

## **Introducción:**

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

1º Como colección de datos numéricos.- Esto es el significado más vulgar de la palabra estadística. Se sobreentiende que dichos 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.

Tenemos muchos ejemplos de este tipo de estadísticas. El Anuario Estadístico publicado por el Instituto Nacional de Estadística, El Anuario de Estadísticas del Trabajo,...

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, iniciaremos el trabajo tomando un grupo numeroso de nacimientos y obtener después la proporción de varones. Es muy frecuente enfrentarnos con fenómenos en los que es muy difícil predecir el resultado; así, no podemos dar una lista ,con las personas que van a morir con una cierta edad, o el sexo de un nuevo ser hasta que transcurra un determinado tiempo de embarazo,...

Por tanto, el objetivo de la estadística es hallar las regularidades que se encuentran en los fenómenos de masa.

## **Población, elementos y caracteres.**

Es obvio que todo estudio estadístico ha de estar referido a un conjunto o colección de personas o cosas. Este conjunto de personas o cosas es lo que denominaremos **población**.

Las personas o cosas que forman parte de la población se denominan **elementos**. En sentido estadístico un elemento puede ser algo con existencia real, como un automóvil o una casa, o algo más abstracto como la temperatura, un voto, o un intervalo de tiempo.

A su vez, cada elemento de la población tiene una serie de características que pueden ser objeto del estudio estadístico. Así por ejemplo si consideramos como elemento a una persona, podemos distinguir en ella los siguientes **caracteres**:

Sexo, Edad, Nivel de estudios, Profesión, Peso, Altura, Color de pelo,Etc.

Luego por tanto de cada elemento de la población podremos estudiar uno o más aspectos cualidades o caracteres.

La población puede ser según su tamaño de dos tipos:

**Población finita:** cuando el número de elementos que la forman es finito, por ejemplo el número de alumnos de un centro de enseñanza, o grupo clase.

**Población infinita:** cuando el número de elementos que la forman es infinito, o tan grande que pudiesen considerarse infinitos.. Como por ejemplo si se realizase un estudio sobre los productos que hay en el mercado. Hay tantos y de tantas calidades que esta población podría considerarse infinita.

Ahora bien, normalmente en un estudio estadístico, no se puede trabajar con todos los elementos de la población sino que se realiza sobre un subconjunto de la misma. Este subconjunto puede ser **una muestra**, cuando se toman un determinado número de elementos de la población, sin que en principio tengan nada en común; o **una subpoblación**, que es el subconjunto de la población formado por los elementos de la población que comparten una determinada característica, por ejemplo de los alumnos del centro la subpoblación formada por los alumnos de 3º ESO, o la subpoblación de los varones.

### **Variables y atributos.**

Como hemos visto, los caracteres de un elemento pueden ser de muy diversos tipos, por lo que los podemos clasificar en: dos grandes clases:

Variables **Cuantitativas**.

Variables **Cualitativas o Atributos**.

Las variables cuantitativas son las que se describen por medio de números, como por ejemplo el peso, Altura, Edad, Número de Suspensos...

A su vez este tipo de variables se puede dividir en dos subclases:

- **Cuantitativas discretas**. Aquellas a las que se les puede asociar un número entero, es decir, aquellas que por su naturaleza no admiten un fraccionamiento de la unidad, por ejemplo número de hermanos, páginas de un libro, etc.
- **Cuantitativas continuas**: Aquellas que no se pueden expresar mediante un número entero, es decir, aquellas que por su naturaleza admiten que entre dos valores cualesquiera la variable pueda tomar cualquier valor intermedio, por ejemplo peso, tiempo. etc.

No obstante en muchos casos el tratamiento estadístico hace que a variables discretas las trabajemos como si fuesen continuas y viceversa.

Los atributos son aquellos caracteres que para su definición precisan de palabras, es decir, no le podemos asignar un número. Por ejemplo Sexo Profesión, Estado Civil, etc.

