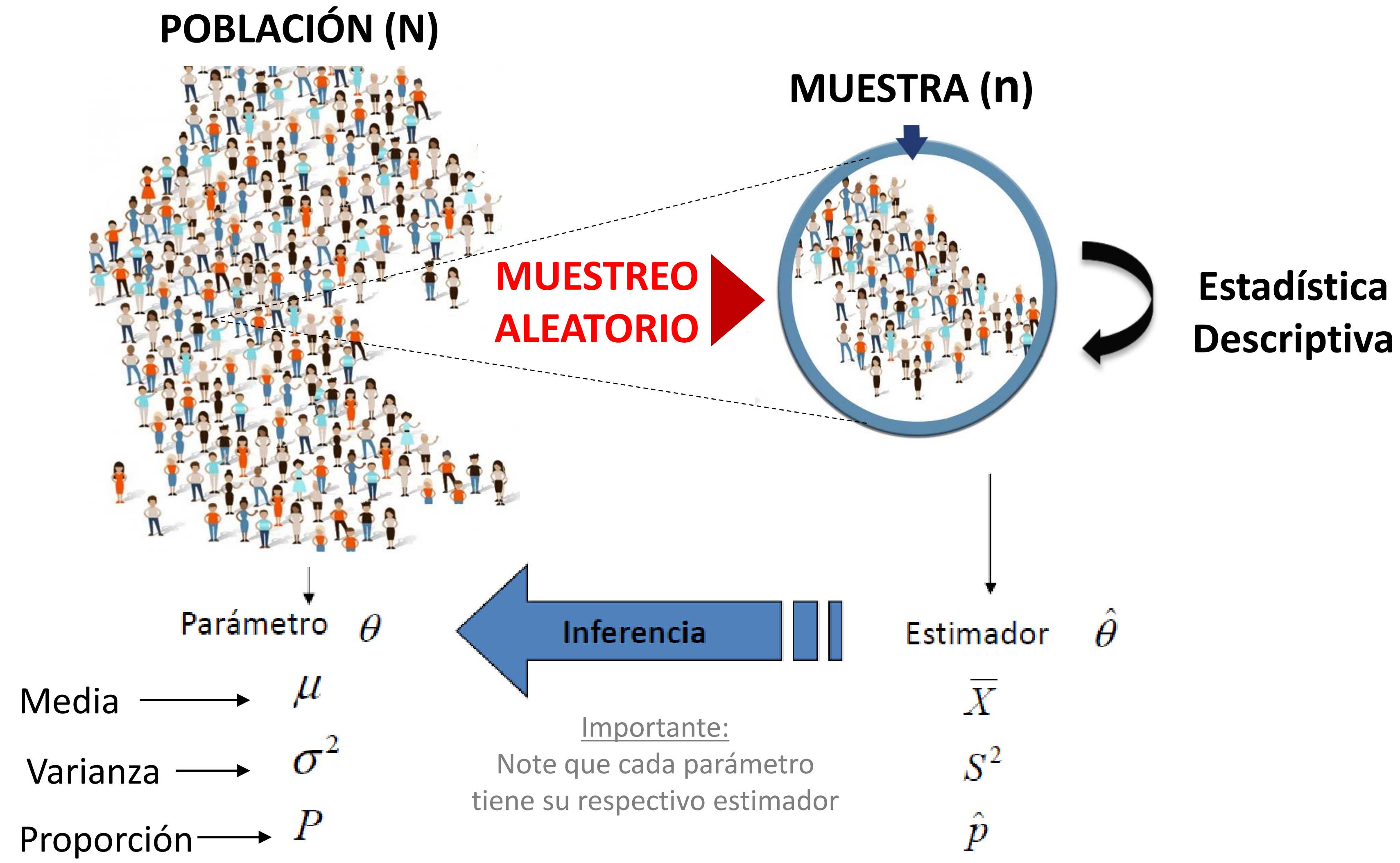
# La Esencia de la Estimación

## La Estadística como herramienta para inferir

*“Por que creer en los resultados obtenidos en una muestra si seguramente al repetir el ejercicio estos presentarán variaciones”*

# Inferencia Estadística

**Relación entre población y muestra, parámetros y estadísticos.**



# Estimación

Procedimiento estadístico en el cual se logra una valoración de un parámetro asociado a una población no medida (en su totalidad), mediante el estudio de una muestra aleatoria

*"Si el procedimiento de muestreo es adecuado y la característica no es muy variable, se esperaría que el estimador de la mejor idea del parámetro"*

## Estimación Puntual

Consiste en la estimación del valor del parámetro mediante un solo valor, obtenido de una formula determinada.

| | Media | Proporción | Varianza | Desviación estándar |
|---|---|---|---|---|
| Estimador | $\bar{x}$ | p | $S^2$ | S |
| Parámetro | μ | π | $\sigma^2$ | σ |

# Parámetros y sus Estimadores

| $\theta$ | $\hat{\theta}$ |
|---|---|
| $\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$ | $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ |
| $\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ | $S^2 = \dfrac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}$ |
| $P = \dfrac{X}{N}$ | $\hat{p} = \dfrac{x}{n}$ |
| $\mu_1 - \mu_2$ | $\overline{X}_1 - \overline{X}_2$ |
| $\pi_1 - \pi_2$ | $p_1 - p_2$ |
| $\sigma_1^2 / \sigma_2^2$ | $S_1^2 / S_2^2$ |

El estadístico $\hat{\theta}$ que se utiliza para obtener una estimación puntual recibe el nombre de ***estimador.***

*Generalmente muestras diferentes conducen a estimaciones diferentes*

# Parámetros y Estimadores

Sea $X$ , una variable aleatoria que representa la **calificación obtenida en una prueba de conocimientos** (escala vigesimal), de los alumnos de la Facultad de Ingeniería, si la población consta de 300 alumnos, entonces:

$$x_1, x_2, x_3, \cdots, x_{300}$$

Asumiéremos que $X$ tiene distribución de probabilidad con media $\boldsymbol{\mu_x}$ y varianza $\boldsymbol{\sigma_x^2}$

Medidas que caracterizan la población.

