EJERCICIOS TEMA 3

1º ESTÁ DESARROLLADO EN LA TEORIA

2º Sea $X$ una v. a. continua con función de densidad $f(x) = \dfrac{e^{-\frac{1}{2}(\ln x-\mu)^2}}{x\sqrt{2\pi}}$ si $x>0$. Se dispone de una m.a.s. suya. Se desea obtener un intervalo de confianza para $\mu$ y para ello se sabe que $\sum_{i=1}^{n} \ln x_i \rightsquigarrow N(n\mu, n)$. Lleve a cabo el proceso completo que conduzca a la obtención del intervalo requerido de amplitud mínima.

Si $\sum_{i=1}^{n} \ln x_i \rightsquigarrow N(n\mu, n) \Rightarrow \dfrac{\sum_{i=1}^{n} \ln x_i - n\mu}{n} \rightsquigarrow N(0,1)$

$\Rightarrow \dfrac{\sum_{i=1}^{n} \ln x_i}{n} - \mu \rightsquigarrow N(0,1)$

Podemos tomar como función pivotal $v(\mu)$

$v(\mu) = \dfrac{\sum_{i=1}^{n} \ln x_i}{n} - \mu$ ya que esta función tan sólo depende de la muestra y del parámetro $\mu$ a estimar. Además $v(\mu) \rightsquigarrow N(0,1)$, su distribución no depende de parámetros desconocidos. Queremos hallar un intervalo $(d_1, d_2)$ tal que:

$P[d_1 < v(\mu) < d_2] = 1-\alpha$ donde $\alpha$ = nivel de significación

Esto es equivalente a poner que $F(d_2) - F(d_1) = 1-\alpha$ siendo $F$ la función de distribución de probabilidad de $N(0,1)$.

Pivotemos la desigualdad $d_1 < v(\mu) < d_2$ para acotar $\mu$:

$d_1 < \dfrac{\sum_{i=1}^{n} \ln x_i}{n} - \mu < d_2 \Rightarrow -d_2 < \mu - \dfrac{\sum_{i=1}^{n} \ln x_i}{n} < -d_1$

$\dfrac{\sum_{i=1}^{n} \ln x_i}{n} - d_2 < \mu < \dfrac{\sum_{i=1}^{n} \ln x_i}{n} - d_1$

El intervalo donde se encuentra $\mu$ es:

$$I = \left( \frac{\sum_{i=1}^{n} \ln x_i}{n} - d_2 \,,\ \frac{\sum_{i=1}^{n} \ln x_i}{n} - d_1 \right)$$

La amplitud de este intervalo es:

$$A(d_1, d_2) = d_2 - d_1$$

El problema que debemos resolver es:

$$\text{Min } A(d_1, d_2) = d_2 - d_1$$

$$\text{s.a: } F(d_2) - F(d_1) = 1 - \alpha$$

Lo resolvemos por el método de los multiplicadores de Lagrange. La función de Lagrange es

$$L(d_1, d_2, \lambda) = d_2 - d_1 - \lambda\left[ F(d_2) - F(d_1) + \alpha - 1 \right]$$

Las condiciones de primer orden para la existencia de mínimo son:

$$\frac{\partial L}{\partial d_1} = -1 + \lambda f(d_1) = 0 \Rightarrow f(d_1) = \frac{1}{\lambda}$$

$$\Rightarrow f(d_1) = f(d_2)$$

$$\frac{\partial L}{\partial d_2} = 1 - \lambda \cdot f(d_2) = 0 \Rightarrow f(d_2) = \frac{1}{\lambda}$$

$$\frac{\partial L}{\partial \lambda} = F(d_2) - F(d_1) + \alpha - 1 = 0$$

donde $f$ es la función de densidad de $N(0,1)$ que es simétrica.

Como $f(d_1) = f(d_2) \Rightarrow \begin{cases} d_1 = d_2 \text{ (absurdo ya que } d_1 < d_2) \\ d_1 = -d_2 \text{ (por ser simétrica)} \end{cases}$

Al ser $d_1 = -d_2$:

$$F(d_2) - F(d_1) = 1 - \alpha \Rightarrow F(d_2) - F(-d_2) = 1 - \alpha \Rightarrow$$

$$F(d_2) - \left[ 1 - F(d_2) \right] = 1 - \alpha \Rightarrow F(d_2) = 1 - \frac{\alpha}{2}$$

Si llamamos $\lambda_2 = z_{\frac{\alpha}{2}}$ al valor en donde $P[X < \lambda_2] = 1 - \frac{\alpha}{2} = F(\lambda_2)$ el intervalo de confianza para $\mu$ es:

$$IC_{1-\alpha}(\mu) = \left( \frac{\sum_{i=1}^{n} \ln x_i}{n} - z_{\frac{\alpha}{2}} , \frac{\sum_{i=1}^{n} \ln x_i}{n} + z_{\frac{\alpha}{2}} \right)$$

ya que $d_1 = - z_{\frac{\alpha}{2}}$.

