

## UNIDAD I

### I.1.- Estadística descriptiva:

Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos. Se puede definirse como aquel método que contiene la recolección, organización, presentación y resumen de una serie de datos. El mencionado resumen puede ser tabular, gráfico o numérico. El análisis que se realiza se limita en sí mismo a los datos recolectados y no se puede realizar inferencia alguna o generalizaciones algunas, acerca de la población de donde provienen esos datos estadísticos.

Una de las ramas de la Estadística más accesible a la mayoría de la población es la Descriptiva. Esta se dedica única y exclusivamente al ordenamiento y tratamiento mecánico de la información para su presentación por medio de tablas y de representaciones gráficas, así como de la obtención de algunos parámetros útiles para la explicación de la información.

La estadística descriptiva analiza, estudia y describe a la totalidad de los individuos de una población, su finalidad es obtener información, analizarla, elaborarla y simplificarla lo necesario para que pueda ser interpretada cómoda y rápidamente y, por tanto, pueda utilizarse eficazmente para el fin que se desee.

El proceso que sigue la estadística descriptiva para el estudio de una cierta población consta de los siguientes pasos:

- 1 Selección de caracteres dignos de ser estudiados.
- 2 Mediante encuesta o medición, obtención del valor de cada individuo en los caracteres seleccionados.
- 3 Elaboración de tablas de frecuencias, mediante la adecuada clasificación de los individuos dentro de cada carácter.
- 4 Representación gráfica de los resultados (elaboración de gráficos estadísticos).
- 5 Obtención de parámetros estadísticos, números que sintetizan los aspectos más relevantes de una distribución estadística.

## Estadística inferencial:

Es aquella rama de la estadística que apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos. Puede definirse como aquella rama de la estadística que hace posible la estimación de una característica de una población o la toma de una decisión referente a una población, fundamentándose sólo en los resultados de la muestra.

La estadística Inferencial, por otro lado, se refiere a la rama de la estadística que trata de los procesos inferenciales, la que a su vez vislumbra la teoría de estimación y prueba de hipótesis. Uno de los primordiales aspectos de la inferencia estadística es el proceso que radica en utilizar estadísticos muestrales para adquirir conclusiones sobre los verdaderos parámetros de la población.

Los requerimientos de los métodos de la inferencia estadística se originan de la necesidad del muestreo. Al tornarse muy grande una población, comúnmente resulta demasiado costoso, prolongado en el tiempo y complicado obtener información de la población completa. Las decisiones con respecto a las características de la población se deben basar en la información contenida en una muestra de esa población. La teoría de la probabilidad suministra el vínculo, determinando la probabilidad de que los resultados provenientes de la muestra reflejen los resultados que se obtendrían de la población.

La fidelidad de cualquier estimación tiene una importancia enorme. Esta precisión depende en gran parte de la forma de tomar la muestra y de la atención que se ponga en que esta muestra suministre una imagen fiable de la población, pero casi nunca la muestra representa la población en toda su plenitud, y de ello resultará un error muestral.

### 1.2.- Finalidad de la estadística

La estadística es una ciencia o método científico que en la actualidad es considerada como un poderoso auxiliar en las investigaciones científicas, que le permite a ésta aprovechar el material cuantitativo.

## Historia de la estadística

Desde el inicio de la civilización han existido formas sencillas de estadística, puesto que en la antigüedad se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o ciertas cosas que eran de importancia en aquellas civilizaciones. El término estadístico es ampliamente percibido y pronunciado a diario desde diversos sectores activos de la sociedad. No obstante, hay una gran diferencia entre el sentido del término cuando se utiliza en el lenguaje corriente, generalmente al anteceder una citación de carácter numérico, y lo que la estadística significa como ciencia.

