# INTRODUCCIÓN

Es necesario que las decisiones que debe tomar el investigador ante la situación de incertidumbre que implica inferir de casos particulares a la generalidad, deben estar respaldadas por la objetividad que garantiza la aplicación del método científico.

De este modo, los resultados obtenidos en situaciones experimentales, serán idealizados de acuerdo a un modelo probabilístico conveniente, permitiendo al investigador medir en términos de probabilidad la incertidumbre que trae aparejada la generalización de sus resultados. En otras palabras, podrá medir y comunicar el 'error' que puede cometer o la confianza que deposita en sus decisiones.

Si la distribución de frecuencias de las observaciones puede asimilarse a la distribución de probabilidad teórica en la cual está basada la aplicación de la metodología inferencial elegida, entonces el investigador podrá estar seguro de que la confianza con que realiza sus estimaciones es adecuada.

## 2. ESTIMADORES POR INTERVALOS

La estimación por intervalos es un procedimiento mediante el cual se puede afirmar, con una determinada probabilidad, que el intervalo (a, b) encierra el verdadero valor del¡ parámetro.

Para realizar una estimación por intervalos se hace la siguiente afirmación:

$$P(a \leq \theta \leq b) = 1 - \alpha$$

donde a y b son **variables aleatorias** que dependen de cierto estimador puntual. El intervalo $(a, b)$ se llama **intervalo de confianza,** ***b - a*** es una medida de la precisión de la estimación y $1 - \alpha$ es una medida de la confianza con la que contamos para efectuar la estimación.

A pesar de la diferencia en el modo en que se expresan los resultados, ambos métodos apuntan a la misma finalidad, conocer los parámetros de la población.

## 2.1 INTERVALOS DE CONFIANZA PARA MEDIAS

Hemos explicado que $\bar{x}$ es el mejor estimador puntual de $\mu$ pero nos sorprendería realmente que la media muestral fuera exactamente igual a $\mu$.

Resultaría más comprensible pensar que el valor aportado por un estimador se ubica en las cercanías del parámetro.

Esta situación sugiere que puede ser más apropiado efectuar un Intervalo alrededor de $\bar{x}$ y establecer una cierta confianza de que $\mu$ esté comprendido en dicho intervalo.

En otras palabras, dada una muestra en particular donde se ha calculado $\bar{x}$ , se puede definir un intervalo alrededor del estadístico media muestral y establecer una cierta probabilidad de que $\mu$ esté comprendido en dicho intervalo.

Trataremos de aclarar este concepto mediante la realización de una analogía con un juego.

Supongamos que en una kermés se juega a tirar un aro de plástico y, para ganar un premio, se debe ensartar, por ejemplo, una botella estratégicamente ubicada. La ficha comprada permite tirar el aro una cantidad determinada de veces.

Sabemos que al tirar el aro, algunas veces acertaremos y otras no. Un buen tirador acertará, por ejemplo, un 90% de las veces que tira el aro.

Podemos asimilar cada tirada del aro con el hecho de seleccionar una muestra aleatoria de una cierta población.

Una vez que se ha seleccionado la muestra, calculado la media en dicha muestra y construido un intervalo de confianza a partir de ese estimador muestral (tirar el aro), acertar a la botella será asimilable al hecho de que este intervalo contenga al verdadero valor del parámetro media poblacional que se está estimando.

En otras palabras, la botella representa al parámetro. El concepto importante ligado a este hecho es que, así como la botella está fija en un lugar determinado del kiosco, el parámetro también es un valor constante, desconocido pero fijo en algún lugar de la distribución de probabilidad poblacional correspondiente a la variable que se está estudiando.

En el juego lo único que se mueve es el aro. En la estimación por intervalos lo que cambia son los extremos del intervalo, pues su construcción depende de una variable aleatoria (la media muestral) que cambia de valor en función de los elementos que fueron seleccionados en la muestra.

Por todo lo analizado, cuando hablemos de probabilidad en la estimación por intervalos de un parámetro poblacional siempre hablaremos de la probabilidad de que el intervalo contenga al parámetro y no de la probabilidad de que el parámetro caiga en un intervalo determinado.

Una regla de oro en este punto es la siguiente:

El **parámetro** es una cantidad desconocida pero fija, el Intervalo es aleatorio.

