

# 10 Distribuciones estadísticas

**E**n la Estadística distinguimos dos partes perfectamente diferenciadas. La primera de ellas se dedica a recoger datos, ordenarlos, simplificarlos, clasificarlos y obtener a partir de ellos un conjunto de valores que los identifican, y que además permiten hacer comparaciones con otros conjuntos de datos y estudiar relaciones entre ellos; a esta parte se le conoce por Estadística Descriptiva.

La otra parte de la Estadística se denomina Inferencia estadística o Estadística Inferencial y tiene por objeto hacer afirmaciones sobre una población a partir de los datos recogidos de una muestra de esa población. En segundo curso de Bachillerato estudiaremos los conceptos básicos de la Inferencia estadística.

Los **objetivos** que nos proponemos alcanzar con el estudio de esta Unidad son los siguientes:

1. Ordenar y disponer los datos de una variable estadística en una tabla de distribución de frecuencias.
2. Representar gráficamente los datos de una tabla de distribución de frecuencias.
3. Calcular y utilizar las medidas de centralización para asignar un número que identifique a los datos de la variable estadística.
4. Calcular y utilizar las medidas de dispersión para conocer el grado de variación de los datos respecto a la media.
5. Conocer la utilidad del coeficiente de variación para comparar distribuciones con la misma dispersión y saber cuál está más concentrada alrededor de la media.

## ÍNDICE DE CONTENIDOS

<b>1. VARIABLES ESTADÍSTICAS</b> .....	<b>243</b>
1.1. Tablas de distribución de frecuencias .....	243
1.2. Representaciones gráficas .....	245
<b>2. MEDIDAS DE CENTRALIZACIÓN DE UNA VARIABLE ESTADÍSTICA</b> .....	<b>247</b>
2.1. Media aritmética .....	248
2.2. Mediana .....	249
2.3. Moda .....	250
2.4. Cuartiles .....	252
<b>3. MEDIDAS DE DISPERSIÓN</b> .....	<b>254</b>
3.1. Recorrido o rango .....	254
3.2. Desviación media .....	255
3.3. Varianza .....	255
3.4. Desviación típica .....	255
<b>4. COEFICIENTE DE VARIACIÓN</b> .....	<b>259</b>

# 1. Variables estadísticas

La estadística estudia una característica o carácter de un conjunto de individuos llamado **población**. Cuando la población es muy grande se recurre a un subconjunto denominado **muestra**.

El carácter en estudio puede adoptar diversas modalidades. A todas las modalidades que puede adoptar un determinado carácter le llamamos **variable estadística**.

La variable estadística puede ser cuantitativa o cualitativa. Por ejemplo, en la población de los alumnos de un instituto: edad, estatura, calificaciones, número de hermanos, etc., son caracteres cuantitativos; es decir, son caracteres contables o medibles. Mientras que el lugar de nacimiento, el nombre de su equipo de fútbol favorito, etc., son caracteres cualitativos.

Los números que sirven para contar o medir un carácter cuantitativo, y que pueden variar con cada individuo, constituyen lo que se llama una **variable estadística cuantitativa**.

Las variables estadísticas cuantitativas pueden ser continuas o discretas. Las variables son continuas si los valores que pueden tomar son los números reales de un intervalo, mientras que las variables discretas sólo toman valores enteros positivos.

## 1.1. Tablas de distribución de frecuencias

Las variables estadísticas, cuantitativas o cualitativas, junto con las frecuencias de cada modalidad constituyen una tabla de distribución de frecuencias o, simplemente, una **distribución de frecuencias**. Estas tablas tienen, además de los valores de las variables y sus frecuencias, otras columnas: la frecuencia relativa, la frecuencia acumulada, etc. Veamos en los ejemplos cómo se construyen.



### Ejemplos

1. En una encuesta realizada a 20 familias, se les preguntó por el número de hijos y se han anotado los siguientes datos: 2, 0, 1, 3, 1, 1, 2, 0, 1, 0, 2, 3, 5, 2, 1, 0, 2, 1, 1, 4. Construye una tabla de distribución de frecuencias absolutas y relativas. ¿ Podrías decir qué porcentaje de familias tiene más de dos hijos? ¿Y qué porcentaje de familias tiene menos de tres hijos?

*Solución.* Llamando  $x_i$  a los valores de la variable estadística,  $f_i$  a sus frecuencias abso-

lutas,  $h_i = \frac{f_i}{n}$  a las frecuencias relativas, donde  $n$  es el número de datos,  $F_i$  a las frecuencias absolutas acumuladas y  $H_i$  a las frecuencias relativas acumuladas podemos confeccionar la tabla siguiente:

# UNIDAD 10

## DISTRIBUCIONES ESTADÍSTICAS

$x_i$	$f_i$	$h_i$	$F_i$	$H_i$
0	4	0,20	4	0,20
1	7	0,35	11	0,55
2	6	0,25	16	0,80
3	2	0,10	18	0,90
4	1	0,05	19	0,95
5	1	0,05	20	1

Las frecuencias relativas expresan porcentajes, la suma de las frecuencias relativas de más de dos hijos son:  $0,10 + 0,05 + 0,05 = 0,20$ , el 20% de las familias de este estudio tiene más de dos hijos.

El porcentaje de familias con menos de tres hijos nos lo da  $H_2$ ,  $H_2 = 0,80 = 80\%$ .

