12 Estadística

INTRODUCCIÓN

La Estadística es la ciencia que estudia los métodos y procedimientos para recoger datos, clasificarlos, analizarlos, tomar decisiones y sacar conclusiones científicas a partir de ellos. Se divide en dos ramas: la Estadística *descriptiva*, que se encarga de la recogida y el análisis de datos pertenecientes a una muestra o a la población, y la Estadística *inductiva*, que se ocupa de generalizar a toda la población los resultados y las conclusiones obtenidos a partir de muestras.

RESUMEN DE LA UNIDAD

- *Estadística:* ciencia que recoge, analiza e interpreta los datos de un conjunto de elementos.
- *Medidas de centralización:* parámetros estadísticos de un conjunto de datos que reflejan la tendencia de los datos a concentrarse alrededor de ciertos valores.
- *Medidas de dispersión:* parámetros estadísticos que reflejan el mayor o menor agrupamiento de un conjunto de datos.

OBJETIVOS	CONTENIDOS	PROCEDIMIENTOS
1. Reconocer y diferenciar los conceptos de población y muestra.	• Estadística. • Población y muestra.	• Distinción de los conceptos de población y muestra.
2. Clasificar variables estadísticas: cuantitativas y cualitativas.	• Variables cualitativas y cuantitativas. • Variables estadísticas discretas y continuas.	• Diferenciación de las variables cualitativas y cuantitativas, y dentro de estas, de las variables discretas y continuas.
3. Obtener la tabla estadística asociada a un conjunto de datos.	• Tablas estadísticas. • Marca de clase.	• Construcción de tablas estadísticas adecuadas al conjunto de datos.
4. Hallar la frecuencia absoluta y relativa de un conjunto de datos.	• Frecuencias absolutas. • Frecuencias relativas.	• Cálculo, a partir de la tabla estadística, de frecuencias absolutas, frecuencias relativas y porcentajes.
5. Construir la tabla de frecuencias acumuladas.	• Frecuencias absolutas acumuladas. • Frecuencias relativas acumuladas.	• Obtención, a partir de la tabla estadística, de frecuencias absolutas acumuladas y de frecuencias relativas acumuladas.
6. Utilizar y analizar los gráficos adecuados para representar datos.	• Gráficos estadísticos: diagrama de barras, histograma y polígono de frecuencias.	• Representación de las variables estadísticas mediante gráficos, diferenciando según el tipo de datos recogidos.
7. Calcular las medidas de centralización de un conjunto de datos.	• Media. • Mediana. • Moda.	• Cálculo e interpretación de la media, la mediana y la moda de un conjunto de datos.
8. Calcular las medidas de dispersión de un conjunto de datos.	• Recorrido. • Desviación media. • Varianza y desviación típica.	• Obtención del recorrido, la desviación media, la varianza y la desviación típica de un conjunto de datos.

12 OBJETIVO 1

RECONOCER Y DIFERENCIAR LOS CONCEPTOS DE POBLACIÓN Y MUESTRA

NOMBRE: ______________________ CURSO: __________ FECHA: __________

- La **Estadística** es la ciencia encargada de recoger, analizar e interpretar los datos relativos a un conjunto de elementos.
- La **población** es el conjunto de elementos sobre los que se va a estudiar un determinado aspecto o característica.
- La **muestra** es una parte de la población. Es importante escoger bien la muestra, ya que esta ha de ser representativa, es decir, debe dar una información correcta y similar a la obtenida si estudiásemos toda la población.
- El **tamaño** de una muestra es el número de elementos que la componen.

EJEMPLO

Los resultados a la pregunta: ¿Cómo clasificarías las desigualdades que actualmente existen entre hombres y mujeres en nuestro país en el ámbito laboral?, del sondeo de opinión sobre «Las mujeres y el empleo» están recogidos en porcentajes (%) en la tabla.

	TOTAL	**SEXO**	
		Hombres	**Mujeres**
Muy grandes	9	6	13
Bastante grandes	45	40	50
Bastante pequeñas	28	32	24
Casi inexistentes	14	19	9
No sabe/No contesta	4	3	4

Junto al sondeo de opinión aparece esta ficha técnica.

Ámbito: territorio español, excluyendo Ceuta y Melilla.
Universo: población española de ambos sexos de 18 años o más.
Tamaño de la muestra: 2.488 entrevistas.
Error muestral: para un nivel de confianza del 95,5 %, el error es del ± 2 %.
Fecha de realización: 23-27 de enero de 1997 (Centro de Investigaciones Sociológicas, CIS).

