ESTADÍSTICA DESCRIPTIVA DE UNA VARIABLE

La Estadística tiene por objeto el desarrollo de técnicas para el conocimiento numérico de un conjunto de datos empíricos (recogidos mediante experimentos o encuestas). Definimos algunos aspectos básicos:

POBLACIÓN: Es el conjunto de todos los elementos que son objeto del estudio.

MUESTRA: Es un subconjunto, extraído de la población, cuyo estudio sirve para inferir características de toda la población.

INDIVIDUO: Es cada uno de los elementos que forman la población o la muestra.

CARACTERES Y VARIABLES: Caracteres son los aspectos que deseamos estudiar en los individuos de una población. Cada carácter puede tomar distintos valores o modalidades. Una variable estadística recorre todos los valores de un cierto carácter.

Clasificación de las variables estadísticas:

- **Cualitativas**: No toman valores numéricos.
- **Cuantitativas discretas**: Toman valores numéricos aislados.
- **Cuantitativas continuas**: Pueden tomar todos los valores de un intervalo.

Ejemplo

Los 13840 estudiantes de una universidad forman una **población**. Cada uno de ellos es un **individuo**. Se pueden analizar múltiples caracteres: sexo, número de años que ha estado matriculado, edad, estatura, … Las **variables** correspondientes son, respectivamente, **cualitativa**, **cuantitativa discreta** y **cuantitativas continuas** las dos últimas.

La **estadística descriptiva** trata de "describir" y analizar algunos caracteres de los individuos de un grupo dado, sin extraer conclusiones para un grupo mayor. Para este estudio, se siguen los pasos:

- Selección de caracteres dignos de estudio.
- Análisis de cada carácter, anotando los valores que toman los individuos en ellos.
- Clasificación y organización en tablas de los resultados obtenidos.
- Cálculo de ciertos valores numéricos (los parámetros estadísticos) a partir de los datos obtenidos.

La **estadística inferencial** trabaja con muestras y pretende, a partir de ellas, "inferir" características de toda la población.

1. DATOS AISLADOS

Al preguntar a un grupo de 18 personas por su edad se obtuvieron los siguientes resultados:

Edad	Número de personas
22	1
23	2
24	3
25	4
26	6
27	2

TABLA DE FRECUENCIAS

DIAGRAMA DE BARRAS

x_i	f_i

x_i: valores de la variable estadística
f_i: frecuencia absoluta o frecuencia

PARÁMETROS ESTADÍSTICOS

x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$

Media: $\overline{x} = \dfrac{\sum x_i \cdot f_i}{\sum f_i} =$

Moda: valor de la variable estadística con mayor frecuencia, $M_o =$

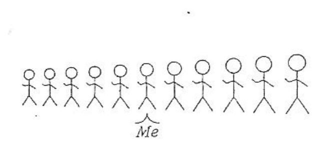
Varianza: $\sigma^2 = \dfrac{\sum (x_i - \overline{x})^2 \cdot f_i}{\sum f_i} = \dfrac{\sum x_i^2 \cdot f_i}{\sum f_i} - \overline{x}^2 =$

Desviación típica: $\sigma = \sqrt{Varianza} =$

<u>Ejercicio 1.- </u>Las notas de Pedro son: 1, 3, 6, 6, 9, 1, 6, 9, 3, 6 y las de Andrés son: 3, 6, 4, 6, 6, 3, 6, 6, 4, 6. Calcula la media, la varianza y la desviación típica en ambos casos y comenta los resultados.

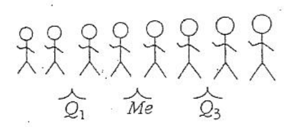
Medidas de posición

- MEDIANA: si los individuos de una población están colocados en orden creciente según la variable que estudiamos, el que ocupa el valor central se llama individuo mediano, y su valor la **mediana**.



