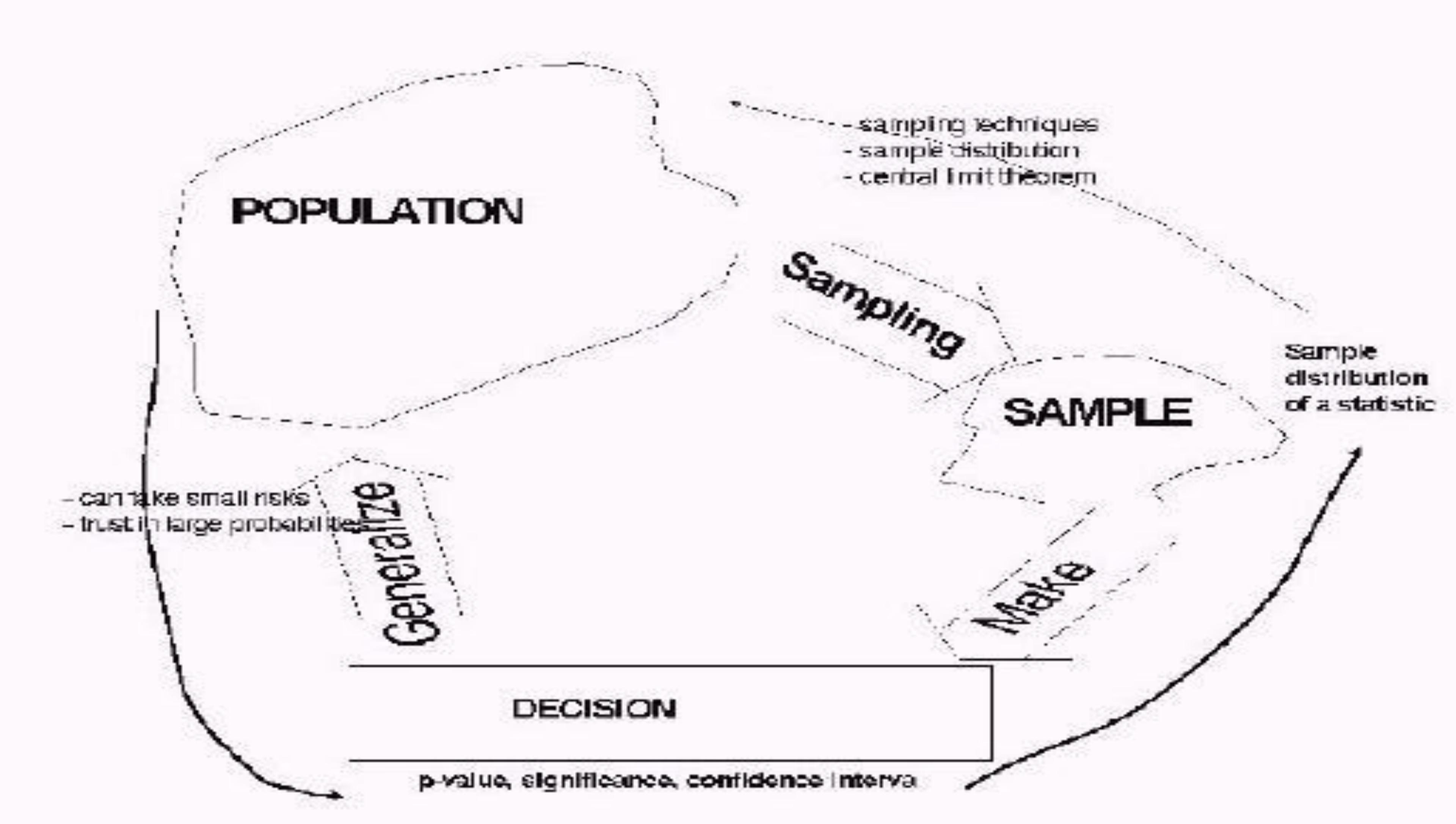
Estimación



Osca Centeno Mora

Índice

Índice

Introducción

- En todo estudio o investigación existe un campo de referencia, universo o **población** al cual se desea generalizar los resultados o conclusiones que se obtengan.

- Por **población** o universo se entiende el total de personas, objetos o mediciones que tienen una característica común.

- "El total de todas las observaciones correspondientes a una cierta característica". Gómez. M.

Introducción

- Las investigaciones y experimentos se llevan a cabo con el propósito de llegar a conclusiones o leyes de carácter general, es decir, buscando resultados que se puedan generalizar a todos los elementos de la población.

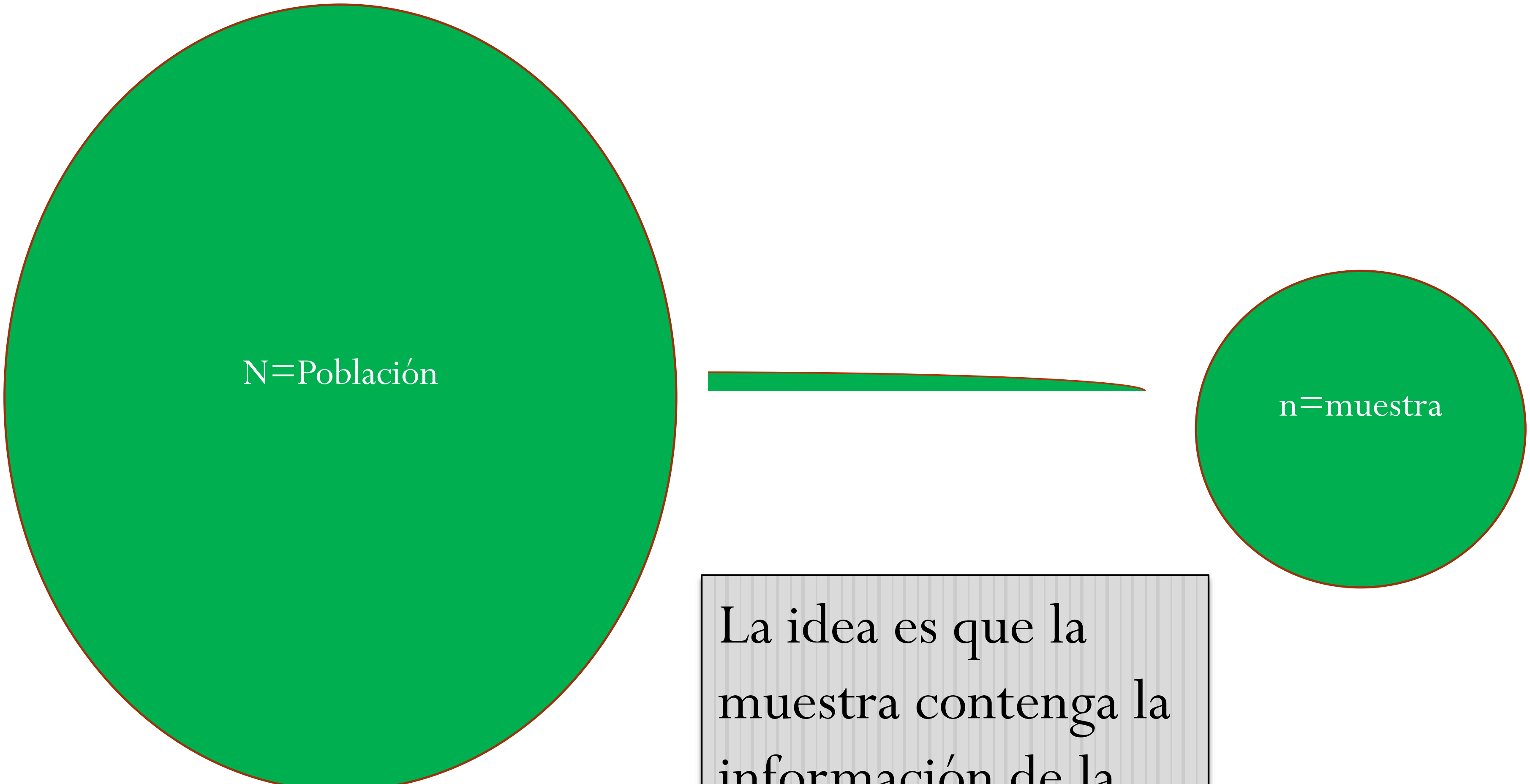
- El sentido común indica que lo mejor sería "censar" a todos los elementos que constituyen la población.

