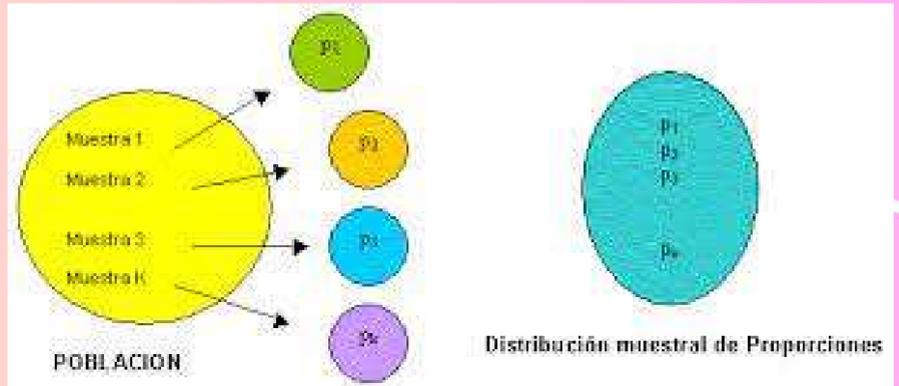
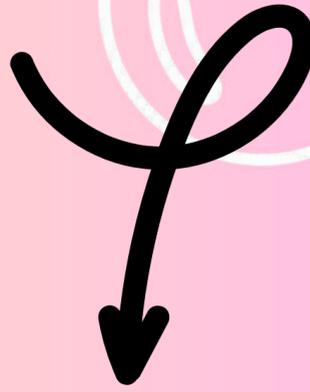




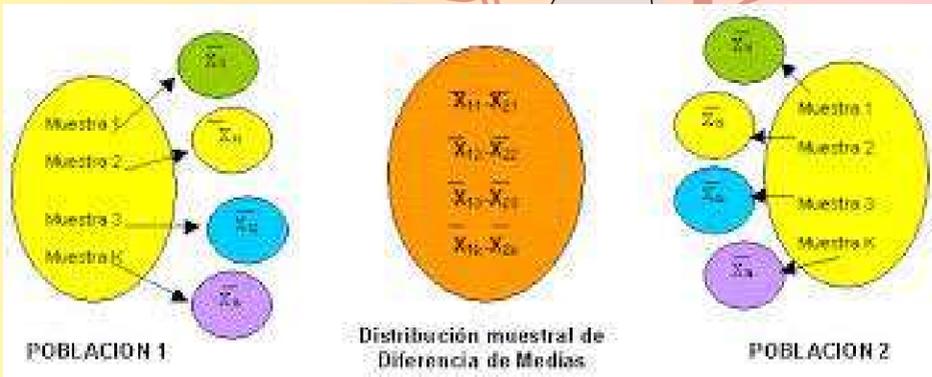
# DISTRIBUCIONES DE MUESTREO



Una distribución de muestral es una distribución de probabilidad de una estadística obtenida de un mayor número de muestras extraídas de una población específica. La distribución muestral de una población dada es la distribución de frecuencias de un rango de resultados diferentes que posiblemente podrían ocurrir para una estadística de una población.



En estadística, una población es el conjunto completo del que se extrae una muestra estadística. Una población puede referirse a un grupo completo de personas, objetos, eventos, visitas al hospital o mediciones. Por tanto, se puede decir que una población es una observación agregada de sujetos agrupados por una característica común.



- Una distribución muestral es una estadística que se obtiene mediante un muestreo repetido de una población más grande.
- Describe un rango de posibles resultados de una estadística, como la media o la moda de alguna variable, ya que realmente existe una población.
- La mayoría de los datos analizados por los investigadores provienen de muestras y no de poblaciones.



## COMPRENSIÓN DE LA DISTRIBUCIÓN DE MUESTRAS

Una gran cantidad de datos extraídos y utilizados por académicos, estadísticos, investigadores, comercializadores, analistas, etc. son en realidad muestras, no poblaciones. Una muestra es un subconjunto de una población. Cada muestra tiene su propia media muestral y la distribución de las medias muestrales se conoce como distribución muestral. Otras estadísticas, como la desviación estándar, la varianza, la proporción y el rango se pueden calcular a partir de datos de muestra. La desviación estándar y la varianza miden la variabilidad de la distribución muestral.

