



---

---

## CONCEPTOS PRELIMINARES

---

---

### 1.1 Naturaleza de la Estadística.

En diversos campos de las ciencias sociales y naturales, el mundo de hoy se enfrenta a un volumen de información que cada vez va en aumento y que es necesario manejar ágil y eficientemente. En muchos casos, la estadística se constituye en una buena alternativa para hacerlo. Evidentemente, la estadística está presente en muchas actividades de la vida diaria de los individuos y por ello se hace necesario que el ciudadano común y corriente maneje de manera apropiada ideas y conceptos básicos del lenguaje estadístico y conozca y comprenda algo acerca del razonamiento estadístico. Por ejemplo, se requiere que pueda usar las herramientas básicas de la estadística como apoyo para asimilar, criticar y contrastar la información recibida y que además las pueda aplicar en el campo del saber donde desarrollará su trabajo.

Algunas veces un análisis estadístico comienza con un conjunto de datos. Por ejemplo, el gobierno colecciona y publica con cierta regularidad datos acerca de los porcentajes de desempleo y de los precios de artículos de consumo básicos de los ciudadanos. La estadística se debe utilizar entonces para resumir y analizar estos datos. Sin embargo, en otras situaciones no hay datos disponibles, entonces en estos casos la estadística también se puede utilizar para diseñar un experimento apropiado para generar datos. El experimento seleccionado dependerá del uso que se quiera hacer de los datos. Bien sea que sean generados a partir de un diseño experimental, o que simplemente estén disponibles o se obtengan de observaciones naturales no planeadas de manera experimental, los datos finalmente se deben describir. Para la descripción de los datos la estadística proporciona diferentes conceptos y sistemas de representación. Las representaciones gráficas y los arreglos tabulares son dos medios muy utilizados para presentar y resumir datos; por otra parte, la localización, la centralidad, la dispersión y la correlación son, algunas de las nociones y conceptos relevantes de la estadística. El uso y aplicación de estos conceptos y representaciones se suele acompañar y materializar en la generación de gráficos de distribuciones y el cálculo e interpretación de medidas de resumen como porcentajes, promedios o desviaciones, todas ellas, estadísticas que se constituyen en diferentes tipos de descripción que se pueden calcular con el fin de caracterizar diversos datos.

### 1.2 ¿Qué es la Estadística?

Hay muchas definiciones posibles de estadística, y una breve revisión de algunas de las dadas por diferentes personas que han aportado trabajos relevantes a la estadística pone de manifiesto el carácter cambiante de su definición en la historia y por ende de dar una definición precisa de ella. En particular, Ross (1996), presenta una selección interesante de algunas definiciones que vale la pena considerar; específicamente, para él “la estadística es el arte de aprender de los datos y trata con la recolección de los datos, su subsiguiente descripción y su análisis, el cual frecuentemente lleva a unas conclusiones”

En 1849, Quetelet (1796-1874) consideraba que la estadística tiene como objetivo presentar una fiel representación de un estado en una época determinada; en esa época todavía predominaba una visión de la estadística asociada al manejo del estado o gobierno de un país. Para Fisher (1890-1962) en 1925, la estadística se podía ver como el estudio de poblaciones, de la variación, o como el estudio de los métodos de reducción de datos; ya al entrar el siglo 20, la concepción de la visión de la estadística se amplía en su aplicación a las ciencias naturales. Para Pearson (1857-1936), en 1936, la estadística era una disciplina científica que trataba con la recolección, análisis e interpretación de datos obtenidos de la observación o experimentación. Esta disciplina tiene una estructura coherente que se fundamenta en la teoría de la probabilidad e incluye muchos diferentes procedimientos que contribuyen a la investigación y desarrollo de la ciencia y la tecnología. W. Weaver (1894-1978) hacia el año 1952, dijo que estadística es el nombre para la ciencia y el arte que trata con inferencias inciertas que utilizan números para conseguir información acerca de alguna cosa de la naturaleza o de la experiencia. Para Porter (1947-), en 1986, la estadística ha llegado a ser conocida en el siglo veinte como la herramienta matemática para el análisis datos de experimentos u observaciones.