Hemos aclarado el concepto de estimación por intervalos, desarrollaremos ahora en detalle su construcción para el parámetro media poblacional $\mu$.

Ya sabemos que el mejor estimador puntual de $\mu$ es $\bar{x}$ , la media muestral y, en consecuencia, lo utilizaremos para la construcción del Intervalo de confianza.

Basándonos en el teorema central del límite podemos establecer:

$$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Pero para poder utilizar la tabla de probabilidades normales debemos estandarizar esta variable aleatoria.

Ya hemos estudiado que para estandarizar una variable, se le debe restar la media y dividirla por la desviación estándar. También hemos estudiado que las variables normales estandarizadas siempre se distribuyen con media 0 y desviación estándar 1.

Entonces, podemos escribir:

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \text{ n } N\,(0,1)$$

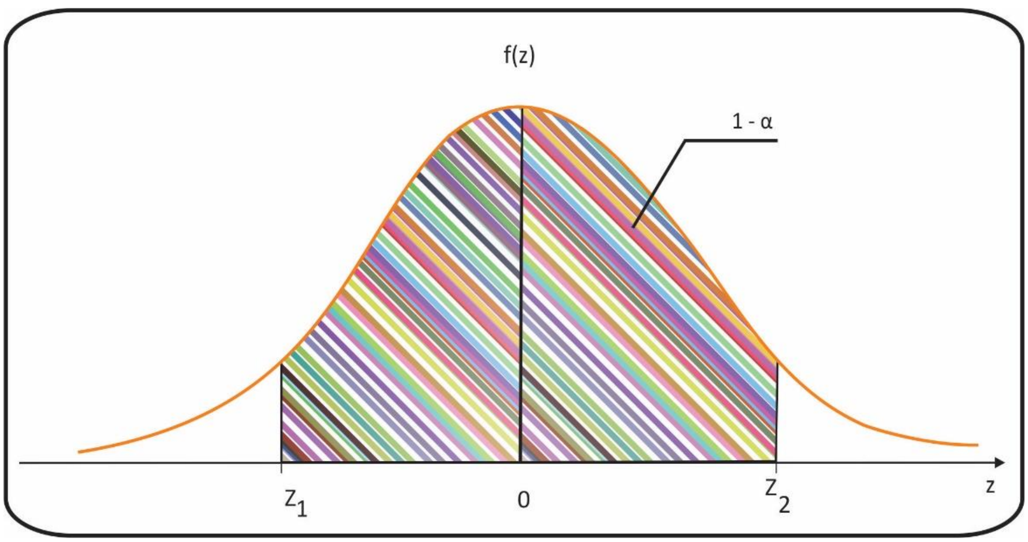
Siendo $z$ una variable normal estandarizada, se deberán buscar dos valores $z_1$ y $z_2$ tales que:

$$P(z_1 \leq z \leq z_2) = 1 - \alpha$$

o lo que es lo mismo:

$$P(z_1 \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_2) = 1 - \alpha$$

Si graficamos la expresión anterior, tenemos:




Ahora bien, ¿Qué significa el coeficiente *1 - α* que hemos establecido en el segundo miembro de la expresión anterior?

Este coeficiente se conoce con el nombre de coeficiente de confianza del que podemos dar la siguiente definición:

El **coeficiente de confianza** es la probabilidad de que un Intervalo contenga al parámetro estimado.

Volvamos al ejemplo de la tirada del aro a una botella fija, y supongamos a un jugador que acierta el 90% de las veces que tira. Podemos decir que tenemos una confianza del 90% de que este jugador acierte en cada tiro que efectúa, o lo que es lo mismo, esperamos que de cada 100 tiros, 90 emboquen la botella.

