LA INFERENCIA ESTADÍSTICA Y LAS TIC (2ª PARTE). ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Rosana Álvarez García (*) y Abel Martín (**)

El objetivo en muchos estudios estadísticos es conocer parámetros poblacionales. A la hora de estudiarlos nos encontramos con la imposibilidad de trabajar con todos los elementos de la población por su excesivo tamaño, falta de tiempo o elevado precio. Para solucionar este problema se trabaja con muestras extraídas de la población, como ya estudiamos en el artículo anterior.

ESTIMACIÓN PUNTUAL

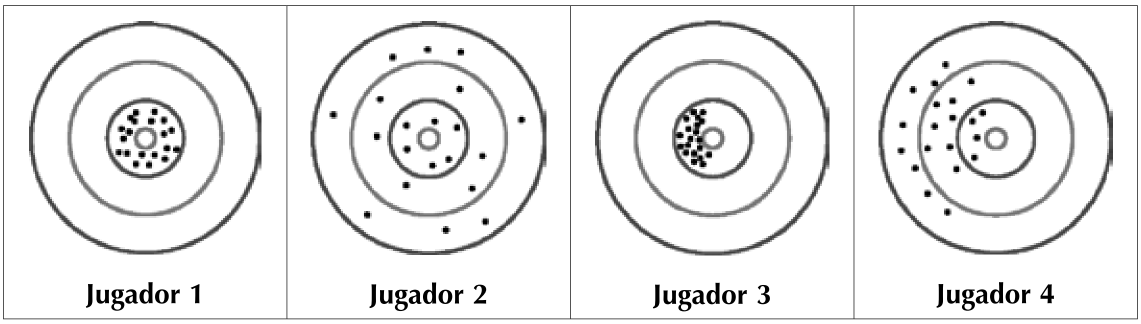
La estimación puntual asigna al parámetro poblacional un único valor. Para su cálculo se emplean los estimadores.

Un estimador es una función de los valores muestrales, es un estadístico que nos permite obtener valores aproximados de los parámetros poblacionales (media, varianza, mediana, moda, ...). El estimador es, por tanto, una variable aleatoria en el muestreo con su correspondiente función de distribución, como ya vimos para la media y proporción en el artículo del número anterior y que ampliaremos brevemente en éste.

Al estimar un parámetro poblacional podemos elegir distintos estimadores pero, ¿cuál es el más adecuado?

Consideremos el siguiente ejemplo.

Tenemos una partida de dardos con cuatro jugadores, cada jugador representa uno de los estimadores elegidos para conocer un parámetro poblacional. El centro de la diana se corresponde con el parámetro poblacional. Una vez realizados los lanzamientos de la partida obtenemos los siguientes resultados:



(*) Profesora de Tecnología del IES Aramo (Oviedo-Asturias).

(**) Profesor de Matemáticas del IES Pérez de Ayala (Oviedo-Asturias).

Analizando estos resultados obtenidos por cada jugador observamos:

Jugador 1 *(estimador 1):* las tiradas no se desvían en ninguna dirección y están muy concentradas.

Jugador 2 *(estimador 2):* las tiradas no se desvían en ninguna dirección y están muy dispersas.

Jugador 3 *(estimador 3):* las tiradas se desvían hacia la izquierda y están muy concentradas.

Jugador 4 *(estimador 4):* las tiradas se desvían hacia la izquierda y están muy dispersas.

Si traducimos estos resultados en términos de estimación podemos asociar cada tirada con una estimación muestral, es decir, las tiradas de cada jugador son las estimaciones realizadas en las muestras estudiadas.

Si el estimador es centrado, no tiene sesgo y lo llamamos insesgado.

Cuanto menor sea la dispersión más eficiente será el estimador. La eficiencia de un estimador centrado es la inversa de su varianza.

En la estimación puntual es necesario tener en cuenta el *error* que se comete. La distribución muestral nos da la distribución de los valores que toma el estimador en las distintas muestras seleccionadas de la población. La media y la varianza son los parámetros utilizados y la desviación típica o error típico de estimación nos da la desviación promedio entre el estimador y el parámetro poblacional.

Como ya vimos en el artículo del número anterior, un estimador es una función de la variable aleatoria que tomará un valor numérico para cada muestra. Teniendo en cuenta todas las muestras extraídas de la población obtenemos un conjunto de valores que siguen una determinada distribución muestral.

Vamos a repasar los estimadores puntuales más utilizados y su distribución en el muestreo.

DISTRIBUCIÓN MUESTRAL DE PROPORCIONES

Para muestras grandes el estimador $\hat{p}$ sigue una distribución normal de parámetros

$$\hat{p} \rightarrow N\left(p,\ \sqrt{\frac{p\cdot q}{n}}\right)$$

Si tipificamos la variable

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{p\cdot q}{n}}}\ N(0,\ 1)$$

DISTRIBUCIÓN MUESTRAL DE LA MEDIA

La distribución muestral de la media en una población normal, sigue una distribución normal de parámetros

$$\overline{X} \rightarrow N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right)$$

(a) Si la varianza poblacional σ es conocida.

Tipificamos la variable

$$Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,\ 1)$$

(b) Si la varianza poblacional σ es desconocida, la estimamos a través de la cuasivarianza

$$\hat{S}^2 = \frac{n}{n-1} \cdot S^2$$

Tipificamos la variable

$$\frac{\overline{X} - \mu}{\frac{\hat{S}}{\sqrt{n}}}$$

sigue una distribución t-Student con (n - 1) grados de libertad.

DISTRIBUCIÓN MUESTRAL DE LA DIFERENCIA DE MEDIAS

Dadas dos variables aleatorias independientes con distribución

$$X \rightarrow N(\mu_1, \sigma_1) \quad e \quad Y \rightarrow N(\mu_2, \sigma_2)$$

La distribución de la nueva variable diferencia de las anteriores $\overline{X} - \overline{Y}$ sigue una distribución normal de parámetros:

(a) Con varianzas poblaciones conocidas

La nueva variable aleatoria $\overline{X} - \overline{Y}$ tiene media $\mu_1 - \mu_2$ y desviación típica $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$. Para poblaciones de gran tamaño sigue una distribución normal de parámetros

$$\overline{X} - \overline{Y} \rightarrow N\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Si tipificamos la variable

$$Z = \left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \rightarrow N(0, 1)$$

(b) Con varianzas poblaciones desconocidas

Podemos distinguir dos casos:

(1) Con varianzas poblacionales desconocidas pero iguales: $\sigma_1 = \sigma_2 = \sigma$. El estadístico viene dado por

$$T = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\hat{S}}{n_1} + \frac{\hat{S}}{n_2}}}$$

siendo $\hat{S}^2$ el promedio de las cuasivarianzas

$$\hat{S} = \sqrt{\frac{(n_1 - 1) \cdot \hat{S}_1^2 + (n_2 + 1) \cdot \hat{S}_2^2}{n_1 + n_2 - 2}}$$

El estadístico final es: $$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot\sqrt{\frac{\left(n_1-1\right)\cdot\hat{S}_1^2+\left(n_2+1\right)\cdot\hat{S}_2^2}{n_1+n_2-2}}}$$

Distribución t-Student con $(n_1 + n_2 - 2)$ grados de libertad.

