



## Resumen de Estadística

### 1. INTRODUCCIÓN A LA ESTADÍSTICA DESCRIPTIVA

Queremos realizar un estudio sobre una propiedad o característica en un determinado conjunto para extraer después conclusiones acerca del comportamiento de esta.

Entonces, la población constituye el conjunto donde vamos a realizar el estudio, y llamamos individuos a cada uno de los elementos de dicho conjunto. Estos individuos pueden ser personas, animales, objetos, ...

A la hora de realizar dicho estudio, podemos considerar toda la población, pero en ocasiones, esta población puede ser muy grande. Entonces seleccionamos una parte de esta población, para realizar en esta el estudio. Este subconjunto se denomina muestra.

Así, una muestra es un subconjunto de la población donde recopilaremos los datos.

Así, realizamos un estudio en una población, para ello podemos seleccionar una muestra y para cada uno de los individuos de la muestra, personas, animales, objetos, etcétera, recogemos datos sobre la propiedad a estudiar. Esta recibe el nombre de carácter estadístico.

Así, un carácter estadístico es la propiedad considerada para cada uno de los individuos al realizar un estudio estadístico.

Los caracteres estadísticos pueden ser de dos tipos:

1. Caracteres cualitativos: Son aquellos que no toman valores numéricos, por ejemplo, el color de los ojos, el grupo sanguíneo, la profesión de un individuo, nivel de estudios, ...
2. Caracteres cuantitativos: Son aquellos que toman valores numéricos, también los llamaremos variables estadísticas.

Así, una variable estadística es un carácter cuantitativo. Estas pueden ser de dos tipos:

- Variables estadísticas discretas, son aquellas variables estadísticas solo pueden tomar un número finito o numerable de valores: edad, número de goles marcados por un equipo de fútbol en una jornada, números de hijos de una familia, ...
- Variables estadísticas continuas, son aquellas que pueden tomar cualquier valor en un intervalo, como la estatura, el peso, el nivel de colesterol de una persona, los kilómetros recorridos por un futbolista durante un partido, ...

Así, a la hora de realizar el estudio, recogeremos los valores del carácter estadístico a estudiar para cada uno de los individuos de la muestra. Entonces, para tener recopilada de forma adecuada toda la información obtenida, crearemos una tabla estadística que incluya toda esta información de forma ordenada.

Estas tablas pueden ser de dos tipos:

1. Tablas de datos simples: donde incluimos todos los datos, y las veces que aparece cada uno de estos.
2. Tablas de datos agrupados intervalos: donde incluimos intervalos que contienen a los datos, y para cada uno de estos intervalos, el número de datos que están en este.

### 2. TABLAS ESTADÍSTICAS DE DATOS SIMPLES

Una tabla de datos simples es un tabla donde se recogen los distintos datos, las veces que se dan estos, y también pueden contener otros valores correspondientes a la variables estadística considerada.

Para empezar, en una tabla de datos simples incluiremos una primera columna donde introduciremos los distintos datos  $x_i$ .

En segundo lugar incluiremos una columna donde introducimos las frecuencias absolutas  $f_i$  correspondientes a los datos de la primera columna. La frecuencia  $f_i$  de un dato  $x_i$  es el número de veces que aparece dicho dato. Al final de esta columna, incluiremos la suma de todas las frecuencias absolutas, que se denota por  $\sum f_i$  y que corresponde con el número total de datos  $N$ .

También podemos incluir una columna con las frecuencias absolutas acumuladas  $F_i$ . La frecuencia absoluta acumulada  $F_i$  de un dato  $x_i$  es su frecuencia absoluta  $f_i$  más la suma de las frecuencias absolutas de los datos menores que este.

Finalmente, podemos también incluir una columna con las frecuencias relativas  $h_i$  de cada uno de los datos. La frecuencia relativa  $h_i$  de un dato  $x_i$  es  $h_i = \frac{f_i}{N}$ .

### 3. TABLAS ESTADÍSTICAS DE DATOS AGRUPADOS EN INTERVALOS

Las tablas de datos agrupados en intervalos se utilizan:

1. Cuando tenemos una variables estadística continua.
2. Cuando existen pocas repeticiones en los datos.

En estos casos, resulta conveniente agrupar los datos en intervalos, dado que aunque esto lleva a perder precisión ya que identificamos cada dato con el intervalo al que pertenece, ganamos en una mejor comprensión de la información que nos dan estos datos.

