UNIVERSIDAD NACIONAL DEL LITORAL

FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

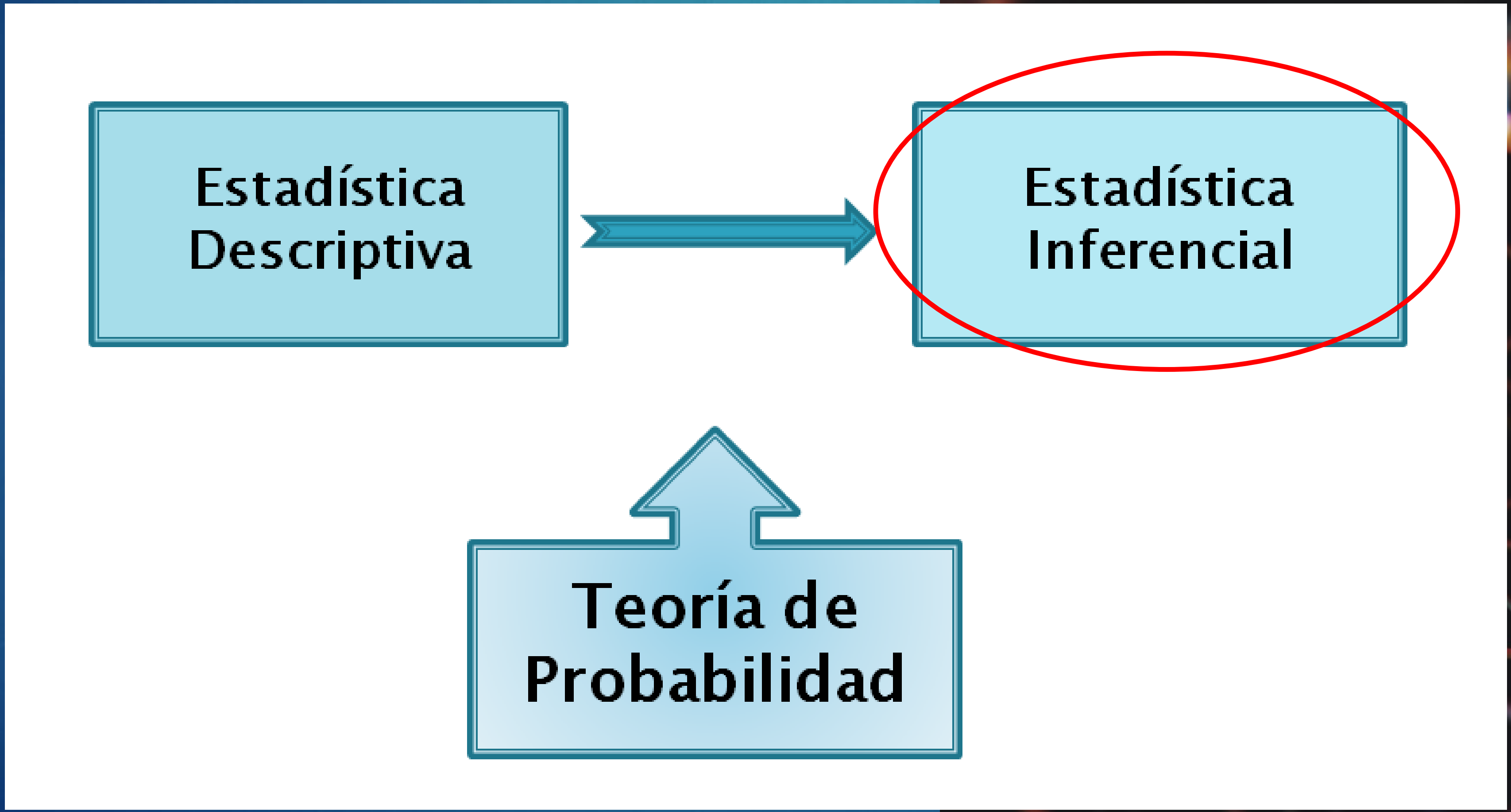
ESTADÍSTICA

INGENIERÍA INFORMÁTICA

TEORÍA

Mg. Susana Vanlesberg

INFERENCIA ESTADÍSTICA



ESTIMACIÓN

La **Teoría de la Estimación** es la parte de la Inferencia Estadística que sirve para conocer o acercarse al valor de los parámetros, características poblacionales, generalmente desconocidos.