- Sin embargo, esto resulta muchas veces imposible por cuestiones de administración, tiempo, grandeza y dinero (recursos).

- Como no se puede "censar" toda la población, se toma una muestra y los resultados de ella se generalizan a toda la población.

Introducción

- Generalización de la información:



La idea es que la muestra contenga la información de la población

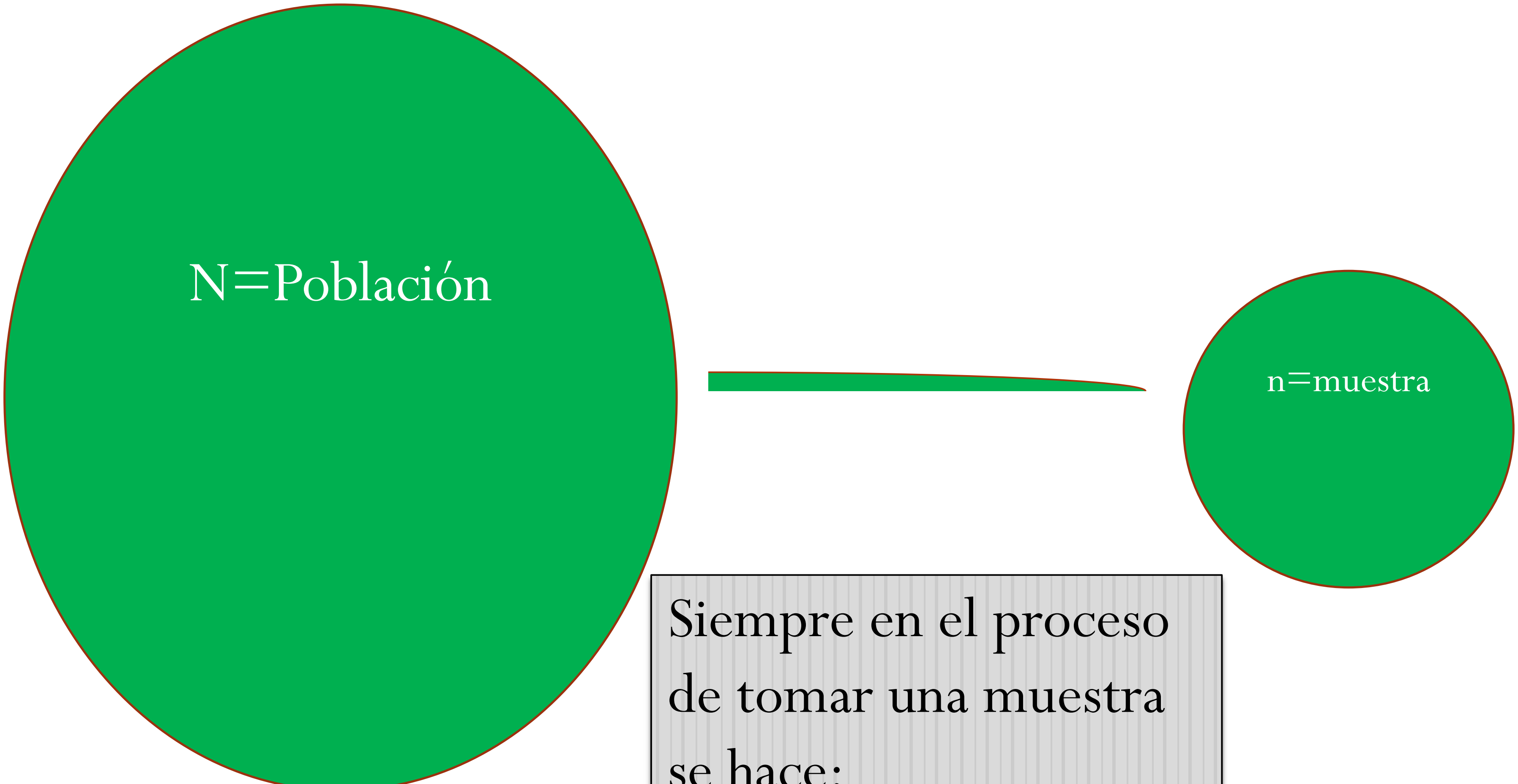
Introducción

- La razón más importante de la muestra es la no necesidad de estudiar toda la población.

- Los resultados que arrojaría una muestra bien seleccionada de tamaño razonable, serán suficientemente precisos para alcanzar los fines prácticos: generalizar a la población.

- ***Inferencia estadística:*** "es el proceso mediante el cual se generalizan los resultados observados en una muestra aleatoria, a la población de la cual se extrajo (población de interés)." Gomez, M.

Introducción



Siempre en el proceso de tomar una muestra se hace:

Inferencia Estadística

Índice

Características de la muestra

- Para tener resultados válidos de generalización, se dice que la muestra debe ser aleatoria o probabilística.

- Esto es, los elementos que constituyen la muestran fueron seleccionados utilizando mecanismos que da a cada elemento de la población una probabilidad conocida de ser incluida en la muestra, y estos fueron seleccionadas al azar.