Asimilando este concepto a la estimación por intervalos, diremos que el coeficiente de confianza es un valor fijado por el investigador antes de comenzar la estimación. Así, si decide trabajar con una confianza del 95% para efectuar la estimación, el razonamiento será el siguiente:

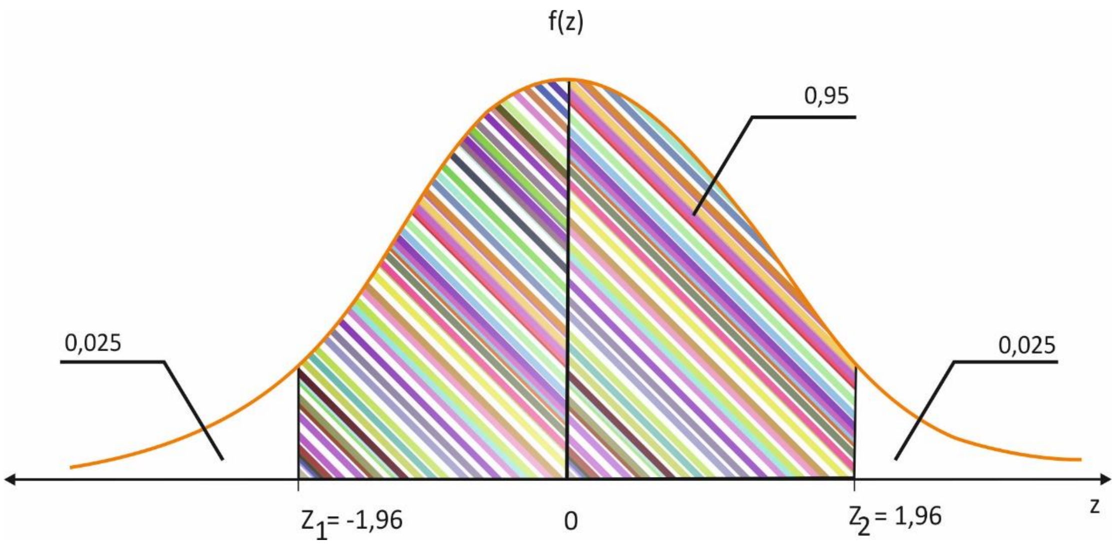
*Sobre 100 muestras aleatorias de un cierto tamaño n de una población, si en cada una se calcula la medía muestral x y, a partir de ellas, se construyen 100 intervalos de confianza para el parámetro que se desea estimar 95 contendrán al verdadero valor del parámetro poblacional, mientras que 5 no lo abarcarán.*

Volvamos ahora al problema de construir el intervalo de confianza para estimar $\mu$.

Ya tenemos un estadístico que liga al parámetro que se desea estimar  con su mejor estimador puntual expresado en la variable $z$ que, como sabemos, tiene distribución normal (0.1).

Una vez fijado el coeficiente de confianza $1 - \alpha$, por ejemplo, igual a 0.95, podemos buscar en la tabla de probabilidades correspondientes a la distribución normal los valores de $z_1$ y $z_2$ que definen un intervalo simétrico de probabilidad igual a 0.95 alrededor de $\mu = 0$.

Gráficamente, tenemos:




$z_2$ es un valor de la variable normal estandarizada que acumula una probabilidad igual a 0.975 y $z_1$ es un valor de la misma variable que acumula una probabilidad de 0.025.

Si buscamos en la tabla de probabilidades, vemos que $z_2 = 1.96$ y $z_1 = -1.96$.

Entonces, podemos escribir:

$$P(-1{,}96 \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq 1{,}96) = 0{,}95$$

Ahora bien, como estamos tratando de estimar al parámetro $\mu$, lo razonable sería despejar convenientemente de modo que quede en el centro del intervalo sólo este parámetro.
Como primera medida llevaremos hacia los extremos del intervalo la cantidad σ / n que, al estar en el denominador de la expresión, pasa al numerador de ambos miembros de la desigualdad.

El intervalo quedará planteado como sigue:

$$P(-1{,}96 \frac{\sigma}{\sqrt{n}} \leq (\bar{x} - \mu) \leq 1{,}96 \frac{\sigma}{\sqrt{n}}) = 0{,}95$$

Debemos ahora despejar $x$ que pasará con signo negativo a ambos lados de la desigualdad.

$$P(-1{,}96 \frac{\sigma}{\sqrt{n}} - \bar{x} \leq -\mu) \leq 1{,}96 \frac{\sigma}{\sqrt{n}} - \bar{x}) = 0{,}95$$

El parámetro $\mu$ ha quedado ahora solo en el centro pero con signo negativo. Para subsanar esto, multiplicamos toda la desigualdad por (- 1).

