

ESTADISTICA

DEFINICIÓN

ESTADISTICA: Ciencia de recolectar describir e interpretar datos La estadística tiene antecedentes históricos en los censos, que consistían en observaciones sistemáticas y periódicas sobre datos de la población para fines de guerra y finanzas realizados años antes de Cristo. En la actualidad la estadística no limita solamente a la recolección de datos, sino a la organización, recopilación y análisis de los mismos con un determinado objetivo.

La estadística se ha convertido en el lenguaje universal de las ciencias. Como potencial usuario de ella, necesitas dominar tanto la “ciencia” como el “arte” de usar correctamente la metodología estadística. El uso cuidadoso de los métodos estadísticos te permitirá obtener información precisa a partir de los datos. Dichos métodos incluyen: 1) definir cuidadosamente la situación, 2) recolectar datos, 3) resumir con precisión los datos y 4) derivar y comunicar conclusiones significativas.

La estadística involucra información, números y gráficos visuales para resumir esta información y su interpretación. La palabra *estadística* tiene diferentes significados para personas de varios antecedentes e intereses. Para algunas personas es un campo de “trucos mágicos” donde una persona trata de abrumar a otros con información y conclusiones incorrectas. Para otros es una forma de recolectar y mostrar información. Y para otros más es una manera de “tomar decisiones ante la incertidumbre”. En la perspectiva apropiada, cada uno de dichos puntos de vista es correcto.

El campo de la estadística puede subdividirse burdamente en dos áreas: estadística descriptiva y estadística inferencial. La **estadística descriptiva** es en lo que piensa la mayoría de las personas cuando escuchan la palabra estadística. En ella se incluye la recolección, presentación y descripción de datos muestrales. El término **estadística inferencial** se refiere a la técnica de interpretar los valores que resultan a partir de las técnicas descriptivas, tomar decisiones y extraer conclusiones acerca de la población.

La estadística es más que sólo números; son datos, lo que se le hace a los datos, lo que se aprende de los datos y las conclusiones resultantes. Se usará la siguiente definición:

La palabra "estadística" suele utilizarse bajo dos significados distintos, a saber:

1º Como colección de datos numéricos. Este es el significado más vulgar de la palabra estadística. Se sobreentiende que estos datos numéricos han de estar presentados de manera ordenada y sistemática. Una información numérica cualquiera puede no constituir una estadística, para merecer este apelativo, los datos han de constituir un conjunto coherente, establecido de forma sistemática y siguiendo un criterio de ordenación.

2º Como ciencia. En este significado, la Estadística estudia el comportamiento de los fenómenos de masas. Como todas las ciencias, busca las características generales de un colectivo y prescinde de las particulares de cada elemento.

Así por ejemplo al investigar el sexo de los nacimientos, se inicia el trabajo tomando un grupo numeroso de nacimientos y obteniendo después la proporción de varones.

Elementos utilizados en el análisis estadístico

- **Individuo o unidad estadística:** es cada uno de los elementos que componen una población. Cabe mencionar que cuando decimos individuo no exclusivamente nos referimos a una persona.
- **Población o Universo:** Es el conjunto de todos los elementos que van a ser estudiados y que permitirán resolver un problema, los cuales presentan una característica común determinada, observable y medible.
- **Parámetro:** Valor numérico que resume todos los datos de una población entera.

Un parámetro es un valor que describe a la población entera. Con frecuencia se usa una letra griega para simbolizar el nombre de un parámetro. Dichos símbolos se asignarán conforme se estudien parámetros específicos.

Para cada parámetro existe un *estadístico muestral correspondiente*. El estadístico describe la muestra de la misma forma que el parámetro describe a la población.

- **Muestra:** Es un subconjunto o una porción de la población. La cual debe tener carácter representativo de lo que se quiere estudiar.
- **Variable:** Característica o fenómeno de una población o muestra que será estudiada, la cual puede tomar diferentes valores. (Más adelante se encuentra información detallada de estas características)
- **Datos:** Son los valores numéricos que han sido recopiladas como resultado de la observación.
- **Valor de datos:** El valor de la variable asociada con un elemento de una población o muestra. Este valor puede ser un número, una palabra o un símbolo.
- **Estadístico:** valor numérico que resume los datos muestrales.

