

ESTADÍSTICA. TABLAS Y GRÁFICOS

1. ESTADÍSTICA: CONCEPTOS BÁSICOS

La *Estadística* surge ante la necesidad de poder tratar y comprender conjuntos numerosos de datos. En sus orígenes históricos, estuvo ligada a cuestiones de Estado (recuentos, censos, etc.) y de ahí proviene su nombre. En la actualidad, los estudios estadísticos están presentes en todos los ámbitos humanos, tanto individuales como colectivos.

Supongamos que queremos saber:

- a) ¿Cuál es la proporción de votantes a un determinado candidato?
- b) ¿Cuántos ordenadores de una cierta marca salen defectuosos?
- c) ¿Cuántas horas duran las bombillas de una determinada marca?

Los anteriores ejemplos tienen todos una característica común: se trata de estudiar un conjunto de elementos, potencialmente muy grande, que llamaremos *población*. Los elementos de la población, llamados *individuos* u *objetos*, pueden ser personas, ordenadores, bombillas, animales, árboles, etc.

En principio, habría que estudiar todos los elementos de la población, pero diversas razones impiden hacerlo, unas veces de carácter intrínseco (las bombillas) y otras de tiempo y dinero. Por ello elegimos una parte de la población que llamamos *muestra* y realizamos el estudio sobre los elementos de la misma. El número de individuos que la componen se llama *tamaño muestral*.

La **Estadística** es la ciencia que se ocupa de la recogida de datos, su organización y análisis; así como de las predicciones que, a partir de estos datos, pueden hacerse.

Los conceptos básicos que aparecen en cualquier estudio estadístico son:

- **Población** es el conjunto formado por todos los elementos que existen para el estudio de una determinada característica.
- **Individuo** u **objeto** es cada elemento de la población.
- **Muestra** es el subconjunto que tomamos de la población para determinar el estudio.
- **Tamaño muestral** es el número de individuos que componen la muestra.

1.1. Caracteres y variables estadísticos

Cada una de las cualidades o propiedades referidas a los elementos de una población objeto de estudio se llama **carácter estadístico**. Las clases de caracteres estadísticos que aparecen en cualquier estudio pueden ser:

- **Caracteres cualitativos** son aquellos que no se pueden medir y se describen con palabras.
- **Caracteres cuantitativos** son aquellos que se pueden medir y expresar con números.

Se llama **variable estadística** al conjunto de valores que toma un carácter estadístico cuantitativo. Puede ser:

- **Discreta** cuando solamente puede tomar un número finito de valores numéricos (valores aislados).
- **Continua** cuando puede tomar cualquier valor dentro de un intervalo.

Observa el siguiente esquema con algunos ejemplos:

CARACTERES ESTADÍSTICOS		
CUALITATIVOS (no se pueden medir)	CUANTITATIVOS (se pueden medir) dan lugar a VARIABLES ESTADÍSTICAS	
	VAR. DISCRETA	VAR. CONTINUA
<u>Ejemplos</u> Color de los ojos Estado civil Deporte preferido	<u>Ejemplos</u> N° de hermanos N° de discos vendidos N° de goles marcados	<u>Ejemplos</u> Estatura de los alumnos Temperatura Peso de los animales

1.2. Encuestas y muestreo

En el estudio de cualquier fenómeno estadístico y para conocer los datos hay que preguntar a un número determinado de individuos.

Se llama **encuesta** a las preguntas que se formulan a un cierto número de individuos de un colectivo o población.

Las encuestas deben contener las preguntas precisas que nos permiten conseguir los datos que nos hacen falta, por ello es importante elaborar bien el cuestionario de preguntas.

Una vez que se ha elaborado una encuesta hay que plantearse a quién se la hacemos. En unos casos se aplica a toda la *población* y, en otros, a una parte suya que llamamos *muestra*. Es conveniente conocer los siguientes criterios básicos en la elección de muestras:

- Cada elemento de la población debe tener igual oportunidad de encontrarse en la muestra.
- Las características de los elementos de la muestra han de reproducir, con la máxima exactitud posible, los de la población.
- Si la muestra es demasiado pequeña puede inducir a errores y si es demasiado grande no resulta manejable.

A las técnicas que nos permiten elegir una muestra se le llama **muestreo**. Un método muy sencillo para elegir una muestra es asignar un número a cada elemento de la población, se introducen los números en un bombo y se extraen al azar tantos números como elementos queramos que tenga la *muestra (tamaño muestral)*. Esta forma de elegir la muestra se denomina **muestreo aleatorio**.

En este curso no veremos cómo escoger las muestras, pero sí cómo saber su tamaño aplicando la técnica del **muestreo proporcional**.

Supongamos que queremos tomar una muestra de personas y preguntarles por su preocupación por el medio ambiente. Veamos cómo debemos realizar el proceso si las personas son de tres provincias distintas.

Llamemos N_1 , N_2 y N_3 a los habitantes de las tres provincias. El problema es repartir el total de la muestra, n , en tres partes n_1 , n_2 y n_3 de forma que se conserve la proporción que guarda la muestra con el número total de personas de la población a estudiar N ($N = N_1 + N_2 + N_3$).

Se trata pues de hacer un reparto proporcional del total de la muestra entre las tres poblaciones, es decir, debe cumplirse que:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{n_1 + n_2 + n_3}{N_1 + N_2 + N_3} = \frac{n}{N}$$

Las muestras así elegidas son **representativas de la población**; eso quiere decir que los resultados que obtengamos a partir de ellas serán análogos a los que obtendríamos en la población.

Ejemplo.- En un centro de Educación Secundaria hay 100 alumnos de 1º de ESO, 80 de 2º, 70 de 3º y 50 de 4º. Se quiere tomar una muestra representativa de 60 alumnos de ESO para hacerles una prueba psicotécnica. ¿Cuántos alumnos habrá que tomar de cada curso?