Cuando los miembros de una desigualdad son multiplicados por (- 1), cambia el sentido de la desigualdad, por lo que tendremos:

$$P\left[ -(-1{,}96 \frac{\sigma}{\sqrt{n}} + \bar{x} \geq \mu \geq -1{,}96 \frac{\sigma}{\sqrt{n}} + \bar{x} \right] = 0{,}95$$

Es decir,

$$P(1{,}96 \frac{\sigma}{\sqrt{n}} + \bar{x} \geq -\mu \geq -1{,}96 \frac{\sigma}{\sqrt{n}} + \bar{x}) = 0{,}95$$

Reacomodando convenientemente el sentido de los signos de la desigualdad ($\geq$ por $\leq$) tendremos;

$$P(\bar{x} - 1{,}96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1{,}96 \frac{\sigma}{\sqrt{n}}) = 0{,}95$$

Y este no es más que el intervalo de confianza para el parámetro $\mu$ cuando trabajamos con una confianza del 95%.

Si aparecen dificultades en la comprensión de los pasos algebraicos que hemos realizado para obtener el intervalo de confianza, no se preocupe demasiado, pues lo realmente importante es saber interpretarlo y comprender el significado de cada uno de sus componentes.

En el intervalo aparece la media de la muestra que, es muy fácil de calcular.

También aparece un valor de $z$ que acumula una cierta probabilidad determinada por el coeficiente de confianza que establece el investigador.

En este caso particular, para una confianza del 95%, este valor de $z$ es 1.96.

Otra cantidad fácilmente reemplazable es n que no es más que el tamaño de la muestra con que se está trabajando.

La medida problemática que aparece es $\sigma$, que ya sabemos identifica a la desviación estándar poblacional.

En situaciones reales de investigación, la población generalmente es grande y $\sigma$ es un parámetro desconocido. Para solucionar este problema, $\sigma$ también debe ser estimado. Su estimador lógico será s, la desviación estándar de la muestra.

Si el tamaño de la muestra es suficientemente grande (algunos autores opinan mayor a 30, otros mayor a 50), no hay problemas en seguir utilizando la distribución de probabilidad normal para medir la confianza de la estimación.

En cambio, si la muestra es chica y no se puede, por razones de costo o de tiempo u otras, aumentar su tamaño, para calcular la confianza de la estimación utilizaremos otra distribución de probabilidad: la distribución correspondiente a la variable *t* de Student, como veremos más adelante.

¿Qué podemos decir de la precisión de la estimación de un parámetro cuando efectuamos un intervalo de confianza?

Si llamamos ***a*** al extremo inferior del intervalo de confianza y ***b*** al extremo superior, la diferencia ***b - a*** nos dará una idea de **precisión de la estimación**.

Partamos del hecho de que una estimación por intervalos tiene sentido sólo si es lo suficientemente precisa.

Por ejemplo, si deseamos estimar el número de camas en hospitales públicos del país y construimos un intervalo donde $a = 5$ y $b = 100$, es evidente que obtendríamos una mayor precisión si los extremos del intervalo fuesen $a = 80$ y $b = 95$.

Lo más probable es que si presentamos un informe con la primera estimación, muchos podrán pensar que esta afirmación no tiene mayor utilidad.

¿De qué depende entonces, la precisión en una estimación por Intervalos?

Si analizamos los extremos de¡ intervalo, vemos que la precisión es directamente proporciona¡ al tamaño de la muestra n e inversamente proporcional a la desviación estándar a o s, ya que

$$b - a = (\bar{x} + z\frac{\sigma}{\sqrt{n}}) - (\bar{x} - z\frac{\sigma}{\sqrt{n}}) = 2z$$

Lógicamente, si la población objetivo de la que partimos para efectuar una investigación es muy variable, hecho que se expresará en una desviación estándar grande, la precisión de la estimación no será muy buena.

Una manera de solucionar esto sería tomar muestras de mayor tamaño, pero ello no siempre es posible.

Cuando se dan estos casos en la práctica cotidiana, es conveniente apelar a algún método de selección de muestras que reduzca la variabilidad y, en consecuencia, aumente la precisión de la estimación. Este tema será discutido oportunamente.

Supongamos, por ejemplo, que se quiere conocer el aumento promedio de peso de niños pertenecientes a un estrato social muy bajo cuando se los alimenta con una dieta fortificante.

La población objetivo fue especificada como los niños de dicho estrato social en la provincia de Córdoba.

