EJERCICIOS U.D.5 INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

5.1- DISTRIBUCIONES EN EL MUESTREO

1. Calcular la probabilidad de que la media de una muestra de 16 elementos extraídos al azar e independientes de una población normal de media 0 y varianza 25 sea en valor absoluto superior a 3.

N = 16, $X \sim N(m=0, \sigma^2=25) \Rightarrow \overline{X} \sim N(0, 5/4)$

$P(|\overline{X}|>3) = 2 \cdot P(\overline{X}>3) = 2 \cdot P(N(0,1)>[(3-0)/(5/4)]) = 2 \cdot P(N(0,1) > 12/5) =$
$=2 \cdot P(N(0,1)>2,4)= 2 \cdot 0,0082 = 0,0164$

2. El tiempo de transferencia de paquetes de un determinado tamaño a través de la red en milisegundos sigue una distribución N(m, σ=4). Si se extrae una muestra aleatoria simple de tamaño 25, calcular la probabilidad de que la diferencia entre la media poblacional y la media muestral sea mayor que 2 en valor absoluto.

Si X es N(m, σ=4) $\Rightarrow$ La media muestral $\overline{X}$ de una muestra aleatoria simple de tamaño 25 es N(m, σ=4/5) , y $\overline{X}$ -m será N(0, σ=4/5).
Por tanto:
$P(|\overline{X}-m|>2)= P(|N(0,4/5)|>2)=P(N(0, 4/5)>2)+P(N(0, 4/5)<-2)=$
$=2P(N(0,4/5)>2)=2 P(N(0,1)>10/4)=2 P(N(0,1)>2.5)= 2 \text{ x } 0.00621=0.01242$

3. Se extrae una muestra de tamaño N de una población normal de desviación típica 4. Calcular cuánto debe valer como mínimo N si se desea una probabilidad inferior al 1% de que la diferencia entre la media muestral y la media poblacional sea en valor absoluto superior a 2.

Si la población es normal entonces $\overline{X}$ es N(m, $\frac{4}{\sqrt{N}}$) y $\frac{\overline{X}-m}{4/\sqrt{N}}$ es N(0,1).

$P(|\overline{X}-m|>2)<0,01 \Rightarrow 2\ P(\overline{X}-m>2)<0,01 \Rightarrow P(\overline{X}-m>2)<0,005 \Rightarrow$

$P(N(0,1)>\frac{2}{4/\sqrt{N}})<0,005 \Rightarrow \frac{\sqrt{N}}{2} \geq 2,58 \Rightarrow N \geq (2x2,58)^2=26,62$

por lo que el valor de N mínimo es 27.

4. El tiempo de transferencia de paquetes de un determinado tamaño a través de la red sigue una distribución normal con varianza 9. Si se extrae una muestra aleatoria simple de tamaño 25, calcular la probabilidad de que la diferencia entre las medias poblacional y muestral sea mayor que 2 en valor absoluto.

X=tiempo de transferencia$\Rightarrow$X es N(m,3)$\Rightarrow$ $\overline{X}$ es N(m,3/5)$\Rightarrow$ $\overline{X}$ -m es N(0,3/5)
$P[|\overline{X} - m| >2)] = 2\ P(Z > 2 / (3 /5)) = 2 \cdot 0,000434 = 0,000868$. Donde Z es N(0,1)

5. Se extrae una muestra de tamaño N de una población normal de desviación típica σ=3. Calcular cuanto debe valer como mínimo N si se desea tener una probabilidad inferior al 1% de que la diferencia entre la media muestral y la media poblacional sea en valor absoluto superior a 1

Como $\frac{\bar{x}-m}{\sigma/\sqrt{N}} = \sqrt{N}\frac{\bar{x}-m}{3}$ ~ N(0,1) ⇒

P(|$\bar{x}$ -m|>1) = $P\left(\left|\sqrt{N}\frac{\bar{x}-m}{3}\right| > \sqrt{N}/3\right) = P\left(|N(0,1)| > \sqrt{N}/3\right)$ ha de ser < 0.01

Mirando en la tabla de la normal se ve que para que se cumpla esto $\sqrt{N}/3$ ha de ser > 2.58 ⇒

$$N > (3x2.58)^2 \Rightarrow N \text{ mínimo}= 60$$

6. Las tres componentes de un vector se distribuyen normal e independientemente con media cero y varianza 4. Hallar el valor medio del cuadrado de la longitud del vector.