Llamamos n_1 a la muestra de alumnos de 1º, n_2 a la de 2º, n_3 a la de 3º y n_4 a la de 4º. Debe cumplirse que:

$$\frac{n_1}{100} = \frac{n_2}{80} = \frac{n_3}{70} = \frac{n_4}{50} = \frac{n_1 + n_2 + n_3 + n_4}{100 + 80 + 70 + 50} = \frac{60}{300}$$

De ahí resulta entonces que:

$$\frac{n_1}{100} = \frac{60}{300} \Rightarrow n_1 = \frac{60 \cdot 100}{300} = 20 ; \quad \frac{n_2}{80} = \frac{60}{300} \Rightarrow n_2 = \frac{60 \cdot 80}{300} = 16$$
$$\frac{n_3}{70} = \frac{60}{300} \Rightarrow n_3 = \frac{60 \cdot 70}{300} = 14 ; \quad \frac{n_4}{50} = \frac{60}{300} \Rightarrow n_4 = \frac{60 \cdot 50}{300} = 10$$

EJERCICIOS

- Clasifica los siguientes caracteres estadísticos según sean cualitativos, variables discretas o variables continuas.
 - Diámetro de una pieza de precisión.
 - La categoría de cada club de fútbol español.
 - Número de acciones vendidas en bolsa.
 - El número de satélites de cada planeta del Sistema Solar.
 - Capacidad del depósito de gasolina de un coche.
 - Tipos de zumos que prefieren los alumnos de tu clase.
 - El número de habitaciones de las viviendas de tu barrio.
 - La profesión que piensan tener los alumnos de tu clase.
 - El peso de cada una de las manzanas producidas por un determinado manzano.
 - Los habitantes de cada una de las provincias españolas.
- En una ciudad hay 1.500 abonados a un canal A de televisión, 1.200 a otro B y 800 a otro C. Se quiere tomar una muestra representativa de 70 abonados para ver sus preferencias. ¿Cuántos abonados a cada canal habrá que encuestar?
- En una encuesta se quiere estudiar una población compuesta por dos millones de habitantes, de los que el 65 % son mujeres, y se va a escoger una muestra de 2.000 personas. ¿Cuántas mujeres deberán figurar en la muestra?
- Los habitantes de derecho que viven en aldeas son aproximadamente 388.000 en la provincia A, 160.000 en la provincia B, 29.000 en C y 23.000 en D. Si queremos entrevistar a 600 personas de las aldeas sobre el ganado, ¿qué muestra tendremos que tomar en cada provincia?

2. TABLAS ESTADÍSTICAS

2.1. Recuento

A partir de la recogida de datos, a través de encuestas o entrevistas, han de ordenarse para un mejor manejo. La forma usual de ordenarlos consiste en realizar un *recuento* y, posteriormente, ordenarlos en una tabla.

El **recuento** en Estadística se realiza de la siguiente forma:

- Se ordenan las cualidades o valores que puede tomar la variable estadística, colocándolos en la primera columna de la tabla.
- En una segunda columna, cada vez que aparece un dato correspondiente a una cualidad o valor se traza una pequeña marca.
- Añadimos una tercera columna donde se anota el número total de marcas trazadas para cada cualidad o valor.

Ejemplo.- Preguntados los alumnos de una clase por el color de la pared de su habitación, han respondido:
«blanco, marrón, blanco, marrón, blanco, blanco, blanco, marrón, marrón, azul, blanco, marrón, blanco, rosa, verde, rosa, blanco, verde, blanco, verde, blanco, azul, azul, amarillo, amarillo».

Color	Recuento	Total
Amarillo	☐	2
Azul	☐	3
Blanco	☑☑	10
Marrón	☑	5
Rosa	☐	2
Verde	☐	3

Ejemplo.- Preguntados por el número de hermanos de cada uno, las respuestas han sido:

2, 3, 3, 2, 2, 3, 3, 1, 1, 2, 2, 3, 1, 1, 4, 2, 2, 3, 4, 2, 1, 2, 4, 2, 5

Nº de hermanos	Recuento	Total
1	☑	5
2	☑☑	10
3	☑	6
4	☐	3
5		1

Como acabamos de ver, cuando se trabaja con caracteres cualitativos o variables discretas, las cualidades o valores de éstas aparecen descritos en la tabla.

En el caso de las variables cuantitativas continuas, o discretas con un número de datos muy grande, los datos deben agruparse en **clases** o **intervalos**.

El valor medio de cada clase o intervalo se llama **marca de clase** y se calcula como la semisuma de los extremos del intervalo.

Para construir intervalos o clases hemos de tener en cuenta los siguientes puntos:

- Todos los intervalos deben tener la misma amplitud. Sólo en el caso de que existan valores muy dispersos tomaremos distintas amplitudes.
- Con el fin de que la clasificación esté bien hecha, los intervalos se deben construir de tal manera que el extremo superior de una clase coincida con el extremo inferior de la siguiente. Así, tomaremos intervalos cerrados por la izquierda y abiertos por la derecha de la forma $[a, b)$, a excepción del último intervalo que será cerrado por ambos extremos, esto es, de la forma $[c, d]$.
- Calculamos la diferencia entre el valor más grande y el más pequeño de la variable a estudio. Esta diferencia se llama **rango** o **recorrido de la variable**.
- Es conveniente que el número de intervalos que tomemos esté entre 5 y 10.
- Calculamos la amplitud de cada intervalo dividiendo el recorrido por el número de intervalos que tomemos.

En el siguiente ejemplo puedes ver cómo se realiza el recuento en estos casos.

Ejemplo.- Los pesos de 60 alumnos que cursan 3º de ESO, en kg, son:

48, 48, 50, 55, 59, 54, 60, 60, 64, 65, 69, 57, 62, 70, 59, 57, 52, 60, 53, 57
53, 45, 54, 47, 59, 60, 57, 66, 62, 55, 52, 63, 67, 58, 49, 47, 54, 62, 57, 53
51, 46, 54, 67, 62, 46, 49, 57, 56, 64, 60, 50, 48, 67, 53, 49, 63, 68, 55, 60

Observamos que los valores extremos son 45 y 70. El recorrido de la variable es $70 - 45 = 25$. Si deseamos tener cinco intervalos, éstos tendrán $25/5 = 5$ de amplitud.

