UNIVERSIDAD
NACIONAL
DE COLOMBIA

Estadística I

Intervalos de Confianza para μ

aproximados

I.C para μ

Mirar si la m.a

No normal, $n \geq 30$

σ^2 conocida

σ^2 desconocida

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\bar{X} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$\hat{P} \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

Normal

σ^2 conocida

σ^2 desconocida

Intervalos de Confianza para μ

Caso II: m.a de una población normal

☑ σ^2 **conocida:**

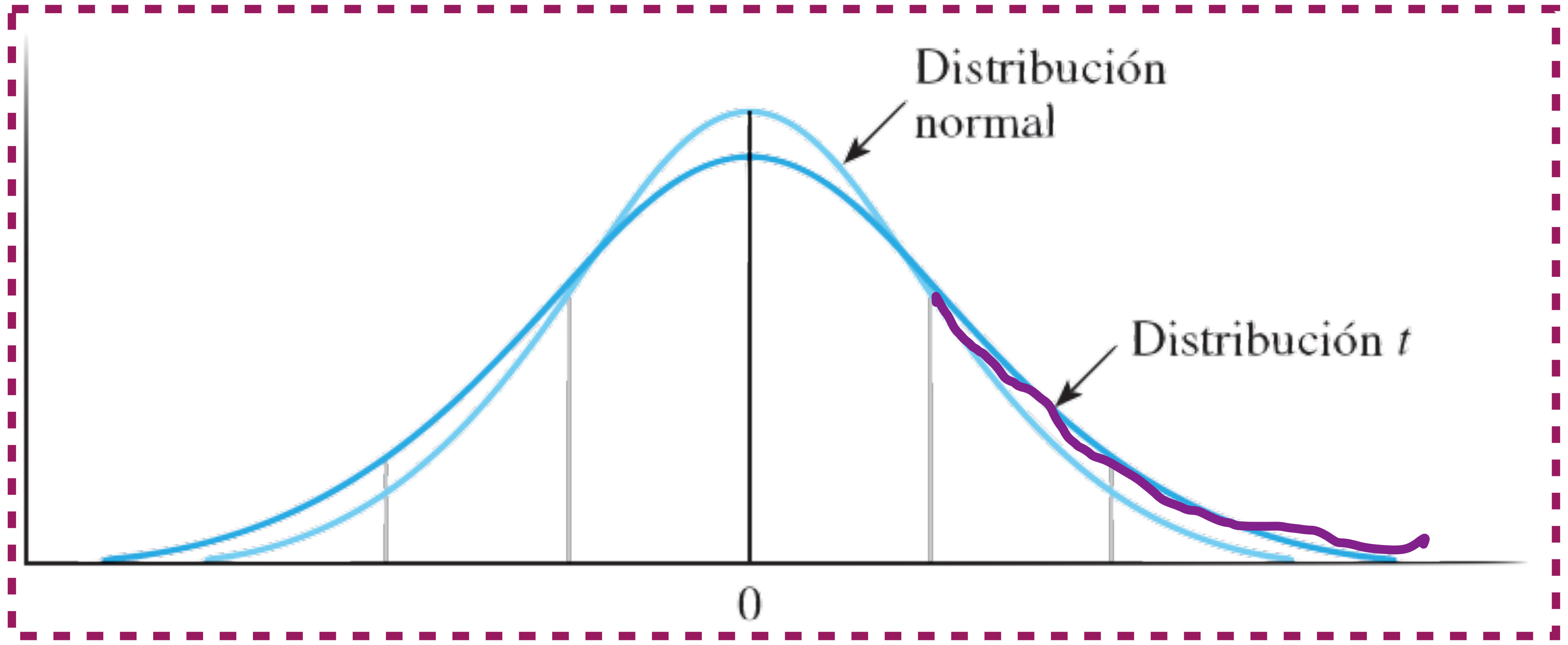
Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal con media μ y varianza σ^2 conocida. Entonces se tiene que:

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

Así que un I.C al $100(1-\alpha)\,\%$ para μ es:

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Intervalos de Confianza para μ



Caso II: m.a de una población normal

☑ σ^2 **desconocida:**

S

Si σ^2 es desconocida, entonces $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ no tiene una distribución normal. Si en la expresión anterior reemplazamos σ por S, el estadístico resultante $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ tiene una distribución cuyas propiedades son similares a las de la distribución normal estándar, pero con colas más pesadas (**distribución T de student**).

