Estadística II

4. Estimación por Intervalos

ADMINISTRACIÓ I DIRECCIÓ D'EMPRESES

Introducción

Estimación Puntual ➡ $\boxed{e = \hat{\vartheta} - \vartheta}$ ➡ Estimación por Intervalo

Estimación por Intervalo para el parámetro ϑ:

$$\boxed{P(\hat{\vartheta}_I \leq \vartheta \leq \hat{\vartheta}_S) = 1 - \alpha \qquad \hat{\vartheta}_I < \hat{\vartheta}_S}$$

- Nivel de Confianza: $1 - \alpha$
- Amplitud del intervalo: $A = \hat{\vartheta}_S - \hat{\vartheta}_I$
- Límites de confianza: $\hat{\vartheta}_S$; $\hat{\vartheta}_I$
- Error de estimación: $e = \hat{\vartheta} - \vartheta$
- Nivel de significación: α

Introducción

Fijado el nivel de confianza en $(1-\alpha)\%$ se trata de determinar los límites del Intervalo de Confianza:

$$\boxed{P(\hat{\vartheta}_I \leq \vartheta \leq \hat{\vartheta}_S) = 1-\alpha \qquad \hat{\vartheta}_I < \hat{\vartheta}_S}$$

Los extremos ($\hat{\vartheta}_S$; $\hat{\vartheta}_I$) dependen de:

- El estimador del parámetro (media muestral, varianza muestral, proporción muestral, etc.)
- La distribución de probabilidad del estimador (Normal, t-Student, Chi-cuadrado, F Snedecor)
- El tamaño de la muestra

Introducción

Procedimiento para obtener un intervalo para un parámetro

Por ejemplo, la media poblacional:

1. Seleccionar el Estimador del parámetro y su distribución

$$X \approx N\left(\mu, \sigma^2\right) \Rightarrow \overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

2. Determinar los valores a y b tales que

$$P(a \leq \overline{X} \leq b) = 1 - \alpha$$

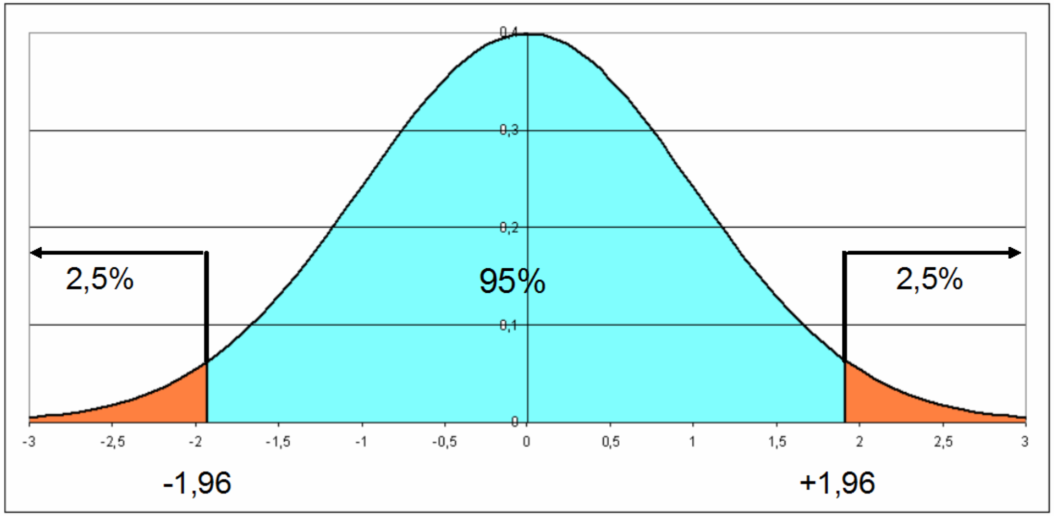
3. Estandarizar los valores

$$P\left(\frac{a-\mu}{\sigma/\sqrt{n}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{b-\mu}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

Introducción

$$P\left(\frac{a-\mu}{\sigma/\sqrt{n}} \le \underbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}_{Z\approx N(0,1)} \le \frac{b-\mu}{\sigma/\sqrt{n}}\right)=1-\alpha$$

$$P\left(-Z_{1-\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le +Z_{1-\alpha/2}\right)=1-\alpha$$



Introducción

$$P\left(-1{,}96 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1{,}96\right) = 0{,}95$$

$$P\left(\overline{X} - 1{,}96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1{,}96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0{,}95$$

