

1.1. El concepto de Estadística.

¿Qué es y para qué sirve?

La Estadística se ocupa de la **recolección, agrupación, presentación, análisis e interpretación de datos.**

A menudo se llaman estadísticas a las listas de estos datos, cosa que crea una cierta ambigüedad, que no debería originarnos confusiones.

La Estadística no son sólo los resultados de encuestas, ni el cálculo de unos porcentajes, la Estadística es un **método científico que pretende sacar conclusiones a partir de unas observaciones hechas.**

¿Cuándo empezó la Estadística?

La Estadística actual es el resultado de la **unión de dos disciplinas** que evolucionaron de forma independiente hasta confluir en el siglo XIX:

- el **Cálculo de Probabilidades**, que nació en el siglo XVII como la teoría matemática de los juegos de azar,
- la “**Estadística**”, o **ciencia del Estado**, que estudia la descripción de datos, y que tiene unas raíces más antiguas, de hecho, tan antiguas como la humanidad (censos de población).

La interacción de ambas líneas de pensamiento da lugar a la **ciencia que estudia cómo obtener conclusiones de la investigación empírica mediante el uso de modelos matemáticos.**

Resumiendo:

La Estadística actúa como **disciplina puente entre los modelos matemáticos y los fenómenos reales.**

Un modelo matemático es una abstracción simplificada de una realidad más compleja y siempre existirá una cierta discrepancia entre lo que se observa y lo previsto por el modelo.

La Estadística proporciona una metodología para evaluar y juzgar estas discrepancias entre la realidad y la teoría.

1.2. Conceptos generales.

Algunas definiciones:

Población estadística: conjunto finito o infinito de elementos, denominados **individuos**, sobre los cuales se realizan observaciones. Ejemplos: *todos los habitantes de cierto lugar, todos los ejemplares de una determinada especie de tortugas, todos los microchips que fabrica una empresa, etc.*

Muestra: subconjunto finito de una población. El número de individuos que forman la muestra se denomina **tamaño muestral**.

Variable o **carácter**: cada una de las características que pueden observarse en un individuo de la muestra. Ejemplos: en una muestra de una población de seres humanos podemos medir: *la altura, la edad, el peso, el sexo, número de hermanos...*; en una muestra de una población de una especie de tortugas podemos medir: *la anchura del caparazón, la longitud del caparazón, la edad...*;

Tipos de variables:

Cualitativas, categóricas (o alfanuméricas):

Pueden tomar valores no cuantificables numéricamente.

Se denomina **categoría** a cada uno de los valores que toma la variable.

Cuantitativas (o numéricas):

Pueden tomar valores cuantificables numéricamente.

Nominales: si no existe ningún orden entre las categorías de la variable. Ejemplos: *el grupo sanguíneo (A ,B ,AB, O); el color de los ojos (azules, verdes, marrones, negros),...*

Hay que distinguir las variables **binarias**, aquéllas que sólo toman dos valores posibles (sí/no, presencia/ausencia de cierto carácter), dentro de las nominales. Ejemplo: *el sexo, ser fumador, tener carné de conducir, ser daltónico,...*

Ordinales: cuando existe un cierto orden entre las categorías de la variable. Ejemplo: *el nivel de estudios (sin estudios, básicos, medios, superiores), el grado de miopía (ausencia, bajo, medio, alto),...*

Discretas: si solamente toman valores aislados (generalmente enteros). Suelen corresponder a contajes. Ejemplos: *el número de hermanos, el número de cafés/día, el número de multas/año,...*

Continuas: potencialmente puede tomar cualquier valor numérico dentro de un intervalo o de una unión de intervalos. Ejemplos: *el tiempo de reacción a un cierto medicamento, el peso de un individuo, la longitud del caparazón de una tortuga,...*

Observaciones:

1. Las variables estadísticas pueden ordenarse según el grado de información que proporcionan:



2. Asignación de **etiquetas** a las categorías de una variable categórica.

Por ejemplo, en la variable “grupo sanguíneo”

A → 1, B → 2, AB → 3, O → 4.

Los valores 1, 2, 3 y 4, no tienen ninguna carga numérica significativa y sirven sólo para **representar las diferentes categorías de la variable**.