Sería casi imposible alimentar con la dieta a todos los niños de la provincia. La única solución a este problema será recurrir a una muestra de, por ejemplo, $n = 40$ niños seleccionados aleatoriamente de la población en cuestión y alimentarlos con la dieta en estudio.

Al cabo de un cierto tiempo se registrarán los aumentos de peso de cada niño y se calculará el promedio en la muestra.

Supongamos que resultó $\bar{x} = 4.300$ Kg.

Utilizaremos esta estimación puntual para efectuar una estimación por intervalos del parámetro $\mu =$ aumento de peso promedio de niños de cierto estrato social en la población.

Supongamos, además, que por estudios previos se conoce que el valor que toma el parámetro varianza poblacional es $\sigma^2 = 4$ Kg$^2$. Podemos utilizar la distribución de probabilidad normal para efectuar la estimación dado que $n = 40$. Con una confianza de $1 - \alpha = 0.95$, el intervalo resultante será:

$$P(4{,}3 - 1{,}96 \frac{2}{\sqrt{40}} \leq \mu \leq 4{,}3 + 1{,}96 \frac{2}{\sqrt{40}}) = 0{,}95$$

$$P(3{,}68 \leq \mu \leq 4{,}92) = 0{,}95$$

Amplitud del intervalo: b - a $= 4.92 - 3.68 = 1.24$

Ahora bien, si en lugar de tomar una muestra de 40 niños, hubiésemos tomado una de $n = 100$ niños, manteniendo todos los demás valores constantes, el intervalo será:

$$P(4{,}3 - 1{,}96 \frac{2}{\sqrt{100}} \leq \mu \leq 4{,}3 + 1{,}96 \frac{2}{\sqrt{100}}) = 0{,}95$$

$$P(3{,}91 \leq \mu \leq 4{,}69) = 0{,}95$$

Amplitud del intervalo: b - a = 4.69 − 3.91 = 0.78

Vemos que, al aumentar el tamaño de la muestra, disminuyó la amplitud del intervalo de confianza, logrando así una mayor precisión en la estimación.

Pero, ¿Qué pasaría ahora si con los mismos datos la varianza poblacional $\sigma^2$ fuese 9 en lugar de 4?

En este caso estamos pensando que la variable aumento de peso en los niños tiene una variabilidad mayor de niño a niño.

Si construimos un nuevo intervalo de confianza, cambiando sólo el valor de la varianza, tendremos:

$$P(4{,}3 - 1{,}96\frac{3}{\sqrt{40}} \leq \mu \leq 4{,}3 + 1{,}96\frac{3}{\sqrt{40}}) = 0{,}95$$

$$P(3{,}37 \leq \mu \leq 5{,}23) = 0{,}95$$

Amplitud del intervalo: b - a = 5.23 − 3.37 = 1.86.

Si comparamos este intervalo con el primero que hemos construido, vemos que ahora se han ampliado los límites provocando una merma en la precisión de la estimación.

Y ¿Qué podemos decir de la confianza de la estimación?

Cuando se efectúa una estimación por intervalos, a medida que se incrementa el nivel de confianza se incrementa también la amplitud del intervalo y, en consecuencia, disminuye la precisión del mismo.

Así, por ejemplo, si decidimos aumentar la confianza trabajando con 1 - $\alpha$ = 0.99, el intervalo sería:

$$P(4{,}3 - 2{,}576\frac{2}{\sqrt{40}} \leq \mu \leq 4{,}3 + 2{,}576\frac{2}{\sqrt{40}}) = 0{,}95$$

$$P(3{,}49 \leq \mu \leq 5{,}11) = 0{,}95$$

Amplitud del intervalo: b - a = 5.11 − 3.49 = 1.62

Mientras que con 1 - $\alpha$ = 0,95 la amplitud del mismo intervalo era de 1.24.

Luego, construyendo el intervalo con un 99% de confianza hemos aumentado nuestra seguridad en la estimación pero a costa de una precisión menor.

*Como síntesis, podemos decir que para reducir la amplitud de un intervalo de confianza y en consecuencia aumentar su precisión, debemos reducir el error estándar de la media muestral x que es* $\sigma$ / *n. Esto puede lograrse solamente disminuyendo la variabilidad de los datos ya sea homogeneizando el material experimental o, sí esto no puede llevarse a cabo, aumentando el tamaño de la muestra.*

Es costumbre utilizar coeficientes de confianza del 90%, 95% y 99%.