Parámetros

$N=300$
$\boldsymbol{\mu_x = 14.8}$
$\boldsymbol{\sigma_x^2 = 3.1}$

Al seleccionar una muestra aleatoria de tamaño seis (n=6):

$x_1 = 13$ $\quad x_4 = 14$
$x_2 = 10$ $\quad x_5 = 11$
$x_3 = 13$ $\quad x_6 = 10$

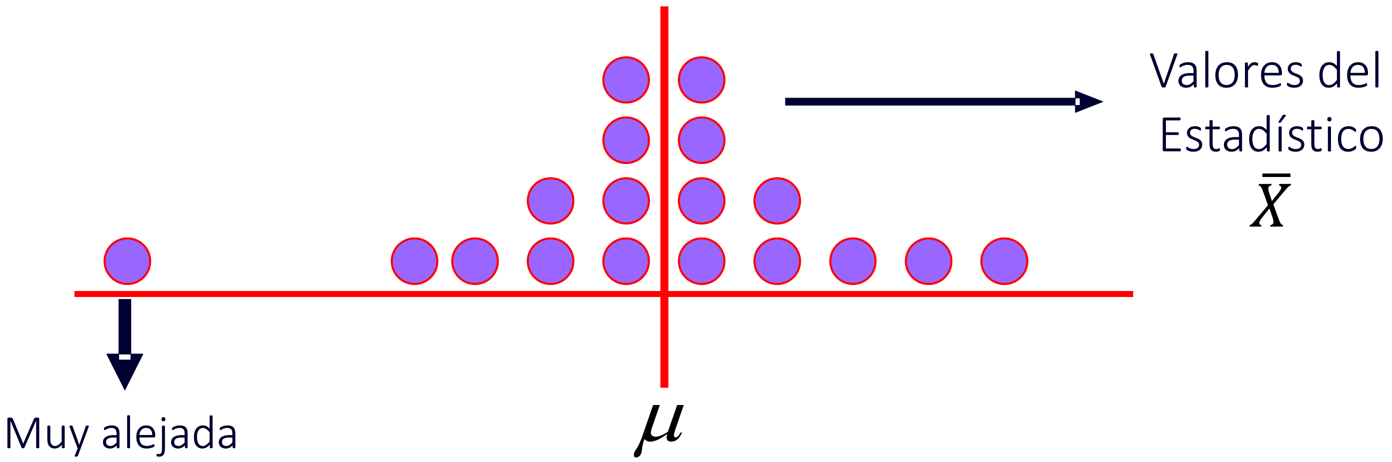
Estimadores

$n=6$
$\boldsymbol{\bar{x} = 11.8}$
$\boldsymbol{s^2 = 2.97}$

¿Que podemos hacer para tener una idea de estos valores?

# Precisión de un Estimador

No se espera que un estimador obtenga ***<u>sin error</u>*** el parámetro poblacional, tampoco que estime el valor de este con exactitud, pero si que *<u>no se aleje mucho del parámetro poblacional.</u>*



# Error de Muestreo

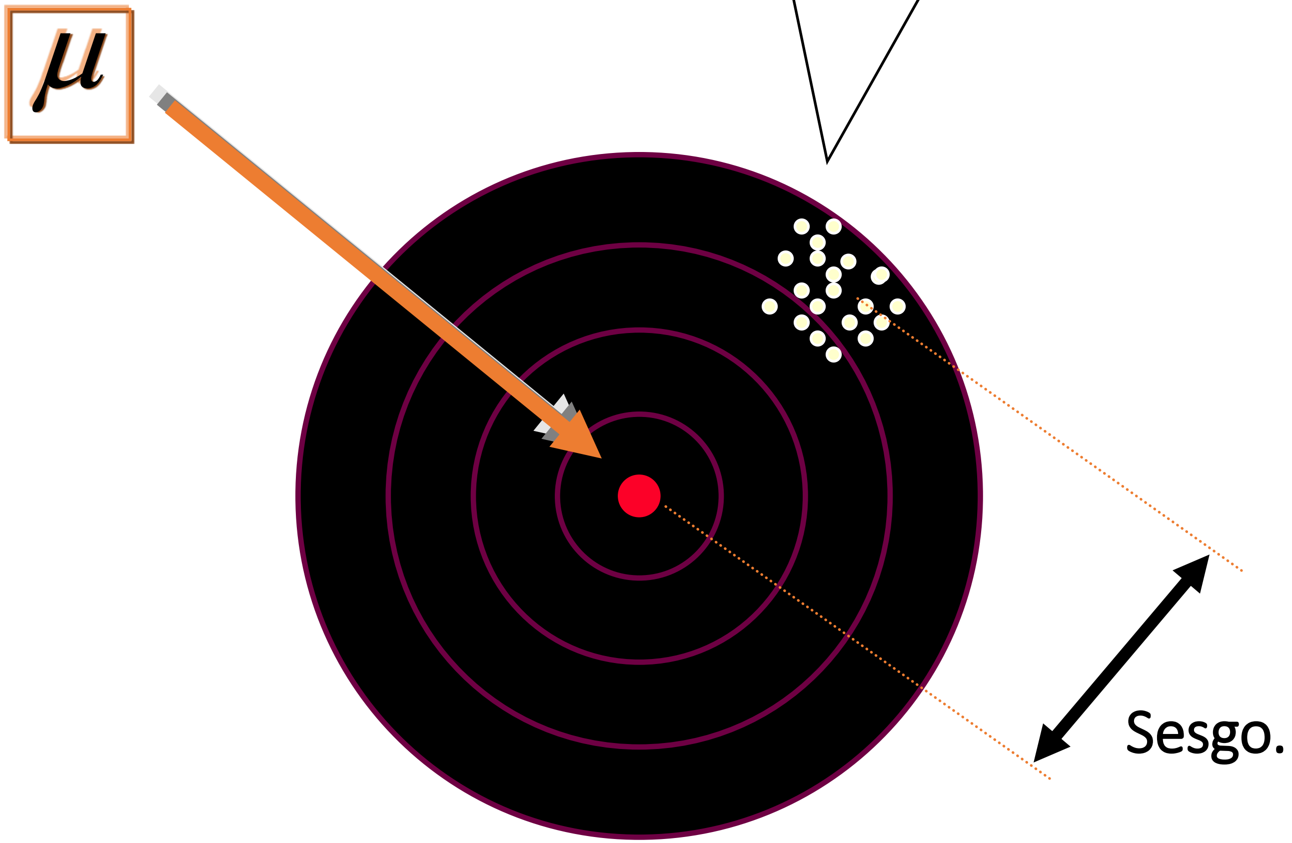
Se define como la diferencia entre el valor del estimador y el valor real del parámetro

$$\left|\bar{X}-\mu\right| \qquad \left|p-\pi\right| \qquad \left|S^2-\sigma^2\right|$$

*“El error de muestreo es inevitable, los métodos estadísticos buscan cuantificarlo y minimizarlo”*

