

Cuaderno de actividades 4º ESO

Estadística Unidimensional ¹

Conceptos básicos.

Cualquier elemento o ente que sea portador de información sobre alguna propiedad en la cual se está interesado se denomina **individuo**.

El conjunto de todos los individuos en los que se desea estudiar alguna propiedad o característica se llama **población**.

Todo subconjunto finito de la población sobre el que se realice el estudio de la propiedad deseada, es una **muestra**. Al número de individuos de este subconjunto se le llama **tamaño** de la muestra.

Ejemplo 1. Para estudiar la evolución del cáncer de mama en la población femenina de un país, se puede considerar que **individuo** es cada una de las mujeres residentes en el mismo, **población** es el conjunto de todas ellas y una **muestra** se obtiene al observar el 1% del censo.

Con mucha frecuencia se consideran como población y muestra, *no* los conjuntos de individuos, *sino las medidas de la característica asociadas a esos individuos*.

Ejemplo 2. En un banco de sangre se experimenta un nuevo sistema para aumentar el período de conservación de la misma. En este caso cada bolsa de sangre es un individuo; la población es el conjunto de todas las bolsas del banco y una muestra se obtiene tomando un cierto número de bolsas para su análisis.

Obsérvese que el concepto de individuo no va asociado necesariamente con el de persona, sino que puede ser algo de naturaleza más abstracta.

Clasificación de los datos.

Cuando los datos, es decir los resultados de las observaciones, no son magnitudes *medibles numéricamente*, sino cualidades o atributos, se dice que se trata de datos **cualitativos**, mientras que en caso contrario se habla de datos **cuantitativos**.

Ejemplo 3. Se observan las causas de muerte de 16 individuos de una cierta población, agrupándolas en las cuatro siguientes: enfermedades cardiovasculares (EC), cáncer (C), accidentes (A) y otras causas (O), habiéndose obtenido los siguientes datos:

EC, EC, A, C, O, A, EC, A, O, C, EC, C, O, C y EC.

Como los resultados no son medibles numéricamente, los datos son cualitativos.

Ejemplo 4. Las notas obtenidas en Matemáticas en una clase de COU han sido:

2, 7, 4, 6, 5, 0, 3, 9, 8, 4, 3, 6, 5 y 8.5.

Se trata de datos cuantitativos.

A su vez los datos cuantitativos se denominan **continuos** si los resultados pueden tomar cualquier valor real dentro de un cierto intervalo, o **discretos**, si sólo pueden tomar ciertos valores particulares.

Ejemplo 5. Del estudio de la estatura de un cierto núcleo de población se han obtenido los siguientes datos:

1.62, 1.78, 1.75, 1.58, 1.83, 1.68 y 1.81 metros.

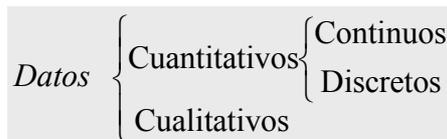
Son datos continuos, pues los individuos de una población pueden tener como estatura cualquier número real en un cierto intervalo.

Ejemplo 6. Del alumbramiento de un conjunto de ratas se ha observado el número de crías, obteniéndose los siguientes valores numéricos:

5, 3, 1, 5, 3, 6, 4, 2, 5, 6, 3, 6, 5, 2, 6, 7 y 3.

Por no ser posibles números no naturales, es evidente que se trata de datos cuantitativos discretos.

Es decir los datos se clasifican:



¹ Se incluyen los contenidos, ya que no se estudiaron el curso pasado en muchos grupos de 3º ESO s.

Los datos pueden provenir del estudio de un sólo carácter o propiedad (caso **unidimensional**) o de varios simultáneamente (caso **multidimensional**).

Variables estadísticas. Frecuencias.

Los caracteres estadísticos de una población son las propiedades o cualidades de los individuos que nos interesa estudiar. Un carácter estadístico divide a la población en clases. A cada una de estas clases se la denomina **modalidad**.