La mayoría de los estadísticos muestrales se encuentran con la ayuda de fórmulas y usualmente se les asignan nombres simbólicos que son letras del alfabeto (por ejemplo, $\bar{x}$, s y r).

- **Experimento:** Actividad planificada cuyos resultados producen un conjunto de datos.

Un experimento incluye las actividades tanto para seleccionar los elementos como para obtener los valores de datos.

Clasificación de la estadística

➤ **Estadística descriptiva o deductiva**

Esta se emplea simplemente para resumir de forma numérica o gráfica un conjunto de datos. Se restringe a describir los datos que se analizan. Si aplicamos las herramientas ofrecidas por la estadística descriptiva a una muestra, solo nos limitaremos a describir los datos encontrados en dicha muestra, no se podrá generalizar la información hacia la población.

➤ **Estadística inferencial o inductiva**

La estadística inferencial permite realizar conclusiones o inferencias, basándose en los datos simplificados y analizados de una muestra hacia la población o universo. Por ejemplo, a partir de una muestra representativa tomada a los habitantes de una ciudad, se podrá inferir la votación de todos los ciudadanos que cumplan los requisitos con un cierto error de aproximación.

La investigación estadística

El proceso de aplicación de la estadística implica una serie de pasos:

- 1) Selección y determinación de la población o muestra y las características contenidas que se desean estudiar. En el caso de que se desee tomar una muestra, es necesario determinar el tamaño de la misma y el tipo de muestreo a realizar (probabilístico o no probabilístico).
- 2) Obtención de los datos. Esta puede ser realizada mediante la observación directa de los elementos, la aplicación de encuestas y entrevistas, y la realización de experimentos.
- 3) Clasificación, tabulación y organización de los datos. La clasificación incluye el tratamiento de los datos considerados anómalos que pueden en un momento dado, falsear un análisis de los indicadores estadísticos. La tabulación implica el resumen de los datos en tablas y gráficos estadísticos.
- 4) Análisis descriptivo de los datos. El análisis se complementa con la obtención de indicadores estadísticos como las medidas: de tendencia central, dispersión, posición y forma.
- 5) Análisis inferencial de los datos. Se aplican técnicas de tratamiento de datos que involucran elementos probabilísticos que permiten inferir conclusiones de una muestra hacia la población (opcional).
- 6) Elaboración de conclusiones. Se construye el informe final.

MUESTREO

El muestreo es un procedimiento estadístico que permite realizar una investigación de cualquier naturaleza (económica, social, educativa, médica) relativa a una población determinada utilizando una cantidad de elementos menor que el total de la población, con lo cual se realiza la tarea con mayor rapidez y a menor costo, pero sin garantizar exactitud, que sólo se consigue efectuando la investigación mediante una metodología censal. La

selección de una muestra constituye una tarea compleja que toman en cuenta una cierta cantidad de principios.

Condiciones para el muestreo

Para realizar una investigación muestral deben cumplirse dos condiciones fundamentales:

I) Representatividad: esta condición implica que la muestra debe integrarse con una proporción de eventos que resulte similar a la composición poblacional. Observe que no se exige para la muestra una composición exactamente igual, sino sólo similar.

II) Confiabilidad: implica que el investigador debe evitar que las respuestas que obtenga de los elementos que integran la muestra contengan errores, omisiones o falseamientos, debido a que, si un elemento de la muestra da una información errónea, ya que cada elemento representa a una determinada cantidad, ese error se multiplica por esa misma cantidad.

Dos métodos comúnmente usados para la recolección de datos son *experimentos y estudios observacionales*. En un **estudio observacional**, el investigador no modifica el entorno y no controla el proceso a observar. Los datos se obtienen al muestrear parte de la población de interés. Las **encuestas** son estudios observacionales de personas.