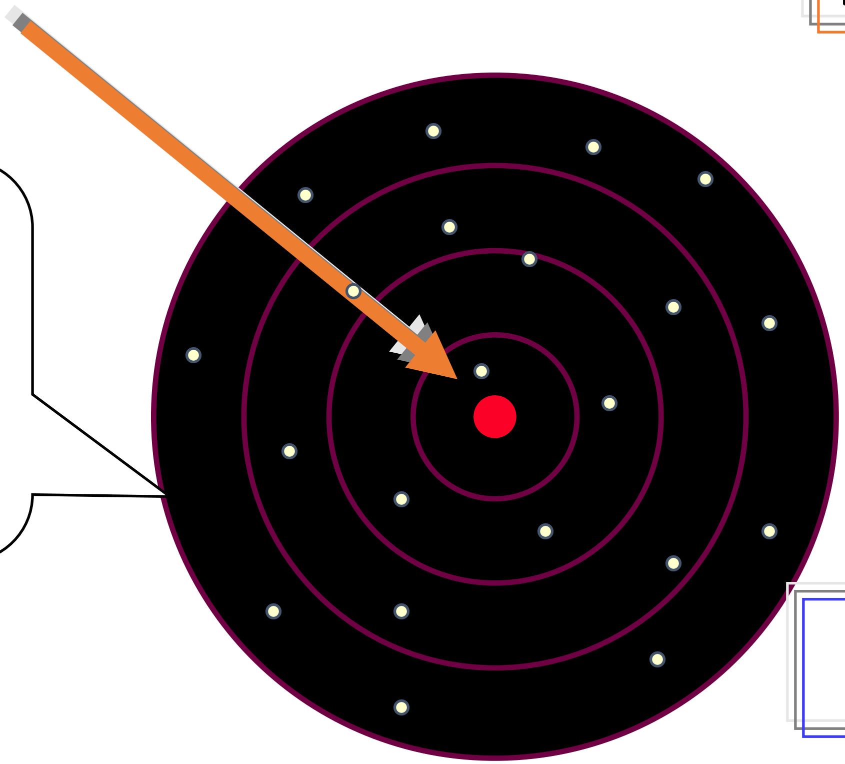
# Hechos claves...

En la práctica
solo hacemos un disparo.
( Corresponde a una muestra)



# Hechos claves…

$\mu$

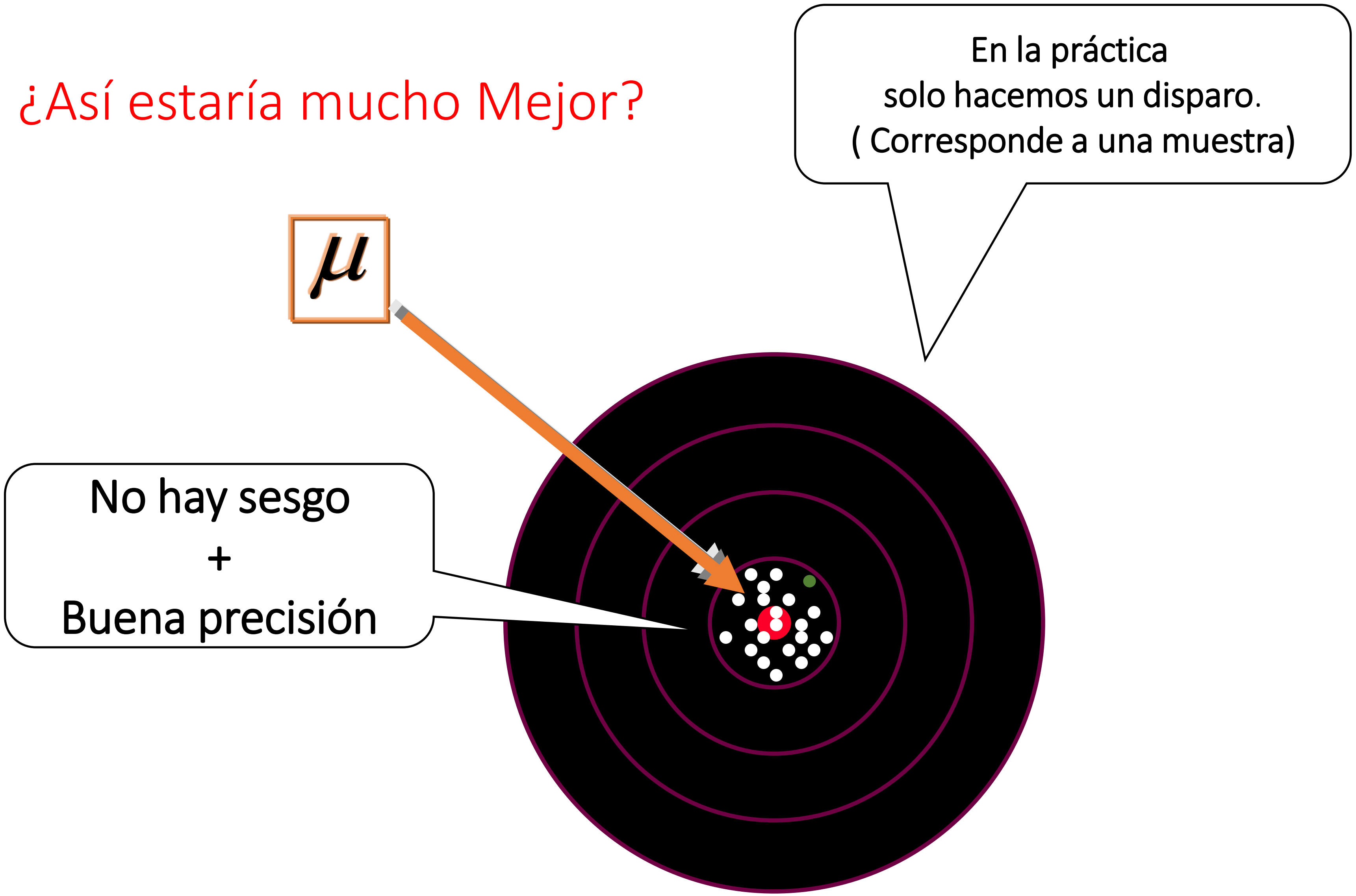


$\bar{x}$
La media Muestral, No tiene sesgo

No hay Sesgo.

