

ESTADÍSTICA

1.- CONCEPTOS BÁSICOS

Estadística Es la ciencia que estudia conjuntos de datos obtenidos de la realidad. Estos datos son interpretados mediante tablas, gráficas, etc.

En estadística se utiliza la siguiente terminología:

Población : Es el conjunto formado por todos los elementos que queremos estudiar.

Por ejemplo, si vamos a estudiar el peso de los jóvenes de 15 años nacidos en España, la población sería precisamente el conjunto formado por dichos jóvenes

Muestra : Es una parte de la población que elegimos para estudiarla.

El nº de elementos de la muestra se llama tamaño de la muestra y cada elemento se llama individuo.

El proceso de elección de la muestra se llama muestreo.

Se tomará una muestra cuando la población sea muy numerosa.

La muestra debe ser representativa de la población.

En el ejemplo del peso de los jóvenes mencionado anteriormente, elegiríamos una muestra formada por sólo algunos jóvenes, por ejemplo, 1000 jóvenes (tamaño 1000).

Para que la muestra fuese representativa de la población habría que tomar hombres y mujeres y de todas las comunidades autónomas.

Si, por ejemplo, queremos estudiar el peso de los alumnos de esta clase no es necesario tomar una muestra.

En este caso tomaríamos toda la población pues hay "pocos" alumnos.

Variable estadística :Es la característica que queremos estudiar de la población. En el ejemplo anterior, la variable estadística es el peso

Las variables estadísticas se clasifican de la siguiente forma:

Cualitativas: Si los valores que toma son cualidades. Por ejemplo, partido político preferido, color del pelo, etc.

Cuantitativas: Si los valores que toma son numéricos. Por ejemplo, nº de hermanos, estatura, peso, edad, etc.

Discretas: Cuando sólo puede tomar valores aislados. Por ejemplo, nº de hermanos, edad, etc.

Continuas: Cuando entre dos valores, aunque estén muy próximos entre sí, siempre se puede tomar otro valor. Por ejemplo, la temperatura, el peso, etc.

Ejercicio 1

Indica, para cada caso, cuál es la población, los individuos, la variable y de qué tipo es.

Indica también si lo más conveniente es estudiar la población o sólo una muestra.

	Población	Individuo	Variable estadística	Tipo de variable	¿Población o Muestra?
La marca de coche de los profesores de Matemáticas del instituto					
Número de asignaturas suspensas de los alumnos de 4º ESO A					
Nombre de los jóvenes españoles de 16 años					
Peso de los bebés nacidos en España en 2011					
Temperatura mínima en Iznalloz durante cada día de la semana pasada					

Ejercicio 2

Nos dicen que la estatura media de los europeos de 30 años es de 1,75 m.

a) Para llegar a este resultado, ¿se ha estudiado toda la población o sólo una muestra? _____

b) Indica algunas características de la muestra para que sea representativa de la población.

2.- FRECUENCIAS Y TABLAS

En estadística, para organizar y analizar los datos se utiliza una tabla, llamada **tabla de frecuencias**

Veamos un ejemplo.

Estas son las edades de un grupo de alumnos:

14, 15, 13, 13, 14 15, 15, 18, 14, 13
15, 13, 14, 15, 16 14, 15, 13, 13, 15

x_i	f_i	F_i	h_i	H_i
13	6	6	30%	30%
14	5	11	25%	55%
15	7	18	35%	90%
16	1	19	5%	95%
18	1	20	5%	100%
Suma total	20 = n		100%	

x_i son los **valores** que aparecen en los datos.

En la tabla se escriben ordenados de menor a mayor.

Si la variable estadística es continua o es discreta con "muchos" valores, para evitar tener que hacer una tabla demasiado grande, agruparemos los datos en intervalos, llamados **clases**, cerrados por la izquierda y abiertos por la derecha.

f_i se llama **frecuencia absoluta** y es el número de veces que aparece cada valor en los datos.