El recuento nos lleva a la tabla siguiente:

<i>Pesos (kg)</i>	<i>Marcas de clase</i>	<i>Recuento</i>	<i>Total</i>
[45, 50)	47'5	☑☑	11
[50, 55)	52'5	☑☑☐	14
[55, 60)	57'5	☑☑☐	14
[60, 65)	62'5	☑☑☐	13
[65, 70]	67'5	☑☐	8

Observa que en el intervalo [45, 50) se contabilizan todos los pesos desde 45 kg (incluido este valor) hasta 50 kg (excluido este valor, que se contabiliza en la siguiente clase). El último intervalo, [65, 70] es cerrado por ambos extremos.

2.2. Frecuencias simples

En lo sucesivo, a los k valores de una variable estadística los notaremos por $x_1, x_2, x_3, \dots, x_k$; y designaremos por N al tamaño de la muestra o la población objeto de estudio.

• Frecuencia absoluta

En las tablas construidas en los ejemplos anteriores, en la columna correspondiente al total se han obtenido unos valores cuyo significado es el número de veces que se presenta cada cualidad o valor de la variable estadística. A estos números les llamamos *frecuencias absolutas*.

Frecuencia absoluta, n_i , de una cualidad o de un valor x_i de la variable estadística es el número total de veces que aparece esta cualidad o valor.

La correspondencia que asocia a cada modalidad o valor de la variable su frecuencia absoluta se llama **distribución estadística**.

La suma de todas las frecuencias absolutas es necesariamente el tamaño de la muestra o la población a estudio:

$$n_1 + n_2 + n_3 + \dots + n_k = \sum_{i=1}^k n_i = N$$

• Frecuencia relativa

En ocasiones nos interesa saber cuál es la proporción del número de individuos con un valor determinado respecto del total. Estos valores reciben el nombre de *frecuencias relativas*.

Frecuencia relativa, f_i , de una cualidad o de un valor x_i de la variable estadística es el cociente que resulta de dividir su frecuencia absoluta entre el número total N de individuos.

$$f_i = \frac{n_i}{N}, \quad i = 1, \dots, k$$

Representa la proporción de éstos sobre el total y verifica $0 \leq f_i \leq 1$

Observa que $f_1 + f_2 + f_3 + \dots + f_k = \frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} + \dots + \frac{n_k}{N} = \frac{n_1 + n_2 + n_3 + \dots + n_k}{N} = \frac{N}{N} = 1$

Luego la suma de todas las frecuencias relativas es la unidad:

$$f_1 + f_2 + f_3 + \dots + f_k = \sum_{i=1}^k f_i = 1$$

- **Frecuencia porcentual**

Las frecuencias relativas representan la proporción de individuos con respecto al total. Estas proporciones se pueden expresar como *porcentajes*, que aparecen muy frecuentemente en las tablas estadísticas.

Frecuencia porcentual o porcentaje, p_i , de una cualidad o de un valor x_i de la variable estadística es el tanto por ciento que representa este valor o cualidad respecto del total. Lógicamente, se calcula multiplicando la frecuencia relativa por 100 y verifica $0 \leq p_i \leq 100$.

$$p_i = f_i \cdot 100, \quad i = 1, \dots, k$$

Obviamente, la suma de todos los porcentajes es 100:

$$p_1 + p_2 + p_3 + \dots + p_k = f_1 \cdot 100 + f_2 \cdot 100 + f_3 \cdot 100 + \dots + f_k \cdot 100 = (f_1 + f_2 + f_3 + \dots + f_k) \cdot 100 = 1 \cdot 100 = 100$$

$$p_1 + p_2 + p_3 + \dots + p_k = \sum_{i=1}^k p_i = 100$$

2.3. Frecuencias acumuladas

A menudo nos interesa conocer cuántos datos estadísticos presentan valores, proporciones o porcentajes que son menores o iguales a uno dado. Para este fin se estudian las *frecuencias acumuladas*.

- **Frecuencia absoluta acumulada**

Frecuencia absoluta acumulada, N_i , de una cualidad o de un valor x_i de la variable estadística es la suma de todas las frecuencias absolutas de los valores menores o iguales a x_i .

$$N_i = n_1 + n_2 + n_3 + \dots + n_i, \quad i = 1, \dots, k$$

- **Frecuencia relativa acumulada**

Frecuencia relativa acumulada, F_i , de una cualidad o de un valor x_i de la variable estadística es la suma de todas las frecuencias relativas de los valores menores o iguales a x_i .

$$F_i = f_1 + f_2 + f_3 + \dots + f_i, \quad i = 1, \dots, k$$

Como $F_i = f_1 + f_2 + f_3 + \dots + f_i = \frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} + \dots + \frac{n_i}{N} = \frac{n_1 + n_2 + n_3 + \dots + n_i}{N} = \frac{N_i}{N}$, podemos también

definir la *frecuencia relativa acumulada*, F_i , de una cualidad o de un valor x_i de la variable estadística como el cociente que resulta de dividir su frecuencia absoluta acumulada N_i entre el número total N de individuos:

$$F_i = \frac{N_i}{N}, \quad i = 1, \dots, k$$

- **Frecuencia porcentual acumulada**

Al igual que anteriormente, estas proporciones se pueden expresar mediante porcentajes.

Frecuencia porcentual acumulada o porcentaje acumulado, P_i , de una cualidad o de un valor x_i de la variable estadística es la suma de las frecuencias porcentuales de los valores menores o iguales a x_i .

$$P_i = p_1 + p_2 + p_3 + \dots + p_i, \quad i = 1, \dots, k$$

Desarrollando esta expresión:

$$P_i = p_1 + p_2 + \dots + p_i = f_1 \cdot 100 + f_2 \cdot 100 + \dots + f_i \cdot 100 = (f_1 + f_2 + \dots + f_i) \cdot 100 = F_i \cdot 100$$

Esto es, la **frecuencia porcentual acumulada**, P_i , de una cualidad o de un valor x_i de la variable estadística es el tanto por ciento que representan los valores menores o iguales a x_i respecto del total. Se pueden hallar multiplicando la frecuencia relativa acumulada por 100.