Cuando el carácter es **cuantitativo** sus diversas modalidades son **medibles**, es decir se les puede asignar un **número**.

Definición 1. Se llama **variable estadística** a la aplicación que a cada modalidad le hace corresponder ese número, es decir su medida.

Ejemplo 7. En el ejemplo 6 la variable estadística toma los valores: 1, 2, 3, 4, 5, 6 y 7.

La variable estadística será **discreta** cuando sólo pueda tomar un nº finito de valores y **continua** cuando pueda tomar todos los valores de un cierto intervalo.

Ejemplo 8. La variable estadística del ejemplo 5 es continua y discreta la del ejemplo 6.

Definición 2. Se llama **frecuencia absoluta** al número de individuos que toman un determinado valor de una variable estadística (o una modalidad de un atributo).

Para variables estadísticas (es decir, datos cuantitativos) puede definir:

Definición 3. Se llama frecuencia absoluta **acumulada** de un valor a la suma de las frecuencias absolutas de todos los valores menores o iguales que él.

Ejemplo 9. En el ejemplo 6 la frecuencia absoluta del 5 (tener 5 crías) es 4. La frecuencia absoluta acumulada del 2 es 3.

Definición 4. Se llama **frecuencia relativa** a la razón entre la frecuencia absoluta y el número total de datos o tamaño de la población.

Definición 5. Se llama frecuencia relativa **acumulada** de un valor de una variable estadística a la suma de las frecuencias relativas de todos los valores menores o iguales que él.

Ejemplo 10. La frecuencia relativa del 5 es $4/17$ y la relativa acumulada del 2 es $3/17$.

Representación de datos: Tablas.

Las dos formas más comunes de representar los datos son las tablas y los gráficos.

Tablas estadísticas

Las tablas estadísticas consisten en masas estructuradas de datos.

Para la construcción de tablas de **datos cuantitativos** pueden tratarse éstos individualmente o agrupándolos en clases

Tratamiento individual

Para variable discreta, o que siendo continua tengamos pocos datos. Lo vemos con un ejemplo.

Ejemplo 11. Las notas de los 20 alumnos de una clase son:

4, 3, 3, 5, 6, 7, 9, 0, 5, 4, 9, 10, 2, 7, 2, 2, 5, 6, 5, 0

Vamos a calcular una tabla:

Variable estadística x_i	Frecuencias absolutas		Frecuencias relativas	
	puntuales n_i	acumuladas N_i	puntuales f_i	acumuladas F_i
0	2	2	1/10	1/10
2	3	5	3/20	5/20=1/4
3	2	7	1/10	7/20
4	2	9	1/10	9/20
5	5	14	1/4	14/20=7/10
7	3	17	3/20	17/20
9	3	20	3/20	20/20=1

Ejercicio 1. En un Instituto hay matriculados 2200 alumnos que se distribuyen por edades en la forma siguiente: 215 de 14 años, 437 de 15, 421 de 16, 396 de 17, 512 de 18, 124 de 19 y 95 de 20. Formar la tabla de distribución y de frecuencias, que incluya frecuencias acumuladas.

Tratamiento por clases

Cuando en la población o muestra que estudiamos existen muchos valores diferentes, es conveniente, aún a costa de perder algo de información, dividir el intervalo de variación en una serie de subintervalos que cubran el total; a cada uno de ellos se le llama una **clase**, a sus extremos, *extremos de clase*, al punto medio de cada clase, **marca de clase** y a la diferencia entre sus extremos, **amplitud** de la clase.

Lo vemos con un ejemplo.