Si todo elemento en la población puede mencionarse o enumerarse y observarse, entonces se compila un **censo**. Sin embargo, los censos se usan rara vez porque con frecuencia son difíciles de compilar y consumen mucho tiempo y por tanto son muy costosos. Imagina la tarea de compilar un censo de cada persona que es un cliente potencial de una empresa de corretaje. En situaciones similares a ésta, por lo general se realiza una *encuesta piloto*.

Cuando seleccionamos una muestra debemos realizar un marco muestral

Marco muestral: lista o conjunto de los elementos que pertenecen a la población de la cual se extraerá la muestra. Es importante que ese marco muestral sea representativo de la población.

Una vez establecido el marco muestral representativo, se procede con la selección de los elementos muestrales del marco muestral. Este proceso de selección se llama **diseño muestral**. Existen muchos tipos diferentes de diseños muestrales; sin embargo, todos ellos encajan en dos categorías: *muestras dirigidas y muestras probabilísticas*.

Una vez establecido el marco muestral representativo, se procede con la selección de los elementos muestrales del marco muestral. Este proceso de selección se llama diseño muestral. Existen diferentes 2 categorías de diseños muestrales. Muestras dirigidas o no probabilísticas y muestras probabilística.

Tipos de muestreo

Hay dos tipos fundamentales de muestreo:

I) Muestreo probabilístico: en este tipo de muestreo cada elemento de la población puede intervenir en la muestra con una determinada probabilidad la que, según el caso, puede resultar igual o diferente para los elementos que integran la población. Los más comunes son:

a) Muestreo simple al azar: consiste en calcular el tamaño de la muestra y luego efectuar la selección, asignando a cada uno de los elementos que componen la población igual probabilidad de aparecer en la muestra. Este método es conveniente cuando los elementos de la población no tienen excesiva variación entre ellos.

b) Muestreo estratificado: conviene aplicar este sistema si los elementos que componen la población tienen una gran diferencia entre ellos. Consiste en dividir a la población en varios estratos, que son grupos de elementos internamente homogéneos pero heterogéneos entre diferentes los diferentes estratos. Realizada esta división cada estrato se selecciona una muestra de elementos. El total de elementos seleccionados está dado por la suma de las cantidades de cada estrato. Ese total resultará menor que si se aplicara el sistema del muestreo simple al azar.

c) Muestreo sistemático: se aplica cuando la población tiene un orden previo no secuencial (es decir, si no se encuentra ordenada de menor a mayor, o viceversa, por ejemplo). En ese caso, este método resulta ser equivalente al muestreo simple al azar, pero mucho más cómodo de aplicar.

II) Muestreo no probabilístico: son las que se seleccionan sobre la base de juzgarse si son representativas o no de la población. La muestra dirigida refleja el juicio del recolector

a) Muestreo por cuotas: en este método la muestra se selecciona de tal modo que su composición respeta una determinada integración (por sexo, edad, nivel social, nivel de estudios) fijada previamente.

b) Muestreo por conveniencia: en este procedimiento se integra la muestra con ciertos elementos que resulten fácilmente accesibles, de modo de cumplir el trabajo de relevamiento con rapidez y sin excesivos costos.

Proceso de la recolección de datos

1. Definir los objetivos de la encuesta o estudio.
Ejemplos: comparar la efectividad de un nuevo medicamento con la efectividad del medicamento estándar; estimar el ingreso doméstico promedio en Estados Unidos.
2. Definir la variable y la población de interés.
Ejemplos: duración del tiempo de recuperación para los pacientes que sufren de una enfermedad particular; ingreso total de los hogares en Estados Unidos.
3. Definir cómo recolectar los datos y los esquemas de medición de datos.
Esto incluye el marco del muestreo, los procedimientos de muestreo, el tamaño muestral y el dispositivo de medición de datos (cuestionario, teléfono, etcétera).
4. Recolección de la muestra. seleccionar los sujetos a muestrear y recolectar datos.
5. Revisar el proceso de muestreo al completar la recolección.