$$P_i = F_i \cdot 100, \quad i = 1, \dots, k$$

En la elaboración de una tabla estadística de frecuencias deben figurar las cualidades o los valores de la variable (las marcas de clase en caso de que se encuentre agrupada en intervalos) y las frecuencias absolutas y relativas. En la mayoría de los casos, es conveniente incluir las frecuencias absolutas acumuladas, las frecuencias relativas acumuladas y los porcentajes. A continuación tienes el modelo de una tabla estadística donde se reflejan todas las frecuencias:

$$N_i = n_1 + n_2 + n_3 + \dots + n_k$$

$$f_i = n_i / N$$

$$F_i = f_1 + f_2 + f_3 + \dots + f_i = N_i / 100$$

$$p_i = f_i \cdot 100$$

$$P_i = p_1 + p_2 + p_3 + \dots + p_i = F_i \cdot 100$$

x_i	n_i	N_i	f_i	F_i	p_i	P_i
x_1	n_1	$N_1 = n_1$	f_1	$F_1 = f_1$	p_1	$P_1 = p_1$
x_2	n_2	N_2	f_2	F_2	p_2	P_2
x_3	n_3	N_3	f_3	F_3	p_3	P_3
...	...	...	...	...	...	...
x_k	n_k	$N_k = N$	f_k	$F_k = 1$	p_k	$P_k = 100$
Total	N		1		100	

Ejemplo.- Vamos a construir la tabla de frecuencias con los datos obtenidos para la variable estadística *número de hermanos* que figura en el epígrafe 2.1.

Nº de hermanos	n_i	N_i	f_i	F_i	p_i	P_i
1	5	5	0'20	0'20	20	20
2	10	15	0'40	0'60	40	60
3	6	21	0'24	0'84	24	84
4	3	24	0'12	0'96	12	96
5	1	25	0'04	1	4	100
Total	25		1		100	

Puedes observar, por ejemplo, en la siguiente tabla que:

- Hay 6 alumnos que tienen 3 hermanos.
(La *frecuencia absoluta* del valor 3 de la variable es 6)
- Hay 21 alumnos que tienen 3 o menos hermanos.
(La *frecuencia absoluta acumulada* del valor 3 de la variable es 21)
- La proporción de alumnos que tienen 3 hermanos es 0'24.
(La *frecuencia relativa* del valor 3 de la variable es 0'24)
- La proporción de alumnos que tienen 3 o menos hermanos es 0'84.
(La *frecuencia relativa acumulada* del valor 3 de la variable es 0'84)
- El 24 % de los alumnos tienen 3 hermanos.
(La *frecuencia porcentual* del valor 3 de la variable es 24)
- El 84 % de los alumnos tienen 3 o menos hermanos.
(La *frecuencia porcentual acumulada* del valor 3 de la variable es 84)

Ejemplo.- A continuación se muestra la tabla de la distribución estadística *color de la pared de la habitación* con los datos obtenidos en el epígrafe 2.1.

<i>Color</i>	n_i	N_i	f_i	F_i	p_i	P_i
Amarillo	2	2	0'08	0'08	8	8
Azul	3	5	0'12	0'20	12	20
Blanco	10	15	0'40	0'60	40	60
Marrón	5	20	0'20	0'80	20	80
Rosa	2	22	0'08	0'88	8	88
Verde	3	25	0'12	1	12	100
Total	25		1		100	

Ejemplo.- Por último, observa la tabla de frecuencias de la variable estadística continua *peso de los alumnos de 3º de ESO* cuyos datos también se obtuvieron en el epígrafe 2.1.

<i>Peso (kg)</i>	<i>Marcas de clase</i>	n_i	N_i	f_i	F_i	p_i	P_i
[45, 50)	47'5	11	11	11/60	11/60	18'33	18'33
[50, 55)	52'5	14	25	14/60	25/60	23'33	41'67
[55, 60)	57'5	14	39	14/60	39/60	23'33	65
[60, 65)	62'5	13	52	13/60	52/60	21'67	86'67
[65, 70]	67'5	8	60	8/60	1	13'33	100
Total		60		1		100	

EJERCICIOS

5. La tabla (que aparece incompleta) resume las calificaciones obtenidas por los 80 alumnos de una clase de cierta Universidad. Completa la tabla con las frecuencias que faltan.

<i>Calificación</i>	n_i	N_i	f_i	F_i	p_i	P_i
Insuficiente			0'375			
Suficiente	20					
Bien	10					
Notable	6					
Sobresaliente						
Total	80					

6. Hemos lanzado al aire 4 monedas 50 veces y hemos anotado el número de caras obtenido en cada lanzamiento. Los resultados son:

3 4 2 2 1 1 1 0 2 3
 0 3 2 2 2 1 3 1 1 2
 4 2 2 1 0 2 2 1 1 0
 3 2 4 2 2 3 1 1 0 2
 3 1 2 4 2 1 0 2 3 2

Haz el recuento y la correspondiente tabla estadística de frecuencias.

7. En un laboratorio científico se midió durante varios años la velocidad de la luz, de forma experimental. Los treinta valores que se obtuvieron, en km/s, fueron:

300.066 300.075 299.400 299.804 299.751 300.012 300.120 299.781 300.380 300.000
 300.021 299.507 300.202 300.043 299.904 299.904 299.825 299.806 299.712 299.875
 299.603 300.090 299.789 300.035 300.082 299.984 299.822 300.119 300.144 299.640

Observa que el mayor valor es 300.380 y el menor 299.400. El recorrido es $300.380 - 299.400 = 980$. Hacemos cinco intervalos, luego la amplitud de cada uno será $980/5 = 196$. Los intervalos que hay que considerar serán entonces $[299.400, 299.596)$, $[299.596, 299.792)$, $[299.792, 299.988)$, $[299.988, 300.184)$ y $[300.184, 300.380]$.

A la vista de lo anterior calcula la tabla estadística con intervalos, marcas de clase y todas las frecuencias correspondientes.

8. La densidad de población, en habitantes por km^2 , de las provincias españolas, en el año 1986, era de:

50 141 54 63 43 48 158 110 13 10 48 105 136 184 226 99 23 24
 32 23 25 34 24 29 22 10 60 21 597 83 29 83 208 65 193 31
 41 59 201 598 89 50 88 345 532 52 12 141 13 21 3.620 3.742

En ocasiones no resulta útil usar intervalos de la misma amplitud, como ocurre en este caso. Utiliza para ello los intervalos $[0, 20)$, $[20, 40)$, $[40, 80)$, $[80, 160)$, $[160, 3.800]$ y elabora la tabla estadística completa de esta distribución.