(2) Con varianzas poblacionales desconocidas y distintas: $\sigma_1 \neq \sigma_2 \neq \sigma$. El estadístico viene dado por

$$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}} \quad t_f$$

Sigue una distribución t-Student con **f** grados de libertad.

Los grados de libertad los calculamos por la aproximación de Welch

$$f = \frac{\left(\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{S}_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{\hat{S}_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

Tomaremos como grados de libertad el valor entero más próximo al resultado obtenido.

ESTIMACIÓN POR INTERVALOS DE CONFIANZA

La estimación puntual se utiliza poco, puesto que carecemos de datos que nos indiquen el grado de fiabilidad del dato muestral que hemos tomado. Es más útil utilizar la estimación por intervalos, que consiste en calcular dos valores que definen el intervalo en el cual estimamos que se encontrará el parámetro poblacional.

La probabilidad de obtener un intervalo a partir de las muestras analizadas que contenga el verdadero valor del parámetro poblacional se la denomina *nivel de confianza* $(1 - \alpha)$, siendo el *nivel de riesgo o de significación* (α) la probabilidad de que no esté contenido.

Si fijamos un nivel de confianza $(1-\alpha) = 0.9$, indica que de cada 100 muestras analizadas en 90 obtendremos un intervalo que contiene el verdadero valor del parámetro poblacional y en 10 obtendremos un intervalo en el que no se encuentra el parámetro poblacional.

Si nuestro intervalo es centrado podemos decir

$$-\lambda_\alpha \leq T \leq \lambda_\alpha$$

Siendo el valor de λ_α la probabilidad de que el valor absoluto de la distribución que sigue el estadístico sea igual al nivel de confianza prefijado de antemano.

Vamos a construir los intervalos de confianza para los estimadores estudiados:

Intervalo de confianza para la media de una población N(μ, σ) con varianza conocida

Buscamos el intervalo de confianza para la media de una variable X. Como vimos, la distribución muestral de la media en una población N(μ, σ), sigue una distribución normal de parámetros

$$\overline{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

El intervalo estará comprendido entre dos valores dados por la distribución del estadístico en el muestreo al nivel de confianza prefijado

$$-\lambda_\alpha \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \lambda_\alpha$$

Despejamos el parámetro buscado

$$-\lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \overline{X}-\mu \leq \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

$$-\lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \overline{X} \leq -\mu \leq \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}} - \overline{X}$$

Multiplicamos por (- 1) y nos queda

$$+\lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}} + \overline{X} \geq \mu \geq -\lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}} + \overline{X}$$

$$\overline{X} - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

λ_α *tendrá en cuenta el nivel de significación y* $\frac{\sigma}{\sqrt{n}}$ *el error debido al muestreo.*

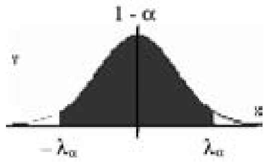
Para calcular el valor de λ_α necesitamos conocer:

(a) La distribución muestral del estadístico, en este caso

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0, 1)$$

(b) El nivel de confianza prefijado $(1 - \alpha)$

$$\lambda_\alpha \,/\, P[|N(0, 1)| \leq \lambda_\alpha] = 1 - \alpha$$



Nuestro intervalo es

$$\left[\overline{X} - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}},\ \ \overline{X} + \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right]$$

ACTIVIDAD 1

En un estudio sobre la estancia media de los pacientes en un hospital se toma una muestra de aleatoria de 30 pacientes y se anota el tiempo de permanencia de cada uno de ellos obteniéndose una media de 8 días. La variable en estudio sigue una distribución normal con varianza poblacional 16. A un nivel de confianza $1 - \alpha = 0.95$, calcula un intervalo de confianza para la media.

Resolución

DATOS:

POBLACIÓN	μ	$\sigma2 = 16$ $\sigma = 4$		$1 - \alpha = 0.95$
MUESTRA	$\bar{x} = 8$	S	n = 30	

La distribución muestral de la media en una población N(μ, σ) con varianza conocida es:

$$\overline{X} \to N\left(\mu, \ \frac{\sigma}{\sqrt{n}}\right)$$

Tipificamos la variable $$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0, 1)$$

Nuestro intervalo centrado es

$$\left[\overline{X} - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \quad \overline{X} + \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right]$$

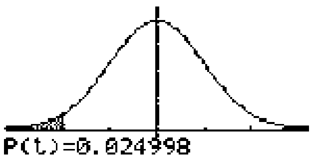
El valor de λ_α / $P[|N(0, 1)| \leq \lambda_\alpha] = 1 - \alpha$ $\qquad P[|N(0, 1)| \leq \lambda_\alpha] = 0.95$

Método I

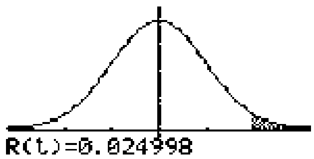
Aplicamos la definición de valor absoluto

$P[|N(0, 1)| \leq \lambda_\alpha] = P[-\lambda_\alpha \leq N(0, 1) \leq \lambda_\alpha] = P[N(0, 1) \leq \lambda_\alpha] - P[N(0, 1) \leq -\lambda_\alpha] = 0.95$

La tabla N(0, 1) sólo nos da valores positivos. Como es simétrica tenemos que

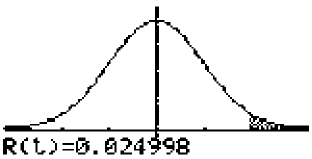


=

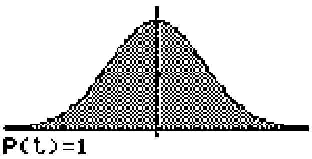


$P[N(0, 1) \leq -\lambda_\alpha]$ = $P[N(0, 1)\ \lambda_\alpha]$

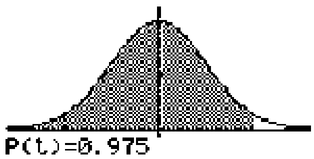
El área total de la curva es la unidad y por tanto



=



-



$P[N(0, 1)\ \lambda_\alpha]$ = 1 - $P[N(0, 1) \leq \lambda_\alpha]$

Sustituímos

$$P[N(0, 1) \leq \lambda_\alpha] - P[N(0, 1) \leq -\lambda_\alpha] = P[N(0, 1) \leq \lambda_\alpha] - \{1 - P[N(0, 1) \leq \lambda_\alpha]\} =$$

$$= P[N(0, 1) \leq \lambda_\alpha] - 1 + P[N(0, 1) \leq \lambda_\alpha] = 2 \cdot P[N(0, 1) \leq \lambda_\alpha] - 1$$