2. Se toma el pulso a un grupo de 30 personas, obteniéndose los datos siguientes: 72, 66, 81, 74, 57, 58, 74, 62, 73, 65, 78, 75, 84, 72, 69, 76, 65, 79, 76, 68, 82, 71, 77, 72, 56, 62, 83, 63, 70, 73. Construir una tabla de distribución de frecuencias.

*Solución.* Se trata de una variable discreta aunque con muchos valores. Cuando los valores de la variable son muy numerosos, los agrupamos en clases o intervalos. ¿Cuántas clases? Para responder a esta pregunta se puede utilizar el sentido común y establecer que no excedan la decena, y desde luego no menos que 5. Existe una fórmula llamada fórmula de Sturges que aconseja que sean:

$$N^{\circ} \text{ de clases} = 1 + 3,32 \cdot \log n,$$

siendo  $n$  el número de datos; siempre que sea posible, nos interesa que las clases sean todas de la misma amplitud.

En nuestro caso, es una variable discreta cuyo valor menor es 56 y el mayor 84, tomamos como rango de los valores de 55 a 84, y como clases: 55 – 59, 60 – 64, 65 – 69, 70 – 74, 75 – 79, 80 – 84. Y construimos la tabla siguiente:

Clases	marca de clase $x_i$	$f_i$	$F_i$
55 – 59	57	3	3
60 – 64	62	3	6
65 – 69	67	5	11
70 – 74	72	9	20
75 – 79	77	6	26
80 – 84	82	4	30
		$n = \sum f_i = 30$	

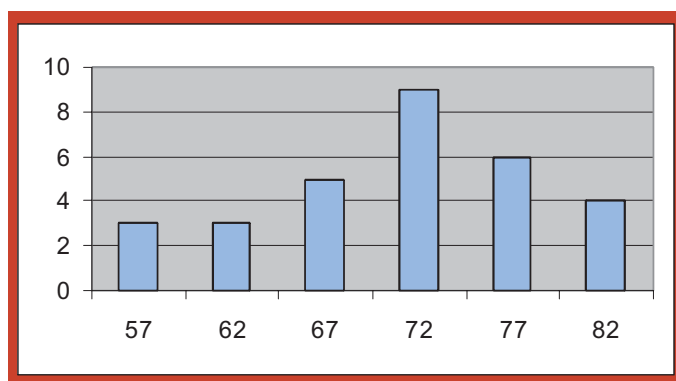
La **marca de clase**,  $x_i$ , se usa como un valor representativo de todos los de la clase, y corresponde al valor central de la clase, es decir, la semisuma de los extremos.

## 1.2. Representaciones gráficas

El objeto de las representaciones gráficas es mostrar de un modo ordenado y agradable la información numérica para facilitar su estudio e interpretación. A partir de las tablas de distribución de frecuencias se construyen representaciones gráficas para este propósito. Veamos algunas muy conocidas.

### ***El diagrama de rectángulos***

Es una gráfica que se emplea para variables cualitativas y cuantitativas discretas. Está formada por tantos rectángulos como valores o modalidades de la variable y todos con igual base. La altura de los rectángulos es igual a la frecuencia absoluta o relativa de cada valor. En el caso del ejemplo 2 anterior el diagrama de rectángulos sería:



### ***Diagrama de sectores***

Se emplea también con variables cualitativas y cuantitativas discretas. Consiste en dividir un círculo en sectores proporcionales a las frecuencias. Para dibujar un sector necesitamos conocer el ángulo central, y que se calcula por la proporción:

$$\frac{\text{ángulo del sector}}{360^\circ} = \frac{\text{frecuencia}}{n^\circ \text{ de datos}}$$

En el caso del ejemplo 1 anterior correspondiente al número de hijos de 20 familias, los ángulos centrales son los siguientes:

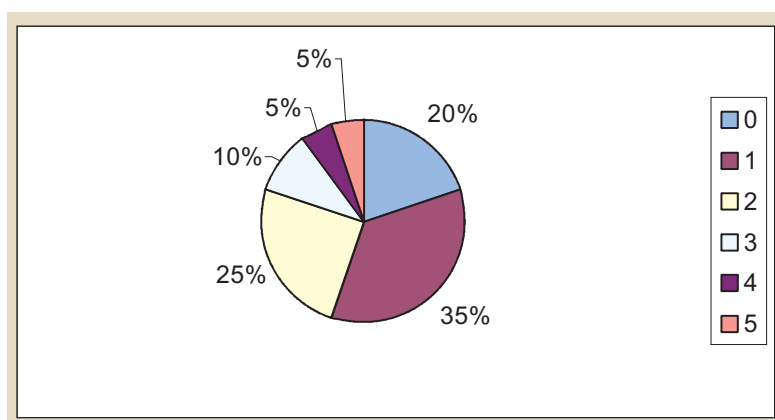
$$\frac{\text{ángulo } 0}{360} = \frac{4}{20}, \text{ ángulo } 0 = \frac{4 \cdot 360}{20} = 72^\circ; \quad \frac{\text{ángulo } 1}{360} = \frac{7}{20}, \text{ ángulo } 1 = \frac{7 \cdot 360}{20} = 126^\circ$$

$$\frac{\text{ángulo } 2}{360} = \frac{5}{20}, \text{ ángulo } 2 = 90^\circ; \quad \text{ángulo } 3 = 36^\circ; \quad \text{ángulo } 4 = 18^\circ; \quad \text{ángulo } 5 = 18^\circ$$