- **No se trata de formas alternativas de estimación, sino complementarias. La estimación puntual representa el primer paso para obtener la estimación por intervalo.**

- **ESTIMADOR PUNTUAL**
- **Se trata de asignar al parámetro un único valor que será un valor aproximado y que depende de la muestra.**
- **La media muestral es un estimador puntual de la media poblacional μ: el valor que toma la media muestral con una muestra dada constituye una estimación para la media poblacional.**

- Un estimador es una función de valores observados (muestra) que no depende de ningún parámetro desconocido.

- Un estimador es un estadístico, y una estimación es uno de sus posibles valores.

- Un estimador puntual de un parámetro θ es simplemente un único valor del parámetro; rara vez coincide con el parámetro a estimar, generalmente es suficiente que el estimador esté "cerca" de la cantidad desconocida.
- *La variable aleatoria* $\hat{\theta}$ *se denomina estimador de* θ *y el valor que toma se llama estimación (puntual) de* θ *.*
- Un estimador es un estadístico y una estimación es cualquiera de sus posibles valores.

Propiedades deseables de los Estimadores

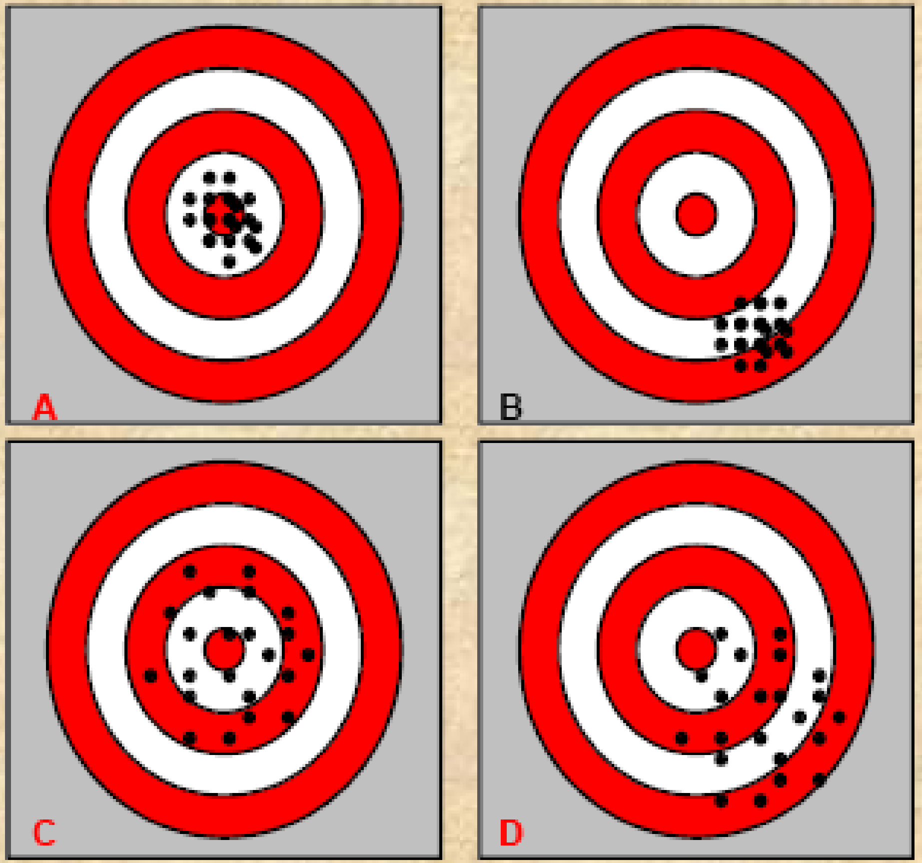
INSESGADO: el estimador $\hat{\theta}$ se dice que es un estimador ***insesgado*** de θ si su valor esperado es igual a θ:

$$E(\hat{\theta}) = \theta$$

EFICIENCIA: si su Varianza es tan pequeña como sea posible. A < varianza del estimador > eficiencia.

CONSISTENCIA: cuando el tamaño de la muestra crece arbitrariamente, el valor estimado se aproxima al parámetro desconocido.