Desarrollo de una distribución de muestreo.

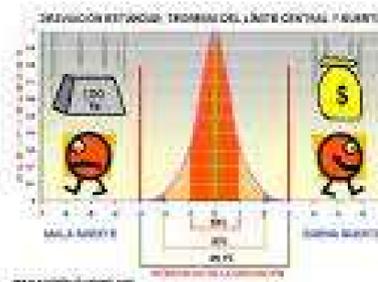
DCOVA

- Supongamos que hay una población...
- Tamaño de la población  $N=4$  A B C
- Variable aleatoria,  $X$ , es la edad de los individuos
- Valores de  $X$ : 18, 20, 22, 24 (años)

# DISTRIBUCIÓN DE MEDIAS MUESTRALES

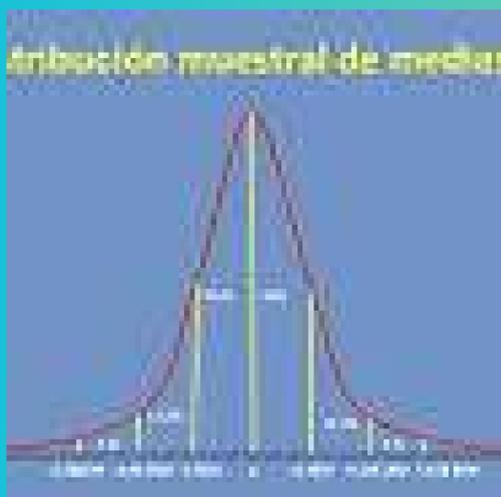
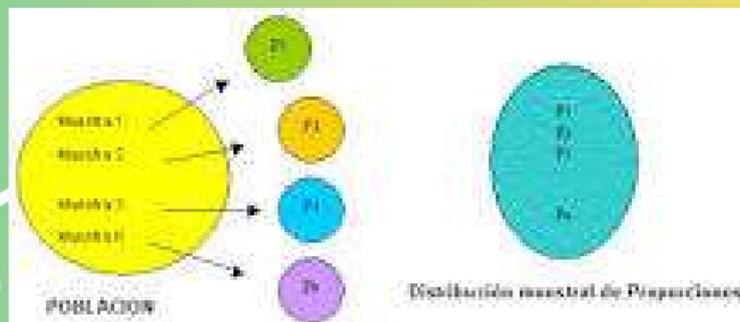
En las últimas entregas hablamos de la distribución de probabilidades de datos individuales. Pero en la práctica cotidiana de la investigación no trabajamos para ubicar en una distribución los datos individuales que ya tenemos; para saber si un valor tiene alta o baja probabilidad de ocurrencia bastaría en este caso con ordenar todos los valores de menor a mayor y ver qué lugar ocupa en la distribución, cuán cercano está a la media o la mediana.

## DISTRIBUCIÓN MUESTRAL DE UNA MEDIA



Cuando hacemos investigación nos interesa inferir si los hallazgos de un grupo de pacientes son similares a los de la población general, o a los de otro grupo, o bien si se trata de valores distintivos. Para inferir si hay o no diferencias es que resulta fundamental trabajar con la distribución muestral de medias.

Cuando en una población se toma una muestra y se mide una variable continua, se obtiene un conjunto de mediciones que puede resumirse en un valor de media. Si se toma otra muestra de la misma medición se obtendrá otra media. Puede intuirse entonces que podemos tomar infinitas muestras y obtener por lo tanto infinitas medias. Esas medias por lo tanto constituyen a su vez una variable continua, que como toda variable continua tiene determinada distribución de probabilidades.



El teorema del límite central nos dice que todas las medias de una variable se distribuyen alrededor de la media de la población: la media de todas las medias es la media poblacional. Notemos que no estamos hablando ahora de datos individuales en torno de la media de una muestra: estamos hablando de medias de muestras en torno de la media poblacional.

