Estimación por intervalos

Jessica Nathaly Pulzara Mora
jessica.pulzara@udea.edu.co

Departamento de ingeniería de sistemas

UNIVERSIDAD DE ANTIOQUIA

- Por ahora solo sabemos hacer estimación puntual.
- Es útil, pero no proporciona información sobre precisión o confiabilidad.
- También es muy útil conocer el rango de posibles valores para el parámetro de interés.

Intervalo de confianza

Para estimar ese rango se requiere:

- El estimador puntual $\hat{\theta}$.
- La distribución de muestreo de $\hat{\theta}$.

Con lo anterior se determina un intervalo que contiene a θ con cierta seguridad. Éste es el **estimador por intervalo**.

- Un estimador de θ por intervalo es un intervalo real (l, u), lo cual signfica que $l < \theta < u$.
- Si se extraen muchas muestras, cada una proporciona un valor diferente para $\hat{\theta}$ y valores diferentes para los extremos l y u.
- Por lo tanto, los extremos también son variables aleatorias, que vamos a denominar L y U.
- (L, U) es un **intervalo aleatorio**.

Si se conoce $\hat{\theta}$ y su distribución de muestreo, es posible calcular lo siguiente:

$$P(L < \theta < U) = 1 - \alpha, \quad \alpha \in (0, 1)$$

α es conocido. Para una muestra particular, se obtiene el intervalo (l, u) donde se espera que esté el verdadero valor de θ.

El intervalo (l, u) se conoce como **intervalo de confianza** al $100 \times (1 - \alpha)\,\%$ para θ.

l y u son llamados **límites de confianza**.

Interpretación

- Los intervalos de confianza (I.C) dan un rango de valores donde está el verdadero valor de θ.

- Con todas las muestras posibles de una población se pueden calcular valores diferentes del estimador $\hat{\theta}$ y valores de los extremos l y u, es decir, se pueden calcular diferentes intervalos.

- De todos los intervalos, se espera que el $100(1-\alpha)\,\%$ contenga en verdadero valor de θ.

Nivel de confianza

Nivel de confianza: $1-\alpha$

- Un 95 % de confianza implica que el 95 % de todas las muestras darían un intervalo que incluye al parámetro verdadero, y el 5 % de las muestras daría un intervalo que no lo incluyen.

- Un $(1-\alpha)100\,\%$ de confianza implica que el $(1-\alpha)100\,\%$ de todas las muestras darían un intervalo que incluye al parámetro verdadero, y el $\alpha 100\,\%$ de las muestras daría un intervalo que no lo incluye.

- Los niveles de confianza más utilizados son 90 %,95 % y 99 %.

- Mientras más alto el nivel de confianza, más fuerte la creencia de que el valor verdadero del parámetro está en el intervalo.