Por ejemplo, el número 7 de la columna f_i , significa que con 15 años, hay 7 alumnos.

La suma de las frecuencias absolutas es igual al nº total de datos "n"

En este caso n = 20, pues hay 20 datos.

F_i se llama **frecuencia absoluta acumulada** y se calcula sumando uno a uno los valores de la columna f_i .

F_i representa el número de datos que hay menores o iguales al valor x_i correspondiente.

Por ejemplo, el número 11 de la columna F_i , significa que con 14 años o menos, hay 11 alumnos.

h_i se llama **frecuencia relativa** y se calcula dividiendo cada f_i entre el nº total de datos, n.

$$h_i = \frac{f_i}{n} \quad \text{La frecuencia relativa se suele expresar en forma de \%}$$

La frecuencia relativa nos indica el % de datos que hay iguales al valor x_i correspondiente.

Por ejemplo, el 30% de la columna h_i significa que con 13 años, hay un 30% de alumnos.

La suma de las frecuencias relativas es siempre igual al 100%

H_i se llama **frecuencia relativa acumulada** y se calcula sumando uno a uno los valores de la columna h_i .

H_i también se puede calcular dividiendo cada valor de la columna F_i entre el nº total de datos, n.

La frecuencia relativa acumulada nos indica el % de datos que hay menores o iguales al valor x_i correspondiente

Por ejemplo, el 95% de la columna H_i significa que hay un 95% de alumnos que tienen 16 años o menos.

$$H_i = \frac{F_i}{n}$$

La frecuencia relativa acumulada se suele expresar en forma de %

Ejercicio 1

Se quiere saber el número de hijos de cada matrimonio de una ciudad. Para ello, se elige una muestra representativa de 50 matrimonios y se obtienen los siguientes datos:

2, 2, 4, 1, 3 5, 3, 2, 1, 6 3, 4, 1, 2, 0 2, 3, 1, 7, 4 2, 3, 0, 5, 1

4, 3, 2, 4, 1 5, 2, 1, 2, 4 0, 3, 3, 2, 6 1, 5, 4, 2, 0 3, 2, 4, 3, 1

a) Haz la tabla de frecuencias.

x_i	f_i	F_i	h_i	H_i

b) Indica el porcentaje de matrimonios que tiene 1 hijo

c) ¿Cuántos matrimonios tienen menos de 5 hijos?

d) Indica el porcentaje de matrimonios que tiene menos de 6 hijos

e) Halla el nº de matrimonios que tienen más de 3 hijos

f) Indica qué porcentaje representa

g) Calcula el nº de matrimonios que tienen 2, 4 ó 6 hijos

h) Indica qué porcentaje representa

Ejercicio 2

Las notas de 25 alumnos en un examen de Matemáticas han sido las siguientes:

7,2 ; 3,5 ; 5,4 ; 6,1 ; 4,5 8,3 ; 5,4 ; 2,5 ; 8 ; 6,5 3,5 ; 7,5 ; 6 ; 5 ; 4
6,5 ; 5,5 ; 4,2 ; 6,7 ; 1,5 5,3 ; 6,4 ; 6,3 ; 5,2 ; 6

a) Haz la tabla de frecuencias tomando intervalos o clases de amplitud 2 empezando por el valor 1 .

clases	fi	Fi	hi	Hi

b) ¿Qué porcentaje de alumnos están aprobados?

c) ¿Cuántos alumnos tienen menos de un 3?

d) ¿Cuántos alumnos tienen un 7 ó más?

e) ¿Qué porcentaje de alumnos está aprobado pero tiene menos de un 7?

Ejercicio 3

Se ha hecho una encuesta sobre la marca de coche en la ciudad de Granada, obteniéndose los siguientes resultados:

Marca	Porcentaje	Nº de personas
SEAT	25	
RENAULT	18,75	
PEUGEOT	28,125	
CITROEN	21,875	
SIN COCHE	

a) Completa la tabla calculando el porcentaje de personas que respondieron que no tenían coche

b) Si hubo 10 personas que respondieron que no tenían coche, ¿a cuántas personas se encuestó?

c) Completa la última columna.

d) Las personas encuestadas, ¿son población o muestra?