Sin embargo, siguiendo a Batanero y Godino (2004), dos definiciones que reflejan de manera más apropiada una concepción más actual del tema son:

"La estadística estudia el comportamiento de los fenómenos llamados de colectivo. Está caracterizada por una información acerca de un colectivo o universo, lo que constituye su objeto material; un modo propio de razonamiento, el método estadístico, lo que constituye su objeto formal y unas previsiones de cara al futuro, lo que implica un ambiente de incertidumbre, que constituyen su objeto o causa final." (Cabriá, 1994).

"La estadística es la ciencia de los datos. Con más precisión, el objeto de la estadística es el razonamiento a partir de datos empíricos. La estadística es una disciplina científica autónoma, que tiene sus métodos específicos de razonamiento. Aunque es una ciencia matemática, no es un subcampo de la Matemática. Aunque es una disciplina metodológica, no es una colección de métodos". (Moore, 1995).

### 1.3 Tipos de Estadística.

Es difícil dividir la estadística en partes separadas, sin embargo, una división clásica hasta hace pocos años ha sido distinguir entre estadística descriptiva y estadística inferencial.

En la **estadística descriptiva** se tiene como fin presentar resúmenes de un conjunto de datos y poner de manifiesto sus características, principalmente mediante representaciones gráficas. La descripción de datos se usa para fines comparativos, y no suele utilizar principios de probabilidad. El interés se centra en describir el conjunto de datos y no se plantea el extender las conclusiones a otros datos diferentes o a una población.

Por el contrario, la **inferencia estadística** estudia los resúmenes de datos con referencia a un modelo de tipo probabilístico. Se supone que el conjunto de datos analizados es una muestra de una población y el interés principal es predecir el comportamiento de la población, a partir de los resultados de la muestra.

Actualmente, las capacidades de cálculo y representación gráfica de los computadores y también de las calculadoras graficadoras posibilitan la obtención de una amplia variedad de gráficos y cálculos estadísticos de una forma sencilla. La tecnología actual con sus posibilidades interactivas, favorecen la introducción, desde los primeros niveles de enseñanza, de la "filosofía" basada en los estudios estadísticos

introducida por Tukey (1977): el **Análisis Exploratorio de Datos** (A. E. D.). Esta última es una perspectiva de análisis intermedia entre la estadística descriptiva y la inferencia y le da un papel bien importante a la visualización por medio de diferentes gráficos.

Bajo esta “filosofía” se presta especial atención al problema de la interpretación de resultados y la generación de hipótesis sobre el problema investigado, a partir de los resultados de los análisis iniciales. Además, esta “filosofía” no sólo se aplica a nivel de estadística elemental. En muchos de los métodos del análisis de datos multivariantes utilizados actualmente también se aplica esta filosofía, para analizar fenómenos físicos o sociales complejos (Godino y Batanero, 1994).

#### 1.4 Variables en Estadística.

La definición de lo que es una variable y los tipos de variables que se suelen distinguir al utilizar la estadística, es uno de los asuntos que debe tener claro un analista de datos. Además, en relación con los tipos de variables es relevante discutir la idea de medición y las escalas o niveles en los que se puede medir una variable. Una **variable estadística** es una característica cuantitativa o cualitativa que se mide o se observa en una población. Para clasificar las variables en Estadística es conveniente hacer tres distinciones:

- a) **Desde lo continuo y discreto:** Se dice que una variable es continua si toma cualquiera de sus valores en un intervalo de números reales, en la unión de intervalos de números reales o, en general, en el conjunto de los números reales. Como ejemplo de este tipo de variable se puede citar la longitud de los tornillos producidos en una fábrica, la estatura de los estudiantes de una escuela, el tiempo gastado por un estudiante para transcribir una página, la distancia recorrida por un ciclista en una hora, etc. Por otra parte, cuando es posible contar los valores que puede tomar una variable o ésta solo toma valores enteros se dice que la variable es discreta. Por ejemplo, el número de estudiantes por curso de un colegio, el número de hijos de cada uno de los empleados de una empresa, la cantidad de autos accidentados en cada departamento de Colombia, la cantidad de asignaturas reprobadas por los estudiantes de un grado particular, etc.
- b) **Desde lo cuantitativo y cualitativo:** Se dice que una variable es **cualitativa**, cuando los valores que toma son atributos o cualidades de los objetos. Por ejemplo: El “estado civil” induce una clasificación natural (soltero, casado, divorciado o separado); los grados o insignias de los oficiales del ejército colombiano permiten hacer un escalafón de acuerdo a la OTAN (General, Mayor general, Brigadier general, Coronel, Teniente coronel, Mayor, Capitán, Teniente y Subteniente); el NIT de las instituciones de educación superior en Colombia; El nombre de las diferentes dependencias u oficinas de una universidad, etc. En contraste, se dice que una variable es **cuantitativa**, cuando los valores que toma son numéricos y representan cantidades con las que se pueden realizar operaciones y comparaciones. Por ejemplo: la estatura de los estudiantes de una escuela, el peso de los niños que ingresan a una clínica pediátrica, el salario de los docentes del distrito, la cantidad de docentes de planta de las diferentes universidades colombianas, etc.
- c) **Desde lo determinístico y aleatorio:** Una variable es **determinística** si su valor corresponde al resultado o salida de un proceso o experimento determinístico y es singularmente determinado por

una entrada dada. Están ligadas a las funciones y fórmulas usadas en las diferentes áreas del conocimiento. Por ejemplo: El volumen de una esfera dada, la velocidad de una partícula que recorre una distancia  $d$  en un tiempo  $t$ , el determinante de una matriz de tamaño  $2 \times 2$ , la distancia recorrida por una pelota de béisbol lanzada a una velocidad inicial  $v$  con un ángulo de inclinación de  $45^\circ$ , las raíces de una ecuación cuadrática dada, etc. Una variable es **aleatoria** cuando sus resultados están asociados a experimentos aleatorios o procesos en donde los resultados son impredecibles. Formalmente son funciones cuyo dominio es el conjunto de los posibles resultados de un experimento aleatorio o probabilístico y cuyo rango es un conjunto de números reales, a su vez se clasifican en discretas, continuas y conjuntas. Por ejemplo: Cantidad de aciertos en un test de 20 preguntas de selección múltiple, la suma de los puntos resultantes cuando se lanzan cuatro dados distinguibles, la cantidad de sellos que aparecen cuando se lanzan tres monedas distinguibles, la cantidad de personas que se deben examinar para encontrar cuatro donantes con sangre A Rh positivo, la cantidad de taxis que pasarán por la intersección de dos avenidas entre las 4:00 pm y las 6:00 pm de un día determinado, etc.

## 1.5 Escalas de medición.

En cuanto a los niveles o escalas en que se puede medir una variable se hará mención a cuatro niveles: nominal, ordinal, de intervalo y de razón.