Nos gustaría menos variabilidad

# ¿Así estaría mucho Mejor?



# Algunos Parámetros y sus Estimadores

| $\theta$ | $\hat{\theta}$ |
|---|---|
| $\mu = \dfrac{\sum_{i=1}^{N} x_i}{N}$ | $\bar{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ |
| $\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$ | $S^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n-1}$ |
| $P = \dfrac{X}{N}$ | $\hat{p} = \dfrac{x}{n}$ |
| $\mu_1 - \mu_2$ | $\bar{X}_1 - \bar{X}_2$ |
| $P_1 - P_2$ | $\hat{p}_1 - \hat{p}_2$ |
| $\sigma_1^2 / \sigma_2^2$ | $S_1^2 / S_2^2$ |

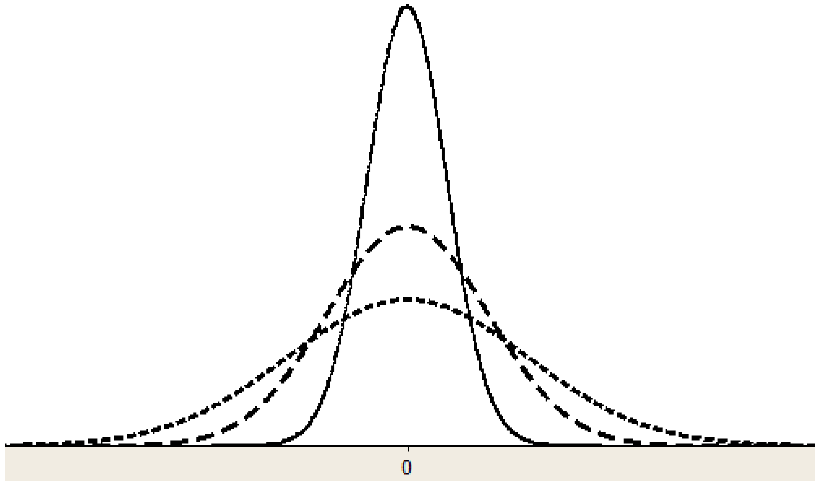
El estadístico $\hat{\theta}$ que se utiliza para obtener una estimación puntual recibe el nombre de *estimador.*

*Generalmente muestras diferentes conducen a estimaciones diferentes*

# Distribución de un Estimador

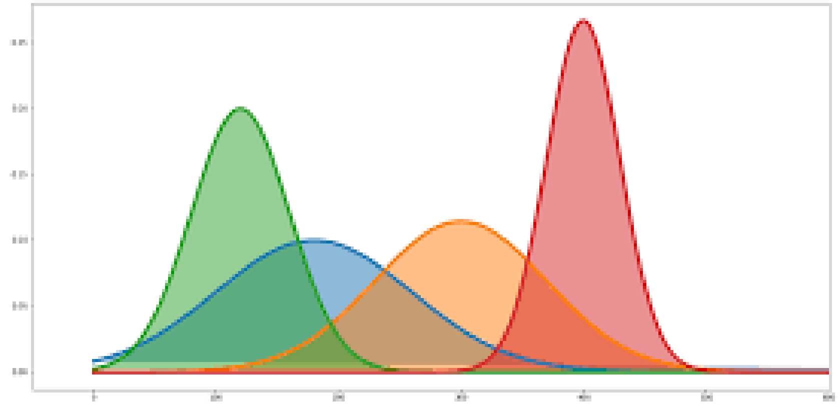
## ¿Un estimador es una variable aleatoria?

El valor que adquiere un estimador es el resultado de un experimento aleatorio, por tanto un estimador es una variable aleatoria y debe tener una distribución de probabilidades asociada.



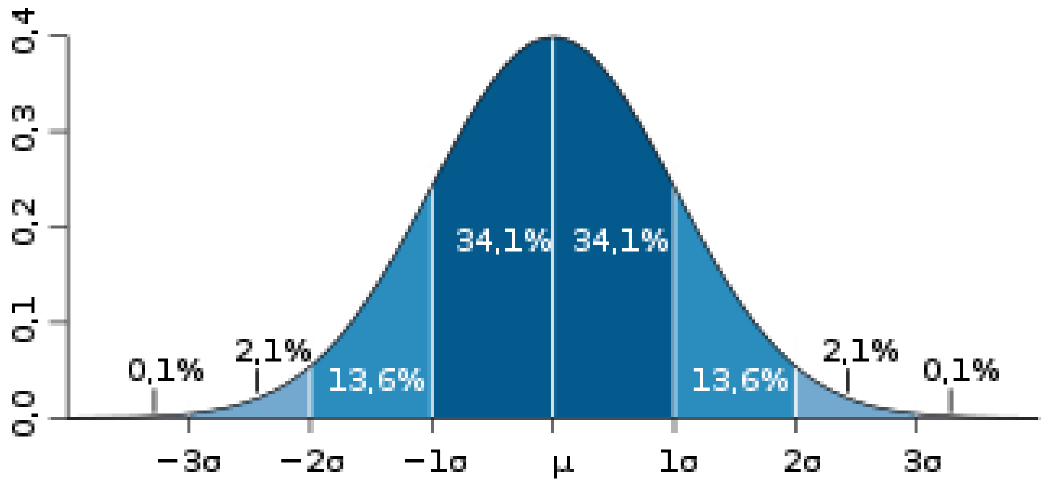

*"La distribución de probabilidad de un estimador recibe el nombre de distribución muestral"*

# Distribuciones muestrales



# La Distribución Normal

Sin duda la distribución continua de probabilidad más importante en el campo de la estadística, por la diversas de aplicaciones que se pueden modelar a través de ella.