Intervalos de Confianza para μ

Caso II: m.a de una población normal

☑ σ^2 **desconocida:**

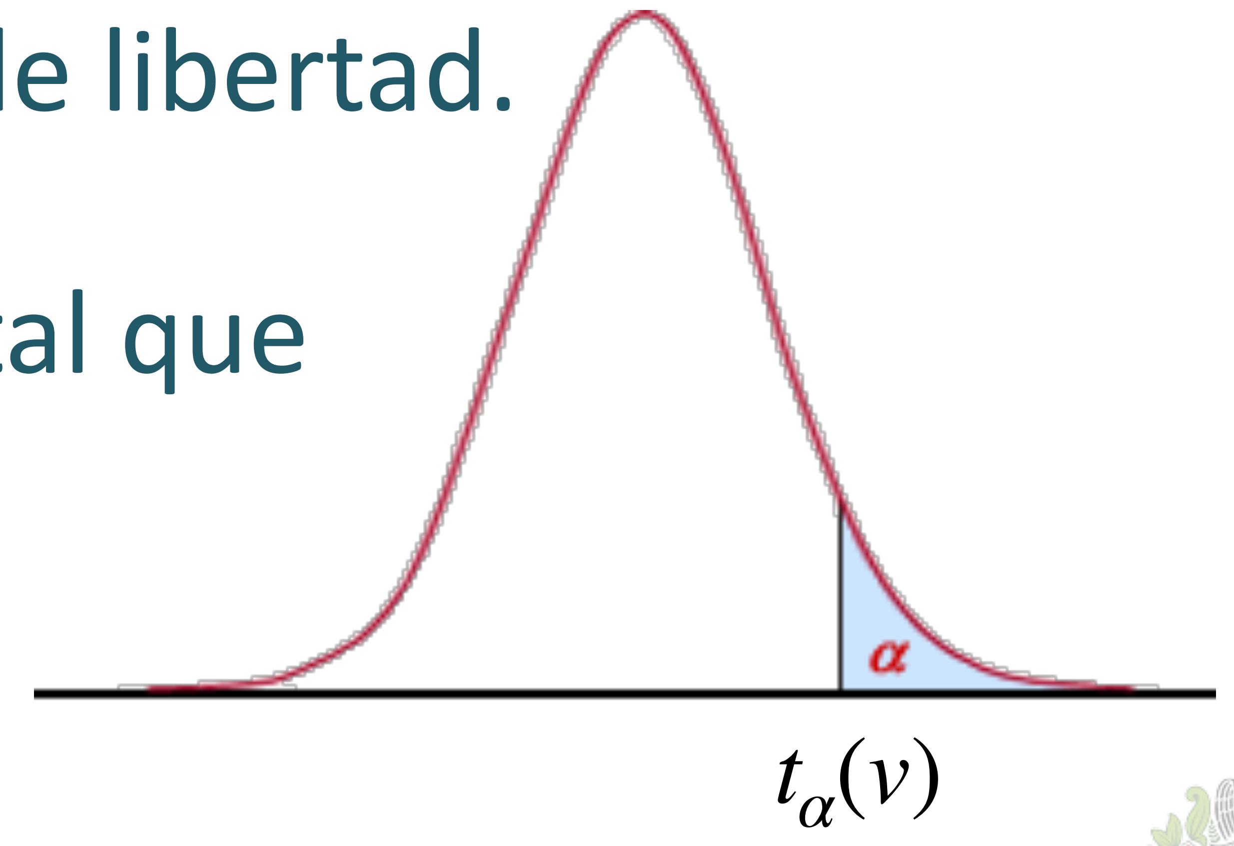
La f.d.p de una variable aleatoria T-student con v grados de libertad es de la forma:

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

Nota: la distribución t-student es el cociente entre una distribución normal estándar y la raíz cuadrada de una Chi-Cuadrado dividida por sus grados de libertad.