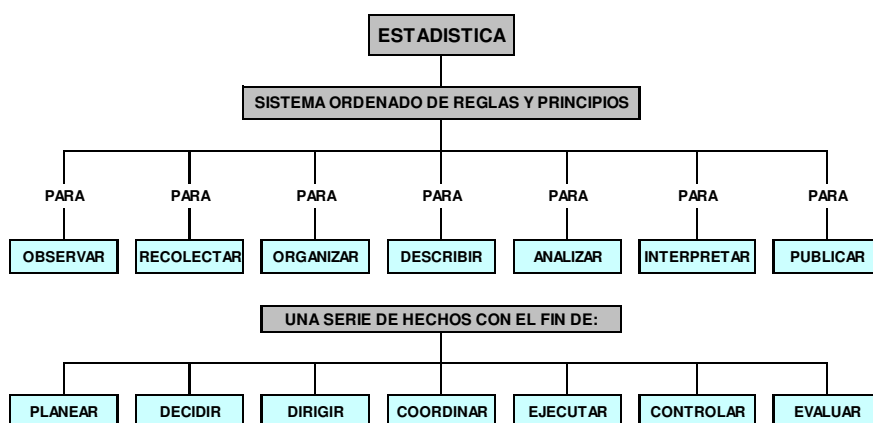
- a) **Nivel de medición nominal:** Es aquel en el que sólo se puede manifestar una relación de pertenencia a las categorías o valores que puede asumir la variable. Se debe señalar que entre los valores de una variable de tipo nominal, no se puede establecer ninguna jerarquía, no se puede trazar ningún ordenamiento. Por ejemplo: el tipo de sangre de los empleados de una empresa (A, AB, B, O), el credo religioso de los gerentes de las empresas privadas del continente (Cristianismo, Taoismo, Budismo, Induismo, Islamismo, Judaísmo, Zoroastrianismo, Confucionismo, Shintoismo, Jainismo, Sikhismo), el cereal cultivado en las diferentes regiones del país (trigo, maíz, centeno, sorgo, cebada, avena, arroz, amaranto, mijo, quinua, etc.), el número de documento de identidad de los nuevos graduados de una universidad, etc.
- b) **Nivel de medición ordinal:** En este se distinguen los diferentes valores de la variable jerarquizándolos simplemente de acuerdo a un rango. Se establece que existe una gradación entre uno y otro valor de la escala, de tal modo que cualquiera de ellos es mayor que el precedente y menor que el que le sigue. Sin embargo, la distancia entre un valor y otro no queda definida sino que es indeterminada. Por ejemplo: la posición de los competidores al finalizar una carrera automovilística (Primero, segundo, tercero, cuarto, etc.), el estrato socioeconómico de los deportistas de la ciudad (uno, dos, tres, etc.), el máximo nivel de formación académica alcanzado por los artistas colombianos (primaria, bachillerato, técnico, tecnólogo, profesional, especialista, magíster, doctorado, postdoctorado).
- c) **Nivel de medición de intervalo:** Es aquel en el que además de poseerse la equivalencia de categorías y el ordenamiento interno entre ellas, se tiene la característica de que la distancia entre dos resultados

cualesquiera está claramente determinada. La diferencia entre dos mediciones tiene sentido y permite comparar las unidades estudiadas. Ejemplos típicos de medición a nivel de intervalo son las escalas termométricas, la temperatura de una persona, el nivel del aceite del motor de un automóvil medido con una varilla graduada, sobrepeso de las personas que asisten a un gimnasio con respecto a una tabla de medidas internacionales, etc.

- d) **Nivel de medición de razón:** Se caracteriza por conservar las propiedades de los casos anteriores pero además se le añade la existencia de un valor cero que tenga un sentido real, con lo que se hacen posibles ciertas operaciones matemáticas. El cociente entre dos mediciones tiene un significado interpretable en función de una proporción o razón entre las dos magnitudes. Esto quiere decir que un valor de 20 en una escala de este tipo es el doble de un valor de 10, o de las dos terceras partes de un valor de 30. Ejemplos de variables medidas en este nivel son la longitud, la masa, la intensidad de corriente eléctrica, cantidad de litros diarios de agua que consumen los deportistas que asisten a los juegos olímpicos, la cantidad de goles anotados por los delanteros de los equipos profesionales del fútbol colombiano, etc.

## 1.6 Ejercicios.

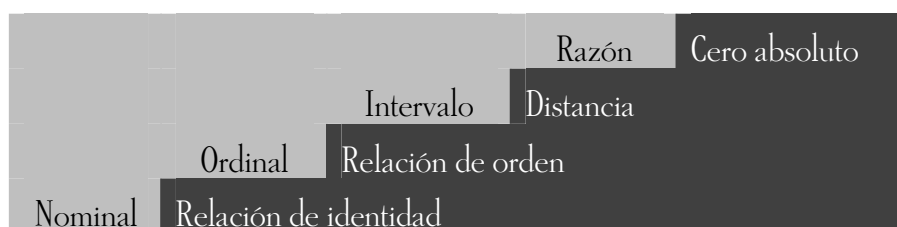
1. Observe cuidadosamente el siguiente diagrama y sugiera una interpretación del mismo para proponer una definición de lo que significa estadística.



2. ¿De qué forma, desde su quehacer diario, puede utilizar la Estadística? Dé su respuesta en términos del diagrama.
3. ¿Qué uso ha observado que tiene la estadística en la información que suministra la prensa escrita? ¿Cómo considera este uso?
4. ¿Por qué cree que es importante que un estudiante de educación básica y educación media tenga conocimientos de estadística?
5. ¿Una variable continua se puede considerar que es una variable cualitativa? Explique.
6. Un experimento aleatorio consiste en lanzar 5 monedas de la misma nominación pero distinguibles, y anotar la figura resultante en cada una de ellas (cara o Sello). Elabore una lista con los posibles

resultados. Elabore una tabla para la función  $X$  o variable aleatoria  $X$  definida como el número Caras resultantes.