Y lo más interesante es que si las muestras son de tamaño suficiente (más de 30 observaciones, aunque algunos hablan de más de 100) las medias se distribuyen alrededor de la media poblacional en forma gaussiana, aún cuando la variable en la población general no tenga esa distribución! En esta curva gaussiana la medida de posición es la media poblacional. Y la medida de dispersión no es el desvío estándar, como en la distribución en una muestra, sino el error estándar de la media (SEM), que es el desvío estándar dividido por la raíz cuadrada del tamaño de la muestra:

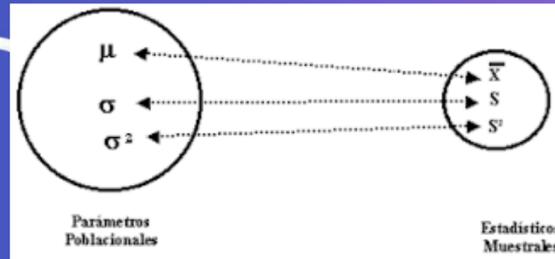
## ¿PARA QUE SIRVE LA DISTRIBUCIÓN MUESTRAL DE MEDIAS?

Entonces, y trabajando con los conceptos de estandarización de la distribución gaussiana y considerando que en una distribución muestral de medias  $z = \frac{\text{media de la muestra} - \text{media de la distribución de medias}}{\text{SEM}}$ , obtenido el valor de  $z$  que corresponde a nuestra media podremos considerar en qué posición se encuentra respecto de la media poblacional, y cuál es la probabilidad de tener valores iguales, menores o mayores.

Así como en una distribución gaussiana de valores individuales entre  $\pm 1$  desvío estándar se engloba aproximadamente el 68% central de las observaciones, entre la media  $\pm 1.96$  desvíos estándar se engloba aproximadamente el 95% central y entre la media  $\pm 2.58$  desvíos estándar se engloban aproximadamente el 99% central de las observaciones, en la distribución muestral de medias entre la media poblacional  $\pm 1$ ,  $\pm 1.96$  y  $\pm 2.58$  SEM se engloban aproximadamente el 68%, el 95% y el 99% central de las medias [figura].

# PARAMETROS MUESTRALES

Un parámetro estadístico o simplemente un estadístico muestral es cualquier valor calculado a partir de la muestra, como por ejemplo la media, varianza o una proporción, que describe a una población y puede ser estimado a partir de una muestra.

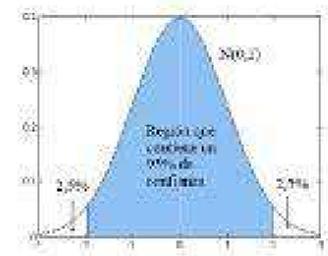


Un estadístico muestral es un tipo de variable aleatoria, y que como tal, tiene una distribución de probabilidad concreta, frecuentemente caracterizada por un conjunto finito de parámetros.

## ESTIMACIÓN ESTADÍSTICA

Una estimación estadística es un proceso mediante el que establecemos qué valor debe tener un parámetro según deducciones que realizamos a partir de estadísticos. En otras palabras, estimar es establecer conclusiones sobre características poblacionales a partir de resultados muestrales.

### Estimación estadística



## ¿QUE ES LA ESTIMACIÓN Y CUAL ES SU IMPORTANCIA?

La estimación es la determinación de un elemento o factor. Esto, usualmente tomando como referencia una base o conjunto de datos. En otras palabras, la estimación es un cálculo que se realiza a partir de la evaluación estadística. Dicho estudio suele efectuarse sobre una muestra y no sobre toda la población objetivo.

## INTERVALOS DE CONFIANZA

Un intervalo de confianza es una técnica de estimación utilizada en inferencia estadística que permite acotar un par o varios pares de valores, dentro de los cuales se encontrará la estimación puntual buscada (con una determinada probabilidad).



el intervalo de confianza no sirve para dar una estimación puntual del parámetro poblacional, si nos va a servir para hacernos una idea aproximada de cuál podría ser el verdadero de este. Nos permite acotar entre dos valores en dónde se encontrará la media de la población.

$$IC_{100(1-\alpha)\%}(\mu) = \left[ \bar{X} - t_{1-\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right]$$

