


14

Estadística



Los censos y recuentos estadísticos se hallan presentes en todas las civilizaciones desde tiempos antiquísimos, pero se limitaban a la recogida de datos y, a lo sumo, a su exposición clara y ordenada.



La estadística como ciencia comienza a des-puntar en el siglo XVII con los estudios demográficos del inglés **John Graunt**. En ellos, los datos se analizaban para obtener conclusiones bien fundamentadas. 

Existen papiros egipcios de hace más de 5000 años donde hay constancia de censos de población y de bienes. Tal era su dedicación a estos asuntos que concibieron una divinidad llamada *Safnkit*, “Señora de los libros”.

También los babilonios guardaban en tablillas de arcilla los recuentos estadísticos que realizaban hasta el punto de que en el siglo VIII a.C. se construyó una biblioteca donde se recopilaban estos documentos.

En distintos pasajes de la Biblia se recogen censos hechos por los judíos. Especialmente en el libro Números, del Pentateuco, donde se describe con detalle el censo realizado por Moisés a la salida de Egipto en el siglo XIV a.C.

Nombre y apellidos: Fecha:

BOC → bocadillo o sándwich
 FRU → pieza de fruta o frutos secos
 BOL → bollo dulce o chuchería
 OTRO → cualquier otra comida
 NADA → no toma nada

BOC	
FRU	
BOL	
OTRO	
NADA	

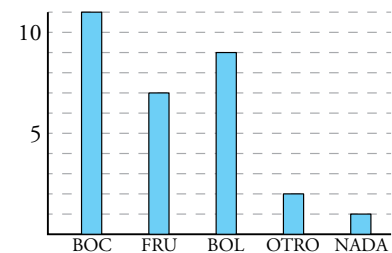
Los dueños de un local próximo a un centro de estudios, con el fin de montar un puesto de bocadillos y otros alimentos, han preguntado a 30 estudiantes del centro qué suelen comer a mitad de mañana. Los resultados son estos:

OTRO, BOC, BOC, BOL, FRU
 FRU, BOL, FRU, BOC, BOC
 FRU, FRU, BOC, BOL, BOC

NADA, BOC, BOL, BOL, BOL
 BOC, BOL, FRU, BOL, BOC
 BOC, OTRO, FRU, BOC, BOL

Con estos datos, confeccionamos una tabla de frecuencias y su gráfica.

ALMUERZO	FRECUENCIA
BOC	11
FRU	7
BOL	9
OTRO	2
NADA	1



En tablas y gráficas estadísticas se recoge la información que se ha ido obteniendo durante un proceso estadístico. Recordemos los pasos que se siguen:

1. Toma de decisiones sobre lo que se quiere estudiar y para qué se necesita.

Los estudios estadísticos en general se realizan por algún motivo: aumentar las ventas de un producto, ganar en competitividad, aumentar la audiencia, mejorar los servicios públicos... En el ejemplo anterior, se quiere estudiar la viabilidad que tiene un puesto de bocadillos junto a un centro de estudios.

2. Enumeración adecuada de los valores que puede tomar la variable.

En este caso, la preferencia mostrada por un estudiante cualquiera siempre encaja en una de las opciones que se le dan (BOC, FRU, BOL, OTRO, NADA).

3. Recogida de datos.

Los datos pueden venir a partir de una encuesta, mediante observación o experimentación, a través de anuarios o páginas oficiales... En nuestro ejemplo, se realiza una encuesta escogiendo aleatoriamente 30 estudiantes del centro.

4. Tabulación y representación gráfica.

Confección de una tabla a partir de los datos recogidos. Habitualmente, puede ser ventajoso acompañarla de una gráfica adecuada.

5. Análisis de los resultados. Conclusiones.

En el ejemplo anterior, los dueños del local posiblemente, a partir de los resultados, ven claro que sí les conviene montar el negocio.

Piensa y practica

1. Los dueños de un cine han preguntado a 30 espectadores por lo que suelen comer durante la película. Las opciones son palomitas (PA); chucherías (CH); nachos o patatas (NCH); otras cosas (OT) y nada (ND).

