

T1. INTRODUCCIÓN

La **población** es el conjunto de individuos que se quiere estudiar, en el que los **individuos** pueden ser tanto objetos (yogures, ruedas, etc.) como personas, niños, programadores, etc.

Una **muestra** es un subconjunto cualquiera de la población.

Una **variable** es una característica de los individuos objeto de nuestro estudio.

Posteriormente, recogeremos la información para **obtener una serie de observaciones**. En esencia, habrá que explorar los datos para encontrar una descripción de la población lo más detallada posible.

- a) **Averiguar la distribución** de la variable (qué valores toma y cómo los toma).
- b) **Calcular algunos resúmenes numéricos** que ayuden a entender cuál es el centro de los valores y cómo se distribuyen dichos valores en torno a ese centro.
- c) **Dibujar gráficos** que ayuden a visualizar los puntos anteriores.

Finalmente, con todo esto **podemos llegar a conclusiones** sobre los datos, hacer comparaciones sobre dos conjuntos diferentes de individuos, plantearnos preguntas más complejas y ver qué preguntas nos podemos hacer con respecto a las características globales de la población.

1.1 TIPOS DE VARIABLES

Variables cualitativas: son aquellas que se expresan como categorías o características de los individuos.

Variables cuantitativas: se expresan de forma numérica. Existen dos tipos:

- a) **Cuantitativas discretas:** sólo toman valores enteros (sin decimales). Ej: número de hijos, personas a favor en una encuesta...
- b) **Cuantitativas continuas:** pueden tomar cualquier valor, incluso con decimales, en un intervalo. Ej: el peso de una persona, estatura, PIB...



1.2 VARIABLES CUALITATIVAS Y VARIABLES NUMÉRICAS DISCRETAS QUE TOMAN UN NÚMERO PEQUEÑO DE VALORES DIFERENTES

Una variable puede tomar k valores posibles, en el que hay que averiguar las características siguientes:

- 1) El **número total de individuos** (N) de los que se disponen datos. $n_1 + n_2 + n_3 + \dots + n_k = N$
- 2) La **frecuencia absoluta** (n_i): el número de individuos para los cuales la variable toma este valor. $f_i = \frac{n_i}{N}$
- 3) La frecuencia absoluta acumulada $N_j = n_1 + n_2 + n_3 + \dots + n_j$
- 4) La **frecuencia relativa** (f_i): la proporción de individuos en los que la variable toma este valor. Se expresará en porcentaje. $f_1 + f_2 + f_3 + \dots + f_k = 1$
- 5) La frecuencia relativa acumulada $F_j = f_1 + f_2 + f_3 + \dots + f_j$

EJEMPLO 1

Ejemplo del tipo de ordenadores

Hemos hecho una encuesta entre estudiantes de la UOC en la que se pregunta el tipo de ordenador que utilizan habitualmente en casa para conectarse al campus virtual. A continuación damos las respuestas abreviadas:

PC, PC, MAC, PCP, PCP, PCP, MAC, O, O, PC
PC, PC, PC, MACP, O, MAC, O, PCP, PCP, PC

	Frecuencia	Frecuencia relativa
Tipo de ordenador	n_i	f_i
1 = PC		
2 = PCP		
3 = MAC		
4 = MACP		
5 = Otros		
Totales		



Con este tendremos la **distribución de frecuencias de la variable**, que es el conjunto de datos que adopta la variable y la frecuencia con que los adopta.

Para representar gráficamente estas variables usaremos el **diagrama de barras** y el **diagrama de sectores**.

