

## TEMA 7: Estadística

### 7.1 CONCEPTOS BÁSICOS

La **Estadística** trata del recuento, ordenación y clasificación de los datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones.

- Un **estudio estadístico** consta de las siguientes fases:

Recogida de datos.

Organización y representación de datos.

Análisis de datos.

Obtención de conclusiones.

- **Población**

Una **población** es el conjunto de todos los elementos a los que se somete a un estudio estadístico.

- **Individuo**

Un **individuo** o **unidad estadística** es cada uno de los elementos que componen la población.

- **Muestra**

Una **muestra** es un conjunto representativo de la población de referencia, el número de individuos de una muestra es menor que el de la población.

- **Muestreo**

El **muestreo** es la reunión de datos que se desea estudiar, obtenidos de una proporción reducida y representativa de la población.

- **Valor**

Un **valor** es cada uno de los distintos resultados que se pueden obtener en un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos dos valores: cara y cruz.

- **Dato**

Un **dato** es cada uno de los valores que se ha obtenido al realizar un estudio estadístico. Si lanzamos una moneda al aire 5 veces obtenemos 5 datos: cara, cara, cruz, cara, cruz.

## **7.2 VARIABLE ESTADISTICA**

Una **variable estadística** es cada una de las **características o cualidades** que poseen los **individuos de una población**.

### **Tipos de variables**

#### **a) Variable cualitativa**

Las **variables cualitativas** se refieren a **características o cualidades** que **no** pueden ser medidas con **números**. Podemos distinguir dos tipos:

##### **Variable cualitativa nominal**

Una **variable cualitativa nominal** presenta **modalidades no numéricas** que **no** admiten un **criterio de orden**. Por ejemplo:

El estado civil, con las siguientes modalidades: soltero, casado, separado, divorciado y viudo.

##### **Variable cualitativa ordinal o variable cuasicuantitativa**

Una **variable cualitativa ordinal** presenta **modalidades no numéricas**, en las que existe un **orden**. Por ejemplo:

La nota en un examen: suspenso, aprobado, notable, sobresaliente.

Puesto conseguido en una prueba deportiva: 1º, 2º, 3º, ...

Medallas de una prueba deportiva: oro, plata, bronce.

#### **b) Variable cuantitativa**

Una **variable cuantitativa** es la que se expresa mediante un **número**, por tanto se pueden realizar **operaciones aritméticas** con ella. Podemos distinguir dos tipos:

##### **Variable discreta**

Una **variable discreta** es aquella que toma **valores aislados**, es decir **no** admite **valores intermedios** entre dos valores específicos. Por ejemplo:

El número de hermanos de 5 amigos: 2, 1, 0, 1, 3.

##### **Variable continua**

Una **variable continua** es aquella que puede tomar **valores comprendidos entre dos números**. Por ejemplo:

La altura de los 5 amigos: 1.73, 1.82, 1.77, 1.69, 1.75.

En la práctica medimos la altura con dos decimales, pero también se podría dar con tres decimales.

### **7.3 DISTRIBUCIÓN DE FRECUENCIAS**

La **distribución de frecuencias** o **tabla de frecuencias** es una **ordenación** en forma de **tabla** de los **datos estadísticos**, asignando a cada **dato** su **frecuencia correspondiente**.

#### **Tipos de frecuencias**

##### **a) Frecuencia absoluta**

La **frecuencia absoluta** es el **número de veces** que aparece un determinado **valor** en un estudio estadístico.

Se representa por  $n_i$ .

La **suma de las frecuencias absolutas** es igual al número total de datos, que se representa por  $N$ .

Para indicar resumidamente estas sumas se utiliza la letra griega  $\Sigma$  (sigma mayúscula) que se lee suma o sumatoria.

$$\sum_{i=1}^n n_i = N$$

##### **b) Frecuencia relativa**

La **frecuencia relativa** es el **cociente** entre la **frecuencia absoluta** de un determinado valor y el **número total de datos**.

Se puede expresar en tantos por ciento y se representa por  $n_i$ .

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

La suma de las frecuencias relativas es igual a 1.

$$\sum_{i=1}^n \frac{n_i}{N} = 1$$

### c) Frecuencia acumulada

La **frecuencia acumulada** es la **suma de las frecuencias absolutas** de todos los **valores inferiores o iguales al valor** considerado.

Se representa por  $N_i$ .

### d) Frecuencia relativa acumulada

La **frecuencia relativa acumulada** es el **cociente** entre la **frecuencia acumulada** de un determinado **valor** y el **número total de datos**. Se puede expresar en tantos por ciento.

### Ejemplo

Durante el mes de julio, en una ciudad se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 29, 29.

En la primera columna de la tabla colocamos la variable ordenada de menor a mayor, en la segunda hacemos el recuento y en la tercera anotamos la frecuencia absoluta.

$x_i$	Recuento	$n_i$	$N_i$	$f_i$	$F_i$
27	I	1	1	0.032	0.032
28	II	2	3	0.065	0.097
29	HHH I	6	9	0.194	0.290
30	HHH II	7	16	0.226	0.516
31	HHH III	8	24	0.258	0.774
32	III	3	27	0.097	0.871
33	III	3	30	0.097	0.968
34	I	1	31	0.032	1
		31		1	

Este tipo de **tablas de frecuencias** se utiliza con **variables discretas**.