Ejemplo 12. Se ha pasado un test de 79 preguntas a 600 personas. El número de respuestas correctas se refleja en la siguiente tabla:

intervalos	m_i	f. abs. puntual	f. abs. acumulada	f. rel. puntual	f. rel. acumulado
[0, 10)	5	40	40	1/15	1/15
[10, 20)	15	60	100	1/10	1/6
[20, 30)	25	75	175	1/8	7/24
[30, 40)	35	90	265	3/20	53/120
[40, 50)	45	105	370	7/40	37/60
[50, 60)	55	85	455	17/120	91/120
[60, 70)	65	80	535	2/15	107/120
[70, 80)	75	65	600	13/120	1
		600		1	

Ejemplo 13. En una Caja de Reclutamiento se toma una muestra de tamaño 30 de los pesos de los mozos correspondientes a un cierto reemplazo, obteniéndose los siguientes datos medidos en kg:

71.9, 63.9, 62.3, 72.5, 78.0, 70.7, 71.4, 60.5, 60.9, 68.2, 88.5, 76.1, 82.1, 63.7, 79.8, 67.5, 50.1, 69.5, 66.1, 47.3, 72.1, 59.8, 93.7, 80.7, 61.2, 64.3, 53.7, 74.7, 96.3, 73.2.

Construir una tabla de frecuencias agrupando los datos en clases de la misma amplitud.

Solución

Si bien no es estrictamente necesario, en general, es conveniente ordenar los datos de menor a mayor. A continuación se presenta la misma muestra ordenada:

47.3, 50.1, 53.7, 59.8, 60.5, 60.9, 61.2, 62.3, 63.7, 63.9, 64.3, 66.1, 67.5, 68.2, 69.5, 70.7, 71.4, 71.9, 72.1, 72.5, 73.2, 74.7, 76.1, 78.0, 79.8, 80.7, 82.1, 88.5, 93.7, 96.3.

Como los valores extremos son 47.3 y 96.3 y el número de clases aconsejado para estos datos es **6** (aplicando la fórmula de Sturges), tomaremos 6 intervalos de amplitud 10, la tabla queda estructurada de la siguiente manera:

clases	Marcas de	frecuencias absolutas	Frecuencias relativas
--------	-----------	-----------------------	-----------------------

	clase	de clase	acumuladas	de clase	acumuladas
45 -55	50	3	3	0.1	0.1
55 -65	60	8	11	0.266	0.366
65 -75	70	11	22	0.366	0.733
75 -85	80	5	27	0.166	0.900
85 -95	90	2	29	0.066	0.966
95 -105	100	1	30	0.033	1
		30		0.997≈1	

Ejercicio 2. El número de personas que viven en cada uno de los portales de una gran barriada es:: 63, 58, 70, 47, 120, 76, 80, 59, 80, 70, 63, 77, 104, 97, 78, 90, 112, 88, 67, 58, 87, 94, 100, 74, 55, 80, 75, 49, 98, 67, 84, 73, 95, 121, 58, 71, 66, 87, 76, 56, 77, 82, 93, 102, 56, 46, 78, 67, 65, 95, 69, 90, 58, 76, 54, 76, 98, 49, 87, 69, 80, 64, 65, 56, 69, 68, 99, 106.

Construye una tabla de frecuencias².

Series cronológicas

Se llaman **series cronológicas** a unas tablas estadísticas que recogen observaciones hechas a lo largo del tiempo, normalmente a intervalos iguales. Es por tanto una serie estadística en que la variable independiente es el **tiempo**. Hay dos tipos: **flujo** cuando sus valores se dan en intervalos y son el resultado de acumular todas las observaciones realizadas durante ese período; y **nivel** si las observaciones son instantáneas.

Ejemplo 14. El número de médicos colegiados en España en el período de 1984 - 1992:

1984	1985	1986	1987	1988	1989	1990	1991	1992
99730	107503	119890	123543	129897	138967	147978	152943	156748

Ejercicio 3. La producción editorial española de libros de sociología y Estadística, en los años que se indica es:

Años	1991	1992	1993	1994	1995	1996	1997
nº	345	487	589	376	479	652	741

Hacer una tabla de frecuencias absolutas y relativas puntuales. Expresar la relativa en porcentajes.