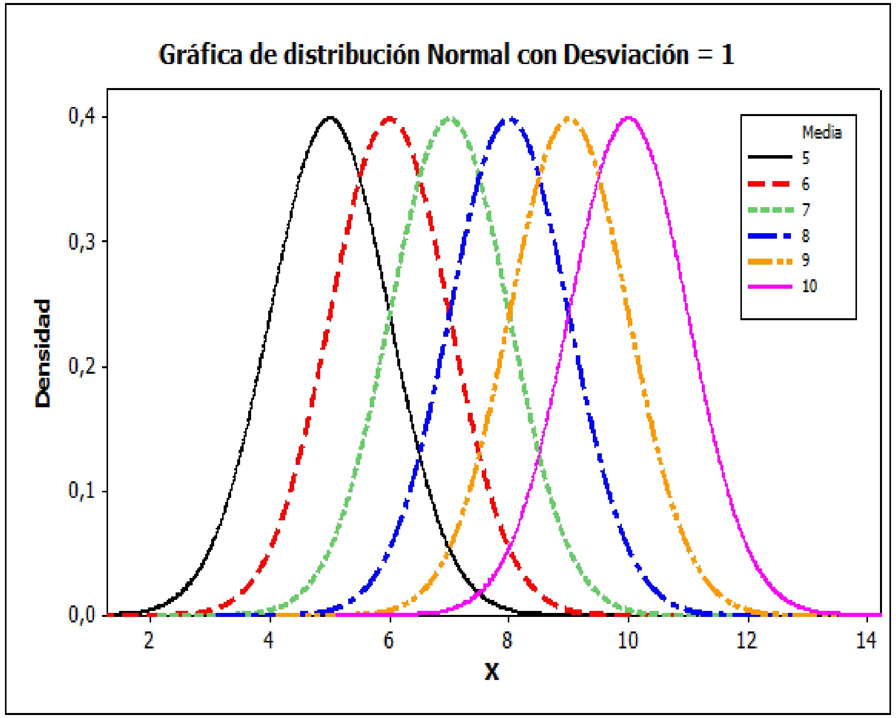

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

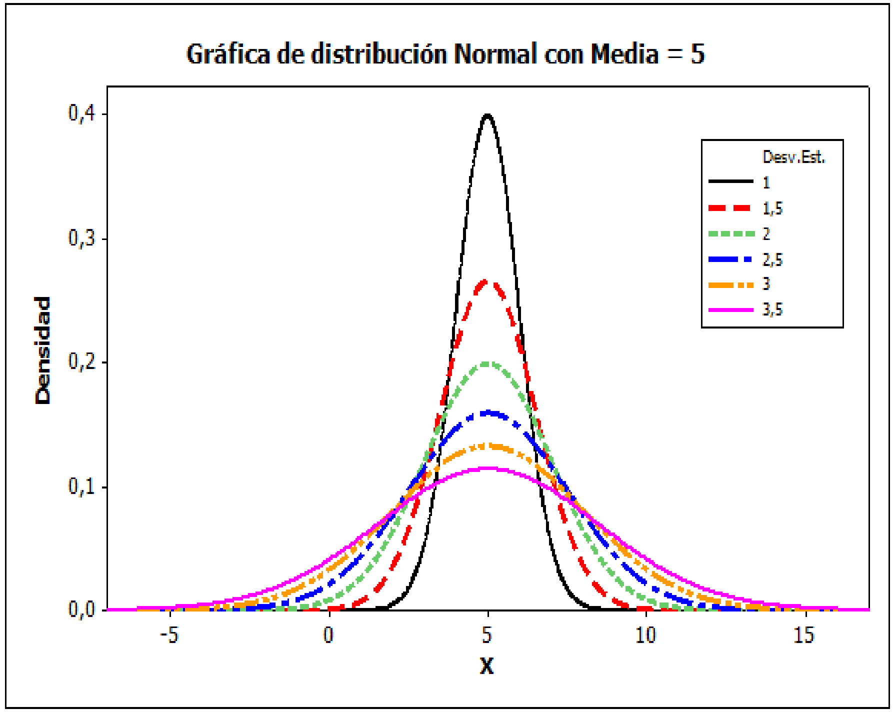
*La distribución de probabilidad normal se caracteriza por:*

- Tiene forma de campana.
- La media aritmética, la mediana y la moda son iguales.
- El área total bajo la curva es de 1.00.
- Es ***simétrica*** respecto de la media.
- La dispersión de la distribución se determina por medio de la desviación estándar **σ**.

# La Distribución Normal

Que papel juegan el promedio (μ) y la desviación (**σ**) en la distribución.




$$\text{Si } X \sim N(\mu;\sigma) \text{ entonces: } E(X)=\mu, V(X)=\sigma^2$$

# Distribución de la Media Muestral ($\bar{x}$)

Si $\bar{\boldsymbol{x}}$ es la media de una muestra aleatoria de tamaño *n* que se toma de una población normal con media $\boldsymbol{\mu_x}$ y varianza $\boldsymbol{\sigma_x^2}$ conocida entonces:

- Propiedades de la Media Muestral

Media de $\bar{\boldsymbol{x}}$

$$\mu_{\bar{x}} = \mu_x$$

Error estándar de $\bar{\boldsymbol{x}}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Para *n* suficientemente grande, la variable aleatoria

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

Tiene una distribución Normal Estándar

$$Z \sim N\left(\mu = 0 \ , \ \sigma = 1\right)$$

## ¿Si la varianza $\boldsymbol{\sigma_x^2}$ no es conocida, que distribución tendrá $\bar{\mathbf{x}}$?

Sea $\{X_1, X_2, .. X_n\}$ una muestra aleatoria extraída de una población normal con media poblacional y varianza desconocida, entonces la variable aleatoria:

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

Sigue una distribución *t* - Student con $n - 1$ grados de libertad.

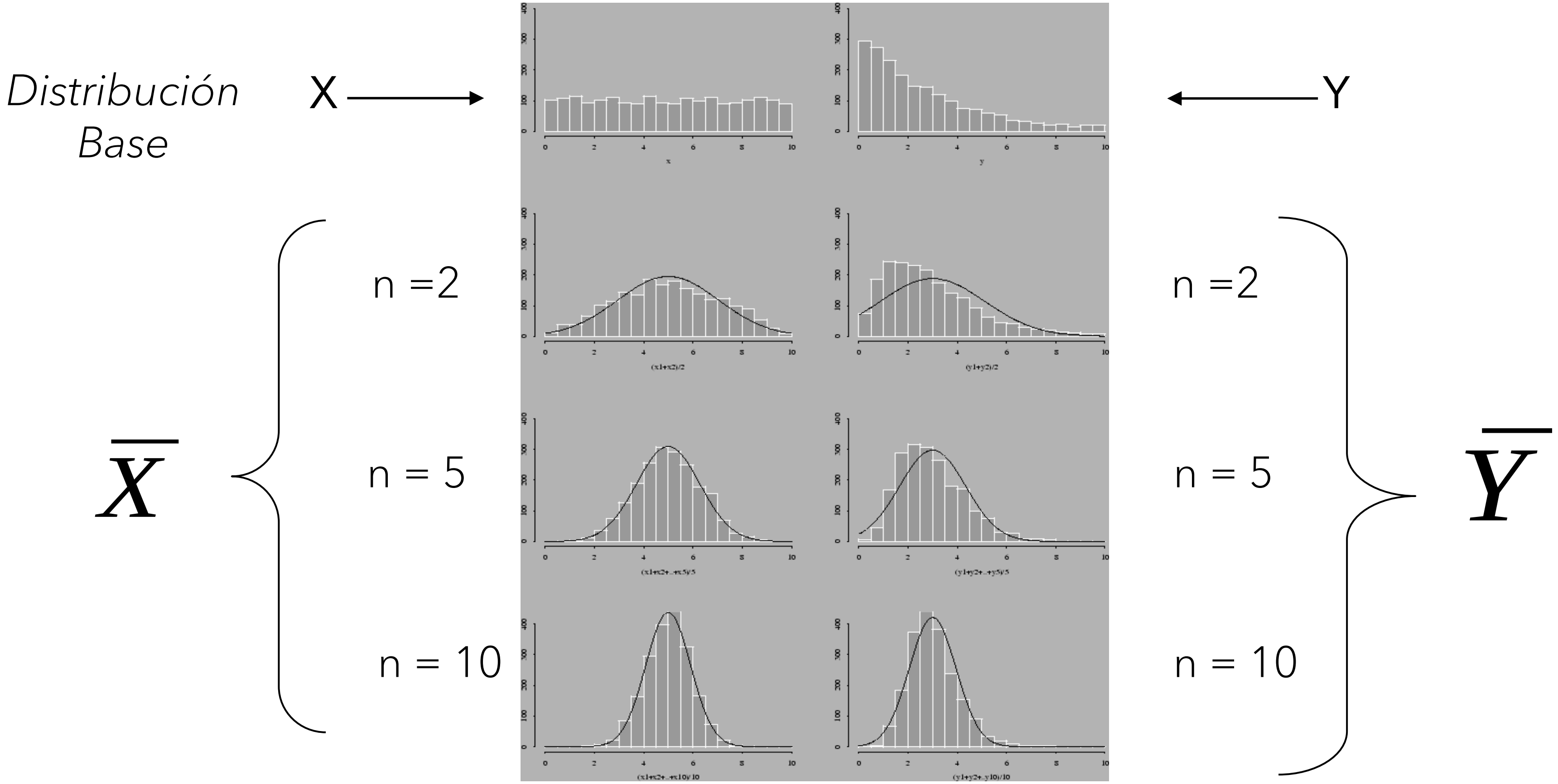
## ¿Si la población no es Normal, que distribución tendrá $\bar{x}$?