9. Completa los datos que faltan en las tablas estadísticas siguientes.

x_i	n_i	N_i	f_i
1	3		0'06
2	4		
3		12	0'10
4	6		0'12
5	8		
6		36	
7		43	
8	3		
9		49	
10			0'02
Total			

x_i	n_i	N_i	p_i
1	4		6'2500
2	6		
3		15	7'8125
4	8		12'5000
5	10		
6		42	
7		50	
8	5		
9		59	
10			7'8125
Total			

x_i	n_i	f_i	F_i
1	3		
2			0'08
3	12		0,20
4			0'35
5			0'53
6	17		
7	16		
8			0'94
9	4		
10			
Total			

10. En una cosecha se escoge una muestra de melocotones para estudiar su peso (gramos). Los resultados obtenidos se muestran en la siguiente tabla.

Peso (gramos)	[119, 136)	[136, 153)	[153, 170)	[170, 187)	[187, 204]
Nº de melocotones	10	10	3	7	10

- ¿Cuál es el carácter estadístico objeto de estudio? ¿Es cualitativo o cuantitativo?
- Define la población, la muestra y el tamaño muestral de la misma.
- Elabora la tabla completa de frecuencias (absolutas, relativas y porcentuales) de la distribución estadística.
- Responde a las siguientes cuestiones:
 - ¿Cuántos melocotones pesan entre 153 y 170 gramos?
 - ¿Qué porcentaje de melocotones tiene un peso inferior a 187 gramos?
 - ¿Qué proporción de melocotones tiene un peso comprendido entre 136 y 153 gramos?
 - ¿Cuántos melocotones pesan menos de 153 gramos?
 - ¿Qué porcentaje de melocotones tiene un peso que oscila entre 170 y 187 gramos?
 - ¿Qué porcentaje de melocotones tiene un peso superior a 153 gramos?

3. REPRESENTACIONES GRÁFICAS

Las tablas estadísticas muestran la información de forma esquemática y están preparadas para cálculos posteriores. Con el fin de hacerla más clara y evidente, la misma información estadística se puede mostrar de forma global y más expresiva utilizando los **gráficos estadísticos**. Los gráficos poseen un fuerte poder de comunicación de los resultados de un estudio estadístico. Mencionaremos, a continuación, los principales tipos de gráficos.

3.1. Diagrama de sectores

Se utiliza para comparar las distintas modalidades de un carácter. Consiste en un círculo dividido en tantos sectores circulares como modalidades tiene el carácter.

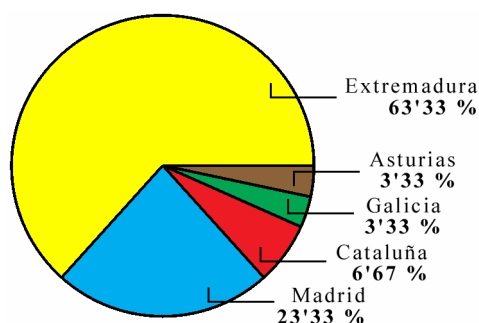
Para construirlo, el ángulo central de cada sector ha de ser proporcional a la frecuencia absoluta correspondiente de la cualidad a la que representa. Por tanto, la amplitud angular correspondiente a cada cualidad o valor de la variable se puede calcular mediante una simple regla de tres directa.

$$\begin{array}{lcl} N & \longleftrightarrow & 360^\circ \\ n_i & \longleftrightarrow & \text{Amplitud de } x_i \end{array} \Rightarrow \boxed{\text{Amplitud de } x_i = \frac{n_i \cdot 360^\circ}{N} = f_i \cdot 360^\circ}$$

En algunos gráficos se suelen incluir los porcentajes que representan las distintas cualidades o valores de la variable.

Ejemplo.- Representamos mediante un diagrama de sectores la distribución estadística que clasifica a los alumnos según la autonomía de nacimiento.

Autonomía	n_i	p_i	Medida del ángulo central
Extremadura	19	63'33	$19/30 \cdot 360^\circ = 228^\circ$
Madrid	7	23'33	$7/30 \cdot 360^\circ = 84^\circ$
Cataluña	2	6'67	$2/30 \cdot 360^\circ = 24^\circ$
Galicia	1	3'33	$1/30 \cdot 360^\circ = 12^\circ$
Asturias	1	3'33	$1/30 \cdot 360^\circ = 12^\circ$
Total	30	100	



3.2. Diagrama de barras y polígono de frecuencias

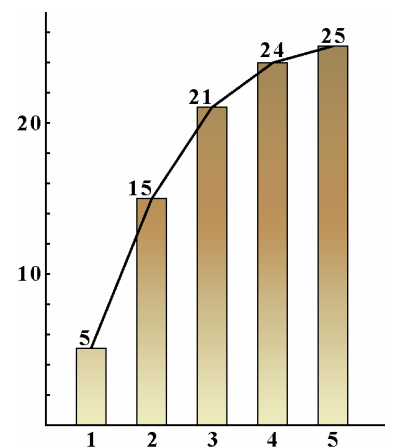
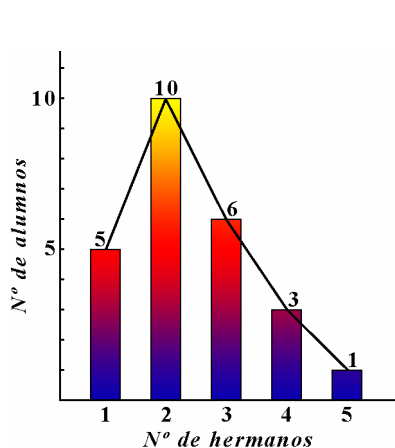
Se utiliza para comparar datos cualitativos o cuantitativos discretos. Para construir un diagrama de barras se representan sobre el eje de abscisas cada modalidad o valor de la variable, y para cada uno de ellos levantamos una barra de altura proporcional a su frecuencia absoluta. En muchas ocasiones se superponen dos o más diagramas con el fin de comparar los datos de diferentes situaciones.

Los polígonos de frecuencias se forman uniendo los extremos de las barras mediante una línea quebrada.