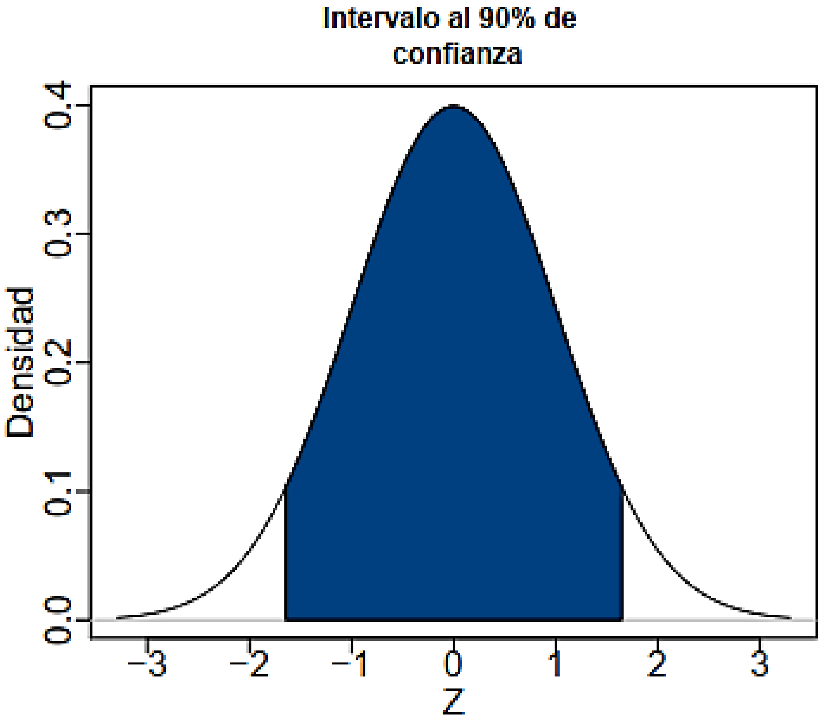
Precisión

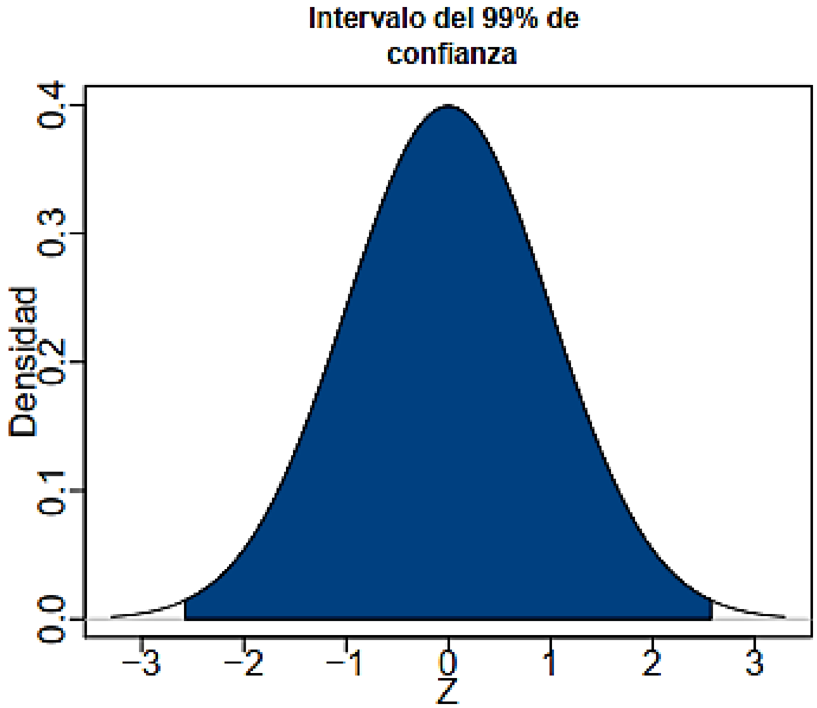
Precisión: se relaciona con la longitud del I.C, da información sobre la precisión del intervalo.

Mientras mayor sea la longitud del I.C, mayor incertidumbre se tendrá sobre el valor estimado.

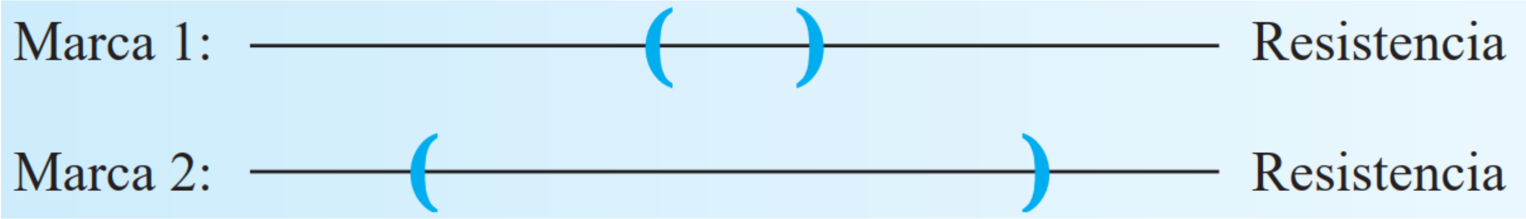
Relación entre la precisión y nivel de confianza

- El **nivel de confianza** y la **precisión** tienen una relación inversa.
- A mayor confianza, más amplio es el intervalo, y menos preciso es.