RESUMEN de las propiedades de un buen estimador



A: Estimador centrado y eficiente;
B: Estimador sesgado y eficiente
C: Estimador centrado e ineficiente;
D: Estimador sesgado e ineficiente

- **Los estimadores puntuales sólo dan una idea de lo que puede valer el parámetro que queremos conocer (estimar), sin conocer cuánto se aproxima el estimador al parámetro; es decir, simplemente proporcionan un valor de los muchos posibles que pueden proponerse como valor del parámetro.**

Estimación por Intervalos de Confianza

- **Dada una variable aleatoria con distribución Normal, se trata de encontrar para el parámetro desconocido θ, un intervalo con un cierto *nivel de probabilidad (confianza)* (1- α). Si se cumple que:**

$$P\left(a \leq \theta \leq b\right) = 1 - \alpha$$

$$P\left(\left|\theta - \hat{\theta}\right| \leq k\,\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

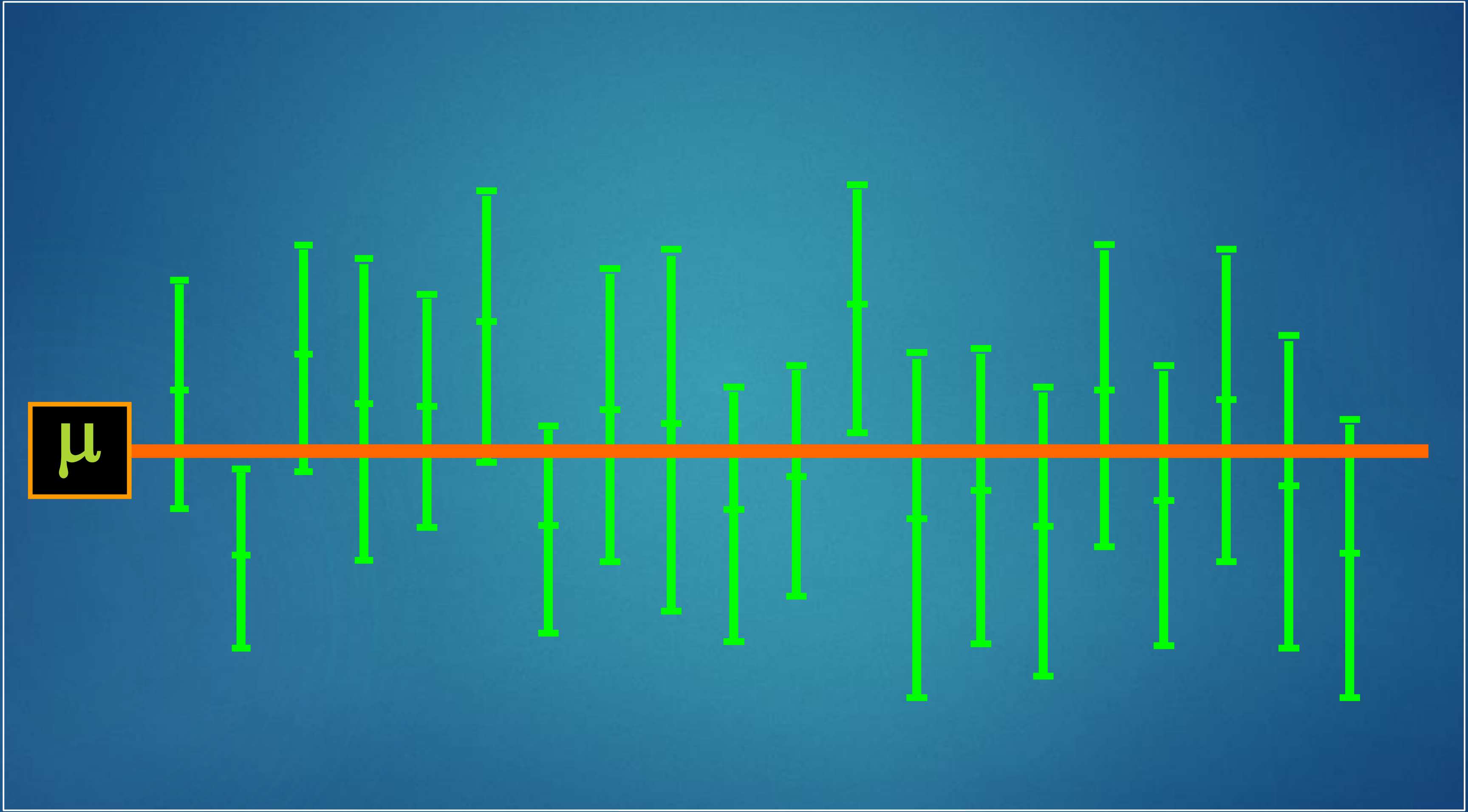
$$P\left(-k\,\sigma_{\hat{\theta}} \leq \theta - \hat{\theta} \leq k\,\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

$$P\left(\hat{\theta} - k\,\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + k\,\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

$$Error\ de\ estimación = k\ \sigma_{\hat{\theta}}$$

- **El significado de la palabra "confianza" es el siguiente:**

- **Si se construyeran muchos intervalos a partir de muestras aleatorias del mismo tamaño de la misma población, entonces 95% de estos intervalos contendrían la media de la población y 5% no la contendrían.**