Intervalos de confianza

## FACTORES DE LOS QUE DEPENDE UN INTERVALO DE CONFIANZA

- **Tamaño de la muestra seleccionada:** Dependiendo de la cantidad de datos que se hayan utilizado para calcular el valor muestral, este se acercará más o menos al verdadero parámetro poblacional.
- **Nivel de confianza:** Nos va a informar en qué porcentaje de casos nuestra estimación acierta. Los niveles habituales son el 95% y el 99%.
- **Margen de error de nuestra estimación:** Este se denomina como alfa y nos informa de la probabilidad que existe de que el valor poblacional esté fuera de nuestro intervalo.
- **Lo estimado en la muestra (media, varianza, diferencia de medias...):** De esto va a depender el estadístico pivote para el cálculo del intervalo.

## EJEMPLO DE INTERVALO DE CONFIANZA PARA LA MEDIA, ASUMIENDO NORMALIDAD Y CONOCIDA LA DESVIACIÓN TÍPICA

- $\bar{X}$  = Media muestral
- $\mu$  = Media poblacional
- $\sigma$  = Desviación típica (conocida)
- $\sqrt{n}$  = Raíz cuadrada del tamaño muestral

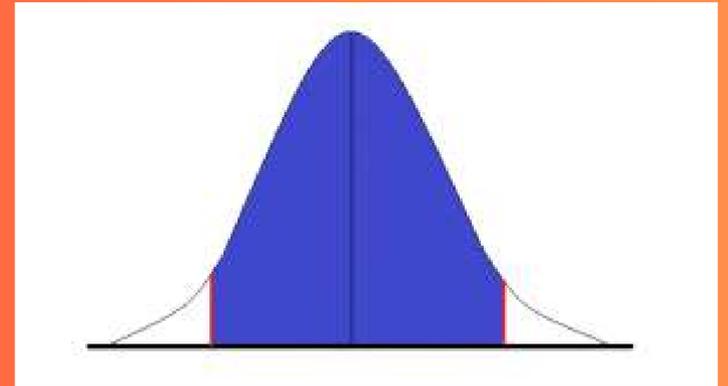
$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

El intervalo resultante sería el siguiente

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

Para obtener el intervalo, no tendríamos más que sustituir los datos en la fórmula del intervalo.

$$\left(4 - \frac{0,55}{\sqrt{10}} * 1,96 < \mu < 4 + \frac{0,55}{\sqrt{10}} * 1,96\right)$$
$$(3,7 < \mu < 4,3)$$



El intervalo de confianza, sería la parte de la distribución que queda sombreada en azul. Los 2 valores acotados por este serían los correspondientes a las 2 líneas de color rojo. La línea central que parte la distribución en 2 sería el verdadero valor poblacional.

Es importante resaltar que en este caso, dado que la función de densidad de la distribución  $N(0,1)$  nos da la probabilidad acumulada (desde la izquierda hasta el valor crítico), tenemos que encontrar el valor que nos deja a la izquierda 0,975% [este es 1,96].

## ERROR ADMITIDO Y TAMAÑO DE LA MUESTRA

Un error de muestreo se produce cuando la muestra utilizada en el estudio no es representativa de toda la población. A menudo se producen errores de muestreo y, por lo tanto, los investigadores siempre calculan un margen de error durante los resultados finales como práctica estadística.

Los errores de muestreo se producen debido a una disparidad en la representatividad de los encuestados. Ocurre principalmente cuando el investigador no planifica su muestra cuidadosamente.

El tamaño de la muestra para que en el intervalo de confianza se cometa un error prefijado al nivel de significación  $\alpha$  es:

$$n = \left[ \frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{E} \right]^2$$

## MUESTREO NO PROBABILÍSTICO

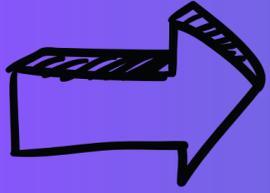
EL MUESTREO NO PROBABILÍSTICO ES UNA TÉCNICA DE MUESTREO EN LA CUAL EL INVESTIGADOR SELECCIONA MUESTRAS BASADAS EN UN JUICIO SUBJETIVO EN LUGAR DE HACER LA SELECCIÓN AL AZAR.