Sean X, Y, y Z las tres componentes del vector. La longitud al cuadrado del mismo será:
$L^2 = X^2 + Y^2 + Z^2$
El valor medio de esa longitud
$E(L^2)=E(X^2+Y^2+Z^2)=4\ E(\frac{X^2}{4}+\frac{Y^2}{4}+\frac{Z^2}{4})$
Como X, Y y Z son variables Normales de media 0 y varianza 4, $X^2/4$, $Y^2/4$ y $Z^2/4$ serán normales tipificadas (de media cero y varianza 1) al cuadrado. Por tanto su suma seguirá, ya que son independientes, una distribución Chi-Cuadrado con 3 grados de libertad. La media de esa suma serán los grados de libertad 3. Con lo que la media del cuadrado de la longitud será:
$E(L^2)=4 \times 3= 12$

7. Calcular la probabilidad de que una χ^2 con 20 grados de libertad sea mayor que 34,17 a partir de la tabla de dicha distribución y también aproximando a una distribución normal.

Con la tabla se busca en la fila de 20 grados de libertad el valor 34,17, y se lee la probabilidad en la columna donde está ese valor. La probabilidad resulta 0,025.
Aproximando a la distribución normal de media 20 y desviación típica $\sqrt{40}$:
$P(\chi^2 >34{,}17)\approx P(N(20, 6{,}32) >34{,}17)=P(N(0, 1)>\frac{34{,}17-20}{6{,}32})=P(N(0, 1) > 2{,}24)=0{,}0125$

8. Se extraen muestras de tamaño 10 de una población normal de varianza 9. Calcular la probabilidad de que la varianza muestral resulta mayor que 18.

Se aplicará el resultado $(N-1)\frac{s^2}{\sigma^2} \sim \chi^2$ con N-1 grados de libertad.
En este caso $9\frac{s^2}{9} \sim \chi^2$ con 9 grados de libertad
$P(s^2>18)=P(\chi^2_9 > 18)$

Buscando en la tabla de la distribución Chi-Cuadrado en la fila de 9 grados de libertad se tiene:
$P(\chi^2_9 > 19{,}023)=0{,}025$
$P(\chi^2_9 > 16{,}919)=0{,}05$

Por tanto la probabilidad que se desea calcular estará entre esos dos valores:
$0{,}025<P(\chi^2_9 > 18)>0{,}05$
Concretamente calculada con el programa Statgraphics da el valor $P(\chi^2_9 > 18)=0{,}035$.

9. Dada una población Normal(m=8, σ=4) y una muestra aleatoria simple de tamaño N=15 extraída de la misma, determinar un intervalo de longitud mínima y tal que la probabilidad de que contenga a la varianza muestral sea 99%.

Se utiliza el resultado $(N-1)\frac{s^2}{\sigma^2} \sim \chi^2$ con N-1 grados de libertad.
En este caso $14\frac{s^2}{16} \sim \chi^2$ con 14 grados de libertad. Con la tabla de la distribución Chi-Cuadrado es posible hallar dos valores g_1 y g_2 tales que
$P(g_1 < 14\frac{s^2}{16} < g_2)=0{,}99 \Rightarrow$
$g_1= 4{,}075$ (se busca en la fila 14 grados de libertad y columna α=0,995)
$g_2= 31{,}319$ (se busca en la fila 14 grados de libertad y columna α=0,005)

Entonces $P(4{,}075 < 14\frac{s^2}{16} < 31{,}319)=0{,}99$ y despejando s^2 se obtiene el intervalo que la contiene con un 99% de probabilidad,

$$\left[16\frac{4{,}075}{14}, 16\frac{31{,}319}{14}\right] = [4{,}657, 35{,}793]$$

10. Calcular un valor x tal que la t de Student con 15 grados de libertad sea mayor que x en valor absoluto con una probabilidad del 10%.

Se quiere obtener x de forma que $P(|t_{15}| > x)=0{,}1$, por lo que $P(t_{15}>x)+P(t_{15}<-x)=0{,}1$. Estas dos probabilidades son iguales al ser la distribución t simétrica respecto al 0 $\Rightarrow$ $2P(t_{15}>x)=0{,}1 \Rightarrow P(t_{15}>x)=0{,}05$. A partir de este resultado se puede buscar x en la tabla de la t de Student, en la fila de 15 grados de libertad y la columna α=0,05, lo que resulta x=1,753

11. Hallar dos valores entre los que se encuentre la t de Student con 12 grados de libertad con una probabilidad del 80%.

