PARÁMETROS ESTADÍSTICOS

Puesto que las representaciones gráficas no siempre consiguen ofrecer una información completa de una serie de datos, es necesario analizar procedimientos numéricos que permitan resumir toda la información del fenómeno en estudio en unos números llamados **parámetros estadísticos**. Los parámetros estadísticos se pueden clasificar en:

a) **Medidas de centralización**.- Que representan a toda la distribución. Buscan características del centro de la distribución. Los más importantes son la media aritmética, la mediana y la moda.

b) **Medidas de posición**.- Indican, una vez ordenados los datos, cuantos elementos quedan a la izquierda o derecha de uno dado: cuartiles, deciles, centiles o percentiles.

c) **Medidas de dispersión**.- Que indican si los valores están agrupados o dispersos. Los más importantes son rango o recorrido, desviación media, la varianza y la desviación típica.

MEDIDAS DE CENTRALIZACIÓN

MEDIA:

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \dots\dots\dots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Cuando los datos vienen dados por una tabla de frecuencias:

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots\dots\dots + x_N f_N}{N} = \frac{\sum_{i=1}^{N} x_i f_i}{N}$$

Cuando los datos están agrupados en intervalos, el valor central de cada intervalo (marca de clase), es el que se asigna a todos los individuos que están en dicho intervalo.

Ejemplo 1:Se ha preguntado a un grupo de 70 alumnos sobre el número de zapatos que calzan, obteniendo los resultados de la siguiente tabla:

N. de calzado	N. de alumnos
35	4
36	15
37	17
38	20
40	10
42	4

En este caso, la media aritmética sería:

$$\overline{x} = \frac{35.4 + 36.15 + 37.17 + 38.20 + 40.10 + 42.4}{70} = \frac{2637}{70} = 37.67$$

Ejemplo 2. El *consumo de carburantes, en litros, de una flota de camiones a lo largo de un día está tabulado en la siguiente tabla de frecuencias:*

Consumo	Camiones
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Como la variable está agrupada en intervalos, tomamos la marca de clase. La media sería:

$$\overline{x} = \frac{5.8 + 15.20 + 25.10 + 35.14 + 45.21 + 55.16 + 65.9}{90} = \frac{3370}{90} = 37.44$$

MODA

Es el valor de la distribución que se repite con mayor frecuencia. Puede no existir o puede no ser única. Las distribuciones que contienen una sola moda se llaman unimodales y las que contienen dos, bimodales. En general, cuando contiene varias modas se denomina multimodal.

- En una representación gráfica, la moda será el rectángulo más alto, en el caso del histograma, y el pico más alto, en el caso del polígono.
  Ejemplo:
  En la distribución de cifras: 2, 3, 3, 3, 5, 5.................la moda es 3
  En la distribución de cifras: 2, 2, 4, 5, 5, 6................ las modas son 2 y 5.
- En el caso de los datos agrupados en intervalos, la moda es aproximadamente el punto medio de la clase que contiene la mayor frecuencia de casos (a la que se le llamaría clase modal)
  Ejemplo:
  De 1 a 3.............6
  De 4 a 6.............15
  De 7 a 8.............10
  De 9 a 11...........6

En este ejemplo, la clase modal es 4-6 y la moda valdrá 5.
Pero si queremos calcular más exactamente la moda (y no de forma aproximada),se busca el intervalo de mayor frecuencia (intervalo o clase modal) y se aplica la fórmula:

$$\boxed{Mo=L_{i-1}+C.\frac{f_i - f_{i-1}}{(f_i - f_{i-1})+(f_i - f_{i+1})}}$$

Donde:

L_{i-1} es el límite inferior del intervalo modal
C es la amplitud del intervalo
f_i es la frecuencia del intervalo modal
f_{i-1} es la frecuencia del intervalo anterior al modal
f_{i+1} es la frecuencia del intervalo posterior al modal

En el ejemplo puesto, sería el intervalo (4,6], y aplicando la fórmula:

$$Mo=4+3.\frac{15-6}{(15-6)+(15-10)}=5.93$$

Otro ejemplo:
El consumo de carburantes, en litros, de una flota de camiones a lo largo de un día está tabulado en la siguiente tabla de frecuencias:

Consumo	Camiones
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Solución:
El intervalo modal sería el (40,50]

$$Mo=40+\frac{21-14}{(21-14)+(21-16)}=45.83$$

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, en la distribución:

6,7,7,7,8,9,10,12,15

M_e=8

Si el número de individuos fuera par, la mediana sería el valor medio de los dos centrales.
Por ejemplo, en la distribución:

6,7,7,7,8,9,10,12,15,16

M_e=8.5

Si los datos están agrupados en intervalos, suponemos que los datos de cada intervalo se reparten uniformemente en él, hemos de buscar el intervalo central (en el que se encuentre el o los valores centrales) y aplicar la fórmula:

$$\boxed{Me = L_{i-1} + C.\frac{\frac{N}{2} - F_{i-1}}{f_i}}$$

Donde L_{i-1} es el límite inferior del intervalo
N es el número total de casos o datos
F_{i-1} es la frecuencia acumulada del intervalo anterior
f_i es la frecuencia absoluta del intervalo
C es la amplitud del intervalo

Ejemplo:

El consumo de carburantes, en litros, de una flota de camiones a lo largo de un día está tabulado en la siguiente tabla de frecuencias:

Consumo	Camiones
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Calcular la mediana.

Solución:

Hemos de buscar el intervalo en el que estén los elementos centrales. Como hay 90 elementos, el intervalo es (40,50]. Aplicamos la fórmula:

$$Me = 40 + 10.\frac{\frac{90}{2} - 44}{21} = 40.48$$

MEDIDAS DE POSICIÓN

CENTILES O PERCENTILES

- Mediana:

Si los individuos de una población están colocados en orden creciente según la variable que estudiamos, el que ocupa e valor central se llama individuo mediano, y su valor la "mediana".

La mediana, M_e , está situada de modo que antes de ella está el 50% de la población y, detrás, el otro 50%.

Por ejemplo, en la distribución:

6,7,7,7,8,9,10,12,15

M_e=8

Si el número de individuos fuera par, la mediana sería el valor medio de los dos centrales.

Por ejemplo, en la distribución:

6,7,7,7,8,9,10,12,15,16

$M_e=8.5$

- Cuartiles:

  Si en vez 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 la 75%.

  Cuartil superior Q_3 es un valor de la variable que deja por debajo de él al 75% de la población, y por encima la 25%.

  Q_2 sería la mediana.

  Por ejemplo, en la distribución:

$$\underbrace{1,\ 2,\ 2}_{25\%}\ ,\ \underbrace{3,\ 4,\ 5}_{25\%}\ ,\ \underbrace{5,\ 5,\ 6}_{25\%}\ ,\ \underbrace{8,\ 9,\ 10}_{25\%}$$

$$\blacktriangle\ Q_1 \qquad \blacktriangle\ M_e \qquad \blacktriangle\ Q_3$$

$$Q_1 = 2.5; \quad M_e = 5; \quad Q_3 = 7$$

- 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 $M_e = p_{50}$

A la mediana, cuartiles y centiles, se les llama medidas de posición.

Veamos unos ejemplos de estas medidas de posición

Ejemplo 1: Calcular M_e, 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

Solución: Hay 17 individuos;

17/2 = 8.5..................la mediana es el valor del individuo 9º..........$M_e=5$

17/4 = 4.25............................5º lugar..$Q_1=4$

17. 3/4 = 12.75.....................13º lugar..$Q_3=7$

17/100 . 10 = 1.7....................2º lugar...$P_{10}=1$

17/100 . 80 = 13.6................14º lugar...$P_{80}=7$

Ejemplo 2: En la siguiente distribución de número de hijos de 110 parejas , halla M_e, Q_1, Q_3, P_{20} y P_{99}

Nº hijos (x_i)	0	1	2	3	4	5	6
f_i	4	18	41	32	11	3	1

Para calcular la mediana, cuartiles y Percentiles en distribuciones dadas por tablas de frecuencias , necesitamos las frecuencias acumuladas y los %

x_i	f_i	F_i	En %
0	4	4	3.6
1	18	22	20
2	41	63	57.3
3	32	95	86.4

4	11	106	96.4
5	3	109	99.1
6	1	110	100

$M_e = P_{50} = 2$ 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.