Se han obtenido estos datos:

PA, PA, CH, NCH, ND
 CH, PA, CH, PA, PA
 PA, ND, PA, NCH, PA
 PA, ND, CH, OT, PA
 NCH, ND, PA, CH, ND
 OT, ND, CH, CH, PA

Construye una tabla de frecuencias y representa los datos en un diagrama de barras.



1. En una clase de 30 estudiantes, la profesora de Lengua ha recomendado 5 libros para leer en verano. A la vuelta de vacaciones, se les ha preguntado cuántos de esos libros han leído. Estos son los resultados:

0 2 2 3 1 1 2 2 1 5
 2 1 2 4 1 1 2 1 5 2
 2 3 4 2 1 2 4 2 1 3

a) Realizar un recuento y confeccionar una tabla de frecuencias.

b) Realizar una representación adecuada de los resultados.

a) Hacemos un recuento y, a partir de él, confeccionamos una tabla de frecuencias donde incluimos la frecuencia relativa y la porcentual:

N.º DE LIBROS LEÍDOS	FRECUENCIA	FRECUENCIA RELATIVA	FRECUENCIA PORCENTUAL
0	1	$1/30 = 0,03$	3 %
1	9	$9/30 = 0,30$	30 %
2	12	$12/30 = 0,40$	40 %
3	3	$3/30 = 0,10$	10 %
4	3	$3/30 = 0,10$	10 %
5	2	$2/30 = 0,07$	7 %
TOTAL	30	1,00	100 %

Mirando las frecuencias porcentuales, podemos decir que el 67 % de los estudiantes ($40 + 10 + 10 + 7 = 67$) ha leído 2 o más libros y que únicamente el 3 % no ha leído ninguno.

b) Diagrama de barras.

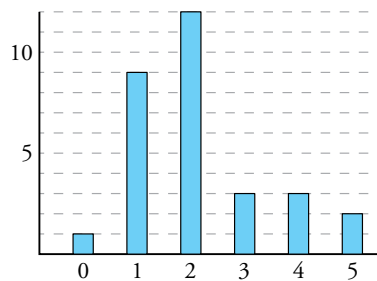
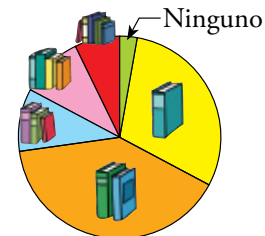


Diagrama de sectores.



2. En la clase de Educación Física se ha pedido a los estudiantes que lancen el balón medicinal lo más lejos que puedan. Estas son las distancias (en cm):

387 355 432 483
 348 497 410 427
 365 392 455 373
 383 413 459 392

a) Confeccionar una tabla de frecuencias agrupando los datos en estos intervalos:

340 - 380 380 - 420
 420 - 460 460 - 500

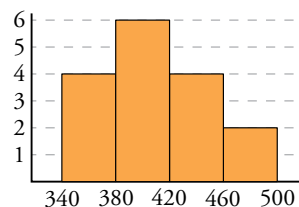
b) Dibujar los resultados en un histograma.

a) Repartiremos los 16 valores en cada uno de los cuatro tramos designados. Por ejemplo, el 387 va al segundo tramo; el 355, al primero; el 432, al tercero... así sucesivamente. Cada vez que ponemos un valor en su correspondiente intervalo, anotamos una rayita. De este modo construimos la siguiente tabla:

INTERVALO	VALOR DEL INTERVALO (MARCA DE CLASE)	RECUESTO	FRECUENCIA
340 - 380	360		4
380 - 420	400		6
420 - 460	440		4
460 - 500	480		2

¿Qué es una marca de clase? Veámoslo: en el primer intervalo hay cuatro individuos cuyos valores son 355, 348, 365 y 373, pero a partir del momento en que construimos la tabla, a los cuatro se le asigna el valor central del intervalo, 360, que se llama marca de clase; y lo mismo diríamos de los demás intervalos.

b) Dibujamos el histograma:



Los parámetros estadísticos son valores que se obtienen a partir de la distribución y que resumen alguna de sus características globales.

La media, la mediana y la moda se llaman **parámetros de centralización** porque son valores alrededor de los cuales se distribuyen los datos.

