

# *ESTADÍSTICA*

## CAPÍTULO 1

### CONCEPTOS Y DEFINICIONES

---

---

*La estadística es el conjunto de métodos o técnicas que se aplican a la recolección, organización, presentación, análisis e interpretación de datos numéricos.*

La estadística consiste en los cuatro pasos fundamentales que se acaban de mencionar: primero se recolectan datos, los cuales, por lo general, en ese primer paso quedan en desorden; se podría decir que se tienen los datos, pero todos revueltos. ¿Por qué? Supóngase, por ejemplo, que se quiere saber el promedio del peso (en kilogramos) de las personas que viajan en avión. Entonces, para recolectar esos datos se les indicará a los viajeros que pasen a pesarse a la báscula. Una persona se encargará de registrar el peso que va dando cada persona. Y evidentemente lo mismo podrá salir en primer lugar una gente con 80 kilos que otra con 60; el siguiente apenas con 48 o con 75 kilos; y así sucesivamente. Es decir, conforme pase a la báscula cada viajero quedarán registrados todos sus pesos, pero sin ningún orden.

Para eso es la segunda etapa, la de organizar los datos, o sea ponerlos en orden, ya sea de menor a mayor o a la inversa, como los pesos de un grupo de personas, o sus edades, o las calificaciones de un grupo escolar, las temperaturas mensuales en una ciudad, etc., o también por intervalos, anotando cuántos datos se registraron de cada valor. Esto suele hacerse a través de tablas.

La presentación de esos datos ya ordenados es para visualizar el comportamiento del grupo en estudio. Suelen hacerse esas presentaciones por medio de gráficas.

Y finalmente la cuarta etapa es la interpretación, que de hecho es realmente el objetivo que se persigue con la estadística. Sin la interpretación de los datos carecería de sentido la estadística. Sería el equivalente a tener muchos datos recolectados y ya.

Cada una de estas cuatro etapas será motivo de estudio en forma individual cada una en diferente capítulo.

La estadística o métodos estadísticos son de gran utilidad en muchas disciplinas, ya que en la vida práctica frecuentemente se necesita extraer conclusiones válidas y confiables respecto de un grupo de individuos u objetos. Por ejemplo, una fábrica de tornillos seguramente requerirá saber el porcentaje de tornillos defectuosos que produce.

La medicina, la biología, la economía, la sociología, los deportes, la astronomía, la docencia, etc., aplican los métodos estadísticos en diferentes niveles de complejidad, desde aquellos que utilizan técnicas muy elaboradas hasta los que solamente requieren tablas de información. En la actualidad la mayoría de las ciencias deben recurrir a la estadística de manera indispensable para su enriquecimiento.

*UNA POBLACIÓN es el conjunto total de individuos u objetos acerca del cual se quiere saber algo. Es el conjunto universal o el todo.*

Ejemplos: Una población puede ser un grupo del 3<sup>er</sup> semestre de la escuela, acerca del cual se desea saber su promedio de calificaciones en un mes determinado, o bien mes por mes; la característica principal es que se trata de todo el grupo, no parte de él. Una población pueden ser los habitantes de una ciudad de la cual se desea saber el índice de enfermedades en las diferentes épocas del año; la característica principal es que se trata de todos los habitantes, no de una parte nada más. Una población pueden ser los trabajadores de una fábrica de los que se desea saber su porcentaje de rendimiento en el trabajo; la característica principal es que se trata de todos los trabajadores, no de unos cuantos.

Una población, dicho con otras palabras, es el conjunto universal, es decir, es el todo.

*UNA MUESTRA es una parte representativa de la población. Es un subconjunto de la población.*

A veces no es conveniente analizar uno a uno todos los elementos de una población, ya sea por razones de costo, de tiempo o de posibilidad, en cuyos casos se seleccionan de manera representativa algunos integrantes de la población. Esos son la *muestra*.