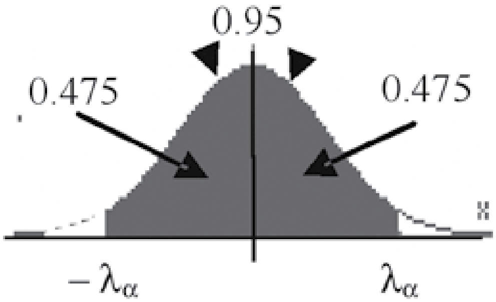
$$2 \cdot P[N(0, 1) \leq \lambda_\alpha] - 1 = 0.95$$

$$2 \cdot P[N(0,1) \le \lambda_\alpha] = 0.95 + 1$$

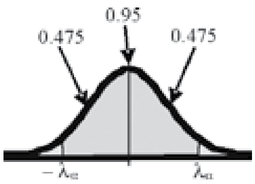
$$P[N(0,1) \le \lambda_\alpha] = (0.95 + 1)/2$$

$$P[N(0,1) \le \lambda_\alpha] = 0.975$$

Método II



Operamos el valor absoluto, buscamos en la tabla de la distribución N(0, 1) y tenemos

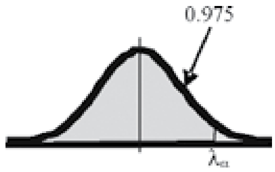


$$P[N(0,1) \le \lambda_\alpha] = 0.975$$

$$\frac{0.95}{2} = 0.475 \quad \rightarrow \quad 0.475 + 0.5 = 0.975$$

El motivo por el que se pone de área 0.975 es porque la superficie buscada para λ_α *es*

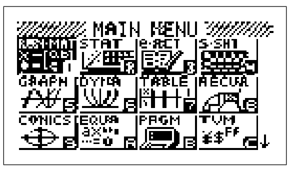
0.5 + 0.475



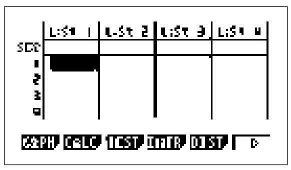
Buscamos en la tabla de la N(0, 1) el valor de la variable que tiene como probabilidad 0.975, por varios métodos:

(a) A través de una opción directa que tiene la calculadora gráfica:

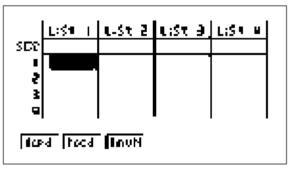
AC/ON



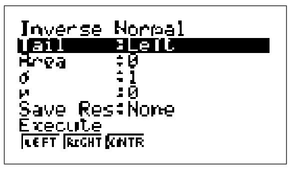
2



F5 F1

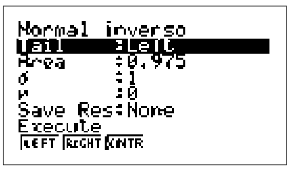


F3

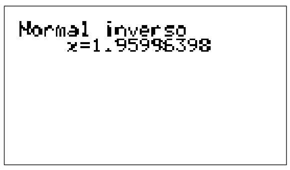


F1

Introducimos los valores que aparecen al margen

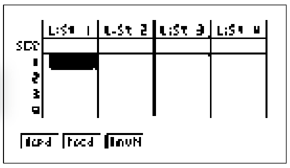


F1



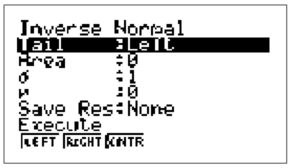
(b) A través de otra opción directa que tiene la calculadora gráfica:

MENU 2 F5 F1



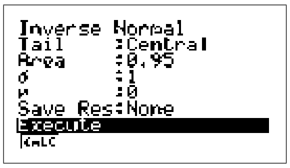
F3

Opción ***CNTR***



F3

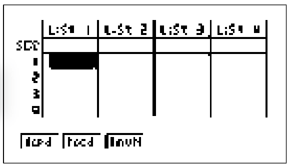
Introducimos los valores que aparecen al margen



F1

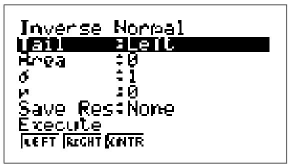
(c) Mediante otra opción directa que tiene la calculadora gráfica:

MENU 2 F5 F1



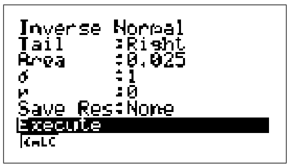
F3

Opción ***RIGHT***

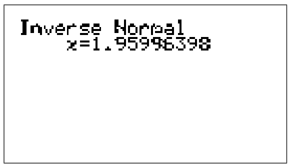


F2

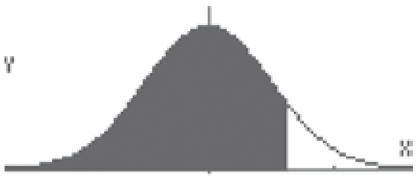
Introducimos los valores que aparecen al margen



F1



(d) Con la ayuda de las tablas de la N(0, 1), buscamos el valor en el que $F(\lambda_\alpha) = 0.975$, "por dentro" de la tabla.



0.975

$P(Z \leq \lambda_\alpha) = 0.975$

$F(\lambda_\alpha) = 0.975$

Tablas $F(1.96) = 0.975$

$\lambda_\alpha = 1.96$

(e) A partir de este momento, mentalmente, tenemos que memorizar este número, ya que es uno de los niveles de confianza más habitualmente utilizados.

Sustituimos en el intervalo:

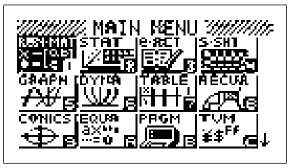
$$\left[8 - 1.96 \cdot \frac{4}{\sqrt{30}},\quad 8 + 1.96 \cdot \frac{4}{\sqrt{30}}\right] = [6.5686,\ 9.4314]$$

La media muestral se encuentra entre esos valores a un nivel de confianza del 95%.

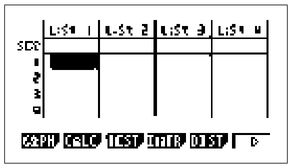
Los valores del *intervalo* también los podemos averiguar directamente con la calculadora gráfica, pudiendo dedicar más tiempo a la reflexión y a la compresión de conceptos, dejando los cálculos numéricos y farragosos para las máquinas.

(a) A través de una opción directa que tiene la calculadora gráfica:

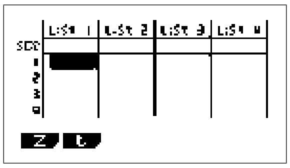
MENU



2

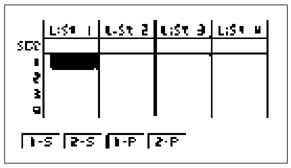


F4



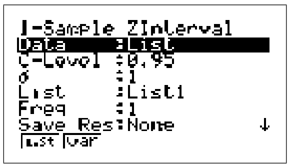
F1

Nuestro estadístico sigue una distribución normal, Z



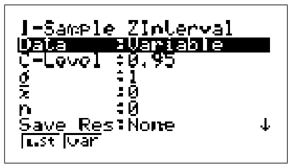
F1

Trabajamos con una ***muestra*** *y buscamos el intervalo para la* ***media*** *1-s. Si trabajásemos con proporciones 1-p*

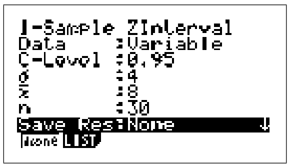


F2

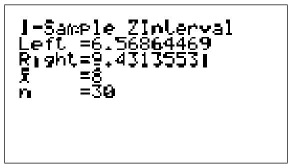
Seleccionamos la opción ***Var****, conocemos los parámetros muestrales*



Introducimos los valores que aparecen a la derecha



EXE



[6.5686, 9.4314]

Intervalo de confianza para la media de una población N(μ, σ) con varianza desconocida

Cuando σ es desconocida ya vimos que

$$\frac{\overline{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} \to t_{n-1}$$

S^2: Cuasivarianza; si n es suficientemente grande podemos aproximar a una N(0, 1).