Generalizar

$$P\left(\overline{X} - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Introducción

$$\mu \in \left[\overline{X} - Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} ; \overline{X} + Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

$$\overline{X} \pm Z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

En distribuciones muestrales simétricas**:**

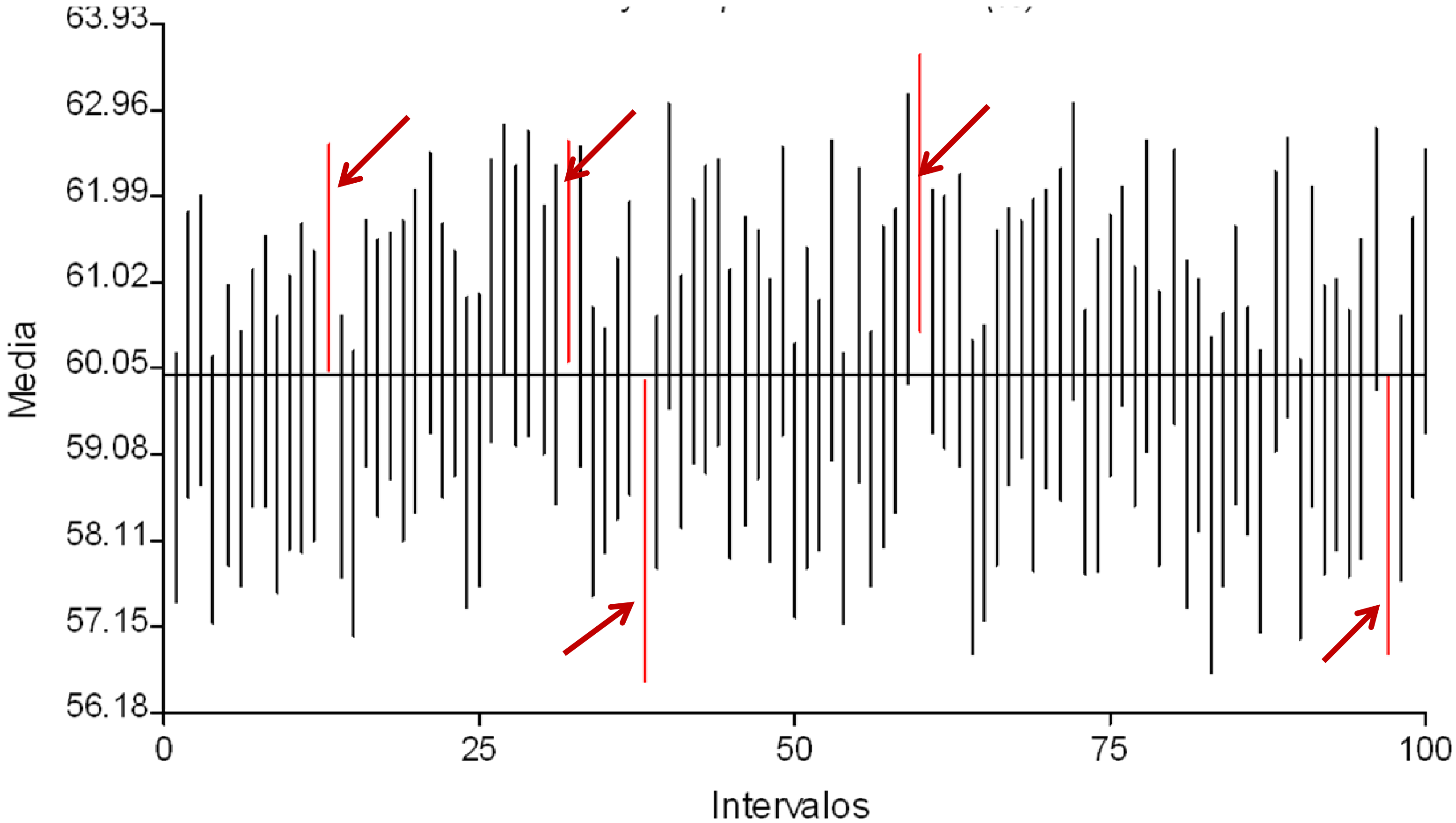
ESTIMADOR ± coef CONFIANZA · Error estándar del Estimador