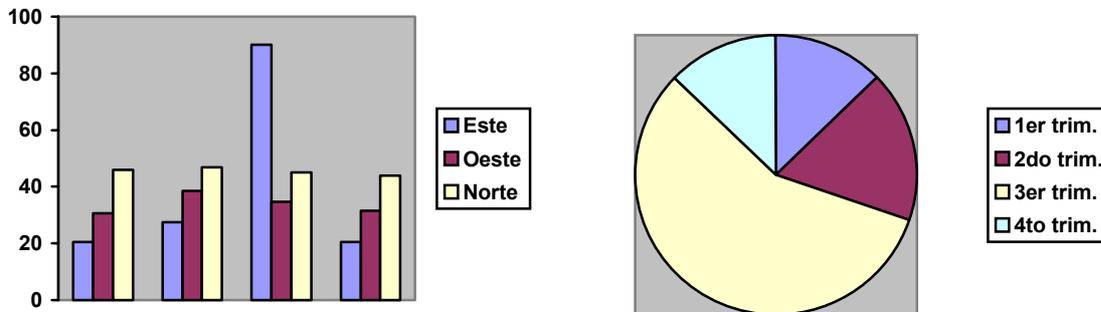
Representación de datos: Gráficos.

² Aunque la variable es discreta conviene agruparlos en clases ya que hay un número muy grande de datos.

Los gráficos no son más que traducciones a un dibujo del contenido de las tablas. La finalidad de los gráficos estadísticos es que la información esté al alcance de personas no expertas, que *entre por los ojos*. Los hay de muy diversos tipos pero todos son muy fáciles de interpretar.

Variables cualitativas

Los más usados son los diagramas de rectángulos (de barras) y los de sectores.



Ejercicio 4. El censo, en miles de cabezas, del ganado en el territorio español, en 1994 fue:

Ganado	Número de cabezas
Bovino	5300
Ovino	18047
Caprino	2601
Porcino	12308
Caballar	264
Mular	153
Asnar	164

Dibujar un diagrama de sectores y otro de rectángulos.

Variables cuantitativas.

Distinguiremos entre variable **discreta** o **continua**.

Tratamiento individual

Para el tratamiento individual los medios de representación más utilizados son el gráfico (o diagrama) de barras, el polígono de frecuencias y los gráficos acumulativos.

Diagrama de barras: Se asocia a una tabla de frecuencias ya sea absoluta o relativa.

Sobre un eje horizontal se representan los valores discretos que toman los datos y sobre cada uno de ellos se coloca una barra vertical (o un rectángulo) de longitud (altura) proporcional a la frecuencia.

Ejemplo 15. Vamos a hacer un diagrama de barras de frecuencias absolutas para el ejemplo 6.

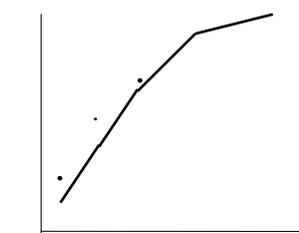
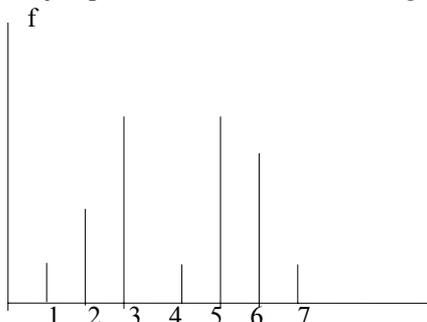
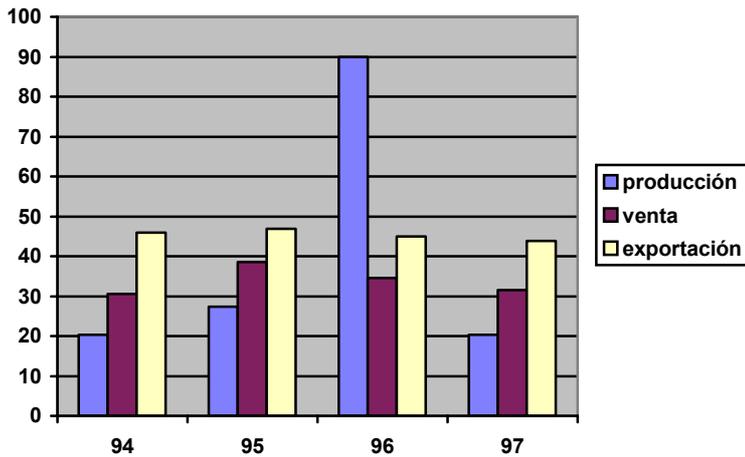


