Clase 14: Intervalos de confianza. Definiciones básicas, propiedades, interpretación. Nivel de confianza, precisión y tamaño muestral

Universidad Nacional de Colombia - Sede Medellín

Intervalos de confianza (IC)

Hasta ahora los estimadores estudiados son puntuales, es decir, exhiben un solo valor como estimación del parámetro de interés. Pero en muchos casos esto no es suficiente. A veces se requiere de un rango de posibles valores para el parámetro de interés, es decir, un intervalo real donde se cree estará el valor del parámetro con una alta confianza.

Sea θ un parámetro de interés y $\hat{\theta}$ un estimador puntual de θ, un estimador de θ por intervalo, es un intervalo real de la forma $(l < \theta < u)$, donde l y u dependen de $\hat{\theta}$ y de la distribución de $\hat{\theta}$.

Cada muestra aleatoria proporcionará un valor diferente para $\hat{\theta}$ y por lo tanto valores diferentes para l y u. Así, los extremos del intervalo en cuestión se convierten en v.as, las cuales denotaremos L y U. El intervalo $(L\,,\ U)$ es llamado Intervalo aleatorio. Usando $\hat{\theta}$ y su distribución se pueden determinar L y U tales que $P(L < \theta < U) = 1 - \alpha$ estando $\alpha \in (0, 1)$ y siendo un valor conocido.

Intervalos de confianza (IC)

El intervalo (l, u) será llamado un intervalo de confianza al $100(1-\alpha)\%$ para θ. Los valores l y u son llamados límites del intervalo de confianza.

Interpretación:
El I. C. al $100\,(1\,-\,\alpha)\,\%$ para θ, se interpreta como: el $100\,(1\,-\,\alpha)\,\%$ de los intervalos contendrá el verdadero valor de θ.

El intervalo $(l\,,\,u)$ se conoce como intervalo Bilateral. Los intervalos $(l\,,\,+\infty)$ ó $(-\infty\,,\,u)$ son llamados Intervalos Unilaterales.

En un IC bilateral la longitud $u\,-\,l$ es una medida de la calidad de la información obtenida. El valor $\theta\,-\,l$ ó $u\,-\,\theta$ se conoce como Precisión del Estimador. Lo ideal es tener IC angostos con una alta confianza.

Intervalos de confianza (IC)

Dada una m.a $X_1, \ldots, X_n$, de una población cuya distribución depende de un parámetro desconocido θ, se define un IC de nivel de confianza $(1-\alpha)$, el intervalo:

$$P\left[L\left(\hat{\theta}\right) < \theta < U\left(\hat{\theta}\right)\right] = 1 - \alpha$$

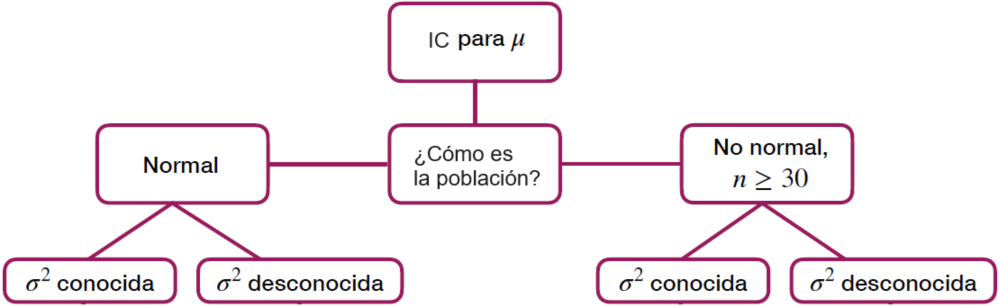
Nivel de Confianza $(1-\alpha)$:

- Un 95% de confianza implica que el 95% de todas las muestras darán un intervalo que incluye el parámetro que se está estimando y sólo el 5% de las muestras darían intervalos de confianza erróneos.
- Los níveles de confianza más utilizados son: 99%, 95% y 90%
- Mientras más alto es el nivel de confianza más fuerte es la creencia de que el valor real del parámetro está dentro del intervalo.

Precisión:

- El ancho del intervalo da información sobre la precisión del IC, es decir, entre más ancho sea el intervalo menos precisión tendrá, lo que indica que hay una gran cantidad de incertidumbre sobre el valor que se está estimando. Entre menor sea la longitud del IC mayor será su precisión.
- La precisión está dada por la longitud del intervalo dividido entre dos.
- Si el tamaño de la muestra n es grande el IC es más preciso. Si $(1-\alpha)$ es grande el IC es menos preciso. Una estrategia es fijar $(1-\alpha)$ y la precisión y encontrar n que permita construir un IC con esas características.

