UNIDAD II

MUESTRA Y DISTRIBUCIONES MUESTRALES

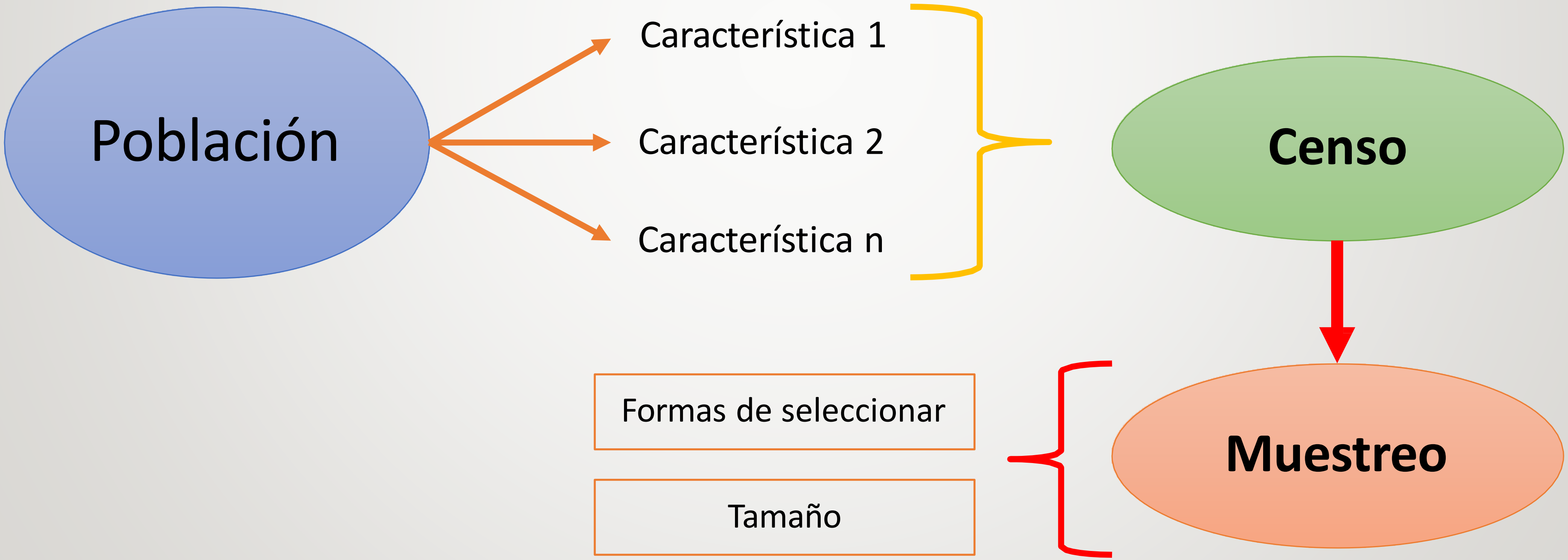
DOCENTE: ING. BLADIMIR RICARDO UREY FERNANDEZ

ESCUELA MILITAR DE INGENIERÍA
"Mcal. Antonio José de Sucre"
Prestigio, Disciplina y Mejores Oportunidades

TEMARIO

INTRODUCCIÓN

En todo estudio que se realice se desea conocer con absoluta verdad y certeza toda la información requerida para tales fines.



1. POBLACIÓN Y MUESTRA

1.1. POBLACIÓN

A) DEFINICIÓN

Población

Todo el conjunto de elementos con características comunes

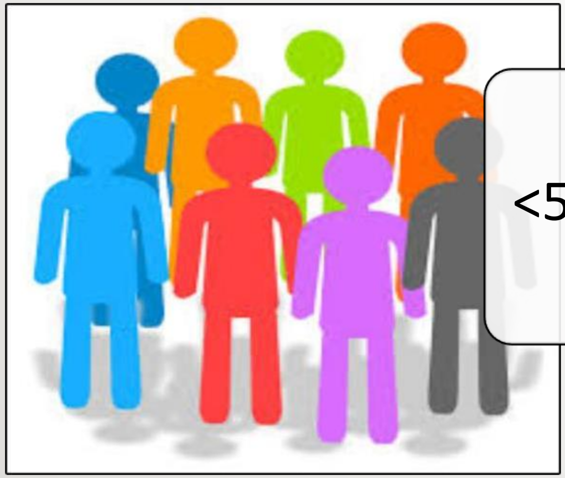
- Personas
- Animales
- Vegetales
- Objeto

1. POBLACIÓN Y MUESTRA

B) TIPOS DE POBLACIÓN

De acuerdo a la magnitud de la población se definen dos tipos de población:

FINITA



<500000

INFINITA

>500000

1. POBLACIÓN Y MUESTRA

C) PARÁMETRO

a) Caracterización numérica de la distribución de la población

- Describe parcial o completamente la función de probabilidad de la población de la variable de interés

b) Caracterización descriptiva de una determinada población

- Una población con determinadas características pueden ser descritas por ciertas medidas descriptivas

1. POBLACIÓN Y MUESTRA

D) CENSO

Población

Parámetro 1
Parámetro 2
Parámetro 3
Parámetro n

CENSO

CENSO

- Costoso
- Larga duración
- Emplea bastantes recursos
- Error = 0

1. POBLACIÓN Y MUESTRA

2. MUESTREO

A) DEFINICIÓN

Para el conocimiento de las características de la población existen métodos opcionales cuyo costo y tiempo de realización se reducen considerablemente.