El intervalo estará comprendido entre dos valores dados por la distribución del estadístico en el muestreo, en nuestro caso una *t de Student* con (n – 1) grados de libertad.

$$-t_{\alpha,\, n-1} \cdot < \frac{\overline{X}-\mu}{\frac{\hat{S}}{\sqrt{n}}} < t_{\alpha,\, n-1}$$

Despejamos μ y nos queda

$$-t_{\alpha,\, n-1} \cdot \frac{\hat{S}}{\sqrt{n}} < \overline{X} - \mu < t_{\alpha,\, n-1} \cdot \frac{\hat{S}}{\sqrt{n}}$$

$$-t_{\alpha,\, n-1} \cdot \frac{\hat{S}}{\sqrt{n}} < \overline{X} - \mu < t_{\alpha,\, n-1} \cdot \frac{\hat{S}}{\sqrt{n}} - \overline{X}$$

Multiplicamos por (- 1) y cambia el sentido de la desigualdad

$$\overline{X} - t_{\alpha,\, n-1} \cdot \frac{\hat{S}}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha,\, n-1} \cdot \frac{\hat{S}}{\sqrt{n}}$$

Nuestro intervalo es

$$\left[\overline{X} - t_{\alpha,n-1} \cdot \frac{\hat{S}}{\sqrt{n}}, \overline{X} + t_{\alpha,n-1} \cdot \frac{\hat{S}}{\sqrt{n}}\right]$$

Siendo $t_{\alpha,\, n-1}$ / $P(|t_{n-1}| < t_\alpha) = 1 - \alpha$

Si las muestras son suficientemente grandes, aproximamos la distribución del estadístico a una distribución N(0, 1).

$$\left[\overline{X} - \lambda_\alpha \cdot \frac{\hat{S}}{\sqrt{n}},\ \overline{X} + \lambda_\alpha \cdot \frac{\hat{S}}{\sqrt{n}}\right]$$

ACTIVIDAD 2

El nivel de protrombina en sangre se distribuye normalmente. Se toma una muestra aleatoria simple de 40 pacientes que se sabe tienen deficiencia en vitamina K. Los resultados obtenidos para esta muestra son, un nivel medio de 18.5 mg / 100ml de plasma y desviación 4 mg / 100 ml de plasma. Calcular un intervalo de confianza para la media a un nivel de confianza del 95%.

(a) Resolución con lápiz, papel y tablas:

DATOS:

POBLACIÓN	μ	σ		$1 - \alpha = 0.95$
MUESTRA	$\bar{x} = 18.5$	S = 4	n = 40	

Debemos calcular el intervalo de confianza para la media de una población N(μ, σ) con varianza desconocida. El intervalo pedido es

$$\left[\overline{X} - t_{\alpha,n-1} \cdot \frac{\hat{S}}{\sqrt{n}},\ \overline{X} + t_{\alpha,n-1} \cdot \frac{\hat{S}}{\sqrt{n}}\right]$$

Conocemos la media y desviación muestral. A partir de la desviación muestral calculamos la cuasivarianza

$$\hat{S}^2 = \frac{n}{n-1} \cdot S^2 = \frac{40}{39} \cdot 16 = 16.41 \rightarrow \hat{S} = 4.051$$

$$\left[18.5 - t_{\alpha,\, n-1} \cdot \frac{4.051}{\sqrt{40}}, 18.5 + t_{\alpha,n-1} \cdot \frac{4.051}{\sqrt{40}}\right]$$

Siendo $t_{\alpha,\, n-1}$, n-1 / $P(|t_{n-1}| < t_\alpha) = 1 - \alpha$

Buscamos en la tabla t-Student y obtenemos el valor de t_α

$P(|t_{39}| < t_\alpha) = 0.95 \rightarrow \qquad t_\alpha = 2.023$

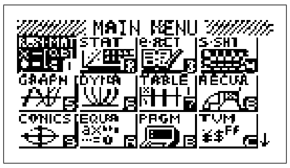
El intervalo pedido es

$$\left[18.5 - 2.023 \cdot \frac{4.051}{\sqrt{40}},\ 18.5 + 2.023 \cdot \frac{4.051}{\sqrt{40}}\right]$$

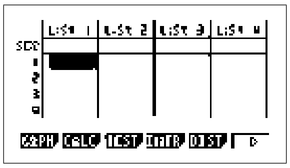
[17.204, 19.79]

(b) A través de una opción directa que tiene la calculadora gráfica:

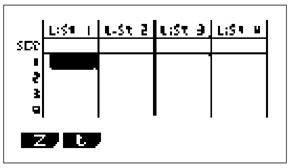
MENU



2

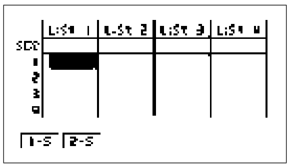


F4



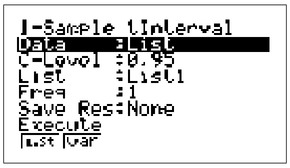
F2

Nuestro estadístico sigue una distribución t-Student



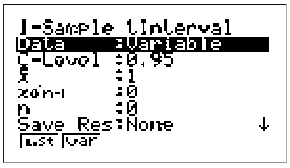
F1

Trabajamos con una ***muestra*** *y buscamos el intervalo para la* ***media*** *1-s.*

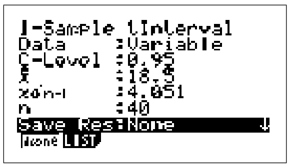


F2

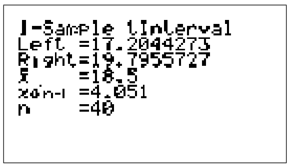
Seleccionamos la opción ***Var****, conocemos los parámetros muestrales*



Introducimos los datos que aparecen a la derecha



EXE



[17.2044273, 19.7955727]

Intervalo de confianza para la proporción de una población N(μ, σ) con varianza desconocida

Para muestras grandes el estimador $\hat{p}$ sigue una Distribución Normal de parámetros

$$\hat{p} \rightarrow N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$$

Si tipificamos la variable

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} \rightarrow N(0, 1)$$

Para conseguir el intervalo de confianza para la proporción situamos el estadístico entre dos valores con un nivel de confianza $1 - \alpha$ prefijado

$$- \lambda_\alpha \leq \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} \leq \lambda_\alpha$$

Si la muestra es grande tomamos como valor aproximado a p el valor.