Recordemos en qué consisten y cómo se calculan.

Ejemplo

A 10 estudiantes les han preguntado por el número de tíos y tías que tienen. Las respuestas han sido las siguientes: 3, 5, 4, 3, 5, 6, 8, 2, 1, 3.

$$\text{Su media es } \bar{x} = \frac{3+5+4+3+5+6+8+2+1+3}{10} = \frac{40}{10} = 4$$

Para hallar la **mediana**, se ordenan los datos de menor a mayor. La mediana es el dato que está en el lugar central. En este caso, 1, 2, 3, 3, **3, 4**, 5, 5, 6, 8, como hay un número par de datos, dos de ellos (3 y 4) ocupan el lugar central. La mediana es el promedio de ellos, $Me = 3,5$.

Su **moda** es $Mo = 3$, porque es el dato más frecuente (está 3 veces).

- La **media** de varias cantidades es la suma de todas ellas dividida por el número de las que hay.
- Se llama **mediana** de un conjunto de datos numéricos al que, colocándolos en orden, ocupa el lugar central. Si hay un número par de datos, se asigna la mediana al valor intermedio entre los dos centrales.
- La **moda** es el dato con mayor frecuencia.

Recuerda

Las **variables cuantitativas** solo toman valores numéricos y las **cualitativas** no toman valores numéricos.

La media y la mediana solo pueden hallarse para variables cuantitativas. La moda puede asignarse a cualquier tipo de variable.

Ejercicio resuelto

Hallar \bar{x} , Me y Mo en estas distribuciones:

a) *N.º de mascotas que tienen 13 estudiantes de una clase.*

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

b) *Estación del año en la que nacieron 10 personas:*

P, V, V, O, P, I, I, V, O, V

a) Media: $\bar{x} = \frac{3+2+0+0+3+2+0+2+0+1+4+6+3}{13} = \frac{26}{13} = 2$

Mediana: Se ordenan los datos: $\underbrace{0, 0, 0, 0, 1, 2, 2, 2, 3, 3, 3, 4, 6}_{13} \rightarrow Me = 2$

Moda: $Mo = 0$ (el dato que está más veces).

b) Es una variable cualitativa y no se le puede asignar ni media ni mediana. Su moda es $Mo = V$ (pues el número de personas nacidas en verano, 4, es mayor que en las demás estaciones).

Piensa y practica

1. Halla \bar{x} , Me y Mo de cada una de las siguientes distribuciones:

a) Grupo sanguíneo de 15 personas:

A, A, B, AB, AB, A, A, B, A, 0, AB, A, A, B, AB

b) Edades de varios estudiantes:

12, 15, 12, 16, 10, 11, 12, 10, 11, 12, 9, 9, 10, 8

c) Número de asignaturas suspensas en la evaluación:

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

Los parámetros de centralización dan una visión muy parcial de la distribución. Deben ser complementados con otros parámetros que informan sobre el grado de dispersión de los datos. Veamos algunos de ellos:

Recorrido o rango

El **recorrido** o **rango** de una distribución es la diferencia entre los valores extremos.

$$\text{RECORRIDO} = \text{valor mayor} - \text{valor menor}$$

En las distribuciones de la página anterior, sus recorridos son:

$$\text{RECORRIDO DE } \textcircled{\text{I}} = 10 - 1 = 9$$

$$\text{RECORRIDO DE } \textcircled{\text{II}} = 30 - 1 = 29$$

Desviación media: un parámetro de dispersión ligado a la media

La **desviación media**, DM, de una distribución es un parámetro asociado a su media; es el promedio de las distancias a la media de los valores de todos los individuos.

Por ejemplo, consideremos la distribución 5, 8, 10, 11, 15, 17 cuya media es 11. Observa que las distancias de los datos a la media son positivas.

DATOS	5	8	10	11	15	17
DISTANCIA A LA MEDIA	6	3	1	0	4	6

De 5 a $\bar{x} = 11$ es $11 - 5 = 6$.

De 15 a $\bar{x} = 11$ es $15 - 11 = 4$.