Con un transportador de ángulos, sobre un círculo, no es difícil dibujar un diagrama de sectores como este:

# UNIDAD 10

## DISTRIBUCIONES ESTADÍSTICAS



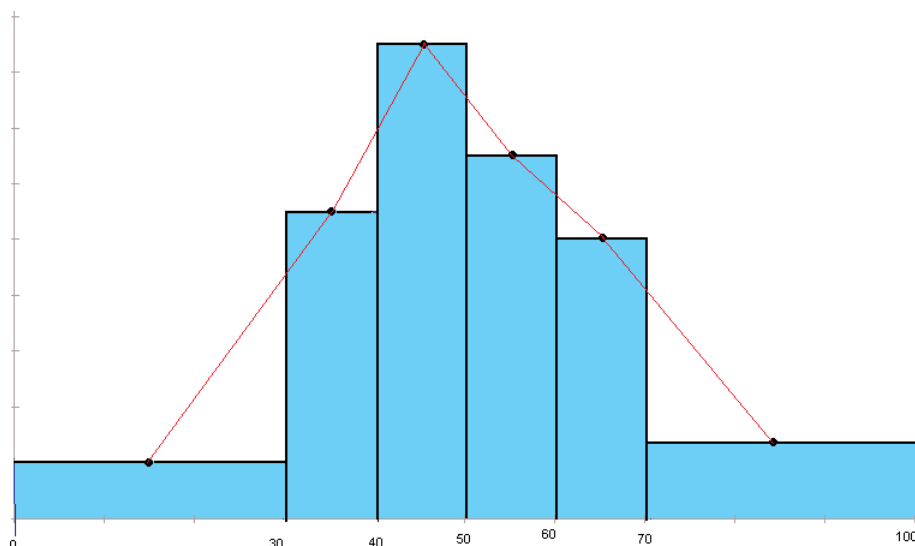
### Histograma

Cuando la variable es cuantitativa y los datos están agrupados en intervalos, se construye un diagrama de rectángulos especial, llamado histograma, en que las áreas de éstos, y no las alturas, indican las frecuencias. Al ser el área la frecuencia, en los histogramas se cumple que: **Frecuencia = amplitud de la clase · altura**

Por ejemplo, si tuviéramos una distribución como:

clases	$f_i$	$h_i$
0 – 30	6	0,10
30 – 40	11	0,17
40 – 50	17	0,26
50 – 60	13	0,20
60 – 70	10	0,15
70 – 100	8	0,12
	65	1

que corresponde a las puntuaciones obtenidas por 65 personas en un test de 100 preguntas, entonces el histograma sería:



Al histograma le hemos añadido **el polígono de frecuencias**, que aparece si unimos por trazos rectos los puntos medios de las bases superiores de los rectángulos del histograma. Cuando el número de datos crece indefinidamente y al mismo tiempo hacemos los intervalos cada vez más pequeños, el polígono de frecuencias tiende a una curva suave y continua que, si se trata de frecuencias relativas, es conocida con el nombre de función densidad, y de la que oiremos hablar en la Unidad 12.



### Actividades

1. En una encuesta realizada a 40 familias, en la que se preguntaba el número de hijos, se obtuvieron los siguientes datos: 1, 1, 2, 0, 1, 0, 2, 3, 0, 2, 3, 5, 0, 1, 0, 2, 2, 0, 1, 3, 1, 1, 0, 5, 0, 1, 0, 2, 1, 1, 0, 2, 0, 1, 3, 1, 1, 2, 0, 4.

Construye una tabla de distribución de frecuencias que incluya frecuencias relativas y porcentajes. ¿Qué porcentaje de familias no tienen hijos? ¿Cuántas familias tienen menos de dos hijos?

2. En un test de 100 preguntas, contestado por 120 alumnos, se han obtenido los siguientes resultados:

Respuestas correctas	Nº de alumnos
[0, 10)	6
[10, 20)	4
[20, 30)	12
[30, 40)	8
[40, 50)	14
[50, 60)	15
[60, 70)	12
[70, 80)	23
[80, 90)	19
[90, 100)	7

- a) Completa la tabla con las columnas de frecuencias relativas, porcentajes y frecuencias acumuladas.
- b) Dibuja el histograma y el polígono de frecuencias de la tabla.
- c) Calcula el porcentaje de alumnos que han contestado a menos de 50 preguntas correctamente.

## 2. Medidas de centralización de una variable estadística

A veces es necesario disponer de un valor numérico que represente la diversidad de valores de una distribución de frecuencias de una variable estadística. A los valores numéricos que cumplen esta función se les llama parámetros centrales o medidas de centralización de una distribución, y, como sabemos, son: **la media, la mediana y la moda.**

### 2.1. Media aritmética

La **media aritmética** de un conjunto de números es el cociente que resulta de dividir la suma de todos los números por el total de éstos. Se representa por  $\bar{x}$ .