Despejamos el parámetro buscado

$$- \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \leq \hat{p} - p \leq \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

$$- \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} - \hat{p} \leq - p \leq \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} - \hat{p}$$

Multiplicamos por -1 y nos queda

$$\hat{p} - \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \leq p \leq \hat{p} + \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}$$

Para calcular el valor de λ_α necesitamos conocer la distribución muestral del estadístico, en este caso

$$\frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}} \to N(0, 1)$$

Y el nivel de confianza prefijado $(1 - \alpha)$

$$\lambda_\alpha \,/\, P[|N(0, 1)| \leq \lambda_\alpha] = 1 - \alpha$$

Nuestro intervalo es

$$\left[\hat{p} - \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \;,\; \hat{p} + \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right]$$

ACTIVIDAD 3

En un población con distribución $N(\mu, \sigma)$ se toma una muestra de 100 unidades de un determinado producto y se observan 15 defectuosas. Calcular el intervalo de confianza para la proporción de unidades defectuosas a un nivel de confianza del 99%.

(a) Resolución con lápiz, papel y tablas:

La proporción muestral de productos defectuosos es

$$P(\text{defectuoso}) = 15/100 = 0.15 \to q = 0.85$$

El intervalo para la proporción viene dada por

$$\left[\hat{p} - \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} \;,\; \hat{p} + \lambda_\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right]$$

$$\left[0.15 - \lambda_\alpha \cdot \sqrt{\frac{0.15 \cdot 0.85}{100}} \;,\; 0.15 + \lambda_\alpha \cdot \sqrt{\frac{0.15 \cdot 0.85}{100}}\right]$$

El valor de λ_α

Siendo $\lambda_\alpha \,/\, P[|N(0, 1)| \leq \lambda_\alpha] = 1 - \alpha \to P[|N(0, 1)| \leq \lambda_\alpha] = 0.99$

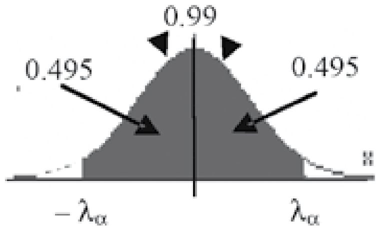
Método I

Como ya vimos antes en el desarrollo del valor absoluto en la distribución N(0, 1)

$$P[|N(0, 1)| \leq \lambda_\alpha] = 1 - 2 \cdot P[N(0, 1) \leq \lambda_\alpha] = 0.99$$

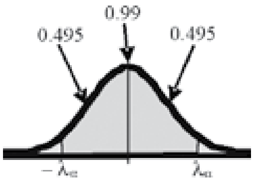
$$P[N(0, 1) \leq \lambda_\alpha] = \frac{0.99 + 1}{2} = 0.995$$

Método II



Operamos el valor absoluto, buscamos en la tabla de la distribución N(0, 1) y tenemos

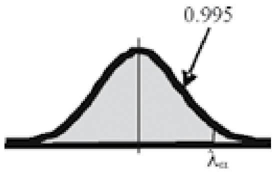
$$P[N(0, 1) \leq \lambda_\alpha] = 0.995$$



$$\frac{0.99}{2} = 0.495 \quad \rightarrow \quad 0.495 + 0.5 = 0.995$$

El motivo por el que se pone de área 0.995 es porque la superficie buscada para λ_α es

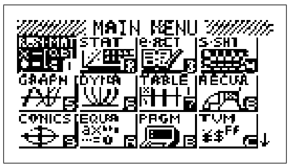
0.5 + 0.495



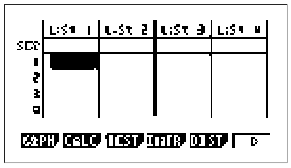
Buscamos en la tabla de la N(0, 1) el valor de la variable que tiene como probabilidad 0.975, por varios métodos (desechamos alguno ya explicado anteriormente):

(a) A través de una opción directa que tiene la calculadora gráfica:

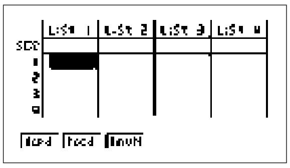
AC/ON



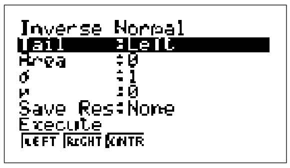
2



F5 F1

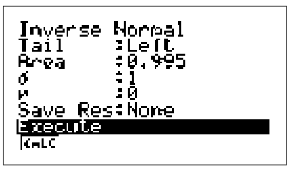


F3

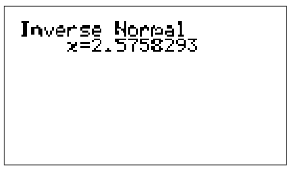


F1

Introducimos los valores que aparecen al margen

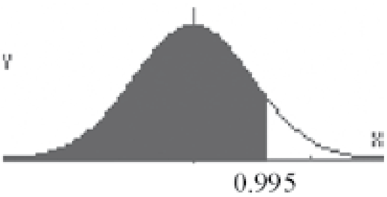


F1



(b) A partir de este momento tenemos que *memorizar* este número, ya que es uno de los niveles de confianza más habitualmente utilizados.

(c) Con la ayuda de las tablas de la N(0, 1), buscamos el valor en el que $F(\lambda_\alpha) = 0.995$, "por dentro" de la tabla.



$$P(Z \leq \lambda_\alpha) = 0.995$$

$$F(\lambda_\alpha) = 0.995$$

Tablas interpolando F(2.5758) = 0.995

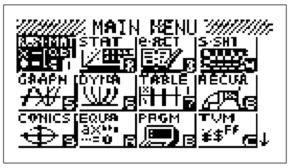
$$\lambda_\alpha = 2.5758$$

$$\left[0.15 - 2.5758293\cdot\sqrt{\frac{0.15\cdot 0.85}{100}}\ ,\quad 0.15 + 2.5758293\cdot\sqrt{\frac{0.15\cdot 0.85}{100}}\right]$$

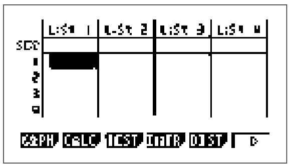
[0.05802449, 0.2419755]

(d) A través de una opción directa que tiene la calculadora gráfica:

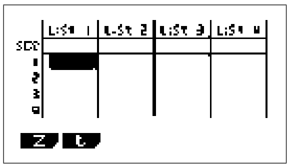
MENU



2

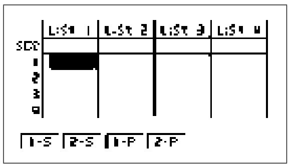


F4



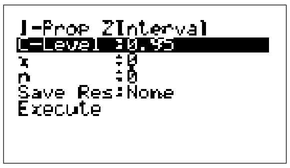
F1

Nuestro estadístico sigue una distribución normal, Z



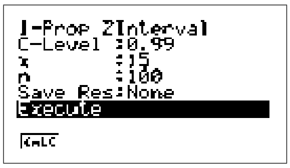
F3

*Trabajamos con una **muestra** y buscamos el intervalo para la **media** 1-p*

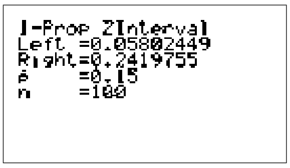


F2

Introducimos los datos que aparecen a la derecha



EXE



[0.05802449, 0.2419755]

