# 1. INTERVALOS DE CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL

Supongamos que X1,...,Xn es una muestra aleatoria de una población normal con media $\mu$ y varianza $\sigma^2$. Sabemos que la media muestral, $\overline{X} = \sum X_i / n$, es un *estimador insesgado* y *consistente* de $\mu$. Sin embargo, no esperamos que la media muestral coincida con $\mu$, en cambio esperamos que esté cerca de $\mu$. Muchas veces, más que un estimador puntual, es más útil especificar un intervalo sobre el que tengamos el grado de confianza de que $\mu$ se encuentre dentro de él. Para obtener dicho intervalo nos basamos en la distribución del estimador puntual.

## 1.1 INTERVALOS DEL 95% DE CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL VARIANZA CONOCIDA

Cuando X1,...,Xn es una muestra aleatoria de una población normal con media $\mu$ y varianza $\sigma^2$ el estimador puntual de $\mu$, $\overline{X}$, tiene distribución normal con media $\mu$ y varianza $\sigma^2$/n y resulta que

$$\sqrt{n}\,\frac{(\overline{X}-\mu)}{\sigma}$$

tiene distribución N (0, 1) (Normal Estándar). Por lo tanto

$$P\left\{-1.96 < \sqrt{n}\,\frac{(\overline{X}-\mu)}{\sigma} < 1.96\right\} = 0.95$$

ó equivalentemente

$$P\left\{\overline{X} - 1.96\,\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

Esto es que el 95% de las veces $\mu$ se encontrará dentro de $1.96\,\frac{\sigma}{\sqrt{n}}$ unidades de la media muestral. Si ahora observamos que $\overline{X}$ vale $\overline{x}$, confiamos con una confianza del 95%, que $\mu$ se encuentre en el intervalo

$$\left(\overline{x} - 1.96\,\frac{\sigma}{\sqrt{n}},\ \overline{x} + 1.96\,\frac{\sigma}{\sqrt{n}}\right) \tag{1}$$

El intervalo anterior es un *intervalo de 95% de confianza para* $\mu$.

**Ejemplo**. Supongamos que, cuando se determina el contenido ($\mu$) de una sustancia en un compuesto, el valor observado ( X ) está normalmente distribuido con media $\mu$ y varianza 4. Esto significa que cuando el contenido verdadero es $\mu$ el valor observado X es $\mu + \varepsilon$ ( $X = \mu + \varepsilon$, modelo de Gauss sin sesgo). La variable aleatoria $\varepsilon$ representa el error de medición y tiene distribución Normal con media 0 y varianza 4. Supongamos que para reducir la varianza de la estimación, se determina el mismo contenido 9 veces y se calcula el promedio. Si los sucesivos valores observados (en las unidades adecuadas) son 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, construyamos un intervalo de confianza para $\mu$.

Si las repeticiones se han realizado en condiciones independientes e idénticas, como $\bar{x} = \frac{81}{9} = 9$ resulta que un intervalo con un nivel de confianza del 95% está dado por

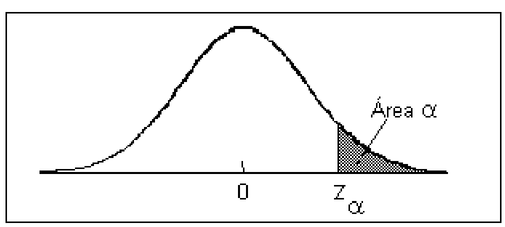
$$\left(9 - 1.96\frac{\sigma}{3}, 9 + 1.96\frac{\sigma}{3}\right) = (7.69, 10.31)$$

Luego, tenemos una confianza del 95%, que el contenido verdadero se encuentre entre 7.69 y 10.31.

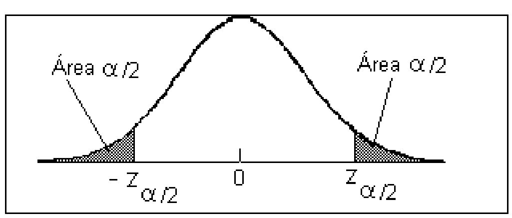
**1.2 INTERVALOS CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL CON VARIANZA CONOCIDA PARA CUALQUIER NIVEL DE CONFIANZA ESPECIFICADO.**

El valor crítico $z_\alpha$ de una distribución normal estándar es el valor que deja a su derecha, bajo la curva de densidad normal estándar, un área $\alpha$:

$$P\{Z > z_\alpha\} = \alpha$$




donde $Z$ es una v.a. con distribución N(0,1).