figura 1

En ocasiones se *superponen dos o más diagramas* para comparar datos:

Ejemplo 16: Producción y venta de automóviles en España:



Polígono de frecuencias: Como el anterior se asocia a una tabla de frecuencias.

Se representan en un sistema cartesiano los puntos aislados y luego se unen por medio de segmentos (poligonal). Se usa sobre todo para frecuencias acumuladas (**figura 1**). También para series cronológicas.

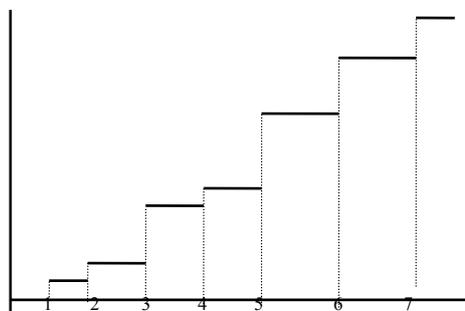
Ejercicio 5. La esperanza de vida al nacimiento ha evolucionado desde 1900, como se refleja en la tabla siguiente:

Años	1900	1910	1920	1930	1940	1950	1960	1970	1980
Varones	33,9	40,9	40,3	48,3	47,1	59,8	67,4	69,6	72,6
Mujeres	35,7	42,6	42,1	51,6	53,2	64,3	72,2	75,1	78,6

Dibujar los polígonos de frecuencias superpuestos para poder compararlos.

Gráficos acumulativos: Se construye a partir del mismo eje horizontal del gráfico de barras, llevando sobre cada valor discreto una vertical de longitud proporcional a la frecuencia acumulada, absoluta o relativa, de dicho valor. Se suele completar el gráfico dándole forma de una escalera de peldaños horizontales.

Ejemplo 16. Gráfico de barras acumulativo



Tratamiento por clases

Cuando las variables son continuas, o discretas agrupadas, los gráficos que más se utilizan son: el histograma de frecuencias y los polígonos de frecuencias (absolutas o relativas)

Histogramas de frecuencias. Sobre el eje de abscisas se marcan los extremos de las sucesivas clases y con base en cada clase se dibuja un rectángulo *de altura proporcional a la frecuencia* (absoluta o relativa) observada en dicha clase.

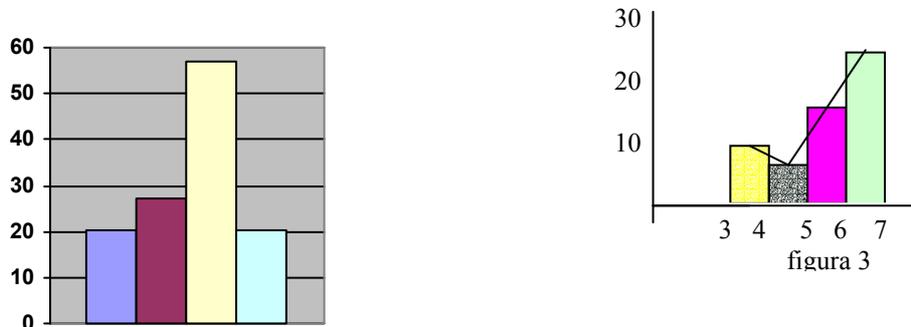


Figura 2

Ejercicio 6. En la siguiente tabla³ se presenta la distribución por edades del número de muertes registradas en España (datos hasta el 30-9-94) a causa del SIDA.