Por este motivo les damos la siguiente tabla que resume los valores de probabilidad de la distribución normal estandarizada para estos niveles de confianza.

| Coeficiente de confianza | Z |
|---|---|
| 0,90 | 1,645 |
| 0,95 | 1,960 |
| 0,99 | 2,576 |

**¿Por qué "intervalos de confianza" y no "de probabilidad"?**

Si observamos las expresiones de los intervalos obtenidos para la media poblacional, de un 95% de confianza o de un 99% de confianza (ver páginas 28 y 29), se puede apreciar que en ellos no está implicada ninguna variable aleatoria, ya que en el centro del intervalo está un parámetro $\mu$ y en los extremos se tienen números obtenidos sumando y restando Z$\sigma$ / n a la media obtenida en la muestra.

Al no existir variable aleatoria en la expresión, no sería correcto hablar de probabilidad.

No obstante, observando el Gráfico de la página 31, se comprende que, cumpliéndose los supuestos (distribución normal de la población, o muestras suficientemente grandes), construyendo los intervalos con ésta metodología (sumar y restar $Z\sigma$ / $n$ a la media muestral), el 95% o 99% (según el valor de $Z$) de los posibles intervalos contendrán al verdadero valor de $\mu$. De allí la expresión "existe entre un 95 (o 99) por ciento de confianza de que el intervalo contenga al parámetro".

En la bibliografía, es común encontrar al intervalo de confianza expresado de la siguiente manera:

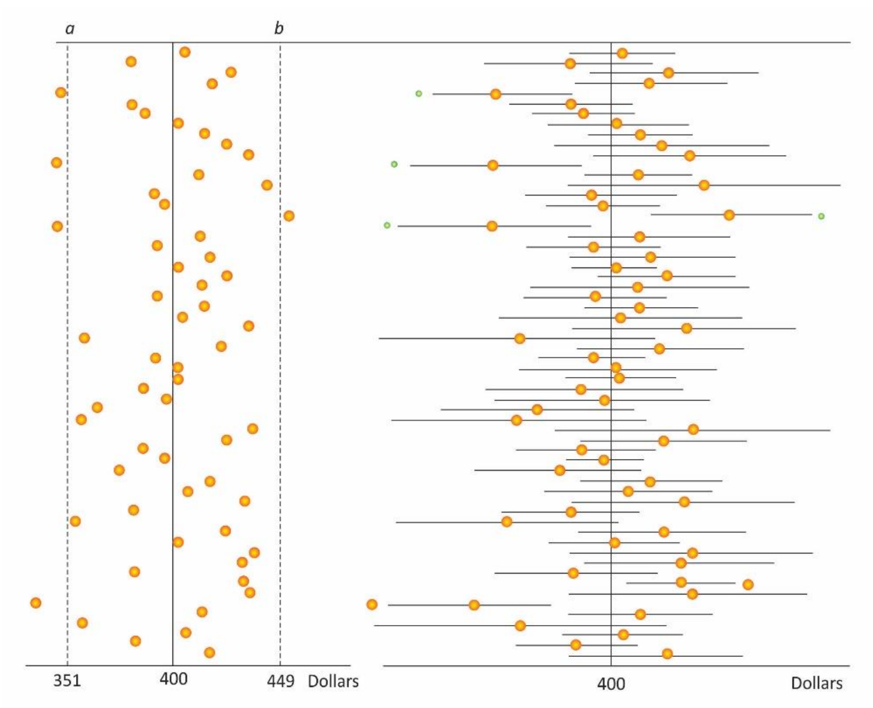
$$a \leq \mu \leq b \text{ con un 95\% de confianza en lugar de } P(a \leq \mu \leq b) = 0{,}95$$

En este Módulo se utiliza esta última forma de expresión, pero debe leerse no como una probabilidad, sino como una medida del nivel de "confianza" de que el parámetro esté incluido en el intervalo. Esto es: si se repitiera el mismo experimento muchas veces, determinando un intervalo cada vez, en un 95% (o 99%) de los casos este intervalo contendría al parámetro verdadero. Por eso, si el experimento se realiza una sola vez, hay un 95% (o 99%) de confianza en que el único intervalo obtenido contenga al parámetro.