Ejemplo 3:
En la fabricación de cierto tipo de bombillas, se han detectado algunas defectuosas. Se han estudiado 200 cajas de 100 bombillas cada una, obteniéndose la siguiente tabla:

Defectuosas	Nº de cajas
1	5
2	15
3	38
4	42
5	49
6	32
7	17
8	2

Calcula la mediana, el cuartil superior y el percentil 20.

Solución: Formemos la tabla de frecuencias acumuladas:

X_i	f_i	F_i	%
1	5	5	2.5
2	15	20	10
3	38	58	29
4	42	100	50
5	49	149	74.5
6	32	181	90.5
7	17	198	99
8	2	200	100

Mediana: Se han ordenado las cajas según el nº de bombillas defectuosas, de menor a mayor. La mediana será la caja que ocupe el lugar central. Como el nº de cajas es par (200), la mediana es el valor medio entre los dos centrales.
La caja nº 100 tiene 4 bombillas defectuosas y la nº 101 tiene 5 bombillas defectuosas. Por tanto, $M_e = (4+5)/2 = 4.5$
El cuartil superior: corresponde al 75% del total : 0.75.200=150. La caja que ocupa el lugar nº 150 tiene 6 bombillas defectuosas. Por tanto, $Q_3 = 6$. El 25% de las cajas tiene 6 o más bombillas defectuosas.
El percentil 20: corresponde al 20% del total: 0.20 . 200 = 40. La caja que ocupa el lugar 40 tiene 3 bombillas defectuosas. Por tanto, $P_{20} = 3$. El 20% de las cajas tiene 3 o menos bombillas defectuosas.

- En caso de una variable agrupada, las fórmulas para hallar centiles, deciles y cuartiles son:

$$C_h = L_{i-1} + \frac{h.\frac{n}{100} - N_{i-1}}{n_i}.c_i$$

donde cada elemento tiene el mismo significado que en el cálculo de la mediana.

Para hallar los cuarteles en este caso, nada más que hay que tener en cuenta que:
$Q_1=C_{25}$ $Q_2=C_{50}=M_e$ $Q_3=C_{75}$

Para hallar los deciles, tendremos en cuenta que:
$D_1=C_{10}$ $D_2=C_{20}$ $D_3=C_{30}$ ………… $D_9=C_{90}$

Veamos algunos ejemplos:

Ejemplo:

El consumo de carburantes, en litros, de una flota de camiones a lo largo de un día está tabulado en la siguiente tabla de frecuencias:

Consumo	Camiones
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Calcular el cuartil superior Q_3, el centil 45, C_{45} y el decil tercero, D_3

- Busquemos el intervalo donde se encuentra el Q_3, que será en el que se encuentre el elemento que deja a su izquierda un 75% de la población (el 75% de 90 es 67,5).

Haciendo la tabla de las frecuencias absolutas acumuladas:

Consumo	Frec. Abs.	F. Ab. acum
(0,10]	8	8
(10,20]	12	20
(20,30]	10	30
(30,40]	14	44
(40,50]	21	65
(50,60]	16	81
(60,70]	9	90

Observamos que el intervalo que buscamos es el (50,60]. Aplicando la fórmula:

$$Q_3=50+\frac{\frac{3.90}{4} - 65}{16}\cdot 10 = 50 + \frac{25}{16} = 50 + 1,5625 = 51,5625$$

