

La aplicación de procedimientos estadísticos es muy antigua. Una de las primeras pruebas de su aplicación se remonta al año 3050 a.C. en el cual se efectuó en Egipto un registro de la riqueza y población con el fin de preparar la construcción de las pirámides. Posteriormente, egipcios, griegos y romanos efectuaron censos esporádicos con fines tributarios, sociales o militares.

En el S. XVI aparecen publicadas compilaciones estadísticas en Francia e Italia de los recursos naturales y otras en Alemania complementarias de mapas y grabados. Durante el s. XVII se comienza a utilizar la palabra Estadística en Alemania y algunos eruditos muestran interés en la Estadística Demográfica como resultado de la utilidad que tenía para los gobernantes conocer si la población aumentaba, decrecía o se mantenía estática.

Una obra muy importante que influyó en el origen de la Estadística fue la de Jacques Bernouilli quien dedicó parte de su obra Arte de la Conjetura a la probabilidad. En ésta aparece un teorema conocido con su nombre que dice que el número de casos favorables a un suceso sobre el número de casos posibles tiende a la probabilidad de ese suceso cuando el número de veces que se repite el experimento tiende a infinito. La generalización de éste teorema dará origen a la ley de los grandes números, que dice que la probabilidad de que la media de una muestra difiera de la media de la población en una cantidad $\leq \varepsilon$ tiende a 0 cuando n tiende a infinito.

Sin embargo durante el s. XVII nadie sospecha que a partir de la teoría de probabilidades se va a originar la Estadística, la cual, si bien al principio fue aplicada a los asuntos de estado, (de donde viene su nombre), actualmente su influencia se extiende a todas las facetas del progreso humano, como la agricultura, negocios, química, biología, enseñanza, medicina, física, psicología, etc.

La Estadística o Métodos Estadísticos como a veces se la llama, está ligada actualmente a los métodos científicos en la toma, organización, recopilación, presentación y análisis de datos, tanto para la deducción de conclusiones como para tomar las decisiones razonables de acuerdo con tales análisis.

Se emplea con tres fines:

- *Descriptiva*: tiene por objeto resumir en un pequeño número de valores característicos, medias, desviaciones típicas, etc. o en gráficos facilitados para una amplia información.
- *Analítica*: tiende a buscar teorías que expliquen los hechos observados. Así, por ejemplo, a partir de numerosas experiencias dedujo Galileo la ley del péndulo.
- *De predicción*: Se pueden predecir los hechos observado con cierta regularidad la naturaleza, cuando se consideran muchas observaciones se presume o intuye que hay cierta limitación en la variedad. Es decir, la frecuencia relativa con que se presenta un suceso puede considerarse como probabilidad para experiencias sucesivas.

Cuando no es posible calcular la probabilidad a priori, se puede deducir observando la frecuencia en la repetición de sucesos.

Def.: La estadística es el estudio de los mejores modos de acumular y analizar datos y de establecer conclusiones acerca del colectivo del que se han recogido tales datos.

Def.: es la ciencia de la recopilación, clasificación, presentación e interpretación de datos.

La estadística es el lenguaje universal de las ciencias. La utilización cuidadosa de los métodos estadísticos permite:

- ☒ describir con precisión los hallazgos de la investigación científica.
- ☒ tomar decisiones.
- ☒ efectuar estimaciones.

Para lo recién enumerado, cuando se enfrenta a los enormes colectivos de posible información que las ciencias de la naturaleza o las ciencias humanas presentan, tiene que:

- Recoger datos selectivamente.
No es un problema superficial y constituye la teoría de muestras. Ej. si quiero indagar riqueza de una población no hago el relevamiento parando a los automovilistas para preguntarles.
- Organizar, resumir y entender la masa de datos recogida.
Incluso en una muestra de 1000 individuos tenemos una gran cantidad de datos que hay que ordenar, agrupar y determinar medias y otros parámetros que describan la muestra de modo resumido y transparente. Ésta es la finalidad de la estadística descriptiva.
- Extraer conclusiones de la información así obtenida.
Una vez que los datos han sido digeridos se busca extraer conclusiones lo que se hace mediante técnicas elaboradas y que constituyen campos tales como:
 - Inferencia estadística - Los colectivos estadísticos se suelen ajustar, muy frecuentemente, a algún tipo particular de una serie no muy amplia de distribuciones específicas. Un problema importante de la inferencia estadística consiste en averiguar, a partir del conocimiento de una muestra de datos tomada de una población, a qué tipo de distribución se ajusta esa población.
 - Estimación - Una vez que sabemos a qué tipo de distribución se ajusta la el colectivo estudiado, el problema consiste en cómo aproximamos con los datos numéricos que tenemos de la muestra al parámetro que caracteriza a la distribución.