**Una Última Reflexión Acerca de la Interpretación de los Intervalos de Confianza**

Para ejemplificar la interpretación del procedimiento de estimación por intervalos supongamos que se tiene una variable $\boldsymbol{X}$ distribuida con media $\mu = 400$. De esta población se extraen 60 muestras de tamaño $n = 16$; con cada muestra se construye un intervalo de confianza del 95%.

Los resultados se pueden representar gráficamente de la siguiente manera:




Cada punto representa la media muestral y cada línea la amplitud del intervalo correspondiente.

En este caso vemos que 6 de tales intervalos (10%) no contienen al verdadero valor del parámetro $\mu$ y 54 (90%) sí lo contienen.

¿Qué reflexión podemos efectuar ante esta situación?

Partimos de una confianza en la estimación de cada intervalo del 95%.

Si esta aseveración fuera correcta, nos indicaría que 57 intervalos deberían contener al parámetro $\mu$.

¿Qué sucedió entonces?

Una explicación podría ser que no se han sacado todas las muestras posibles ($N^n$). Es decir que habría que completar el muestreo y luego determinar si esta situación se mantiene.

Otra explicación podría estar en el hecho de tener una variable $\boldsymbol{X}$, cuyo promedio en la población es de $\mu = 400$, pero de la cual no conocemos su distribución poblacional.

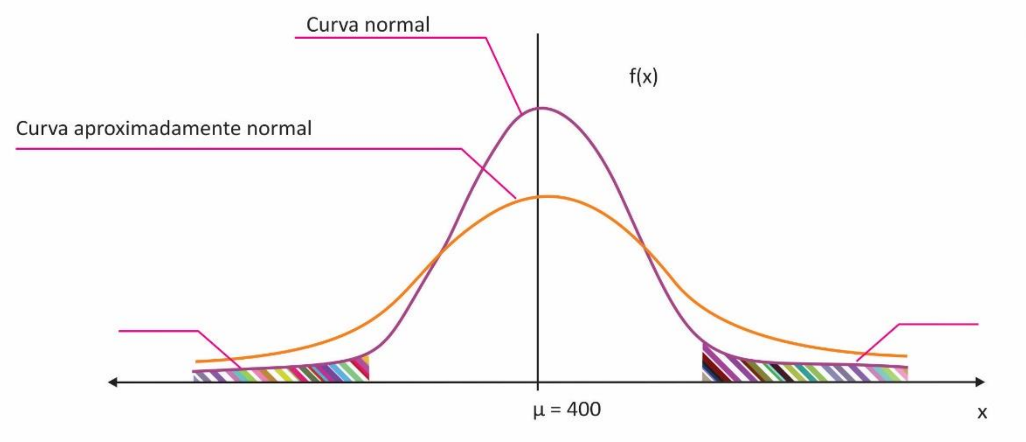
No obstante, en base al teorema central del límite sabemos que la variable aleatoria media muestral se distribuye normalmente a medida que aumenta el tamaño de la muestra.

En este caso, es evidente que la distribución de la variable media muestral con muestras de tamaño $n = 16$ no ha logrado una perfecta simetría.

Ante esta situación, construimos intervalos de confianza basados en áreas de probabilidad de la distribución normal, que no coinciden exactamente con las áreas de probabilidad de la distribución de la media muestral de nuestro ejemplo.

Entonces, creemos trabajar con una confianza en nuestra estimación del 95% pero en realidad, al no cumplirse estrictamente el supuesto de distribución normal de la variable en estudio ($\boldsymbol{X}$), esta confianza no es real.

Gráficamente la situación podría ser la siguiente:




En la práctica de investigación es muy común encontrar situaciones como la descripta, en las que se violan supuestos esenciales a la aplicación de metodologías estadísticas. Por eso creemos necesario efectuar una última reflexión:

"Siempre que debamos realizar inferencias tengamos cuidado con la violación de los supuestos que subyacen a las diferentes metodologías".

"Los conocimientos del investigador suelen ser la mejor herramienta en la prevención de conclusiones erróneas".

Trate de resolver ahora las siguientes situaciones prácticas de investigación aplicando los conceptos de estimación por intervalos del parámetro poblacional $\mu$ aprendidos.