  La **mediana**, Me, está situada de modo que antes de ella está el 50% de la población y, detrás, el otro 50%.
  Por ejemplo: 6, 7, 7, 7, **8**, 9, 10, 12, 15 $\rightarrow$mediana: Me=8. Si el número de individuos es par, la mediana es el valor medio de los dos centrales. Por ejemplo: 6, 7, 7, 7, **8, 9**, 10, 12, 15, 16 $\rightarrow$ mediana: Me=8,5.

- CUARTILES: si en lugar de partir la totalidad de los individuos en dos mitades, lo hacemos en cuatro partes iguales (todas ellas con el mismo número de individuos), los dos nuevos puntos de separación se llaman **cuartiles**.



  Cuartil inferior, Q_1, es un valor de la variable que deja por debajo de él al 25% de la población, y por encima al 75%. El **cuartil superior**, Q_3, deja debajo al 75% y encima al 25%.
  Se designan por Q_1 y Q_3 porque la mediana sería el Q_2.
  Por ejemplo, en la distribución

  $$\underbrace{1,\ 2,\ 2}_{25\%}\ ,\ \underset{Q_1}{\uparrow}\ \underbrace{3,\ 4,\ 5}_{25\%}\ ,\ \underset{Me}{\uparrow}\ \underbrace{5,\ 5,\ 6}_{25\%}\ ,\ \underset{Q_3}{\uparrow}\ \underbrace{8,\ 9,\ 10}_{25\%}$$

  estos parámetros toman los valores siguientes: $Q_1 = 2,5$; Me=5; $Q_3 = 7$

- CENTILES O PERCENTILES: si partimos la población en 100 partes y señalamos el lugar que deja debajo k de ellas, el valor de la variable correspondiente a ese lugar se designa por P_k y se denomina **centil** k o **percentil** k.
  La mediana es Me=P_{50} y los cuartiles $Q_1 = P_{25}$, $Q_3 = P_{75}$.
  Mediana, cuartiles y centiles, se llaman **medidas de posición**.

<u>EJERCICIO RESUELTO.-</u> Calcular Me, Q_1, Q_3, P_{10} y P_{80} en la distribución: 1, 1, 2, 3, 4, 4, 5, 5, 5, 5, 6, 7, 7, 7, 8, 9, 10

Hay 17 individuos.
$17/2 = 8,5 \rightarrow$ La mediana es el valor del individuo 9º: Me=5
$17/4 = 4,25 \rightarrow$ 5º lugar, $Q_1 = 4$
$17 \cdot (3/4) = 12,75 \rightarrow$ 13º lugar, $Q_3 = 7$
$17 \cdot (10/100) = 1,7 \rightarrow$ 2º lugar, $P_{10} = 1$
$17 \cdot (80/100) = 13,6 \rightarrow$ 14º lugar, $P_{80} = 7$

<u>Ejercicio 2.-</u> La altura, en centímetros, de un grupo de alumnos y alumnas de una misma clase es: 150, 169, 171, 172, 172, 175, 181, 182, 183, 177, 179, 176, 184, 158. Calcula Me, Q_1, Q_3, P_{60}.

- FRECUENCIAS ACUMULADAS
  Para calcular la mediana, los cuartiles y los demás percentiles en distribuciones dadas por tablas de frecuencias, se necesita el concepto de **frecuencia acumulada**.
  En una distribución de frecuencias, se llama frecuencia acumulada, F_i, correspondiente al valor i-ésimo, x_i, a la suma de la frecuencia de ese valor con todas las anteriores:
  $F_i = f_1 + f_2 + ... + f_i$

Por ejemplo, en la distribución de frecuencias dada en el margen, obtenemos:

x_i	f_i
0	4
1	18
2	41
3	32
4	11
5	3
6	1

x_i	f_i	F_i	en %
0	4	4	3,6
1	18	4+18=22	20
2	41	4+18+41=22+41=63	57,3
3	32	4+18+41+32=63+32=95	86,4
4	11	4+18+41+32+11=95+11=106	96,4
5	3	4+18+41+32+11+3=106+3=109	99,1
6	1	4+18+41+32+11+3+1=109+1=110	100

La expresión en % de las frecuencias acumuladas nos permite obtener fácilmente los percentiles.