**Teorema del Límite Central**

La distribución Normal Estándar *N*(0,1)
Se considera una **buena aproximación** cuando $n \geq 30$

# Teorema del Límite Central

Si $\bar{x}$ es la media de una muestra aleatoria de tamaño n obtenida desde una población con cualquier distribución, con media $\mu_x$ y varianza $\sigma_x^2$ , entonces cuando **n es grande** , $\bar{x}$ sigue una distribución normal $\mu_{\bar{x}} = \mu_x$ y $\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$




*"La aproximación es buena cuando n≥30"*

# Distribución muestral de la media ($\bar{x}$)

Si $\boldsymbol{\sigma}$ es conocido
ó $\boldsymbol{n \geq 30}$

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

Si $\boldsymbol{\sigma}$ es desconocido
y $\boldsymbol{n < 30}$

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

# Distribución de la Proporción Muestral (p)

Consideremos una población en la que existe una proporción **π** de elementos que tienen el atributo *A* (o pertenecen a la categoría *A* ).

Si se toma una muestra aleatoria de **n** elementos, de esa población y se calcula el número $\boldsymbol{n_A}$ de elementos con el atributo **A** , entonces:

$$p = \frac{n_A}{n}$$

## PROPIEDADES DE LA PROPORCIÓN MUESTRAL

1. $\mu_\pi = \pi$

2. $\sigma_\pi = \sqrt{\frac{\pi(1-\pi)}{n}}$

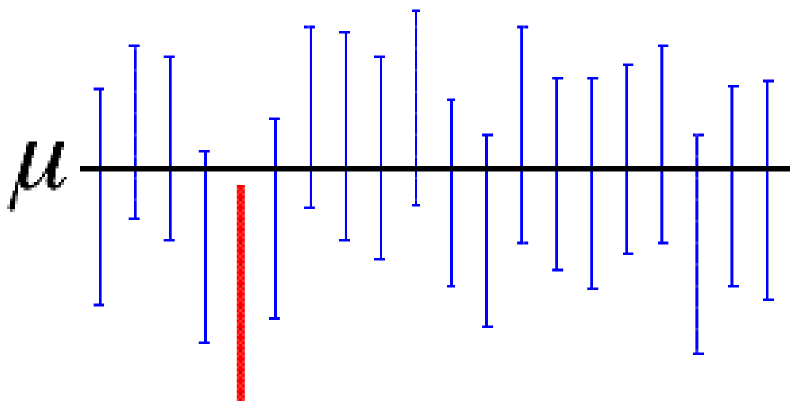
Error estándar de ***p***

3. Para *n* suficientemente grande, la variable aleatoria

$$Z = \frac{P - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$$

Se distribuye aproximadamente como una $N(0,1)$. Se considera una buena aproximación cuando $n*p \geq 5$ y $n*(1-p) \geq 5$ (Teorema del Límite Central).

# Estimación por Intervalos de confianza

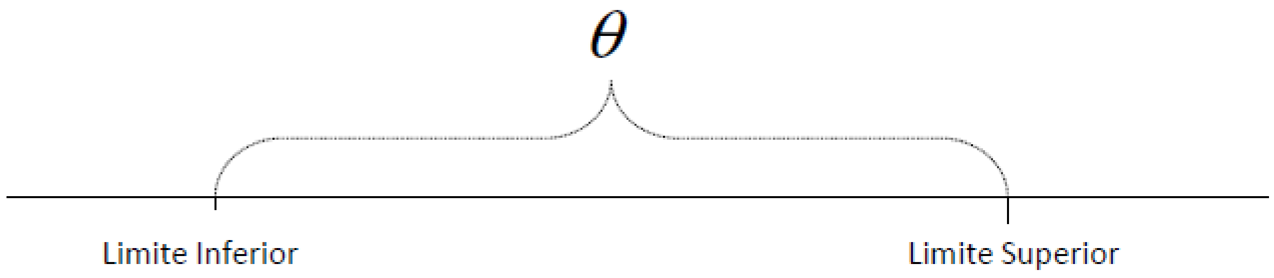


# Estimación por  Intervalos de confianza

Dado que los estimadores puntuales pocas veces serán iguales a los parámetros que se desean estimar, es posible darse mayor libertad utilizando estimadores por intervalos que reciben el nombre de *intervalos de confianza*.

**Definición:**

Un intervalo de confianza es un intervalo **estimado** en el cual se espera se encuentre el valor del parámetro.

