\hat{p}$ y la exigencia impuesta es: $P\{|\hat{p}-p|<0{,}03\}=0{,}90$

Si la muestra se realiza con reposición (tamaño poblacional infinito), el tamaño muestral necesario, suponiendo que es elevado, se calculará como:

$$n_\infty = \frac{z_{\alpha/2}^2 p(1-p)}{e^2}$$

Dado el nivel de confianza 1-α=0,90, podemos obtener el percentil adecuado:

$$\frac{\alpha}{2}=0{,}05 \Rightarrow z_{0{,}05}=1{,}6449$$

En el peor de los casos, la máxima varianza de una Bernoulli se alcanza cuando p=0,50.

El tamaño para un muestreo con reposición sería:

$$n_\infty = \frac{z_{\alpha/2}^2 p(1-p)}{e^2} = \frac{1{,}6449^2\cdot 0{,}5\cdot 0{,}5}{0{,}03^2} = 751{,}58 \cong 752 \text{ contactos}$$

Puesto que la población es finita, el muestreo adecuado sería sin reposición y deberíamos realizar la corrección:

$$n = \frac{n_\infty}{1+\frac{n_\infty}{N}} = \frac{752}{1+\frac{752}{480}} = 292{,}99 \cong 293 \text{ contactos}$$

Deberíamos ponernos en contacto con 293 de los potenciales clientes para estimar con las garantías requeridas la proporción de nuevos clientes.

B) Sea X= número de contratos firmados en una semana y sabemos que X~P(λ)

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

Función de probabilidad de un dato muestral: $P\{X=X_i\}=e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}$

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:

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

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

El estimador máximo verosímil del número medio de contratos firmados en una semana (λ) es la media muestral. Puesto que tenemos datos de 13 semanas, tamaño muestral n=13, y el número de contratos totales son 8; entonces la estimación máximo verosímil del número de contratos semanales firmados es:

$$\hat{\lambda} = \frac{8}{13} = 0{,}6153$$

C) Sea F=facturación de una obra y tiene media μ y desviación típica σ.

La variable aleatoria que mide los beneficios es una transformación lineal (cambio de escala de la variable facturación). B=beneficio neto de una obra y se define B=0,35F, por lo tanto, E[B]=0,35μ y V[B]=$0{,}35^2\sigma^2$.

Nos piden estimar E[B]=beneficio neto medio de una obra, así que necesitamos estimar μ=facturación media de una obra. Propongo como estimador la media muestral, es decir, $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} F_i = \overline{F}$ y dado que conozco que en una muestra de tamaño n=50 obras la facturación global (sumado todas las facturaciones) es 250.000 euros; entonces tengo la estimación de la facturación media:

$$\hat{\mu} = \frac{250000}{50} = 5000€$$

Esta estimación me conduce a estimar el beneficio neto medio de una obra tomando el porcentaje adecuado:

$$\overline{B} = \widehat{E[B]} = 0{,}35 \cdot \hat{\mu} = 0{,}35 \cdot 5000 = 1750 \text{ euros}$$

Dado que la media muestral hemos demostrado que es un estimador insesgado y consistente para la media poblacional, independientemente de la v.a. poblacional, entonces estas propiedades se conservan al realizar un cambio de escala. Así que el estimador $\overline{B} = 0{,}35 \cdot \hat{\mu}$ mantiene las propiedades de insesgadez y consistencia.

La insesgadez se cumple porque coincide la media del estimador con el parámetro que queremos estimar:

$$E[\bar{B}] = E[0{,}35\hat{\mu}] = 0{,}35E[\hat{\mu}] = 0{,}35\mu = E[B]$$

La consistencia se cumple porque es insesgado y su varianza se anula cuando el tamaño muestral tiende a infinito:

$$V[\bar{B}] = V[0{,}35\hat{\mu}] = 0{,}35^2 V[\hat{\mu}] = 0{,}35^2 \frac{\sigma^2}{n}$$

$$V[\bar{B}] = 0{,}35^2 \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0 \text{ porque es una constante divida por infinito}$$

PROBLEMA 3

Sea X=consumo diario de frutas y verduras en un comedor escolar y sabemos que X~N(μ,σ). La muestra tiene tamaño n=16 y con los datos recogidos calculamos la media y la cuasidesviación típica muestrales:

$$\bar{X} = 360{,}50 \text{ gramos y } S_1 = 58{,}81 \text{ gramos}$$

A) Puesto que es una normal con desviación típica poblacional desconocida y el tamaño muestral es pequeño, entonces el intervalo de confianza a un nivel (1-α)100% viene dado por: $\left(\bar{X} - t_{n-1;\alpha/2}\frac{S_1}{\sqrt{n}}, \bar{X} + t_{n-1;\alpha/2}\frac{S_1}{\sqrt{n}}\right)$