El error del ± 2 % significa que a la respuesta de «Muy grandes», que es el 9 % en la muestra (2.488 casos), la respuesta en la población sería del 9 ± 2 %; es decir, entre un 7 % y un 11 % de las personas contestarían «Muy grandes», afirmándolo en el 95,5 % de las estimaciones (nivel de confianza).

En los estudios estadísticos se eligen muestras en lugar de poblaciones cuando estas son muy amplias, por motivos económicos, por la rapidez en conocer los resultados, etc.

1 **Hazle esa misma pregunta a tus compañeros de clase y construye una tabla similar a la anterior, pero sin calcular porcentajes, es decir, apuntando cuántos compañeros han dado cada una de las respuestas y su género.**

CLASIFICAR VARIABLES ESTADÍSTICAS: CUANTITATIVAS Y CUALITATIVAS

NOMBRE: ________________ CURSO: ________ FECHA: ________

Una **variable estadística** es cualquier característica o aspecto de los elementos de una población o de una muestra que se puede estudiar.

Las variables estadísticas pueden ser:

- **Variables cuantitativas:** se pueden medir y se expresan mediante números. A su vez, pueden ser discretas o continuas.
  - Las **variables cuantitativas discretas** toman un número determinado de valores.
  - Las **variables cuantitativas continuas** pueden tomar cualquier valor comprendido entre dos valores dados.
- **Variables cualitativas:** no se pueden medir y se expresan mediante cualidades o descripciones.

EJEMPLO

Señala, en cada caso, qué *tipo de variable* es, y di si es más conveniente estudiar la *población* o una *muestra*.

a) La estatura de los 20 alumnos de una clase: *variable cuantitativa continua*, y estudiamos la *población*.

b) Los efectos de un nuevo medicamento en el ser humano: *variable cualitativa*, y estudiamos una *muestra* de la población.

c) La talla de pantalones de los varones de una Comunidad Autónoma: *variable cuantitativa discreta*, y estudiamos una *muestra*.

d) Las aficiones deportivas de los alumnos de un instituto: *variable cualitativa*, y podemos estudiar una *muestra* de alumnos de los diferentes cursos.

e) El color del pelo de los alumnos de una clase: *variable cualitativa*, y en este caso es conveniente estudiar la *población*.

1 **Señala en cada caso lo que corresponda.**

VARIABLE	CUANTITATIVA		CUALITATIVA	POBLACIÓN	MUESTRA
	Discreta	Continua			
Profesión del padre					
Número de personas que viven en cada piso de un edificio					
Número de llamadas realizadas desde un teléfono al día					
Equipo de fútbol preferido por cada alumno de una clase					
Temperaturas medidas a lo largo de una semana					
El peso de cada uno de los 20 alumnos de una clase					

ADAPTACIÓN CURRICULAR

12 OBJETIVO 3
OBTENER LA TABLA ESTADÍSTICA ASOCIADA A UN CONJUNTO DE DATOS

NOMBRE: ______________________ CURSO: __________ FECHA: __________

Las **tablas estadísticas** sirven para ordenar y estudiar los datos de una variable estadística.

Si la variable es **discreta**, es decir, si tenemos un conjunto de datos pequeño, se forma una tabla con dos columnas. En una de las columnas se colocan los distintos valores de la variable, y en la otra columna se indica el número de veces que aparece cada uno de ellos.

Si la variable es **continua**, y tenemos un conjunto de datos grande:

1.º Se halla el **recorrido** de la variable, o la diferencia entre sus valores mayor y menor.
2.º Se agrupan los valores en **intervalos de igual amplitud**.
3.º Se establece la **marca de clase**, que es el punto medio de cada intervalo.
4.º Se hace el **recuento** de cada uno de los datos.

EJEMPLO

Las notas obtenidas en un examen de Matemáticas por los 20 alumnos de una clase de 4.º ESO, han sido: 6, 5, 3, 1, 2, 5, 6, 5, 9, 8, 7, 4, 9, 10, 7, 7, 8, 6, 5 y 5. Ordena estos datos en una tabla.

El número de valores que puede tomar la variable es pequeño, y es una variable discreta.

Para recoger los datos en una tabla, ponemos en la primera columna los posibles valores de las notas, que en este caso es la variable estadística, y en la segunda columna, el número de veces que ha salido cada una de ellas.

NOTAS	1	2	3	4	5	6	7	8	9	10
RECUENTO	1	1	1	1	5	3	3	2	2	1

EJEMPLO

El número de personas que viven en cada uno de los edificios de una calle son:

69, 85, 139, 114, 103, 84, 97, 133, 155, 127, 110, 138, 94, 143, 106, 99, 80, 74, 102, 93, 128, 78, 86, 104, 121, 137, 89, 107, 92 y 101

Haz una tabla, el recuento y obtén las marcas de clase.