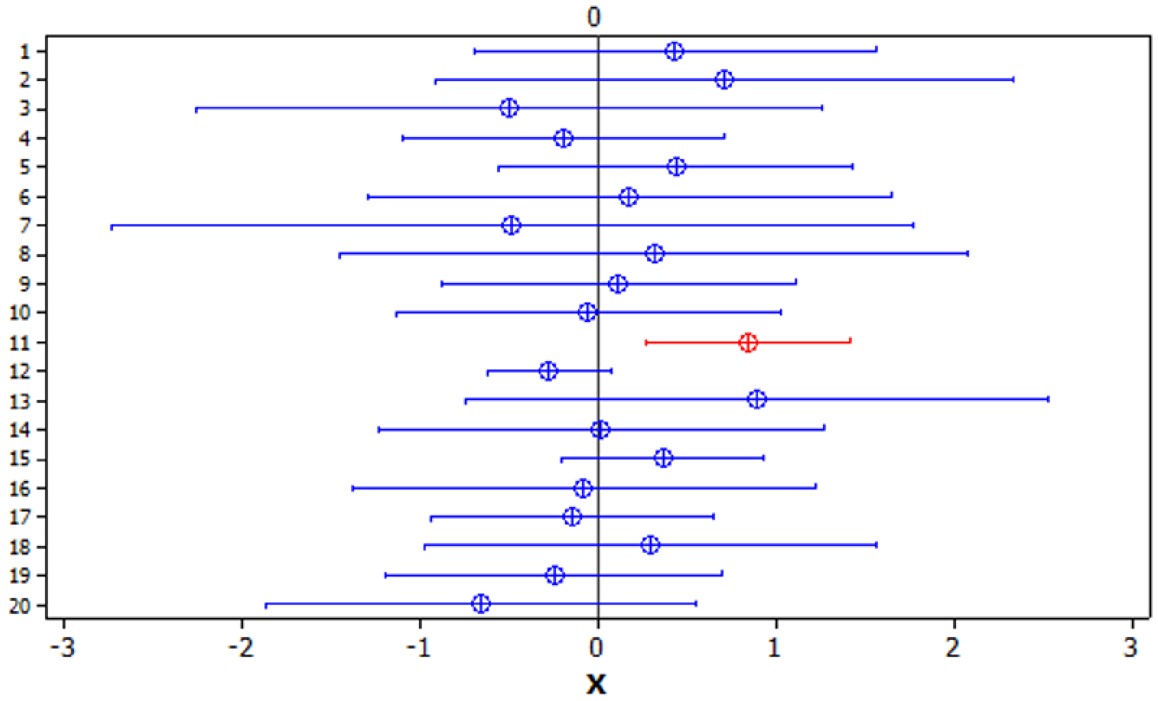


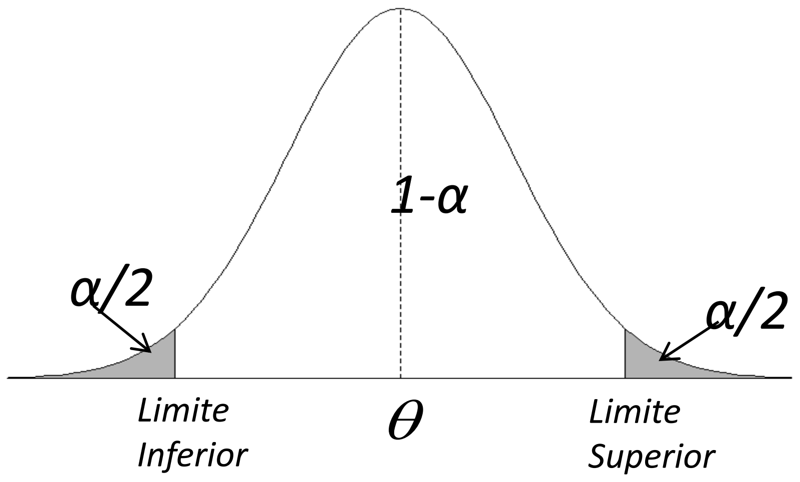
# Estimación por Intervalos de confianza

La idea es la siguiente:

Si se toman diferentes muestras de la población de referencia, y se efectúa el calculo de los intervalos de confianza, se espera que aproximadamente el (1-$\alpha$)% de estos intervalos contengan el verdadero valor del parámetro.

La cantidad **1-$\alpha$** recibe el nombre de Nivel de Confianza donde $\alpha$ es nivel de error (nivel de significancia).



# Intervalos de confianza

Pasos en la construcción de un intervalo de confianza:

1. Obtener una muestra aleatoria de la población de interés.
2. Calcular el estimador correspondiente al parámetro de interés.
3. Fijar el nivel de confianza deseado.
4. Construir un intervalo mediante el siguiente planteamiento:

$$IC\langle\theta\rangle_{(1-\alpha)\%} = [LI\ ;\ LS]$$

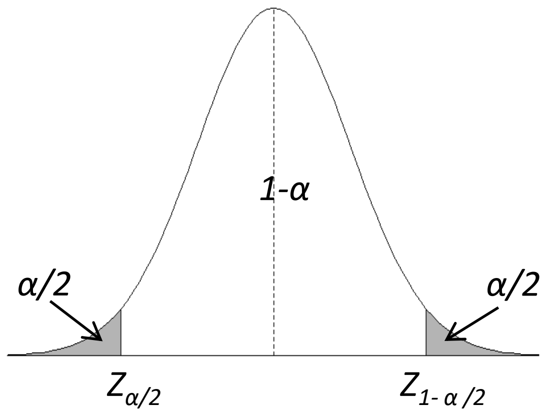
# Intervalos de confianza para la Media

$$IC\langle \mu \rangle_{(1-\alpha)\%} = \left[ \bar{X} \pm \left( Z * \frac{\sigma}{\sqrt{n}} \right) \right]$$

$$Li = \bar{X} - \left( Z * \frac{\sigma}{\sqrt{n}} \right)$$

$$Ls = \bar{X} + \left( Z * \frac{\sigma}{\sqrt{n}} \right)$$

Desviación Conocida o *n≥30* se utiliza la distribución Normal Estándar




Lo cual indica que existe una probabilidad del (1- $\alpha$)% de que la media poblacional se encuentre comprendida en el intervalo estimado.

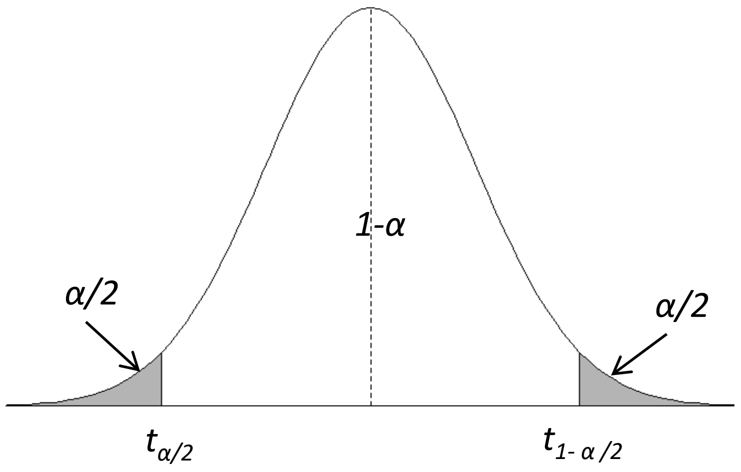
# Intervalos de confianza para la Media

$$IC\langle\mu\rangle_{(1-\alpha)\%} = \left[\bar{X} \pm \left(t * \frac{s}{\sqrt{n}}\right)\right]$$

Desviación Desconocida y *n<30* se utiliza la distribución *t*-Student

$$Li = \bar{X} - \left(t * \frac{s}{\sqrt{n}}\right)$$

$$Ls = \bar{X} + \left(t * \frac{s}{\sqrt{n}}\right)$$




Lo cual indica que existe una probabilidad del (1- $\alpha$)% de que la media poblacional se encuentre comprendida en el intervalo estimado.