- OBTENCIÓN DE PERCENTILES EN TABLAS DE FRECUENCIAS

  Para hallar el percentil P_k en una tabla de frecuencias, se obtienen las frecuencias acumuladas y se expresan en %. El percentil P_k es el valor para el cual la frecuencia acumulada correspondiente supera el k%.

  En caso de que una de ellas coincida con $k\%$, se toma como P_k el valor intermedio entre ése y el siguiente.

  Por ejemplo, en la tabla de arriba obtenemos Me, Q_1, Q_3, P_{99} y P_{20}:

  Me= $P_{50} = 2$ porque para $x_i = 2$ la F_i supera el 50%

  $Q_1 = P_{25} = 2$ porque para $x_i = 2$ la F_i supera el 25%

  $Q_3 = P_{75} = 3$ porque para $x_i = 3$ la F_i supera el 75%

  $P_{99} = 5$ porque para $x_i = 5$ la F_i supera el 99%

  $P_{20} = 1{,}5$ porque para $x_i = 1$ la F_i iguala el 20%. Por tanto, el valor 1,5 es superior al 20% de la población e inferior al 80% restante.

Ejercicio 3.- En la siguiente distribución de notas, halla M_e, Q_1, Q_3, p_{80}, p_{90} y p_{99}.

NOTAS	1	2	3	4	5	6	7	8	9	10
Nº DE ALUMNOS	7	15	41	52	104	69	26	13	19	14

Ejercicio 4.- Halla la mediana, los cuartiles y el percentil 60 para la siguiente distribución, correspondiente a las notas obtenidas en un test que hizo un grupo de estudiantes: 25-22-27-30-23-22-31-18-24-25-32-35-20-28-30

2. DATOS AGRUPADOS EN INTERVALOS

2.1 INTERVALOS DE LA MISMA AMPLITUD

En una clase de Educación Física de 4º ESO se ha cronometrado el tiempo, en segundos, que tarda cada alumno/a en recorrer cierta distancia fija. Los resultados son:

Tiempo	Número de alumnos
$[8,9)$	11
$[9,10)$	8
$[10,11)$	5
$[11,12)$	1
$[12,13)$	3
$[13,14)$	0
$[14,15)$	2

HISTOGRAMA

2.2 INTERVALOS DE DISTINTA AMPLITUD

Al preguntar a un grupo de personas cuánto tiempo dedicaron a ver televisión durante un fin de semana, se obtuvieron los siguientes resultados:

Tiempo(horas)	Número de personas
$[0;0,5)$	10
$[0,5;1,5)$	10
$[1,5;2,5)$	18
$[2,5;4)$	12
$[4,8)$	12

HISTOGRAMA

Si los intervalos no tienen la misma amplitud se calculan las alturas de las barras necesarias para que las superficies de las barras coincidan con las frecuencias:

Intervalo	f_i
$[0;0,5)$	10
$[0,5;1,5)$	10
$[1,5;2,5)$	18
$[2,5;4)$	12
$[4,8)$	12

$f_1 = 0,5 \cdot h_1 \Rightarrow h_1 =$

PARÁMETROS ESTADÍSTICOS

En un autobús escolar se les pregunta a los alumnos y alumnas por el tiempo que tardan en llegar de su casa al autobús. Los resultados se recogen en la siguiente tabla:

Tiempo (minutos)	$[0,5)$	$[5,10)$	$[10,15)$	$[15,20)$	$[20,25)$
Nº de alumnos	20	13	18	5	4

Intervalos	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$

x_i: marca de clase (punto medio del intervalo)

Media:

Varianza:

Desviación típica:

Ejercicio 5.- Halla la media, la varianza y la desviación típica en el ejemplo:

Tiempo(horas)	Número de personas
$[0;0,5)$	10
$[0,5;1,5)$	10
$[1,5;2,5)$	18
$[2,5;4)$	12
$[4,8)$	12