INTERVALO DE CONFIANZA

I- Concepto de Intervalo de Confianza.

En el contexto de estimar un parámetro poblacional, un intervalo de confianza es un rango de valores (calculado en una muestra) en el cual se encuentra el verdadero valor del parámetro, con una probabilidad determinada.

La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina **nivel de confianza**, y se denota 1-α. La probabilidad de equivocarnos se llama **nivel de significancia** y se simboliza α. Generalmente se construyen intervalos con confianza 1-α=95% (o significancia α=5%). Menos frecuentes son los intervalos con α=10% o α=1%.

Para construir un intervalo de confianza, se puede comprobar que la distribución Normal Estándar cumple:

$$P(-1.96 < z < 1.96) = 0.95$$

(lo anterior se puede comprobar con una tabla de probabilidades o un programa computacional que calcule probabilidades normales).

Luego, si una variable X tiene distribución N(μ,σ), entonces el 95% de las veces se cumple:

$$-1.96 \leq \frac{\bar{X}-\mu}{\sigma}\sqrt{n} \leq 1.96$$

Despejando μ en la ecuación se tiene:

$$\bar{X} - 1.96 \times \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}$$

El resultado es un intervalo que incluye al μ el 95% de las veces. Es decir, es un **intervalo de confianza al 95% para la media** μ cuando la variable X es normal y σ^2 es conocido.

En forma general, Si el nivel de confianza es 1-α, y se conoce σ, el intervalo de confianza de μ es: $\overline{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

II- Intervalo de confianza para un promedio:

Generalmente, cuando se quiere construir un intervalo de confianza para la media poblacional μ, la varianza poblacional σ^2es desconocida, por lo que el intervalo para μ construido al final de II es muy poco práctico.

Si en el intervalo se reemplaza la desviación estándar poblacional σ por la desviación estándar muestral **s**, el intervalo de confianza toma la forma:

$$\bar{X} - 1.96 \times \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \times \frac{s}{\sqrt{n}}$$

La cual es una buena aproximación para el intervalo de confianza de 95% para μ con σ^2 desconocido. Esta aproximación es mejor en la medida que el tamaño muestral sea grande, es decir, *n>30*

Cuando el tamaño muestral es pequeño, (*n≤30*) el intervalo de confianza requiere utilizar la distribución t de Student (con n-1 grados de libertad, siendo n el tamaño de la muestra), en vez de la distribución normal (por ejemplo, para un intervalo de 95% de confianza, los límites del intervalo ya no serán construidos usando el valor 1,96).

En forma general si no se conoce σ se sustituye por la desviación estándar de la muestra, es decir, por s, si además el tamaño de la muestra es *n>30*. El intervalo de confianza para un nivel de confianza de 1-α, de la media poblacional es:

$$\bar{X} - Z_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{s}{\sqrt{n}}$$

En el caso de que *n≤30*, el intervalo será:

$$\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}$$ con *n-1* grados de libertad para la distribución t

Ejemplo:
Los siguientes datos son los puntajes obtenidos para 45 personas de una escala de depresión (mayor puntaje significa mayor depresión).

2	5	6	8	8	9	9	10	11
11	11	13	13	14	14	14	14	14
14	15	15	16	16	16	16	16	16
16	16	17	17	17	18	18	18	19
19	19	19	19	19	19	19	20	20

Para construir un intervalo de confianza para el puntaje promedio poblacional, asumamos que los datos tienen distribución normal, con varianza poblacional σ^2 desconocida. Como σ^2 es desconocido, lo estimamos por s^2=18,7. Luego, un intervalo de confianza aproximado es:

$$\bar{X} - 1.96 \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{S}{\sqrt{n}} \Rightarrow 14.5 - 1.96 \frac{4.3}{\sqrt{45}} \leq \mu \leq 14.5 + 1.96 \frac{4.3}{\sqrt{45}}$$

Luego, el intervalo de confianza para μ es (13,2 , 15,8). Es decir, el puntaje promedio poblacional se encuentra entre 13,2 y 15,8 con una confianza 95%.

III. Intervalo de Confianza para una Proporción.

En este caso, interesa construir un intervalo de confianza para una proporción o un porcentaje poblacional (por ejemplo, el porcentaje de personas con hipertensión, fumadoras, etc.)

Si el tamaño muestral n es grande, el Teorema Central del Límite nos asegura que:

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

O bien:

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)}}\sqrt{n} \sim N(0,1)$$

Donde p es el porcentaje de personas con la característica de interés en la población (o sea, es el parámetro de interés) y $\hat{p}$ es su estimador muestral.

Luego, procediendo en forma análoga al caso de la media, podemos construir un intervalo de 95% de confianza para la proporción poblacional p.

$$\hat{p} - 1.96 \times \sqrt{\hat{p} \times (1-\hat{p})/n} \leq p \leq \hat{p} - 1.96 \times \sqrt{\hat{p} \times (1-\hat{p})/n}$$

En forma general:

$\hat{p} - Z_{\alpha/2}(\sqrt{\hat{p}(1-\hat{p})/n}) \leq p \leq \hat{p} + Z_{\alpha/2}(\sqrt{\hat{p}(1-\hat{p})/n})$ con un nivel de confianza de 1-α

Ejemplo:
En un estudio de prevalencia de factores de riesgo en una cohorte de 412 mujeres mayores de 15 años en la Región Metropolitana, se encontró que el 17.6% eran hipertensas. Un intervalo de 95% de confianza para la proporción de mujeres hipertensas en la Región Metropolitana está dado por:

$$0.176 - 1.96\sqrt{0.176(1-0.176)/412} \leq p \leq 0.176 + 1.96\sqrt{0.176(1-0.176)/412}$$

Luego, la proporción de hipertensas varía entre (0,139 , 0,212) con una confianza de 95%.

IV. Uso de Intervalos de Confianza para verificar Hipótesis.

Los intervalos de confianza permiten verificar hipótesis planteadas respecto a parámetros poblacionales.

Por ejemplo, supongamos que se plantea la hipótesis de que el promedio de peso de nacimiento de cierta población es igual a la media nacional de 3250 gramos.

Al tomar una muestra de 30 recién nacidos de la población en estudio, se obtuvo:

$\bar{x}$= 2930
s= 450
n= 30

Al construir un intervalo de 95% de confianza para la media poblacional, se obtiene:

$$2930 - 1.96 \times \frac{450}{\sqrt{30}} \leq \mu \leq 2930 + 1.96 \times \frac{450}{\sqrt{30}}$$

Luego, el peso de nacimiento varía entre 2769 y 3091 gramos, con una confianza de 95%.
Como el intervalo no incluye el valor μ=3250 gramos planteado en la hipótesis, entonces esta es rechazada con confianza 95% (o un valor p menor a 0,5).