Intervalo de confianza para la diferencia de medias

(a) Con varianzas poblaciones conocidas:

Si tenemos dos poblaciones que siguen distribuciones N(μ_1, σ_1) y N(μ_2, σ_2), con σ_1 y σ_2 conocidas. El estadístico de la diferencia muestral de medias sigue la distribución

$$\overline{X} - \overline{Y} \rightarrow N\left(\mu_1 - \mu_2,\ \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Si tipificamos la variable

$$Z = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow N(0,\ 1)$$

El intervalo de confianza a un nivel de confianza prefijado $(1 - \alpha)$ será

$$-\lambda_\alpha \leq \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \leq \lambda_\alpha$$

Despejamos el parámetro buscado

$$-\lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leq (\overline{X} - \overline{Y}) - (\mu_1 - \mu_2) \leq \lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$-\lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leq (\overline{X} - \overline{Y}) - (\mu_1 - \mu_2) \leq \lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} - (\overline{X} - \overline{Y})$$

Multiplicamos por (-1) y nos queda

$$(\overline{X} - \overline{Y}) - \lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\overline{X} - \overline{Y}) + \lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Para calcular el valor de λ_α necesitamos conocer la distribución muestral del estadístico, en este caso

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \rightarrow N(0,\ 1)$$

λ_α al nivel de confianza prefijado $(1 - \alpha)$ es:

$$\lambda_\alpha \ /\ P[|N(0,\ 1)| \leq \lambda_\alpha] = 1 - \alpha$$

Nuestro intervalo es

$$\left[(\overline{X}-\overline{Y})-\lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\quad (\overline{X}-\overline{Y})+\lambda_\alpha \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

(b) Con varianzas poblaciones desconocidas:

Si tenemos dos poblaciones que siguen distribuciones $N(\mu_1,\ \sigma_1)$ y $N(\mu_2,\ \sigma_2)$, con σ_1 y σ_2 desconocidas, debemos considerar los dos casos ya vistos al estudiar la distribución en el muestreo:

(1) Con varianzas poblacionales desconocidas pero iguales: $\sigma_1 = \sigma_2 = \sigma$. El estadístico viene dado por

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2+1)\cdot \hat{S}_2^2}{n_1+n_2-2}}} \quad t_{n_1+n_2-2}$$

Sigue una distribución *t-Student* con $(n_1 + n_2 - 2)$ grados de libertad.

El intervalo de confianza a un nivel de confianza prefijado $(1 - \alpha)$ será

$$-t_\alpha \leq \frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2+1)\cdot \hat{S}_2^2}{n_1+n_2-2}}} \leq t_\alpha$$

Despejamos el parámetro buscado

$$-t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}} \leq (\overline{X} - \overline{Y}) - (\mu_1 - \mu_2) \leq t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}}$$

$$-t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}} \le (\overline{X} - \overline{Y}) - (\mu_1 - \mu_2) \le \; t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}} - (\overline{X} - \overline{Y})$$

Multiplicamos por (- 1) y nos queda

$$(\overline{X} - \overline{Y}) - t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}} - (\overline{X} - \overline{Y}) \le (\mu_1 - \mu_2) \le (\overline{X} - \overline{Y}) + t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}}$$

Para calcular el valor de λ_α necesitamos conocer la distribución muestral del estadístico, en este caso

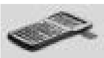
$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2+1)\cdot \hat{S}_2^2}{n_1+n_2-2}}} \to t_{n1\,+\,n2\,-\,2}$$

t_α al nivel de confianza prefijado (1 - α) es

Siendo $t_{\alpha\,,\,n1\,+\,n2\,-\,2}$ / $P(|\, t_{n1\,+\,n2\,-\,2} \,| < t_\alpha) = 1 - \alpha$

Nuestro intervalo es

$$\left[(\overline{X} - \overline{Y}) - t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}},\;\; (\overline{X} - \overline{Y}) + t_\alpha \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \cdot \sqrt{\frac{(n_1-1)\cdot \hat{S}_1^2+(n_2-1)\cdot \hat{S}_2^2}{n_1+n_2-2}} \right]$$



NOTA: en la calculadora gráfica la *homecedasticidad* debe indicarse por lo que se activa la opción *"Pooled"* (ON); se dice que el agrupamiento se encuentra *"en efecto"*.

(2) Con varianzas poblacionales desconocidas y distintas: $\sigma_1 \neq \sigma_2 \neq \sigma$. El estadístico viene dado por

$$\frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}} \to t_f$$

Sigue una distribución *t-Student* con *f* grados de libertad.

El intervalo de confianza en el mismo que en el caso anterior. Cambian los grados de libertad, que se calculan utilizando la aproximación de Welch, tomando el valor entero más próximo al resultado obtenido

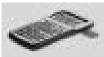
$$f = \frac{\left(\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{S}_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{\hat{S}_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

t_α al nivel de confianza prefijado (1 - α) es

Siendo $t_{\alpha,\,f}$ / $P(|t_f| < t_\alpha) = 1 - \alpha$

Nuestro intervalo es

$$\left[(\overline{X}-\overline{Y})-t_\alpha\cdot\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}},\ (\overline{X}-\overline{Y})+t_\alpha\cdot\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}\right]$$



NOTA: en la calculadora gráfica se dice que el agrupamiento NO se encuentra *"en efecto"* por lo que se desactiva la opción *"Pooled"* (OFF). Para calcular los grados de libertad utiliza otro tipo de aproximación y es capaz de manejar un número de grados de libertad *"no enteros"*, por lo que los límites del intervalo no salen exactamente iguales a los obtenidos con lápiz y papel.

ACTIVIDAD 4

En el estudio de la edad media en el comienzo de la reproducción de una determinada especie, se han considerado dos series procedentes de diferentes cruces genéticos con 7 y 8 individuos respectivamente, obteniéndose:

Serie I	$\overline{x} = 7.514$	$S_1 = 0.657$
Serie II	$\overline{x} = 7.5$	$S_2 = 0.5744$

Supuesto que la variable en estudio sigue una distribución normal, determina el intervalo de confianza al 99% para:

- La diferencia de medias poblacionales si las varianzas poblacionales son respectivamente 0.51^2 y 0.33^2.
- La diferencia de medias poblacionales si las varianzas poblacionales son desconocidas e iguales.
- La diferencia de medias poblacionales si las varianzas poblacionales son desconocidas y distintas.