**COSTEABILIDAD:** Por ejemplo, supóngase que se desea estimar el porcentaje de tornillos defectuosos que se producen en una fábrica en cierto período. Si se fabrican 500 mil piezas a la semana no es costeable pagar a un trabajador para que se dedique a analizar uno a uno todos los tornillos. Entonces se seleccionan algunos de los tornillos bajo ciertas técnicas que proporciona la estadística, se calcula el porcentaje de tornillos defectuosos detectados en esa muestra y el resultado se hace extensivo a toda la población. Si en la muestra el 5% resultó defectuoso, debe esperarse que en toda la población el promedio sea igual o muy próximo a ese 5%.

**DESTRUCCIÓN:** En otros casos no se puede analizar toda la población porque ésta se destruiría. Por ejemplo, supóngase que se desea saber el tiempo promedio de duración de ciertas piezas antes de que sufran desgaste considerable. Una manera de hacerlo sería probando pieza por pieza, pero eso implica que haya que someter a todas las piezas a la prueba y esperar a que se desgasten hasta su inutilización para obtener su tiempo y promediar; entonces se tendría ciertamente la información exacta del tiempo promedio de duración de las piezas fabricadas, pero a costa de haberlas destruido. De manera que para evitar lo anterior, se recurre a la muestra.

**TEMPORALIDAD :** Existen también ciertos eventos que dependen del tiempo, siendo la población total los eventos del pasado, los del presente y los que acontecerán en el futuro, por lo tanto resulta imposible analizar a los del futuro. En tales casos la muestra vienen siendo los sucesos ya verificados (del pasado). Por ejemplo, se realiza un análisis durante 50 años sobre las temperaturas promedio en las diferentes épocas del año en una ciudad; al observarse que en Mayo las temperaturas siempre oscilaron entre los 29 y 32 grados, podrá hacerse una especie de "predicción" para los años futuros en el sentido que se repetirán las mismas temperaturas en el mes de Mayo. Aquí la muestra son los cincuenta años ya pasados que fueron estudiados, mientras que la población la forman todos los años, incluyendo a los que están por venir.

**IMPOSIBILIDAD:** Finalmente, hay casos en que resulta imposible estudiar a todos los elementos de una población, por lo que se recurre a la muestra. Sería el caso, por ejemplo, del biólogo que desea descubrir rasgos de comportamiento de alguna especie animal como podría ser su longevidad, o sea, los años que vive. No es posible investigar uno a uno a todos los animales, por lo que se recurre a una muestra y de ellos se sacan conclusiones que se extienden a toda la población.

Las dos características más importantes que debe tener una muestra son: uno, [ser aleatorias](#), esto es que cada miembro de la población tenga igual oportunidad de salir en la muestra, o sea la misma probabilidad de ser escogido; y, dos, [ser representativas](#) de la población.

Al proceso de seleccionar o de obtener una muestra se le llama *muestreo*. La teoría del muestreo abarca un capítulo completo dentro de la estadística; sin embargo, como esto queda fuera del presente programa, no se abundará en este tema más allá que la simple mención.

Una muestra, dicho en otras palabras, es un subconjunto de la población, es decir, es una parte del todo.

*LA ESTADÍSTICA DESCRIPTIVA es la que analiza a toda la población y sus conclusiones están en función de los rasgos o datos extraídos de dicha población. Es la que trabaja con toda la población, no con muestras.*

Por ejemplo, el promedio de calificaciones del grupo "A" del 3<sup>er</sup> semestre en el mes de noviembre fue de 7.12 en Matemáticas. Esta conclusión se debe a que se consideraron las calificaciones de todos los alumnos, no de unos cuantos. Es decir, el promedio obtenido de 7.12 está describiendo en forma real y exacta a ese grupo.

Otro ejemplo: al efectuar el censo en la ciudad, se obtuvo que el 23% de la población no había terminado sus estudios de secundaria. Como el censo se realiza casa por casa, es decir, considerando uno por uno a todos los habitantes, la conclusión anterior describe en forma real a esa población.