Estos gráficos son muy utilizados aquellos que representan las frecuencias acumuladas en el estudio del crecimiento de determinados fenómenos.

Ejemplo.- Representamos mediante un diagrama de barras y un polígono de frecuencias la distribución que clasifica a los alumnos según el número de hermanos.

Nº de hermanos	1	2	3	4	5
n_i	5	10	6	3	1
N_i	5	15	21	24	25



Nota.- El hecho de que las frecuencias relativas representen la proporción de cada cualidad o valor de la variable y la suma de éstas sea igual a la unidad, nos permite comparar gráficamente distintas distribuciones entre sí. Para ello se han de construir los diagramas de barras usando las frecuencias relativas. Piensa que no es lo mismo que haya 5 alumnos con un hermano en una clase de 25 que en otra clase de 20.

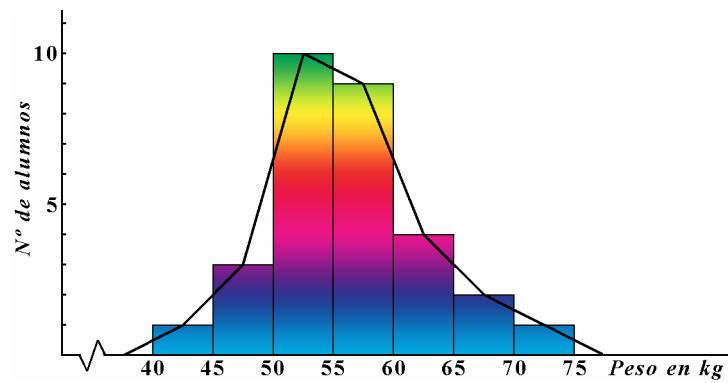
3.3. Histograma y polígono de frecuencias

Son análogos a los diagramas de barras y se utilizan para distribuciones de variable estadística continua o para distribuciones de variable estadística discreta cuyos datos han sido agrupados en clases. Para construirlo, se representan sobre el eje de abscisas los extremos de las clases y se levantan unos rectángulos yuxtapuestos de base la amplitud del intervalo y de altura proporcional a la frecuencia absoluta correspondiente, siempre que todos los intervalos tengan la misma amplitud. En el caso de que los intervalos tengan distinta amplitud, las alturas de los rectángulos se han de calcular de modo que sus áreas sean proporcionales a las correspondientes frecuencias. Al igual que mencionamos antes, para comparar gráficamente distintas distribuciones construiremos los histogramas usando las frecuencias relativas.

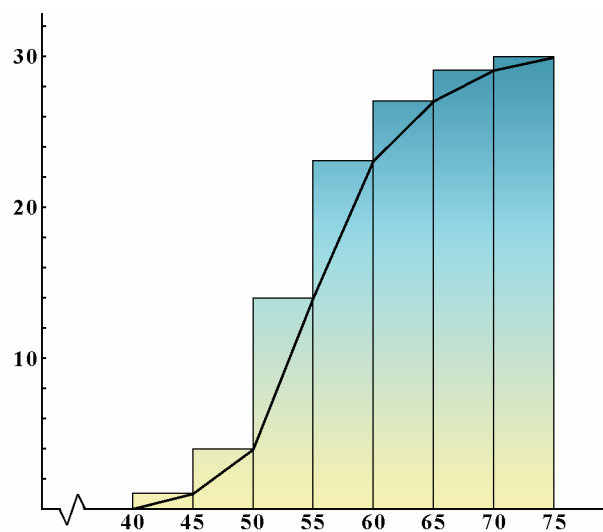
El polígono de frecuencias se obtiene al unir los puntos medios de los lados superiores de cada rectángulo. Con el fin de que el área encerrada bajo el polígono de frecuencias sea igual a la suma de las áreas de los rectángulos, se une el extremo por la izquierda del polígono con la marca de la clase anterior; análogamente, con el extremo por la derecha.

Ejemplo.- En el siguiente histograma y polígono de frecuencias se ha representado la distribución estadística que clasifica a los alumnos según su peso en kilogramos.

Peso (kg)	[40, 45)	[45, 50)	[50, 55)	[55, 60)	[60, 65)	[65, 70)	[70, 75]
n_i	1	3	10	9	4	2	1
N_i	1	4	14	23	27	29	30



Histograma y polígono de frecuencias



Histograma y polígono de frecuencias acumuladas

EJERCICIOS

11. Indica la representación gráfica (diagrama de sectores, diagrama de barras o histograma) que resulta más adecuada para cada una de las siguientes situaciones.

a) Altura de los árboles de un parque.	b) Deporte que practican tus compañeros.
c) Peso de los alumnos de tu clase.	d) Talla de calzado de los miembros de tu familia.
e) Profesión de vuestros padres.	f) Distancia de vuestras casas al colegio.
g) Número de habitaciones de vuestras casas.	h) Idioma que estudiáis.
12. El ajuste de empleo previsto en los sectores de reconversión más implicados viene dado por la siguiente tabla. Representa gráficamente la distribución mediante un diagrama de sectores.

Sector	Ajuste previsto
Construcción naval	19.660
Siderurgia integral	17.090
Electrodomésticos	11.320
Alimentación	10.091
Textil	7.668
Calzado	3.705

13. Se ha comparado el gasto familiar en España en los años 1970 y 1993 con el fin de observar si han cambiado los hábitos de consumo, y se ha obtenido la siguiente tabla. Representa dos diagramas de sectores y compara ambas distribuciones.

Porcentaje sobre el gasto total	1970	1993
Alimentos, bebidas y tabaco	34'8	20'1
Vestido y calzado	9'8	8'1
Alquileres, calefacción y alumbrado	13'6	13'1
Muebles y ajuar doméstico	8'3	6'5
Gasto sanitario	4'4	4'7
Transportes y comunicaciones	9'0	15'2
Espectáculos, enseñanza y cultura	6'0	6'6
Hostelería y turismo	14'1	25'7

14. Durante el mes de julio, en una determinada ciudad, se han registrado las siguientes temperaturas máximas:

32, 31, 28, 29, 33, 32, 31, 30, 31, 31, 27, 28, 29, 29, 30, 32, 31, 31, 30, 30, 29, 29, 30, 30, 31, 30, 31, 34, 33, 33, 28

- Forma la tabla de frecuencias.
 - Representa gráficamente la distribución mediante un diagrama de barras y un polígono de frecuencias.
 - Representa también el diagrama de barras acumulado y el polígono de frecuencias acumulado.
15. A 1.000 alumnos de Enseñanza Primaria se les ha aplicado un test sobre satisfacción por el colegio. Los resultados obtenidos son los que figuran a continuación. Forma la tabla estadística y representa la situación mediante un diagrama de barras.