Si los valores de una variable estadística no tienen frecuencias, y son:  $x_1, x_2, \dots, x_n$ , es decir  $n$  valores, la media aritmética se calcula por la fórmula,

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

Si los datos vienen con frecuencias y los  $n$  datos se distribuyen en  $k$  valores de la variable, o de las marcas de clase cuando están agrupados, y éstos son  $x_1, x_2, \dots, x_k$ ; y  $f_1, f_2, \dots, f_k$ , las frecuencias respectivas, donde  $f_1 + f_2 + \dots + f_k = n$ , entonces la media aritmética se calcula con la fórmula:

$$\bar{x} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \dots + x_k \cdot f_k}{n} = \frac{\sum_{i=1}^k x_i \cdot f_i}{n}$$

La media aritmética es la medida o parámetro de centralización más empleado. Con todo, está influida por los valores extremos de una distribución o por algún valor extravagante, en estos casos pierde algo de significado. Refleja, sin embargo, todas las alteraciones que sufran los datos:

- Si se suma la misma cantidad a todos los valores de la variable, la media resulta aumentada en esa cantidad.
- Si se multiplican todos los valores de la variable por el mismo número, la media resulta multiplicada por ese número.

Existe además otra media de un conjunto de  $n$  números, que se llama **media ponderada**, y se calcula cuando los datos no tienen todos el mismo peso. La media ponderada se obtiene sumando todos los productos de cada valor por su peso dividiendo el resultado por la suma de los pesos. La fórmula es:

$$\bar{x}_p = \frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n}{p_1 + p_2 + \dots + p_n} = \frac{\sum_{i=1}^n x_i \cdot p_i}{\sum_{i=1}^n p_i}$$

## 2.2. Mediana

**Mediana** es el valor de la variable estadística, suponiendo que los datos estén ordenados, que ocupa la posición central; es decir, deja a su izquierda el mismo número de datos que a su derecha. Se simboliza por  $Me$ .

En una distribución sencilla de variable discreta, la mediana corresponde al valor central si el número de datos es impar, pero si el número de datos es par, la mediana es la media de los dos valores centrales.

En: 0, 0, 1, 1, 1, 1, 2, 2, 2,  $Me = 1$ .

En: 1, 2, 3, 4, 5, 6, 7, 8,  $Me = \frac{4+5}{2} = 4,5$

En el caso de una distribución con muchos datos, pero sin agrupar, la mediana es el primer valor cuya frecuencia absoluta acumulada está por encima de la mitad de los datos.

Por ejemplo, en la distribución 

$x_i$	0	1	2	3	4
$f_i$	5	6	9	4	3

,  $n = 27$ ,  $n/2 = 13,5$  y las frecuencias absolutas acumuladas son: 5, 11, 20, 24, 27, la primera frecuencia acumulada por encima del mitad de los datos es 20, y corresponde a  $x_3 = 2$ , entonces  $Me = 2$ .

Sin embargo, en la distribución 

$x_i$	0	1	2	3	4
$f_i$	5	7	6	4	2

,  $n = 24$ ,  $n/2 = 12$  y la frecuencia acumulada correspondiente al valor 1 es también 12, en este caso tomamos como mediana la media entre los dos valores cuyas frecuencias acumuladas son igual y mayor que  $n/2$ ,  $Me = \frac{1+2}{2} = 1,5$

En el caso de una variable continua, cuyos datos se encuentran agrupados en intervalos, se llama **clase mediana** a aquella cuya frecuencia absoluta acumulada sobrepasa la mitad de los datos.

Como valor aproximado de la mediana puede tomarse la marca de clase de la clase mediana, pero si queremos mayor precisión, entonces la mediana es un número dentro de la clase mediana que se obtiene por la fórmula:

$$Me = e_i + \frac{\frac{n}{2} - F_{Me-1}}{f_{Me}} \cdot c$$



# UNIDAD 10

## DISTRIBUCIONES ESTADÍSTICAS

En dicha fórmula, tenemos que:

↪  $e_i$  es el extremo inferior de la clase mediana;

↪  $c$  es la amplitud de la clase mediana;

↪  $n$  es el número total de datos;

↪  $f_{Me}$  es la frecuencia absoluta de la clase mediana;

↪  $F_{Me-1}$  es la frecuencia absoluta acumulada inmediatamente anterior a la de la clase mediana.

La principal ventaja de la mediana como medida de centralización es que no está influida por los valores extremos, ni por datos extravagantes, aunque tiene serios inconvenientes: no es fácil su empleo en operaciones algebraicas y no tiene en cuenta el valor de todos los datos.

## 2.3. Moda

Se conoce como **Moda** de una variable estadística al valor que tiene mayor frecuencia absoluta. Se simboliza por  $Mo$ .

Cuando la variable es discreta la moda se obtiene buscando el valor de la variable que tiene mayor frecuencia. A veces, la moda no es única, es decir, la distribución puede tener dos, tres o más modas, en cuyo caso recibe el nombre de bimodal, trimodal, etc.

Si los datos se encuentran agrupados en intervalos, la clase de mayor frecuencia se llama **clase modal**. Es frecuente tomar como moda la marca de clase de la clase modal, pero si queremos mayor precisión, entonces el cálculo de la moda se hace por la fórmula:

$$Mo = e_i + \frac{f_{Mo} - f_{Mo-1}}{(f_{Mo} - f_{Mo-1}) + (f_{Mo} - f_{Mo+1})} \cdot c$$

en la cual se tiene que:

↪  $e_i$  es el extremo inferior de la clase modal;

↪  $c$  es la amplitud de la clase modal;

↪  $f_{Mo}$ ,  $f_{Mo-1}$ ,  $f_{Mo+1}$  son, las frecuencias absolutas de la clase modal, la clase anterior a la clase modal y la posterior a la clase modal.