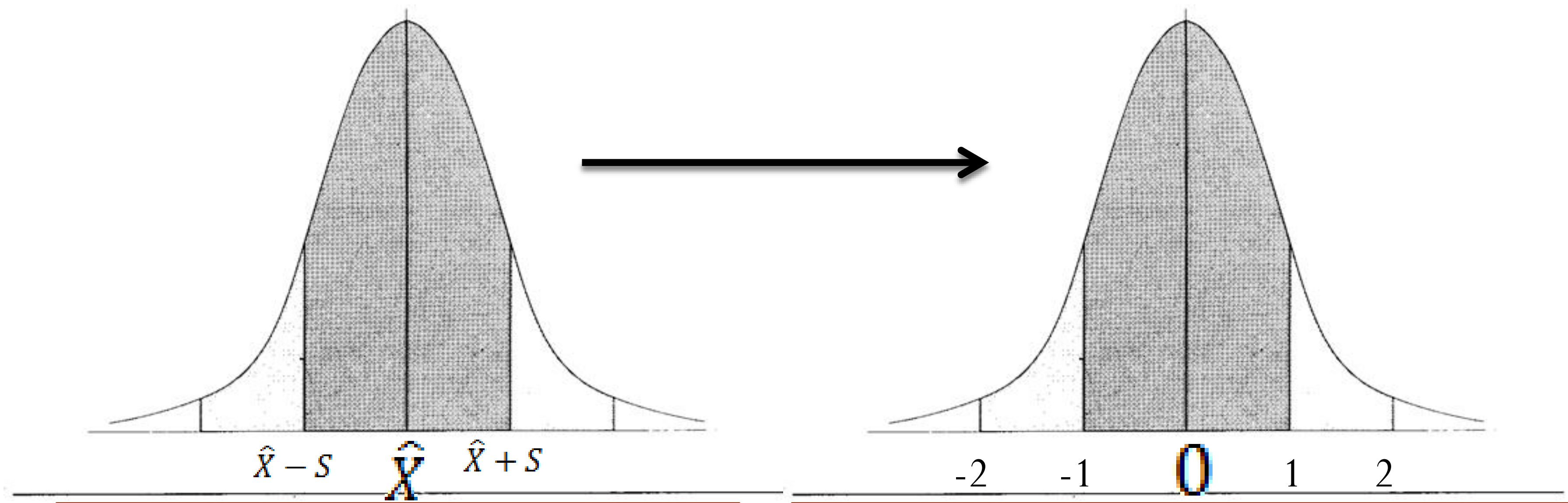
- Al ser la muestra aleatoria, esto permite que esta posea una distribución probabilística y es posible aplicar la teoría de las probabilidades para evaluar la confiabilidad de las inferencias.

Aplicación de la curva normal estándar
$N(0,1)$

Características de la muestra

n=muestra

Al ser una muestra aleatoria, se puede aplicar la teoría de probabilidad para obtener inferencias confiables



Información de la muestra

Aplicación de las probabilidades

El error asociado a la muestra

- Cada muestra posee un error asociado a la generalización (predicho probabilísticamente).

- Esto es, como no se toma en cuenta a todos los individuos de la población, entonces hay una cierta diferencia entre el valor verdadero que si quiere (el de la población), y el que se obtuvo con la muestra.

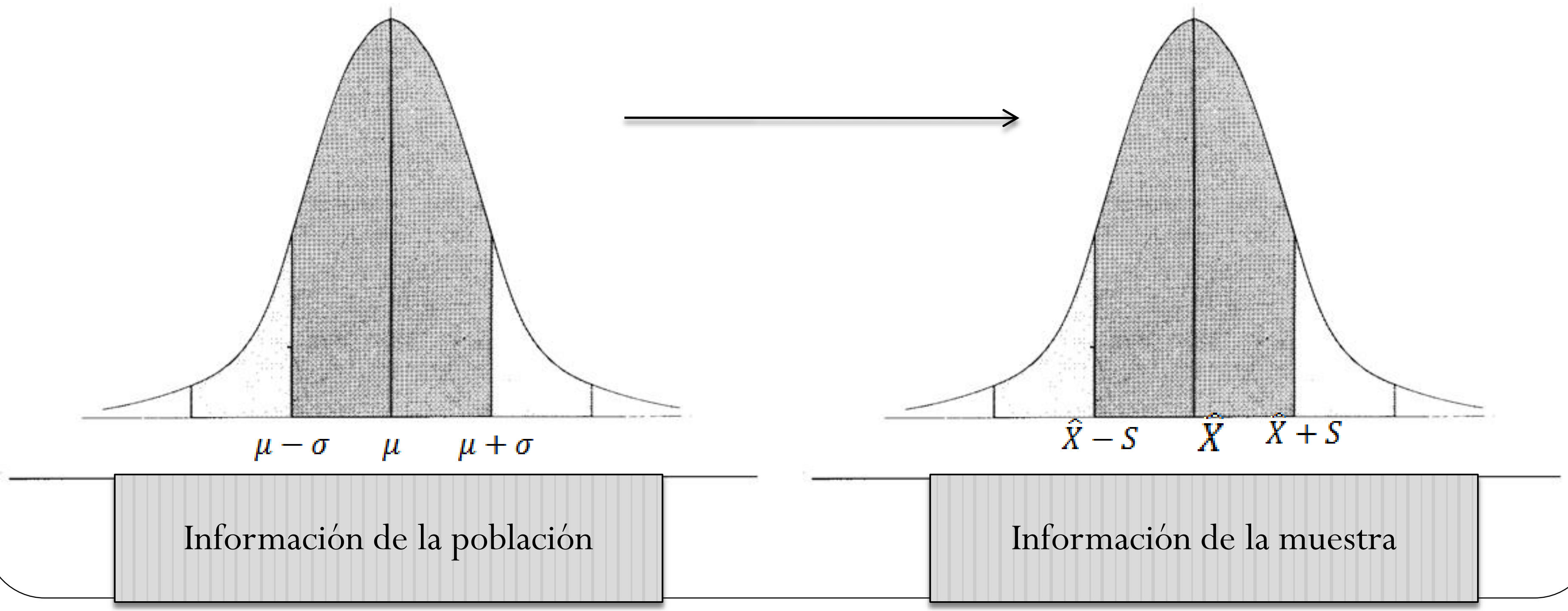
- Mediante la teoría de probabilidades, si y sólo si se tomó una muestra aleatoria, es posible, es posible medir ese error entre el valor real de la población y el de la muestra.

- A este error se le conoce como **error de muestreo**.

El error asociado a la muestra

La diferencia entre el valor real de la población población y el de la muestra se denominada error de muestreo.

Para el caso del promedio: $\mu - \bar{x}$ = error de muestreo



Muestra no aleatorio o al azar

- Si la muestra es seleccionada en forma intencional o de juicio, los errores no tendrán un comportamiento aleatorio, sino que dependerán de los prejuicios y de tendencias conscientes o inconscientes del que hizo la selección.
- En estas circunstancias no es posible aplicar la teoría de las probabilidades para evaluar las inferencias o generalizaciones.
- Al no tener una forma objetiva de evaluar los errores de muestreo, la inferencia estadística no puede ser aplicada a las muestras no aleatorias.
- Por lo tanto, por esta y otras razones, se debe aplicar una muestra aleatoria.