ESTIMADOR ± error de estimación

Interpretación Nivel de Confianza

<u>Simulación: Intervalo de confianza para la media poblacional</u>

Media Poblacional 60,05 / 100 muestras / Nivel de confianza 95%

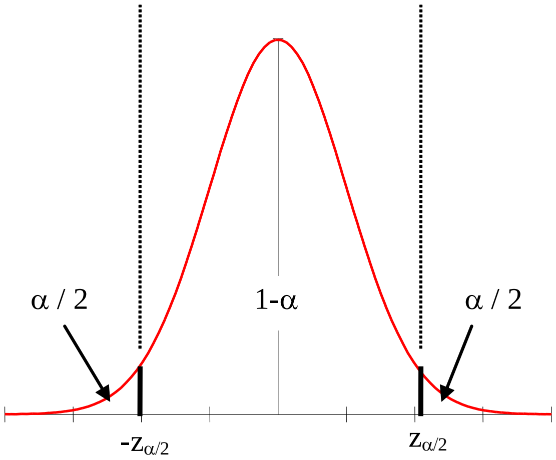


Definición de intervalo de confianza

I.C. PARA LA MEDIA EN POBLACIONES NORMALES.

A.-La varianza poblacional es conocida:

$$\bar{X} \approx N\left(\mu, \sigma^2/n\right) \Rightarrow z_{\bar{x}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \approx N(0;1)$$



$$P\left(-z_{\alpha/2} < z_{\bar{x}} < z_{\alpha/2}\right) = 1-\alpha$$

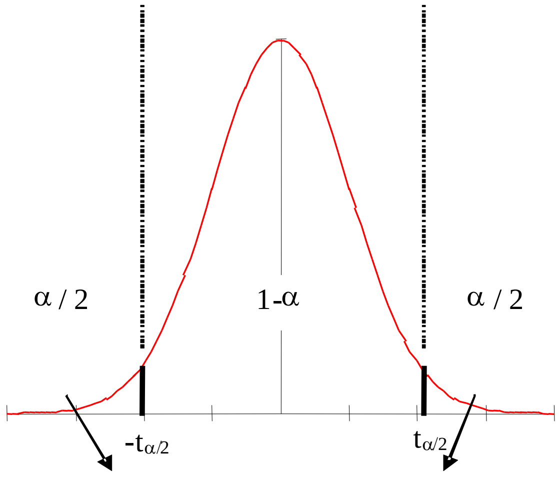
$$\boxed{P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha}$$

Definición de intervalo de confianza

INTERVALOS DE CONFIANZA PARA LA MEDIA EN POBLACIONES NORMALES.

B.-La varianza poblacional es desconocida:

$$\bar{X} \approx N\left(\mu, \sigma^2/n\right) \quad \Rightarrow \quad t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \approx t_{(n-1)}$$



$$P\left(-t_{(n-1),\alpha/2} < t_{\bar{x}} < t_{(n-1),\alpha/2}\right) = 1-\alpha$$

$$\boxed{P\left(\bar{x} - t_{(n-1)\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1)\alpha/2}\frac{s}{\sqrt{n}}\right) = 1-\alpha}$$

Definición de intervalo de confianza

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS POBLACIONALES.

A.-Las varianzas poblacionales conocidas:

$$\left(\overline{X} - \overline{Y}\right) \approx N\left(\mu_x + \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right) \Rightarrow Z = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_x + \mu_y\right)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \approx N(0,1)$$

$$\boxed{P\left((\bar{x} - \bar{y}) - z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} < \mu_x - \mu_y < (\bar{x} - \bar{y}) + z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right) = 1 - \alpha}$$

Definición de intervalo de confianza

B.-Las varianzas poblacionales son iguales, pero no conocidas:

$$\left(\overline{X} - \overline{Y}\right) \approx N\left(\mu_x + \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right) \Rightarrow t = \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_x + \mu_y\right)}{\sqrt{S_p^2 \cdot \left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} \approx t_{(n_x + n_y - 2)}$$

Varianza muestral ponderada, $S_p^2 = \dfrac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$

$$\boxed{P\left((\bar{x} - \bar{y}) - t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)} < (\mu_x - \mu_y) < (\bar{x} - \bar{y}) + t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_x} + \frac{1}{n_y}\right)} \right) = 1 - \alpha}$$