7. Un experimento aleatorio consiste en lanzar 5 dados normales distinguibles, y anotar los números resultantes en la cara superior de cada uno de ellos. Elabore una tabla para la función  $X$  o variable aleatoria  $X$  definida como la suma resultante. Elabore una tabla para la función  $Y$  o variable aleatoria  $Y$  definida como la cantidad de “5” resultantes.
8. De un grupo de 5 deportistas, entre los cuales hay 3 beisbolistas y 2 ciclistas, se eligen al azar 3 de ellos y se anotan sus nombres. Elabore una lista de las posibles elecciones o selecciones. Elabore una tabla para la variable aleatoria  $X$  definida como el número de beisbolistas que resulta en una selección.
9. Un experimento consiste en lanzar un dado dos veces y anotar la pareja de números  $(X,Y)$ , donde  $X$  es la cantidad de pares que resultan en el primer lanzamiento,  $Y$  es la cantidad de pares entre los dos lanzamientos. Escriba la población de posibles parejas resultantes.
10. Un experimento consiste en lanzar cinco monedas distinguibles y anotar el resultado que aparece en la cara visible de cada moneda. Las monedas no tienen Caras y Sellos, sino “Cuatros” y “Cincos”. Escriba la población de posibles resultados y las posibles sumas resultantes.
11. En un campeonato de ajedrez participan ocho jugadores: Lasker, Alekhine, Capablanca, Botvinnik, Fischer, Spaski, Kárpov, Kaspárov. Determine la población de posibles partidos que deben realizarse, en cada una de las siguientes metodologías de eliminatoria: (a) Juegan todos contra todos un solo partido. (b) Juegan todos contra todos dos partidos: uno con fichas blancas y otro con negras. (c) Se juegan 4 partidos en la primera ronda y se eliminan cuatro, luego se juegan dos partidos en la segunda ronda y se eliminan dos, y en la tercera ronda se define el campeón.
12. Interprete la siguiente representación acerca de las escalas de medición.



13. Clasifique las siguientes variables según su tipo y nivel de medición.
  - a) Número de estafilococos por mililitro.
  - b) Número de habitantes por kilómetro cuadrado.
  - c) Número de ciruelas por árbol.
  - d) Duración de cierta marca de tubos fluorescentes.
14. Dé otros ejemplos de variables que se puedan ubicar en cada una de las escalas de medición consideradas en este capítulo.

15. Discutir y establecer el nivel de medición de los siguientes casos:

- a) Clases de bacterias en una muestra de agua contaminada.
- b) Latitud de una ciudad.
- c) Rango militar en el ejército colombiano.
- d) Intensidad de corriente eléctrica.
- e) Densidad de diferentes muestras de un mismo líquido.
- f) Tamaño de gaseosa que se vende en restaurantes de comida rápida.
- g) Velocidad de un automóvil al pasar por la calle 100 con 7ª.
- h) Salario mensual de los profesores de cierta universidad.
- i) Tiempo de duración de cierta marca de tubos fluorescente.
- j) Estrato socio-económico indicado en el recibo de energía.
- k) Temperatura de Bogotá a las 12 meridiano.
- l) Ubicación temporal de un acontecimiento histórico.
- m) Tipos de productos enlatados para la alimentación.
- n) Altitud de ciudades de Colombia.
- o) Número telefónico de los estudiantes de Estadística.
- p) Primer beneficiario de una póliza de seguros.
- q) Día de la semana en la que más estudia un estudiante.
- r) Talla de calzado que utiliza un estudiante.
- s) Grado de escolaridad de un trabajador.
- t) Hora GMT.
- u) Medición de un taxímetro.
- v) Sintonía de un programa de televisión.
- w) Índice de masa corporal.
- x) Magnitud de los terremotos ocurridos en los últimos cinco años.
- y) Intensidad del sonido de los instrumentos musicales.
- z) Puntaje obtenido en un Test de medición de CI.