3.- GRÁFICOS ESTADÍSTICOS

Los datos obtenidos en un estudio estadístico los podemos representar con diferentes gráficos. Los gráficos nos ayudan a analizar los datos a simple vista.

Diagrama de barras

Se suele utilizar para variables discretas con "pocos" valores o para variables cualitativas

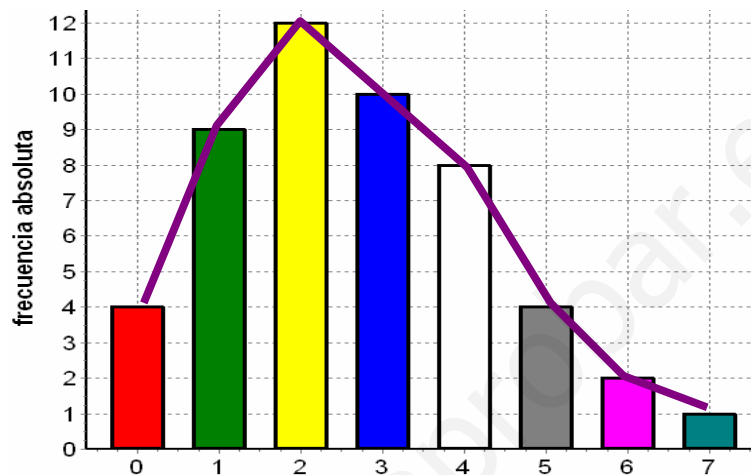
Se representan los valores x_i en un eje horizontal y para cada valor x_i se dibuja una barra cuya altura sea la frecuencia de x_i .

Las barras deben ser de la misma anchura y debemos dibujarlas separadas

Uniendo los extremos superiores de las barras por su punto medio, se obtiene una línea quebrada llamada **polígono de frecuencias**

Por ejemplo, para los datos correspondientes al nº de hijos de 50 matrimonios

X_i	f_i
0	4
1	9
2	12
3	10
4	8
5	4
6	2
7	1
Total	50 = n



Histograma

Se suele utilizar cuando los datos los agrupamos en intervalos o clases, es decir en el caso en que la variable es continua o discreta con "muchos" valores

Se representan los intervalos en un eje horizontal y para cada intervalo se dibuja una barra cuya altura sea la frecuencia en dicho intervalo.

Las barras debemos dibujarlas juntas

Uniendo los extremos superiores de las barras por su punto medio, se obtiene una línea quebrada llamada **polígono de frecuencias**

Por ejemplo, para los datos correspondientes a las notas en Inglés de 20 alumnos la tabla de frecuencias absolutas es la siguiente:

clases	f_i
[2,3)	3
[3,4)	2
[4,5)	3
[5,6)	5
[6,7)	3
[7,8)	4
Total	20 = n

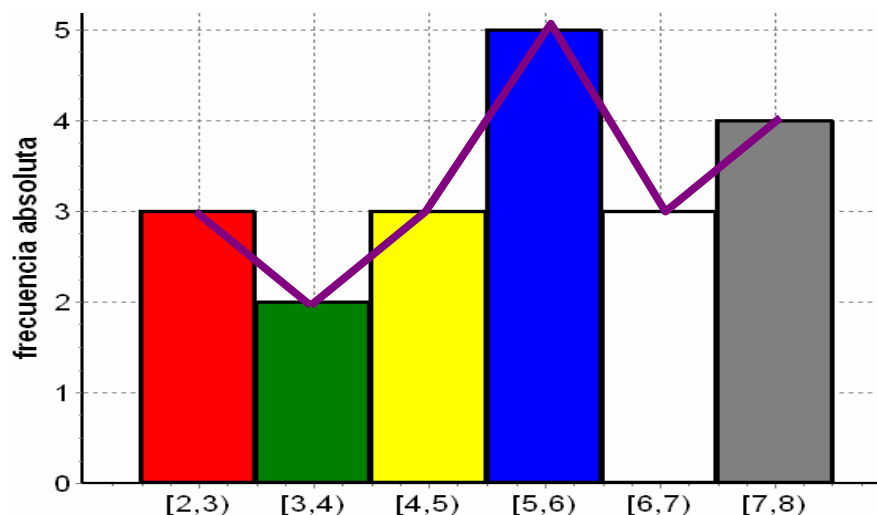


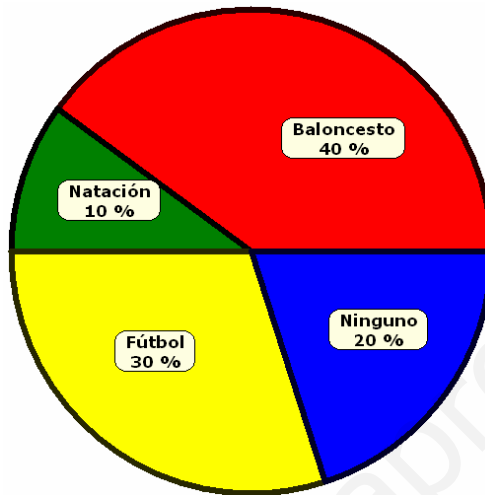
Diagrama de sectores

Se suele utilizar para variables discretas con "pocos" valores, cuando el nº de intervalos es pequeño o para variables cualitativas

Se dibuja un círculo y se divide en tantos sectores como valores o intervalos haya

Ejemplo: En una clase de 30 alumnos, 12 juegan a baloncesto, 3 practican la natación, 9 juegan al fútbol y el resto no practica ningún deporte

Deporte	f_i	h_i (%)	Ángulo del sector
Baloncesto	12	40	40% de $360^\circ = 144^\circ$
Natación	3	10	10% de $360^\circ = 36^\circ$
Fútbol	9	30	30% de $360^\circ = 108^\circ$
Ninguno	6	20	20% de $360^\circ = 72^\circ$
Total	$30 = n$	100	360°



Ejercicios propuestos:

Pág. 216: 7, 8 y 9

Pág. 217: 11 y 12 a)

Pág. 224: el 31

4.- MEDIDAS ESTADÍSTICAS

Los parámetros estadísticos son unos valores que se suelen calcular para analizar conjuntamente los datos

Media aritmética

Es la suma de todos los datos dividida entre el nº total de datos. Se calcula con la fórmula

$$\bar{x} = \frac{\sum(x_i \cdot f_i)}{n}$$

Si los datos están agrupados en intervalos, se toma como x_i el punto medio de cada intervalo. Este valor se llama **marca de clase**

La media aritmética sólo se puede calcular para variables cuantitativas

Ejemplos:

Número de hijos de 50 matrimonios

x_i	f_i	$x_i \cdot f_i$
0	4	0
1	9	9
2	12	24
3	10	30
4	8	32
5	4	20
6	2	12
7	1	7
Total	$50 = n$	134

Notas de 20 alumnos en un examen

clases	x_i (marca de clase)	f_i	$x_i \cdot f_i$
[2,3)	2,5	3	7,5
[3,4)	3,5	2	7
[4,5)	4,5	3	13,5
[5,6)	5,5	5	27,5
[6,7)	6,5	3	19,5
[7,8)	7,5	4	30
Total		$20 = n$	105

$$\bar{x} = \frac{\sum(x_i \cdot f_i)}{n} = \frac{134}{50} = 2,68$$

$$\bar{x} = \frac{\sum(x_i \cdot f_i)}{n} = \frac{105}{20} = 5,25$$

Moda

Es el valor que más se repite en los datos; es decir, el que tiene mayor frecuencia absoluta

Si los datos están agrupados en intervalos se toma el intervalo de mayor frecuencia (intervalo o clase modal).