Muestra no aleatorio o al azar

- La preferencia por las muestras aleatorias para la realización de inferencias obedece a tres razones básicas:

1. Eliminan los sesgos o errores de selección.

2. Producen errores aleatorios de muestreo que son medibles utilizando modelos probabilísticos.

3. El error de muestreo puede hacerse tan pequeño como se quiera, aumentando el tamaño de la muestra.

Índice

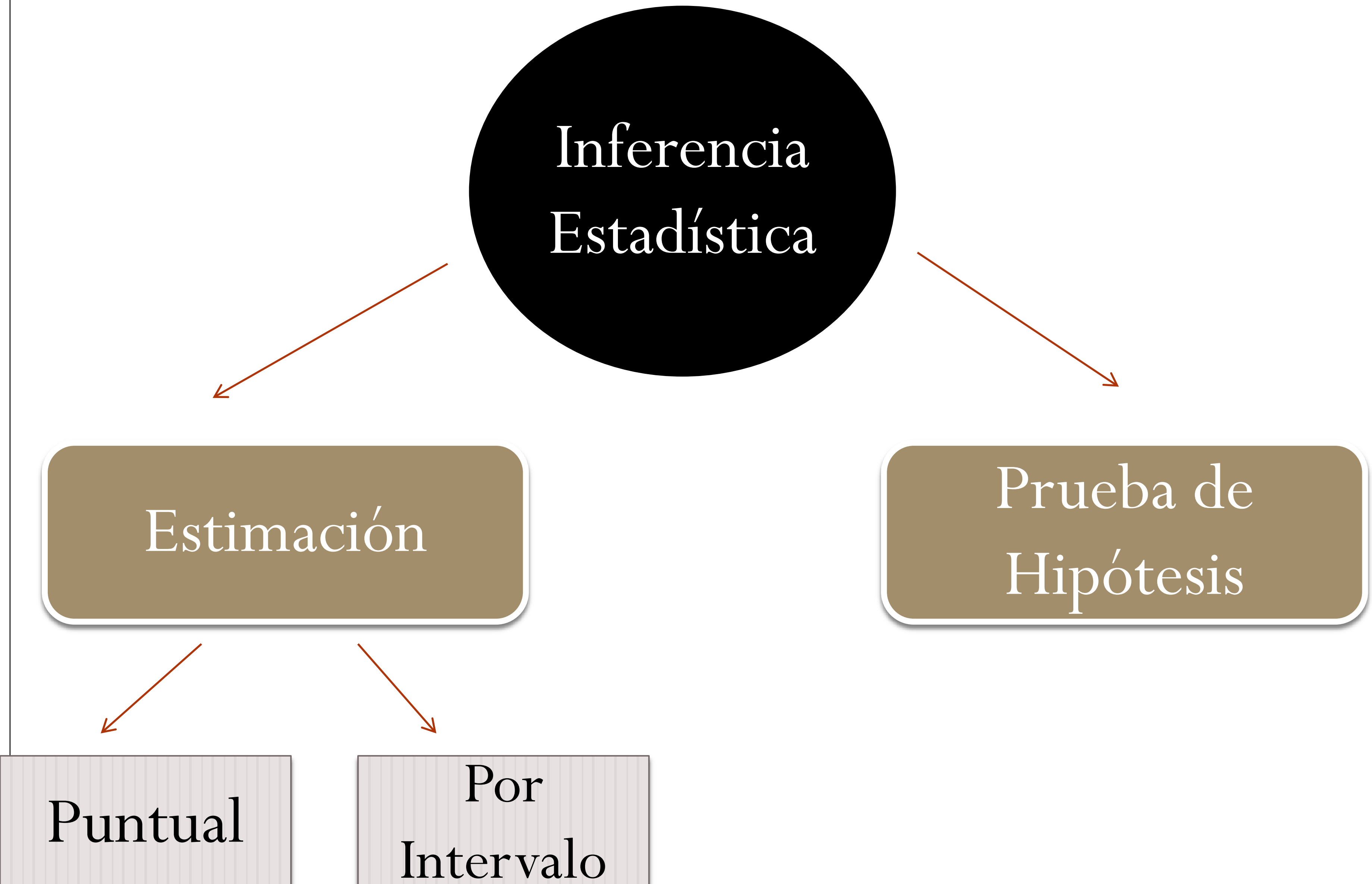
Inferencia estadística

- Dentro de la inferencia estadística se distinguen dos tipos de problemas:

1. Uno en que interesa ***estimar (estimación)*** características o propiedades de la población –promedio, variancia- de los elementos de ciertas características.
2. El otro es cuando interesa someter a ***prueba de hipótesis*** cierta característica de la población.

- Mientras que en la estimación interesa saber el rango de valores que podría tener cierto parámetro, en la prueba de hipótesis intersa si un determinado parámetro podría tener un cierto valor.

Inferencia estadística



Parámetros y Estimadores

- Las medidas que representan características o propiedades de la población, y que fueron obtenidas para toda la población, se denominan "valores poblacionales" o "Parámetros". Algunos a citar son:

Parámetro	Definición
μ	Promedio de la población
σ^2	Variancia de la población
σ	Desviación estándar de la población
Me	Mediana de la población
P	Proporción de la población
ρ	Coeficiente de población

Parámetros y Estimadores

- Generalmente, los valores poblacionales no son conocidos y una de las tareas básicas de la inferencia estadística es su ***estimación***.

- La estimación es el cálculo a partir de los datos de la muestra, de un valor numérico que será considerado, para efectos prácticos, como el valor de la población.

- Entonces, ***un estimador*** es una función de los valores de la muestra que se utiliza para aproximar o estimar el valor poblacional.

- La ***estimación*** es el valor numérico producido por una muestra específica (es el proceso de cálculo con un respectivo valor).

Parámetros y Estimadores

- Resumiendo:

Datos obtenidos para toda la población

N=Población

Parámetros

Datos obtenidos para una muestra

n=muestra

Estimadores

Parámetros y Estimadores

- Los estimadores más famosos son los del promedio (x barra), la variancia (S^2) y el de la desviación estándar (S). A continuación se presentan algunos parámetros y estimadores

ESTADÍSTICO	PARÁMETRO	ESTIMADOR
Media aritmética	μ	$\bar{x}$
Desviación estándar	σ_x	s
Variancia	σ^2	s^2
Proporción	P	$\hat{p}$
Coeficiente de correlación	ρ	r

Índice

Estimación puntual del parámetro μ

- Para estimar el valor del promedio de cierta población (μ) se podría utilizar el estimador $\bar{x}$. Esto sería una estimación puntual de μ.