↓

MUESTREO

1. POBLACIÓN Y MUESTRA

Población

Censo

Muestra

Muestreo

Reconstruir modelos reducidos de la población total

- Bajo costo
- Poco tiempo
- Existe error

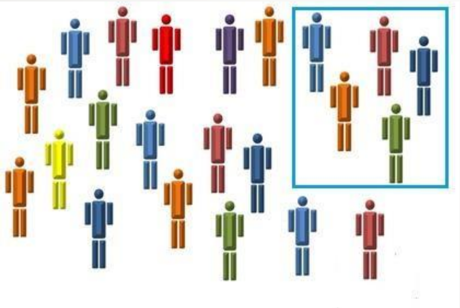
1. POBLACIÓN Y MUESTRA

B) MUESTRA PROBABILISTICA

Muestra

Subconjunto representativo de una organización

Probabilística



No Probabilística

Todos tienen la misma probabilidad de ser parte de la muestra

No todos tienen la misma probabilidad de ser parte de la muestra

1. POBLACIÓN Y MUESTRA

EJEMPLO 1:

- Si se tiene una florería que cuenta con 100 000 semillas que se desean comercializar, de la cual se sabe que pueden producir flores blancas o rojas.
- El objetivo para la gerencia es averiguar cuantas de estas 100 000 semillas producirán flores rojas.

1. POBLACIÓN Y MUESTRA

RESOLUCIÓN EJEMPLO 1:

1. Para dar una respuesta correcta, se debería sembrar todas las semillas y observar el numero de las que producen flores rojas.
2. Como ello es imposible, puesto que se desea vender todas las semillas y aunque no se quisiera venderlas, el obtener una respuesta requerirá invertir mucho esfuerzo y dinero.
3. La solución será emplear unas cuantas semillas y basados en los resultados aparecidos, hacer una afirmación sobre el numero de flores rojas que se tendrá del total restante de semillas.

1. POBLACIÓN Y MUESTRA

C) ESTADÍGRAFO

El estadígrafo es cualquier función de las variables que se observaron en la muestra, de manera que, esta función no contiene cantidades desconocidas.