Intervalo al 90% de
confianza
Densidad
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
Z


Intervalo del 99% de
confianza
Densidad
0.0
0.1
0.2
0.3
0.4
−3
−2
−1
0
1
2
3
Z

Relación entre la precisión y nivel de confianza



Obtención del intervalo de confianza

- $\hat{\theta}$ es un estimador, por lo tanto, depende de los datos y del parámetro de la distribución de la población $(f(x;\theta))$:

$$\hat{\theta} = g(X_1, \ldots, X_n, \theta)$$

- Entonces, se pueden encontrar unas consantes a y b tales que:

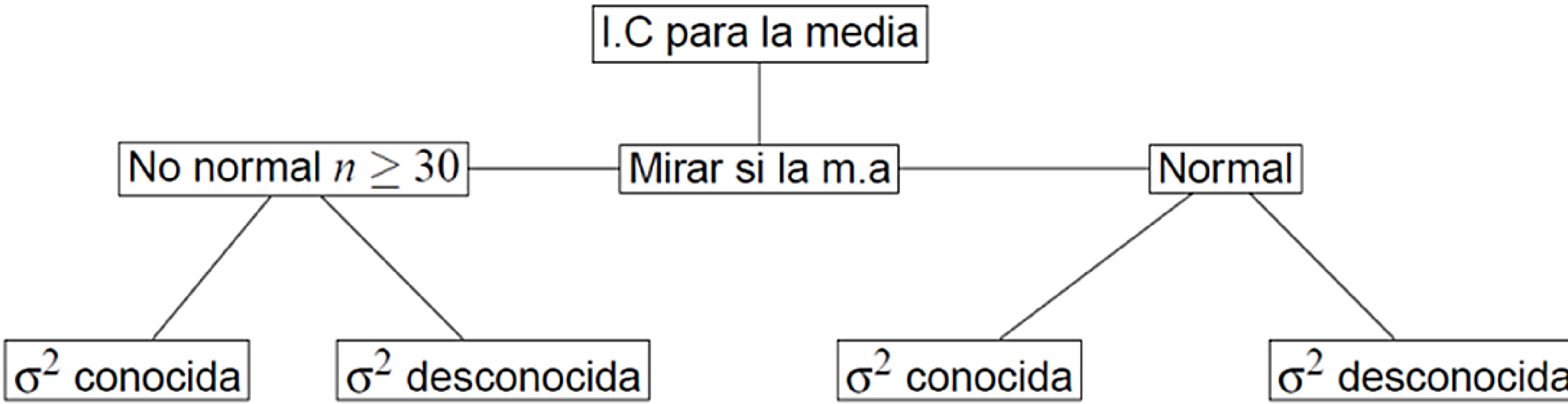
$$P(a < \hat{\theta} < b) = 1 - \alpha$$

- Si se puede despejar la desigualdad del paréntesis, entonces se obtiene:

$$\begin{aligned} P(L < \theta < U) &= 1 - \alpha \\ L &= L(X_1, \ldots, X_n; a) \\ U &= U(X_1, \ldots, X_n; b) \end{aligned}$$

- Para la muestra se obtienen $l = l(X_1, \ldots, X_n)$ y $u = u(X_1, \ldots, X_n)$.