A DIFERENCIA EN EL MUESTREO PROBABILÍSTICO, DONDE CADA MIEMBRO DE LA POBLACIÓN TIENE UNA POSIBILIDAD CONOCIDA DE SER SELECCIONADO, EN EL MUESTREO NO PROBABILÍSTICO, NO TODOS LOS MIEMBROS DE LA POBLACIÓN TIENEN LA OPORTUNIDAD DE PARTICIPAR EN EL ESTUDIO

EL MUESTREO NO PROBABILÍSTICO ES MÁS ÚTIL PARA ESTUDIOS EXPLORATORIOS COMO LA ENCUESTA PILOTO (UNA ENCUESTA QUE SE IMPLEMENTA EN UNA MUESTRA MÁS PEQUEÑA, EN COMPARACIÓN CON EL TAMAÑO DE MUESTRA PREDETERMINADO).

EL MUESTREO NO PROBABILÍSTICO ES UN MÉTODO MENOS ESTRICTO, ESTE MÉTODO DE MUESTREO DEPENDE EN GRAN MEDIDA DE LA EXPERIENCIA DE LOS INVESTIGADORES. EL MUESTREO NO PROBABILÍSTICO COMÚNMENTE SE LLEVA A CABO MEDIANTE MÉTODOS DE OBSERVACIÓN, Y SE UTILIZA AMPLIAMENTE EN LA INVESTIGACIÓN CUALITATIVA.



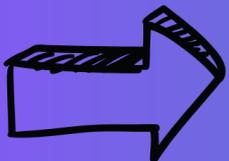
## TIPOS DE MUESTREO NO PROBABILÍSTICO

- Muestreo por conveniencia
- Muestreo consecutivo
- Muestreo por cuotas
- Muestreo intencional o por juicio
- Muestreo de bola de nieve



### ¿CUANDO USAR EL MUESTREO NO PROBABILÍSTICO?

- Este tipo de muestreo se usa para indicar si existe un rasgo o característica particular en una población.
- Esta técnica de muestreo es ampliamente utilizada cuando los investigadores realizan investigaciones cualitativas, estudios piloto o investigación exploratoria.
- El muestreo no probabilístico se usa cuando los investigadores tienen un tiempo limitado para llevar a cabo la investigación o tienen limitaciones presupuestarias.
- El muestreo no probabilístico se realiza para observar si un tema en particular necesita un análisis en profundidad. Mejora tu muestreo siguiendo estos consejos



## GRÁFICO O DIAGRAMA DE CONTROL

UN GRÁFICO DE CONTROL ES UNA HERRAMIENTA UTILIZADA PARA DISTINGUIR LAS VARIACIONES DEBIDAS A CAUSAS ASIGNABLES O ESPECIALES A PARTIR DE LAS VARIACIONES ALEATORIAS INHERENTES AL PROCESO.

LAS VARIACIONES ALEATORIAS SE REPITEN CASUALMENTE DENTRO DE LOS LÍMITES PREDECIBLES.

LAS VARIACIONES DEBIDAS A CAUSAS ASIGNABLES O ESPECIALES INDICAN QUE ES NECESARIO IDENTIFICAR, INVESTIGAR Y PONER BAJO CONTROL ALGUNOS FACTORES QUE AFECTAN AL PROCESO.

LA CONSTRUCCIÓN DE GRÁFICOS DE CONTROL ESTÁ BASADA EN LA ESTADÍSTICA MATEMÁTICA. LOS GRÁFICOS DE CONTROL EMPLEAN DATOS DE OPERACIÓN PARA ESTABLECER LÍMITES DENTRO DE LOS CUALES SE ESPERA HACER OBSERVACIONES FUTURAS, SI EL PROCESO DEMUESTRA NO HABER SIDO AFECTADO POR CAUSAS ASIGNABLES O ESPECIALES.



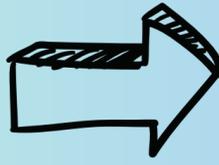
LÍMITE SUPERIOR DE CONTROL (LSC): ES EL MAYOR VALOR ACEPTADO EN EL PROCESO.

LÍMITE INFERIOR DE CONTROL (LIC): ES EL VALOR MÁS PEQUEÑO QUE SE ACEPTA EN EL PROCESO.