**3º Sea una población uniforme en el intervalo $(0,b)$ con $b$ desconocido. Demuestre que $\left(\frac{\max\{x_1, x_2 \cdots x_n\}}{b}\right)^n \rightsquigarrow U(0,1)$ si $x_1, x_2, \dots x_n$ es una m.a.s. suya. Construya un intervalo de confianza al 90% para $b$ teniendo en cuenta este resultado. Respecto de la cuestión de encontrar el intervalo óptimo, únicamente plantee el problema a solucionar, pero no lo resuelva.**

Sabemos que la función de densidad de $X$ es

$$f(x) = \begin{cases} \frac{1}{b} & \text{si } 0 < x < b \\ 0 & \text{en el resto} \end{cases}$$

Veamos que $Y = \frac{X}{b} \rightsquigarrow U(0,1)$. Si $F$ es la función de distribución de $Y \Rightarrow F(y) = P[Y \leq y] = P[\frac{X}{b} \leq y] =$

$$= P[X \leq by] = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^{by} \frac{1}{b}\, dx & \text{si } 0 < x < b \\ 1 & \text{si } x \geq b \end{cases} \text{, es decir}$$

$$F(y) = \begin{cases} 0 & \text{si } y < 0 \\ y & \text{si } 0 < y < 1 \\ 1 & \text{si } y \geq 1 \end{cases}$$

La función de densidad de $Y$ es:

$$F'(y) = f(y) = \begin{cases} 1 & \text{si } 0 < y < 1 \\ 0 & \text{en el resto} \end{cases} \Rightarrow Y \rightsquigarrow U(0,1)$$

Al ser $T = \left[\dfrac{\max\{x_1, x_2, \ldots x_n\}}{b}\right]^n = \left[\max\left\{\dfrac{x_1}{b}, \dfrac{x_2}{b}, \ldots \dfrac{x_n}{b}\right\}\right]^n$

y cada $\dfrac{x_i}{b} \rightsquigarrow U(0,1)$. Entonces la función de distribución de $T$ es:

$$F(t) = P[T \leq t] = P\left[\max\left\{\frac{x_1}{b}, \frac{x_2}{b}, \ldots, \frac{x_n}{b}\right\}^n \leq t\right]$$

$$P\left[\max\left\{\frac{x_1}{b}, \frac{x_2}{b}, \ldots \frac{x_n}{b}\right\} < t^{1/n}\right] =$$

$$= P\left[\frac{x_1}{b} < t^{1/n}, \frac{x_2}{b} < t^{1/n}, \ldots \frac{x_n}{b} < t^{1/n}\right] = \prod_{i=1}^{n} P\left[\frac{x_i}{b} \leq t^{1/n}\right] = (t^{1/n})^n = t$$

si $0 < t < 1$

Como cada $\dfrac{x_i}{b} \rightsquigarrow U(0,1)$

$$P\left[\frac{x_i}{b} \leq t^{1/n}\right] = \int_0^{t^{1/n}} 1\, dx = x\Big]_0^{t} = t^{1/n}.$$ Por tanto

$$P[T \leq t] = F(t) = \begin{cases} 0 & \text{si } t^{1/n} < 0 \\ (t^{1/n})^n & 0 < t < 1 \\ 1 & \text{si } t \geq 1 \end{cases} = \begin{cases} 0 & \text{si } t \leq 0 \\ t & \text{si } 0 < t < 1 \\ 1 & \text{si } t \geq 1 \end{cases}$$

Luego $F'(t) = f(t) = \begin{cases} 1 & \text{si } 0 < t < 1 \\ 0 & \text{en el resto} \end{cases} \Rightarrow T \rightsquigarrow U(0,1)$

Podemos tomar como función pivotal

$v(b) = \left(\dfrac{\max\{x_1, x_2, \ldots x_n\}}{b}\right)^n$ ya que tan solo depende de la muestra y del parámetro a estimar $b$. Además su distribución de probabilidad $U(0,1)$ no depende de ningun parametro desconocido.
Si el nivel de significación es $\alpha$, pretendemos hallar un intervalo $(d_1, d_2)$ tal que:

$$P[d_1 < v(b) < d_2] = 1 - \alpha$$

Pivotamos la desigualdad $d_1 < v(b) < d_2$ para acotar b

$$d_1 < \left[ \frac{\max\{x_1, x_2, \dots x_n\}}{b} \right]^n < d_2 \Rightarrow$$

$$d_2^{-1/n} < \frac{b}{\max\{x_1, x_2, \dots x_n\}} < d_1^{-1/n}$$

$$\frac{\max\{x_1, x_2, \dots x_n\}}{\sqrt[n]{d_2}} < b < \frac{\max\{x_1, x_2, \dots x_n\}}{\sqrt[n]{d_1}}$$

los intervalos en donde se encuentra b son de la forma:

$$I = \left( \frac{\max\{x_1, x_2, \dots x_n\}}{\sqrt[n]{d_2}}, \frac{\max\{x_1, x_2, \dots x_n\}}{\sqrt[n]{d_1}} \right)$$

la amplitud de este intervalo es

$$A(d_1, d_2) = \max\{x_1, x_2, \dots x_n\} \left( \frac{1}{\sqrt[n]{d_1}} - \frac{1}{\sqrt[n]{d_2}} \right)$$