En los tres pasos la base matemática utilizada a fin de proceder adecuadamente es la teoría de la probabilidad.

Podemos dividir el campo de la Estadística aproximadamente en dos áreas:

- **Estadística Descriptiva:** incluye la recopilación, presentación y descripción de datos. Su objetivo es describir un conjunto de datos.
- **Estadística Inferencial:** se refiere a la técnica de interpretar los valores resultantes de las técnicas descriptivas y a su posterior utilización para tomar decisiones.

Población es el conjunto de todos los elementos que serán objeto de estudio.

Muestra es un subconjunto de la población

Variable es la característica de interés acerca de cada elemento de una población o muestra.

Podemos clasificar las variables en:

✍ **Cuantitativas:** cuantifica, o sea, cuenta o mide. (peso, longitud, edad, tiempo, etc.)

Podemos clasificarlas en:

- **Discretas:** sólo puede tomar valores aislados, por lo que entre dos valores posible no siempre existen valores intermedios. (notas, nº hijos)
- **Continuas:** podría tomar todos los valores de un intervalo, por lo que siempre existen infinitos elementos entre dos posibles, teniendo en cuenta como única limitación la precisión del instrumento con el que se mide. (peso)

Un conteo siempre produce datos numéricos discretos, por ejemplo el número de animales hallados en un lugar es 2, 3, etc., pero no 3,25.

✍ **Cualitativas:** categoriza, describe el elemento de una población. (estado civil, marca, etc.)

Podemos clasificarlas en:

- **Nominales:** todas las clases son equivalentes (marcas de celulares, estado civil)
- **Ordinales:** Las clases pueden ser ordenadas (nivel educativo, rendimiento)

Dato es el valor de la variable asociado a un elemento de una población o una muestra.

Experimento es la actividad realizada según un plan definido cuyos resultados producen un conjunto de datos.

Frecuencia es el número de veces que aparece una categoría o valor de una variable en la muestra.

Frecuencia relativa es la proporción que representa la frecuencia absoluta con respecto al número de observaciones.

Parámetro es la característica numérica de una población.

Estadística es la característica numérica de una muestra.

Medibilidad y variabilidad: siempre se espera que ocurra variabilidad en un conjunto de datos experimentales. No importa de qué variable se trate, siempre existe variabilidad en la respuesta numérica si el instrumento de medición es suficientemente preciso. Uno de los objetivos en el análisis estadístico es la medición de la variabilidad, por ej. en controles de calidad.

RECOPIACIÓN DE DATOS

Uno de los primeros problemas con los que se enfrenta la Estadística es la recopilación de los datos. Se deben tener buenas técnicas de muestreo puesto que las inferencias obtenidas finalmente estarán basadas en las estadísticas calculadas a partir de los datos muestrales. La obtención de datos para el análisis estadístico es un proceso integral que incluye las siguientes etapas:

1. Definición de los objetivos del estudio observacional o experimento.
2. Definición de la variable y la población de interés.
Ej.: ingreso familiar total de la ciudad
3. Definición de los métodos para la obtención y medición de datos.
Esto incluye los procedimientos de muestreo, la determinación del tamaño de la muestra y la forma en que se obtendrán los datos.
4. Determinación de las técnicas descriptivas o inferenciales que sean apropiadas para el análisis de datos.

Marco de muestreo: es una lista de elementos pertenecientes a la población de donde se extrae la muestra. Debe ser idéntico a la población.

Diseño de la muestra o plan de muestreo: es el conjunto de procedimientos utilizados para seleccionar los elementos de la muestra. Se los puede clasificar en dos categorías:

-  **Muestras no probabilísticas:** son las que se seleccionan a partir de su consideración como "típicas". Se eligen los elementos que se cree son representativos de la población.
-  **Muestras probabilísticas:** son las muestras donde los elementos por seleccionar se toman de acuerdo con leyes probabilísticas. Cada elemento en una población tiene cierta probabilidad de ser seleccionado como parte de la muestra.

Los procedimientos de la estadística inferencial requieren muestras probabilísticas. Algunas de estas son:

- *Muestra aleatoria simple*: es aquella seleccionada de manera que cada elemento de la población tiene la misma probabilidad de ser elegido. De manera equivalente, todas las muestras de tamaño n tienen la misma probabilidad de ser elegidas. El empleo de un generador de números aleatorios o de una tabla de números aleatorios es el procedimiento idóneo para seleccionar una muestra aleatoria simple.
- *Muestra sistemática*: es aquella donde se selecciona toda k -ésimo elemento en el marco de muestreo después de un inicio aleatorio entre los primeros k elementos. Por tanto se utiliza la tabla de números aleatorios solo la primera vez. Para seleccionar un porcentaje X como muestra sistemática es necesario escoger 1 de cada $100/X$ elementos en el marco del muestreo. Por ejemplo, si se quiere el 4%, luego que se tiene el primero escogido aleatoriamente entre los 25 primeros, se va de 25 en 25.