Gusta mucho	8 %
Gusta	44 %
No gusta	33 %
No gusta nada	15 %

16. Los siguientes datos son una muestra del número de ejemplares vendidos (en miles) de un periódico en 21 puntos de venta diferentes durante un mes:

5, 7, 4, 10, 9, 6, 9, 7, 8, 11, 6, 8, 10, 11, 6, 10, 10, 8, 11, 6, 5

Construye la tabla de frecuencias necesaria para representar la distribución mediante un diagrama de barras y un polígono de frecuencias. Representa también el diagrama de barras y polígono de frecuencias acumulado.

17. Se han obtenido las pulsaciones de un equipo de atletas después de una carrera. Los datos obtenidos han sido los siguientes:

Pulsaciones	[70, 75)	[75, 80)	[80, 85)	[85, 90)	[90, 95)	[95, 100]
Número de atletas	3	3	7	10	12	8

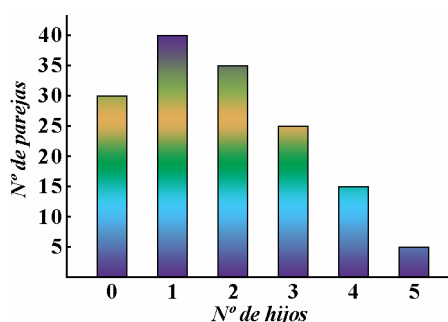
Forma la tabla de frecuencias y representa gráficamente la distribución mediante un histograma y un histograma de frecuencias acumuladas. Dibuja también los respectivos polígonos de frecuencias.

18. La duración en segundos de las llamadas de una empresa tomadas de un recibo son las siguientes:

120 131 142 157 15 27 94 7 21 32 238 210 48 214 156
 57 62 12 49 58 149 210 56 139 24 64 31 23 204 147
 120 131 97 84 61 32 15 58 69 234 13 66 54 7 64
 58 62 169 42 134 114 73 201 93 231 15 56 179 32 124

- Agrupar los datos en 7 clases.
- Forma la tabla de frecuencias completa.
- Representa el histograma y el polígono de frecuencias.
- Representa el histograma acumulado y el polígono de frecuencias acumulado.

19. Un estudio estadístico sobre el número de hijos de las parejas de una ciudad reflejó los resultados que muestra el siguiente diagrama de barras.



- Define la variable estadística en estudio. ¿Cuáles son los individuos de la muestra?
- ¿De qué tipo es dicha variable? ¿Qué valores toma?
- Elabora una tabla donde solamente aparezcan los valores de la variable y sus correspondientes frecuencias absolutas simples.
- Responde a las siguientes preguntas:
 - ¿De qué tamaño es la muestra del estudio?
 - ¿Cuál es el número de hijos más frecuente en las parejas de esa muestra? ¿Qué porcentaje de las parejas tiene ese número de hijos?
 - ¿Qué porcentaje de parejas tiene dos o menos hijos? ¿Y cuatro o más hijos?

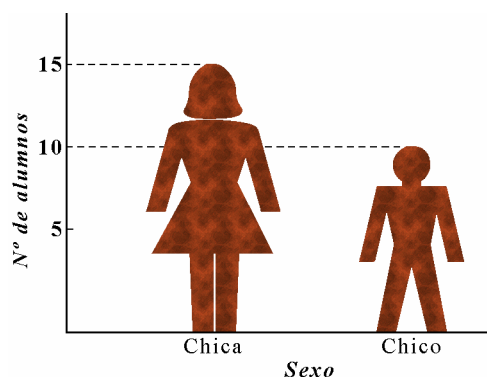
3.4. Otras representaciones gráficas

• Pictogramas

Los pictogramas representan la variable mediante un dibujo cuyo tamaño debe ser proporcional a la frecuencia. Son gráficos poco precisos, aunque fáciles de interpretar a simple vista.

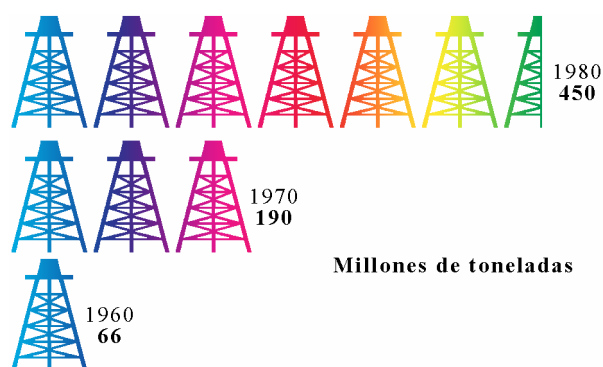
Al margen se muestra un pictograma que nos proporciona el sexo de los alumnos de una clase.

Hemos dibujado una figura alusiva al tema de altura igual a la frecuencia absoluta correspondiente a cada modalidad.



No obstante, estos gráficos son poco fiables, ya que es muy difícil representar datos porcentualmente exactos a través de un dibujo.

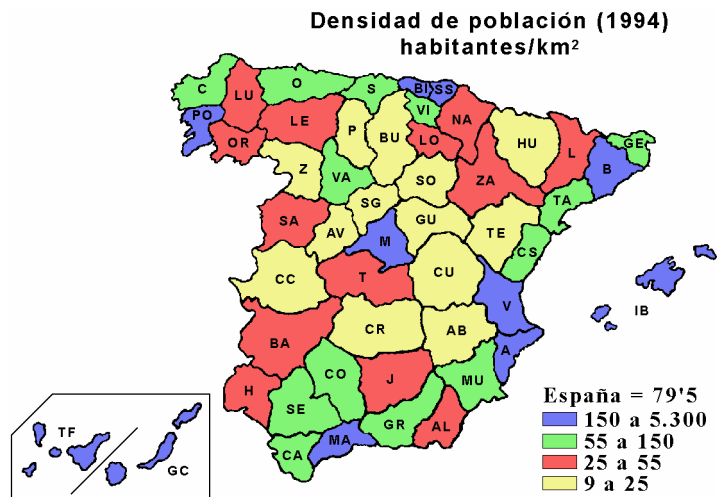
En el siguiente ejemplo puedes ver la producción de petróleo de Arabia Saudí (millones de toneladas).