Como $v(b) \rightsquigarrow U(0,1)$

$P(d_1 < v(b) < d_2) = F(d_2) - F(d_1)$ siendo F la función de distribución de $U(0,1)$, es decir

$$F(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 < t < 1 \\ 1 & \text{si } t > 1 \end{cases}$$

luego $P(d_1 < v(b) < d_2) = F(d_2) - F(d_1) = d_2 - d_1$

El problema que nos encontramos es

Min

$$A(d_1, d_2) = \max\{x_1, x_2, \dots x_n\} \cdot \left( \frac{1}{\sqrt[n]{d_1}} - \frac{1}{\sqrt[n]{d_2}} \right)$$

s.a:

$$d_2 - d_1 = 1 - \alpha$$

la función de Lagrange es

$L(\lambda_1, \lambda_2, \lambda) = \max\{x_1, x_2, \dots x_n\} \left( \frac{1}{\sqrt[n]{\lambda_1}} - \frac{1}{\sqrt[n]{\lambda_2}} \right) - \lambda(\lambda_2 - \lambda_1 + \alpha - 1)$

Las condiciones de primer orden para la existencia de mínimo son:

$$\frac{\partial L}{\partial \lambda_1} = -\frac{1}{n} \cdot \max\{x_1, x_2 \dots x_n\} \cdot \lambda_1^{-\frac{1}{n}-1} + \lambda = 0$$

$$\frac{\partial L}{\partial \lambda_2} = -\frac{1}{n} \cdot \max\{x_1, x_2, \dots x_n\} \cdot \lambda_2^{-\frac{1}{n}-1} - \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = \lambda_2 - \lambda_1 + \alpha - 1 = 0$$

De estas ecuaciones $\lambda_1 = \lambda_2$ Como $\alpha = 0'1$ y $\lambda_2 - \lambda_1 + \alpha - 1 = 0 \Rightarrow$ ¡Contradicción! (nos avisa que no intentemos resolverlo el enunciado)

Entonces ha de ser:

$$P\left[ \frac{\max\{x_1, x_2, \dots x_n\}}{\sqrt[n]{\lambda_2}} < b < \frac{\max\{x_1, x_2, \dots x_n\}}{\sqrt[n]{\lambda_1}} \right] = 0'90$$

4º Sea $x_1, x_2, \dots x_n$ una m.a.s. de una distribución exponencial $Exp(\lambda)$.

a) Determine la función de densidad de las tres variables siguientes: $S = \sum_{i=1}^{n} x_i$, $T = \frac{S}{\lambda}$, $Y = \lambda \cdot S$ ¿En qué casos se trata de distribuciones conocidas?

- $X_i \rightsquigarrow Exp(\lambda)$ $\forall i = 1 \dots n \Rightarrow X_i \rightsquigarrow G(\lambda, 1)$ $\forall i = 1 \dots n$. Como la distribución gamma es reproductiva $S = \sum_{i=1}^{n} X_i \rightsquigarrow G(\lambda, n)$. Luego $S \rightsquigarrow G(\lambda, n)$
- Si $F$ es la función de distribución de $R = \lambda X$

$F(r) = P[R \leq r] = P[\lambda X \leq r] = P[X \leq \frac{r}{\lambda}]$

$$F(r) = \int_0^{\frac{r}{\lambda}} \lambda \cdot e^{-\lambda x} dx = -e^{-\lambda x}\Big]_0^{\frac{r}{\lambda}} = 1 - e^{-r} \Rightarrow$$

$\Rightarrow f(r) = F'(r) = e^{-r}$ si $r > 0$

$$f(r) = \begin{cases} e^{-r} & \text{si } r > 0 \\ 0 & \text{en el resto} \end{cases} \Rightarrow R \rightsquigarrow Exp(1). \Rightarrow$$

$$\Rightarrow \lambda X_i \rightsquigarrow Exp(1) = G(1,1) \Rightarrow \sum_{i=1}^{n} \lambda X_i = \lambda \cdot S \rightsquigarrow G(1,n)$$

Luego $Y = \lambda S \rightsquigarrow G(1,n)$

- Sea $T = \frac{S}{\lambda}$ y $F$ la función de distribución de $T$

Entonces:

$$F(t) = P[T \leq t] = P[S \leq \lambda t] = \int_0^{\lambda t} \frac{\lambda^n}{\Gamma(n)} \cdot e^{-\lambda x} \cdot x^{n-1} dx \quad \text{si } x > 0$$

Por tanto la función de densidad de $T$ es:

$$F'(t) = f(t) = \frac{\lambda^n}{\Gamma(n)} \cdot e^{-\lambda^2 t} \cdot t^{n-1} \cdot \lambda^{n-1} \cdot \quad \text{si } t > 0$$

Luego la función de densidad de $T$ es:

$$f(t) = \begin{cases} \frac{\lambda^{2n-1}}{\Gamma(n)} \cdot e^{-\lambda^2 t} \cdot t^{n-1} & \text{si } t > 0 \\ 0 & \text{en el resto} \end{cases}$$

$T$ no tiene una distribución conocida.

b) ¿Serviría alguna de estas variables como estadístico pivote para calcular un intervalo de confianza de $\lambda$? En caso afirmativo desarrolle todo el proceso necesario hasta obtener el intervalo de confianza de un nivel $p$. Si la distribución que hubiera que manejar, al querer calcular ese intervalo no está en las tablas, deje planteado el intervalo tanto de manera exacta como usando alguna aproximación asintótica que tenga sentido aplicar (NOTA: no se preocupe del tema de la obtención del intervalo óptimo de menor amplitud, asegúrese solamente de que el intervalo que obtiene tenga el nivel de confianza deseado).

Sabemos que $Y = \lambda S = \lambda \cdot \sum_{i=1}^{n} X_i \sim G(1, n)$

Podemos tomar como función pivotal $v(\lambda) = \lambda \cdot \sum_{i=1}^{n} X_i$ ya que tan solo depende del parámetro a estimar $\lambda$ y de la muestra. Además su función de distribución $G(1,n)$ no contiene parámetros desconocidos.