La moda tiene las mismas ventajas e inconvenientes que la mediana: no está influida por los valores extremos, ni por datos extravagantes, pero no tiene en cuenta el valor de todos los datos, ni se puede operar algebraicamente con ella.



## Ejemplo

El tiempo en minutos que tardan 87 empleados de una empresa en llegar desde su casa al trabajo se muestra en la tabla siguiente:

Minutos	(0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)	[30, 35]
empleados	6	7	18	22	17	11	6

- Calcula la duración media de los trayectos de casa al trabajo de los 87 empleados.
- Calcula la mediana correspondiente a la variable duración del trayecto casa-trabajo.
- Calcula la moda correspondiente a la variable duración del trayecto casa-trabajo.

*Solución.* a) Construimos una tabla para hallar la duración media

Minutos	Marca de clase, $x_i$	$f_i$	$x_i \cdot f_i$
(0, 5)	2,5	6	15
[5, 10)	7,5	7	52,5
[10, 15)	12,5	18	225
[15, 20)	17,5	22	385
[20, 25)	22,5	17	382,5
[25, 30)	27,5	11	302,5
[30, 35]	32,5	6	195
	Totales	87	1557,5

$$\bar{x} = \frac{1557,5}{87} = 17,9022$$

Por tanto, la duración media es 17,9 minutos.

Con una calculadora científica es todavía más fácil, con las teclas **MODE** · ponemos la calculadora en modo estadístico, en la pantalla aparece SD (standard deviation), y ya están activas las teclas escritas en azul.

Borramos, antes que nada los datos de la memoria, por si los hubiera, con la tecla **SHIFT** **SAC** e introducimos los valores y sus frecuencias

2.5 × 6 **DATA** 7.5 × 7 **DATA** 12.5 × 18 **DATA** 17.5 × 22 **DATA** 22.5 × 17 **DATA**

27.5 × 11 **DATA** 32.5 × 6 **DATA**

Una vez introducidos los datos las teclas **SHIFT** y  $\bar{x}$  dan la media.

- Como  $n/2 = 43,5$  la clase mediana es [15, 20), la primera cuya frecuencia acumulada sobrepasa a la mitad de los datos, la marca de clase correspondiente, 17,5, será la mediana. Si queremos afinar más, por la fórmula de la mediana, obtenemos:

$$Me = 15 + \frac{\frac{87}{2} - 31}{22} \cdot 5 = 17,84$$

Es decir, el 50% de los empleados tarda más de 17,84 minutos y el 50% menos que esa cantidad.

- c) La mayor frecuencia es 22 y la clase modal será también [15, 20). Un primer valor para la moda sería la marca de clase 17,5. Si queremos más precisión, usamos la fórmula de la moda y obtenemos:

$$Mo = 15 + \frac{22 - 18}{(22 - 18) + (22 - 17)} \cdot 5 = 17,2222$$

El valor 17,22 es el que tiene mayor frecuencia.

## 2.4. Cuartiles

Si ordenamos los datos de menor a mayor, se llaman **cuartiles** a los valores de la variable que dividen al conjunto de datos en cuatro partes iguales. El primero se simboliza por  $Q_1$  y deja a su izquierda el 25% de los datos y a su derecha el 75%, el segundo cuartil coincide con la mediana,  $Q_2 = Me$ , y el tercer cuartil se simboliza por  $Q_3$  y deja a su izquierda el 75% de los datos y a su derecha el 25%.

El cálculo de los cuartiles primero y tercero es similar al de la mediana y se obtienen por las fórmulas:

$$Q_1 = e_1 + \frac{\frac{n}{4} - F_{Q_1-1}}{f_{Q_1}} \cdot c$$

$$Q_3 = e_3 + \frac{\frac{3n}{4} - F_{Q_3-1}}{f_{Q_3}} \cdot c$$

en donde:

- ☞  $e_1$  o  $e_3$  es el extremo inferior de la clase a la que pertenece el cuartil;
- ☞  $c$  es la amplitud de la clase a la que pertenece el cuartil;
- ☞  $n$  es el número total de datos;
- ☞  $f_{Q_1}$  o  $f_{Q_3}$  es la frecuencia absoluta de la clase a la que pertenece el cuartil;
- ☞  $F_{Q_1-1}$  o  $F_{Q_3-1}$  es la frecuencia absoluta acumulada inmediatamente anterior a la de la clase a la que pertenece el cuartil.

Como en el caso de la mediana, la clase a la que pertenece  $Q_1$  es aquella cuya frecuencia absoluta acumulada sobrepasa la cuarta parte de los datos, y la clase a la que pertenece  $Q_3$  es aquella cuya frecuencia absoluta acumulada sobrepasa las tres cuartas partes de los datos.



## Ejemplo

Un test de 100 preguntas, realizado a 90 alumnos, ha arrojado las puntuaciones que figuran en la tabla adjunta.

Hallar  $Q_1$ ,  $Q_2$  o  $Me$ ,  $Q_3$ .

*Solución.* La cuarta parte, la mitad y las tres cuartas partes de los datos son: 22,5, 45 y 67,5; por lo tanto,  $Q_1$  pertenece a la clase 60 – 69, la clase mediana es 70 – 79 y  $Q_3$  pertenece a 80 – 89.