En este caso, el número de posibles valores que puede tomar la variable es grande, pues varía entre 69 y 155. Agrupamos los datos en intervalos. Para ello, hallamos el recorrido (diferencia entre el mayor y el menor valor):

$$155 - 69 = 86$$

Tomaremos 6 intervalos de amplitud 15 ($6 \cdot 15 = 90 > 86$), empezando por el menor valor: 69.

INTERVALO	**MARCA DE CLASE**	**RECUENTO**
[69, 84)	76,5	4
[84, 99)	91,5	8
[99, 114)	106,5	8
[114, 129)	121,5	4
[129, 144)	136,5	5
[144, 159)	151,5	1

Las marcas de clase son: $\frac{69+84}{2} = 76,5$ $\frac{99+114}{2} = 106,5$ $\frac{129+144}{2} = 136,5$

$\frac{84+99}{2} = 91,5$ $\frac{114+129}{2} = 121,5$ $\frac{144+159}{2} = 151,5$

OBJETIVO 4

HALLAR LA FRECUENCIA ABSOLUTA Y RELATIVA DE UN CONJUNTO DE DATOS

NOMBRE: ________________ CURSO: ________ FECHA: ________

La **frecuencia absoluta** $\boldsymbol{f_i}$ de un conjunto de datos es el número de veces que se repite cada valor de la variable x_i en el total de los datos.

La **frecuencia relativa** $\boldsymbol{h_i}$ es el cociente entre la frecuencia absoluta y el número total de datos:

$$\boldsymbol{h_i = \frac{f_i}{n}}$$

La frecuencia relativa es siempre un número comprendido entre 0 y 1.

La suma de todas las frecuencias absolutas es igual al número total de datos, *n*. La suma de todas las frecuencias relativas es 1.

Multiplicando la frecuencia relativa por 100, obtenemos el porcentaje (%).

EJEMPLO

Con los datos de las notas del examen de Matemáticas del ejemplo anterior, construye una tabla de frecuencias y porcentajes.

En la segunda columna colocamos el recuento, es decir, el número de veces que aparece cada valor. Este recuento se llama *frecuencia absoluta*.

En la tercera columna colocamos el cociente entre la frecuencia absoluta y el número total de datos (20). Este número se llama *frecuencia relativa*.

$$h_1 = \frac{1}{20} = 0{,}05 \qquad h_6 = \frac{3}{20} = 0{,}05$$

$$h_2 = \frac{1}{20} = 0{,}05 \qquad h_7 = \frac{3}{20} = 0{,}05$$

$$h_3 = \frac{1}{20} = 0{,}25 \qquad h_8 = \frac{2}{20} = 0{,}15$$

$$h_4 = \frac{1}{20} = 0{,}15 \qquad h_9 = \frac{2}{20} = 0{,}10$$

$$h_5 = \frac{5}{20} = 0{,}10 \qquad h_{10} = \frac{1}{20} = 0{,}05$$

x_i	f_i	h_i	%
1	1	0,05	5
2	1	0,05	5
3	1	0,05	5
4	1	0,05	5
5	5	0,25	25
6	3	0,15	15
7	3	0,15	15
8	2	0,10	10
9	2	0,10	10
10	1	0,05	5
Suma	20	1	100

En la cuarta columna colocamos el porcentaje, que es el resultado de multiplicar por 100 cada valor de la frecuencia relativa h_i.

1 **Se ha lanzado un dado 20 veces, obteniendo los siguientes resultados:**
2, 3, 5, 4, 2, 4, 6, 5, 5, 1, 2, 3, 1, 4, 1, 5, 4, 6, 3 y 3. Construye una tabla con las frecuencias absoluta y relativa y los porcentajes.

ADAPTACIÓN CURRICULAR

EJEMPLO

Con los datos del ejemplo anterior del número de habitantes de cada edificio construye la tabla de frecuencias absolutas, relativas y porcentajes.

En la primera columna colocamos los valores de la variable (número de habitantes por edificio), agrupados en 6 *intervalos* de amplitud 15; en la segunda columna ponemos la *marca de clase* de cada intervalo; en la tercera columna indicamos la *frecuencia absoluta*; en la cuarta, la *frecuencia relativa*, y en la quinta, el *porcentaje*.

