

Inferencia Estadística

Población y muestra

Conceptos y definiciones

Población:

Es la colección de toda la posible información que caracteriza a un fenómeno. Es cualquier colección ya sea de un número finito de mediciones o una colección grande, virtualmente infinita de datos acerca de algo de interés.

Muestra:

Es un subconjunto **representativo** seleccionado de una población. Una buena muestra es aquella que refleja las características esenciales de la población de la cual se obtuvo.

Representativo:

En las técnicas de muestreo, la finalidad es garantizar que cada elemento de la población tenga una oportunidad igual e independiente de ser incluida en la muestra, tales procesos de muestreo conducen a lo que llamaremos una muestra aleatoria

Muestra aleatoria:

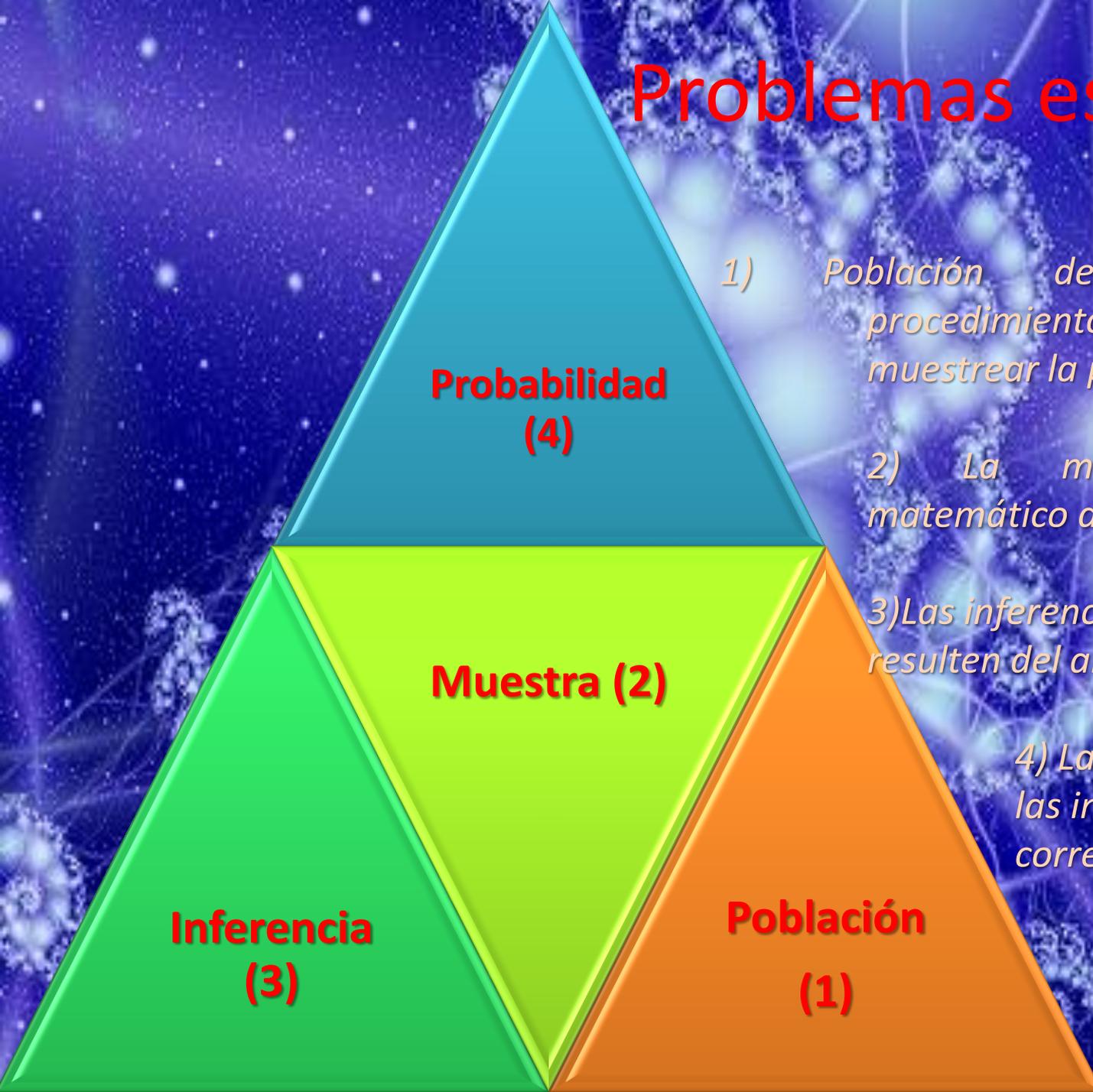
Se usan para calcular ciertas características de la muestra. Estas características son estadísticas, las cuales nos permitirán hacer la inferencia a cerca de las características de la población, y estas reciben el nombre de parámetros