Consideraremos intervalos de la forma  $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$ . Cada uno de estos intervalos se identifica por su marca de clase del intervalo, que es  $x_i = \frac{a+b}{2}$ .

Para la obtención de la tabla de datos agrupados en intervalos, seguiremos los siguientes pasos:

1. En una primera columna incluiremos los intervalos.
2. En una segunda columna incluiremos las marcas de clase correspondientes a los intervalos.
3. En una tercera columna incluiremos las frecuencias absolutas  $f_i$ .
4. También podemos incluir nuevas columnas para las frecuencias absolutas acumuladas  $F_i$  y para las frecuencias relativas  $h_i$ .

### 4. GRÁFICOS ESTADÍSTICOS

Para los gráficos estadísticos de una variable estadística, distinguiremos si trabajamos esta con una tabla de datos simples o para una tabla de datos agrupados en intervalos.

Para una tabla de datos simples, obtendremos el diagrama de barras para las frecuencias absolutas y el diagrama de barras para las frecuencia absolutas acumuladas.

Para el diagrama de barras para las frecuencias absolutas, dibujaremos unos ejes, en el eje horizontal incluiremos los distintos datos, en el vertical valores para las frecuencias absolutas. Entonces, para cada dato incluiremos una barra vertical cuya altura corresponda con su frecuencia absoluta. Si unimos los puntos superiores de cada una de las barras, obtenemos el polígono de frecuencias de nuestra variables estadística.

El diagrama de barras para las frecuencias absolutas acumuladas se obtendrá de la misma forma que el anterior, correspondiendo ahora el eje vertical a las frecuencias absolutas acumuladas y siendo la altura de la barra correspondiente a cada dato su frecuencia absoluta acumulada.

Para una tabla de datos agrupados en intervalos obtendremos los histogramas de frecuencia absolutas y frecuencias absolutas acumuladas, y los diagramas de sectores.

Para los histogramas dibujaremos dos ejes, en el horizontal incluiremos los extremos de los intervalos y en el vertical valores para las frecuencias absolutas o frecuencias absolutas acumuladas, según corresponda. Entonces, a cada intervalo le haremos corresponden un rectángulo cuya base será dicho intervalo y altura su frecuencia absoluta o frecuencia absoluta acumulada, según corresponda.

A partir del histograma correspondiente a las frecuencias absolutas, podemos obtener el polígono de frecuencias, uniendo los puntos medios de los lados superiores de los rectángulos obtenidos en el histograma de frecuencias absolutas.

Finalmente, el diagrama de sectores es una circunferencia donde a cada intervalo le hacemos corresponder un ángulo que es directamente proporcional a su frecuencia absoluta.

### 5. PARÁMETROS DE CENTRALIZACIÓN

Los parámetros de centralización que estudiaremos son la moda, la mediana y la media.

La moda, que denotamos por  $M_0$ , es el dato  $x_i$  con mayor frecuencia absoluta. En el caso de que existan varios datos con mayor frecuencia absoluta, diremos que estos son las modas de la variable estadística.

Si tenemos dos modas, diremos que la variables estadística bimodal, si tenemos tres diremos que la variable estadística es trimodal ...

La mediana, que denotaremos por  $M_e$ , es el valor que, una vez ordenados los datos de menor a mayor incluyendo repeticiones, ocupa la posición central. En el caso de que sean dos los datos que ocupan la posición central, la mediana será la semisuma de estos, esto es, el resultado de sumar estos datos y dividir el resultado por 2.

Cuando el número de datos es grande, deberemos calcular la mediana utilizando la tabla de datos simples. Para ello añadiremos la columna con las frecuencias absolutas acumuladas  $F_i$ .

Entonces calcularemos  $\frac{N}{2}$ :

1. Si este valor no aparece entre las frecuencias absolutas acumuladas, la mediana es el primer dato cuya frecuencia absoluta acumulada supera a dicho valor.
2. Si este valor aparece como una frecuencia absoluta acumulada, la mediana será la semisuma del dato correspondiente a esta frecuencia absoluta y el siguiente dato.