INTERVALO	x_i	f_i	h_i	%
[69, 84)	76,5	4	4/30 = 0,13	13
[84, 99)	91,5	8	8/30 = 0,27	27
[99, 114)	106,5	8	8/30 = 0,27	27
[114, 129)	121,5	4	4/30 = 0,13	13
[129, 144)	136,5	5	5/30 = 0,17	17
[144, 159)	151,5	1	1/30 = 0,03	3
Suma		30	1	100

2 **El peso (en kilos) de una muestra de 30 individuos, escogidos al azar, es: 59, 69, 74, 70, 68, 85, 83, 75, 56, 92, 86, 94, 58, 61, 74, 77, 79, 67, 84, 73, 82, 74, 79, 80, 81, 65, 60, 59, 73 y 62. Agrupa los datos en intervalos y construye una tabla con las frecuencias absoluta y relativa y los porcentajes.**

Hay que calcular el recorrido de la variable (peso, en este caso). Para ello, observamos cuáles son los valores menor y mayor.

valor **menor** = 56 valor **mayor** = 94 **recorrido** = 94 − 56 = 38

Podemos tomar 5 intervalos de amplitud 10, ya que $5 \cdot 10 = 50 > 38$.

INTERVALO	x_i	f_i	h_i	%
[50, 60)				
[60, 70)				
[70, 80)				
[80, 90)				
[90, 100)				
Suma				

OBJETIVO 5

CONSTRUIR LA TABLA DE FRECUENCIAS ACUMULADAS

NOMBRE: ______________________ CURSO: __________ FECHA: __________

- La **frecuencia absoluta acumulada F_i** de un valor x_i es la suma de las frecuencias f_i de todos los valores menores o iguales que él.
- La **frecuencia relativa acumulada H_i** de un valor x_i es el cociente entre la frecuencia absoluta acumulada y el número total de datos: $H_i = \frac{F_i}{N}$.

EJEMPLO

Con los datos de las notas del examen de Matemáticas del ejemplo anterior, construye una tabla de frecuencias absolutas acumuladas y frecuencias relativas acumuladas.

Obtenemos la frecuencia absoluta acumulada de cada valor:

$F_1 = f_1 = 1$

$F_2 = F_1 + f_2 = 1 + 1 = 2$

$F_3 = F_2 + f_3 = 2 + 1 = 3$

$F_4 = F_3 + f_4 = 3 + 1 = 4$

$F_5 = F_4 + f_5 = 4 + 5 = 9$

$F_6 = F_5 + f_6 = 9 + 3 = 12$

$F_7 = F_6 + f_7 = 12 + 3 = 15$

$F_8 = F_7 + f_8 = 15 + 2 = 17$

$F_9 = F_8 + f_9 = 17 + 2 = 19$

$F_{10} = F_9 + f_{10} = 19 + 1 = 20$

Calculamos la frecuencia relativa acumulada de los distintos valores:

$$H_1 = \frac{F_1}{N} = \frac{1}{20} = 0{,}05$$

$$H_2 = \frac{F_2}{N} = H_1 + h_2 = 0{,}05 + 0{,}05 = 0{,}10$$

$$H_3 = \frac{F_3}{N} = H_2 + h_3 = 0{,}10 + 0{,}05 = 0{,}15$$

$$H_4 = \frac{F_4}{N} = H_3 + h_4 = 0{,}15 + 0{,}05 = 0{,}20$$

$$H_5 = \frac{F_5}{N} = H_4 + h_5 = 0{,}20 + 0{,}25 = 0{,}45$$

$$H_6 = \frac{F_6}{N} = H_5 + h_6 = 0{,}45 + 0{,}15 = 0{,}60$$

$$H_7 = \frac{F_7}{N} = H_6 + h_7 = 0{,}60 + 0{,}15 = 0{,}75$$

$$H_8 = \frac{F_8}{N} = H_7 + h_8 = 0{,}75 + 0{,}10 = 0{,}85$$

$$H_9 = \frac{F_9}{N} = H_8 + h_9 = 0{,}85 + 0{,}10 = 0{,}95$$

$$H_{10} = \frac{F_{10}}{N} = H_9 + h_{10} = 0{,}95 + 0{,}05 = 1$$

DATOS	FRECUENCIA ABSOLUTA (f_i)	FRECUENCIA ABSOLUTA ACUMULADA (F_i)	FRECUENCIA RELATIVA (h_i)	FRECUENCIA RELATIVA ACUMULADA (H_i)
1	1	1	0,05	0,05
2	1	2	0,05	0,10
3	1	3	0,05	0,15
4	1	4	0,05	0,20
5	5	9	0,25	0,45
6	3	12	0,15	0,60
7	3	15	0,15	0,75
8	2	17	0,10	0,85
9	2	19	0,10	0,95
10	1	20	0,05	1

12

1 **Un dado se ha lanzado 20 veces, obteniéndose los siguientes resultados.**

2, 3, 5, 4, 2, 4, 6, 5, 5, 1, 2, 3, 1, 4, 1, 5, 4, 6, 3 y 3

Construye la tabla de frecuencias con las columnas de las frecuencias absolutas y relativas acumuladas.