Un intervalo de longitud mínimo que contenga el 80% de la distribución t_{12}, estará centrado en el cero y sus dos límites serán –x y x de forma que:
$P(-x < t_{12} < x)=0{,}8$ lo que implica que $P(t_{12} >x)=0{,}1$. Se puede encontrar x en la tabla de la t con 12 grados de libertad y en la columna α=0,1, lo que da como resultado x=1,356. Por tanto el intervalo puede ser [-1,356, 1,356]

12. El tiempo de respuesta de un sistema informático sigue una distribución normal de media m=5. Se han realizado N=20 pruebas para medir dicho tiempo. Si la varianza muestral s^2 resulta igual a 2,5, calcula un intervalo que contenga el valor de la media muestral $\bar{X}$ con una probabilidad del 95%.

Como la distribución en la población es normal se puede aplicar el resultado $\frac{\bar{X}-m}{s/\sqrt{N}} \sim t_{N-1}$, que en este caso $\frac{\bar{X}-5}{\sqrt{2,5}/\sqrt{20}} \sim t_{19}$. Con la tabla de la distribución t de Student es posible hallar dos valores –t y t tales que,

$P(-t < t_{19} < t) = P(-t < \frac{\bar{X}-5}{\sqrt{2,5}/\sqrt{20}} < t) = 0,95$. t se busca con 19 grados de libertad en la columna α=0,025 ⇒ t= 2,093 ⇒ $P(-2,093 < \frac{\bar{X}-5}{\sqrt{2,5}/\sqrt{20}} < 2,093) = 0,95$, y despejando $\bar{X}$, resulta $P(5 - 2,093\sqrt{\frac{2,5}{20}} < \bar{X} < 5 + 2,093\sqrt{\frac{2,5}{20}})=0,95$. El intervalo que contiene la media muestral con probabilidad del 95% resulta: [4,26, 5,74]

13. El voltaje de un determinado tipo de circuito electrónico sigue una distribución normal de media m=120. Se ha extraído una muestra de N=10 circuitos, y la desviación típica muestral ha resultado s=1,5. Calcula la probabilidad de que la media muestral resulte mayor que 121.

Se desea calcular $P(\bar{X}>121)$, se va a utilizar el resultado $\frac{\bar{X}-m}{s/\sqrt{N}} \sim t_{N-1}$, donde sustituyendo los datos disponibles $\frac{\bar{X}-120}{1,5/\sqrt{10}} \sim t_9$.

$P(\bar{X}>121)=P(\frac{\bar{X}-120}{1,5/\sqrt{10}} > \frac{121-120}{1,5/\sqrt{10}})= P(t_9 > 2,108)$

Buscando en la tabla t de Student con 9 grados de libertad se obtiene:
$P(t_9 > 2,262) = 0,025$
$P(t_9 > 1,833) = 0,05$
Por tanto la probabilidad que se desea calcular estará entre los valores,
$0,025 < P(\bar{X}>121) < 0,05$. Con el Statgraphics resulta $P(\bar{X}>121)=P(t_9>2,108)=0,032$

14. Se tiene un grupo de 1500 personas, ¿cuántos sujetos tendrían una puntuación entre 0,28 y 8,1, si la puntuación es una variable que sigue una distribución F con 8 y 6 grados de libertad ?

El número de sujetos con puntuación entre 0,28 y 8,1 se puede calcular con
1500 P(0,28 < puntuación < 8,1)
Como la puntuación sigue una distribución $F_{8,6}$
$P(0,28<\text{puntuación}<8,1)=P(0,28<F_{8,6}<8,1)=P(F_{8,6}<8,1)-P(F_{8,6}<0,28)=$
$=1-P(F_{8,6}>8,1)-P(F_{6,8}>\frac{1}{0,28}) = 1 - 0,01 - P(F_{6,8} > 3,58) = 0,99 - 0,05 = 0,94 \Rightarrow$
Número de sujetos con puntuación entre 0,28 y 8,1=1500 x 0,94 =1410

15. La temperatura que soporta un determinado tipo de equipo informático sigue distribución normal. Si se extraen dos muestras aleatorias de equipos de tamaños N_1=10 y N_2=18 de la población de equipos, calcula la probabilidad de que la primera varianza muestral resulte más de 2,5 el valor de la segunda varianza muestral.

Se pide $P(s_1^2 > 2{,}5\, s_2^2)$. Como la distribución de la variable es normal y ambas muestras proceden de la misma población se cumple que $\frac{s_1^2}{s_2^2} \sim F_{9,17} \Rightarrow$

$P(s_1^2 > 2{,}5\, s_2^2) = P\left(\frac{s_1^2}{s_2^2} > 2{,}5\right) = P(F_{9,17} > 2{,}5) \approx 0{,}05$ (ya que en la tabla de la F entrando con grados de libertad de numerados 9 y denominador 17, se lee para α=0,05 el valor 2,49).