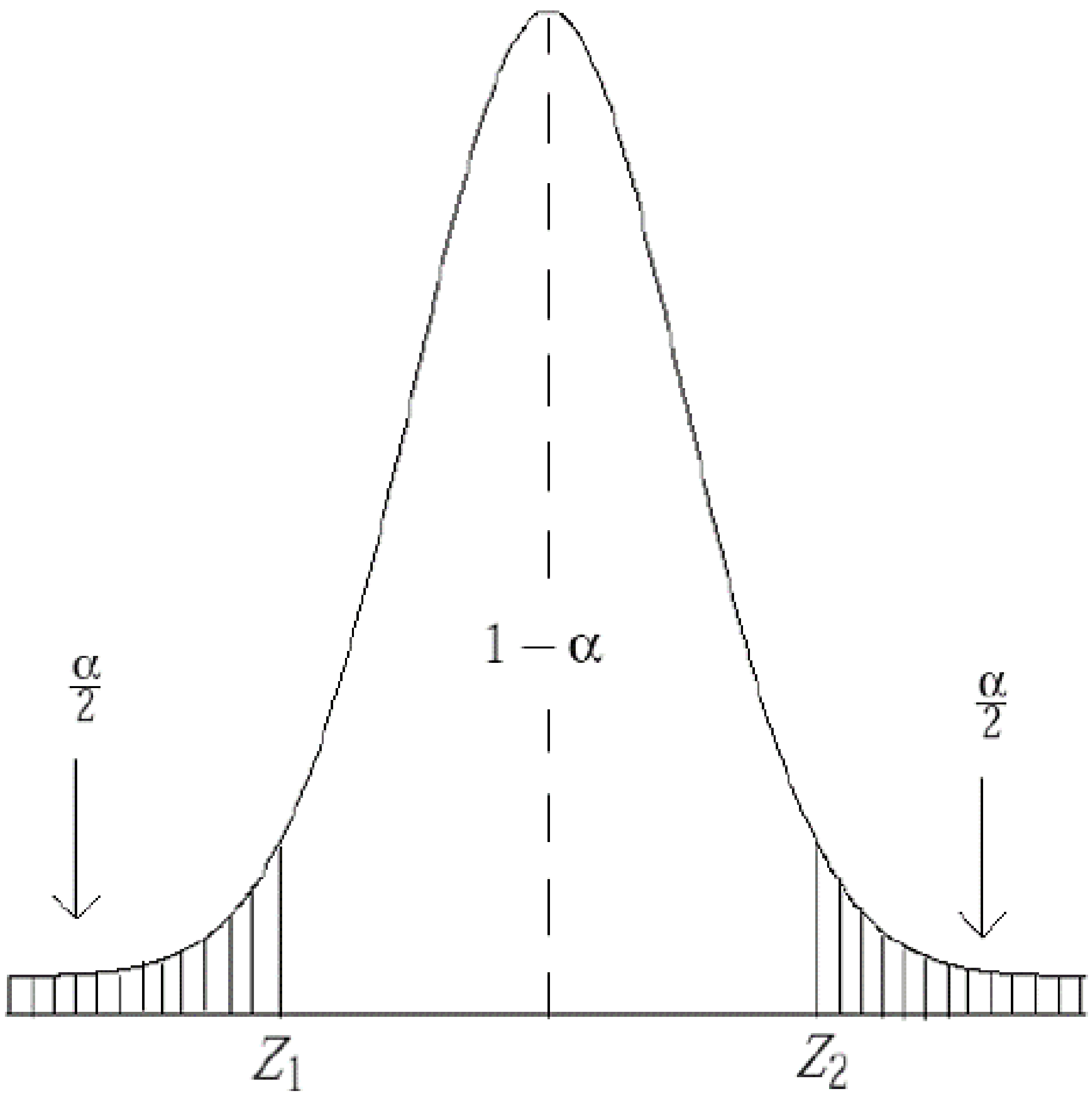
I.C para la media muestral


I.C para la media
Mirar si la m.a
No normal $n \geq 30$
Normal
σ^2 conocida
σ^2 desconocida
σ^2 conocida
σ^2 desconocida

Caso 1

Sea $X_1, X_2, \ldots, X_n$ una muestra que viene de una población NO normal, con n grande y σ^2 **conocida**.Hallamos L y U tales que el intervalo tenga una confianza $100(1-\alpha)\,\%$ para μ:

$$P(L < \mu < U) = 1 - \alpha$$


$\frac{\alpha}{2}$
$1-\alpha$
$\frac{\alpha}{2}$
Z_1
Z_2

I.C para el caso 1:

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

ó

$$\bar{X} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Interpretación

Suponga que se construyó un intervalo para $\bar{X}$, obtenida con una muestra de tamaño n. Este intervalo va de $\bar{X} - 120$ a $\bar{X} + 120$ y la probabilidad de que contenga a μ es 95 %. Esto se escribe así:

$$P(\bar{X} - 120 < \mu < \bar{X} + 120) = 0.95$$

Interpretación: cuando se extraigan varias muestras del mismo tamaño de una población y se calculan los extremos de $(\bar{X} - 120, \bar{X} + 120)$, el 95 % de esos intervalos contiene a μ.