Si $T \sim t_{(v)}$, el cuantil α superior, el cual se denotará $t_\alpha(v)$ es tal que

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



Intervalos de Confianza para μ

Caso II: m.a de una población normal

σ^2 desconocida:

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal con media μ y varianza σ^2 desconocida. Entonces se tiene que:

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

Así que un I.C al $100(1-\alpha)\,\%$ para μ es:

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

UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza para μ

- I.C para μ
  - Mirar si la m.a
    - No normal, $n \geq 30$
      - σ^2 conocida: $\bar{X} \pm Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
      - σ^2 desconocida: $\bar{X} \pm Z_{\alpha/2}\dfrac{S}{\sqrt{n}}$
    - Normal
      - σ^2 conocida: $\bar{X} \pm Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
      - σ^2 desconocida: $\bar{X} \pm t_{\alpha/2,(n-1)}\dfrac{S}{\sqrt{n}}$

$$\hat{P} \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}$$

Intervalos de Confianza para μ

Ejemplo:

Se supone que la vida útil de una batería usada en un marcapasos cardíaco tiene una distribución normal. Una muestra aleatoria de 10 baterías se somete a una prueba de vida útil acelerada, manteniendo las baterías operando de manera continua a una temperatura elevada hasta que ocurra una falla. Se obtienen los siguientes tiempos de vida:

25.5	26.1
26.8	23.2
24.2	28.4
25.0	27.8
27.3	25.7

a. ¿Qué se puede concluir con una confianza del 90% sobre la vida útil media en la prueba acelerada?
b. El fabricante quiere asegurarse de que la vida útil media de la batería en la prueba excede las 25 h ¿Qué conclusión se puede sacar con la información obtenida en el literal a) que sea de interés para el fabricante?

Intervalos de Confianza

x_i : Vida útil de la batería i

$X \sim N(\mu, \sigma^2)$

$$\bar{X} = \frac{\sum_{i=1}^{10} x_i}{n} = \frac{\sum_{i=1}^{10} x_i}{10} = 26$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{10} (x_i - \bar{x})^2}$$

$$S = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (x_i - 26)^2} = 1.62$$

$1-\alpha = 0.9$

$\Rightarrow \alpha = 0.1$

$\Rightarrow \alpha/2 = 0.05$

$$t_{0.05}(n-1) = t_{0.05}(9) = 1.833$$

Así, el I.C del 90% para μ es:

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

$$26 \pm (1.833) \frac{1.62}{\sqrt{10}} \Rightarrow (25.06, 26.93)$$

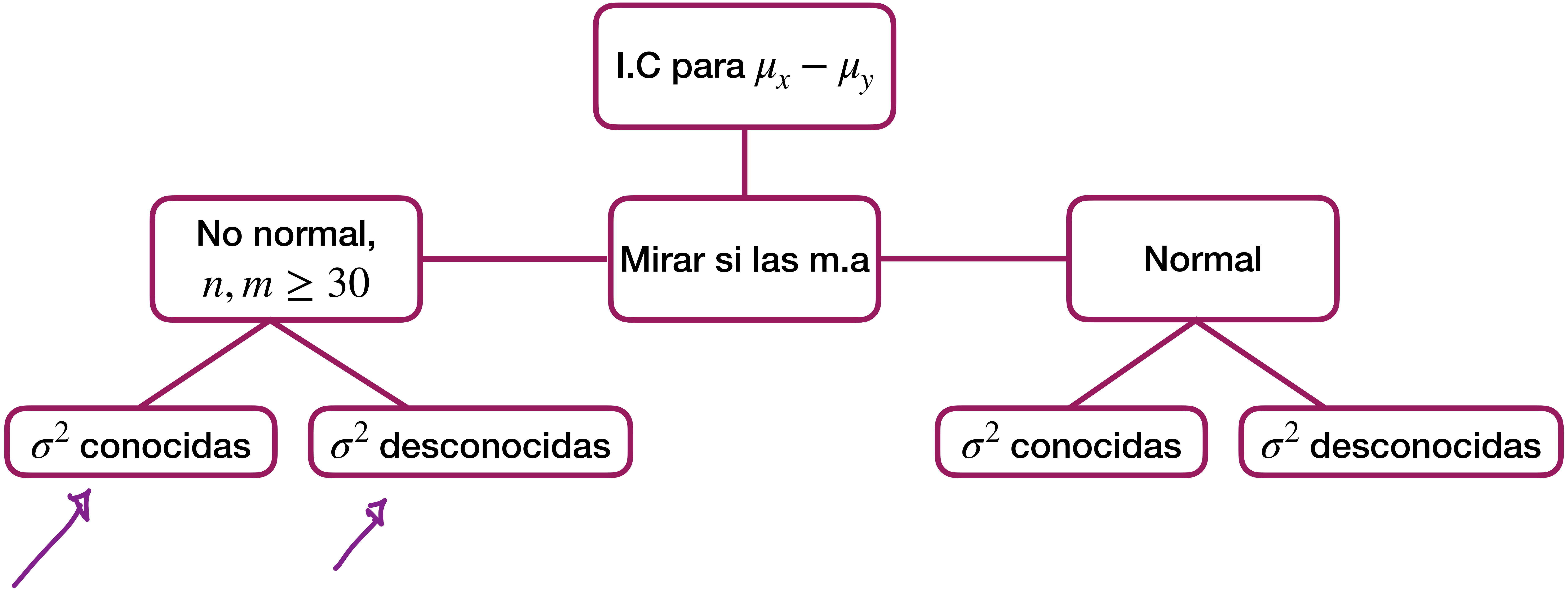
UNIVERSIDAD NACIONAL DE COLOMBIA

Intervalos de Confianza

I.C para la diferencia de Medias

En muchas ocasiones surge la necesidad de comparar el comportamiento de una variable en dos poblaciones por medio de sus medias. Por ejemplo mirar si el comportamiento medio es más grande o más pequeño que en la otra población.

I.C para la diferencia de medias $\mu_x - \mu_y$



I.C para la diferencia de medias $\mu_x - \mu_y$

$\mu \to \bar{x}$ $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Caso I: poblaciones no normales con muestras grandes

Varianzas conocidas:

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población con media μ_x y varianza σ_x^2 conocida. Sea $Y_1, Y_2, \ldots, Y_m$ otra m.a de una población con media μ_y y varianza σ_y^2 conocida. El interés recae en construir un intervalo al $100(1-\alpha)\,\%$ para $\mu_x - \mu_y$.

$\bar{X} - \bar{Y}$

Como ambas muestras son independientes, $\bar{X}$ y $\bar{Y}$ también lo son, así que:

$$\begin{cases} E\left[\bar{X}-\bar{Y}\right] = E\left[\bar{X}\right] - E\left[\bar{Y}\right] = \mu_x - \mu_y \\ V\left[\bar{X}-\bar{Y}\right] = V\left[\bar{X}\right] + V\left[\bar{Y}\right] = \dfrac{\sigma_x^2}{n} + \dfrac{\sigma_y^2}{m} \end{cases}$$

UNIVERSIDAD NACIONAL DE COLOMBIA

I.C para la diferencia de medias $\mu_x - \mu_y$

Caso I: poblaciones no normales con muestras grandes

Varianzas conocidas:

Por el TLC

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim N(0,1)$$

Así que un I.C aproximado al $100(1-\alpha)\%$ para $\mu_x - \mu_y$ es:

$$\boxed{(\bar{X} - \bar{Y}) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