- Ejemplo: se quiere estimar el promedio de edad de todos las personas de la Universidad de Costa Rica (μ). Se selecciona una única muestra para estimar el promedio ($\bar{x}$), y el resultado fue de 25.7 años.

$$\begin{tikzcd} \mu \arrow[r] \arrow[d] & \boxed{?????} \arrow[d] \\ \bar{x} \arrow[r] & \boxed{25.7} \end{tikzcd}$$

Estimación puntual del parámetro μ

- Sin embargo, antes se había discutido que, al no tomar toda la información de la población, se podría producir una diferencia imporante entre el valor real (μ), y el del estimador ($\bar{x}$).

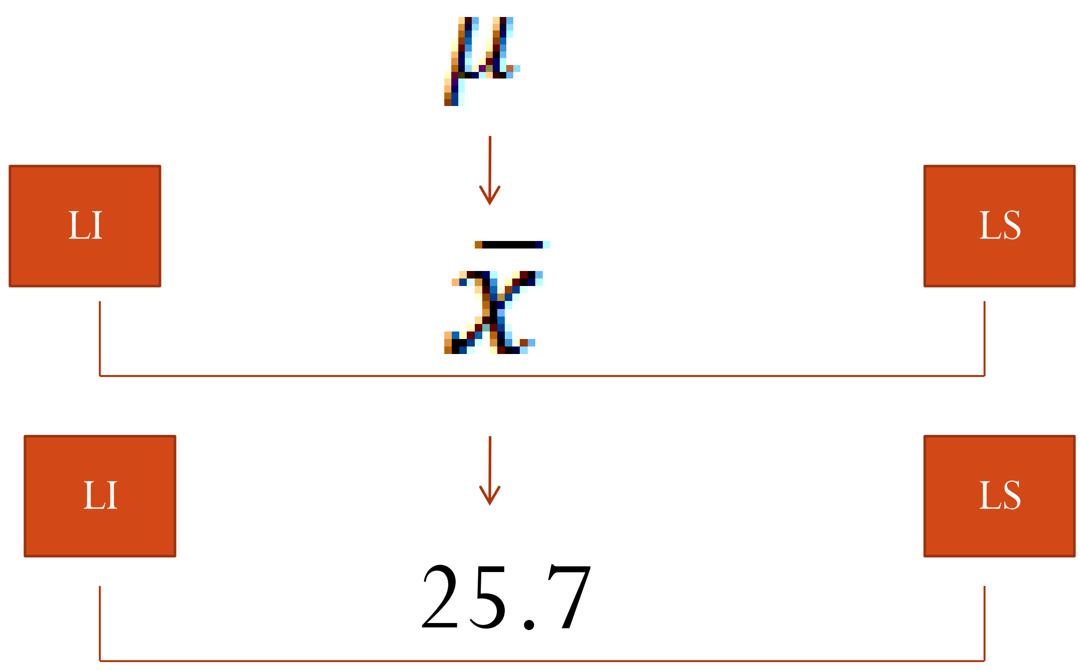
Error de muestreo: μ-X = E.M.

- De ahí la importancia de contar con una idea acerca del error aosociado con la estimación que se ha realizado.

- La estimación puntual es insuficiente si no se posee alguna medida medida de confianza.

Estimación puntual del parámetro μ

- El problema se resuelve utilizando la ***estimación por intervalos***, o sea construyendo, a partir de la estimación puntual, un intervalo dentro de cuyos límites, se espera, con una determinada grado de confianza, que esté contenido el verdadero valor poblacional.



Estimación puntual del parámetro μ

- Para nuestro presente caso, la estimación del promedio (25.7 años), podría estar muy alejada del valor real de la población de interés (μ), o poco alejada, no se sabe.

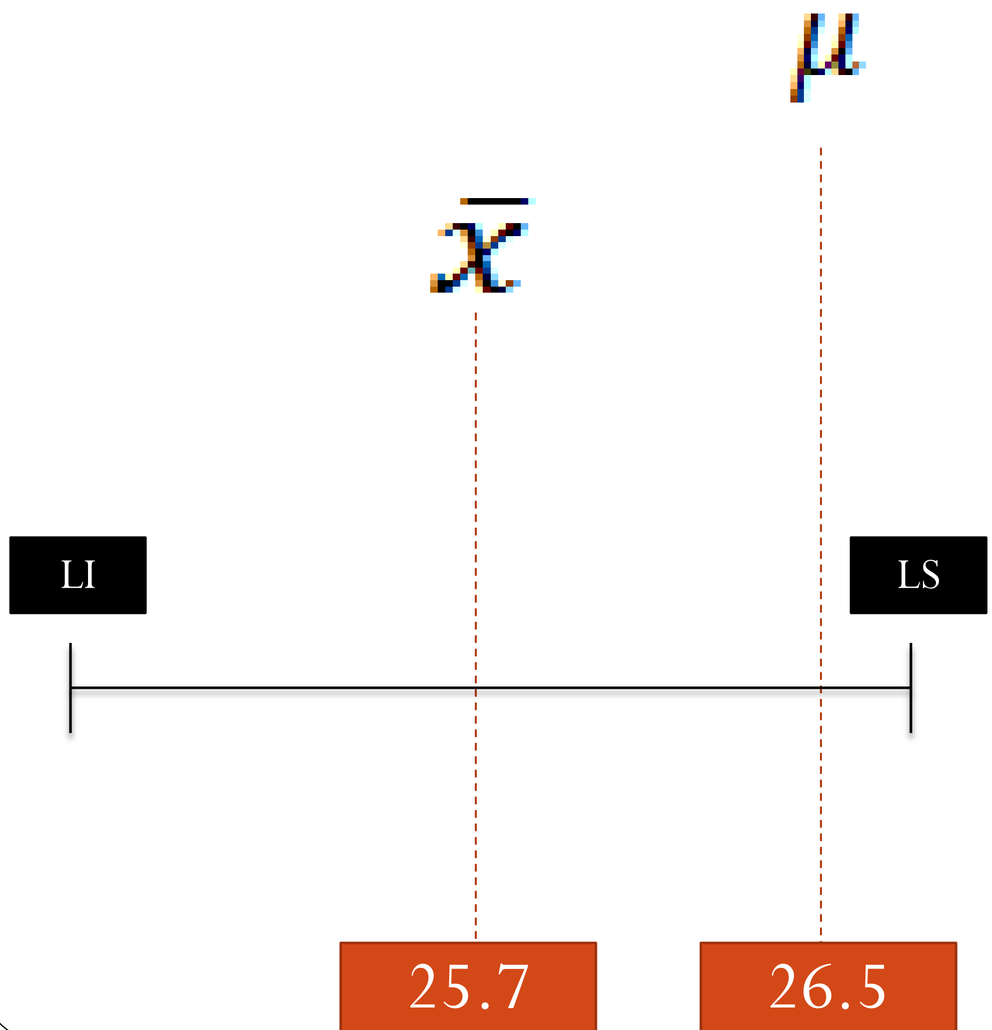
- Para superar esta dificultad, esta incertidumbre, puede usarse la estimación por intervalos.