μ

- ***Puede ser que el intervalo obtenido pertenezca al 95% de los intervalos que contienen la media de la población.***

- ***O bien formar parte del 5% de los intervalos que no la contienen.***

Intervalos para parámetros

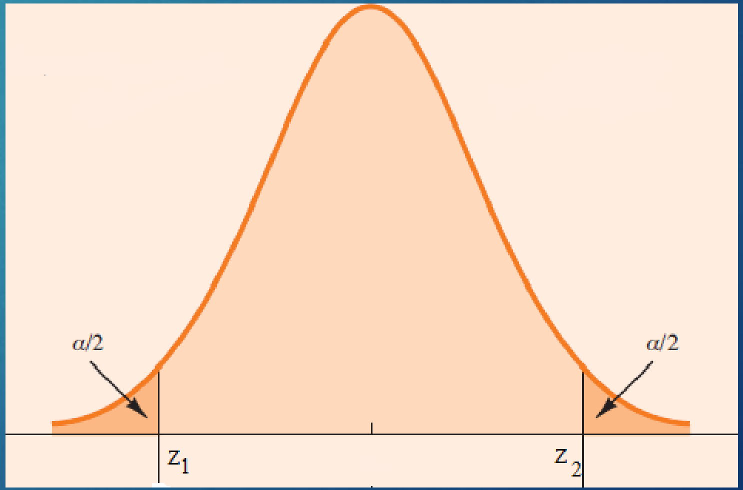
- **Para la media poblacional**
  - Población normal
    - σ conocido
    - σ desconocido: muestra grande –muestra chica

- **Varianza poblacional σ^2 conocida**
- **Estimador media muestral. Su distribución muestral es la siguiente:**

$$x \to N(\mu;\sigma)$$

$$\bar{x} \to N\left(\mu;\frac{\sigma}{\sqrt{n}}\right)$$

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



- **Varianza poblacional σ² desconocida, muestra grande**
- **Estimador media muestral. Su distribución muestral es:**

$$\bar{x} \sim N\left(\mu; \frac{S}{\sqrt{n}}\right)$$

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

- **Varianza poblacional σ^2 desconocida, muestra chica**
- **Estimador media muestral. Su distribución muestral es la siguiente:**

$$t_{n-1} = \frac{\bar{x} - \mu}{\frac{S'}{\sqrt{n}}}$$

$$P\left(\bar{x} - t_{n-1,1-\frac{\alpha}{2}} \frac{S'}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,1-\frac{\alpha}{2}} \frac{S'}{\sqrt{n}} \right) = 1 - \alpha$$

Para la varianza poblacional

- **Recordar la distribución muestral de la varianza muestral:**

$$\frac{(n-1)\, S'^2}{\sigma^2} \sim \chi^2_{n-1} \quad ó \quad bien \qquad \frac{n\, S^2}{\sigma^2} \sim \chi^2_{n-1}$$

- **Luego**

$$P\left(\chi_1^2 < \chi^2 < \chi_2^2\right) = 1-\alpha$$

$$P\left(\chi_1^2 < \frac{nS^2}{\sigma^2} < \chi_2^2\right) = 1-\alpha$$

$$P\left(\frac{nS^2}{\chi^2_{1-\frac{\alpha}{2};n-1}} \le \sigma^2 \le \frac{nS^2}{\chi^2_{\frac{\alpha}{2};n-1}}\right) = 1-\alpha$$

Para la proporción poblacional

$$p \sim N\left(\pi \; ; \; \sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

$$\left(p \pm |z| \sqrt{\frac{\pi\,(1-\pi)}{n}}\right)$$

INTERVALOS PARA DOS POBLACIONES

- Siempre recordar la distribución muestral de los estimadores y luego se generan los intervalos correspondientes