EJEMPLO

Con los datos de los habitantes de cada edificio del ejemplo anterior, construye una tabla de frecuencias absolutas y relativas acumuladas.

INTERVALO	x_i	f_i	F_i	h_i	%
[69, 84)	76,5	4	4	4/30 = 0,13	0,13
[84, 99)	91,5	8	12	8/30 = 0,27	0,40
[99, 114)	106,5	8	20	8/30 = 0,27	0,67
[114, 129)	121,5	4	24	4/30 = 0,13	0,80
[129, 144)	136,5	5	29	5/30 = 0,17	0,97
[144, 159)	151,5	1	30	1/30 = 0,03	1
Suma		30		1	

2 **Con los datos del peso de las 30 personas del ejemplo anterior, completa la tabla con las frecuencias absolutas y relativas acumuladas.**

INTERVALO	x_i	f_i	F_i	h_i	%
[50, 60)	55				
[60, 70)					
[70, 80)					
[80, 90)					
[90, 100)					
Suma					

OBJETIVO 6

UTILIZAR Y ANALIZAR LOS GRÁFICOS ADECUADOS PARA REPRESENTAR DATOS

NOMBRE: ____________________ CURSO: __________ FECHA: __________

Los **gráficos** representan la información que contienen las tablas estadísticas. Según sea la variable, se usa un tipo u otro de gráfico.

VARIABLES DISCRETAS

- **Diagrama de barras:** para representar datos cuantitativos o cualitativos discretos.
  Sobre el eje *X* se señalan los valores de la variable y se levantan barras de altura igual a la frecuencia representada (absoluta, absoluta acumulada, relativa o relativa acumulada).
- El **polígono de frecuencias** es una línea poligonal que se obtiene al trazar, en el diagrama de barras, una línea desde cada extremo de una barra hasta el extremo de la barra siguiente.

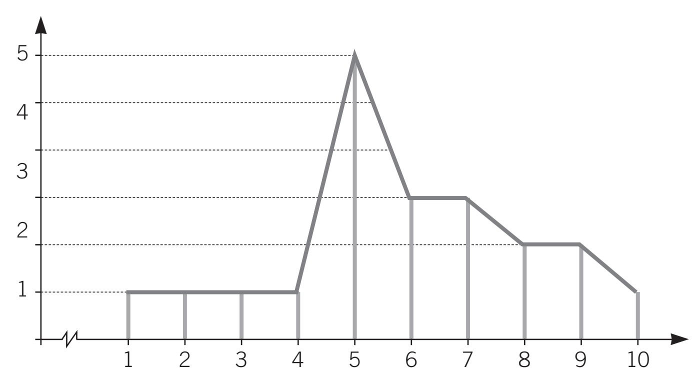
VARIABLES CONTINUAS

- **Histograma:** para representar variables cuantitativas continuas.
  Se señalan sobre el eje horizontal los extremos de los intervalos y se levantan rectángulos de altura igual que la frecuencia que se va representar.
- El **polígono de frecuencias** se obtiene al unir los puntos medios de los lados superiores de los rectángulos del histograma.

EJEMPLO

Con las frecuencias absolutas del ejemplo de las notas en el examen de Matemáticas, construye el diagrama de barras y el polígono de frecuencias.

x_i	f_i
1	1
2	1
3	1
4	1
5	5
6	3
7	3
8	2
9	2
10	1
Suma	20

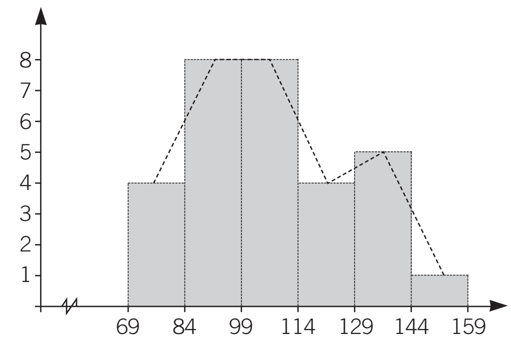


1 **Con los datos del ejemplo del lanzamiento del dado, construye el diagrama de barras y el polígono de frecuencias acumuladas.**

EJEMPLO

Con las frecuencias absolutas del ejemplo anterior de los habitantes de cada edificio, construye el histograma y el polígono de frecuencias.