- Busquemos ahora el intervalo donde queda el elemento que deja a su izquierda al 45% de la población: (45% de 90 es 40,5). Observando la columna de las frecuencias absolutas acumuladas, el intervalo (30,40] contiene el C_{45}. Aplicando la fórmula:

$$C_{45}=30+\frac{45.\frac{90}{100}-30}{14}\cdot 10=30+7{,}5=37{,}5$$

- Calculemos el $D_3=P_{30}$

Busquemos el intervalo donde se encuentra el elemento que deja a su izquierda el 30% de la población (30% de 90 es 27). Observando la tabla de frecuencias, el intervalo (20,30] contiene al individuo que ocupa el lugar 27, y aplicando la fórmula de los percentiles:

$$D_3=P_{30}=20+\frac{30.\frac{90}{100}-20}{10}=27$$

MEDIDAS DE DISPERSIÓN

RANGO O RECORRIDO

Es la diferencia entre el valor mayor y el menor si la variable es no agrupada. Si la variable es agrupada, se calcula la diferencia entre el límite superior del último intervalo y el límite inferior del primer intervalo.
El valor del recorrido sólo tiene en cuenta los valores extremos; en su valor no influyen los demás elementos de la distribución.
Cuanto menor es el rango o recorrido de una distribución, mayor es el grado de representatividad de los valores centrales
Ejemplo: Mercedes y Paco miden 169 y 171 respectivamente. Ana y Luís es otra pareja que miden 145 y 195 respectivamente.

Ambas distribuciones tienen la misma media: 170, pero evidentemente nadie los confundirían por la calle.
El rango de la pareja Mercedes y Paco: 171-169=2
El rango de la pareja Ana y Luís:195-145=50
Diremos por tanto que la 2ª pareja está más dispersa que la 1ª

Ejemplo 1:
Se ha preguntado a un grupo de 70 alumnos sobre el número de zapatos que calzan, obteniendo los resultados de la siguiente tabla:

N. de calzado	N. de alumnos
35	4
36	15
37	17
38	20
40	10
42	4

El rango o recorrido será 42-35 = 7
Ejemplo 2:
El consumo de carburantes, en litros, de una flota de camiones a lo largo de un día está tabulado en la siguiente tabla de frecuencias:

Consumo	Camiones
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Al ser agrupada, el rango o recorrido es 70-0 = 70

DESVIACIÓN MEDIA

Es la media de las desviaciones de los valores de la variable respecto a la media de la distribución (siendo la desviación respecto de la media: $|x_i - \overline{x}|$):

$$DM=\frac{|x_1-\overline{x}|.f_1+|x_2-\overline{x}|.f_2+..........+|x_k-\overline{x}|.f_k}{N}=\frac{\sum_{i=1}^{k}|x_i-\overline{x}|.f_i}{N}$$

Es una medida muy poco utilizada por la complicación de su cálculo.

Si la DM es muy pequeña, indica que hay una gran concentración de valores en torno a la media.

Si la variable está agrupada en intervalos, tomamos x_i la marca de la clase.

VARIANZA

Es la media aritmética de los cuadrados de las desviaciones respecto de la media. Se representa por σ^2, y viene dada por la expresión:

$$\sigma^2=\frac{(x_1-\overline{x})^2.f_1+(x_2-\overline{x})^2.f_2+\cdots\cdots+(x_n-\overline{x})^2.f_n}{f_1+f_2+...........+f_n}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2.f_i}{\sum_{i=1}^{n}f_i}=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2.f_i}{N}$$

(vamos a obtener una segunda expresión para σ^2 que vamos a utilizar con frecuencia)

$$\frac{(x_1-\overline{x})^2f_1+(x_2-\overline{x})^2f_2+.......+(x_n-\overline{x})^2f_n}{f_1+f_2+.....+f_n}=\frac{x_1^2.f_1+x_2^2.f_2+....+x_n^2.f_n}{f_1+f_2+.....+f_n}+\frac{\overline{x}^2f_1+\overline{x}^2f_2+...+\overline{x}^2f_n}{f_1+f_2+...+f_n}-$$

$$-\frac{2x_1\overline{x}f_1+2x_2\overline{x}f_2+....+2x_n\overline{x}f_n}{f_1+f_2+...+f_n}=\frac{\sum x_i^2.f_i}{\sum f_i}+\overline{x}^2-2\overline{x}\frac{\sum x_if_i}{\sum f_i}=\frac{\sum x_i^2.f_i}{\sum f_i}+\overline{x}^2-2\overline{x}^2=\frac{\sum x_i^2.f_i}{\sum f_i}-\overline{x}^2$$

DESVIACIÓN TÍPICA

Es la raíz cuadrada de la varianza. Se representa por S

$$S = \sqrt{\frac{(x_1 - \bar{x})^2.f_1 + (x_2 - \bar{x})^2.f_2 + \cdots + (x_n - \bar{x})^2.f_n}{f_1 + f_2 + \ldots\ldots\ldots + f_n}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2.f_i}{\sum_{i=1}^{n} f_i}} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2.f_i}{N}}$$

Es la unidad de dispersión más usada. Es siempre positiva. Se calcula directamente en las calculadoras científicas.