CLASIFICACIÓN DE LAS VARIABLES

Es muy probable que una persona con conocimientos en Estadística que realiza una encuesta desee desarrollar un instrumento que le permita hacer varias preguntas y manejar diversos fenómenos o características. A estos fenómenos o características se les denomina variables aleatorias.

Según la forma en que se expresen las variables, se dividen en:

 **Variable cualitativa, categórica o atributo** Variable que describe o jerarquiza un elemento de una población. Son las variables que expresan distintas cualidades, características o modalidad. Cada modalidad que se presenta se denomina atributo o categoría y la medición consiste en una clasificación de dichos atributos. Las variables cualitativas pueden ser ordinales y nominales.

A) Variable nominal: Variable cualitativa que caracteriza (describe o nombra) un elemento de una población. No sólo las operaciones aritméticas no son significativas para los datos que resultan de una variable nominal, tampoco puede asignarse un orden a las categorías. En esta variable los valores no pueden ser sometidos a un criterio de orden como por ejemplo los colores, el lugar de residencia, etc.

B) Variable ordinal: Variable cualitativa que incorpora una posición ordenada o clasificación. La variable puede tomar distintos valores ordenados siguiendo una escala establecida, aunque no es necesario que el intervalo entre mediciones sea uniforme. Las variables ordinales pueden ser dicotómicas cuando sólo pueden tomar dos valores posibles como sí y no, hombre y mujer. Las variables ordinales son politómicas cuando pueden adquirir tres o más valores como, por ejemplo, leve, moderado, grave.

Ejemplo:

Una muestra de cuatro clientes de un salón de belleza se encuesta por su “color de cabello”, “ciudad de origen” y “nivel de satisfacción” con los resultados de su tratamiento en el salón. Las tres variables son ejemplos de variables (atributo) cualitativo porque describen alguna característica de la persona, y todas las personas con el mismo atributo pertenecen a la misma categoría. Los datos recolectados fueron {rubio, café, negro, café}, {Brighton, Columbus, Albany, Jacksonville} y {muy satisfecho, satisfecho, un poco satisfecho, no satisfecho}.

Variable nominal:

En la encuesta de cuatro clientes del salón de belleza, dos de las variables, “color de cabello” y “ciudad de origen”, son ejemplos de variables nominales porque ambas mencionan alguna característica de la persona y no serían significativas para encontrar el promedio muestral al sumar y dividir entre 4. Por ejemplo $(\text{rubio} + \text{café} + \text{negro} + \text{café})/4$ es indefinido. Más aún, el color de cabello y la ciudad de origen no tienen un orden en sus categorías.

Variable ordinal:

En la encuesta de cuatro clientes del salón de belleza, la variable “nivel de satisfacción” es un ejemplo de una variable ordinal porque sí incorpora una clasificación ordenada: “muy satisfecho” se clasifica adelante de “satisfecho”, que clasifica adelante de “un poco satisfecho”. Otra ilustración de una variable ordinal es la clasificación de cinco imágenes de paisajes de acuerdo con la preferencia de alguien: Primera elección, segunda elección, etcétera.

 **Variable cuantitativa o numérica** Variable que cuantifica un elemento de una población. Son aquellas variables que pueden expresarse en forma numérica. Se dividen en discretas y continuas.

A) Variable discreta Variable cuantitativa que puede asumir un número contable de valores. Intuitivamente, la variable discreta puede asumir cualquier valor correspondiente a puntos aislados a lo largo de un intervalo lineal. Esto es: entre dos valores cualesquiera existe un intervalo. Son respuestas numéricas que surgen de un proceso de conteo, siendo siempre un número entero. Como, por ejemplo: Número de asignaturas inscritas en el primer semestre, número de integrantes del grupo familiar, Número de salas de clases.

B) Variable continua Variable cuantitativa que puede asumir un número incontable de valores. Intuitivamente, la variable continua puede asumir cualquier valor a lo largo de un intervalo lineal, incluido todo posible valor entre dos valores cualesquiera. Son respuestas numéricas que surgen de un proceso de medición, las cuales pueden tomar valores entre dos números enteros. Como, por ejemplo: la estatura y/o el peso de personas, la temperatura en una cierta región, etc.