Interpretación

Supongamos que $\bar{X} = 200$

- $P(\bar{X} - 120 < \mu < \bar{X} + 120) = 0.95$ **NO QUIERE DECIR** que la probabilidad de que μ esté entre 80 y 320 es 95 %.
- **Interpretación:** Significa que con un 95 % de confianza el intervalo $(80, 320)$ contiene a μ

Tamaño de muestra

El tamaño de muestra para que la media muestral se encuentre alejada a ε unidades de la media poblacional con una confianza de $1 - \alpha$:

$$P(|\bar{X} - \mu| < \varepsilon) = 1 - \alpha$$

de donde:

$$\boxed{n = \left(\frac{z_{\alpha/2}\sigma}{\varepsilon}\right)^2}$$

Ejemplo

Un fabricante de anillos para pistón ha encontrado con la experiencia que la dispersión de los diámetros es aproximadamente 0.05mm. Él escoge 35 anillos al azar y mide sus diámetros, y calcula un promedio de 74.04 mm.

1. Calcule un I.C al 95 % para el diámetro promedio verdadero de los anillos.
2. Si desea un intervalo con una precisión de 0.01 y una confianza del 95 %, ¿cuál es el tamaño de la muestra que debe usar?

Caso 2

Sea $X_1, X_2, \ldots, X_n$ una muestra que viene de una población NO normal, con n grande y σ^2 **desconocida**. El intervalo con una confianza $100(1-\alpha)\,\%$ para μ es:

$$\mu \in \left(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

ó

$$\bar{X} \mp z_{\alpha/2}\frac{S}{\sqrt{n}}$$

Ejemplo

Se selecciona una muestra aleatoria de 50 botellas de jarabe para la tos de cierta marca y se determina el contenido de alcohol de cada botella. El contenido promedio y la desviación estándar obtenidas fueron 8.6 y 2.87 respectivamente.

- Encuentre el I.C al 95 % para el contenido promedio real.
- ¿Se puede afirmar que el contenido promedio de alcohol en el jarabe para la tos es mayor que 8.5?

Caso 3

Sea $X_1, X_2, \ldots, X_n$ una muestra que viene de una población normal y σ^2 **conocida**. El intervalo con una confianza $100(1-\alpha)\,\%$ para μ es:

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

ó

$$\bar{X} \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Caso 4

Sea $X_1, X_2, \ldots, X_n$ una muestra que viene de una población normal y σ^2 **desconocida** y **muestra pequeña**. En este caso no se cumple el TLC.

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

tiene otra distribución. Si se reemplaza σ por S, ahora llamaremos a la variable T:

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

T tiene una distribución t-student, que es similar a la normal, pero con las **colas más pesadas**

Distribución T-student

La f.d.p para una varaible que se distribuye $t(\nu)$ es:

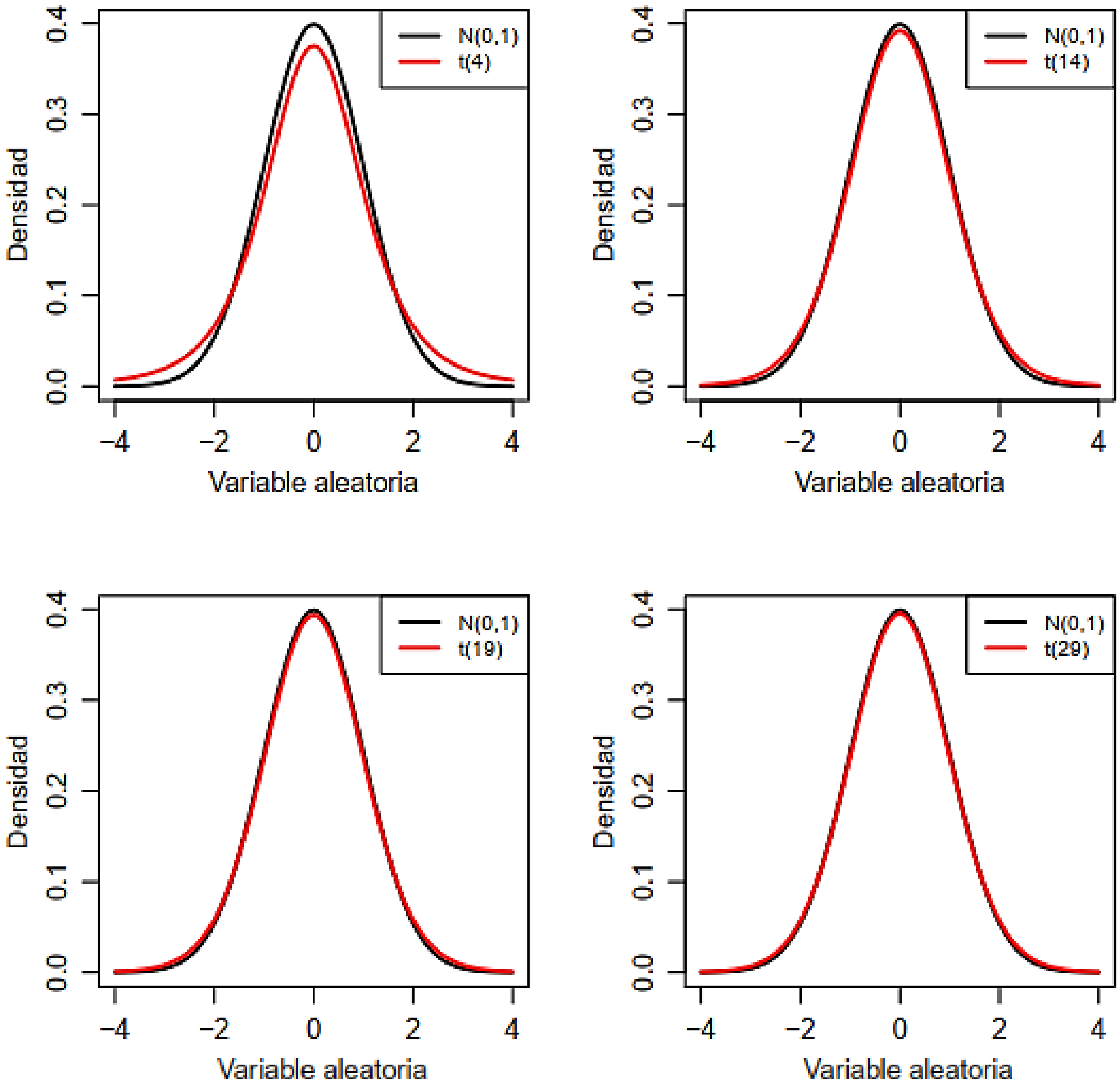
$$f(t) = \frac{\Gamma\frac{\nu+1}{2}}{\frac{\nu}{2}\nu\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}} \quad , \quad \nu > 0, t \in \mathbb{R}$$

$\Gamma(x)$ es la función gamma, ν son grados de libertad. Para el caso que nos interesa, se sabe que

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

.

Osea que $\nu = n-1$ para la variable T.