Nuevamente consideremos una muestra aleatoria de una población normal con media $\mu$ y varianza $\sigma^2$ ( X1,...,Xn ) luego la variable aleatoria $\sqrt{n}\,\frac{(\overline{X}-\mu)}{\sigma}$ tiene distribución N (0, 1) Por lo tanto

$$P\left\{-z_{\alpha/2} < \sqrt{n}\,\frac{(\overline{X}-\mu)}{\sigma} < z_{\alpha/2}\right\} = 1-\alpha$$

ó equivalentemente

$$P\left\{\overline{X} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

Luego, un intervalo para $\mu$ con el 100*(1-$\alpha$)% de confianza está dado por

$$\left(\bar{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right) \qquad (2)$$

siendo $\bar{x}$ el valor observado de la media muestral.

**Observación**: El intervalo dado en (2) no es aleatorio. Esto significa que $\mu$ pertenece ó no pertenece al intervalo construido y no lo sabemos. Confiamos que sí pertenezca con un nivel de confianza del 95% cuando $\alpha$ = 0.05.

## 1.3 TAMAÑO DE MUESTRA NECESARIO PARA LA OBTENCIÓN DE UN INTERVALO DE CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL CON LONGITUD PREFIJADA

La longitud del intervalo de confianza dado por la ecuación (2) es

$$2\, z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}$$

Supongamos que nos interesa afirmar con un nivel de confianza del 99% que $\mu$ se encuentra dentro de un intervalo de longitud $L$ , ¿cuán grande tiene que ser n?.

Como el nivel de confianza es 99%, $\alpha$=0.01. Por lo tanto $\alpha$/2 = 0.005, $Z_{0.005}$ = 2.58 y la longitud del intervalo de confianza es

$$5.16\,\frac{\sigma}{\sqrt{n}}$$

Para que la longitud sea $L$ debemos elegir n de manera que

$$5.16 \frac{\sigma}{\sqrt{n}} = L$$

o sea

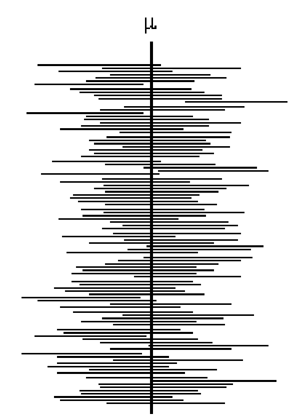
$$n = (5.16\sigma / L)^2$$

En general $n$ no será entero y se elige el menor que cumple $n \geq (5.16\sigma / L)^2$

**Ejemplo** (continuación). Si interesa obtener un intervalo con un nivel de confianza del 99% de longitud $L$ =1, como $\sigma^2$ = 4 resulta $n$ ≥ ( 5.16 * 2 )$^2$ = 106,5. Luego deberá elegirse $n$ = 107.

**OBSERVACIÓN**

Un intervalo de confianza para $\mu$, es un rango de valores entre los cuales confiamos se encuentre $\mu$. ¿Por qué confiamos? Si construimos un intervalo del 95% de confianza significa que 95 de cada 100 veces que utilicemos la ecuación (1) para calcularlo, el intervalo obtenido contendrá al verdadero valor $\mu$. Confiamos que nuestro intervalo sea uno de esos 95 intervalos ¨buenos¨.



Intervalos de Confianza repetidos para μ

Una vez que hemos construido el intervalo, la probabilidad de que $\mu$ pertenezca a dicho intervalo es 0 (si $\mu$ no pertenece) ó 1 (si $\mu$ pertenece) pero no lo sabemos. Es decir, deja de tener sentido plantear una probabilidad cuando hemos hallado el intervalo resultante.