Queremos hallar un intervalo $(\lambda_1, \lambda_2)$ tal que:

$$P[\lambda_1 < v(\lambda) < \lambda_2] = p$$

Pivotamos la desigualdad $\lambda_1 < v(\lambda) < \lambda_2$ para acotar $\lambda$:

$$\lambda_1 < \lambda \cdot \sum_{i=1}^{n} X_i < \lambda_2$$ Como $X_i \sim Exp(\lambda) \Rightarrow X_i > 0$

$$\Rightarrow \sum_{i=1}^{n} X_i > 0 \Rightarrow \frac{\lambda_1}{\sum_{i=1}^{n} X_i} < \lambda < \frac{\lambda_2}{\sum_{i=1}^{n} X_i}$$, es decir:

$$\frac{\lambda_1}{S} < \lambda < \frac{\lambda_2}{S}$$

Los intervalos en los que se encuentra $\lambda$ son de la forma $\left(\frac{\lambda_1}{S}, \frac{\lambda_2}{S}\right)$

La amplitud de este intervalo es: $A(\lambda_1, \lambda_2) = \frac{1}{S}(\lambda_2 - \lambda_1)$

Como ha de ser $P(\lambda_1 < v(\lambda) < \lambda_2) = p \Rightarrow$
$\Rightarrow F(\lambda_2) - F(\lambda_1) = p$ siendo $F$ la función de distribución de $G(1, n)$. El problema que tenemos que resolver es:

$$\text{Min } A(\lambda_1, \lambda_2) = \frac{1}{S}(\lambda_2 - \lambda_1)$$

$$\text{s.a. } F(\lambda_2) - F(\lambda_1) = p$$

La función de Lagrange es:

$$L(\lambda_1, \lambda_2, \xi) = \frac{1}{S}(\lambda_2 - \lambda_1) - \xi\,(F(\lambda_2) - F(\lambda_1) - p)$$

Las condiciones de primer orden para la existencia de mínimo son:

$$\frac{\partial L}{\partial \lambda_1} = -\frac{1}{S} + \xi \cdot f(\lambda_1) = 0$$

$$\frac{\partial L}{\partial \lambda_2} = \frac{1}{S} - \xi \cdot f(\lambda_2) = 0$$

$$\frac{\partial L}{\partial \xi} = F(\lambda_2) - F(\lambda_1) - p = 0$$

Siendo $f$ la función de densidad de $G(1, n)$

Obtenemos que $f(\lambda_1) = f(\lambda_2) = \frac{1}{\xi S}$. Como la distribución gamma no es simétrica, imponemos la condición de que la probabilidad que deja $\lambda_1$ a su izquierda sea $\frac{p}{2}$ y la que deja $\lambda_2$ a su derecha sea también $\frac{p}{2}$, es decir:

$$P[X < \lambda_1] = F(\lambda_1) = \frac{p}{2}$$

$$P[X > \lambda_2] = 1 - P[X < \lambda_2] = 1 - F(\lambda_2) = \frac{p}{2} \Rightarrow F(\lambda_2) = 1 - \frac{p}{2}$$

De aquí podríamos obtener $\lambda_1$ y $\lambda_2$. El intervalo de confianza sería:

$$IC_p(\lambda) = \left(\frac{\lambda_1}{S}, \frac{\lambda_2}{S}\right)$$

Aproximación asintótica.

Hemos visto que si $X \rightsquigarrow Exp(\lambda) \Rightarrow \lambda X \rightsquigarrow Exp(1)$ lo cual indica que $E[X] = \mu = 1$; $Var[X] = 1 = \sigma^2 \Rightarrow \sigma = 1$.

Si $X_1, X_2, \ldots X_n$ es una m.a.s. y $n$ es grande, por el teorema central del límite:

$$\frac{\sum \lambda X_i}{n} = \lambda \bar{X} = \frac{T}{n} \rightsquigarrow N(1, 1)$$

Es decir que $\frac{\lambda S}{n} \rightsquigarrow N(1,1) \Rightarrow \frac{\lambda S}{n} - 1 \rightsquigarrow N(0,1)$

Podemos tomar como función pivotal $v(\lambda) = \frac{\lambda S}{n} - 1$ ya que depende tan sólo de la muestra y del parámetro a estimar $\lambda$. Además su distribución $N(0,1)$ no depende de ningún parámetro desconocido.

Queremos hallar un intervalo $(d_1, d_2)$ tal que

$P[d_1 < v(\lambda) < d_2] = p \Leftrightarrow F(d_2) - F(d_1) = p$ siendo $F$ la función de distribución de $N(0,1)$.