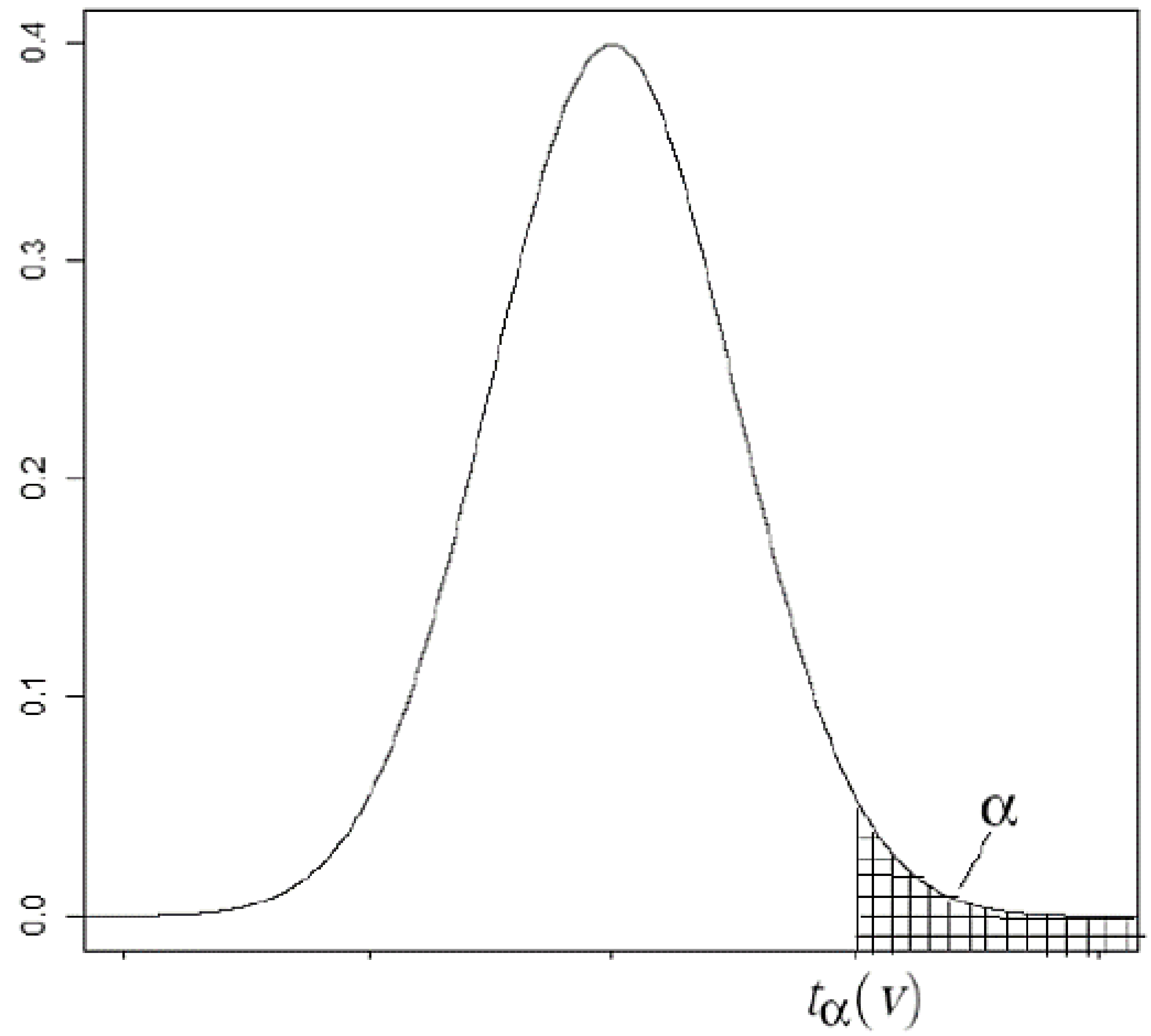

Densidad
Variable aleatoria
N(0,1)
t(4)
t(14)
t(19)
t(29)
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4

Observaciones

- A medida que aumentan el tamaño muestral (y por lo tanto, los grados de libertad), la curva de la t se parece más a la de la normal.

- Para una variable $T \sim t(\nu)$, la tabla reporta el cuantil superior:

$$P(T > t(\nu)) = \alpha$$


0.4
0.3
0.2
0.1
0.0
α
tα(v)

Caso 4

$$\mu \in \left(\bar{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \right)$$