Edad en años	<3	3-9	10-12	13-14	15-19	20-24	25-29	30-34	35-39	40-49	50-59	60-69
Nº de muertes	411	171	35	31	247	2888	8576	7640	3292	2552	909	544

a) Construye la tabla de frecuencias relativas agrupando los datos en las siguientes categorías de edad: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59 y 60-69 años.

b) Representa gráficamente la información obtenida en el apartado a) mediante un **histograma**.

Polígono de frecuencias

Se asocia a cada clase un punto del plano cartesiano, de abscisa el valor de la marca de clase y de ordenada la frecuencia observada en dicha clase. Uniendo los puntos resulta una línea quebrada que se denomina polígono de frecuencias (**figura 3**)

Polígono de frecuencias acumuladas.

Partiendo del valor cero en el extremo izquierdo de la primera clase, el polígono acumulado va tomando en los sucesivos extremos derechos de las clases un valor igual a la frecuencia acumulada. Uniendo los puntos así obtenidos resulta el polígono acumulativo de frecuencias (**figura 4**).

³ Esta serie cronológica es un **flujo**.

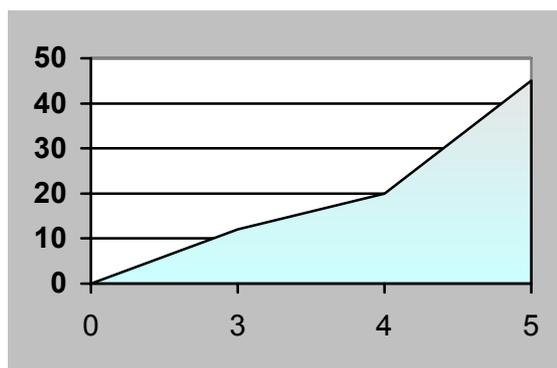


figura 4

Ejercicio 7. Los jugadores de un determinado equipo de baloncesto se clasifican, por altura, según la tabla siguiente:

Altura	1,70-175	1,75-1,80	1,80-185	185-190	1,90-1,95	1,95-2,00
Nº de jugadores	1	3	4	8	5	2

Dibujar el polígono de frecuencias absolutas acumulativo.

Parámetros estadísticos.

Las tablas estadísticas son una forma organizada de dar toda la información, todos los datos de que disponemos.

Con las gráficas estadísticas se pierde algo de información, pero el mensaje “entra por los ojos”, que es lo que se pretende.

En cualquiera de los dos casos, la cantidad de datos que se dan es excesiva para que sea operativo, por ejemplo para la comparación con otras distribuciones.

Por ello se definen los **parámetros estadísticos**, que nos van a servir para resumir en números aspectos relevantes de la distribución, que puedan dar una idea de la misma o permitir compararlas con otras.

Clases de parámetros estadísticos

♦ Medidas de centralización: media (ya conocida), moda (el valor que se presenta con más frecuencia) y mediana (el valor del individuo que ocuparía el lugar central si se colocaran ordenados de menor a mayor). Tienen como misión representar con un número a la serie estadística bajo el punto de vista de su posición.

♦ Medidas de dispersión: rango o recorrido, desviación media, varianza, desviación típica, coeficientes de Pearson... Sirven para medir el grado de alejamiento de los datos respecto de una medida central.

♦ Medidas de posición: cuartiles, deciles, centiles o percentiles. Señalan la situación de algunos valores importantes de la distribución.

En la ordenación que se hizo para la mediana se llaman cuartiles primero, segundo y tercero a los que superan exactamente al 25%, 50% y 75% de los valores. El segundo cuartil es la mediana. Para su obtención se usan los *diagramas de cajas*

Para el cálculo **práctico** de muchos parámetros estadísticos se utilizan tablas que facilitan dichos cálculos. Utilizaremos dos tipos que llamamos Tabla I y Tabla II. Las vemos, así como las **fórmulas** para hallar los parámetros estadísticos más usuales, en los ejemplos resueltos.