Intervalos de confianza para la media



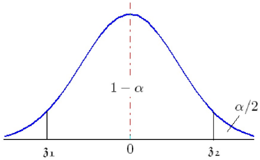
IC para μ con σ^2 conocida y población normal

Caso 1: Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población **normal** $N(\mu, \sigma^2)$ con media μ desconocida y **varianza** σ^2 **conocida**.
Vamos a obtener L y U tales que $P(L < \mu < U) = 1 - \alpha$ dado un valor de α.

$$
\begin{aligned}
P(L < \mu < U) &= P(-U < -\mu < -L) \\
&= P\left(\bar{X} - U < \bar{X} - \mu < \bar{X} - L\right) \\
&= P\left(\frac{\bar{X} - U}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{\bar{X} - L}{\sigma/\sqrt{n}}\right) \\
&= P(z_1 < Z < z_2) = 1 - \alpha
\end{aligned}
$$

Haciendo

$$P(Z < z_1) = P(Z > z_2) = \alpha/2\,.$$



Entonces

$$\mathfrak{z}_2 = Z_{\alpha/2} \quad = \frac{\bar{X} - L}{\sigma/\sqrt{n}} \quad \text{y} \quad \mathfrak{z}_1 = -Z_{\alpha/2} = \frac{\bar{X} - U}{\sigma/\sqrt{n}}$$

Así,

$$L = \bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{y} \quad U = \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ .$$

Para una muestra particular se calcula $\bar{x}$ y así, un IC al $100(1-\alpha)\%$ para μ es:

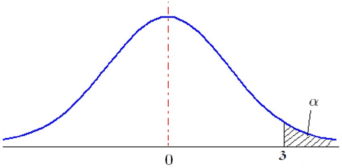
$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\, , \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)\ .$$

Intervalos unilaterales para la media

Para obtener intervalos de confianza unilaterales, de la forma $(l, +\infty)$ o $(-\infty, u)$ se procede así:

Para calcular un IC unilateral superior se debe hallar L tal que $P(\mu > L) = 1 - \alpha$ dado un valor de α, luego

$$
\begin{aligned}
P(\mu > L) &= P(-\mu < -L) = P\left(\bar{X} - \mu < \bar{X} - L\right) \\
&= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - L}{\frac{\sigma}{\sqrt{n}}}\right) \\
&= P(Z < \mathfrak{z}) = 1 - \alpha
\end{aligned}
$$



donde

$$\mathfrak{z} = \frac{\bar{X} - L}{\frac{\sigma}{\sqrt{n}}}$$

Como $\mathfrak{z} = z_\alpha$, despejando L se tiene que $L = \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}}$, donde z_α se obtiene de R. Para una muestra particular se calcula $\bar{x}$ y así, un Intervalo de Confianza Unilateral al $100(1-\alpha)\%$ para μ es

$$\left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty\right) .$$

Similarmente, un Intervalo de confianza unilateral inferior para μ al $100(1-\alpha)\%$ es de la forma:

$$\left(-\infty, \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) .$$

Distribución t-student

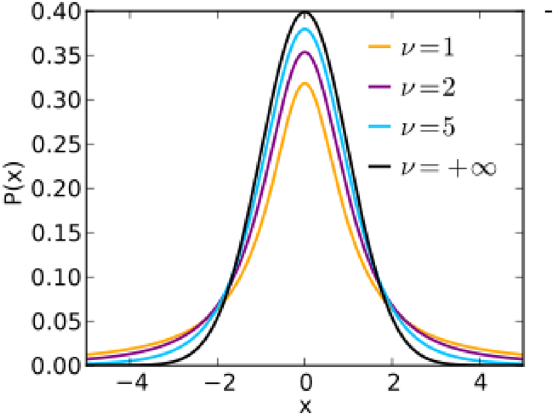
Definición

Sea T una variable aleatoria continua. Diremos que T tiene una distribución t-Student, si su f.d.p. es de la forma:

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\left(\frac{v}{2}\right)\sqrt{\pi\, v}}\left[1+\frac{t^2}{v}\right]^{-\frac{(v+1)}{2}} \quad ; \quad v > 0\,; \quad t \in \mathbb{R}\,.$$

Por notación se escribe $T \sim t\,(v)$. El valor esperado y la varianza son:

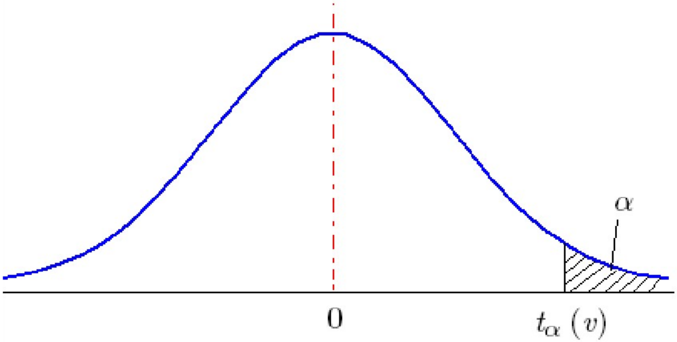
$$E\,[T] = 0 \qquad Var\,[T] = \frac{v}{v-2}\,; \quad v > 2\,.$$



El parámetro de esta distribución se conoce como *grados de libertad* y es usualmente denotado v.