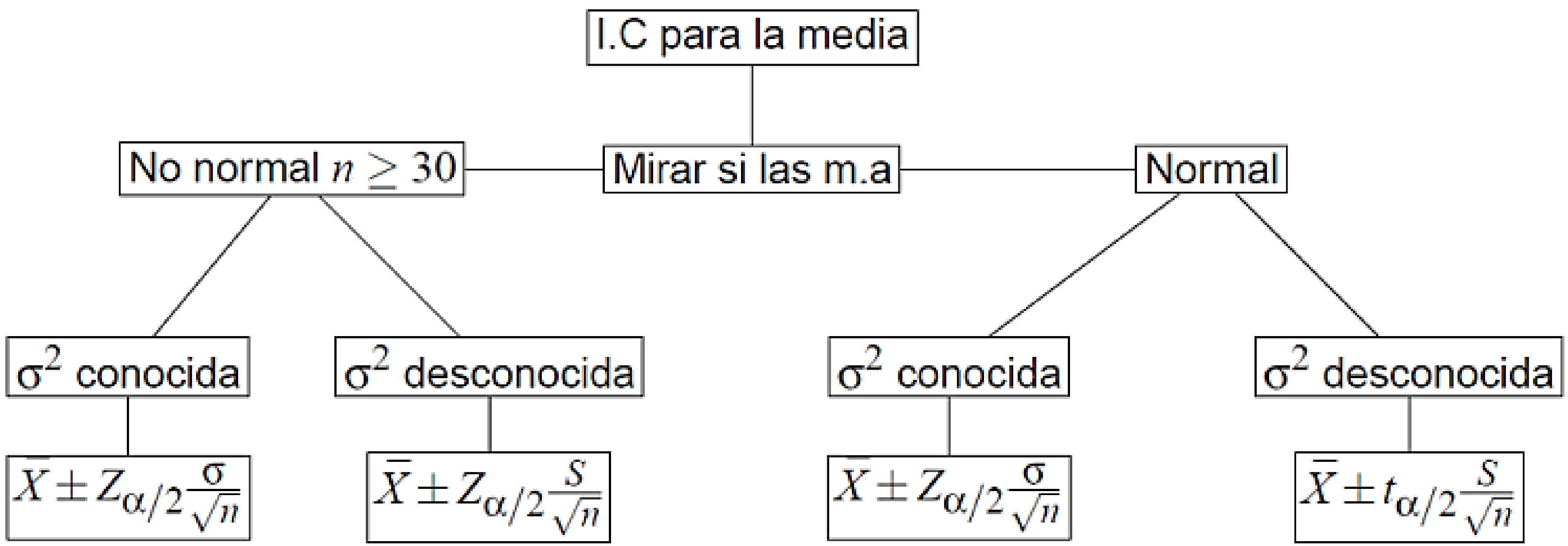
ó

$$\bar{X} \mp t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}$$

Ejemplo

Una máquina produce varillas de metal utilizadas en la suspensión de automóviles. De la experiencia, se ha encontrado que los diámetros de este tipo de varillas se distribuye aproximadamente normal. Se toma una muestra de 15 varillas y se miden sus respectivos diámetros. Se obtiene un diámetro promedio de 8.234 cm y una desviación estándar de 0.0253 cm. Estime el diámetro promedio real de estas varillas usando un I.C al 95 %.

Resumen Intervalos de confianza para la media



I.C para la proporción

Distribución de la proporción

Sea X una variable aleatoria tal que $X \sim Binom(n, p)$, si n es grande, el TLC garantiza lo siguiente

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

$\hat{p} = \frac{X}{n}$ es un estimador insesgado para p, entonces:

$$\frac{X - np}{\sqrt{np(1-p)}} = \frac{n\left(\frac{X}{n} - p\right)}{\sqrt{n}\sqrt{\frac{X}{n}\left(1 - \frac{X}{n}\right)}} = \frac{\frac{X}{n} - p}{\sqrt{\frac{\frac{X}{n}\left(1-\frac{X}{n}\right)}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

I.C para la proporción

Sea $X_1, X_2, \ldots, X_n$ una muestra que viene de cierta población, con n **grande**. Hallemos L y U tales que el intervalo tenga una confianza $100(1-\alpha)\,\%$ para p:

$$P(L < p < U) = 1 - \alpha$$

I.C **aproximado** para p, con n muy grande es:

$$p \in \left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

ó

$$\hat{p} \mp z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Ejemplo

Se lleva a cabo un estudio para determinar la efectividad de una vacuna. Se administra la vacuna a una muestra de 3000 sujetos y 13 contraen la enfermedad. Obtenga un I.C al 90 % para la proporción real de sujetos vacunados que contraen la enfermedad.