Ejemplo 1:

Se ha preguntado a un grupo de 70 alumnos sobre el número de zapatos que calzan, obteniendo los resultados de la siguiente tabla:

N. de calzado	N. de alumnos
35	4
36	15
37	17
38	20
40	10
42	4

Haced un cuadro donde aparezca la desviación media, la desviación típica y la varianza.

x_i	f_i	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert.f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2.f_i$
35	4	|35-37.67|=2.67	10.68	7.1298	28.5192
36	15	|36-37.67|=1.67	25.05	2.7889	41.8335
37	17	|37-37.67|=0.67	11.39	0.4489	7.6313
38	20	|38-37.67|=0.33	6.6	0.1089	2.178
40	10	|40-37.67|=2.33	23.3	5.4289	54.289
42	4	|42-37.67|=4.33	17.32	18.7489	74.9956
	70		**94.34**	**34.6543**	**209.4466**

Por tanto:

La desviación media DM = 94.34/70 = 1.348

La varianza $S^2 = 209.4466/70 = 2.992$
La desviación típica $S = 1.72973$

Ejemplo 2:
El consumo de carburantes, en litros, de una flota de camiones a lo largo de un día está tabulado en la siguiente tabla de frecuencias:

Consumo	Camiones
(0,10]	8
(10,20]	12
(20,30]	10
(30,40]	14
(40,50]	21
(50,60]	16
(60,70]	9

Haced un cuadro donde aparezca la desviación media, la desviación típica y la varianza.
Calculamos $\bar{x} = 37.44$

Consumo	x_i	Camiones=f_i	$\|x_i-\bar{x}\|$	$\|x_i-\bar{x}\|.f_i$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^2.f_i$
(0,10]	5	8	32.44	259.52	1052.3536	8418.848
(10,20]	15	12	22.44	269.28	503.5536	6042.6432
(20,30]	25	10	12.44	124.4	154.7536	1547.536
(30,40]	35	14	2.44	34.16	5.9536	83.3504
(40,50]	45	21	7.56	158.78	57.1536	1200.2256
(50,60]	55	16	17.56	280.96	308.3536	4933.6576
(60,70]	65	9	27.56	248.04	759.5536	6835.9824
		90		**1375.12**		**29062.241**

$DM = 1375.12 / 90 = 15.279$
$S^2 = 29062.241 / 90 = 322.9138$
$S = 17.9698$

Ejemplo 3:
Se ha anotado el peso de 88 personas, obteniéndose los siguientes resultados:

Peso (Kg)	[38,44)	[44,50)	[50,56)	[56,62)	[62,68)	[68,74)	[74,80)
Nº personas	7	8	15	25	18	9	6

Calcula el rango, la desviación media, la varianza y la desviación típica.

Calculemos primeramente la media aritmética:
$= 5204/88 = 59{,}14$

Datos			Media	Desviación	Desviación media		Varianza	
Clases	Marcas de clase x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$	$f_i \lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$	$f_i (x_i - \bar{x})^2$
[38-44)	41	7	287	−18,14	18,14	126,98	329,05	2 303,35
[44-50)	47	8	376	−12,14	12,14	97,12	147,38	1 179,04
[50-56)	53	15	795	−6,14	6,14	92,1	37,70	565,5
[56-62)	59	25	1 475	−0,14	0,14	3,50	0,0196	0,49
[62-68)	65	18	1 170	5,86	5,86	105,48	34,34	618,12
[68-74)	71	9	639	11,86	11,86	106,74	140,65	1 265,85
[74-80)	77	6	462	17,86	17,86	107,16	318,98	1 913,88
		88	5 204			639,08		7 846,23

Rango: rango=80 - 38 = 42 Kg
Desviación media: D = 639,08/88 = 7,26 Kg
Varianza: s^2= 7846,23/88 = 89,16 Kg^2
Desviación típica: s = 9,44 Kg

EJERCICIOS RESUELTOS

1) Un inspector de autobuses toma nota de los minutos de retraso con que llegan los autobuses a una parada. Su trabajo queda reflejado en el siguiente diagrama de barras:

Halla la varianza y el rango.