- Al decir que el valor de 25.7 años puede estar entre dos valor (llamados límites inferior y superior), se tendría una mayor confianza de reportar el resultado.

- De ahí la gran importancia de la estimación por intervalos.

Estimación puntual del parámetro μ

- La utilidad de los intervalos de confianza:



Aunque el estimador no es igual que el parámetro, con el cálculo de los intervalos de confianza, tenemos que el intervalo está conteniendo al valor poblacional μ

Índice

1 Introducción

2 Características de la muestra

3 Inferencia estadística

4 Estimación puntual y por intervalos

5 Límites de confianza

Límites de confianza para μ

- Supóngase que la distribución de la población es una curva normal con media μ y variancia σ^2 (desviación estándar σ), y de esta se tomó una muestra de tamaño n.

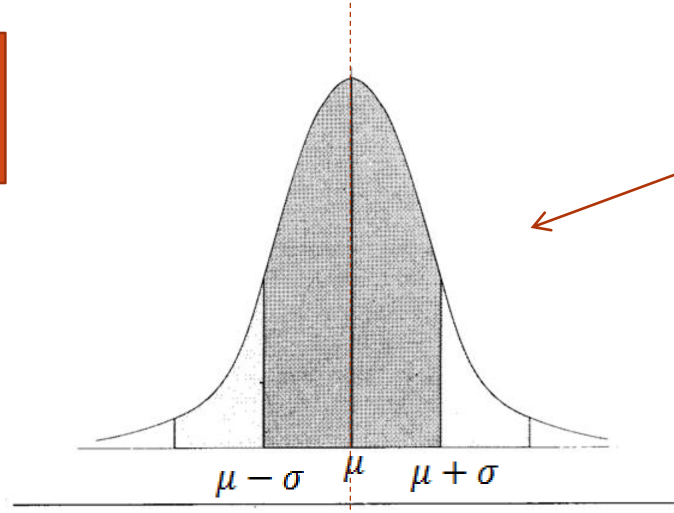
- Ahora, se desea saber la posible distribución alrededor de la media muestral ($\bar{x}$) que se obtuvo.

- Por teoría sabemos entonces que la distribución muestral tendrá una media de $\bar{x}$ y una variancia de σ^2/n (desviación estándar $\sigma/\sqrt{n}$).

- Si la curva poblacional tiene una distribución $N(\mu,\sigma)$, la muestral tiene una curva de $N(\bar{x},\sigma/\sqrt{n})$

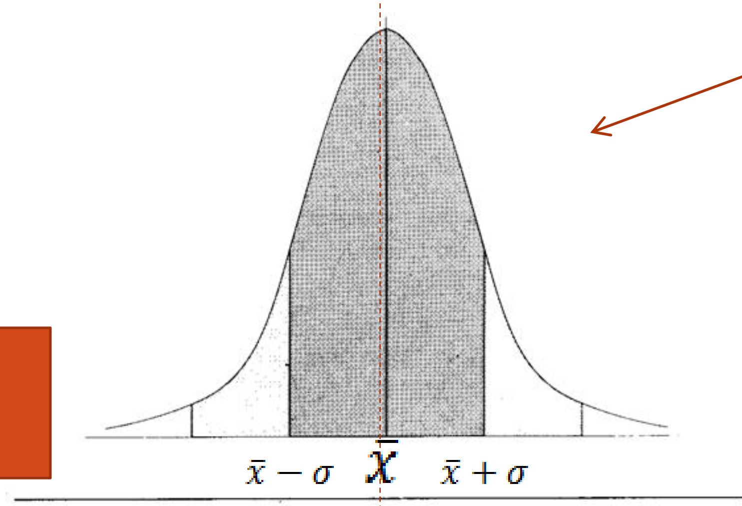
Límites de confianza para μ

Distribución de la población



$$\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Se extrae muestra de tamaña "n"



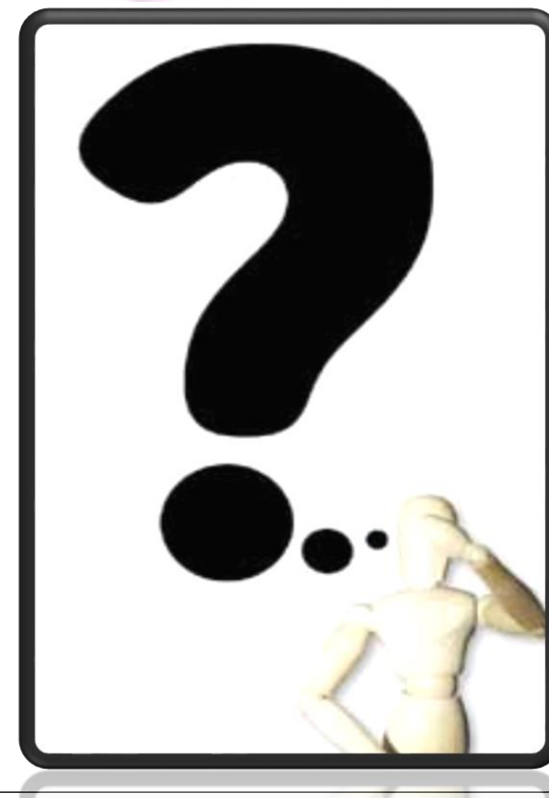
$$\frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma/\sqrt{n}}\right)^2}$$

Distribución de la muestra

Límites de confianza para μ

- Lo deseable es que el valor del promedio muestral ($\bar{x}$), fuera exactamente igual como el del poblacional (μ), como se dio en la filmina anterior.

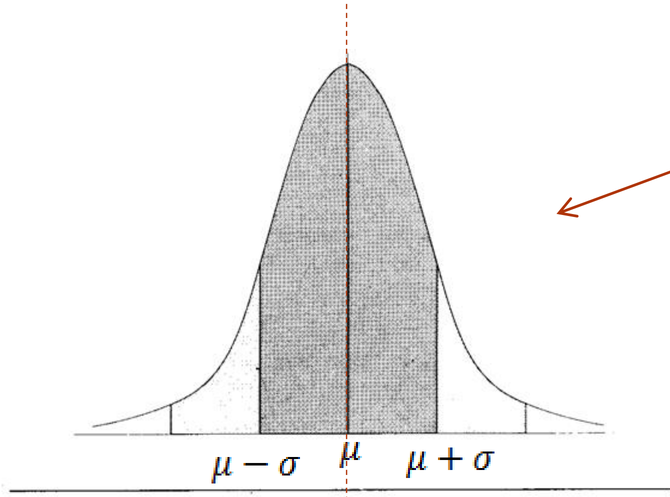
- Como casi nunca se sabe el valor verdadero de μ, entonces lo ideal sería brindar posibles límites de confianza en donde se podría esperar que estos límites contengan al parámetro μ, con un determinado error probabilístico asociado.