Si $T \sim t(v)$, el cuantil superior α de T, el cual se denota $t_\alpha(v)$ es tal que

$$P(T > t_\alpha(v)) = \alpha\,.$$



IC para μ con σ^2 desconocida y población normal

Caso 2: Suponga que $X_1, \cdots, X_n$ es una m.a. de una $N(\mu, \sigma^2)$. Si σ^2 es conocida se sabe que:

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0\,,\,1).$$

Pero si σ **es desconocida** y n (tamaño de muestra) es pequeña, la distribución de $\bar{X}$ estandarizada NO es normal. Si en la expresión anterior, reemplazamos σ por S, el estadístico resultante es:

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}\,,$$

Cuya distribución es la distribución t-student o distribución t, la cual tiene propiedades similares de la normal estándar, pero con colas más pesadas (más alargadas). La distribución t es una distribución de probabilidad que estima el valor de la media de una muestra pequeña extraída de una población que sigue una distribución normal y de la cual no conocemos su desviación típica.

Si $X_1, \cdots, X_n$ es una m.a. de una $N(\mu, \sigma^2)$ con σ^2 desconocida y n pequeño, se puede mostrar que

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}; \quad \text{con} \quad n-1 \quad \text{grados de libertad}$$

Realizando un proceso similar al caso de muestras aleatorias con σ^2 conocida, se puede encontar que un IC al $100(1-\alpha)\%$ para μ es de la forma:

$$\left(\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right).$$

Para hallar $t_{\alpha/2,(n-1)}$ en **R** se usa la instrucción:
`qt(p=alpha/2, df=n-1, lower.tail=FALSE)`.

IC unilaterales para μ

Un Intervalo de confianza Unilateral superior al $100(1-\alpha)\%$ para μ, con σ^2 desconocida, es de la forma:

$$\left(\bar{x} - t_{\alpha,n-1}\,\frac{s}{\sqrt{n}}\,,\,+\infty\right)\,.$$

Un Intervalo de confianza unilateral inferior para μ al $100(1-\alpha)\%$, con σ^2 desconocida, es de la forma:

$$\left(-\infty\,,\,\bar{x} + t_{\alpha,n-1}\,\frac{s}{\sqrt{n}}\right)\,.$$

Para hallar $t_{\alpha,n-1}$ en **R** se usa la instrucción:
`qt(p=alpha, df=n-1, lower.tail=FALSE).`

IC para μ con σ^2 conocida, población no normal y n grande

Caso 3: Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población **no normal**, con media μ desconocida, **varianza** σ^2 **conocida** y $n \geq 30$.
Por el teorema del límite central se tiene que:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \overbrace{n \to +\infty}^{aprox} \quad N(0, 1).$$

Luego un IC aproximado al $100(1-\alpha)\%$ para μ está dado por:

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

IC para μ con σ^2 desconocida, población no normal y n grande

Caso 4: Sea $X_1, \cdots, X_n$ una muestra aleatoria de una población **no normal**, con media μ desconocida, **varianza** σ^2 **desconocida** y $n \geq 30$.
Por el teorema del límite central y usando el hecho de que $\sigma \approx S$ se tiene que:

$$Z = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \quad \overbrace{n \to +\infty}^{aprox} \quad N(0, 1) .$$

Luego un IC aproximado al $100(1-\alpha)\%$ para μ está dado por:

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

IC unilaterales para μ sin normalidad

Un IC unilateral superior aproximado al $100(1-\alpha)\%$ para μ, con σ^2 **conocida**, es de la forma:

$$\left(\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}}, +\infty\right) \; .$$

Un IC unilateral inferior aproximado para μ al $100(1-\alpha)\%$, con σ^2 **conocida**, es de la forma:

$$\left(-\infty, \bar{x} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) \; .$$

IC unilaterales para μ sin normalidad

Un Intervalo de Confianza Unilateral superior aproximado al $100(1-\alpha)\%$ para μ, con σ^2 **desconocida**, es de la forma:

$$\left(\bar{x} - z_\alpha \frac{s}{\sqrt{n}}, +\infty\right) .$$

Un Intervalo de confianza unilateral inferior aproximado para μ al $100(1-\alpha)\%$, con σ^2 **desconocida**, es de la forma:

$$\left(-\infty, \bar{x} + z_\alpha \frac{s}{\sqrt{n}}\right) .$$

Resumen IC para μ