Pivotamos la desigualdad $d_1 < v(\lambda) < d_2$ para acotar $\lambda$:

$$d_1 < \frac{\lambda S}{n} - 1 < d_2 \Rightarrow 1 + d_1 < \frac{\lambda S}{n} < 1 + d_2 \Rightarrow$$

$$\Rightarrow \frac{n(1+d_1)}{S} < \lambda < \frac{n(1+d_2)}{S}$$

Los intervalos en los que se encuentra $\lambda$ son de la forma:

$$I = \left( \frac{n(1+d_1)}{S}, \frac{n(1+d_2)}{S} \right)$$

La amplitud del intervalo es:

$$A(d_1, d_2) = \frac{n(d_2 - d_1)}{S}$$

El problema que tenemos que resolver es:

$$\text{Min } A(d_1, d_2) = \frac{n}{S}(d_2 - d_1)$$

$$\text{s.a: } F(d_2) - F(d_1) = p$$

La función de Lagrange es:

$$L(d_1, d_2, \xi) = \frac{n}{S}(d_2 - d_1) + \xi(F(d_2) - F(d_1) - p)$$

Las condiciones de primer orden para la existencia de mínimo son:

$$\frac{\partial L}{\partial d_1} = -\frac{n}{S} + \xi \cdot f(d_1) = 0$$

$$\frac{\partial L}{\partial \lambda_2} = \frac{n}{S} - \xi \cdot f(\lambda_2) = 0$$

$$\frac{\partial L}{\partial \xi} = F(\lambda_2) - F(\lambda_1) - p = 0$$

donde $f$ es la función de densidad de $N(0,1)$

Se tiene que $f(\lambda_2) = f(\lambda_1) = \frac{nS}{\xi}$. Como la función de densidad $f(x)$ de $N(0,1)$ es simétrica, si

$$f(\lambda_2) = f(\lambda_1) \Rightarrow \begin{cases} \lambda_2 = \lambda_1 & (\text{absurdo ya que } \lambda_1 < \lambda_2) \\ \lambda_1 = -\lambda_2 \end{cases}$$

$$F(\lambda_2) - F(-\lambda_2) = p \Rightarrow F(\lambda_2) - 1 + F(\lambda_2) = p \Rightarrow$$

$$\Rightarrow F(\lambda_2) = \frac{1+p}{2} \qquad \lambda_2 = z_{\frac{1+p}{2}} \Rightarrow \lambda_1 = -z_{\frac{1+p}{2}}$$

El intervalo para $\lambda$ con confianza $p$ es

$$IC_p(\lambda) = \left( \frac{n(1 - z_{\frac{1+p}{2}})}{S}, \frac{n(1 + z_{\frac{1+p}{2}})}{S} \right)$$

5º Se realiza una comparación entre las probabilidades $p_A$ y $p_B$ de aprobar dos asignaturas A y B; en concreto se trabaja con la diferencia $p_A - p_B$. Una muestra aleatoria simple de estudiantes de A de tamaño 50 arroja una cifra de 30 aprobados. Una de B, de tamaño 75, muestra 36 aprobados. Ambas muestras son independientes entre sí. Se usa como estimador de cada una de las probabilidades, la proporción de aprobados de cada muestra, y como estimador de la diferencia entre ambas, la diferencia entre esas proporciones muestrales. Empleando resultados asintóticos adecuados para cada proporción muestral, halle una distribución asintótica para la diferencia entre las proporciones muestrales. Por ahora use tamaños muestrales cualesquiera $n_A$ y $n_B$. Emplee esa distribución para hallar un intervalo aleatorio de nivel p asintoticamente válido para $p_A - p_B$, siga empleando tamaños genéricos para las muestras. Aplíquelo sobre las muestras disponibles y obtenga un intervalo de confianza de tamaño 95%

$X$: proporción de aprobados de la asignatura A
$Y$: " " " " " " B

Tomamos $x_1, x_2, \ldots x_{n_A}$ m.a.s de $X$
$y_1, y_2 \ldots y_{n_B}$ " " $Y$

Por el TCL, si $n_A$ y $n_B$ son grandes:

$$\bar{X} \rightsquigarrow N\left(\mu_A, \frac{\sigma_{\cancel{X}}}{\sqrt{n_A}}\right) \qquad \bar{Y} \rightsquigarrow N\left(\mu_B, \frac{\sigma_{\cancel{Y}}}{\sqrt{n_B}}\right)$$

No conocemos ni $\sigma_A$ ni $\sigma_B$. Las estimamos a partir de las expresiones:

$$s^2_x = \frac{n_A - 1}{n_A} \cdot \sigma^2_x \Rightarrow \sigma_x = \frac{\sqrt{n_A}}{\sqrt{n_A - 1}} s_x$$

$$s^2_y = \frac{n_B - 1}{n_B} \cdot \sigma^2_y \Rightarrow \sigma_y = \frac{\sqrt{n_B}}{\sqrt{n_B - 1}} s_y$$

Por tanto podemos poner

$$\bar{X} \rightsquigarrow N\left(\mu_{\cancel{x}}, \frac{s_x}{\sqrt{n_A - 1}}\right) \quad \bar{Y} \rightsquigarrow N\left(\mu_y, \frac{s_y}{\sqrt{n_B - 1}}\right)$$

El intervalo de confianza al nivel p es:

$$IC_p(\mu_x - \mu_y) = \left((\bar{x} - \bar{y}) - z_{\frac{1-p}{2}} \sqrt{\frac{s^2_x}{n_A - 1} + \frac{s^2_y}{n_B - 1}}, (\bar{x} - \bar{y}) + z_{\frac{1-p}{2}} \sqrt{\frac{s^2_x}{n_A - 1} + \frac{s^2_y}{n_B - 1}}\right)$$

En una muestra $\bar{x} = \frac{30}{50} = 0'6 \quad \bar{y} = \frac{36}{75} = 0'48$, $p = 0'95$

$$IC_p(\mu_x - \mu_y) = \left(-0'12 - z_{0'025} \sqrt{\frac{s^2_x}{n_A - 1} + \frac{s^2_y}{n_B - 1}}, -0'12 + z_{0'025} \sqrt{\frac{s^2_x}{n_A - \cancel{B}} + \frac{s^2_y}{n_B - 1}}\right)$$

6º Se quiere estimar la media $\mu$ de una variable aleatoria X que se distribuye según una ley normal de varianza conocida $\sigma^2 = 6'25$ con la ayuda de una muestra de $n = 100$ observaciones independientes. La media observada en la muestra (media muestral) es 4'3. Determine un intervalo de nivel de confianza del 95% para la media poblacional.