# Ejemplo

Se estudia la temperatura de flexión bajo carga de una formulación de un tubo de plástico ABS. Se elabora una muestra de 12 observaciones midiendo las temperaturas de flexión (en °F) en cada una de ellas.

El ingeniero encargado del proceso desea saber cual es la temperatura media a la cual el tubo de plástico, sufre un determinado valor de deformación.

| Formulación 1 | |
|---|---|
| 206 | 185 |
| 188 | 189 |
| 205 | 192 |
| 187 | 210 |
| 193 | 194 |
| 207 | 178 |

$n = 12$
$\bar{x} = 194.5$
$s = 10.18$

Para determinar lo anterior, se desea estimar un intervalo de confianza para la temperatura media al 95% de confianza. **¿Cual sería el intervalo si se considera un nivel de confianza del 95%?**

Como σ es desconocido
*n<30* se utiliza la
Distribución *t*-Student

$$IC\langle\mu\rangle_{(1-\alpha)\%} = \left[\bar{X} \pm \left(t * \frac{s}{\sqrt{n}}\right)\right]$$

$$1 - \alpha = 0.95$$
$$\alpha = 0.05$$
$$\alpha/2 = 0.025$$

$$\longrightarrow \quad t_{\left(\frac{\alpha}{2};\, n-1\right)} = t_{(0.025\,;11)} = -2.20$$

`=INV.T(0,025;11)`

$$IC\langle\mu\rangle_{95\%} = \left[194.5 \pm \left(2.20 * \frac{10.18}{\sqrt{12}}\right)\right]$$

$$IC\langle\mu\rangle_{95\%} = [188; 201]$$

Con un nivel de confianza del 95%, la temperatura de felxión bajo carga de la formulación ABS estará entre 188 y 201 °F.

# Ejercicio

Cierta firma encargada del control de la contaminación de un rio, desea estimar la cantidad promedio de oxigeno disuelto, para ello ha tomado 10 observaciones cuyos resultados son los siguientes:

**2.6, 3.6, 3.1, 2.6, 2.7, 3.9, 2.4, 2.7, 2.5, 2.3**

El ingeniero encargado dice poseer fuertes indicios de que esta variable sigue una distribución normal con desviación estándar de 1.2.

a. Estime un intervalo de confianza para la cantidad promedio de oxigeno disuelto al 95% de confianza.
b. Como calcularía el intervalo si la desviación estándar no se conoce y debe estimarse desde la muestra?

# ¿De que depende la amplitud del intervalo?

$$IC\langle\mu\rangle_{(1-\alpha)\%} = \left[\bar{X} \pm Z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right]$$

$$LI: \bar{X} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$LS: \bar{X} + Z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

La amplitud del intervalo es:

$$2 * Z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \longrightarrow$$

En la amplitud intervienen:

*Tamaño de muestra

*Variabilidad de las observaciones

*Nivel de confianza

# ¿Cómo controlar la amplitud del Intervalo?

Es preferible un intervalo más estrecho debido a la precisión adicional que proporciona. Hay dos métodos principales para lograr un intervalo más preciso:

**1) Reducir el nivel de confianza:**

Disminuye la amplitud del intervalo de confianza pero aumenta la probabilidad de error.

**2) incrementar el tamaño de muestra:**

Incrementando el tamaño de muestra se puede reducir el error estándar, lo cual conlleva a un intervalo más preciso.

# Intervalo para la Proporción $\pi$

Porcentaje de unidades defectuosas.

Prevalencia de una enfermedad.

$$IC\langle \pi \rangle_{(1-\alpha)\%} = \left[ p \pm \left( Z\sqrt{\frac{p(1-p)}{n}} \right) \right]$$

$$Li = p - \left( Z\sqrt{\frac{p(1-p)}{n}} \right)$$

$$Ls = p + \left( Z\sqrt{\frac{p(1-p)}{n}} \right)$$

Cuando **n** es lo suficientemente grande se utiliza la distribución Normal Estándar

# Ejemplo

El propietario de West End Kwick Fill Gas Station desea determinar la proporción de clientes que utilizan tarjeta de crédito o débito para pagar la gasolina en el área de las bombas. Entrevistó a 100 clientes y descubre que 80 pagaron en ella.

$$n = 100$$
$$x = 80$$
$$p = \frac{x}{n} = \frac{80}{100} = 0.80$$

Con un nivel de confianza de 95%, calcule un intervalo de confianza para la proporción poblacional.

n*p = 100*0.8 = 80 > 5
n*(1-p) = 100*0.2=20 > 5
Distribución ***Z***

$$IC\langle\pi\rangle_{(1-\alpha)\%} = \left[p \pm \left(Z\sqrt{\frac{p(1-p)}{n}}\right)\right]$$

$$1-\alpha = 0.95$$
$$\alpha = 0.05$$
$$\alpha/2 = 0.025$$

$$\longrightarrow Z_{\left(\frac{\alpha}{2}\right)} = Z_{(0.025)} = -1.96$$

```
=INV.NORM.ESTAND(0,025)
```

$$IC\langle\pi\rangle_{(1-\alpha)\%} = \left[0.80 \pm \left(1.96\sqrt{\frac{0.80*0.20}{100}}\right)\right]$$

$$IC\langle\pi\rangle_{(1-\alpha)\%} = [0.72;0.88]$$

Con un nivel de confianza del 95%, la proporción de clientes que pagan con tarjeta esta entre el 0.72 y 0.88.

## Ejercicio

Schadek Silkscreen Printing, Inc., compra tazas de plástico para imprimir en ellas logotipos de eventos deportivos, graduaciones, cumpleaños u otras ocasiones importantes. Zack Schadek, el propietario, recibió un envío grande esta mañana. Para asegurarse de la calidad del envío, seleccionó una muestra aleatoria de 300 tazas. Halló que 15 estaban defectuosas.

Construya un intervalo de confianza de 90% de la proporción de tazas defectuosas.