Puede haber más de una moda o puede que no haya moda porque todos los valores tengan la misma frecuencia absoluta

Ejemplos:

Número de horas que ven semanalmente la televisión un grupo de jóvenes

Horas	Nº de jóvenes
1	2
2	3
3	5
4	3
5	2

La moda es 3 horas

Edad de los pacientes que acuden a una consulta médica

Edades	Nº de pacientes
[0,10)	7
[10,20)	10
[20,30)	10
[30,40)	18
[40,50)	12
[50,60)	30

El intervalo o clase modal es [50,60)

Equipo de fútbol preferido

Equipo	Nº de personas
Real Madrid	12
Atletico de Madrid	5
F.C. Barcelona	12
Sevilla F.C.	3

Hay 2 modas: Real Madrid y F.C. Barcelona. Es una distribución bimodal

Estatura (en metros) de 30 personas

Estatura	Nº de personas
1,60	6
1,65	6
1,72	6
1,75	6
1,84	6

No hay moda, pues todos los valores tienen la misma frecuencia absoluta

Cuartiles y mediana

Cuando los datos están ordenados de menor a mayor, los cuartiles son tres valores Q_1 , Q_2 , Q_3 que dividen los datos en 4 partes iguales

El primer cuartil, **Q_1** , es el primer valor x_i cuya H_i supera el **25%**

El segundo cuartil o **mediana**, **$Me = Q_2$** , es el primer valor x_i cuya H_i supera el **50%**

El tercer cuartil, **Q_3** , es el primer valor x_i cuya H_i supera el **75%**

Si los datos estuviesen agrupados en intervalos se toma como x_i la marca de clase

Los cuartiles sólo se pueden calcular para variables cuantitativas

Cuando hay "pocos" datos: Si el nº de datos es impar, la mediana es el dato central
Si el nº de datos es par, la mediana es la media aritmética de los 2 datos centrales

Ejemplos:

Edades de 9 personas:
15, 12, 17, 15, 14, 14, 17, 15, 15

Se ordenan de menor a mayor

12, 14, 14, 15, 15, 15, 15, 17, 17

$Me = 15$

Notas de 12 alumnos:

7, 4, 6, 5, 7, 7, 8, 5, 8, 4, 4, 5

Se ordenan de menor a mayor

4, 4, 4, 5, 5, 5, 6, 7, 7, 7, 8, 8

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

Nº de hijos de 50 matrimonios

x_i	f_i	F_i	H_i
0	4	4	8
1	9	13	26
2	12	25	50
3	10	35	70
4	8	43	86
5	4	47	94
6	2	49	98
7	1	50	100
	50 = n		

$Q_1 = 1$ $Q_2 = Me = 3$ $Q_3 = 4$

Notas en el examen de Inglés de 20 alumnos

clases	x_i	f_i	F_i	H_i
[2,3)	2,5	3	3	15
[3,4)	3,5	2	5	25
[4,5)	4,5	3	8	40
[5,6)	5,5	5	13	65
[6,7)	6,5	3	16	80
[7,8)	7,5	4	20	100
Total		20 = n		

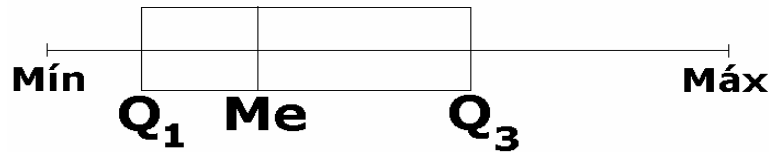
$Q_1 = 4,5$ $Q_2 = Me = 5,5$ $Q_3 = 6,5$

Representación gráfica de los cuartiles

Los cuartiles se suelen representar en un diagrama, llamado **diagrama de caja**

Para dibujar el diagrama de caja, se calculan los valores mínimo y máximo de x_i así como los cuartiles.