$X \rightsquigarrow N(\mu, 2'25)$ $\quad n = 100 \quad \bar{x} = 4'3 \quad \alpha = 0'05$

según hemos visto en teoría

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - Z_{\frac{\alpha}{2}} \cdot \sigma, \bar{x} + Z_{\frac{\alpha}{2}} \cdot \sigma\right).$$ Por tanto:

$$IC_{95\%}(\mu) = (4'3 - Z_{0'025} \cdot 2'25, 4'3 + Z_{0'025} \cdot 2'25)$$

$$IC_{95\%}(\mu) = (4'3 - 1'96 \cdot 2'25, 4'3 + 1'96 \cdot 2'25) = (3'82, 4'79)$$

7º Examinados los datos acerca del gasto trimestral en vestir de 9 estudiantes, que de antemano podemos suponer que siguen una distribución normal, se obtuvieron los siguientes resultados muestrales: media = = 157'82 €, desviación típica = 38'89 €. Halle un intervalo al 90% de confianza que contenga al gasto medio trimestral en vestir en esa población.

Según hemos visto en teoría, el intervalo de confianza con $\sigma$ desconocido viene dado por:

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}, \bar{x} + t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}\right)$$

En este caso:

$$IC_{90\%}(\mu) = \left(157'82 - t_{8,0'05} \cdot \frac{38'89}{\sqrt{8}}, 157'82 + t_{8,0'05} \cdot \frac{38'89}{\sqrt{8}}\right)$$

$$IC_{90\%}(\mu) = \left(157'82 - 1'86 \cdot \frac{38'39}{\sqrt{8}}, 157'82 + 1'86 \cdot \frac{38'89}{\sqrt{8}}\right)$$

$$IC_{90\%}(\mu) = (132'22, 183'40)$$

b) Razone intuitivamente y demuestre recurriendo a fórmulas precisas, pero sin hacer cálculos de tipo numérico de ninguna clase, qué pasaría, en relación con el intervalo hallado en el apartado anterior, en cada una de las circunstancias siguientes:

b1) Si tuviéramos una muestra con la misma media y desviación típica muestrales pero con un tamaño superior.

La amplitud del intervalo de confianza anterior es, para la muestra de tamaño n:

$$A_n = 2\, t_{n-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}$$

Si la muestra fuera de tamaño m, la amplitud del intervalo sería:

$$A_m = 2\, t_{m-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{m-1}}$$

Si $n < m \Rightarrow \frac{1}{\sqrt{m-1}} < \frac{1}{\sqrt{n-1}}$

$t_{m-1, \frac{\alpha}{2}} < t_{n-1, \frac{\alpha}{2}}$ (t-Student decrece en $(0, +\infty)$)

Por tanto: $t_{m-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{m-1}} < t_{n-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}} \Rightarrow A_m < A_n$,

Si la muestra aumenta en tamaño, la amplitud del intervalo disminuye con lo cual aumenta la precisión al estimar la media poblacional $\mu$.

b2) Si tuviéramos que hallar con la misma muestra inicial, un intervalo de mayor nivel de confianza para la media poblacional.

Para un nivel de significación $\alpha$ (o un nivel de confianza $1-\alpha$) la amplitud del intervalo es:

$$A_\alpha = 2\, t_{n-1, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}$$

Si aumentamos el nivel de confianza a $1-\beta$, entonces $\beta < \alpha \Rightarrow t_{n-1,\frac{\beta}{2}} > t_{n-1,\frac{\alpha}{2}}$. La amplitud sería ahora $A_\beta = 2\, t_{n-1,\frac{\beta}{2}} > A_\alpha$. Podemos afirmar que:

Si aumenta el nivel de precisión, aumenta también la amplitud del intervalo de confianza con lo cual la precisión al estimar la media poblacional $\mu$ disminuye.

b3) Si tuviéramos la misma muestra inicial pero supiéramos que la desviación típica de 38'89 € es realmente la desviación típica poblacional.

Como la desviación típica poblacional se conoce ahora el intervalo de confianza de nivel $1-\alpha$ es:

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \,,\, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

La amplitud de este intervalo es:

$$A = 2\, z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < 2\, t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}} \quad \text{ya que } \sigma = S$$

y $z_{\frac{\alpha}{2}} < t_{n-1,\frac{\alpha}{2}}$

La amplitud del intervalo disminuye con lo que aumenta la precisión al estimar $\mu$.