Generalidades de la estadística inferencial

En estadística, la inferencia es de carácter *inductivo*, dado que se proyecta de lo específico a lo general (muestra)-> (población). Debido a esta naturaleza inductiva, siempre existe la posibilidad de error, lo cual nos indica que es imposible tener el 100% de seguridad sobre una proposición que esté basada en la inferencia estadística. Sin embargo, existe una medida de la confiabilidad de la estadística, esta confiabilidad se mide en ~~términos de probabilidad~~.

(In other words) Para cada inferencia estadística se estima su probabilidad de éxito o fracaso (sea correcta o incorrecta)

Problemas estadísticos



1) *Población de interés y el procedimiento empleado para muestrear la población.*

2) *La muestra y análisis matemático de la información.*

3) *Las inferencias estadísticas que resulten del análisis de la muestra*

4) *La probabilidad de que las inferencias sean correctas*

Dos enfoques

Clásico:

Descansa únicamente en la evidencia muestral. Este enfoque también es desarrollado como una teoría del muestreo.

Bayesiano:

Utiliza la evidencia estadística y la combina con otro tipo de información, generalmente proporcionada por el investigador del problema. Esta información descansa fundamentalmente en la creencia o convicción de quien realiza la inferencia estadística, respecto a las incertidumbres y errores estadísticos

Muestra Aleatoria

Existen diferentes métodos para investigar u observar una población (observación exhaustiva o censo, sub-población, muestra y observación mixta), Nos referimos a la observación parcial mediante una muestra y diremos que se ha investigado la población a partir de una muestra cuando los elementos que componen la muestra no reúnen ninguna característica esencial que los diferencie de los restantes, representando a toda la población. La muestra debe ser representativa de toda la población y, por tanto, tendrá características similares a las que se observarían en la población entera.

La razón principal para investigar una muestra en lugar de la población completa es que al levantar (recolectar) información para toda la población daría lugar a un costo muy elevado, tanto en recursos económicos como en tiempo. Incluso para ciertos casos, donde los recursos fuesen suficientes para investigar la población completa, es preferible investigar sólo una muestra representativa.

Ejemplo:

Supongamos que como director comercial deseas conocer la opinión sobre un nuevo producto de limpieza que sacarás al mercado. No sería correcto que limitaras la correspondiente encuesta a sus amigos y a y vecinos, pues tales personas no reflejarían la opinión de toda la población ya que la muestra no sería representativa de toda la población, ni tampoco sería aleatoria. Para evitar estos problemas y poder realizar una inferencia correcta sobre la población a partir de una muestra, es necesario que se verifique representatividad y la aleatoriedad de la muestra.



El objetivo del muestreo es seleccionar una muestra que garantice con un costo razonable una buena representatividad de la población.

El procedimiento de selección de la muestra conduce a diferentes tipos de muestreo, en poblaciones finitas. Existen dos tipos de Muestreo



Con reemplazamiento



Consiste en seleccionar, por mecanismos aleatorios, los elementos de la población que entran a formar parte de la muestra, pero de tal manera que cuando se observa la característica que estamos investigando, del primer elemento seleccionado, se devuelve el elemento a la población, se selecciona un segundo elemento entre todos los elementos de la población, se anota la característica que se está investigando y se devuelve nuevamente a la población, y así sucesivamente. Este procedimiento permite que un elemento de la población pueda ser seleccionado en más de una ocasión para formar parte de una muestra, pues la selección se realiza con reemplazamiento es decir, con devolución del elemento seleccionado a la población.

sin reemplazamiento



Los elementos de la población que entran a formar parte de la muestra también se seleccionan aleatoriamente, pero después de observar la característica que estamos investigando no se devuelve el elemento de nuevo a la población, con lo cual no pueden volver a ser seleccionados como ocurría en el muestreo con reemplazamiento.