Ejemplo 17. Construir la **tabla I** con los datos del ejemplo 11

x_i	n_i	$x_i n_i$	$ x_i - \bar{x} $	$ x_i - \bar{x} n_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
0	2	0	4,65	9,30	21,62	43,24
2	3	6	2,65	7,95	7,02	21,06
3	2	6	1,65	3,30	2,72	5,44
4	2	8	0,65	1,30	0,42	0,84
5	5	25	0,35	1,75	0,12	0,60
7	3	21	2,35	7,05	5,52	16,56
9	3	27	4,35	13,05	18,92	56,76

20 93

Observación. Para esta tabla se necesitaba la media : $\bar{x} = \frac{\sum n_i x_i}{\sum n_i} = 93/20=4,65$

Ejemplo 18. Construir la tabla II con los datos del ejemplo 13.

clases	Marcas de clase x_i	frecuencia n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$
45 -55	50	3	150	2500	7500
55 -65	60	8	480	3600	28800
65 -75	70	11	770	4900	53900
75 -85	80	5	400	6400	32000
85 -95	90	2	180	8100	16200
95 -105	100	1	100	10000	10000

Ejemplo 19. a) Hallar la media y la varianza de la variable cuyos valores y frecuencias absolutas vienen dadas en la tabla adjunta

Valores de la variable	3	5	4	2	0	8	7
frecuencias	1	3	4	1	3	1	2

b) Representar gráficamente los datos en un diagrama de barras.

Solución a) la media es:

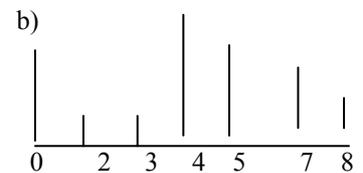
La varianza: $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i} = \frac{\sum x_i^2 n_i}{\sum n_i} - \bar{x}^2$, foma que se usa en la tabla II

x_i	n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$
0	3	0	0	0
2	1	2	4	4
3	1	3	9	9
4	4	16	16	64
5	3	15	25	75
7	2	14	49	98
8	1	8	64	64
	15	58		314

Se tiene :

$$\bar{x} = \frac{58}{15} = 3,87$$

$$\sigma^2 = \frac{314}{15} - (3,87)^2 = 5,96$$



Ejercicio 8. Los pacientes que acuden a una consulta médica se distribuyen, según la edad, en una tabla:

X(edad)	[0, 10)	[10, 20)	[20,30)	[30, 40)	[40, 50)	[50,60)
N (frecuencia)	7	10	30	18	12	3

Se pide:

- El histograma de frecuencias.
- La media, desviación típica, mediana y moda.
- Porcentaje de pacientes menores de 40 años que acuden a la consulta.

Ejemplo 20. Completar los datos que faltan en la siguiente tabla estadística, donde f, F y f_r representan, respectivamente, la frecuencia absoluta, acumulada y relativa:

x	f	F	f_r
1	4		0,08
2	4		
3		16	0,16
4	7		0,14
5	5	28	
6		38	
7	7	45	
8			

b) Calcula la media, mediana y moda de esta distribución

Solución

a) La frecuencia relativa de 1 es $0,08 = \frac{4}{N}$, de donde $N = 50$, lo que nos permite completar la tabla.

x	f	F	f_r
1	4	4	0,08
2	4	8	0,08
3	8	16	0,16
4	7	23	0,14
5	5	28	0,10
6	10	38	0,20
7	7	45	0,14
8	5	50	0,10

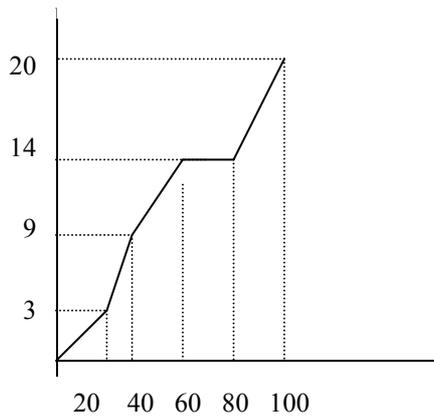
b) la media $\bar{x} = 4,76$; la mediana es 5 y la moda es 6.

