

UNIDAD N° 1

ESTADÍSTICA DESCRIPTIVA

Población. Muestra. Ramas de la Estadística.
Variables. Escalas de medición.

UNIDAD N° 1: ESTADÍSTICA DESCRIPTIVA

INTRODUCCIÓN

La estadística es la ciencia que proporciona técnicas y métodos para la recolección, sistematización y análisis de datos con el fin de extraer conclusiones y tomar decisiones frente a situaciones de incertidumbre. Esta incertidumbre está determinada por lo que llamaremos variabilidad de los datos producto de las mediciones desde donde estos provienen.

Las técnicas y métodos que utiliza la estadística podemos clasificarlos en dos grandes grupos y determinan dos grandes ramas de esta ciencia llamadas *estadística descriptiva* y *estadística inferencial*.

DEFINICIONES PREVIAS

Población: Es el conjunto de la totalidad de elementos bajo estudio. A la población la podemos clasificar en finitas (cuando tienen un número limitado de elementos) e infinitas (cuando tienen un número ilimitado de elementos). Cuando podemos realizar un estudio sobre toda la población este recibe el nombre de **censo** o enumeración completa.

Muestra: Es un subconjunto de la población que se selecciona ante la imposibilidad de estudiar a toda la población. Existen técnicas que nos permiten determinar cómo seleccionar este subconjunto de la población para que sea representativo de esta, conocidas como técnicas de muestreo.

Para que los resultados sean confiables, la muestra debe ser:

- Representativa (debe reflejar las características de la población).
- Adecuada en tamaño (ni demasiado pequeña, ni innecesariamente grande).
- Seleccionada de forma imparcial (por ejemplo, con métodos aleatorios).

Unidad elemental: Es cada uno de los elementos individuales que componen la población o la muestra y sobre los que se realizan las observaciones.

RAMAS DE LA ESTADÍSTICA

Cuando recopilamos datos, por ejemplo, de una muestra de una población, lo que nos interesa es poder organizarlos y resumirlos para poder extraer información y tomar decisiones. Los métodos que se utilizan para obtener este tipo de información corresponden a la **estadística descriptiva**. Algunos de estos métodos son de naturaleza gráfica como los diagramas de punto, histogramas, diagrama de caja, entre otros y nos permiten obtener de forma visual una síntesis de los datos. Otros métodos consisten en obtener medidas numéricas como, por ejemplo, la media, varianza, coeficiente de correlación, etc.

En determinadas ocasiones podemos utilizar la información aportada por los datos de una muestra para obtener conclusiones sobre la población a la que pertenece. Las técnicas que se utilizan para realizar este tipo de generalizaciones se agrupan dentro de la **estadística inferencial**.

VARIABLES

Cuando realizamos un experimento u observaciones para extraer datos y analizarlos, estos pueden ser de dos tipos: categóricos (cualitativos) o numéricos (cuantitativos).

A la característica que estamos estudiando sea cuantitativa o cualitativa, la llamamos variable y entonces tenemos que las variables pueden clasificarse en:

Variables cualitativas: toman valores no numéricos. Son aquellas donde los datos observados se pueden clasificar en categorías. Las categorías pueden o no tener un orden natural. Son también llamadas variables categóricas. Por ejemplo: lugar de nacimiento, profesión, estado civil, género, etc. Dentro de este tipo de variables podemos distinguir:

- **Variables cualitativas ordinales:** las categorías implican un orden. Por ejemplo, la escala de calificación conceptual (Excelente, Muy bueno, bueno, etc.) o las medallas obtenidas en una competencia (Oro, Plata, Bronce).
- **Variables cualitativas nominales:** las categorías no implican un orden. Por ejemplo, color de ojos, lugar de nacimiento.

Variables cuantitativas: toman valores numéricos que son mediciones (longitud, peso) o frecuencias (cantidad). Dentro de estas variables podemos distinguir:

- **Variables cuantitativas discretas:** toman una cantidad numerable finita o infinita contable de valores, por ejemplo, el número de hojas de un árbol o la cantidad de artículos vendidos en un negocio. Son el resultado de un conteo.
- **Variables cuantitativas continuas:** puede tomar infinitos valores entre dos dados, por ejemplo, la altura de un grupo de personas o el tiempo necesario para realizar un determinado proceso.

En la mayoría de los problemas prácticos las *variables continuas* representan *datos medidos*, como serían todos los posibles pesos, alturas, temperaturas, distancias o periodos de vida; en tanto que las *variables discretas* representan *datos por conteo*, como el número de artículos defectuosos en una muestra o el número de accidentes en una autopista por año.