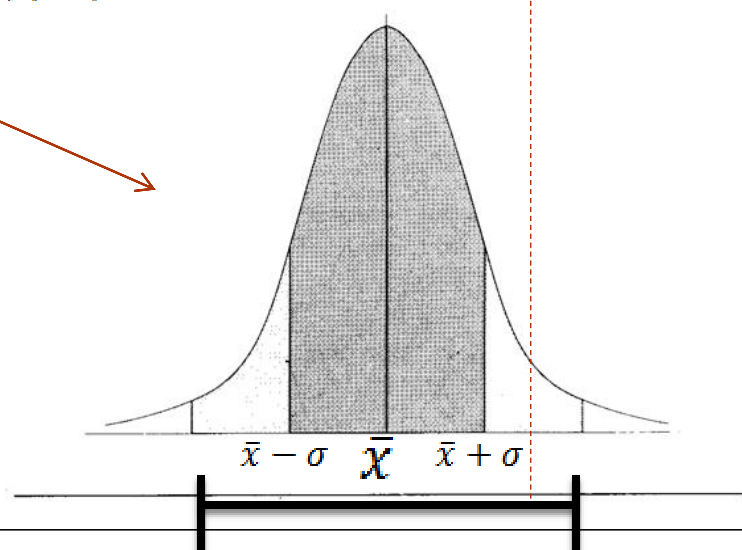
- Para este error probabilístico utilizaremos la curva normal estándar y sus valores de Z.

Límites de confianza para μ

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}$$



$$\frac{1}{\frac{\sigma}{\sqrt{n}}\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma/\sqrt{n}}\right)^{2}}$$



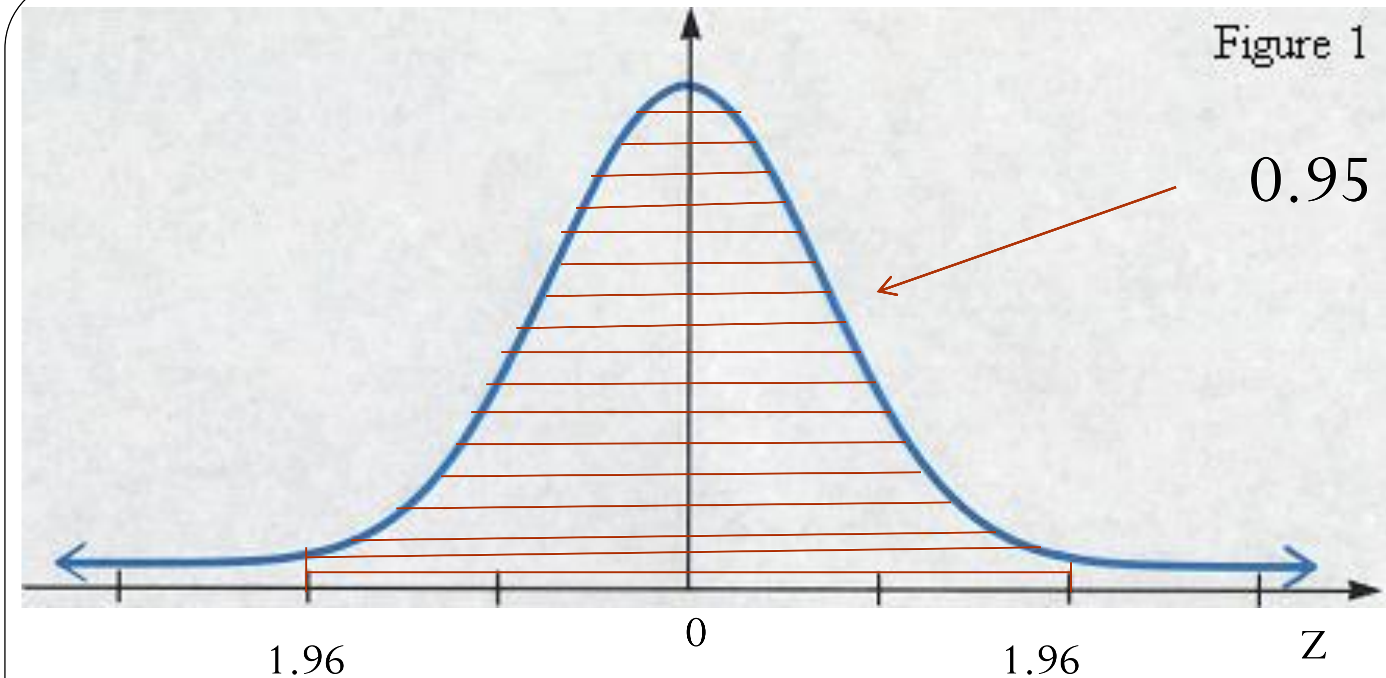
Límites de confianza para μ

- Ahora nos preguntamos: ¿dentro de qué valores se encontrará el 95% de los valores de $\bar{x}$?

- Se sabe, por la teoría estadística, que en una curva normal estándar, el 95% central del área bajo la curva está entre

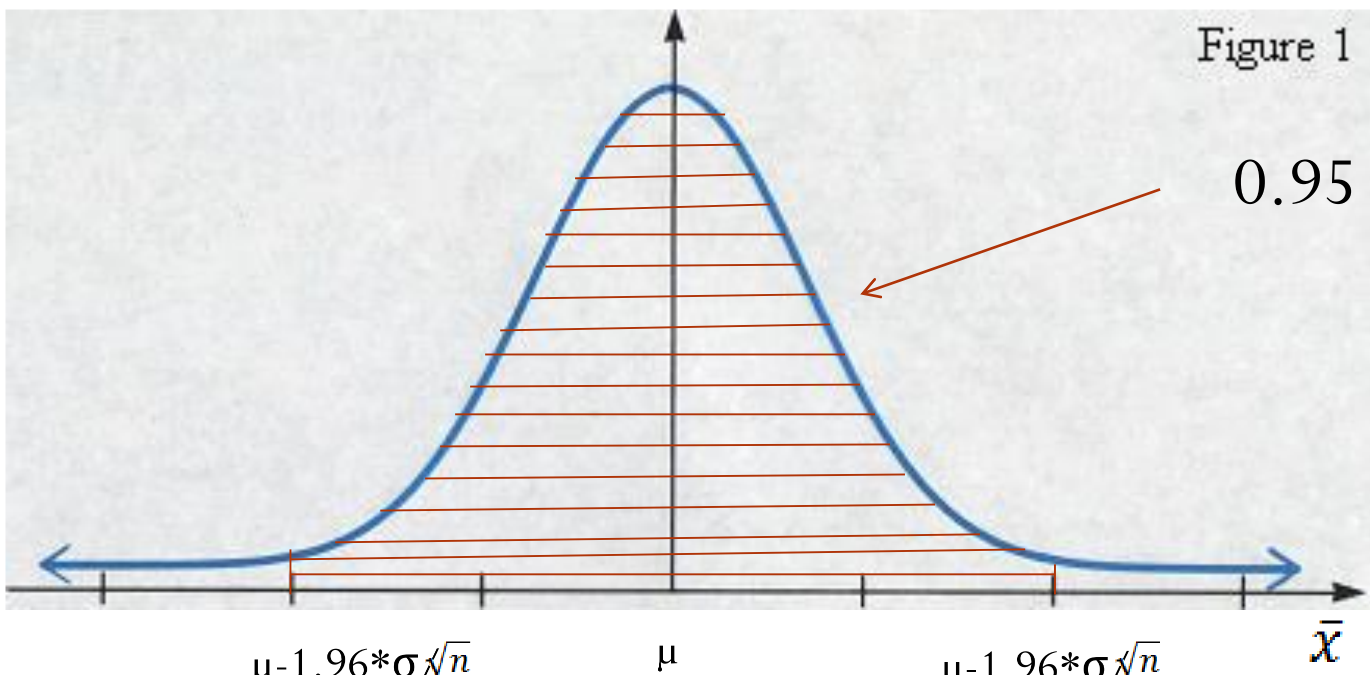
  -1.96 y +1.96; por lo tanto, en la distribución de $\bar{x}$, el 95% de los valores se encontrarán entre:

  "μ-1.96*σ/$\sqrt{n}$ " y "μ+1.96*σ/



Normal Estándar

Normal con μ y σ



Límites de confianza para μ

- La probabilidad de que $\bar{x}$ asuma un valor en el intervalo

$\mu \pm 1.96*\sigma/\sqrt{n}$ es de 0.95.

- Esto puede expresarse simbólicamente en la siguiente forma:

$$P\left[\mu - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \bar{x} \leq \mu + 1.96 * \frac{\sigma}{\sqrt{n}}\right] = 0.95$$

- De la expresión anterior se deriva la fórmula para el cálculo de un intervalo de confianza del 95% para μ. Mediante a una sencilla sencilla manipulación algebraica obtendríamos los intervalos del confianza con un 95%

$$P\left[\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right] = 0.95$$

Límites de confianza para μ

- Los límites de confianza se generan de la presente fórmula:

$$P\left[\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right] = 0.95$$

LI: Límite Inferior

LS: Límite Superior

$$P[LI \leq \mu \leq LS] = 0.95$$

Límites de confianza para μ

- Tenemos entonces que:

LI = límite inferior $= \bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$

LS = límite superior $= \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}$

Entonces, el objetivo de la estimación por intervalos es calcular los límites de confianza (el LI y LS), para decir que, con una probabilidad del 95%, el intervalo va a contener al verdadero promedo poblacional μ.

Ejemplo de estimación por intervalos

- Se quiere estimar el promedio de edad (μ) de todos las personas de la Santa Paula.

- Se selecciona una muestra (n) de tamaño 50 para estimar el promedio de edad (μ), y el resultado fue de 25.7 años ($\bar{x}$), con una desviación estándar de 12 (σ).

- Dado que no se seleccionó a todas las personas de la Santa Paula, para tener más confianza en las estimaciones, se quiere utilizar los intervalos de confianza con un 95% de confiabilidad, esto es, con 5% de aceptar un error.

μ= ?????	σ = 12	Confianza=95%
$\bar{x}$ = 25.7	n= 50	Error=5%

Ejemplo de estimación por intervalos

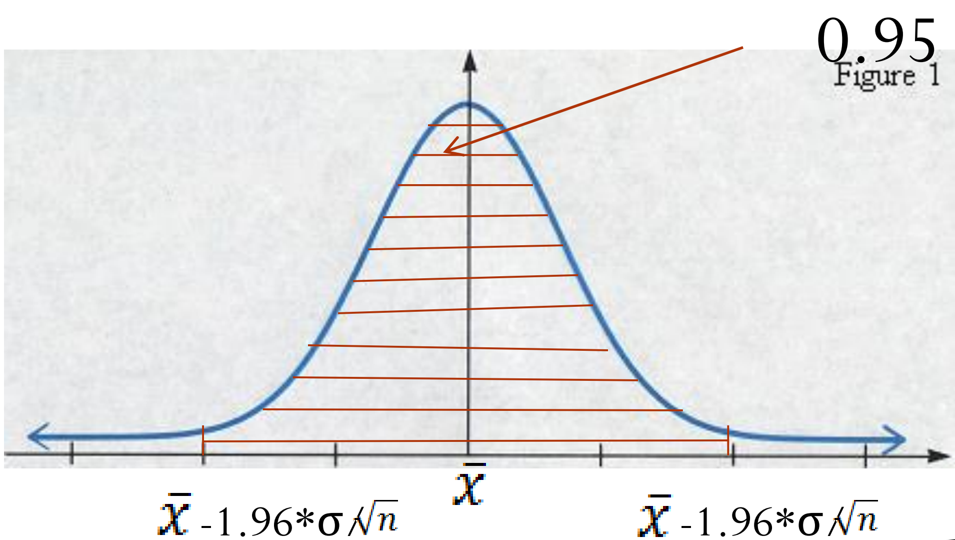
- Entonces necesitamos calcular los intervalos de confianza.

$$Límite\ inferior: LI = \bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$$

$$Límite\ superior: LS = \bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}}$$

- Y sabemos que, este límite tiene un 95% de confianza, por la fórmula antes vista:

$$P\left[\bar{x} - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 * \frac{\sigma}{\sqrt{n}}\right] =$$



Ejemplo de estimación por intervalos

- Solución:

$$Límite\ inferior: LI = 25.7 - 1.96 * \frac{12}{\sqrt{50}} = 22.37$$

$$Límite\ superior: LI = 25.7 + 1.96 * \frac{12}{\sqrt{50}} = 29.03$$

Corroborar:

$$\frac{22.37 + 29.03}{2} = 25.7$$

$$Límite\ inferior = 22.37$$

$$Límite\ superior = 29.03$$

Intervalo de confianza:

[22.37 ; 29.03]

- Interpretación: Se tiene una confianza del 95% de que intervalo calculado contenga al verdadero valor de la edad promedio en la Universidad Santa Paula.

Reseñas finales

- La utilidad de la estimación por intervalos se utiliza en todos los campos en donde se aplica la estadística.

- En general, está es la técnica más utilizada.

- El conocer un rango de valores que podría tener cierto parámetro es de gran utilidad, ya que no siempre se puede der del todo preciso.

- Gran parte de la teoría de probabilidades, la curva normal y la estandarización residen en la estimación por intervalos.


arte
FIN