Para la diferencia de medias poblacionales

- **Desvíos poblacionales conocidos**

$$P\left((\bar{x} - \bar{y}) \pm \left| z_{1-\frac{\alpha}{2}} \right| \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right) = 1 - \alpha$$

- **Desvíos poblacionales desconocidos y muestras grandes:**

$$P\left(\left(\bar{x}-\bar{y}\right) \pm \left| z_{1-\frac{\alpha}{2}} \right| \sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}} \right) = 1-\alpha$$

- **Desvíos poblacionales desconocidos e iguales y muestras chicas**

$$P\left((\bar{x}-\bar{y}) \pm \left| t_{1-\frac{\alpha}{2};\nu} \right| S_p \sqrt{\frac{1}{n_x}+\frac{1}{n_y}} \right) = 1-\alpha$$

$$S_p = \sqrt{\frac{{S_x}^2 n_x + {S_y}^2 n_y}{n_x + n_y - 2}}$$

$$\nu = n_x - n_y - 2$$

- Desvíos poblacionales desconocidos y distintos - muestras chicas

$$P\left(\left(\bar{x}-\bar{y}\right)\pm\left|t_{1-\frac{\alpha}{2};\nu}\right|\sqrt{\frac{S_x'^2}{n_x}+\frac{S_y'^2}{n_y}}\right)=1-\alpha$$

$$\nu=\frac{\left(\frac{S'^2_x}{n_x}+\frac{S'^2_y}{n_y}\right)^2}{\frac{\left(\frac{S'^2_x}{n_x}\right)^2}{n_x+1}+\frac{\left(\frac{S'^2_y}{n_y}\right)^2}{n_y+1}}-2$$

Para la razón de varianzas poblacionales

$$P\left(\frac{S_1^2}{S_2^2}F_{1-\frac{\alpha}{2};n_2-1;n_1-1}\leq\frac{\sigma_1^2}{\sigma_2^2}\leq\frac{S_1^2}{S_2^2}F_{\frac{\alpha}{2};n_2-1;n_1-1}\right)=1-\alpha$$

$$F_{1-\frac{\alpha}{2};n_2-1;n_1-1}=\frac{1}{F_{\frac{\alpha}{2};n_2-1;n_1-1}}$$

$$P\left(\frac{S_1}{S_2}\sqrt{\frac{1}{F_{\frac{\alpha}{2};n_2-1;n_1-1}}}\leq\frac{\sigma_1}{\sigma_2}\leq\frac{S_1}{S_2}\sqrt{F_{\frac{\alpha}{2};n_2-1;n_1-1}}\right)=1-\alpha$$

PARA LA DIFERENCIA DE PROPORCIONES POBLACIONALES

$$P\left(p_1 - p_2 \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}\right) = 1-\alpha$$

RESUMEN

Parámetro	Supuesto	Estimador puntual	Distribución del estimador	Intervalo de confianza
μ	σ *conocido*	$\bar{x}$	$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	$\left(\bar{x} \pm \lvert Z_{\left(1-\frac{\alpha}{2}\right)} \rvert \frac{S}{\sqrt{n}}\right)$
μ	σ *desconocido*, $n > 30$	$\bar{x}$	$\bar{x} \sim N\left(\mu, \frac{S}{\sqrt{n}}\right)$	$\left(\bar{x} \pm \lvert Z_{\left(1-\frac{\alpha}{2}\right)} \rvert \frac{S}{\sqrt{n}}\right)$
μ	σ *desconocido*, $n < 30$	$\bar{x}$	$\frac{x-\mu}{\frac{S'}{\sqrt{n}}} = t_{(n-1)}$ ó $\frac{\bar{x}-\mu}{\frac{S}{\sqrt{n-1}}} = t_{n-1}$	$\left(\bar{x} \pm \lvert t_{\left(1-\frac{\alpha}{2};n-1\right)} \rvert \frac{S'}{\sqrt{n}}\right)$
σ^2	Población normal	S^2	$\left(\frac{n.S^2}{\sigma^2} \sim \chi^2_{(n-1)}\ \text{ó}\ \frac{(n-1).S'^2}{\sigma^2} \sim \chi^2_{(n-1)}\right)$	$\left(\frac{n.S}{\chi^2_{\left(1-\frac{\alpha}{2};n-1\right)}} \leq \sigma^2 \leq \frac{n.S}{\chi^2_{\left(\frac{\alpha}{2};n-1\right)}}\right)$
π	Población normal	p	$Z = \frac{p-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$	$\left(p \pm \lvert Z_{\left(1-\frac{\alpha}{2}\right)} \rvert \sqrt{\frac{\pi(1-\pi)}{n}}\right)$
$\mu_x - \mu_y$	σ_x, σ_y *conocidas*	$\bar{x} \cdot \bar{y}$	$\left(\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y, \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)\right)$	$\left((\bar{x} - \bar{y}) \pm \lvert Z_{\left(1-\frac{\alpha}{2}\right)} \rvert \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}\right)$