I.C para la diferencia de medias $\mu_x - \mu_y$

Caso I: poblaciones no normales con muestras grandes

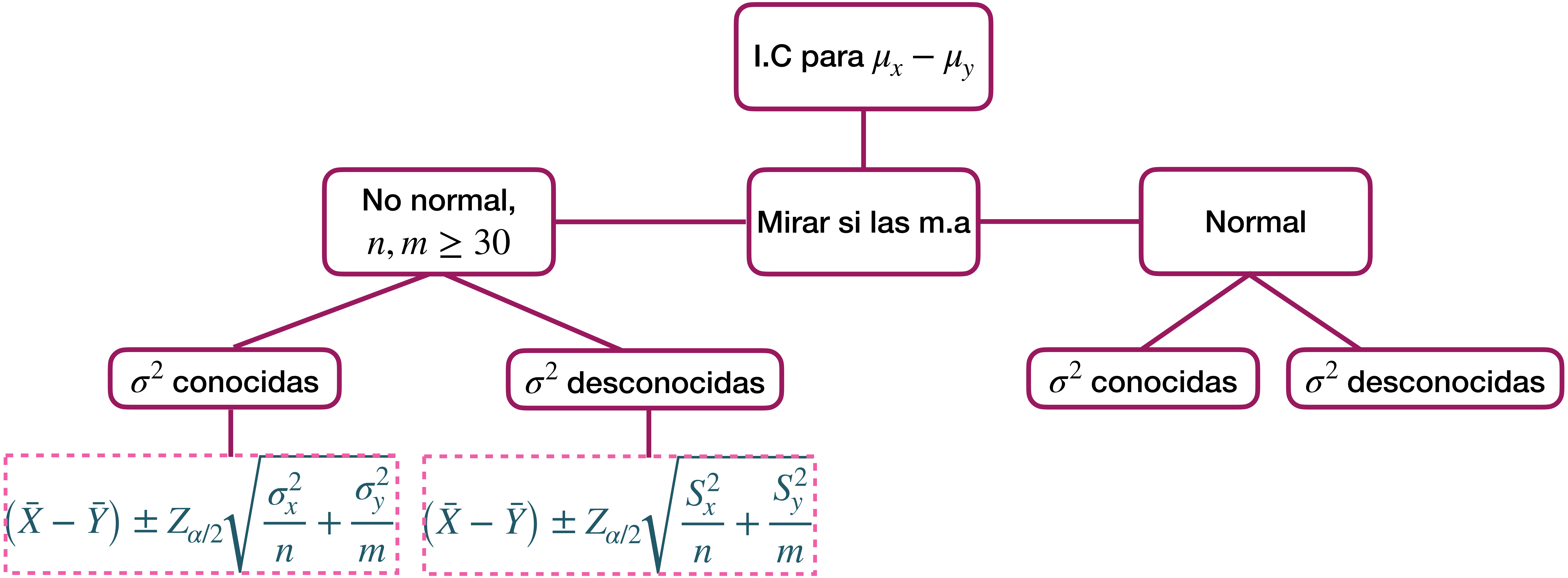
☑ Varianzas desconocidas:

Como $n, m \to \infty$ se puede usar S_x^2 y S_y^2 así que un I.C aproximado al $100(1-\alpha)\,\%$ para $\mu_x - \mu_y$ es:

$$(\bar{X} - \bar{Y}) \pm Z_{\alpha/2}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$$

UNIVERSIDAD NACIONAL DE COLOMBIA

I.C para la diferencia de medias $\mu_x - \mu_y$



I.C para la diferencia de medias $\mu_x - \mu_y$

Caso II: poblaciones normales

☑ **Varianzas conocidas:**

Sea $X_1, X_2, \ldots, X_n$ y $Y_1, Y_2, \ldots, Y_m$ dos muestras aleatorias de dos distribuciones normales independientes, con medias μ_x y μ_y, varianzas σ_x^2 y σ_y^2 conocidas. El Intervalo de confianza al $100(1-\alpha)\,\%$ para $\mu_x - \mu_y$ está dado por:

$$\left(\bar{X} - \bar{Y}\right) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

UNIVERSIDAD NACIONAL DE COLOMBIA

I.C para la diferencia de medias $\mu_x - \mu_y$

Caso II: poblaciones normales

Varianzas desconocidas:

Si σ_x^2 y σ_y^2 son desconocidas, es necesario usar una distribución que tenga en cuenta el desconocimiento de las varianzas poblacionales bajo normalidad. Esa es la distribución T de student. Dependiendo de la relación entre las varianzas poblacionales, se distinguen dos casos:

UNIVERSIDAD NACIONAL DE COLOMBIA

I.C para la diferencia de medias $\mu_x - \mu_y$

Caso II: poblaciones normales

☑ **Varianzas desconocidas:**

▶ Si $\sigma_x^2 = \sigma_y^2$:

Para hallar un I.C al $100(1-\alpha)\,\%$ se debe usar el hecho de que

$$\frac{(\bar{X}-\bar{Y}) - \left(\mu_x - \mu_y\right)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$$

Donde $S_p = \sqrt{\dfrac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}}$, así un I.C al $100(1-\alpha)\,\%$ para $\mu_x - \mu_y$ es:

$$(\bar{X}-\bar{Y}) \pm t_{\alpha/2,(n+m-2)} S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$$

UNIVERSIDAD NACIONAL DE COLOMBIA

I.C para la diferencia de medias $\mu_x - \mu_y$

Caso II: poblaciones normales

Varianzas desconocidas:

- Si $\sigma_x^2 \neq \sigma_y^2$:

En este caso se tiene que:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim t_v \quad \text{donde} \quad \boxed{v = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{\left(\frac{S_x^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_y^2}{m}\right)^2}{m-1}}}$$

Así un I.C al $100(1-\alpha)\,\%$ para $\mu_x - \mu_y$ es:

$$\underbrace{(\bar{X} - \bar{Y})} \pm t_{\alpha/2,(v)} \underbrace{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

I.C para la diferencia de medias $\mu_x - \mu_y$

I.C para $\mu_x - \mu_y$

Mirar si las m.a

- No normal, $n, m \geq 30$
  - σ^2 conocidas

    $$\left(\bar{X} - \bar{Y}\right) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

  - σ^2 desconocidas

    $$\left(\bar{X} - \bar{Y}\right) \pm Z_{\alpha/2}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$$

- Normal
  - σ^2 conocidas

    $$\left(\bar{X} - \bar{Y}\right) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

  - σ^2 desconocidas

    ▶ Si $\sigma_x^2 = \sigma_y^2$:

    $$\left(\bar{X} - \bar{Y}\right) \pm t_{\alpha/2,(n+m-2)} S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$$

    ▶ Si $\sigma_x^2 \neq \sigma_y^2$:

    $$\left(\bar{X} - \bar{Y}\right) \pm t_{\alpha/2,(v)}\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}$$

I.C para la diferencia de medias $\mu_x - \mu_y$

Ejemplo:

Dos técnicos de control de calidad midieron el terminado superficial de una pieza metálica, obteniendo los datos que se muestran abajo. Suponga que las mediciones tienen una distribución normal.

Técnico 1	1.45	1.37	1.21	1.54	1.48	1.29	1.34

Técnico 2	1.54	1.41	1.56	1.37	1.20	1.31	1.27	1.35

Suponiendo que las varianzas de las mediciones de los dos técnicos son iguales, ¿es posible concluir que no existe una diferencia real entre las mediciones medias del terminado superficial hechas por los dos técnicos? (Use un nivel de confianza del 95%)

I.C para la diferencia de medias $\mu_x - \mu_y$

X_i: mediciones terminado sop. técnico 1. Y

Y_i: " " " " 2. X

$X, Y \sim$ Normal.