8. Se han estudiado las tallas de los individuos de una cierta población normal a través de una m.a.s. de 40 de ellos. Los valores muestrales para la media y la desviación típica fueron respectivamente 1'68m y 8 cm. Obtenga mediante el método correcto un intervalo de confianza al 95% para la media poblacional. Haga lo mismo pero suponiendo que $n = 400$. Repita los dos ejercicios anteriores tomando la desviación típica muestral como si fuera el verdadero valor poblacional comentando y comparando los resultados entre sí. (Nota: Si los gl. de la t de Student no aparecen en la tabla, emplee los que más se aproximen)

a) No conocemos la varianza poblacional. El intervalo de confianza al 95% $(1-\alpha = 0'95 \Rightarrow \alpha = 0'05)$ de la media poblacional $\mu$ viene dado por

$$IC_{95\%}(\mu) = \left(\bar{x} - t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}, \bar{x} + t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}\right)$$

$\bar{x} = 1'68 m.$; $S = 0'08 m.$, $n = 40$ $\quad t_{n-1,\frac{\alpha}{2}} = t_{39,0'025} = 2'021$

$$IC_{95\%}(\mu) = \left(1'68 - 2'021 \cdot \frac{0'08}{\sqrt{39}}, 1'68 + 2'021 \cdot \frac{0'08}{\sqrt{39}}\right) = (1'65, 1'71)$$

b) Si es $n = 400$ permaneciendo todo lo demás igual:

$$IC_{95\%}(\mu) = \left(1'68 - 1'96 \cdot \frac{0'08}{\sqrt{399}}, 1'68 + 1'96 \cdot \frac{0'08}{\sqrt{339}}\right) = (1'67, 1'69)$$

se reduce la amplitud del intervalo con respecto al del apartado a) con lo que aumenta la precisión al estimar $\mu$.

c) Si suponemos conocida la desviación típica poblacional $\sigma = 0'08$, ahora el intervalo de confianza viene dado por:

$$IC_{95\%}(\mu) = \left(\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$IC_{95\%}(\mu) = \left(1'68 - z_{0'025} \cdot \frac{0'08}{\sqrt{40}}, 1'68 + z_{0'025} \cdot \frac{0'08}{\sqrt{40}}\right)$$

$$IC_{95\%}(\mu) = \left(1'68 - 1'96 \cdot \frac{0'08}{\sqrt{40}}, 1'68 + 1'96 \cdot \frac{0'08}{\sqrt{40}}\right) = (1'6552, 1'7048)$$

d) Si tomamos como tamaño de la muestra $n = 400$

$$IC_{95\%}(\mu) = \left(1'68 - 1'96 \cdot \frac{0'08}{20}, 1'68 + 1'96 \cdot \frac{0'08}{20}\right) = (1'6721, 1'6878)$$

9. Para investigar el coeficiente de inteligencia medio de cierta población estudiantil se propuso un test a 26 estudiantes. Los resultados obtenidos fueron $\bar{x} = 86$, y $S = 10'2$. Determine un intervalo de confianza a un nivel del 98% para la media poblacional suponiendo distribuciones normales.

Como la varianza poblacional es desconocida, el intervalo de confianza de nivel $1-\alpha$ es:

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}, \bar{x} + t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}\right)$$

$n = 26$, $\bar{x} = 86$, $S = 10'2$, $t_{25,0'01} = 2'485$

$$IC_{98\%}(\mu) = \left(86 - 2'485 \cdot \frac{10'2}{5}, 86 + 2'485 \cdot \frac{10'2}{5}\right) = (80'9306, 91'0694)$$

10. Se ha practicado un test a los alumnos de una universidad. Del total de ejercicios se ha escogido y analizado una muestra formada por 100 de ellos, que ha dado como resultado una cuasidesviación típica de 289 puntos. Si las puntuaciones siguen una ley normal y si se ha determinado un intervalo de confianza para la media, con límites en los valores 502 y 566, determine el nivel de confianza al que se ha hecho. Halle otro intervalo pero esta vez con un nivel de confianza del 98'4%

Como no conocemos la varianza muestral el intervalo de confianza con un nivel de significación $\alpha$ es:

$$IC_{1-\alpha}(\mu) = \left(\bar{x} - t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n-1}}, \bar{x} + t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}\right)$$

La amplitud de este intervalo es $A = 2\, t_{n-1,\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}$

$S_c = \sqrt{S_c^2} = 289 \Rightarrow S_c^2 = 289^2$. Como

$$S^2 = \frac{n-1}{n} S_c^2 \Rightarrow S^2 = \frac{99}{100} \cdot 289^2 \Rightarrow S = 289\sqrt{\frac{99}{100}} = 287'55$$

Si los extremos del intervalo son $(502, 566)$ entonces su amplitud es $A = 64$

$$64 = 2 \cdot t_{99, \frac{\alpha}{2}} \cdot \frac{287'55}{\sqrt{99}} \Rightarrow 64 = 2\, t_{99, \frac{\alpha}{2}} \cdot 28'9$$

$t_{99, \frac{\alpha}{2}} = 1'107271$. Es muy difícil encontrarlo en la t-Student. La t-Student se aproxima a una normal. Buscamos en $N(0,1)$ el valor $Z_{\frac{\alpha}{2}} = 1'107271$

$$\Rightarrow P[X > Z_{\frac{\alpha}{2}}] = \frac{\alpha}{2} \Rightarrow P[X \leq Z_{\frac{\alpha}{2}}] = 1 - \frac{\alpha}{2} \Rightarrow$$

$$\Rightarrow P[X \leq 1'11] = 1 - \frac{\alpha}{2} \Rightarrow 0'8665 = 1 - \frac{\alpha}{2}$$

$\frac{\alpha}{2} = 1 - 0'8665 \Rightarrow \frac{\alpha}{2} = 0'1335 \Rightarrow \alpha = 0'267$. Por tanto el nivel de confianza es $1 - \alpha = 0'733 = 73'3\%$

Vamos a calcular $\bar{x}$. Como:

$$502 = \bar{x} - t_{99, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}} \Rightarrow \bar{x} = 502 + t_{99, \frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n-1}}$$

$\bar{x} = 502 + 1'107271 \cdot 28'9 = 534$.