Después se dibuja una caja, cuyos extremos son Q_1 y Q_3 , que indica donde se concentran el 50% de los datos y una línea central que marca la mediana.



Esta representación nos permite saber qué datos son atípicos y si la distribución de datos es simétrica respecto de la mediana

Los datos fuera de la caja se alejan de la mayoría de datos

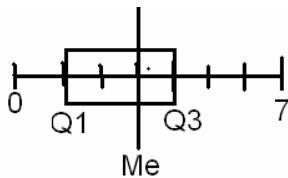
Si la caja está desplazada hacia la izquierda o hacia la derecha respecto de la mediana significa que la distribución de datos es asimétrica

Ejemplos:

Nº de hijos de 50 matrimonios

Min = 0 Máx. = 7

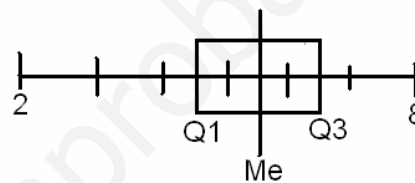
$Q_1 = 1$ $Q_2 = Me = 3$ $Q_3 = 4$



Notas en el examen de Inglés de 20 alumnos

Min = 2 Máx. = 8

$Q_1 = 4,5$ $Q_2 = Me = 5,5$ $Q_3 = 6,5$



Recorrido o rango Es la diferencia entre el mayor y el menor valor de x_i

Varianza Se calcula con la fórmula

$$\sigma^2 = \frac{\sum (x_i^2 \cdot f_i)}{n} - \bar{x}^2$$

Coefficiente de variación

$$CV = \frac{\sigma}{\bar{x}}$$

Desviación típica Se calcula con la fórmula

$$\sigma = \sqrt{\sigma^2}$$

Estos parámetros nos indican si los datos están más o menos dispersos con respecto a la media

En el intervalo $(\bar{x} - \sigma, \bar{x} + \sigma)$ se encuentra, aproximadamente los 2/3 de todos los datos

Número de hijos de 50 matrimonios

Ejemplos:

Notas de 20 alumnos en un examen

x_i	f_i	$x_i \cdot f_i$	$x_i^2 f_i$
0	4	0	0
1	9	9	9
2	12	24	48
3	10	30	90
4	8	32	128
5	4	20	100
6	2	12	72
7	1	7	49
Total	50 = n	134	496

clases	x_i (marca de clase)	f_i	$x_i \cdot f_i$	$x_i^2 f_i$
[2,3)	2,5	3	7,5	18,75
[3,4)	3,5	2	7	24,5
[4,5)	4,5	3	13,5	60,75
[5,6)	5,5	5	27,5	151,25
[6,7)	6,5	3	19,5	126,75
[7,8)	7,5	4	30	225
Total		20 = n	105	607

Recorrido = 7 - 0 = 7

$$\bar{x} = \frac{\sum (x_i \cdot f_i)}{n} = \frac{134}{50} = 2,68$$

$$\sigma^2 = \frac{\sum (x_i^2 \cdot f_i)}{n} - \bar{x}^2 = \frac{496}{50} - 2,68^2 = 2,7376$$

$$\sigma = \sqrt{\sigma^2} = 1,6546 \quad CV = \frac{\sigma}{\bar{x}} = \frac{1,6546}{2,68} = 0,6174$$

Recorrido = 8 - 2 = 6

$$\bar{x} = \frac{\sum (x_i \cdot f_i)}{n} = \frac{105}{20} = 5,25$$

$$\sigma^2 = \frac{\sum (x_i^2 \cdot f_i)}{n} - \bar{x}^2 = \frac{607}{20} - 5,25^2 = 2,7875$$

$$\sigma = \sqrt{\sigma^2} = 1,6696 \quad CV = \frac{\sigma}{\bar{x}} = \frac{1,6696}{5,25} = 0,318$$