### **7.3.1 Distribución de frecuencias agrupadas**

La **distribución de frecuencias agrupadas** o **tabla con datos agrupados** se emplea si las **variables** toman un **número grande de valores** o la **variable es continua**.

Se **agrupan** los **valores** en **intervalos** que tengan la **misma amplitud** denominados **clases**. A cada **clase** se le asigna su **frecuencia correspondiente**.

- **Límites de la clase**

Cada **clase** está **delimitada** por el **límite inferior de la clase** y el **límite superior de la clase**.

- **Amplitud de la clase**

La **amplitud de la clase** es la **diferencia** entre el **límite superior e inferior** de la **clase**.

- **Marca de clase**

La **marca de clase** es el **punto medio** de cada **intervalo** y es el **valor** que representa a todo el **intervalo** para el **cálculo** de algunos **parámetros**.

- **Construcción de una tabla de datos agrupados**

3, 15, 24, 28, 33, 35, 38, 42, 43, 38, 36, 34, 29, 25, 17, 7, 34, 36, 39, 44, 31, 26, 20, 11, 13, 22, 27, 47, 39, 37, 34, 32, 35, 28, 38, 41, 48, 15, 32, 13.

1° Se localizan los valores menor y mayor de la distribución. En este caso son 3 y 48.

2° Se restan y se busca un número entero un poco mayor que la diferencia y que sea divisible por el número de intervalos que queremos establecer.

Es conveniente que el número de intervalos oscile entre 6 y 15.

En este caso,  $48 - 3 = 45$ , incrementamos el número hasta  $50 : 5 = 10$  intervalos.

Se forman los intervalos teniendo presente que el límite inferior de una clase pertenece al intervalo, pero el límite superior no pertenece intervalo, se cuenta en el siguiente intervalo.

	$c_i$	$n_i$	$N_i$	$f_i$	$F_i$
[0, 5)	2.5	1	1	0.025	0.025
[5, 10)	7.5	1	2	0.025	0.050
[10, 15)	12.5	3	5	0.075	0.125
[15, 20)	17.5	3	8	0.075	0.200
[20, 25)	22.5	3	11	0.075	0.2775
[25, 30)	27.5	6	17	0.150	0.425
[30, 35)	32.5	7	24	0.175	0.600
[35, 40)	37.5	10	34	0.250	0.850
[40, 45)	42.5	4	38	0.100	0.950
[45, 50)	47.5	2	40	0.050	1
		<b>40</b>		<b>1</b>	

## **7.4 PARAMETROS ESTADÍSTICOS**

### **7.4.1 Definición de parámetro estadístico**

Un **parámetro estadístico** es un **número** que se obtiene a partir de los **datos** de una **distribución estadística**.

Los **parámetros estadísticos** sirven para sintetizar la información dada por una tabla o por una gráfica.

### **7.4.1 Tipos de parámetros estadísticos**

Hay **dos tipos parámetros estadísticos**:

De centralización.

De dispersión.

## a) Medidas de centralización

Nos indican en torno a qué valor (centro) se distribuyen los datos.

Las **medidas de centralización** son:

### Media aritmética

La **media** es el valor **promedio** de la distribución.

$$\bar{x} = \sum_{i=1}^n x_i \cdot f_i = \sum_{i=1}^n \frac{x_i \cdot n_i}{N}$$

### Mediana

La **mediana** es la **puntuación** de la escala que **separa la mitad superior** de la distribución y **la inferior**, es decir divide la serie de datos en **dos partes iguales**.

Si **N es Impar**, hay un término central, el término  $X_{\frac{N+1}{2}}$  que será el valor de la mediana.

Si **N es Par**, hay dos términos centrales,  $X_{\frac{N}{2}}$ ,  $X_{\frac{N}{2}+1}$  la mediana será la media de esos dos valores

### Moda

La **moda** es el **valor** que **más se repite** en una distribución.

## b) Medidas de dispersión

Las **medidas de dispersión** nos informan sobre cuanto se alejan del centro los valores de la distribución.

Las **medidas de dispersión** son:

### Rango o recorrido

El **rango** es la **diferencia** entre el **mayor** y el **menor** de los **datos** de una distribución estadística.

## Varianza

La **varianza** es la **media aritmética del cuadrado de las desviaciones** respecto a la **media**.

Es la media de los cuadrados de las desviaciones, y la denotaremos por  $S_x^2$  o también por  $\sigma_x^2$ .

$$S_x^2 = \sigma_x^2 = \sum_{i=1}^n \frac{d_i^2 \cdot n_i}{N} = \sum_{i=1}^n \frac{(x_i - \bar{x})^2 \cdot n_i}{N}$$

$$S_x^2 = \sigma_x^2 = \sum_{i=1}^n \frac{x_i^2 \cdot n_i}{N} - \bar{x}^2$$

## Desviación típica

La **desviación típica** es la **raíz cuadrada de la varianza**.

$$S_x = \sigma_x = \sqrt{\sum_{i=1}^n \frac{d_i^2 \cdot n_i}{N}} = \sqrt{\sum_{i=1}^n \frac{x_i^2 \cdot n_i}{N} - \bar{x}^2}$$