Población

Parámetro 1

Parámetro n

Constantes

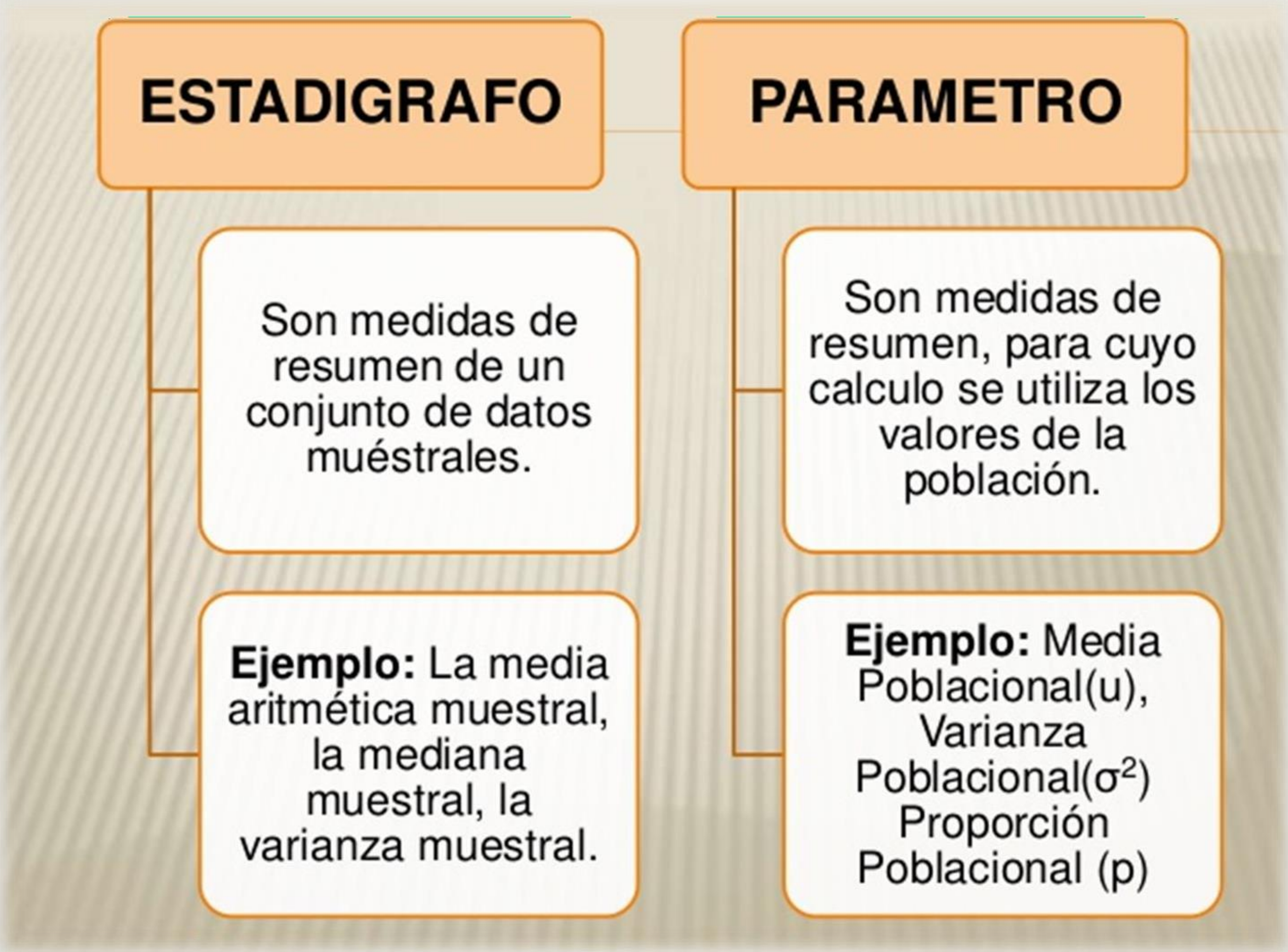
Muestra

Estadígrafo 1

Estadígrafo n

Variables aleatorias

1. POBLACIÓN Y MUESTRA



2. DISTRIBUCIONES MUESTRALES

2.1. DISEÑO DE UNA MUESTRA

Metodología para

tomar muestras

Secuencia de pasos

Paso 1

Paso 2

Paso 3

Paso n

2. DISTRIBUCIONES MUESTRALES

2.1.1. Criterios para evaluar el diseño de una muestra

- Criterios de evaluación
  - Fiabilidad — Error
  - Efectividad — Al menor costo

2. DISTRIBUCIONES MUESTRALES

2.2. TIPOS DE MUESTREO

2.2.1. Muestreo Aleatorio

Muestreo aleatorio

- Simple
- Sistemático
- Por conglomerado
- Estratificado

Se puede combinar

2. DISTRIBUCIONES MUESTRALES

A. Muestreo aleatorio Simple

Condiciones de aplicaciones

- Unidades fáciles de identificar
- Población pequeña
- Población con características de interés homogéneo

En el muestreo aleatorio simple la selección de los elementos se efectúa en una sola etapa y en forma directa, pudiendo ser con o sin reemplazo.

2. DISTRIBUCIONES MUESTRALES

Pasos :

- Determinar el tamaño de la población y el tamaño de la muestra
- Numerar los individuos de 1 a N
- Tirar unidades al azar (probabilidad igual)
- Aplicar entrevista o encuesta

Ejemplo 1:

Estudio sobre los estudiantes del 3er Semestre de la carrera de Ingenieria en Sistemas Electronicos que no les gusta la Asignatura de Estadística II.

2. DISTRIBUCIONES MUESTRALES

Con reemplazo

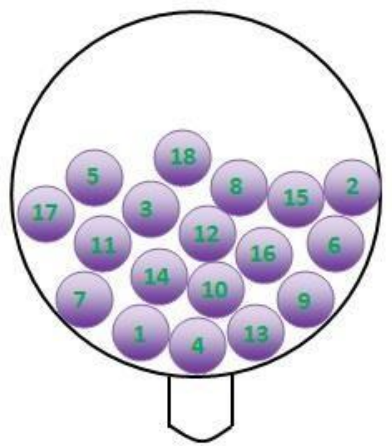
Cada elemento de la muestra posee la misma probabilidad de ser elegida

Cada uno es reintegrado a la población de la cual fue extraída

Sin reemplazo

Cada elemento posee la misma probabilidad de ser escogida

Considerando que la probabilidad de que un elemento sea extraído dependerá de los que anteriormente hayan sido elegidos



2. DISTRIBUCIONES MUESTRALES

B. Muestreo aleatorio Sistemático

Condiciones de aplicaciones

- Unidades fáciles de identificar
- Población pequeña
- Población con características de interés heterogéneo

Este procedimiento consiste en obtener una muestra tomando cada **k-ésima** unidad de la población, tras numerar las unidades de la población u ordenarlas de alguna manera. La letra **k** representa un número entero llamado razón de muestreo, coeficiente de elevación o salto es igual a: **k=N/n**