- **Cartogramas**

Son representaciones gráficas de unidades geográficas diferenciadas por colores, puntos o rayas. Al aplicar un color para representar una variable estadística, su graduación muestra la frecuencia porcentual con la que dicha variable se distribuye.

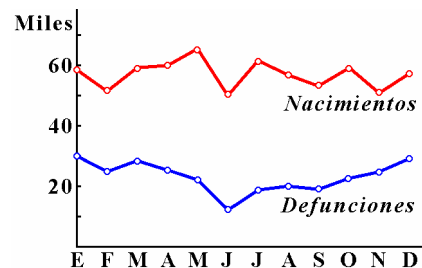
El cartograma del margen describe la densidad de la población española en el año 1994.



- **Series temporales**

Estos gráficos se utilizan para observar la variación de un dato a lo largo de un período de tiempo. A veces deben superponerse dos o más series cronológicas cuyo estudio quiere compararse.

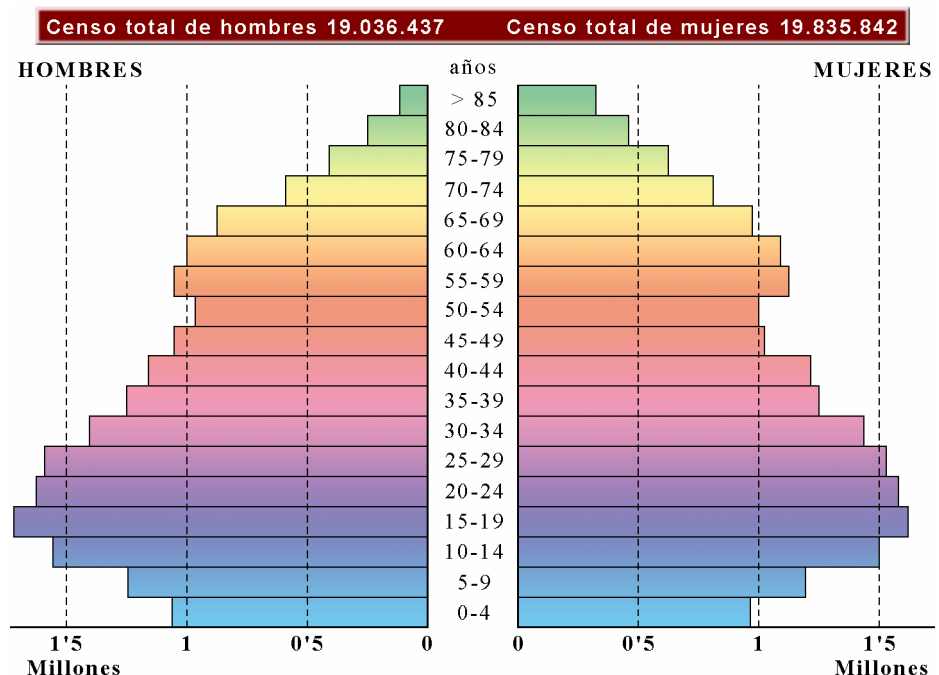
El gráfico siguiente expresa, en miles, los nacimientos y defunciones que se han producido en un determinado año.



- **Pirámides de población**

Son gráficos de población muy útiles para conocer la estructura de la población de un país determinado, según el sexo y la edad. La variable edad se toma sobre el eje vertical y las frecuencias o porcentajes de mujeres (derecha) y hombres (izquierda) se toman sobre el eje horizontal.

Es decir, una pirámide de población consiste en dos histogramas, cuya variable estadística está en el eje vertical y las frecuencias o porcentajes en el eje horizontal.



El estudio detallado de las pirámides de población aporta numerosos datos sobre aspectos sociológicos ligados a dicha población. En la imagen puedes ver la pirámide de población de España en el año 1991.

EJERCICIOS

20. Analiza el gráfico y responde a las preguntas siguientes.



- ¿Cuál es la población? ¿Cuál es su tamaño? ¿Quiénes son los individuos?
- ¿Cuál es la variable estadística? ¿De qué tipo es?
- Haz una tabla estadística en la que figuren los valores o cualidades de la variable estadística, las frecuencias absolutas, las frecuencias relativas y los porcentajes.

21. El tráfico de pasajeros y mercancías del sistema ferroviario ha evolucionado según la siguiente tabla.

Año	Viajeros (millones)	Mercancías (miles de toneladas)
1945	100	25.992
1950	107	29.758
1955	117	34.963
1960	109	34.302
1965	174	30.028
1970	164	30.838

Representa dos series temporales, una para los viajeros y otra para las mercancías.

22. Observa el cartograma del epígrafe 3.4 y responde a las siguientes preguntas.

- ¿Qué es densidad de población?
- ¿Qué provincias españolas presentan una densidad de población superior a 55 habitantes/km²?
- Busca en un libro apropiado la extensión de tu provincia y realiza, a partir del gráfico, una valoración estimativa de su número de habitantes en el año 1994.

23. Analiza la pirámide de población que aparece en el epígrafe 3.4 y responde a las siguientes cuestiones.

- ¿Cuál es el grupo de edad más numeroso en 1991 si se consideran conjuntamente los dos sexos? ¿Y si se consideran por separado?
- ¿Nacen más niños que niñas, o al revés? ¿En qué grupos de edad hay más varones que mujeres? ¿Y en cuáles hay más mujeres que hombres?
- Existe un *valle*, es decir, un valor mínimo relativo respecto de los valores anteriores y posteriores, tanto entre las mujeres como en los varones, para los grupos de edad entre los 45 y 55 años. ¿Encuentras alguna explicación a ese fenómeno? ¿Qué ocurrió en España hace, aproximadamente, 50 años (recuerda que la pirámide es del año 1991)?