INTERVALO	x_i	f_i
[69, 84)	76,5	4
[84, 99)	91,5	8
[99, 114)	106,5	8
[114, 129)	121,5	4
[129, 144)	136,5	5
[144, 159)	151,5	1



2 **Con los datos del peso de las 30 personas, construye el histograma y el polígono de frecuencias.**

INTERVALO	x_i	f_i
[50, 60)	55	4
[60, 70)	65	7
[70, 80)	75	10
[80, 90)	85	7
[90, 100)	95	2
Suma		

GRÁFICO DE SECTORES

Se divide un círculo en sectores de ángulo proporcional a la frecuencia absoluta de cada valor de la variable estadística.

EJEMPLO

Con las frecuencias absolutas del ejemplo de las notas en el examen de Matemáticas, construye el diagrama de sectores.

x_i	f_i
1	1
2	1
3	1
4	1
5	5
6	3
7	3
8	2
9	2
10	1
Suma	20

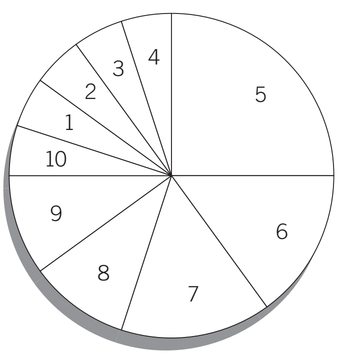
Para hallar la amplitud de cada sector, aplicamos con cada valor una regla de tres:

$$\left.\begin{array}{l}\text{Si } 360° \longrightarrow 20 \text{ casos}\\ \quad x \longrightarrow 1 \text{ caso}\end{array}\right\} \rightarrow x = \frac{360 \cdot 1}{20} = 18°$$

$$\left.\begin{array}{l}\text{Si } 360° \longrightarrow 20 \text{ casos}\\ \quad x \longrightarrow 5 \text{ casos}\end{array}\right\} \rightarrow x = \frac{360 \cdot 5}{20} = 90°$$

$$\left.\begin{array}{l}\text{Si } 360° \longrightarrow 20 \text{ casos}\\ \quad x \longrightarrow 3 \text{ casos}\end{array}\right\} \rightarrow x = \frac{360 \cdot 3}{20} = 54°$$

$$\left.\begin{array}{l}\text{Si } 360° \longrightarrow 20 \text{ casos}\\ \quad x \longrightarrow 2 \text{ casos}\end{array}\right\} \rightarrow x = \frac{360 \cdot 2}{20} = 36°$$



3 **Representa en un gráfico de sectores las frecuencias del ejemplo de los habitantes de cada edificio.**

INTERVALO	x_i	f_i
[69, 84)	76,5	4
[84, 99)	91,5	8
[99, 114)	106,5	8
[114, 129)	121,5	4
[129, 144)	136,5	5
[144, 159)	151,5	1

12 OBJETIVO 7
CALCULAR LAS MEDIDAS DE CENTRALIZACIÓN DE UN CONJUNTO DE DATOS

NOMBRE: ______________________ CURSO: __________ FECHA: __________

La media, la mediana y la moda son medidas de centralización, ya que reflejan la tendencia de los datos a concentrarse alrededor de ciertos valores.

La **media**, $\bar{x}$, de un conjunto de datos $x_1, x_2, ..., x_n$ es: $\bar{x} = \dfrac{x_1f_1 + x_2f_2 + ... + x_nf_n}{f_1 + f_2 + ... + f_n}$

Si los datos están agrupados en intervalos, el valor x_i es la marca de clase de cada intervalo.

EJEMPLO

Halla la nota media de las notas obtenidas por los 20 alumnos en el examen de Matemáticas.

x_i	1	2	3	4	5	6	7	8	9	10	Suma
f_i	1	1	1	1	5	3	3	2	2	1	20
$x_i \cdot f_i$	1	2	3	4	25	18	21	16	18	10	118

$$\bar{x} = \frac{\Sigma x_i \cdot f_i}{\Sigma f_i} = \frac{\Sigma x_i \cdot f_i}{N} = \frac{118}{20} = 5{,}9$$

1 **Determina la media de habitantes en cada edificio de una calle del ejemplo anterior.**