2. DISTRIBUCIONES MUESTRALES

- **Pasos :**
- Determinar el tamaño de la población y el tamaño de la mu

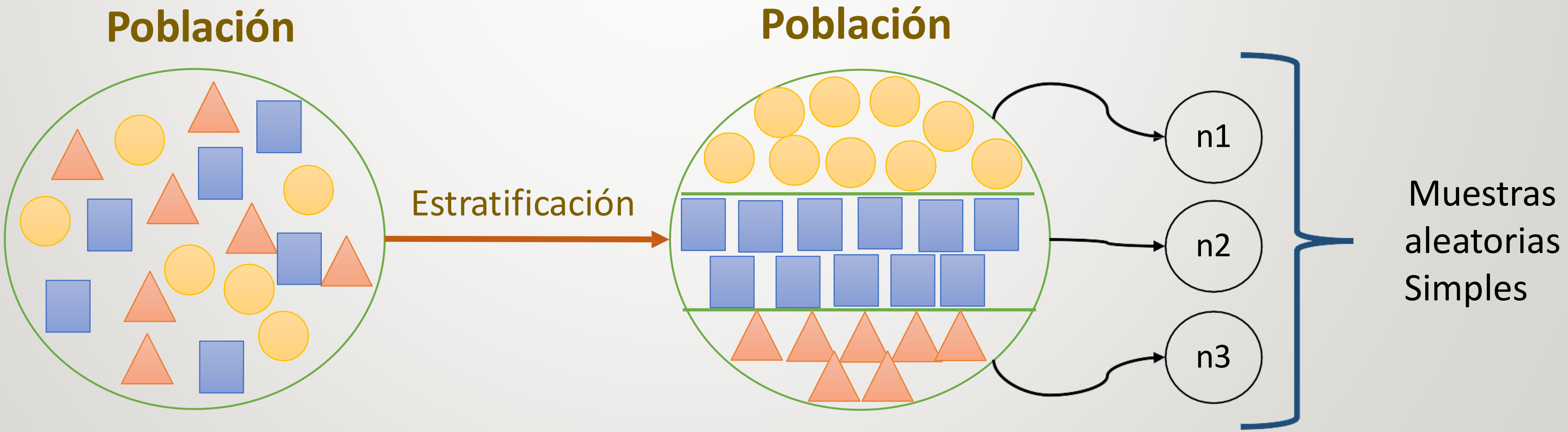
- Numerar los individuos de 1 a N
- Calcular el salto “k”
- Elegir la muestra (primero se elige al azar)
- **Ejemplo 2:**
- Estudio sobre las edad de los estudiantes de 3er semestre de la carrera de Ingeniería en Sistemas Electronicos.

2. DISTRIBUCIONES MUESTRALES

C. Muestreo aleatorio Estratificado

Condiciones de aplicaciones

- Población grande
- Población con características de interés heterogéneo
- Dividir la población en clases o grupos llamados estratos

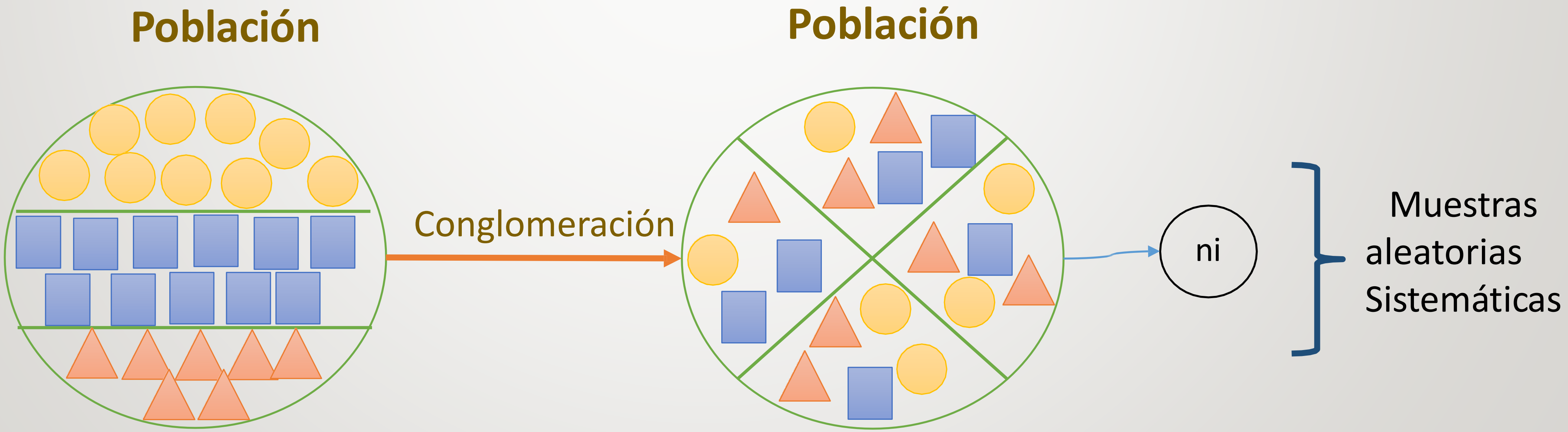


2. DISTRIBUCIONES MUESTRALES

D. Muestreo aleatorio por conglomerados

Condiciones de aplicaciones

- Población grande
- Población con características de interés homogéneo
- Dividir la población en áreas llamadas conglomerados