A su vez las podemos clasificar en:

- **Ordenables**: Aquellas que sugieren una ordenación, por ejemplo la graduación militar, El nivel de estudios, etc.
- **No ordenables**: Aquellas que sólo admiten una mera ordenación alfabética, pero no establece orden por su naturaleza, por ejemplo el color de pelo, sexo, estado civil, etc.

### **Distintos Tipos de Frecuencia:**

Una de los primeros pasos que se realizan en cualquier estudio estadístico es la tabulación de resultados, es decir, recoger la información de la muestra resumida en una tabla en la que a cada valor de la variable se le asocian determinados números que representan el número de veces que ha aparecido, su proporción con respecto a otros valores de la variable, etc. Estos números se denominan *frecuencias*: Así tenemos los siguientes tipos de frecuencia:

1. [Frecuencia absoluta](#)
2. [Frecuencia relativa](#)
3. [Porcentaje](#)

4. [Frecuencia absoluta acumulada](#)
  5. [Frecuencia relativa acumulada](#)
  6. [Porcentaje acumulado](#)
  7. [Ejemplo](#)
- 

### **Frecuencia absoluta:**

La frecuencia absoluta de una variable estadística es el número de veces que aparece en la muestra dicho valor de la variable, la representaremos por  $n_i$

### **Frecuencia relativa:**

La frecuencia absoluta, es una medida que está influida por el tamaño de la muestra, al aumentar el tamaño de la muestra aumentará también el tamaño de la frecuencia absoluta. Esto hace que no sea una medida útil para poder comparar. Para esto es necesario introducir el concepto de *frecuencia relativa*, que es el cociente entre la frecuencia absoluta y el tamaño de la muestra. La denotaremos por  $f_i$

$$f_i = \frac{n_i}{N}$$

Donde  $N$  = Tamaño de la muestra

### **Porcentaje:**

La frecuencia relativa es un *tanto por uno*, sin embargo, hoy día es bastante frecuente hablar siempre en términos de *tantos por ciento o porcentajes*, por lo que esta medida resulta de multiplicar la frecuencia relativa por 100. La denotaremos por  $p_i$ .

$$p_i = f_i \cdot 100 \%$$

### **Frecuencia Absoluta Acumulada:**

Para poder calcular este tipo de frecuencias hay que tener en cuenta que la variable estadística ha de ser cuantitativa o cualitativa ordenable. En otro caso no tiene mucho sentido el cálculo de esta frecuencia. La frecuencia absoluta acumulada de un valor de la variable, es el número de veces que ha aparecido en la muestra un valor menor o igual que el de la variable y lo representaremos por  $N_i$ .

### **Frecuencia Relativa Acumulada:**

Al igual que en el caso anterior la frecuencia relativa acumulada es la frecuencia absoluta acumulada dividido por el tamaño de la muestra, y la denotaremos por  $F_i$

$$F_i = \frac{N_i}{N}$$

### **Porcentaje Acumulado:**

Análogamente se define el Porcentaje Acumulado y lo vamos a denotar por  $P_i$  como la frecuencia relativa acumulada por 100.

$$P_i = F_i \cdot 100 \%$$

Veamos esto con un ejemplo: Tomamos para ello los datos relativos a las personas activas.

Personas Activas	Número Familias					
Xi	ni	Fi	pi	Ni	Fi	Pi
1	16	16/50	32%	16	16/50	32%
2	20	20/50	40%	36	36/50	72%
3	9	9/50	18%	45	45/50	90%
4	5	5/50	10%	50	50/50	100%
Total	50					

En este ejemplo se puede ver fácilmente como se calculan estas frecuencias.

### Accidentes estadísticos



El naufragio del *Prestige* no ha sido una desgraciada casualidad. Se trata, por contra, de una cuestión estadística: cada año pasan frente a las costas gallegas unos 13.000 barcos con mercancías peligrosas.

En los últimos 30 años, el litoral de la comunidad ha sufrido siete de las once mayores catástrofes marítimas en Europa. Galicia no ha tomado nota de este dato y la realidad se ha empeñado en demostrar, nuevamente, que no dispone ni de los medios técnicos ni de la formación para afrontar un trágico episodio que se repite, aproximadamente, cada cuatro años. Para evitar estos accidentes es necesario tener en cuenta la información estadística y adoptar un plan de acción preventiva específico, basado en ella. Si no, la estadística nos dice que el siguiente *Prestige* ya está a punto de zarpar.