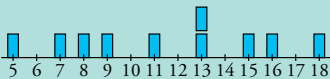
El promedio de las distancias a la media se calcula así:

$$\text{DM} = \frac{6 + 3 + 1 + 0 + 4 + 6}{6} = \frac{20}{6} = 3,33$$

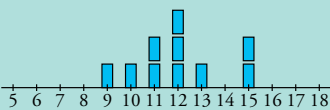
Ejercicio resuelto

Hallar la desviación media de las siguientes distribuciones:

$\textcircled{\text{III}}$ 5, 7, 8, 9, 11, 13, 13, 15, 16, 18



$\textcircled{\text{IV}}$ 9, 10, 11, 11, 12, 12, 12, 13, 15, 15



$\textcircled{\text{III}}$ Su media es $\bar{x} = 11,5$.

DATOS	5	7	8	9	11	13	13	15	16	18	
DISTANCIA A 11,5	6,5	4,5	3,5	2,5	0,5	1,5	1,5	3,5	4,5	6,5	SUMA → 35

$$\text{Desviación media: DM} = \frac{\text{suma de las distancias a } \bar{x}}{10} = \frac{35}{10} = 3,5$$

$\textcircled{\text{IV}}$ Su media es $\bar{x} = 12$.

DATOS	9	10	11	11	12	12	12	13	15	15	
DISTANCIA A 12	3	2	1	1	0	0	0	1	3	3	SUMA → 14

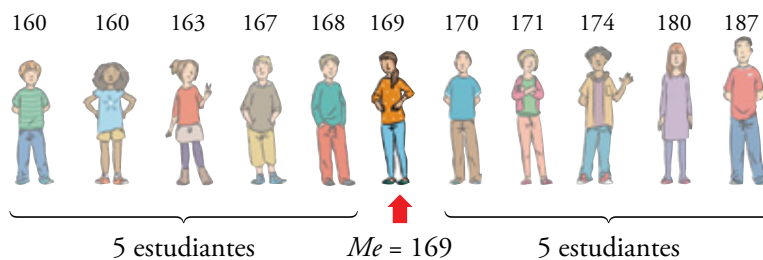
$$\text{Desviación media: DM} = \frac{14}{10} = 1,4$$

Los datos de $\textcircled{\text{III}}$ (DM = 3,5) están más dispersos que los de $\textcircled{\text{IV}}$ (DM = 1,4).

Piensa y practica

- Halla el recorrido, la DM y la desviación típica en las distribuciones A, B y C de la actividad 2 de la página anterior.

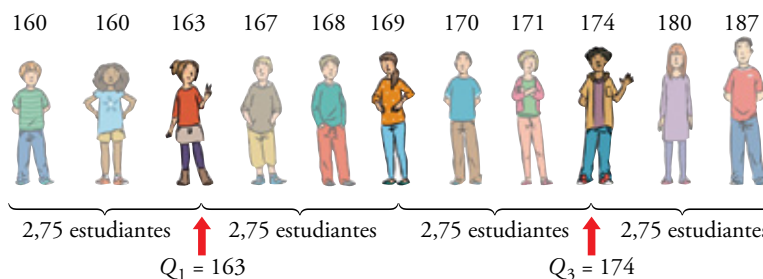
Los 11 componentes de un grupo se han colocado en fila ordenados según sus estaturas. Como ya sabes, la mediana es la medida del que ocupa el lugar central; en este caso, el sexto lugar.



¿Qué tiene que ver el **sexto lugar** con el número de individuos, 11? La mitad de 11 es 5,5, y por tanto, la mediana correspondería a la medida de la persona que ocupa el lugar 5,5. Como las personas no se pueden partir, será la persona que va después de la quinta, es decir, la sexta.

Si en lugar de dividir este grupo en 2 partes lo hacemos en 4 partes iguales, ¿dónde quedarían las particiones? Como hay 11 personas, la cuarta parte es $11 : 4 = 2,75$.

Q_1 deja a su izquierda “2,75 personas”, Q_3 deja a su izquierda “8,25 personas” (3 · 2,75 = 8,25), por tanto, lo situamos en el valor correspondiente a la tercera persona, es decir, en 163. Q_3 deja a su izquierda “8,25 personas” (3 · 2,75 = 8,25), por tanto, lo situamos en el valor correspondiente a la novena persona, es decir, en 174.