INTERVALO	x_i	f_i	$x_i \cdot f_i$
[69, 84)	76,5	4	
[84, 99)	91,5	8	
[99, 114)	106,5	8	
[114, 129)	121,5	4	
[129, 144)	136,5	5	
[144, 159)	151,5	1	

2 **Halla la media de los datos del peso de las 30 personas del ejemplo anterior.**

INTERVALO	x_i	f_i
[50, 60)	55	4
[60, 70)	65	7
[70, 80)	75	10
[80, 90)	85	7
[90, 100)	95	2

- La **mediana** de un conjunto de datos es el valor tal que, ordenados los datos de forma creciente, la mitad son menores que él y la otra mitad son mayores que él. Se representa por ***Me***.
- Si el conjunto de datos es un número impar, la mediana es el valor central, colocados los datos de forma creciente.
- Si el conjunto de datos es par, la mediana es la media de los dos valores centrales.
- La **moda** de un conjunto de datos es el valor de la variable o el intervalo que más se repite. Se representa por ***Mo***.
- El valor de la moda puede no ser único, es decir, puede haber varios valores de la variable o intervalos que se repitan por igual.

EJEMPLO

En el ejemplo del examen de Matemáticas, el número de notas es par, y la mediana será:

$$Me = \frac{5+6}{2} = 5{,}5$$

La moda es el valor que más se repite: $Mo = 5$.

3 **Halla la mediana y la moda en el ejemplo de los habitantes de cada edificio, tomando para ello las marcas de clase.**

INTERVALO	x_i	f_i
[69, 84)	76,5	4
[84, 99)	91,5	8
[99, 114)	106,5	8
[114, 129)	121,5	4
[129, 144)	136,5	5
[144, 159)	151,5	1

4 **Obtén la mediana y la moda de los datos del peso de las 30 personas del ejemplo anterior.**

INTERVALO	x_i	f_i
[50, 60)	55	4
[60, 70)	65	7
[70, 80)	75	10
[80, 90)	85	7
[90, 100)	95	2

12 OBJETIVO 8
CALCULAR LAS MEDIDAS DE DISPERSIÓN DE UN CONJUNTO DE DATOS

NOMBRE: ______________________ CURSO: __________ FECHA: __________

- Las **medidas de dispersión** son las medidas estadísticas que indican el mayor o menor grado de agrupamiento de los valores que forman un conjunto de datos.
- El rango o recorrido, la desviación, la desviación media, la varianza y la desviación típica son medidas de dispersión.
- El **rango o recorrido** se calcula como la diferencia entre el valor mayor y el valor menor de la variable estadística.
- La **desviación respecto a la media** es la diferencia entre cada valor de la variable y la media de dicha variable.

  La suma de las desviaciones es siempre cero.

EJEMPLO

En una ciudad hay dos coros, A y B, formados por 9 personas cada uno, cuyas edades son:

CORO A	10	10	20	30	30	30	40	50	50
CORO B	25	25	30	30	30	30	30	35	35

Halla la media de edad de cada coro, la mediana, la moda, el recorrido y la desviación.

- La media de cada coro es:

$$\bar{x}_A = \frac{10 + 10 + 20 + 30 + 30 + 30 + 40 + 50 + 50}{9} = \frac{270}{9} = 30 \text{ años}$$

$$\bar{x}_B = \frac{25 + 25 + 30 + 30 + 30 + 30 + 30 + 35 + 35}{9} = \frac{270}{9} = 30 \text{ años}$$

- La mediana es 30 en ambos casos, ya que ocupa el lugar central en cada serie.
- La moda es también igual en las dos series, 30.

  Se observa que los dos coros tienen los tres promedios iguales, pero son muy desiguales respecto a las edades. El coro A tiene dos niños de 10 años y dos personas mayores de 50 años, es decir, hay cuatro valores muy alejados de la edad media; en cambio, en el coro B, las edades son más homogéneas, pues todas están próximas a la media.

  Si queremos tener en cuenta estos aspectos, hay que medir el grado de separación o de dispersión de los datos con respecto a la media.
- El recorrido del coro A es: $50 - 10 = 40$ años, mientras que el recorrido del coro B es: $35 - 25 = 10$ años.
- Para los componentes de cada coro, las desviaciones son:

CORO A	10	10	20	30	30	30	40	50	50
$x - 30$	−20	−20	−10	0	0	0	10	20	20

CORO B	25	25	30	30	30	30	30	35	35
$x - 30$	−5	−5	0	0	0	0	0	5	5

Observa que, en ambos casos, la suma de las desviaciones es cero.