La media de una variable estadística  $X$  es  $\bar{x} = \frac{\sum x_i \cdot f_i}{N}$

## 6. PARÁMETROS DE DISPERSIÓN

Los parámetros de dispersión son valores numéricos que miden cómo de separados están los datos. Estudiaremos los siguientes:

1. Rango o recorrido, que denotaremos por Rango, es el dato mayor menos el dato menor.
2. Desviación media:

$$DM = \frac{\sum |x_i - \bar{x}| \cdot f_i}{N}$$

3. Varianza:

$$VAR = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{N} = \frac{\sum x_i^2 f_i}{N} - \bar{x}^2$$

4. Desviación típica:  $DT = \sqrt{VAR}$ .
5. Coeficiente de variación, que es el dato adecuado para comparar las dispersión entre distintas variables estadísticas:

$$CV = \frac{\sigma}{\bar{x}}$$

## 7. PARÁMETROS DE POSICIÓN

Además de la mediana, los parámetros de posición que estudiaremos son los cuartiles, deciles y percentiles.

Para los cuartiles, dividiremos los datos en cuartos.

El primer cuartil, que denotaremos por  $C_1$ , es el dato que deja por debajo de él la  $\frac{1}{4}$  parte de los datos.

El segundo cuartil, que denotaremos por  $C_2$ , es el dato que deja por debajo de él las  $\frac{2}{4} = \frac{1}{2}$  parte de los datos.

El tercer cuartil, que denotaremos por  $C_3$ , es el dato que deja por debajo de él las  $\frac{3}{4}$  parte de los datos.

Para la obtención de los cuartiles, trabajaremos con una tabla de datos simples donde incluiremos la columna de las frecuencias absolutas acumuladas.

Así, asignaremos a los cuartiles primero, segundo y tercero el valor  $\frac{N}{4}$ ,  $\frac{2N}{4}$  y  $\frac{3N}{4}$  respectivamente.

Entonces, para la obtención de un cierto cuartil, buscaremos el valor obtenido en el párrafo anterior en la columna de las frecuencias absolutas acumuladas.

1. Si este valor no aparece entre las frecuencias absolutas acumuladas, el cuartil correspondiente es el primer dato cuya frecuencia absoluta acumulada supera a dicho valor.
2. Si este valor aparece como una frecuencia absoluta acumulada, el cuartil correspondiente sería la semisuma del dato correspondiente y el siguiente dato, esto es, sumamos estos dos datos y dividimos entre 2.

Para los deciles, trabajaremos de la misma forma dividiendo los datos ahora en 10 partes. Así, obtendremos 9 deciles, que denotaremos por  $D_1, D_2, D_3, \dots, D_9$ .

Así, el decil  $D_k$  es el dato que deja por debajo de él las  $\frac{k}{10}$  partes de los datos.

Podemos obtener este decil a partir de la tabla de datos agrupados en intervalos, de la misma forma que hemos obtenido los cuartiles, trabajando ahora con la fracción  $\frac{kN}{10}$ .

Para los percentiles, trabajaremos de la misma forma dividiendo los datos ahora en 100 partes. Así, obtendremos 99 percentiles, que denotaremos por  $P_1, P_2, P_3, \dots, P_{99}$ .

Así, el percentil  $P_k$  es el dato que deja por debajo de él las  $\frac{k}{100}$  partes de los datos.

Podemos obtener este percentil a partir de la tabla de datos agrupados en intervalos, de la misma forma que hemos obtenido los percentiles, trabajando ahora con la fracción  $\frac{kN}{100}$ .

## 8. PARÁMETROS PARA DATOS AGRUPADOS EN INTERVALOS

En primer lugar, para la moda y los parámetros de posición correspondientes a una variable estadística que viene dada por una tabla de datos agrupados en intervalos, trabajaremos de la misma forma que lo hacíamos para las tablas de datos simples, obteniendo ahora intervalos como resultado. También podemos considerar un valor numérico para los parámetros de posición, que será la marca de clase del intervalo correspondiente.

En cuanto al resto de parámetros, trabajaremos con las marcas de clase, para obtener estos parámetros como en el caso de las tablas de datos simples.