Definición de intervalo de confianza

C.-Las varianzas poblacionales son distintas y no conocidas

$n_x < 30$ y $n_y < 30$:

$$\left(\overline{X}-\overline{Y}\right) \approx N\left(\mu_x + \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right) \Rightarrow t = \frac{\left(\overline{X}-\overline{Y}\right) - \left(\mu_x + \mu_y\right)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}} \approx t_{(v)}$$

$$v = \frac{\left(\frac{S_x^2}{n_x} + \frac{S_x^2}{n_y}\right)^2}{\frac{\left(S_x^2 / n_x\right)^2}{n_x - 1} + \frac{\left(S_y^2 / n_y\right)^2}{n_y - 1}}$$

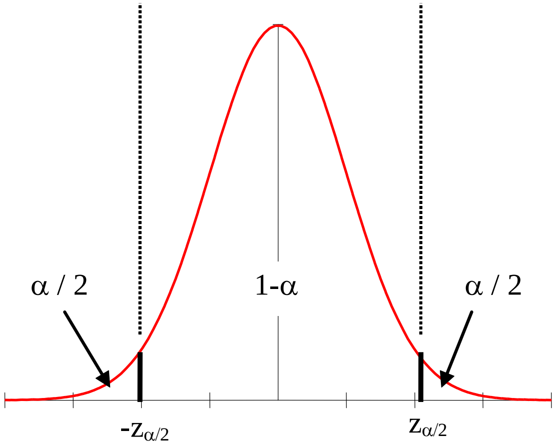
$$P\left((\bar{x} - \bar{y}) - t_{(\gamma)\alpha/2} \sqrt{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)} < (\mu_x - \mu_y) < (\bar{x} - \bar{y}) + t_{(\gamma)\alpha/2} \sqrt{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)} \right) = 1 - \alpha$$

Definición de intervalo de confianza

INTERVALO DE CONFIANZA DE LA PROPORCIÓN POBLACIONAL.

$$\hat{p} \approx N\left(p, \frac{p \cdot q}{n} \right) \Rightarrow Z = \frac{\hat{p} - p}{\sqrt{p \cdot q / n}} \approx N(0,1)$$

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{p \cdot q}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{p \cdot q}{n}} \right) = 1 - \alpha$$



$$\boxed{P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \right) = 1 - \alpha}$$

Definición de intervalo de confianza

INTERVALO DE CONFIANZA PARA DIFERENCIA DE PROPORCIONES POBLACIONALES.

$$(\hat{p}_x - \hat{p}_y) \approx N\left(p_x - p_y, \frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y} \right)$$

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{p_x q_x}{n_x} + \frac{p_y q_y}{n_y}}} \approx N(0,1)$$

$$P\left((\hat{p}_x - \hat{p}_y) - z_{\alpha/2}\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}} < p_x - p_y < (\hat{p}_x - \hat{p}_y) + z_{\alpha/2}\sqrt{\frac{\hat{p}_x \hat{q}_x}{n_x} + \frac{\hat{p}_y \hat{q}_y}{n_y}} \right) = 1 - \alpha$$