ESCALAS DE MEDICIÓN

En estadística, medir significa asignar un valor a una característica de los elementos que estamos estudiando, siguiendo ciertas reglas. Ese valor puede ser un número o una categoría, y su finalidad es representar la información de manera tal que se pueda analizar, comparar y comunicar con precisión.

Por ejemplo:

- Cuando registramos el tiempo de ejecución de un script en milisegundos, estamos midiendo una duración.
- Cuando asignamos a un usuario una categoría como “satisfecho” o “insatisfecho”, también estamos midiendo (en este caso, una opinión) usando categorías.

Medir no es simplemente observar o notar una característica. Es transformarla en un dato legible y utilizable para el análisis estadístico.

Toda medición implica:

- Una característica observable (Variable).

- Una regla de asignación (por ejemplo, registrar en segundos, o elegir entre dos opciones).
- Un resultado: un valor que representa esa característica para cada caso individual.

Medir correctamente nos permite organizar los datos, analizarlos con métodos adecuados, identificar patrones y fundamentar decisiones con base estadística.

Las cuatro escalas de medición

Cuando medimos una variable, no solo estamos asignándole un dato: también estamos definiendo cómo puede interpretarse ese dato y qué tipo de análisis podemos hacer con él.

Imaginemos que las escalas de medición son niveles de profundidad en la forma de representar y trabajar con los datos. A medida que avanzamos de una escala a otra, vamos ganando en precisión, posibilidades de análisis y riqueza informativa.

Escala Nominal: nombrar para distinguir

La escala nominal es la más básica: sirve para identificar o clasificar los elementos dentro de grupos o categorías. No hay orden, jerarquía ni distancia entre los valores. Simplemente, cada categoría distingue algo de lo demás.

Ejemplos:

- Tipo de base de datos utilizada: relacional, documental, orientada a objetos.
- Lenguaje de programación preferido: Python, Java, C++.

Este tipo de datos nos permite contar cuántos elementos hay en cada grupo, pero no permite establecer un orden ni calcular diferencias entre ellos.

Escala Ordinal: reconocer un orden

La escala ordinal también clasifica, pero ahora los valores tienen un orden natural. Podemos decir que uno es "más" o "menos" que otro, aunque no sepamos exactamente cuánto más.

Ejemplos:

- Nivel de prioridad de un ticket en un sistema de soporte: Alta, Media, Baja.
- Clasificación de desempeño: Excelente, Bueno, Regular, Insuficiente.

Esta escala es útil cuando queremos establecer jerarquías o rankings, aunque no podamos hacer operaciones aritméticas entre los valores.

Escala de Intervalo: medir diferencias

Con la escala de intervalo, además de clasificar y ordenar los datos, podemos medir las diferencias entre los valores. Es decir, sabemos cuánto más o cuánto menos vale un dato respecto a otro. Sin embargo, hay una limitación importante: el cero no representa ausencia de lo que se mide, sino que es un punto arbitrario de referencia dentro de la escala.

Ejemplos:

- Temperatura en grados Celsius o Fahrenheit.
- Fechas en un calendario (años).

En estas variables, podemos sumar, restar y calcular promedios, porque las diferencias entre valores son significativas. Pero no tiene sentido comparar proporcionalmente. Por ejemplo:

No podemos decir que 30°C es el doble de 15°C , porque 0°C no significa “sin temperatura”.

Tampoco tiene sentido decir que el año 2000 es el doble del año 1000, en todo caso podemos decir que hay una diferencia de 1000 años entre ambos.

Escala de Razón: comparar y multiplicar

Finalmente, la escala de razón es la más completa. Permite clasificar, ordenar, medir diferencias y, además, tiene un cero absoluto, lo que nos permite realizar operaciones como multiplicar y dividir los valores.

Ejemplos:

- Tiempo de respuesta en milisegundos.
- Cantidad de líneas de código en un programa.
- Espacio de almacenamiento en megabytes.

Lo que distingue a esta escala es que permite comparar valores de manera proporcional, ya que el cero representa la ausencia total de lo que se mide. Por eso tiene sentido afirmar que una ejecución que tarda 4 ms tarda el doble de una que tarda 2 ms,

o que un archivo de 1 GB ocupa solo una cuarta parte del espacio de uno de 4 GB.

¿Cómo se relacionan las escalas con los tipos de variables?

Es importante tener presente que:

Las escalas **nominal y ordinal** se utilizan para describir variables cualitativas, ya que sus valores son categorías o atributos.

Las escalas **de intervalo y de razón** se aplican a variables cuantitativas, cuyos valores son numéricos y permiten realizar operaciones matemáticas.

Identificar adecuadamente el tipo de variable y su escala de medición es fundamental para no cometer errores al seleccionar las herramientas estadísticas adecuadas y garantizar decisiones basadas en análisis correcto.