LÍMITE CENTRAL DE CONTROL (LC): ES LA LÍNEA CENTRAL DEL GRÁFICO. MIENTRAS MÁS CERCA ESTÉN LOS PUNTOS A LA LÍNEA, MÁS ESTABLE ES EL PROCESO.

## CAUSAS ASIGNABLES

FACTORES (GENERALMENTE NUMEROSOS, PERO INDIVIDUALMENTE DE RELATIVA IMPORTANCIA) QUE SE PUEDEN DETECTAR E IDENTIFICAR COMO CAUSANTES DE UN CAMBIO EN UNA CARACTERÍSTICA DE LA CALIDAD O NIVEL DEL PROCESO.



## CAUSAS ALEATORIAS

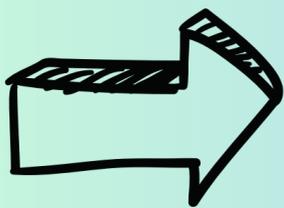
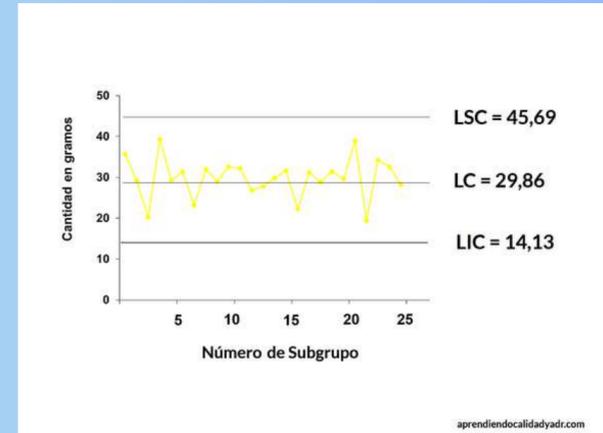
FACTORES GENERALMENTE NUMEROSOS, PERO POCO IMPORTANTES, QUE CONTRIBUYEN A LA VARIACIÓN Y NO HAN SIDO NECESARIAMENTE IDENTIFICADOS.

¿PARA QUÉ SIRVE UN GRÁFICO O DIAGRAMA DE CONTROL?

**Diagnóstico:** Para evaluar la estabilidad de un proceso.

**Control:** Para determinar cuándo es necesario ajustar un proceso y cuándo se debe dejar tal y como está.

**Confirmación:** Para confirmar la mejora de un proceso.



## TIPOS DE GRÁFICOS DE CONTROL

- Gráfico de control por variables
- Gráfico de control por atributos



- SELECCIONAR LA CARACTERÍSTICA OBJETO DE ANÁLISIS EN EL GRÁFICO DE CONTROL.
- SELECCIONAR EL TIPO APROPIADO DE GRÁFICO DE CONTROL.
- DECIDIR EL SUBGRUPO (UNA PEQUEÑA RECOPIACIÓN DE ARTÍCULOS, EN EL MARCO DE LOS CUALES LAS VARIACIONES SE DEBEN PROBABLE Y ÚNICAMENTE AL AZAR), SUS DIMENSIONES, Y LA FRECUENCIA DE MUESTREO DEL SUBGRUPO.
- RECOLECTAR Y REGISTRAR DATOS SOBRE 20 Ó 25 SUBGRUPOS POR LO MENOS, O UTILIZAR DATOS REGISTRADOS PREVIAMENTE.
- CALCULAR ESTADÍSTICAMENTE LAS CARACTERÍSTICAS DE CADA MUESTRA DEL SUBGRUPO.
- CALCULAR LOS LÍMITES DE CONTROL SOBRE LA BASE DE LAS ESTADÍSTICAS DE LAS MUESTRAS DE SUBGRUPOS.
- CONSTRUIR UN GRÁFICO Y PLOTEAR LAS ESTADÍSTICAS DEL SUBGRUPO.
- EXAMINAR EL PLOTEO POR SI HAY PUNTOS FUERA DE LOS LÍMITES DE CONTROL Y PATRONES QUE INDIQUEN LA PRESENCIA DE CAUSAS ASIGNABLES O ESPECIALES.
- DECIDIR LAS ACCIONES A TOMAR EN EL FUTURO.