La razón o razones que motivaron al hombre en un momento de su desarrollo a tomar en cuenta datos con propósitos estadísticos, posiblemente se encuentra si se toma en cuenta que es difícil suponer un organismo social, sea cual fuere la época, sin la necesidad, casi instintiva, de recoger aquellos hechos que aparecen como actos esenciales de la vida; y así, al ubicarnos en una etapa del desarrollo de la estadística podemos especular que se convirtió en una aritmética estatal para asistir al gobernante que necesitaba conocer la riqueza y el número de los súbditos entre otros, con el objeto de recaudar impuestos o presupuestar la guerra.

Desde los comienzos de la civilización han existido formas sencillas de estadística, pues ya se utilizaban representaciones gráficas y otros símbolos en pieles, rocas, palos de madera y paredes de cuevas para contar el número de personas, animales o cosas. Hacia el año 3000 a.C. los babilonios usaban pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y sobre las especies vendidas o cambiadas mediante trueque.

### **I.3.- conceptos básicos**

Universo:

En estadística es el nombre específico que recibe particularmente en la investigación social la operación dentro de la delimitación del campo de investigación que tienen por objeto la determinación del conjunto de unidades de observaciones del conjunto de unidades de observación que van a ser investigadas. Para muchos investigadores él

termino universo y población son sinónima. En general, el universo es la totalidad de elementos o características que conforman el ámbito de un estudio o investigación.

**Población:**

En estadística el concepto de población va más allá de lo que comúnmente se conoce como tal. En términos estadísticos, población es un conjunto finito o infinito de personas, animales o cosas que presentan características comunes, sobre los cuales se quiere efectuar un estudio determinado. En otras palabras, la población se define como la totalidad de los valores posibles (mediciones o conteos) de una característica particular de un grupo especificado de personas, animales o cosas que se desean estudiar en un momento determinado. Así, se puede hablar de la población de habitantes de un país, de la población de estudiantes universitarios de la zona sur del Estado Anzoátegui, de la población de casas de la Urbanización Los Ríos de la ciudad de El Tigre, el rendimiento académico de los estudiantes del IUTJAA, el número de carros marca Corola de la ciudad de El Tigre, la estatura de un grupo alumnos del IUTJAA, la talla, etc.

**Muestra:**

La muestra es un subconjunto de la población, seleccionado de tal forma, que sea representativo de la población en estudio, obteniéndose con el fin de investigar alguna o algunas de las propiedades de la población de la cual procede. En otras palabras es una parte de la población que sirve para representarla. Según el DRAE, es una parte o porción extraída de un conjunto por métodos que permiten considerarla como representativa del mismo. Entonces, una muestra no es más que una parte de la población que sirve para representarla. La muestra debe obtenerse de la población que se desea estudiar; una muestra debe ser definida sobre la base de la población determinada, y las conclusiones que se obtengan de dicha muestra sólo podrán referirse a la población en referencia.

**Muestreo:**

Es el procedimiento mediante el cual se obtiene una o más muestras de una población determinada. Existen dos tipos de muestreos a saber:

**Los Parámetros:**

Son cualquiera característica que se pueda medir y cuya medición se lleve a cabo sobre todos los elementos que integran una población determinada, los mismos suelen

representarse con letras griegas. El valor de un parámetro poblacional es un valor fijo en un momento dado. Ejemplo: La media Aritmética =  $\mu$  (miu), La desviación Típica =  $\sigma$ , (Sigma) etcétera.

Dato estadístico:

Es un conjunto de valores numéricos que tienen relación significativa entre sí. Los mismos pueden ser comparados, analizados e interpretados en una investigación cualquiera. Se puede afirmar que son las expresiones numéricas obtenidas como consecuencia de observar un individuo de la población; por lo tanto, son las características que se han tomado en cuenta de cualquiera población para una investigación determinada.

Frecuencia:

La frecuencia es el número de veces que se repite (aparece) el mismo dato estadístico en un conjunto de observaciones de una investigación determinada, las frecuencias se les designan con las letras  $f_i$ , y por lo general se les llaman frecuencias absolutas.