*LA ESTADÍSTICA INFERENCIAL es la que a partir del análisis de una muestra, sus conclusiones se generalizan sobre todos los individuos de la población.*

Por ejemplo, se analizan 300 animales de la misma especie y de ellos se obtiene que la edad promedio de vida de dicha especie es de 12 años, a pesar de que no se estudiaron todos los animales de tal especie.

Otro ejemplo: en una muestra aleatoria de 450 transistores producidos en una fábrica, 14 de ellos resultaron defectuosos. Se concluye, es decir se infiere, que el 3.1% de los 12 000 transistores que forman la producción total deben, o pueden, ser defectuosos.

De tal forma que la estadística sirve fundamentalmente para tres cosas: primero, para describir el comportamiento o ciertos rasgos de una población a partir del estudio realizado a cada uno de los

integrantes de dicha población; segundo, para hacer generalizaciones sobre toda una población a partir de rasgos comunes en una muestra; y, tercero, para predecir características de sucesos futuros.

Debe entenderse que cuando se habla de "predicción" es en cierto sentido un tanto simbólica, pues la interpretación de los datos no va a llevar a contemplar de manera infalible el futuro, sino simplemente con un alto grado de certeza que así sucederá.

*Una variable es **CONTINUA** cuando puede tomar cualquier valor dentro un rango especificado, es decir, valores enteros y no enteros.*

Por ejemplo, si  $x$  representa el peso de las personas,  $x$  es una variable continua porque puede tomar cualquier valor dentro de un rango específico, como  $20 \leq x \leq 120$ . Significa que la variable  $x$  puede tomar cualquier valor entre 20 y 120, como  $x = 35$ ;  $x = 56.3$ ;  $x = 77.8$ , etc.

*Una variable es **DISCRETA** o **DISCONTINUA** cuando toma valores no continuos, es decir valores enteros solamente.*

Por ejemplo, el número de personas con casa propia no puede ser 4.5, sino que debe ser un número entero necesariamente. El número de materias reprobadas de un alumno es una variable discreta porque debe ser un número entero, no puede ser decimal.

***UN PARÁMETRO** es todo valor numérico obtenido a partir de una característica que sea medible de una población.*

Por ejemplo, en una fábrica de 970 trabajadores, el salario promedio es de \$7500.00 mensuales. Se trata de un parámetro porque ese dato numérico resultó de un análisis hecho a una población.

***UN ESTADÍSTICO** es todo valor numérico obtenido a partir de una característica que sea medible de una muestra.*

Por ejemplo, un medicamento provoca reacción alérgica en tres de cada veinte pacientes que lo toman. Es un estadístico porque esa información numérica resultó de un estudio sobre una muestra.

### CUESTIONARIO 1

- 1) ¿Qué es la estadística?
- 2) ¿Cuáles son las cuatro etapas de la estadística?
- 3) ¿Qué es una población?
- 4) Citar dos ejemplos, diferentes a los del libro, de poblaciones.
- 5) ¿Qué es una muestra?
- 6) Citar dos ejemplos, diferentes a los del libro, de muestras.
- 7) Citar un ejemplo, diferente al del libro, en el que no resulta costeable analizar a todos los individuos de una población.
- 8) Citar un ejemplo, diferente al del libro, en el que no se pueden analizar a todos los individuos de una población porque ésta se destruiría.
- 9) Al conjunto universal, es decir, al todo, al que abarca a todos los individuos, ¿cómo se le llama?
- 10) Al subconjunto de la población, ¿cómo se le llama?
- 11) ¿Qué es la estadística descriptiva?
- 12) ¿Qué es la estadística inferencial?
- 13) Cuando se hace estadística analizando a todos los individuos de una población, ¿qué tipo de estadística es?
- 14) Cuando se hace estadística analizando una muestra de la que se generaliza hacia toda la población, ¿qué tipo de estadística es?
- 15) ¿Qué es una variable continua?
- 16) ¿Qué es una variable discreta?
- 17) ¿Qué es un estadístico?