2. DISTRIBUCIONES MUESTRALES

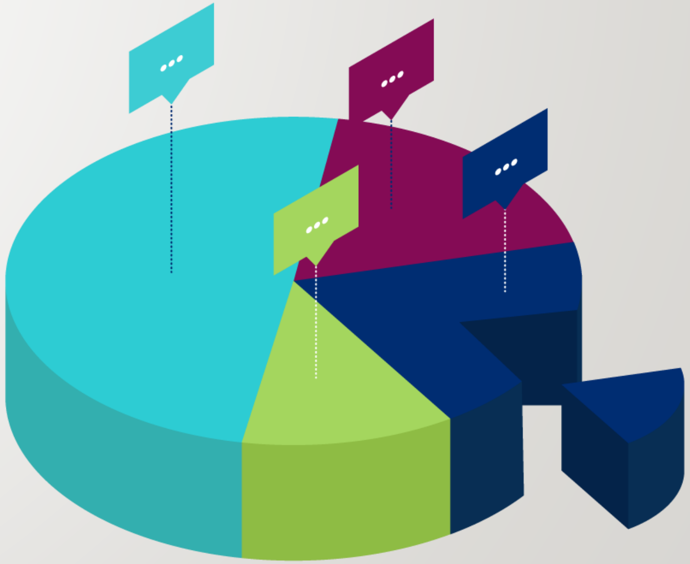
2.2.2. Muestreo No Aleatorio

Muestreo no aleatorio

Solo una parte de la población puede ser parte de la muestra

Se puede combinar

Generalmente, las instituciones oficiales tienden a emplear muestreos aleatorios y las instituciones de opinión, mayormente privadas, emplean el muestreo no aleatorio; ello en virtud a la disponibilidad de información y el costo que ello representa.



2. DISTRIBUCIONES MUESTRALES

Muestreo no aleatorio

Muestreo opinático

Se selecciona la muestra según el objetivo de la investigación

Muestreo por cuotas

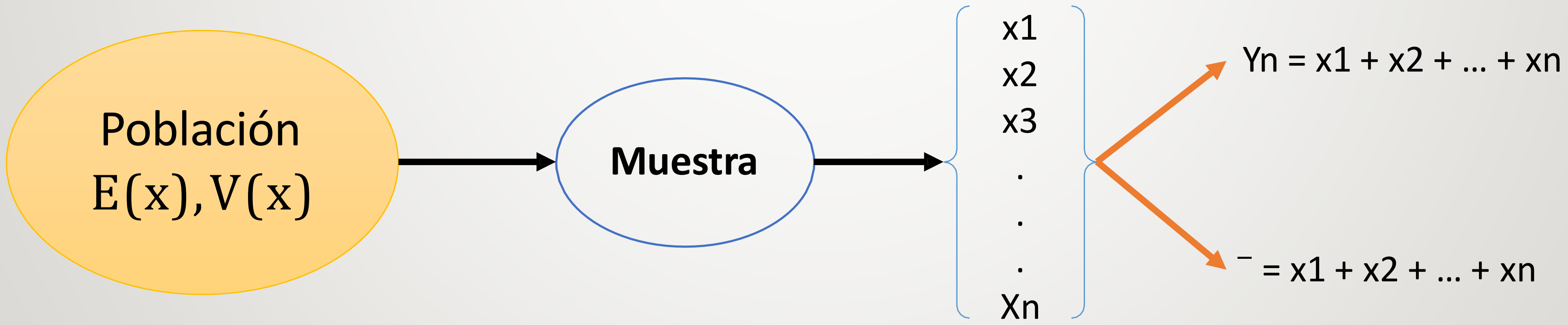
Facilitar el perfil de las personas que tiene que entrevistar

Eje.: Estudio sobre:
- Alcohol
- Drogas
- Tabaco
- Etc.

Eje.: Estudio sobre la leche en polvo:
- Mujer
- Madre
- Con niño en edad de lactancia

3. TEOREMA CENTRAL DEL LÍMITE

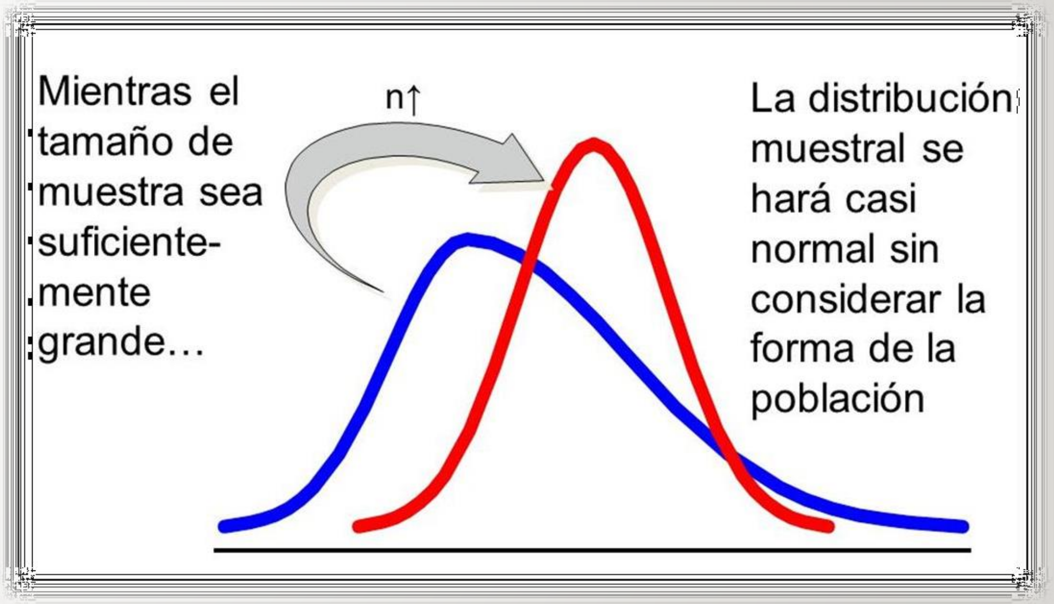
- El teorema central del límite **(TCL)** es una teoría estadística que establece que, dada una **muestra** suficientemente grande de la población, la distribución de las medias muéstrales seguirá una distribución normal.