Distribución de Frecuencia:

En estadística existe una relación con cantidades, números agrupados o no, los cuales poseen entre sí características similares. Existen investigaciones relacionadas con los precios de los productos de la dieta diaria, la estatura y el peso de un grupo de individuos, los salarios de los empleados, los grados de temperatura del medio ambiente, las calificaciones de los estudiantes, etc., que pueden adquirir diferentes valores gracias a una unidad apropiada, que recibe el nombre de variable. La representación numérica de las variables se denomina dato estadístico. La distribución de frecuencia es una disposición tabular de datos estadísticos, ordenados ascendente o descendentemente, con la frecuencia ( $f_i$ ) de cada dato. Las distribuciones de frecuencias pueden ser para datos no agrupados y para datos agrupados o de intervalos de clase.

Distribución de frecuencia para datos no Agrupados:

Es aquella distribución que indica las frecuencias con que aparecen los datos estadísticos, desde el menor de ellos hasta el mayor de ese conjunto sin que se haya hecho ninguna modificación al tamaño de las unidades originales. En estas distribuciones cada dato mantiene su propia identidad después que la distribución de frecuencia se ha elaborado.

En estas distribuciones los valores de cada variable han sido solamente reagrupados, siguiendo un orden lógico con sus respectivas frecuencias.

Distribución de frecuencia de clase o de datos Agrupados:

Es aquella distribución en la que las disposiciones tabulares de los datos estadísticos se encuentran ordenados en clases y con la frecuencia de cada clase; es decir, los datos originales de varios valores adyacentes del conjunto se combinan para formar un intervalo de clase. No existen normas establecidas para determinar cuándo es apropiado utilizar datos agrupados o datos no agrupados; sin embargo, se sugiere que cuando el número total de datos (N) es igual o superior 50 y además el rango o recorrido de la serie de datos es mayor de 20, entonces, se utilizará la distribución de frecuencia para datos agrupados, también se utilizará este tipo de distribución cuando se requiera elaborar gráficos lineales como el histograma, el polígono de frecuencia o la ojiva.

La razón fundamental para utilizar la distribución de frecuencia de clases es proporcionar mejor comunicación acerca del patrón establecido en los datos y facilitar la manipulación de los mismos. Los datos se agrupan en clases con el fin de sintetizar, resumir, condensar o hacer que la información obtenida de una investigación sea manejable con mayor facilidad.

Componentes de una distribución de frecuencia de clase

1.- Rango o Amplitud total (recorrido).- Es el límite dentro del cual están comprendidos todos los valores de la serie de datos, en otras palabras, es el número de diferentes valores que toma la variable en un estudio o investigación dada. Es la diferencia entre el valor máximo de una variable y el valor mínimo que ésta toma en una investigación cualquiera. El rango es el tamaño del intervalo en el cual se ubican todos los valores que pueden tomar los diferentes datos de la serie de valores, desde el menor de ellos hasta el valor mayor estando incluidos ambos extremos. El rango de una distribución de frecuencia se designa con la letra R.

2.- Clase o Intervalo de clase.- Son divisiones o categorías en las cuales se agrupan un conjunto de datos ordenados con características comunes. En otras palabras, son fraccionamientos del rango o recorrido de la serie de valores para reunir los datos que presentan valores comprendidos entre dos límites. Para organizar los valores de la serie de datos hay que determinar un número de clases que sea conveniente. En otras palabras,

que ese número de intervalos no origine un número pequeño de clases ni muy grande. Un número de clases pequeño puede ocultar la naturaleza natural de los valores y un número muy alto puede provocar demasiados detalles como para observar alguna información de gran utilidad en la investigación.

#### Tamaño de los Intervalos de Clase

Los intervalos de clase pueden ser de tres tipos, según el tamaño que estos presenten en una distribución de frecuencia:

a) Clases de igual tamaño, b) clases desiguales de tamaño y c) clases abiertas.