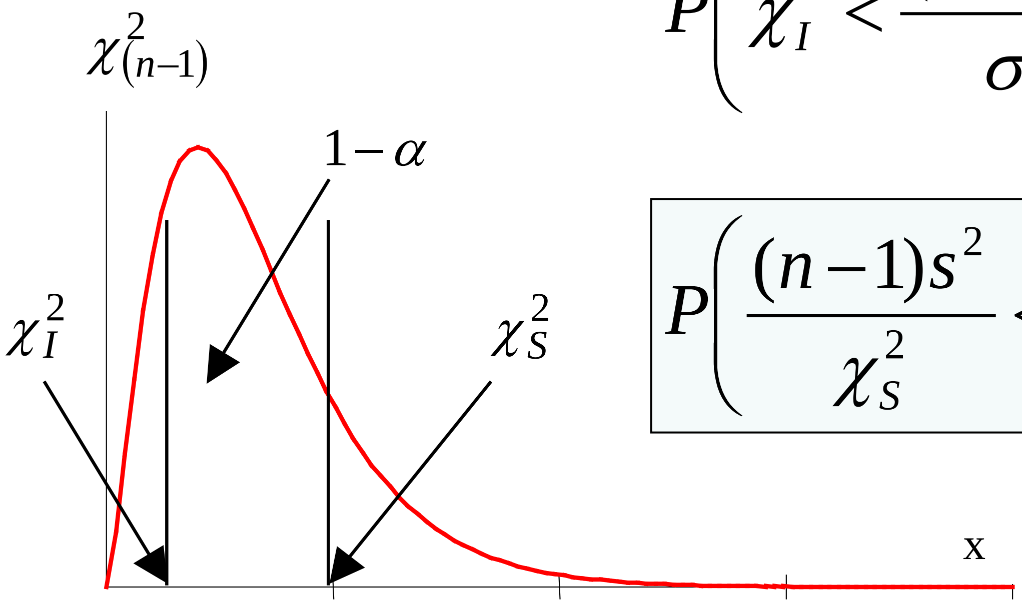
Definición de intervalo de confianza

INTERVALO DE CONFIANZA PARA LA VARIANZA POBLACIONAL

Población Normal

$$\frac{(n-1)\cdot S^2}{\sigma^2} = \chi^2_{(n-1)}$$

$$P\left(\chi_I^2 < \frac{(n-1)s^2}{\sigma^2} < \chi_S^2\right) = 1-\alpha$$



$$P\left(\frac{(n-1)s^2}{\chi_S^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_I^2}\right) = 1-\alpha$$

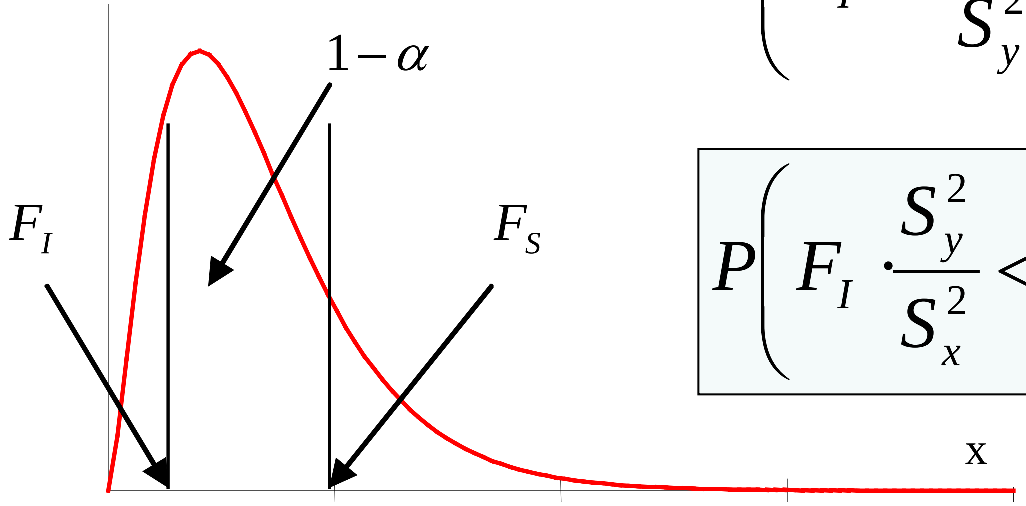
Definición de intervalo de confianza

INTERVALO DE CONFIANZA PARA EL COCIENTE DE VARIANZAS POBLACIONALES

Poblaciones Normales

$$\frac{S_x^2}{S_y^2}\cdot\frac{\sigma_y^2}{\sigma_x^2} \approx F_{(n_x-1;n_y-1)}$$