Calificación	Nº de estudiantes
0 - 39	2
40 - 49	3
50 - 59	9
60 - 69	17
70 - 79	32
80 - 89	19
90 - 100	8

Empleamos la fórmula que permite calcular los cuartiles,  $Q_1 = 60 + \frac{\frac{90}{4} - 14}{17} \cdot 10 = 65$

$$Me = Q_2 = 70 + \frac{\frac{90}{2} - 31}{32} \cdot 10 = 74,37 \qquad Q_3 = 80 + \frac{\frac{90 \cdot 3}{4} - 63}{19} \cdot 10 = 82,36$$

Entre 65 y 82,36 están el 50% de las puntuaciones centrales registradas, a la diferencia  $Q_3 - Q_1$  se denomina el recorrido intercuartílico  $82,36 - 65 = 17,64$

A los cuartiles, y a otras medidas semejantes llamadas deciles y percentiles, también se les conoce como medidas de posición porque permiten situar un dato con respecto al resto de la distribución.



## Actividades

- Una asignatura se aprueba con un examen y un trabajo. El trabajo supone el 30% de la nota. ¿Qué nota tendrá que sacar un alumno en el examen si en el trabajo obtuvo un 8?
- Un profesor ha realizado cuatro exámenes a sus alumnos: el segundo cuenta el doble que el primero, el tercero el triple que el primero y el cuarto el doble que el primero. ¿Cuál será la nota final de un alumno que obtuvo en cada examen: 3, 5, 4 y 6?
- Para determinar la nota de selectividad, el 60% corresponde a la media de los dos cursos de Bachillerato y el 40% a la calificación obtenida en las pruebas de acceso. Si un alumno tenía 6,3 como media del Bachillerato y consiguió 5,2 en las pruebas de acceso, ¿cuál será su nota de selectividad? ¿Qué nota mínima debe conseguir en las pruebas de acceso un alumno que tiene de media de Bachillerato un 5,2?
- Las alturas en centímetros de 12 chicos son: 158, 162, 176, 184, 167, 196, 158, 165, 176, 184, 198, 165. Halla la media, mediana y moda.
- Se toma el pulso a un grupo de 30 personas, obteniéndose los datos siguientes: 72, 66, 81, 74, 57, 58, 74, 62, 73, 65, 78, 75, 84, 72, 69, 76, 65, 79, 76, 68, 82, 71, 77, 72, 56, 62, 83, 63, 70, 73. Calcular la media, la mediana, la moda y el primer cuartil.

# UNIDAD 10

## DISTRIBUCIONES ESTADÍSTICAS

8. Se han anotado las duraciones de 50 viajes de ferrocarril entre dos ciudades y se ha confeccionado la siguiente tabla:

Duración viaje en minutos	Frecuencia
60 – 69	25
70 – 79	10
80 – 89	9
90 – 99	4
100 – 109	2

Calcula la duración media del viaje, la mediana, la moda y el tercer cuartil.

9. Un pediatra ha registrado el peso de 45 niños y niñas de 6 meses, y ha confeccionado la tabla siguiente:

Peso en kg	Frecuencia
[6, 7)	5
[7, 8)	11
[8, 9)	18
[9, 10)	8
[10, 11)	3

Calcula el peso medio, la mediana y el recorrido intercuartilico de la variable peso de todas esas criaturas.

10. Las temperaturas máximas en una ciudad durante el mes de mayo han sido: 26, 26, 30, 30, 31, 27, 28, 27, 27, 26, 26, 37, 27, 28, 28, 29, 27, 26, 25, 22, 28, 28, 27, 28, 29, 30, 26, 25, 27, 29, 37.

Calcula la media moda y la mediana.

Edad	Nº de accidentes
[20, 25)	1231
[25, 30)	610
[30, 35)	442
[35, 40)	336
[40, 50)	173
[50, 60)	95
[60, 90)	30

11. Se han registrado los accidentes de tráfico y las edades de los conductores durante cierto periodo de tiempo en un determinado país obteniéndose la siguiente distribución por edades. Calcula los tres cuartiles.

## 3. Medidas de dispersión

Puede ocurrir que dos distribuciones distintas tengan los parámetros centrales muy parecidos o iguales y sin embargo las distribuciones ser completamente diferentes. Por tanto, se necesitan otras medidas que nos indiquen el grado de variación de los datos respecto a la media. Estas nuevas medidas, que llamamos parámetros o medidas de dispersión, informan del grado de dispersión o de concentración de los datos en relación a la media. Veamos cuales son.

### 3.1. Recorrido o rango

El **recorrido o rango** de una variable estadística es la diferencia entre el mayor y el menor valor de los datos. Se simboliza  $R$ , es decir,

$$R = x_{\max} - x_{\min}$$

## 3.2. Desviación media

Se define la **desviación media** de una variable estadística a la suma de los productos de los valores absolutos de las desviaciones de los valores o marcas de clase respecto de la media aritmética multiplicados por sus frecuencias y divididos por el total de los datos. Se simboliza por  $DM$  y viene dada por:

$$DM = \frac{\sum_{i=1}^k |x_i - \bar{x}| \cdot f_i}{n}$$

## 3.3. Varianza

La **varianza** es la media aritmética de los cuadrados de las desviaciones de todos los valores de la variable estadística o marcas de clase respecto de la media, y se simboliza por  $s^2$ .