#### 3.- Amplitud de Clase, Longitud o Ancho de una Clase

La amplitud o longitud de una clase es el número de valores o variables que concurren a una clase determinada. La amplitud de clase se designa con las letras  $l_c$ . Existen diversos criterios para determinar la amplitud de clases, ante esa diversidad de criterios, se ha considerado que lo más importante es dar un ancho o longitud de clase a todos los intervalos de tal manera que respondan a la naturaleza de los datos y al objetivo que se persigue y esto se logra con la práctica.

#### 4.-Punto medio o Marca de clase

El centro de la clase, es el valor de los datos que se ubica en la posición central de la clase y representa todos los demás valores de esa clase. Este valor se utiliza para el cálculo de la media aritmética.

#### 5.-Frecuencia de clase

La frecuencia de clase se le denomina frecuencia absoluta y se le designa con las letras  $f_i$ . Es el número total de valores de las variables que se encuentran presente en una clase determinada, de una distribución de frecuencia de clase.

#### 6.- Frecuencia Relativa

La frecuencia relativa es aquella que resulta de dividir cada uno de los  $f_i$  de las clases de una distribución de frecuencia de clase entre el número total de datos ( $N$ ) de la serie de valores. Estas frecuencias se designan con las letras  $fr$ ; si cada  $fr$  se multiplica por 100 se obtiene la frecuencia relativa porcentual ( $fr \%$ ).

## 7.-Frecuencias acumuladas

Las frecuencias acumuladas de una distribución de frecuencias son aquellas que se obtienen de las sumas sucesivas de las  $f_i$  que integran cada una de las clases de una distribución de frecuencia de clase, esto se logra cuando la acumulación de las frecuencias se realiza tomando en cuenta la primera clase hasta alcanzar la última. Las frecuencias acumuladas se designan con las letras  $f_a$ . Las frecuencias acumuladas pueden ser menor que ( $f_a < que$ ) y frecuencias acumuladas mayor que ( $f_a > que$ ).

## 8.- Frecuencia acumulada relativa

La frecuencia acumulada relativa es aquella que resulta de dividir cada una de las  $f_a$  de las diferentes clases que integran una distribución de frecuencia de clase entre el número total de datos ( $N$ ) de la serie de valores, estas frecuencias se designan con las letras  $f_{ar}$ . Si las  $f_{ar}$  se multiplican por 100 se obtienen las frecuencias acumuladas relativas porcentuales y las mismas se designan así:  $f_{ar} \%$ .

## La mediana

La mediana ( $M_d$ ) es una medida de posición que divide a la serie de valores en dos partes iguales, un cincuenta por ciento que es mayor o igual a esta y otro cincuenta por ciento que es menor o igual que ella. Es por lo tanto, un parámetro que está en el medio del ordenamiento o arreglo de los datos organizados, entonces, la mediana divide la distribución en una forma tal que a cada lado de la misma queda un número igual de datos.

Para encontrar la mediana en una serie de datos no agrupados, lo primero que se hace es ordenar los datos en una forma creciente o decreciente y luego se ubica la posición que esta ocupa en esa serie de datos; para ello hay que determinar si la serie de datos es par o impar, luego el número que se obtiene indica el lugar o posición que ocupa la mediana en la serie de valores, luego la mediana será el número que ocupe el lugar de la posición encontrada.

## La moda

La moda es la medida de posición que indica la magnitud del valor que se presenta con más frecuencia en una serie de datos; es pues, el valor de la variable que más se repite en un conjunto de datos. De las medidas de posición la moda es la que se determina con

mayor facilidad, ya que se puede obtener por una simple observación de los datos en estudio, puesto que la moda es el dato que se observa con mayor frecuencia. La moda se designa con las letras Mo.