Muestreo

Entonces, si tenemos una población de N elementos y queremos seleccionar una muestra de tamaño n resulta que la probabilidad de que un elemento de la población sea seleccionado en la primera extracción para formar parte de la muestra será en ambos tipos de muestreo.

$$P(n = 1) = \frac{1}{N}$$

Sin embargo, en la selección del segundo elemento las probabilidades son diferentes, pues en el muestreo con reemplazamiento se mantiene la probabilidad igual, mientras que para el muestreo sin reemplazamiento la probabilidad cambia

Con reemplazamiento

$$P(n = 2) = \frac{1}{N}$$

Sin reemplazamiento

$$P(n = 2) = \frac{1}{N-1}$$

Vemos pues que en el muestreo con reemplazamiento la probabilidad de seleccionar uno a uno los n elementos de la muestra permanece constante y en el muestreo sin reemplazamiento no sucede lo mismo ya que en cada extracción no se devuelve el elemento a la población y esta va disminuyendo a medida que se selecciona la muestra, siendo los tamaños poblacionales $N, N-1, N-2, \dots, N-(n-1)$.

Observación importante

Si el tamaño de la población es infinito o muy grande, entonces el tamaño de la muestra n en comparación con ese tamaño N infinito o muy grande de la población es prácticamente despreciable, y entonces no existe diferencia significativa entre ambos tipos de muestreo.

No sucede lo mismo cuando el tamaño N de la población es finito, dando lugar a tratamientos diferentes.

Muestra Aleatoria Simple (MAS)

Consideremos una población, cuya **función de distribución** esta dada por $F(x)$, la cual está constituida por un **número infinito** de posibles valores, de una cierta característica medible X .

Esta característica puede ser, por ejemplo:

- el tiempo de espera para recibir un servicio.
- el valor de las ventas de un determinado producto.

Entonces para seleccionar una muestra aleatoria de tamaño n de esta población se diseña un experimento. De tal manera que la primera realización de ese experimento nos proporciona la observación X_1 , de la característica medible X , **repitiendo** sucesivamente el experimento bajo las mismas condiciones, para todos los factores controlables, tendremos las n observaciones:

$$X_1, X_2, \dots, X_n$$

Elementos
de la
Muestra
Aleatoria
Simple

Muestra Aleatoria Simple (MAS)

Cada observación X_i correspondiente a la repetición i -ésima del experimento es una variable aleatoria cuya distribución de probabilidad es idéntica a la de la población de característica X , para todo $i = 1, 2, \dots, n$.

Si la población consta de un número finito de elementos, por ejemplo, personas, viviendas, establecimientos comerciales, etc., y realizamos un **muestreo aleatorio con reemplazamiento**, reiterando el proceso n veces tendríamos las n observaciones:

$$X_1, X_2, \dots, X_n$$

de la característica medible X de la población, que constituyen la muestra aleatoria simple.

Cada una de estas observaciones, X_1, X_2, \dots, X_n , también es una variable aleatoria cuya **función de probabilidad es idéntica** a la de la población, ya que cada selección de un elemento que da lugar a una observación procede de la población original.

Variables aleatorias,
independientes y
uniformemente
distribuidas

Muestra Aleatoria Simple (MAS)

Si en la población con un número finito de elementos, se seleccionan análogamente n elementos sin reemplazamiento, tendríamos una muestra aleatoria sin reemplazamiento de observaciones de la característica X que estamos investigando

$$X_1, X_2, \dots, X_n$$

Definición de MAS

Una muestra aleatoria simple de tamaño n de una población X está constituida por un conjunto de n variables aleatorias X_1, X_2, \dots, X_n independientes y uniformemente distribuidas a la población X , es decir, está constituida por un conjunto de observaciones muestrales independientes e idénticamente distribuidas.

Cuando el experimento se realiza, a cada una de las variables aleatorias se le asignará un valor numérico.

$$X_1 = x_1 \cdots X_n = x_n$$

Es decir, tendremos la realización de la muestra y diremos que ha sido seleccionada una muestra.

VARIABLES aleatorias,
no independientes y
uniformemente
distribuidas