Si los valores de la variable son  $x_1, x_2, \dots, x_n$ , y no se repite ninguno, entonces

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

La varianza,  $s^2$ , también se puede expresar por:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2x_i \cdot \bar{x} + \sum_{i=1}^n \bar{x}^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \frac{2\bar{x} \sum_{i=1}^n x_i}{n} + \frac{\sum_{i=1}^n \bar{x}^2}{n} =$$

$$\text{que también se puede expresar por: } = \frac{\sum_{i=1}^n x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

Si los valores (o las marcas de clase) vienen con frecuencias, la varianza se calcula con las fórmulas que figuran a continuación que, como vimos anteriormente en el caso sencillo, son equivalentes:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot f_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^n x_i^2 \cdot f_i}{n} - \bar{x}^2$$

La varianza es siempre positiva, pero tiene un pequeño inconveniente no se expresa en las mismas unidades que los datos; es decir, si por ejemplo los datos son en cm la varianza sería  $\text{cm}^2$ .

## 3.4. Desviación típica

Para subsanar el pequeño defecto de la varianza se define la **desviación típica** de una variable estadística como la raíz cuadrada positiva de la varianza. Y ésta sí tiene las mismas unidades que los datos.

# UNIDAD 10

## DISTRIBUCIONES ESTADÍSTICAS

La representamos por la letra  $s$  y viene expresada por:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{n}} = \sqrt{\frac{\sum_{i=1}^k x_i^2 \cdot f_i}{n} - \bar{x}^2}$$

La desviación típica es el parámetro de dispersión más utilizado. Y tiene algunas propiedades interesantes:

- ✓ Si sumamos una cantidad constante a todos los valores de la variable, la desviación típica no varía.
- ✓ Si multiplicamos todos los valores de la variable por la misma cantidad, la desviación típica queda multiplicada por esa cantidad.



### Ejemplos

1. Se han registrado las distancias en km, desde su casa al trabajo, de 40 empleados de una empresa situada en un polígono industrial, y han resultado los siguientes datos: 3, 5, 10, 15, 20, 25, 3, 6, 12, 18, 23, 28, 4, 8, 14, 15, 22, 25, 8, 19, 10, 20, 9, 12, 16, 11, 15, 13, 18, 23, 10, 17, 22, 12, 15, 25, 14, 18, 20, 22.

Hallar la media, la varianza y la desviación típica

*Solución.* Agrupamos los datos en clases, de amplitud 5 km, que incluyan el extremo izquierdo del intervalo, pero no el derecho. Construimos con los datos suministrados la tabla siguiente:

Clases en km	Marcas de clase $x_i$	$f_i$	$x_i \cdot f_i$
[0, 5)	2,5	3	7,5
[5, 10)	7,5	5	37,5
[10, 15)	12,5	10	125,5
[15, 20)	17,5	10	175,5
[20, 25)	22,5	8	180
[25, 30)	27,5	4	110
		$\sum f_i = 40$	$\sum x_i \cdot f_i = 630$

La media de las distancias será:

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{630}{40} = 15,75 \text{ kilómetros}$$

La varianza es la media de los cuadrados de las desviaciones, o separaciones, de cada uno de los valores de la variable respecto a la media aritmética.

La desviación típica de una distribución estadística es la raíz cuadrada de la varianza.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{n}} = \sqrt{\frac{\sum_{i=1}^k x_i^2 \cdot f_i}{n} - \bar{x}^2}$$

Completamos el cuadro anterior con la columna  $x_i \cdot f_i$ . Luego multiplicando la primera columna de la tabla por la tercera se obtiene la columna encabezada por  $x_i^2 \cdot f_i$ .

$x_i$	$f_i$	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
2,5	3	7,5	18,75
7,5	5	37,5	281,25
12,5	10	125,5	1562,5
17,5	10	175,5	3062,5
22,5	8	180	4050
27,5	4	110	3025
	$\sum f_i = 40$		$\sum x_i^2 \cdot f_i = 12000$

La varianza  $s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{n} = \frac{\sum x_i^2 \cdot f_i}{n} - \bar{x}^2 = \frac{12000}{40} - 15,75^2 = 51,9375$ .

La desviación típica  $s = \sqrt{s^2} = \sqrt{51,9375} = 7,20676\dots$

2. Al pesar 32 profesores de un Instituto se obtuvieron los siguientes datos en kg: 48,5, 52, 58, 60, 65, 70,5, 77, 83, 89, 49, 53, 59, 61,5, 68, 74, 78,5, 55, 62, 67, 73, 63, 66, 71,5, 62, 65, 72, 64, 66, 71, 63,5, 69, 68,5.

Calcular la media, la varianza y la desviación típica.

*Solución.* El mayor valor es 89 y el menor 48,5. Los agrupamos en clases o intervalos semicerrados, de 5 kg, comenzando en 45 y terminando en 90. Hacemos un cuadro con las clases, las marcas de clase y la frecuencia.

Clases	$x_i$	$f_i$
[45, 50)	47,5	2
[50, 55)	52,5	3
[55, 60)	57,5	2
[60, 65)	62,5	7
[65, 70)	67,5	8
[70, 75)	72,5	6
[75, 80)	77,5	2
[80, 85)	82,5	1
[85, 90)	87,5	1
		$\sum f_i = 32$

Este ejemplo lo vamos a resolver con ayuda de una calculadora.