Resolución: Formemos la siguiente tabla:

x_i	f_i	$x_i f_i$	$f_i (x_i - \bar{x})^2$
0	3	0	337,08
5	12	60	376,32
10	14	140	5,04
15	7	105	135,52
20	3	60	265,08
25	2	50	414,72
30	1	30	376,36
	42	445	1 910,12

Media, $\bar{x} = 445/42 = 10.6$ minutos

Rango = 30 – 0 = 30 minutos

$S^2 = 1910.12 / 42 = 45.47 \text{ min}^2$

2) Elaborad una tabla de frecuencias con las estaturas de 40 adolescentes:

168 160 167 175 175 167 168 158 149 160
178 166 158 163 171 162 165 163 156 174
160 165 154 163 165 161 162 166 163 159
170 165 150 167 164 165 173 164 169 170

Resolución: El nº de valores distintos que hay es grande (mayor que 20), por eso lo adecuado es clasificarlos en intervalos. Para ello procedemos así:

- localizamos los valores extremos: el menor 149 y el mayor 178. Hallamos su diferencia: 178-149=29 (este es el valor del recorrido).
- Puesto que el nº de datos es pequeño (solo 40), decidimos que el nº de intevalos sea pequeño (por ej, 6). Buscamos un nº mayor que el recorrido y que sea múltiplo de 6, por ej. 30 (el recorrido era 29). De este modo, cada uno de los seis intevalos tendrá una longitud igual a 5

- Formamos los intervalos comenzando por un nº algo menor que el 149 y de modo que los seis intervalos abarquen a la totalidad de los datos.
- Repartimos los cuarenta datos en los seis intervalos. (Es conveniente tomar los intervalos con extremos no enteros para que no haya duda de si un valor pertenece a un intervalo o al siguiente.

Intervalos	frecuencias
(148.5, 153.5]	2
(153.5, 158.5]	4
(158.5, 163.5]	11
(163.5, 168.5]	14
(168.5, 173.5]	5
(173.5, 178.5]	4

3) Calcula la media y la desviación típica del ejercicio anterior.
Resolución: construyamos la siguiente tabla

Intervalos	x_i	f_i	f_ix_i	$f_ix_i^2$
(148.5, 153.5]	151	2	302	45602
(153.5, 158.5]	156	4	624	97344
(158.5, 163.5]	161	11	1771	285131
(163.5, 168.5]	166	14	2324	385784
(168.5, 173.5]	171	5	855	146205
(173.5, 178.5]	176	4	704	123904
		40	6580	1083970

Media: $\bar{x} = \dfrac{\sum f_i.x_i}{\sum f_i} = \dfrac{6580}{40} = 164.5\text{cm}$

Varianza: $\sigma^2 = \dfrac{\sum f_i.x_i^2}{\sum f_i} - \bar{x}^2 = \dfrac{1083970}{40} - 164.5^2 = 39$

Desviación Típica: $\sigma = \sqrt{39} = 6.24\text{cm}$

Vemos en este ejemplo la ventaja de la segunda expresión de σ^2 para hallar su valor numérico a partir de una tabla de frecuencias.

4) . Construir la tabla estadística de las edades de las personas que acuden a un logopeda a lo largo de un mes, sabiendo que son:

3,	2,	11,	13,	4,	3,	2,	4,	5,	6,	7,	3,
4,	5,	3,	2,	5,	6,	27,	15,	4,	21,	12,	4,
3,	6,	29,	13,	6,	17,	6,	13,	6,	5,	12,	26

Como hay 36 datos, el número de clases que debemos formar puede ser aproximadamente la raíz cuadrada de 36, es decir 6 clases. Si el intervalo los extendemos de 0 hasta 30, al dividir por 6 se tiene que la amplitud de cada clase debe ser 5.