(a) Resolución con lápiz, papel y tablas:

Intervalo de confianza para la diferencia de medias con varianzas poblacionales conocidas. El intervalo es

$$\left[(7.514-7.5)-\lambda_\alpha\cdot\sqrt{\frac{0.51^2}{7}+\frac{0.33^2}{8}},\ (7.514-7.5)+\lambda_\alpha\cdot\sqrt{\frac{0.51^2}{7}+\frac{0.33^2}{8}}\right]$$

Siendo λ_α / $P[|N(0,1)| \le \lambda_\alpha] = 1-\alpha$

$P[|N(0,1)| \le \lambda_\alpha] = 0.99$

 $\lambda_\alpha = \mathbf{2.5758}$

El intervalo es

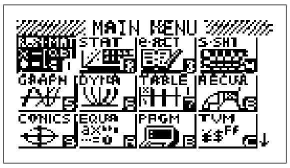
$$\left[(7.514-7.5)-2.5758293\cdot\sqrt{\frac{0.51^2}{7}+\frac{0.33^2}{8}},\ (7.514-7.5)+2.5758293\cdot\sqrt{\frac{0.51^2}{7}+\frac{0.33^2}{8}}\right]$$

El intervalo para la diferencia de medias en una poblacional con varianzas poblacionales conocidas

[- 0.5663889, 0.59438894]

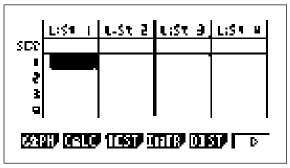
(b) A través de una opción directa que tiene la calculadora gráfica:

MENU

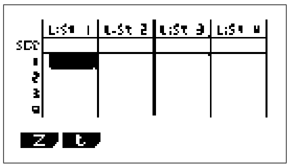


2

Opción ***CNTR***

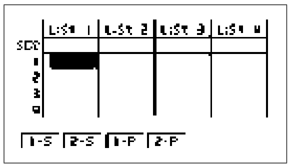


F4



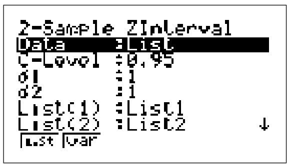
F1

Nuestro estadístico sigue una distribución normal Z



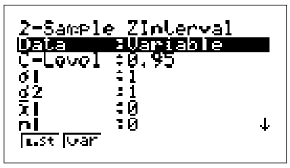
F3

Trabajamos con dos muestras y buscamos el intervalo para la media ***2-s***



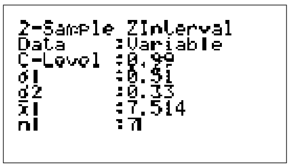
F2

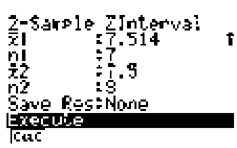
Seleccionamos la opción ***Var****; conocemos los parámetros muestrales*



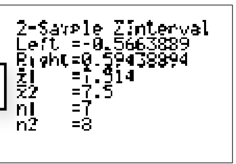
EXE

Introducimos los datos que aparecen a la derecha





EXE



[- 0.5663889, 0.59438894]

(c) Resolución con lápiz, papel y tablas:

Intervalo de confianza para la diferencia de medias con varianzas poblacionales desconocidas pero iguales.

Nuestro intervalo es

$$\left[(\bar{X}-\bar{Y})-t_\alpha\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot\sqrt{\frac{(n_1-1)\cdot\hat{S}_1^2+(n_2-1)\cdot\hat{S}_2^2}{n_1+n_2-2}},\ (\bar{X}-\bar{Y})+t_\alpha\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot\sqrt{\frac{(n_1-1)\cdot\hat{S}_1^2+(n_2-1)\cdot\hat{S}_2^2}{n_1+n_2-2}}\right]$$

Las cuasivarianzas muestrales son

$$\text{Serie I: } \hat{S}_1^2=\frac{n}{n-1}\cdot S^2 = 7/6\cdot 0.657^2 = 0.50359 \rightarrow \hat{S}_1 = 0.7096$$

$$\text{Serie II: } \hat{S}_2^2=\frac{n}{n-1}\cdot S^2 = 8/7\cdot 0.5744^2 = 0.37706 \rightarrow \hat{S}_2 = 0.6141$$

Sustituimos

$$\left[(7.514-7.5)-t_\alpha\cdot\sqrt{\frac{1}{7}+\frac{1}{8}}\cdot\sqrt{\frac{6\cdot 0.50359+7\cdot 0.37706}{7+8-2}},\ (70514-7.5)+t_\alpha\cdot\sqrt{\frac{1}{7}+\frac{1}{8}}\cdot\sqrt{\frac{6\cdot 0.50359+7\cdot 0.37706}{7+8-2}}\right]$$

Siendo $t_{\alpha\ ,\ n1\ +\ n2\ -\ 2}$ / $P(|\ t_{n1}+{}_{n2}-{}_2\ |< t_\alpha) = 1 - \alpha$

$P(|t_{13}|< t_\alpha) = 0.99$

Buscamos en la tabla *t-Student* y obtenemos t_α

$t_\alpha = 3.012$

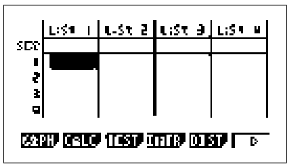
$$\left[(7.514-7.5)-3.012\sqrt{\frac{1}{7}+\frac{1}{8}}\cdot\sqrt{\frac{6\cdot 0.50359+7\cdot 0.37706}{7+8-2}},\ (70514-7.5)+3.012\sqrt{\frac{1}{7}+\frac{1}{8}}\cdot\sqrt{\frac{6\cdot 0.50359+7\cdot 0.37706}{7+8-2}}\right]$$

El intervalo para la diferencia de medias en una población con varianzas poblacionales desconocidas e iguales es:

[-1.014679181, 1.042679181]

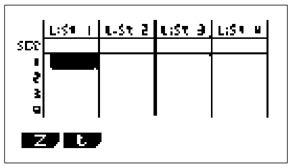
(d) A través de una opción directa que tiene la calculadora gráfica:

MENU **2**



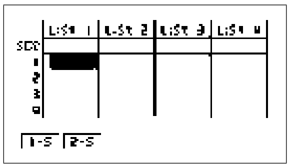
F4

*Opción **CNTR***



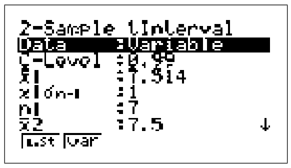
F2

*Nuestro estadístico sigue una distribución **t-Student***

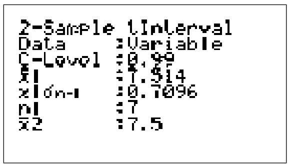


F2

*Trabajamos con dos muestras y buscamos el intervalo para la media **2-s***

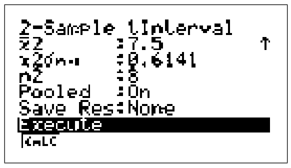


*Seleccionamos la opción **Var**, conocemos los estadísticos muestrales. Introducimos los datos que aparecen a la derecha*

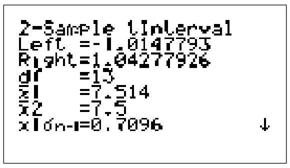


▼ ... ▼

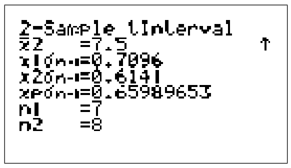
(Pooled: On)



EXE



▼ ... ▼



[-1.0147793, 1.04277926]

(e) Resolución con lápiz, papel y tablas:

Intervalo de confianza para la diferencia de medias con varianzas poblacionales desconocidas y distintas.