Recorrido intercuartílico

El **recorrido intercuartílico** es la distancia entre los cuartiles primero y tercero:

$$\text{Recorrido intercuartílico} = Q_3 - Q_1$$

La mediana juega el papel de parámetro de centralización. El recorrido intercuartílico es un parámetro de dispersión asociado a ella.

Estos dos nuevos puntos se llaman **cuartiles**: Q_1 es el primer cuartil y Q_3 es el tercer cuartil. El segundo es la mediana: $Q_2 = Me$. Los cuartiles dividen la población en 4 partes iguales. La mediana y los cuartiles son **medidas de posición**.

Ejercicio resuelto

Calcular la mediana y los cuartiles de la siguiente distribución de notas de 13 estudiantes:

1, 1, 2, 3, 4, 4, 5, 6, 7, 7, 7, 8, 10

Es claro que la mediana es $Me = 5$: 1, 1, 2, 3, 4, 4, **5**, 6, 7, 7, 7, 8, 10
Como hay 13 estudiantes, la cuarta parte es 3,25.

El primer cuartil tiene 3,25 estudiantes por debajo de él, por tanto corresponde a la nota del cuarto estudiante, $Q_1 = 3$; el tercer cuartil tiene 9,75 estudiantes por debajo de él, por lo que corresponde a la nota del décimo, $Q_3 = 7$.

Piensa y practica

1. Calcula la mediana y los cuartiles de: 13, 12, 15, 19, 12, 12, 13, 14, 15, 14, 13, 18, 17, 9, 8.

En la web Actividades guiadas para practicar los parámetros de posición.

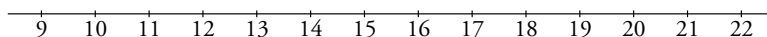
Diagrama de caja

Una forma de representar los parámetros de posición, mediana y cuartiles es lo que se denomina **diagrama de caja**, también conocido por diagrama de caja y bigotes.

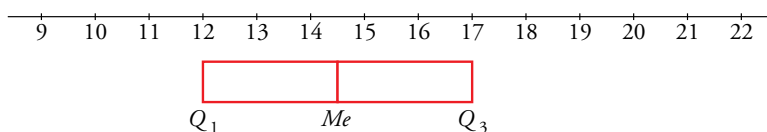
Ejemplo

En el banquete de una boda han juntado a todos los primos en una mesa. Sus edades ordenadas son las siguientes: 9, 9, 10, 14, 14, 14, 15, 16, 17, 17, 19, 22. Observamos que $Me = 14,5$, $Q_1 = 12$ y $Q_3 = 17$. Para representar estos datos en un diagrama de caja, damos estos pasos:

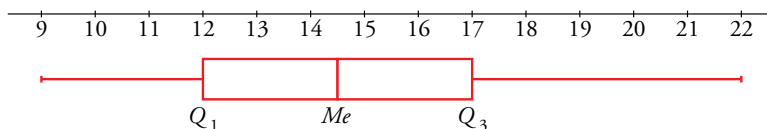
1. Trazamos una recta numérica que comprenda todos los valores que aparecen, del 9 al 22.



2. Dibujamos una caja que vaya de $Q_1 = 12$ a $Q_3 = 17$. Dividimos la caja en dos partes por el valor $Me = 14,5$.



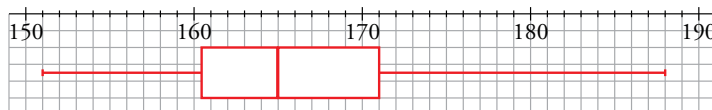
3. Desde el lateral izquierdo de la caja trazamos una recta horizontal hacia la izquierda que llegue hasta el menor valor, 9, y desde el derecho trazamos otra recta horizontal hacia la derecha que llegue hasta el mayor valor, 22.



Ejercicio resuelto

Las estaturas de un grupo de jóvenes se representan mediante el diagrama de caja de la derecha.

Interpretar e indicar cuáles son los parámetros de posición.



A la vista del diagrama podemos decir que $Q_1 = 160,5$, $Me = 165$ y $Q_3 = 171$. Además, el más bajo mide 151 cm, y el más alto, 188 cm.