1 **En un examen se han obtenido las siguientes notas: 3, 5, 7, 2, 9, 5 y 3. Obtén el recorrido y la desviación.**

- La **desviación media (*DM*)** es la media de los valores absolutos de las desviaciones.

$$DM = \frac{\Sigma |x_i - \overline{x}|}{N}$$

Si los datos vienen dados con sus frecuencias, la desviación media es:

$$DM = \frac{\Sigma |x_i - \overline{x}| \cdot f_i}{N}$$

- La **varianza** es la media de los valores absolutos de las desviaciones.
  Mide las desviaciones respecto de la media al cuadrado.

- La **desviación típica** es la raíz cuadrada de la varianza. Se designa con la letra σ.

  Para datos simples, su fórmula es: $\sigma = \sqrt{\dfrac{\Sigma(x_i - \overline{x})^2}{N}}$

  Para datos agrupados, su fórmula es: $\sigma = \sqrt{\dfrac{\Sigma(x_i - \overline{x})^2 \cdot f_i}{N}}$

EJEMPLO

Halla la desviación media, la varianza y la desviación típica de los dos coros del ejemplo anterior.

En este caso, los datos no vienen dados con sus frecuencias (la frecuencia de cada dato es 1).

- Las desviaciones medias son:

$$DM_A = \frac{|-20| + |-20| + |-10| + 0 + 0 + 0 + |10| + |20| + |20|}{9} = \frac{100}{9} = 11{,}11$$

$$DM_B = \frac{|25 - 30| + |25 - 30| + 0 + 0 + 0 + 0 + 0 + |35 - 30| + |35 - 30|}{9} = \frac{20}{9} = 2{,}22$$

Se observa que la desviación media del coro A es mayor que la del coro B.

- Las varianzas y las desviaciones típicas de ambos coros son:

$$\text{Varianza del coro A} = \frac{d_1^2 + d_2^2 + d_3^2 + d_4^2 + d_5^2 + d_6^2 + d_7^2 + d_8^2 + d_9^2}{9} =$$

$$= \frac{(-20)^2 + (-20)^2 + (-10)^2 + 0^2 + 0^2 + 0^2 + 0^2 + 10^2 + 20^2 + 20^2}{9} =$$

$$= \frac{400 + 400 + 100 + 0 + 0 + 0 + 100 + 400 + 400}{9} = \frac{1.800}{9} = 200$$

Desviación típica: $\sigma_A = \sqrt{\text{varianza}} = \sqrt{200} = 14{,}14$

$$\text{Varianza del coro B} = \frac{(-5)^2 + (-5)^2 + 0^2 + 0^2 + 0^2 + 0^2 + 5^2 + 5^2}{9} =$$

$$= \frac{25 + 25 + 0 + 0 + 0 + 0 + 0 + 25 + 25}{9} = \frac{100}{9} = 11{,}11$$

Desviación típica: $\sigma_B = \sqrt{\text{varianza}} = \sqrt{11{,}11} = 3{,}33$

Se observa que la desviación típica en el coro A es bastante mayor que en el coro B, es decir, $\sigma_A > \sigma_B$.
Esto refleja que la dispersión de las edades en el coro A es mucho mayor que en el coro B.
La desviación típica aumenta a medida que se incrementa la dispersión de los datos.

ADAPTACIÓN CURRICULAR

2 **En la tabla aparecen los resultados obtenidos en una prueba de 120 preguntas. Halla la desviación media, la varianza y la desviación típica.**

INTERVALO	x_i	f_i
[0, 30)	15	12
[30, 60)	45	20
[60, 90)	75	10
[90, 120)	105	7

3 **Las alturas (en cm) de los alumnos de una clase de 4.º ESO se distribuyen según la tabla.**

INTERVALOS DE ALTURAS	FRECUENCIAS ABSOLUTAS
[145, 150)	51
[150, 155)	95
[155, 160)	141
[160, 165)	152
[165, 170)	120
[170, 175)	88
[175, 180)	58

Resuelve.

a) Completa la tabla con las frecuencias absolutas acumuladas y las frecuencias relativas.

b) Dibuja el histograma y el polígono de frecuencias correspondiente.

c) Calcula la media aritmética.

d) Determina el intervalo modal y toma como moda la marca de clase correspondiente.

e) Calcula la mediana.

f) Halla la desviación típica.

(Sugerencia: averigua el intervalo cuya frecuencia absoluta acumulada** F_i **es inmediatamente superior que la mitad del número de datos. Ese es el intervalo en el que se encuentra la mediana, y su marca de clase se puede tomar como valor aproximado de la mediana.)