Nuestro intervalo es $\left[(\bar{X}-\bar{Y})-t_\alpha\cdot\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}},\ (\bar{X}-\bar{Y})+t_\alpha\cdot\sqrt{\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}}\right]$

Las cuasivarianzas muestrales, ya calculadas en el apartado anterior son

Serie I: $\hat{S}_1^2=\frac{n}{n-1}\cdot S^2$ = 7/6 · 0.657^2 = 0.50359 → $\hat{S}_1$ = 0.7096

Serie II: $\hat{S}_2^2=\frac{n}{n-1}\cdot S^2$ = 8/7 ·0.5744^2 = 0.37706 → $\hat{S}_2$ = 0.6141

Sustituimos

$$\left[(7.514-7.5)-t_\alpha\cdot\sqrt{\frac{0.5036}{7}+\frac{0.377}{8}},\ (7.514-7.5)+t_\alpha\cdot\sqrt{\frac{0.5036}{7}+\frac{0.377}{8}}\right]$$

Siendo $t_{\alpha,\ f}$ / P(|t_f|< t_α) = 1 - α

Calculamos los grados de libertad por la aproximación de Welch

$$f = \frac{\left(\frac{\hat{S}_1^2}{n_1}+\frac{\hat{S}_2^2}{n_2}\right)^2}{\frac{\left(\frac{\hat{S}_1^2}{n_1}\right)^2}{n_1+1}+\frac{\left(\frac{\hat{S}_2^2}{n_2}\right)^2}{n_2+1}}-2 \rightarrow f = \frac{\left(\frac{0.5036}{7}+\frac{0.377}{8}\right)^2}{\frac{\left(\frac{0.5036}{7}\right)^2}{8}+\frac{\left(\frac{0.377}{8}\right)^2}{9}}-2 = 13.86 \approx 14$$

Buscamos en la tabla t-Student y obtenemos t_α

$$P(|t_{13}| < t_\alpha) = 0.99$$

$$t_\alpha = 2.977$$

$$\left[(7.514-7.5)-2.977\cdot\sqrt{\frac{0.5036}{7}+\frac{0.377}{8}},\ (7.514-7.5)+2.977\cdot\sqrt{\frac{0.5036}{7}+\frac{0.377}{8}}\right]$$

El intervalo para la diferencia de medias en una poblacional con varianzas poblacionales desconocidas y distintas es:

[- 1.01327, 1.041276]

(f) A través de una opción directa que tiene la calculadora gráfica:

DATOS (Pooled: Off)		RESULTADOS	
2-Sample tInterval Data :Variable C-Level :0.99 $\bar{x}1$:7.514 x1σn-1 :0.7096 n1 :7 $\bar{x}2$:7.5	2-Sample tInterval $\bar{x}2$:7.5 x2σn-1 :0.6141 n2 :8 Pooled :Off Save Res:None Execute CALC	2-Sample tInterval Left =-1.0397555 Right=1.06775547 df =12.0171179 $\bar{x}1$ =7.514 $\bar{x}2$ =7.5 x1σn-1=0.7096	2-Sample tInterval $\bar{x}1$ =7.514 $\bar{x}2$ =7.5 x1σn-1=0.7096 x2σn-1=0.6141 n1 =7 n2 =8

[- 1.039755, 1.0677554]

Nota: observamos que nos devuelve 12.017 grados de libertad.

ACTIVIDADES DE CONSOLIDACIÓN

1. A partir de una muestra de tamaño 10 se obtuvo el intervalo (3, 4.5) con confianza 0.95 para la media μ de una población normal. ¿Cuál es la probabilidad de que el parámetro μ esté en ese intervalo?

2. Dados dos intervalos de confianza I1, I2, al nivel 0.95 para la media de una población normal con varianza conocida. ¿Pueden tener distinta amplitud en función del tamaño muestral?. Justificar la respuesta.

3. Una enfermedad hereditaria puede desarrollarse en los individuos a distintas edades. La enfermedad puede ser producida por dos genes A y B indistintamente, cuya presencia se manifiesta cuando aparece la enfermedad. Se sabe que la edad de aparición de la enfermedad es una v.a. con media μ y varianza σ^2 cuando es producida por A.

   (a) Hallar un intervalo de confianza con α = 0.95 para μ utilizando una muestra aleatoria simple de tamaño 20 a partir de X, suponiendo que la varianza es 9. ¿y para n = 36?

4. Una característica de una población sigue una distribución normal N(μ , √240). Si se extrae una muestra aleatoria simple de tamaño 100, ¿cuál es el coeficiente de confianza del intervalo (23 , 28) obtenido para la media μ si la media muestral es 25.5?

PROPUESTAS DE ACTIVIDADES INDAGATORIAS TIPO TEST

1. Cuando se ha obtenido la muestra, los límites del intervalo son:

- [ ] Puntos fijos que contienen el parámetro.
- [ ] Puntos aleatorios que contienen el parámetro.
- [ ] Puntos fijos que pueden o no contener el parámetro.
- [ ] Puntos aleatorios que pueden o no contener el parámetro.
- [ ] Ninguna es verdadera.

2. El tamaño de una muestra no está determinado por:

- [ ] La amplitud del intervalo de confianza.
- [ ] La media aritmética.
- [ ] El nivel de confianza.
- [ ] La varianza en caso de variables cuantitativas.
- [ ] La proporción en caso de variables que siguen distribución binomial.

3. Si una variable que se distribuye normalmente presenta una media de 100 y una desviación estándar de 10. ¿Cuál es la probabilidad de encontrar un sujeto en el intervalo 80–120?

- [ ] 95%
- [ ] 65%
- [ ] No se puede calcular.
- [ ] 78.6%
- [ ] 68.7%

4. La utilización de estadísticos como medida aproximada de los parámetros es un proceso de:

- [ ] Randomización.
- [ ] Estimación.
- [ ] Cuantificación.
- [ ] Agrupamiento.
- [ ] Ninguno.

5. Si aumentamos el intervalo de confianza en una estimación de una media, ¿qué ocurre con la seguridad?

- [ ] Aumenta.
- [ ] Disminuye.

- [ ] No se modifica.
- [ ] Puede aumentar o disminuir.
- [ ] Depende del tipo de variable.

6. En lo referente al tamaño de una muestra es cierto que:

- [ ] El tamaño no influye en el resultado del estudio.
- [ ] Una muestra del 19.6% de la población es suficiente en todos los casos.
- [ ] Cada experimento tiene su tamaño idóneo.
- [ ] Depende fundamentalmente de la media de la variable en estudio.
- [ ] Todas son ciertas.

7. En una muestra de 100 pacientes se ha observado un tiempo medio de supervivencia de 36 meses, con un error estándar de la media de 4 meses. ¿Cuál de las siguientes afirmaciones es correcta?

- [ ] El 95% de los sujetos sobrevivieron entre 32 y 40 meses.
- [ ] El 95% de los sujetos sobrevivieron entre 28 y 44 meses.
- [ ] Se tiene un 95% de confianza de que el verdadero tiempo medio de supervivencia se sitúe entre 32 y 40 meses.
- [ ] Se tiene un 95% de confianza de que el verdadero tiempo medio de supervivencia se sitúe entre 28 y 44 meses.
- [ ] El tamaño de la muestra es insuficiente para estimar el tiempo medio de supervivencia.

Para ampliar: www.aulamatematica.com