Por tanto, un 25% de los jóvenes mide entre 151 cm y 160,5 cm; otro 25%, entre 160,5 y 165 cm; otro 25%, entre 165 cm y 171 cm, y el último 25% (los más altos) miden entre 171 cm y 188 cm.

Piensa y practica

2. Representa mediante un diagrama de caja y bigotes las siguientes calificaciones de 35 individuos:

0 3 3 3 4 4 4 4 4 5 5 5 6 6 6
6 7 7 7 7 7 7 7 8 8 8 8 8 8 8
9 9 9 10 10

3. El siguiente diagrama de caja representa la distribución de las notas de una clase de 30 estudiantes.



Interpétalo e indica los parámetros de posición.

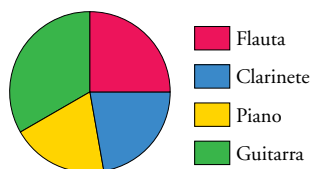
Ejercicios y problemas

Tablas y gráficas estadísticas

1. Este diagrama de sectores representa los 24 estudiantes de una clase de 2.º de ESO, según se queden o no a comer en el colegio:



- a) ¿Qué fracción de los estudiantes se queda a comer?
 b) ¿Qué porcentaje no se queda nunca?
 c) ¿Qué tanto por ciento se queda a veces?
2. En clase de Música, cada alumno tiene que elegir un instrumento entre cuatro posibles. La distribución de los alumnos según el instrumento elegido viene dada por este diagrama de sectores:



- a) ¿Cuál es el instrumento más elegido? ¿Y el menos?
 b) ¿Hay algún instrumento que lo hayan elegido exactamente el 25% de la clase?
 c) Sabiendo que los alumnos que han elegido cada instrumento son 7, 8, 9 y 12, ¿qué número corresponde a cada uno de ellos?
3. Los miembros de la sección de infantiles de un club de atletismo han anotado sus pesos en un cuaderno:

48, 52, 36, 53, 44 39, 40, 53, 55, 47
 54, 49, 57, 44, 54 41, 54, 48, 48, 52
 51, 47, 58, 53, 50 46, 45, 38, 39, 49

- a) Confecciona una tabla de frecuencias en la que los datos se repartan en los siguientes intervalos:

35,5 - 42,5; 42,5 - 49,5; 49,5 - 56,5; 56,5 - 63,5

Recuerda que la marca de clase de cada intervalo es su valor central, es decir, la suma de sus extremos dividida entre dos.

- b) Representa los resultados en un histograma.

4. En un festival de música han tocado cinco grupos. A la salida del recinto, se ha pasado una pequeña encuesta en la que se pregunta el número de conciertos a los que se ha asistido. Los resultados vienen dados en la siguiente tabla:

N.º DE CONCIERTOS	N.º DE PERSONAS
1	63
2	185
3	167
4	32
5	13

- a) ¿A cuánta gente le han preguntado?
 b) Copia la tabla en tu cuaderno y añade la columna de frecuencias porcentuales.
 c) ¿Qué porcentaje de gente ha estado en menos de 4 conciertos? ¿Y en más de 3 conciertos?
 d) ¿Qué porcentaje de gente ha asistido al menos a 2 conciertos?

5. Una pequeña pensión ha anotado cada día de un determinado mes el número de habitaciones que tienen ocupadas. Estos son los resultados:

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

- a) Confecciona una tabla de frecuencias absolutas, relativas y porcentuales.
 b) Representa los datos en un diagrama de barras.
 c) Dibuja el correspondiente diagrama de sectores.