   Para n=16 y 1-α=0,95, calculamos el percentil de la distribución t-Student correspondiente: $t_{n-1;\alpha/2} = t_{15;0{,}025} = 2{,}131$. Por lo tanto, el intervalo queda expresado como:

$$\left(\bar{X} - t_{n-1;\alpha/2}\frac{S_1}{\sqrt{n}}, \bar{X} + t_{n-1;\alpha/2}\frac{S_1}{\sqrt{n}}\right) = \left(360{,}5 \pm 2{,}131\frac{58{,}81}{\sqrt{16}}\right)$$
$$= (329.16, 391.84)$$

   El consumo medio de frutas y verduras en un comedor escolar es una cantidad entre 329,16 y 391,84 gramos.

B) El contraste de hipótesis se plantea como hipótesis nula la situación recomendado mientras que la alternativa será una dieta ineficiente, es decir, la cantidad media consumida es menor:

$$H_0 : \mu \geq 380{,}25$$
$$H_1 : \mu < 380{,}25$$

   El error de tipo I es rechazar la hipótesis nula cuando realmente es cierta. Por lo tanto, se piensa que la dieta no es correcta y debe aumentarse la cantidad de fruta y verdura. Van a buscar otra dieta encareciendo la actual y los niños tendrán que consumir más fruta y verdura cuando no hace falta y no les gusta en general.

   El error de tipo II es no rechazar la hipótesis nula cuando realmente es cierta la alternativa. Por lo tanto, el Gobierno de Aragón no cambia la dieta cuando es necesario y los niños no consumen la cantidad necesaria.

C) La media muestral obtenida en una muestra de 16 centros es de 360,50 gramos con una cuasidesviación típica de 58,81 gramos. Puesto que el tamaño muestral n=16 es pequeño tenemos el siguiente estadístico de contraste y su región crítica según las hipótesis del apartado anterior..

Hipótesis nula	Hipótesis alternativa	Región crítica	Estadístico
$\mu \geq 380{,}25$	$\mu < 380{,}25$	$T < -t_{n-1;\alpha}$	$T = \dfrac{\overline{X} - \mu_0}{S_1}\sqrt{n}$

El estadístico toma el siguiente valor:

$$T = \frac{\overline{X} - \mu_0}{S_1}\sqrt{n} = \frac{360{,}5 - 380{,}25}{58{,}81}\sqrt{16} = -1{,}34$$

Para un nivel de significación del 1%, el percentil es $t_{n-1;\alpha} = t_{15;0{,}01} = 2{,}602$

Es evidente que **<u>no estamos en la región crítica:</u>** $T = -1{,}34 > -t_{n-1;\alpha} = -2{,}602$ y, por lo tanto, no tenemos evidencia para rechazar la hipótesis nula. Creemos que el consumo medio de frutas y verduras en los comedores escolares de Aragón se adecúa a la recomendación del Ministerio de Sanidad y Consumo.

El p-valor será: p-valor= $P\{t_{15} < -1{,}34\} = P\{t_{15} > 1{,}34\} \cong 0{,}10$

La probabilidad de observar un valor más extremo que el dado por nuestra muestra es aproximadamente del 10%, por lo tanto, no tenemos evidencia en contra de la hipótesis nula. Por lo que no rechazaríamos dicha hipótesis nula y creeríamos que el consumo medio es el adecuado y nuestros escolares reciben una dieta acorde a los recomendaciones del Ministerio.

PROBLEMA 4

A) Una variable aleatoria es una aplicación que asigna un número real a cada suceso de un experimento aleatorio, es decir, una cuantificación numérica de los posibles resultados de un fenómeno aleatorio. Se dice que es variable aleatoria discreta si utiliza un conjunto finito o infinito numerable de números reales mientras que se dice v.a. continua si toma valores en un conjunto infinito de la recta real.

B) La selección de una muestra para su estudio se realiza principalmente por los siguientes motivos: la población es infinita o demasiado grande, el estudio implica la destrucción del elemento, las características poblacionales cambian si el estudio se prolonga demasiado tiempo y abaratar costes monetarios y temporales.

C) El nivel de confianza de un intervalo es la probabilidad que garantizamos para que dentro del intervalo se encuentre el verdadero valor del parámetro que queremos estimar. Si el intervalo está calculado con un nivel del 95%, significa que el 95% de los intervalos, es decir, el 95% de las muestras extraídas, construidos contendrán en su interior al verdadero valor del parámetro.