Este sería un ejemplo de tabla estadística para una variable estadística continua (la edad de una persona no tiene por qué ser un número entero).

Clases	Marcas de clase	f_i	h_i	F_i	H_i
$[0,5)$	2'5	13	$\frac{13}{36}$	13	$\frac{13}{36}$
$[5,10)$	7'5	11	$\frac{11}{36}$	24	$\frac{24}{36}$
$[10,15)$	12'5	6	$\frac{6}{36}$	30	$\frac{30}{36}$
$[15,20)$	17'5	2	$\frac{2}{36}$	32	$\frac{32}{36}$
$[20,25)$	22'5	1	$\frac{1}{36}$	33	$\frac{33}{36}$
$[25,30)$	27'5	3	$\frac{3}{36}$	36	1
		36	**1**		

5- Las calificaciones en la asignatura de historia de los/as 40 alumnos/as de una clase viene dada por la siguiente tabla:

Calificaciones	1	2	3	4	5	6	7	8	9
Nº de alumnos/as	2	2	4	5	8	9	3	4	3

Calcula la mediana.

x_i	f_i	F_i
1	2	2
2	2	4
3	4	8
4	5	$13 < 20$
5	8	$21 > 20$
6	9	30
7	3	33
8	4	37
9	3	40
	40	

La mediana es $M_e = 5$, dado que es el primer valor de la variable cuya frecuencia absoluta acumulada, 21, excede la mitad del número de datos, 20.

6- Consideremos la siguiente tabla de frecuencias:

x_i	3	6	7	8	9
f_i	15	20	15	40	10

Calcula la mediana.

x_i	f_i	F_i

3	15	15
6	20	35
7	15	$50 \le 50$
8	40	$90 > 50$
9	10	100
	100	

Como 50 coincide con la frecuencia acumulada del valor 7, la mediana vendrá dada por la semisuma de 7 y el valor siguiente, 8. Por tanto $M_e = \frac{7+8}{2} = 7'5$.

7- Se ha aplicado un test sobre satisfacción en el trabajo a 88 empleados/as de una fábrica, obteniéndose los siguientes resultados:

Puntuaciones	Nº de trabajadores/as
$[38, 44)$	7
$[44, 50)$	8
$[50, 56)$	15
$[56, 62)$	25
$[62, 68)$	18
$[68, 74)$	9
$[74, 80)$	6

Calcula la mediana.

x_i	f_i	F_i
$[38, 44)$	7	7
$[44, 50)$	8	15
$[50, 56)$	15	$30 < 44$
$[56, 62)$	25	$55 > 44$
$[62, 68)$	18	73
$[68, 74)$	9	82
$[74, 80)$	6	88
	88	

La clase mediana es el intervalo $[56, 62)$.

$$M_e = 56 + 6 \cdot \frac{\frac{88}{2} - 30}{25} = 59'36$$

<u>Observaciones</u>

1. La mediana es particularmente útil en los siguientes casos:

a) Cuando entre los datos existe alguno muy extremo que afecta a la media.

b) Cuando los datos están agrupados en clases y alguna de ellas es abierta.

2. Como consecuencia de definición de mediana, se tiene que el 50% de los datos son menores o iguales que ella y el 50% restante son mayores o iguales.

3. La mediana depende del orden de los datos y no de su valor.

8- Las calificaciones en la asignatura de historia del arte de los 40 alumnos/as de una clase viene dada por la siguiente tabla:

Calificaciones	1	2	3	4	5	6	7	8	9
Nº de alumnos/as	2	2	4	5	8	9	3	4	3

Rango: 9-1=8

Varianza y desviación típica:

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
1	2	2	2
2	2	4	8
3	4	12	36
4	5	20	80
5	8	40	200
6	9	54	324
7	3	21	147
8	4	32	256
9	3	27	243
	40	**212**	**1296**

Media: $\bar{x} = 5'3$

Varianza: $s^2 = \frac{1296}{40} - (5'3)^2 = 4'31$

Desviación típica: $s = \sqrt{4'31} = 2'08$