$F_{(n_x-1;n_y-1)}$

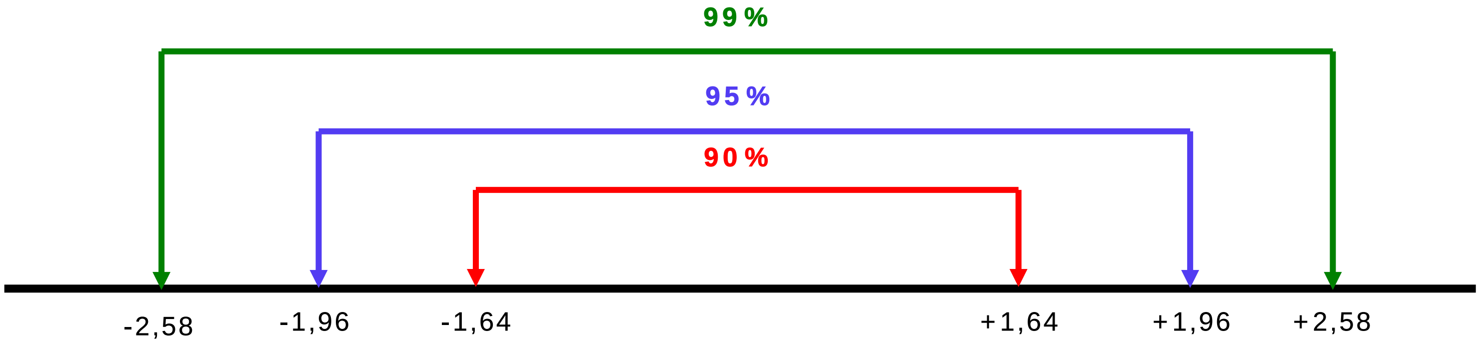


$$P\left(F_I < \frac{S_x^2}{S_y^2}\cdot\frac{\sigma_y^2}{\sigma_x^2} < F_S\right) = 1-\alpha$$

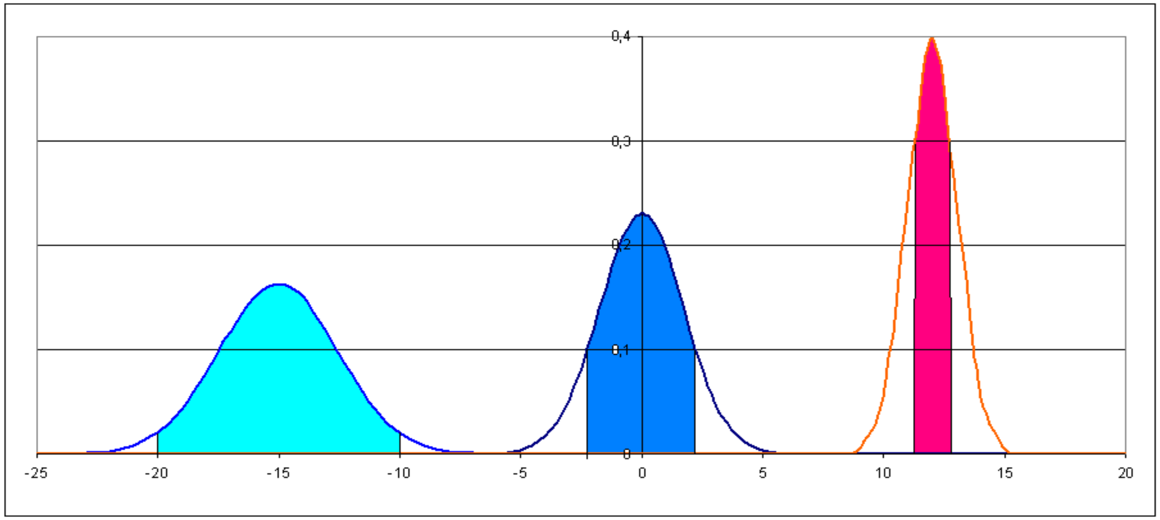
$$\boxed{P\left(F_I \cdot\frac{S_y^2}{S_x^2} < \frac{\sigma_y^2}{\sigma_x^2} < F_S \cdot\frac{S_y^2}{S_x^2}\right) = 1-\alpha}$$

Control del Intervalo de Confianza

- **Nivel de Confianza**



- **Tamaño Muestral**



Control del Intervalo de Confianza

1- I.C. Media Poblacional

(a)Población Normal-Varianza conocida

$$P\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Fija Amplitud

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

$$\boxed{n = \frac{4z_{\alpha/2}^2\sigma^2}{A^2}}$$

Fija Error Estimación

$$e = |\bar{x} - \mu| = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\boxed{n = \frac{z_{\alpha/2}^2\sigma^2}{e^2}}$$

Control del Intervalo de Confianza

(b)Población Normal-Varianza desconocida

$$P\left(\bar{x} - t_{(n-1)\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1)\alpha/2}\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

$$\boxed{n = \frac{4t^2_{(n-1),\alpha/2}S^2}{A^2}}$$

$$\boxed{n = \frac{t^2_{(n-1),\alpha/2}S^2}{e^2}}$$

1- <u>I.C. Proporción Poblacional</u>

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{pq}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{pq}{n}}\right) = 1-\alpha$$

Fija Amplitud

$$\boxed{n = \frac{4z^2_{\alpha/2}pq}{A^2}}$$

Fija Error Estimación

$$\boxed{n = \frac{p(1-p)z^2_{\alpha/2}}{e^2}}$$

"$pq = 0{,}5\ 0{,}5 = 0{,}25$"