3. TEOREMA CENTRAL DEL LÍMITE

Además, el **TCL** afirma que a medida que el tamaño de la muestra se incrementa, la media muestral se acercará a la media de la población.

Por tanto, mediante el **TCL** podemos definir la distribución de la media muestral de una determinada población con una varianza conocida. De manera que la distribución seguirá una distribución normal si el tamaño de la muestra es lo suficientemente grande.



3. TEOREMA CENTRAL DEL LÍMITE

Sean: **x1, x2, x3,………, xn** un conjunto de **n** variables aleatorias independientes e igualmente distribuidas, tal que la media **μ** y la varianza finita $\boldsymbol{\sigma^2}$, tienen un valor finito para **i= 1, 2, 3, ….., n**.

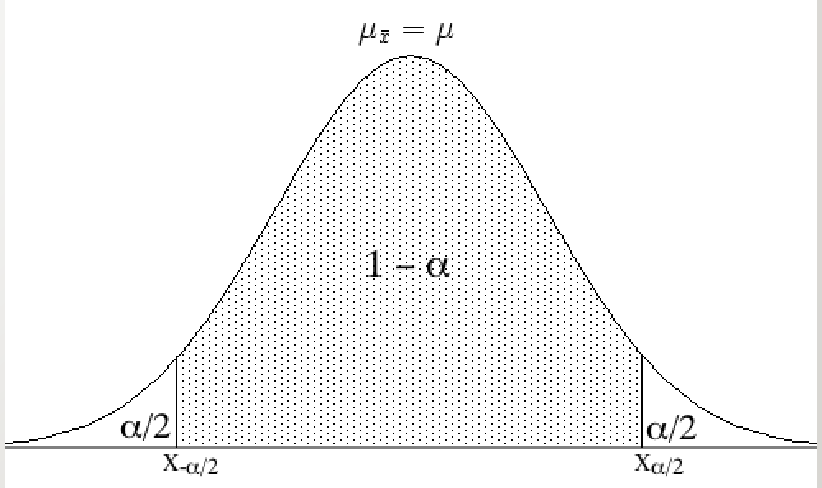
Si:

$$Yn = x1 + x2 + x3 + \ \dots\dots. + xn$$

Con valor esperado y varianza:

$$\boldsymbol{E(Yn)} = n\mu$$

$$\boldsymbol{V(Yn)} = n\sigma^2$$



3. TEOREMA CENTRAL DEL LÍMITE

Entonces la variable aleatoria z, estandarizada de la siguiente manera:

$$\boxed{z = \frac{y - n\mu}{\sigma\sqrt{n}}}$$

Se aproxima a una Distribución Normal con media igual a 0 y varianza igual a 1, siempre y cuando "n" tienda al infinito. Esto significa que la suma de un numero grande $(n \geq 30)$ de variables aleatorias tendrá una Distribución Normal Standard.

3. TEOREMA CENTRAL DEL LÍMITE

En otras palabras, para n grande $(n \geq 30)$, la variable aleatoria es:

$$\boxed{z = \frac{\bar{\ } - \mu_x}{\sigma_x}}$$

- Sin importar el modelo de probabilidad a partir del cual se obtuvo la muestra.
- Donde:

$$\mu_x = Media\ de\ la\ distribución\ muestral\ de\ \bar{\ }$$
$$\sigma_x{}^2 = Varianza\ de\ la\ distribución\ muestral\ de\ \bar{\ }$$
$$\sigma_x = Desviación\ standard\ de\ la\ distribución\ muestral\ de\ \bar{\ }$$

4. DISTRIBUCIÓN DE FUNCIONES DE VARIABLES ALEATORIAS

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso, definido sobre la variable la probabilidad de que dicho suceso ocurra.

La distribución muestral de un estadígrafo (¯) es la distribución de probabilidad que expresa la relación funcional entre cada uno de los valores del estadígrafo y su correspondiente probabilidad.