El intervalo de confianza al $98'4\%$ es:

$$IC_{98'4\%}(\mu) = (534 - t_{99, 0'008} \cdot 28'9,\ 534 + t_{99, 0'008} \cdot 28'9)$$

Como no podemos buscar en las tablas de t-Student el valor $t_{99, 0'008}$, buscamos $Z_{0'008}$, $P[X < Z_{0'008}] = 0'992$

$\Rightarrow Z_{0'008} = 2'41. \simeq t_{99, 0'008}$

Por tanto:

$$IC_{98'4\%}(\mu) = (534 - 2'41 \cdot 28'9,\ 534 + 2'41 \cdot 28'9) = (464'35, 603'64)$$

11. En una prueba de lanzamiento de un cierto tipo de producto informático se han obtenido los siguientes niveles de ventas en 6 mercados de prueba distintos e independientes: 99, 100, 98, 97, 96, 98. Supuesto que en todos los mercados las ventas siguen la misma distribución normal, determine un intervalo de confianza para la desviación típica, a un nivel del 95%, supuesto que:

a) La media es conocida y vale $\mu = 99$

b) La media es desconocida.

a) Como $X \rightsquigarrow N(99, \sigma)$, tomamos una m.a.s $X_1, X_2, \dots X_n$. Entonces por el lema de Fisher:

$$\bar{X} \rightsquigarrow N\left(99, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \frac{\bar{X}-99}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$$

Podemos tomar como función pivotal $v(\sigma) = \dfrac{\sqrt{n}(\bar{x}-99)}{\sigma}$

En nuestro caso $n = 6$ y $\bar{x} = 98$ por lo que podemos tomar como función pivotal:

$$v(\sigma) = \frac{-\sqrt{6}}{\sigma}$$

Queremos hallar un intervalo $(\lambda_1, \lambda_2)$ tal que

$$P[\lambda_1 < v(\sigma) < \lambda_2] = 0'95$$

Vamos a pivotar la inecuación $\lambda_1 < v(\sigma) < \lambda_2$ para acotar $\sigma$:

$$\lambda_1 < \frac{-\sqrt{6}}{\sigma} < \lambda_2 \Rightarrow \frac{1}{\lambda_2} < \frac{\sigma}{-\sqrt{6}} < \frac{1}{\lambda_1} \Rightarrow$$

$$\Rightarrow \boxed{\frac{-\sqrt{6}}{\lambda_1} < \sigma < \frac{-\sqrt{6}}{\lambda_2}}$$

La amplitud del intervalo es $\Delta = \dfrac{\sqrt{6}}{\lambda_1} - \dfrac{\sqrt{6}}{\lambda_2}$

La expresión $P[\lambda_1 < v(\sigma) < \lambda_2] = 0'95$ es equivalente a $F(\lambda_2) - F(\lambda_1) = 0'95$ siendo $F$ la función de distribución de $N(0,1)$

El problema que tenemos que resolver es:

$$\text{Min } A(d_1, d_2) = \frac{\sqrt{6}}{d_1} - \frac{\sqrt{6}}{d_2}$$

$$\text{s.a: } F(d_2) - F(d_1) = 0'95$$

La función de Lagrange es:

$$L(d_1, d_2, \lambda) = \frac{\sqrt{6}}{d_1} - \frac{\sqrt{6}}{d_2} - \lambda\left(F(d_2) - F(d_1) - 0'95\right)$$

Las condiciones de primer orden para la existencia de mínimo son:

$$\frac{\partial L}{\partial d_1} = -\frac{\sqrt{6}}{d_1^2} + \lambda f(d_1) = 0$$

$$\frac{\partial L}{\partial d_2} = \frac{\sqrt{6}}{d_2^2} - \lambda f(d_2) = 0$$

$$\frac{\partial L}{\partial \lambda} = F(d_2) - F(d_1) - 0'95 = 0$$

$f$ es la función de densidad de $N(0,1)$

obtenemos que $d_1^2 f(d_1) = d_2^2 \cdot f(d_2) = \frac{\sqrt{6}}{\lambda}$. Esta ecuación tiene, entre otras soluciones, $d_1 = d_2$ (no sirve ya que $d_1 < d_2$) $d_1 = -d_2$ (ya que $f$ es simétrica)

Entonces $F(d_2) - F(-d_2) = 0'95 \Rightarrow F(d_2) - (1 - F(d_2)) = 0'95$

$2F(d_2) = 1'95 \Rightarrow F(d_2) = 0'975 = 1 - \frac{\alpha}{2}$, $\alpha = 0'05$

Entonces $d_2 = z_{\frac{\alpha}{2}} = z_{0'025}$, $P[X \leq z_{0'025}] = 1 - 0'025 = 0'975$

Se tiene que $z_{\frac{\alpha}{2}} = z_{0'025} = 1'96$. Por tanto el intervalo de confianza es:

$$IC_{95\%}(\sigma) = \left( -\frac{\sqrt{6}}{+1'96}, \frac{-\sqrt{6}}{-1'96} \right)$$