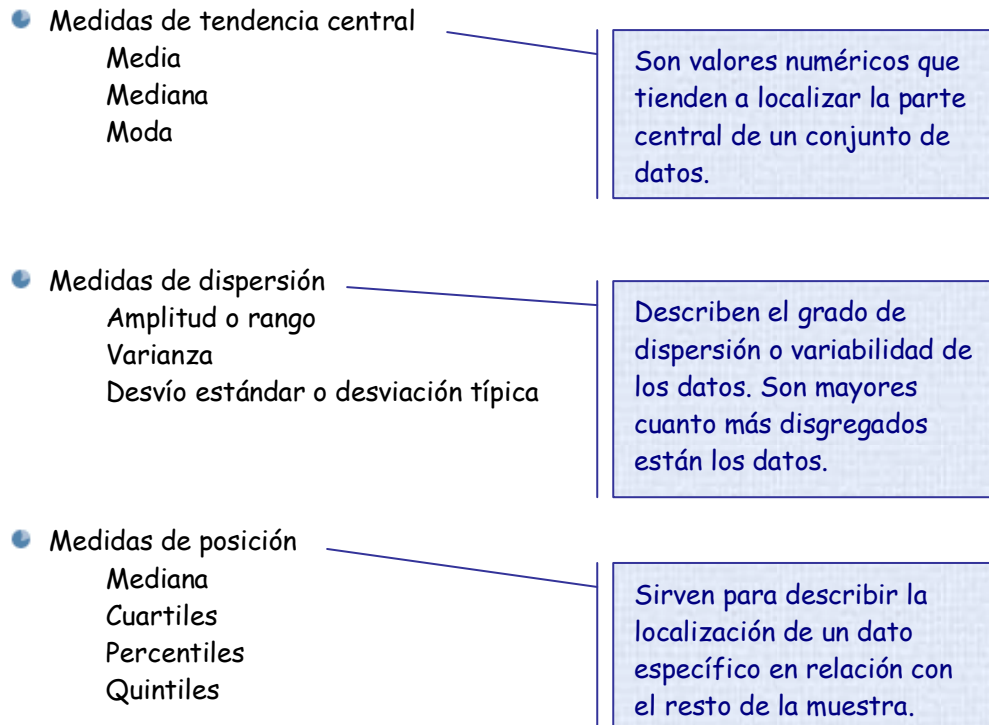
Algunas veces cierta característica de una población de gran tamaño, permite dividirla en subpoblaciones, que reciben el nombre de estratos. Cada estrato se muestrea por separado.

- *Muestra estratificada*: es aquella que se obtiene estratificando primero al marco de muestreo y seleccionando después un número fijo de elementos de cada estrato mediante un muestreo aleatorio simple.
- *Muestra por conglomerado*: es aquella que se obtienen estratificando primero el marco de muestreo y seleccionando después todos los elementos de algunos estratos.

El plan de muestreo que se utilice en una situación particular depende de varios factores:

1. de la naturaleza tanto de la población como de la variable.
2. de la facilidad o dificultad del muestreo.
3. del costo del muestreo.

CÁLCULO DE ESTADÍSTICAS DESCRIPTIVAS



Las medidas numéricas que se obtienen a partir de los datos pretenden resumir las características de una distribución en números con un significado preciso. Es deseable que éstas medidas numéricas presenten las siguientes propiedades:

- Deben ser definidas en forma objetiva, dos personas distintas deben obtener el mismo resultado.
- Deben depender de todas las observaciones porque si no lo hacen ignoran parte de la información contenida en los datos.
- Deben tener un significado concreto, la interpretación de una medida debe ser simple e inmediata.
- Deben ser simples de calcular.
- Deben tener propiedades simples que permitan un fácil análisis teórico y su manipulación algebraica.
- Deben ser poco sensibles a valores extremos, la medida descriptiva no se debe afectar por la aparición en los datos de valores aberrantes que pueden surgir como resultado de algún error de medición.

En general las medidas descriptivas no cumplen todas estas propiedades, por lo tanto deben elegirse aquellas que tengan las propiedades que más necesitemos según sea lo que demande un problema en particular.

Las medidas descriptivas en la mayoría de los casos se evalúan en una muestra con el propósito de estimar un parámetro poblacional.