$\sigma_x^2 = \sigma_Y^2$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{\sum_{i=1}^{7} x_i}{7} = 1.383$$

$$n = 7$$

$$S_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{7}(x_i - \bar{x})^2} = 0.115$$

$$\bar{Y} = 1.376$$

$$m = 8$$

$$S_Y = 0.125$$

$$1-\alpha = 0.95$$

$$\alpha/2 = 0.025$$

$$t_{\alpha/2}(n+m-2) = t_{0.025}^{(7+8-2)}$$

$$= t_{0.025,\underbrace{13}} = \underbrace{2.160}$$

$$(\underbrace{\bar{X}-\bar{Y}}) \pm t_{\alpha/2}(n+m-2)\, S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$S_p = \sqrt{\frac{(n-1)S_x^2 + (m-1)S_Y^2}{n+m-2}} = \sqrt{\frac{(7-1)(0.115)^2 + (8-1)(0.125)^2}{13}} = \underbrace{0.120}$$

I.C para la diferencia de medias $\underbrace{\mu_x - \mu_y}$

$$(1.383 - 1.376) \pm (2.160)(0.120)\sqrt{\frac{1}{7}+\frac{1}{8}}$$

$$(-0.127,\ 0.1471)$$

$- , -$ $\quad$ $- , +$ $\quad$ $+ , +$

$(\quad) (\quad \underset{0}{|} \quad) (\quad)$

1. $\mu_x - \mu_y < 0 \Rightarrow \underline{\mu_x < \mu_y}$
2. $\mu_x - \mu_y > 0 \Rightarrow \mu_x > \mu_y$
3. $\mu_x - \mu_y = 0 \Rightarrow \mu_x = \mu_y$

$X - T1$
$Y - T2.$ $\longrightarrow (+ \ , + \ )$ $\boxed{T1 > T2}$

$\mu_x - \mu_y > 0 \Rightarrow \underbrace{\mu_x > \mu_y}$

$X \to T2$
$Y \to T1$ $\longrightarrow (- \ , - \ )$ $\mu_x - \mu_y < 0 \Rightarrow \mu_x < \mu_y$ $\boxed{T2 < T1}$

I.C para la diferencia de medias $\mu_x - \mu_y$

Ejemplo:

Se usan dos procesos de endurecimiento diferentes -1) templado en agua salada y 2) templado en aceite- en muestras de un tipo particular de aleación metálica. Se tomaron 10 muestras de la dureza obtenida con el proceso 1, las cuales arrojaron una dureza promedio de 147.6 y una desviación estándar de 4.97 y 10 muestras de la dureza obtenida con el proceso 2 arrojando esta vez una dureza promedio de 149.4 y una desviación estándar de 5.46. Suponiendo que las durezas con los 2 procesos se distribuyen normal con varianzas diferentes. Construya un intervalo de confianza del 98% para la diferencia entre la dureza media obtenida con en proceso 1 y la dureza media obtenida con el proceso 2. ¿Se puede afirmar que la dureza media obtenida con el proceso 1 es inferior a la dureza media obtenida con el proceso 2 en más de una unidad?

μ_x μ_y $(-0.\ldots, -0.\ldots)$ $(-1.\ldots, -2.\ldots)$

I.C para la diferencia de medias $\mu_x - \mu_y$

X: dureza obtenida proceso 1. ~ Normales

Y: " " " 2.

$n = 10$, $\bar{X} = 147.6$ $S_x = 4.97$ $\sigma_x^2 \neq \sigma_y^2$ $t_{0.01(18)} = 2.552$

$m = 10$, $\bar{Y} = 149.4$ $S_Y = 5.46$ $1-\alpha = 0.98$

$\alpha/2 = 0.01$

Un I.C para $\mu_x - \mu_y$ es:

$$(\bar{X}-\bar{Y}) \pm t_{\alpha/2}(\nu)\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}} \Rightarrow (147.6-149.4) \pm 2.552\sqrt{\frac{(4.97)^2}{10}+\frac{(5.46)^2}{10}}$$

$$\Rightarrow (-7.75,\ 4.15)$$

$$v = \frac{\left(\frac{S_x^2}{n} + \frac{S_y^2}{m}\right)^2}{\frac{\left(\frac{S_x^2}{n}\right)^2}{n-1} + \frac{\left(\frac{S_y^2}{m}\right)^2}{m-1}} = 17.84 \approx 18$$

Gracias

Clase preparada por: Verónica Guarín