#### Desviación típica o estándar

Es la medida de dispersión más utilizada en las investigaciones por ser la más estable de todas, ya que para su cálculo se utilizan todos los desvíos con respecto a la media aritmética de las observaciones, y además, se toman en cuenta los signos de esos desvíos. Se le designa con la letra castellana S cuando se trabaja con una muestra y con la letra griega minúscula s (Sigma) cuando se trabaja con una población. Es importante destacar que cuando se hace referencia a la población él número de datos se expresa con N y cuando se refiere a la muestra él número de datos se expresa con n. La desviación típica se define como:

#### Interpretación de la desviación estándar

La desviación típica como medida absoluta de dispersión, es la que mejor nos proporciona la variación de los datos con respecto a la media aritmética, su valor se encuentra en relación directa con la dispersión de los datos, a mayor dispersión de ellos, mayor desviación típica, y a menor dispersión, menor desviación típica.

#### Varianza

Es otra de las variaciones absolutas y la misma se define como el cuadrado de la desviación típica; viene expresada con las mismas letras de la desviación típica pero elevada al cuadrado, así  $S^2$  y  $s^2$ . Las fórmulas para calcular la varianza son las mismas utilizadas por la desviación típica, exceptuando las respectivas raíces, las cuales desaparecen al estar elevados el primer miembro al cuadrado

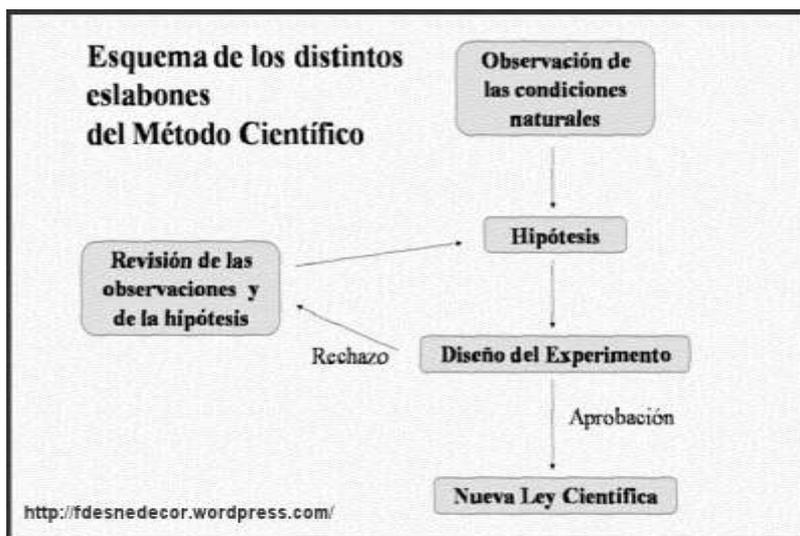
#### La Estadística dentro del Método Científico

La estadística no se puede utilizar como una caja mágica para extraer certezas, donde se introducen datos y se extraen leyes. La estadística, en el contexto de probabilidades y técnicas de inferencia, es incapaz por sí misma de suplantar al Método Científico, sólo es un gran apoyo.

## ¿Cómo ayuda la estadística en el Método Científico?

Definimos el Método Científico como un método o conjunto sistematizado de procesos en los que se basa la ciencia para explicar cualquier fenómeno y las leyes que los administran.

En la siguiente imagen os muestro, muy esquematizado, el proceso que se sigue al aplicar el Método Científico.



La estadística descriptiva es la herramienta más útil en la etapa de **observación**, ya que nos permite extraer información para realizar nuestras hipótesis fundadas en estos resultados. También es utilizada para valorar los resultados del **experimento**.

La estadística analítica se utiliza a partir de la **observación**, ya que dependiendo de los datos observados, se utilizará una técnica u otra, y por supuesto en el proceso del **experimento**, ya que su diseño dependerá en cierta medida de las técnicas estadísticas más apropiadas, además, la estadística analítica es el primer y principal razonamiento válido.

Como vemos, la estadística proporciona un gran apoyo al Método Científico en las fases de observación y experimentación, pero en el proceso de hipótesis y en el de la obtención de una ley científica son otras las bases