## 1.4 INTERVALOS CONFIANZA PARA LA MEDIA DE UNA POBLACIÓN NORMAL CON VARIANZA DESCONOCIDA

Hemos desarrollado los intervalos de confianza anteriores basándonos en que $\sqrt{n}\,\frac{(\overline{X}-\mu)}{\sigma}$ tiene distribución N(0,1). Como no conocemos la varianza la estimamos por

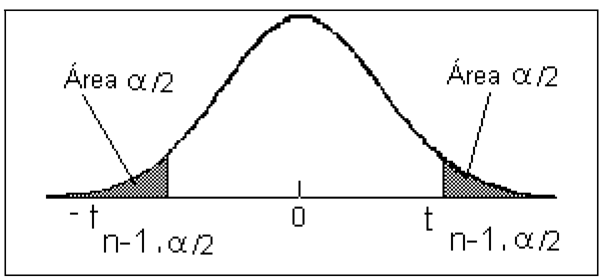
$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1},$$

sabemos que la varianza muestral $S^2$ es un estimador insesgado y consistente de $\sigma^2$.

Para hallar los intervalos de confianza para la media de una población normal utilizaremos el siguiente resultado:

$$T_{n-1} = \sqrt{n}\,\frac{(\overline{X}-\mu)}{S}$$

tiene distribución t con n-1 grados de libertad.




Por lo tanto

$$P\left\{-t_{n-1,\alpha/2} < \sqrt{n}\,\frac{(\overline{X}-\mu)}{S} < t_{n-1,\alpha/2}\right\} = 1-\alpha$$

ó equivalentemente

$$P\left\{\overline{X} - t_{n-1,\alpha/2}\,\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{n-1,\alpha/2}\,\frac{S}{\sqrt{n}}\right\} = 1-\alpha$$

Por lo tanto si observamos que $\overline{X} = \overline{x}$ y $S = s$ diremos que

$$\mu \in \left( \overline{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} ,\ \overline{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \right) \tag{3}$$

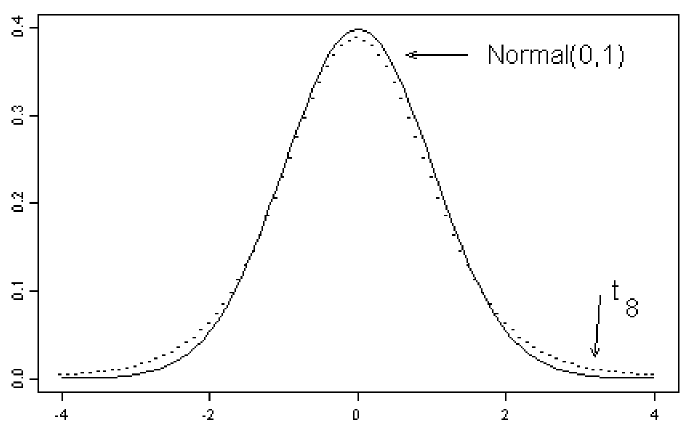
con un 100*(1-$\alpha$)% de confianza

**Ejemplo**. Continuación.
Supongamos ahora que cuando se determina el contenido ($\mu$) de una sustancia en un compuesto, el valor observado está normalmente distribuido con media $\mu$ y *varianza desconocida*. Si consideramos los valores obtenidos en las mismas 9 determinaciones y *estimamos la varianza* resultan: $\overline{x} = 9$ y $s^2$ = 9.5 ó $s$ = 3.082

Como $t_{8,0.025}$ =2.306, de (3), un intervalo del 95% confianza para $\mu$ es

$$\left( 9 - 2.306 \frac{3.082}{3}, 9 + 2.306 \frac{3.082}{3} \right) = (6.63, 11.37)$$

En el ejemplo, suponemos que los datos provienen de una distribución Normal, como el desvío es desconocido (lo estimamos por s), el estadístico en el que basamos el intervalo tiene **distribución t con n-1 (8) grados de libertad**.




La distribución t es una distribución de probabilidad simétrica alrededor del cero y con forma de campana, similar a la curva Normal. A diferencia de la Normal, su dispersión depende de los **grados de libertad**. A medida que aumentan los grados de libertad las curvas t tienden a ser indistinguibles de la curva Normal estándar.