Parámetros estadísticos

6. Halla la media, la mediana, la moda, el recorrido y la desviación media de las siguientes distribuciones:

- a) 1, 1, 1, 3, 4, 5, 7, 8, 10, 10
 b) 1, 1, 2, 3, 4, 5, 6, 6, 8, 9, 10
 c) 4, 2, 2, 4, 3, 10, 3, 4, 4, 5, 2, 5

7. Halla la media, la mediana, el recorrido, la desviación media y los cuartiles de las siguientes distribuciones:

- a) 1, 3, 8, 9, 4, 1, 1, 7, 10, 10
 b) 1, 3, 5, 4, 2, 8, 9, 6, 10, 6

Ejercicios y problemas

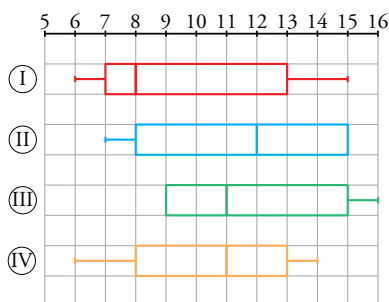
8. El número de errores que tuvieron en un test un grupo de estudiantes fueron:

1, 1, 2, 2, 4, 5, 5, 8, 8, 9

Halla la mediana y los cuartiles primero y tercero, y haz un diagrama de caja con esos datos.

9. Los tiempos que un grupo de personas han empleado en hacer un test se distribuyen entre 0 y 50 minutos. Construye el diagrama de caja sabiendo que $Q_1 = 23$, $Me = 34$ y $Q_3 = 39$.

10. Indica la mediana y los cuartiles en cada uno de los siguientes diagramas de caja:



11. Calcula los cuartiles de las distribuciones del ejercicio 6 y representa cada una de ellas mediante un diagrama de caja.

12. Lanzamos un dado 40 veces. Estos son los resultados:

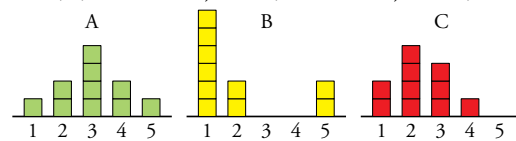
3	5	1	2	5	5	3	4	6	2
4	3	6	4	1	6	4	2	6	1
4	3	5	6	2	1	5	6	6	2
4	2	3	2	6	5	4	1	6	1

Calcula la media y la moda de la distribución.

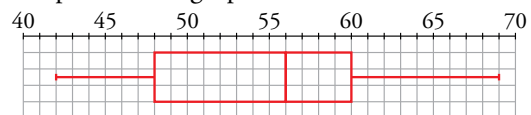
Resuelve problemas

13. Asocia cada par de parámetros con su correspondiente gráfica.

a) $\bar{x} = 2,3$; $Me = 2$ b) $\bar{x} = 3$; $Me = 3$ c) $\bar{x} = 2$; $Me = 1$



14. Este diagrama de caja representa la distribución de los pesos de un grupo de estudiantes de una clase.



Completa estas frases observando el diagrama:

- a) El 50% de los estudiantes de esta clase pesa ... o menos.
 b) El 25% de los estudiantes pesa ... o menos.
 c) El 25% de los estudiantes pesa ... o más.
 d) El 50% de los pesos centrales varía entre ... y ...
 e) El 75% de los estudiantes pesa ... o más.
 f) El ... de los estudiantes de esta clase pesa 60 kg o menos.
15. La edad media de un grupo de diez personas es 13 años.
- a) ¿Es posible que al incorporarse una persona al grupo haga que la edad media sea 10 años? Explica por qué.
 b) Bruno, el profesor, se ha incorporado al grupo y la media ahora es de 15 años. ¿Qué edad tiene Bruno?
 c) ¿Qué edad debe tener una persona para aumentar en 1 año la media del grupo inicial?

Autoevaluación

1. En un club de jugadores de videojuegos se ha anotado el número de partidas ganadas por cada uno de sus miembros. Estos son los resultados:

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

- a) Confecciona con estos datos una tabla de frecuencias absolutas, relativas y porcentuales.
 b) Representalos en un diagrama de barras.

2. Calcula la media, la mediana, la moda y la desviación media de las siguientes distribuciones:

a) 10, 12, 19, 15, 8, 10, 10
 b) 0, 3, 3, 3, 3, 4, 5
 c) 4, 2, 3, 4, 2, 3, 4, 5, 6, 3, 5, 7

3. Halla la mediana y los cuartiles de esta distribución:
 23, 25, 26, 28, 31, 31, 34, 36, 36, 37, 38, 38, 39, 40
 Representa los datos obtenidos en un diagrama de caja y bigotes.