- 1º) Establecemos el modo estadístico. Pulsamos **MODE** . Aparece en pantalla las letras SD.
- 2º) Borramos el contenido de la memoria estadística **SHIFT** **SAC** .
- 3º) Introducimos los valores de la tabla:

47.5 × 2 **DATA** 52.5 × 3 **DATA** 57.5 × 2 **DATA** 62.5 × 7 **DATA** 67.5 × 8 **DATA**  
72.5 × 6 **DATA** 77.5 × 2 **DATA** 82.5 **DATA** 87.5 **DATA**



# UNIDAD 10

## DISTRIBUCIONES ESTADÍSTICAS

4º) Con **SHIFT**  $\bar{x}$  obtenemos la media,  $\bar{x} = 65,78125$

Con **SHIFT**  $\sigma_n$  obtenemos la desviación típica,  $s = 9,2372\dots$

Con **SHIFT**  $\sigma_n$  **SHIFT**  $x^2$  obtenemos la varianza,  $s^2 = 85,3271\dots$

En las calculadoras aparece la tecla  $\sigma_{n-1}$  que corresponde a un parámetro cuyo cuadrado se llama cuasivarianza y tiene importancia en el estudio de las estimaciones estadísticas, pero que no es nuestro caso. En otros modelos y marcas de calculadoras debe consultarse el manual de uso, aunque todas tienen instrucciones similares.



### Actividades

12. Las alturas en centímetros de 12 chicos son: 158, 162, 176, 184, 167, 196, 158, 165, 176, 184, 198, 165. Hallar la varianza y la desviación típica.
13. Se toma el pulso a un grupo de 30 personas, obteniéndose los datos siguientes: 72, 66, 81, 74, 57, 58, 74, 62, 73, 65, 78, 75, 84, 72, 69, 76, 65, 79, 76, 68, 82, 71, 77, 72, 56, 62, 83, 63, 70, 73. Calcular la desviación media, la varianza y la desviación típica.
14. Las temperaturas máximas en una ciudad durante el mes de mayo han sido: 26, 26, 30, 30, 31, 27, 28, 27, 27, 26, 26, 37, 27, 28, 28, 29, 27, 26, 25, 22, 28, 28, 27, 28, 29, 30, 26, 25, 27, 29, 37. Calcula la varianza y la desviación típica

15. El sueldo mensual de 35 empleados de una empresa viene dado por:

Sueldo(euros)	550	780	975	1100	1600	1800
Nº empleados	10	8	7	5	3	2

Calcular la media, la varianza y la desviación típica.

16. Se han anotado las duraciones de 50 viajes de ferrocarril entre dos ciudades y se ha confeccionado la siguiente tabla:

Duración viaje en minutos	Frecuencia
60 – 69	25
70 – 79	10
80 – 89	9
90 – 99	4
100 – 109	2

Como ya sabes la duración media del viaje, calcula la varianza y la desviación típica.

## 4. Coeficiente de variación

Al comparar la dispersión de dos distribuciones, puede ocurrir que tengan la misma desviación típica y que las dispersiones sean totalmente diferentes. No es lo mismo una desviación típica de 2 kg en una población pollos que en una vacas. Por lo que vamos a definir un parámetro que nos permite medir la desviación típica, tomando la media como unidad, y consiste en calcular el cociente  $\frac{S}{\bar{X}}$ .

A este cociente se llama **coeficiente de variación** y suele expresarse en tantos por cien, por lo tanto habrá que multiplicarlo por 100, y lo simbolizamos por  $V_p$ ,

$$V_p = \frac{S}{\bar{X}} \cdot 100$$

El coeficiente de variación mide, en tantos por cien, la desviación de una dispersión con respecto a la media. Cuanto más pequeño sea el coeficiente de variación, los datos estarán más concentrados alrededor de la media, y ésta resultará más significativa.



### Ejemplo

Tenemos dos marcas de bombillas, A y B, elegimos dos muestras de 10 bombillas de cada marca y anotamos la duración, en horas, de cada una de ellas, resultando:

A	260	285	250	350	290	280	245	265	230	235
B	250	310	270	290	325	300	280	285	240	315

Calcula el coeficiente de variación de cada muestra.

*Solución.* Empleando la calculadora obtenemos que en las bombillas de la marca A,  $\bar{x}_A = 269$  y  $s_A = 33,3$ , mientras que en las bombillas de la marca B,  $\bar{x}_B = 286,5$  y  $s_B = 26,17$ .

Los coeficientes de variación serán:

$$V_p = \frac{33,3}{269} \cdot 100 = 12,37\%, \text{ para la marca A, y } V_p = \frac{26,17}{286,5} \cdot 100 = 9,13\% \text{ para la marca}$$

B. La marca B tiene una duración media mayor y además la dispersión es menor con respecto a la media.



### Actividades

- Las alturas en cm de los jugadores de dos equipos de fútbol son:  
Equipo A: 187, 170, 183, 177, 182, 174, 187, 180, 172, 177, 178, 187, 179, 174, 188, 175, 181.  
Equipo B: 188, 170, 176, 178, 185, 180, 185, 177, 173, 174, 174, 179, 182, 171, 191, 188, 168, 165.  
Calcula el coeficiente de variación de cada equipo y compara su dispersión
- El peso de 5 chicos es: 55, 63, 57, 66 y 65 kg, y sus alturas respectivas: 175, 168, 174, 179 y 181 cm. Halla el coeficiente de variación de cada grupo de datos y compara su dispersión.