DIAGRAMA DE SECTORES

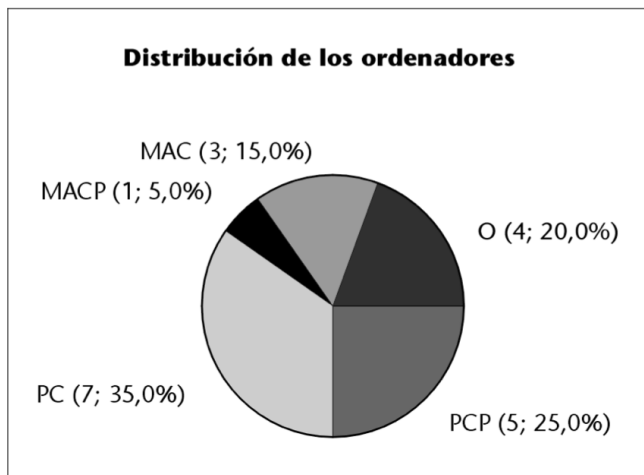
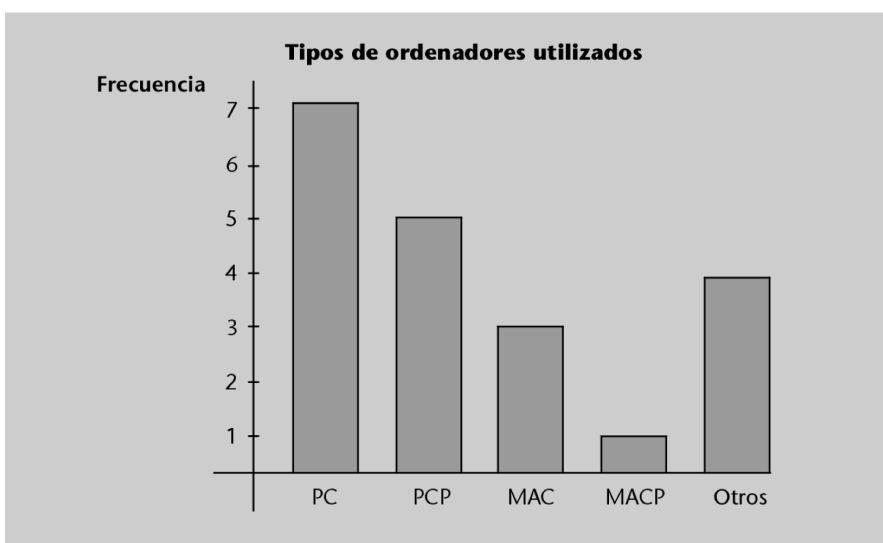


DIAGRAMA DE BARRAS



EJEMPLO 2

Ejemplo de las notas de estadística

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

	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Nota	n_i	f_i	N_i	F_i
0	-		-	
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
Totales				

1.3 VARIABLES NUMÉRICAS

Lo primero que hay que averiguar es:

- 1) **Cuántos individuos** aparecen en el estudio (N).
- 2) El **máximo** (máx) y el **mínimo** (min).
- 3) El **rango de la variable**, es decir, la diferencia entre el valor máximo y el mínimo.

Para representar gráficamente estas variables utilizaremos el **diagrama de puntos** y el **diagrama de tallo y hojas**.

EJEMPLO 3

Ejemplo de los días de hospitalización

Se ha elaborado un estudio sobre los días de hospitalización de un grupo de pacientes sometidos a un mismo tratamiento, estudio del que se han obtenido los datos siguientes: 15, 15, 3, 46, 623, 126, 64 días.

DIAGRAMA DE PUNTOS

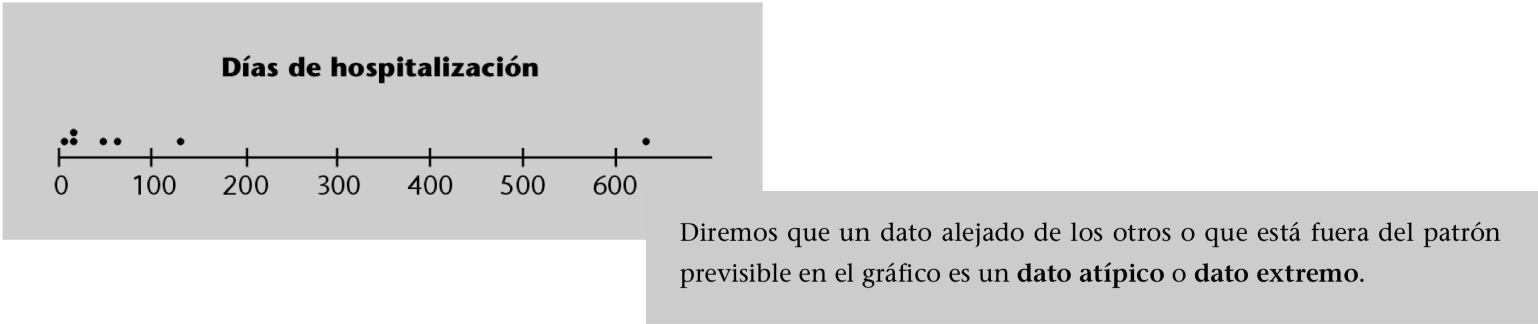


DIAGRAMA DE TALLO Y HOJAS

EJEMPLO 4

Ejemplo del jugador de ajedrez

Consideremos el número de minutos que un jugador de ajedrez de gran nivel ha necesitado para ganar al programa de ordenador *Deep Yellow* en quince partidas consecutivas:

54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22

2	25
3	45
4	1166679
5	449
6	0

2	2
2	5
3	4
3	5
4	11
4	66679
5	44
5	9
6	0

1.4 VARIABLES CONTINUAS E HISTOGRAMA

Nos encontramos ante una variable con muchos valores diferentes. Son variables que conducen a la necesidad de **definir una distribución de frecuencias**. Para obtener esta expresión, haremos los siguiente:

- a) **Agrupamos las observaciones en intervalos** llamados clases. Los intervalos deben ser adyacentes (sin dejar espacio entre ellos) y deben cubrir todo el rango, desde el mínimo hasta el máximo.
- b) **Calculamos el punto medio de cada intervalo**, llamado marca de clase.
- c) Calculamos la **frecuencia absoluta** de cada una y su **frecuencia relativa**.
- d) Si es preciso, también se puede calcular la **frecuencia absoluta acumulada** y la **frecuencia relativa acumulada**.

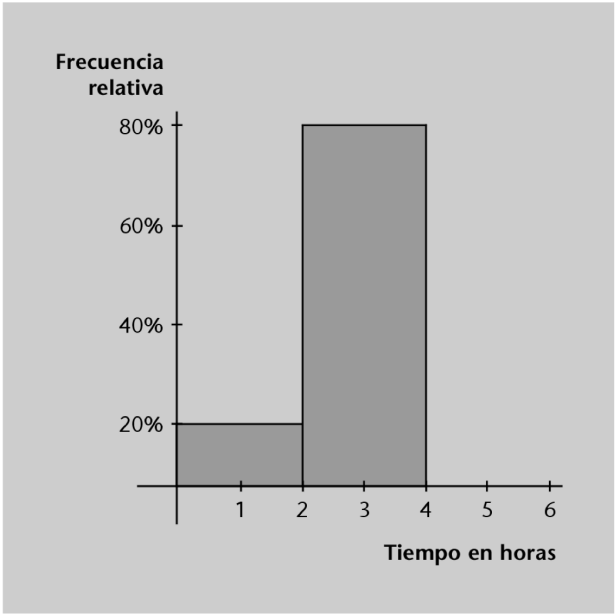
Para representar gráficamente dichas variables se utilizan los **histogramas**.

EJEMPLO 5

Ejemplo de elaboración de un diagrama de frecuencias relativas

Se ha llevado a cabo un estudio sobre el número de horas diarias que los estudiantes de la UOC ven la televisión y se ha constatado que el 20% la ve menos de 2 horas y que el 80% restante la ve 2 horas o más, pero menos de 4 horas. Representamos los datos en forma de tabla:

Clases	f_i
[0,2)	20%
[2,4)	80%

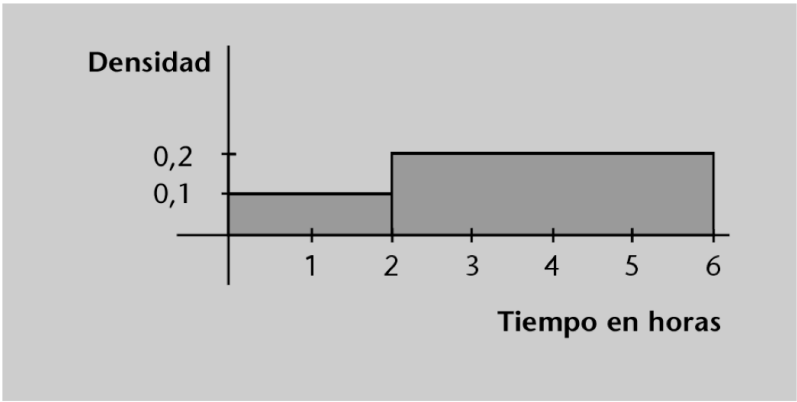


EJEMPLO 6

Ejemplo de elaboración de un diagrama de densidad

Otro estudio sobre el número de horas diarias que los estudiantes de la UOC ven la televisión ha constatado que el 20% de los estudiantes la ve menos de 2 horas y que el 80% restante la ve 2 horas o más, pero menos de 6. Cuando representamos los datos en una tabla, obtenemos lo siguiente:

Clases	f_i
[0,2)	20%
[2,6)	80%



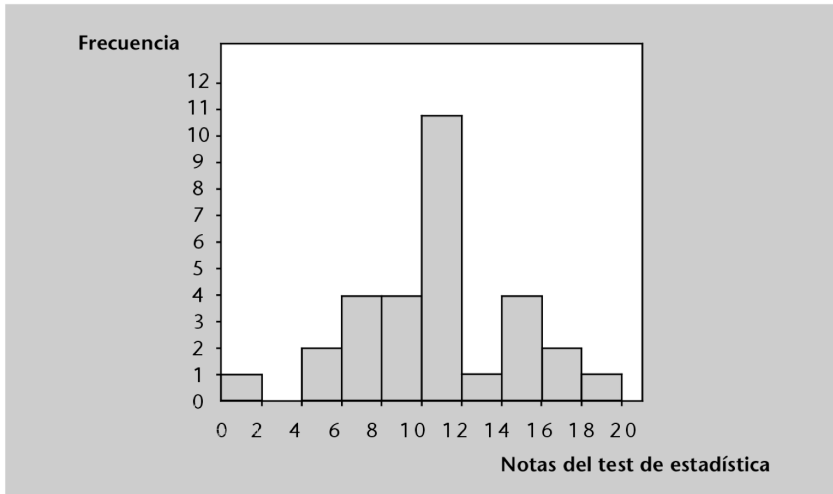
EJEMPLO 7

Los resultados de un test de estadística

A continuación mostramos las notas de un test de estadística pasado a treinta estudiantes de un curso de la UOC durante el semestre de otoño. La puntuación posible del test va de 0 a 20 puntos.

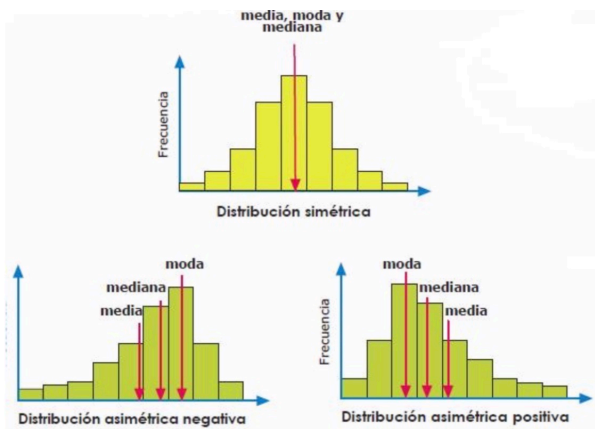
11,5	7	5,25	19	8,5	11	10	6	15,5	11,75
9,75	15	16	16	7,5	8	11,5	0	10	11
13	10,25	10	8	14	6	14,5	10	5,6	10

		Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
Intervalo	Marca de clase	n_i	f_i	N_i	F^i
[0,2)					
[2,4)					
[4,6)					
[6,8)					
[8,10)					
[10,12)					
[12,14)					
[14,16)					
[16,18)					
[18,20)					
Totales					



1.5 INTERPRETACIÓN HISTOGRAMA

La simetría: el histograma será simétrico si es posible trazar un eje vertical de manera que **la parte** que hay a la **derecha del eje sea** aproximadamente **igual a la** imagen reflejada en un espejo de la **parte izquierda**.



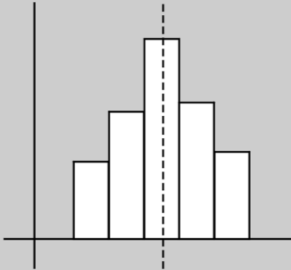
Los picos: corresponden a **intervalos en los que se tienden a concentrar los valores** de la variable (son intervalos con frecuencias grandes). Si sólo hay un pico destacado, diremos que es un caso unimodal; si tenemos dos picos de altura similar, será un caso bimodal.

Las colas: si no hay simetría, es posible que uno de los **lados** del histograma se extienda **mucho más lejos que el otro**. En este caso diremos que hay una cola en el lado más extenso (llamado caso de cola larga) o que hay asimetría hacia este mismo lado.

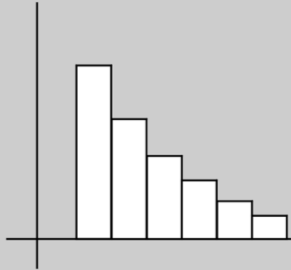
Hay que **detectar los datos extremos**, las clases vacías y si estas clases vacías **separan a la población** en grupos.

Elementos de interpretación de los histogramas según la forma

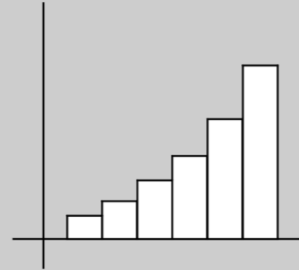
Simetría



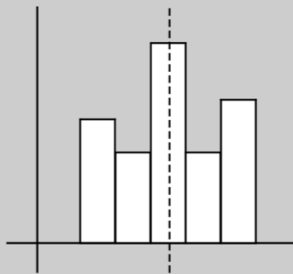
Asimetría a la derecha;
cola a la derecha



Asimetría a la izquierda;
cola a la izquierda



Unimodal



Bimodal