$\mu_x - \mu_y$	σ_x, σ_y *desconocidas* *pero iguales* $n < 30$	$\bar{x} - \bar{y}$	$\dfrac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{Sw \cdot \sqrt{\dfrac{1}{n_x}+\dfrac{1}{n_y}}} = t_{(n_x+n_y-2)}$ $con \quad Sw = \sqrt{\dfrac{(n_x-1)\cdot S'^2_x + (n_y-1)\cdot S'^2_y}{n_x+n_y-2}}$ $ó \quad con \quad Sw = \sqrt{\dfrac{n_x \cdot S^2_x + n_y \cdot S^2_y}{n_x+n_y-2}}$	$\left((\bar{x}-\bar{y}) \pm \lvert t_{\left(1-\frac{\alpha}{2}\right)} \rvert \cdot S_w \cdot \sqrt{\dfrac{1}{n_x}+\dfrac{1}{n_y}} \right)$
$\mu_x - \mu_y$	σ_x, σ_y *desconocidas* *y distintas* $n < 30$	$\bar{x} - \bar{y}$	$\dfrac{\bar{x}-\bar{y}-(\mu_x-\mu_y)}{\sqrt{\dfrac{S'^2_x}{n_x}+\dfrac{S'^2_y}{n_y}}} = t_v$ $con \quad v = \dfrac{\left(\dfrac{S'^2_x}{n_x}+\dfrac{S'^2_y}{n_y}\right)^2}{\dfrac{\left(\dfrac{S'^2_x}{n_x}\right)^2}{n_x+1}+\dfrac{\left(\dfrac{S'^2_y}{n_y}\right)^2}{n_y+1}} - 2$	$\left((\bar{x}-\bar{y}) \pm \lvert t_{\left(1-\frac{\alpha}{2}\right),v} \rvert \sqrt{\dfrac{S^2_x}{n_x}+\dfrac{S^2_y}{n_y}} \right)$

$\pi_1 - \pi_2$	*Poblaciones normales*	$p_1 - p_2$	$p_1 - p_2 \sim N\left(\pi_1 - \pi_1; \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}\right)$
$\frac{\sigma_x^2}{\sigma_y^2}$	*Poblaciones normales*	$\frac{s_x^2}{s_y^2}$	$F = \frac{\frac{S_x^2}{\sigma_x^2}}{\frac{S_y^2}{\sigma_y^2}}$

APLICAMOS

- **La varianza en un proceso de producción es un indicador importante de la calidad del proceso. Las varianzas grandes representan una oportunidad para mejorarlo, buscando formas de reducir la varianza del proceso.**
- **Se tiene información de un proceso de llenado en una industria, y basados en lo dicho se quiere investigar si el proceso está funcionando correctamente.**
- **Podría determinar si existe una diferencia significativa entre las varianzas de los pesos de lo que se envasa y que se realiza con dos máquinas diferentes.**
- **¿Alguna de las dos máquinas representa una oportunidad para mejorar la calidad del proceso?**

- **Máquina 1:** 2.95 3.45 3.50 3.75 3.48 3.26 3.33 3.20

3.16 3.20 3.22 3.38 3.90 3.36 3.25 3.28

3.20 3.22 2.98 3.45 3.70 3.34 3.18 3.35

3.12

- **Máquina 2:** 3.22 3.30 3.34 3.28 3.29 3.25 3.30 3.27

3.38 3.34 3.35 3.19 3.35 3.05 3.36 3.28

3.30 3.28 3.30 3.20 3.16 3.33

- **Los jóvenes usan Internet intensamente, se dispone de información: 87% de los jóvenes entre 12 y 17 años son usuarios de la red. En una muestra de usuarios de Internet de esta edad, 9% votó por MySpace como el sitio más popular de la Web.**

- **Si en este estudio participaron 1400 jóvenes, ¿cuál es la proporción real de quienes consideran que este sitio es el más popular?**