Ejercicio 9. a) Hallar la media y la varianza de la variable cuyos valores y frecuencias absolutas vienen dadas en la tabla adjunta

Valores de la variable	3	5	4	2	0	8	7
frecuencias	1	3	4	1	3	1	2

b) Representar gráficamente los datos en un diagrama de barras.

Ejemplo 21. Se considera una distribución de datos agrupados en intervalos cuyo polígono de frecuencias acumuladas es el de la figura.



b) la media.

Solución

a)

x_i	n_i	N_i
20	3	3
40	6	9
60	5	14
80	0	14
100	6	20

b)
$$\bar{x} = \frac{20 \cdot 3 + 40 \cdot 6 + 60 \cdot 5 + 80 \cdot 0 + 100 \cdot 6}{20} = 60$$

Calcula:

a) Tabla de distribución de frecuencias acumuladas.

Ejercicio 10. En la fabricación de un cierto tipo de clavos, aparecen un cierto nº de ellos defectuosos. Se han estudiado 200 lotes de 500 clavos cada uno obteniendo:

Clavos defectuosos	1	2	3	4	5	6	7	8
nº de lotes	5	15	38	42	49	32	17	2

a) Calcular la mediana y la desviación típica.

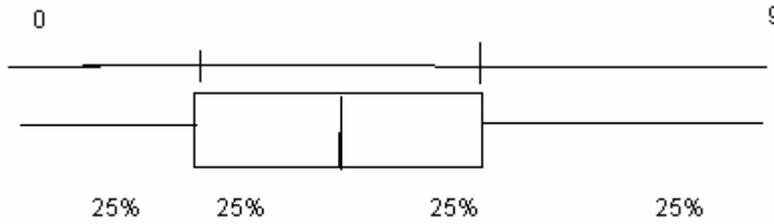
b) Dibuja un diagrama de barras.

Ejemplo 22. Representa mediante un diagrama de cajas las siguientes calificaciones de 20 alumnos.

0, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 5, 5, 5, 7, 7, 7, 8, 8, 9

Como es múltiplo de 4, $20:4 = 5$, Q_1 , M_e y Q_3 , serán los valores que hay entre el 5º y 6º, 10º y 11º, 15º y 16º, es decir

$Q_1 = 3,5$, $M_e = 5$ y $Q_3 = 7$



Ejercicio 11. Los pesos de un grupo de alumnos de 2º Bach son:

63, 58, 70, 57, 56, 76, 80, 59, 80, 70, 63, 77, 84, 77, 78, 90, 72, 88, 67, 58, 87, 94, 80, 74, 55, 80, 75, 59, 81, 67, 84, 73, 65, 71, 58, 71, 66, 87, 76, 56, 77, 82, 73, 67, 56, 46, 78, 67, 65, 65, 69, 80, 58, 76, 54, 76, 78, 49, 87, 69, 80, 64, 65, 56, 69, 68, 69, 64.

Representa la distribución mediante un diagrama de caja.

Ejemplo 23. En el estudio de un cierto fenómeno se obtiene la siguiente tabla:

x_i	7	10	12	16	19	20	21
n_i	6	7	16	17	22	19	17

Calcula los cuartiles Q_1 y Q_3 correspondientes..Haz un diagrama de caja.

Solución

x_i	n_i	N_i
7	6	6
10	7	13
12	16	29
16	17	46
19	22	68
20	19	87
21	17	104

Se tiene: $\sum n_i = 104$, y $\frac{104}{4} = 26$, que corresponde al dato 12; $3 \cdot 26 = 78$, correspondiente al dato 20. Luego:

$Q_1 = 12$,	$Q_3 = 20$
--------------	------------