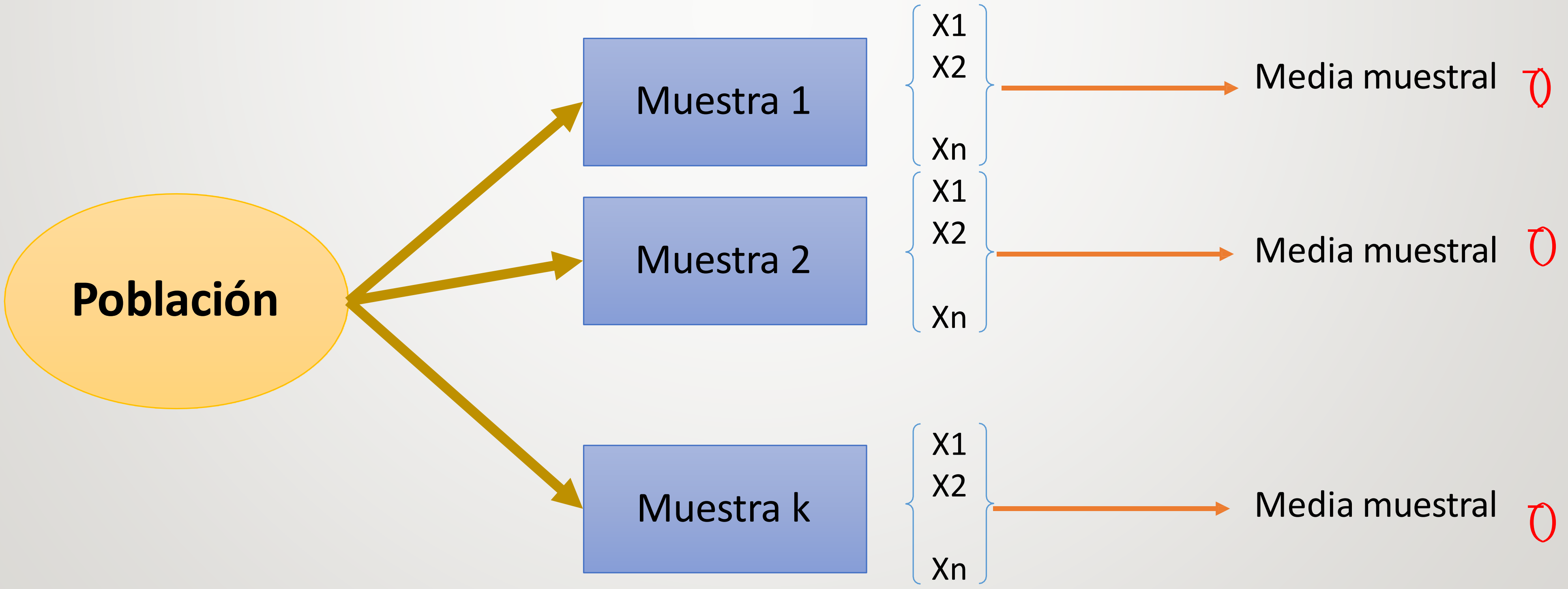
X vs **p(x)**

4. DISTRIBUCIÓN DE FUNCIONES DE VARIABLES ALEATORIAS

Cuando se hablan de:

Parámetros → Población (μ)

Estadígrafos → Muestra (¯)



4. DISTRIBUCIÓN DE FUNCIONES DE VARIABLES ALEATORIAS

4.1. DISTRIBUCIÓN MUESTRAL DE $\bar{X}$ (PROMEDIO MUESTRAL)

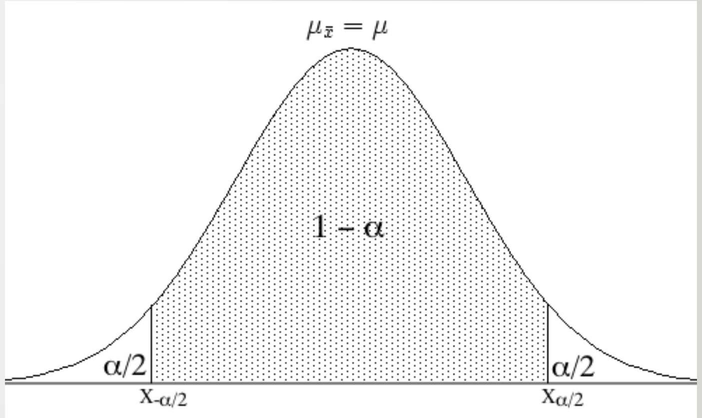
Uno de los estadígrafos más importantes es el promedio de un conjunto de variables aleatorias e independientemente distribuidas, llamado también **promedio** o **media muestral**.

"Este estadígrafo tiene un papel muy importante en problemas de decisiones para medias poblacionales desconocidas."

4. DISTRIBUCIÓN DE FUNCIONES DE VARIABLES ALEATORIAS

Por tanto, si: **x1, x2, x3,………, xn** es una muestra aleatoria de **n** variables aleatorias independientes e igualmente distribuidas, tal que la media **μ** y la varianza finita $\boldsymbol{\sigma^2}$, tienen un valor finito para **i= 1, 2, 3, ….., n**.

La media muestral como:



$$\bar{} = \frac{x1 + x2 + \ldots\ldots + xn}{n} = \sum_{i=1}^{n} \frac{xi}{n}$$

Si se aplica muestreo **con reemplazo**, entonces se cumple que:

$$\boldsymbol{\mu_x = E(x) = \mu}$$

$$\boldsymbol{\sigma_x{}^2 = V(x) = \frac{\sigma^2}{n}}$$

4. DISTRIBUCIÓN DE FUNCIONES DE VARIABLES ALEATORIAS

Para el caso del muestreo **sin reemplazo**, se tiene:

$$\mu = E\left(\;\right) = \mu$$

$$\sigma_x{}^2 = V\left(\;\right) = \frac{(N-n)\,\sigma^2}{(N-1)\,n}$$

En la que:

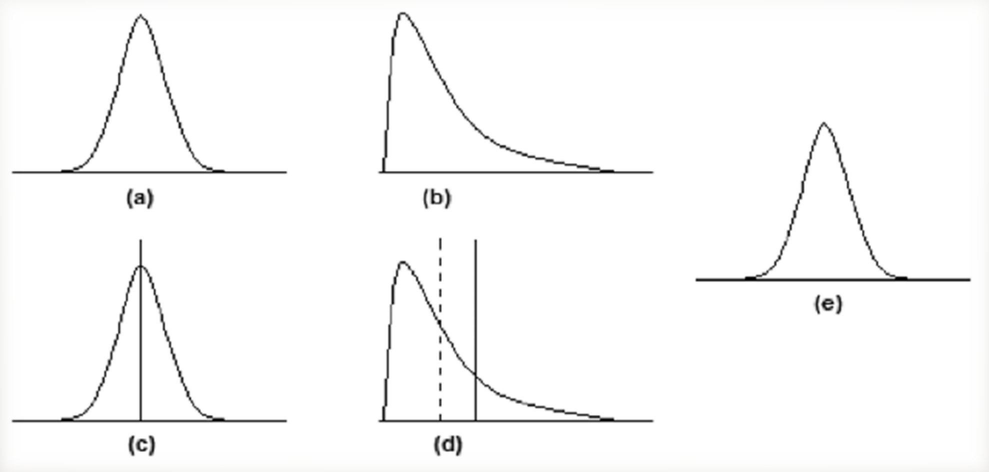
N = Número de elementos de la población

(N-n)/(N-1) = Corrección finita de la población

4. DISTRIBUCIÓN DE FUNCIONES DE VARIABLES ALEATORIAS

En estadística , distribuciones asintóticas , o distribuciones de muestra grande , es un marco para evaluar las propiedades de los estimadores y pruebas estadísticas.

Dentro de este marco, generalmente se supone que el tamaño de la muestra n crece **indefinidamente**; las propiedades de los estimadores y pruebas se evalúan entonces en el límite cuando **n** → **∞** .



5. DISTRIBUCIONES ASINTÓTICAS

5.1. TEOREMA O DESIGUALDAD DE TCHEBYCHEFF

- Si una variable aleatoria "x" tiene una **distribución de probabilidad conocida**, se podrá conocer la **media (μ)** y la **varianza (σ^2)**.
- Y, si se conoce la **media (μ)** y la **varianza (σ^2)**, no se puede determinar la distribución de probabilidad.

Sin embargo, se puede calcular un límite superior o inferior para la probabilidad del tipo:

$$\boxed{p(|x-\mu| < k\sigma)}$$

5. DISTRIBUCIONES ASINTÓTICAS

Si la variable aleatoria x con función de probabilidad f(x) (generalmente desconocida) tiene media y varianza conocidos, entonces para cualquier k>1, se cumple que:

$$\boxed{p_{(|x-\mu| < k\sigma)} \geq 1 - \frac{1}{k^2}}$$

La ecuación indica que la probabilidad de que x tome un **valor dentro del intervalo** $(\mu - k\sigma\ ;\ \mu + k\sigma)$ es por lo menos $1 - \frac{1}{k^2}$

5. DISTRIBUCIONES ASINTÓTICAS

Puesto que $(|x-\mu| \geq k\sigma)$ y $(|x-\mu| < k\sigma)$ son eventos complementarios también se cumple:

$$\boxed{p(|x-\mu| \geq k\sigma) < \frac{1}{k^2}}$$

Lo anterior significa que la probabilidad de que x tome algún valor **fuera del intervalo** $(\mu - k\sigma;\ \mu + k\sigma)$ es a lo más $\frac{1}{k^2}$

5. DISTRIBUCIONES ASINTÓTICAS

5.2. LEY DE LOS GRANDES NÚMEROS

- El teorema de Tchebycheff se aplica a la variable x, pero si este teorema se aplicara a la variable ¯, esta aplicación se denomina **Ley de los grandes números.**

$$x \longrightarrow {}^{-}$$

Teorema de Tchebycheff

Ley de los Grandes Números

Sean: x1, x2, x3,….,xn; **n** variables aleatorias independientes e igualmente distribuidas, tales que la **media** y la **varianza** tienen un valor finito y considerando que $^{-} = \sum_{i=1}^{n} \frac{xi}{n}$ es un buen estimador de media y población.

5. DISTRIBUCIONES ASINTÓTICAS

A partir del Teorema de Tchebycheff para población:

$$p(|x - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Aplicando a la variable aleatoria¯, se tiene:

$$\boxed{p(|\bar{x} - \mu_{\bar{x}}| < k\sigma_{x}) \geq 1 - \frac{1}{k^2}}$$

Puesto que:

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

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

5. DISTRIBUCIONES ASINTÓTICAS

Expresado de otra forma:

$$p\left(\lceil - \mu \mid < \frac{k\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2}$$

Obteniendo el error:

$$\boxed{e = \frac{k\sigma}{\sqrt{n}}}$$

De lo que se deduce:

$$p(\lceil - \mu \mid < e) \geq 1 - \frac{1}{k^2}$$

Por lo tanto:

$$\boxed{n \geq \left(\frac{k}{e}\right)^2}$$

6. IMPLICACIONES EN LA INVESTIGACIÓN

El acto de investigación es una acción individual y colectiva a la vez; este punto intenta dar cuenta de algunas cuestiones relacionadas con el sujeto de la investigación y su relación con la verdad.

EMI

ESCUELA MILITAR DE INGENIERÍA

"Mcal. Antonio José de Sucre"

Prestigio, Disciplina y Mejores Oportunidades

Muchas Gracias!