## ACTIVIDAD 1

**Optativa, sin entrega**

*Algunas publicaciones médicas aseguran que la vitamina C puede resultar de utilidad en la reducción del colesterol de las paredes internas de las arterias, disminuyendo así la posibilidad de ataques cardíacos. Para corroborar esta afirmación, un investigador médico observó el nivel de colesterol en 50 pacientes (todos ellos registraban niveles de colesterol mayores a los normales y, en consecuencia, conformaban un grupo de riesgo al infarto).*

*Durante un mes sometió a estos pacientes a una dosis diaria de 500 mg. de vitamina C. Al finalizar el tratamiento, volvió a medir los niveles de colesterol de cada paciente registrando la disminución observada en cada uno de ellos. De estos registros obtuvo un promedio de $x = 64{,}3$ mg. por 100 ml. con una desviación estándar $s = 18{,}9$ mg. por 100 ml.*

***Con la información proporcionada por la muestra de pacientes seleccionados, estime por intervalos el verdadero valor de la disminución promedio de colesterol en la población objetivo considerada.***

Veremos ahora un caso muy común en la práctica, que se presenta cuando el investigador está obligado a trabajar con muestras chicas y pretende estimar la media poblacional con $\sigma^2$ desconocida.

**Estimación por intervalos**

En estos casos, dado que no se conoce la varianza poblacional, para estimar por intervalos al parámetro media poblacional, debemos recurrir a un nuevo estadístico distribuido como una variable "*t*" de Student.

El estadístico $t$ de Student surge de estandarizar la variable media muestral, pero tomando a $S^2$ como estimador puntual de la varianza poblacional.

Luego:

$$\frac{(\bar{x} - \mu)}{s/\sqrt{n}} = t$$

Esta variable ya no se distribuye normalmente cuando el tamaño de la muestra es chico, en la práctica $\leq 30$.

Entonces:

$$t = \frac{(\bar{x} - \mu)}{s/\sqrt{n}} \sim t_{n-1}$$

donde n - 1, es el denominador con el que se calculó la varianza muestral $S^2$. Este valor corresponde a los grados de libertad de la variable "*t*" de Student.

Ahora estamos en condiciones de establecer el correspondiente Intervalo de confianza para el parámetro media poblacional cuando trabajamos con muestras chicas.

Podemos escribir, luego:

Despejando convenientemente, (y recordando que $t_{\alpha/2}$ es un número negativo), obtendremos:

$$P(x - t_{n-1,1-\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq x - t_{n-1-\alpha/2} \frac{S}{\sqrt{n}} = 1 - \alpha$$

Por ejemplo, supongamos que un investigador desea estimar el rendimiento promedio de grasa que contiene la leche producida por vacas de cierta raza en un período de tiempo determinado.

Para ello extrae una muestra de 10 vacas lecheras, obteniendo:

$$X = 36.4 \qquad S^2/n = 264.04/10 = 26.404 \quad S/\sqrt{n} = 5.138 \text{ kg}$$

Supongamos que no sabemos nada acerca de los parámetros poblacionales y los únicos datos que posee el investigador para llevar a cabo su investigación son la media y la varianza muestral. Luego deberá estimar a $\sigma^2$ utilizando la varianza muestral $S^2$.

Como la muestra es chica y no conocemos la varianza poblacional $\sigma^2$, el estadístico para confeccionar el correspondiente intervalo de confianza de $\mu$, tendrá distribución t de Student con *n - 1* grados de libertad.

Para un nivel de confianza del 95% y buscando convenientemente en la tabla de probabilidades de la distribución *t* de Student, obtenemos:

$$t_{9;\,0{,}025} = -\ 2.262 \text{ y } t_{9;\,0{,}975} = 2.262$$

El intervalo de confianza será:

$$P(36.4 - 2.262 \text{ x } 5.138 \leq \mu \leq 36.4 + 2.262 \text{ x } 5.138) = 0.95$$
$$P(36.4 - 11{,}62 \leq \mu \leq 36.4 + 11{,}62) = 0.95$$
$$P(24.78 \leq \mu \leq 48.02) = 0.95$$

Podemos concluir diciendo que existe un 95% de confianza de que este intervalo contenga al verdadero valor del parámetro poblacional $\mu$ (promedio de grasa en la población de vacas lecheras consideradas).