### La modernización de los datos



Cada día existe una mayor demanda de información para áreas geográficas pequeñas. A los tradicionales demandantes, administración pública e investigadores, se ha sumado el sector privado, ya que las empresas quieren saber cómo son sus potenciales clientes, para dirigir sus campañas de publicidad, y abren establecimientos en zonas donde la población tiene determinada edad y condición socioeconómica.

La propia administración precisa de información muy desagregada para sus tareas de planificación. Esta demanda genera peticiones de información de un volumen considerable, por ejemplo, la simple pregunta sobre los datos de la población por grupos quinquenales de edad y sexo, para todos los municipios de España, genera una respuesta que consiste en 300.000 datos, y, para secciones censales, supera el millón. ¡Y no es más que una pregunta! Unido al problema del volumen está el de la calidad, y el de la disponibilidad. Para que una fuente de datos permita gran desagregación territorial, deberá ser exhaustiva. El INE es un gran productor de información infraprovincial, ya que realiza, cada diez años, los censos Agrario, Edificios, Locales, Población y Viviendas, y en esta década, por primera vez, establecerá un plan de difusión para áreas pequeñas. Ello pasa por desarrollar una aplicación informática que gestione un callejero y lo relacione con la ingente cantidad de datos que ello exige.

### Ejercicios

#### Lectura de libros

Se pregunta a 15 alumnos por el número de libros que leen al año obteniéndose los siguientes resultados:

2 5 3 4 9 1 0 3 3 5 3 2 3 1 6

Calcula la frecuencia absoluta, la frecuencia relativa y el porcentaje de alumnos que leen 3 libros al año.

#### paso 1 →

La frecuencia absoluta es simplemente contar cuántos alumnos han respondido que leen 3 libros al año. Por tanto: Frecuencia absoluta= 5

#### paso 2 →

La frecuencia relativa tiene en cuenta no solo cuántos han respondido que leen tres libros al año sino también a cuánta gente hemos preguntado. Así:

$$f = \frac{5}{15} = \frac{1}{3}$$

#### paso 3 →

Calculo, por último el porcentaje de gente que ha respondido que lee 3 libros al año:

$$\% = \frac{5}{15} \cdot 100 \approx 33.3\%$$

### Deportistas

Se ha preguntado a 50 personas por el número de deportes que practican obteniéndose los siguientes resultados:

N.deportes	0	1	2	3	4	5	6
$F_i$	15	12	10	8	2	2	1

Calcula los porcentajes de gente que practican 0, 1, 2, 3, 4, 5 y 6 deportes respectivamente. ¿Cuánto suman estos porcentajes?

#### paso 1

Para calcular el porcentaje de gente que practica 0 deportes tengo que dividir el número de personas que no practica ningún deporte entre el número de personas a las que he hecho la encuesta y multiplicar el resultado por cien. Así.

$$\% (0 \text{ deportes}) = \frac{15}{50} \cdot 100 = 0.3\%$$

#### paso 2

De igual forma:

$$\% (1 \text{ deportes}) = \frac{12}{50} \cdot 100 = 0.24\%$$

$$\% (2 \text{ deportes}) = \frac{10}{50} \cdot 100 = 0.2\%$$

$$\% (3 \text{ deportes}) = \frac{8}{50} \cdot 100 = 0.16\%$$

$$\% (4 \text{ deportes}) = \frac{2}{50} \cdot 100 = 0.04\%$$

$$\% (5 \text{ deportes}) = \frac{2}{50} \cdot 100 = 0.04\%$$

$$\% (6 \text{ deportes}) = \frac{1}{50} \cdot 100 = 0.02\%$$

#### paso 3

Los porcentajes siempre han de sumar 1. No obstante lo compruebo:

$$0.3 + 0.24 + 0.2 + 0.16 + 0.04 + 0.04 + 0.02 = 1$$