# 11 Estadística

## Presentación de la unidad

- Con esta unidad comenzamos un nuevo bloque: estadística y probabilidad. No es la primera vez que nuestros estudiantes se encuentran con la estadística; ya en tercero de ESO (y en cursos anteriores) se familiarizaron con los conceptos básicos, como la idea de población y muestra, las variables estadísticas, el proceso que se sigue en estadística, la confección de una tabla de frecuencias y algunos parámetros estadísticos (media, mediana, moda, recorrido, desviación media, varianza, desviación típica, coeficiente de variación).
- En esta unidad, además de repasar los conceptos anteriores y recordar las nociones generales (idea de población, muestra, variables estadísticas…), se introducen las dos ramas de la estadística: estadística descriptiva y estadística inferencial.
- Continúa la unidad con un repaso de las tablas de frecuencias y de algunos parámetros estadísticos (media, varianza, desviación típica y coeficiente de variación).
- Hay una profundización en el tratamiento estadístico de datos agrupados en intervalos. Es importante que los estudiantes comprendan la necesidad de agrupar los datos en intervalos cuando la variable es continua, o cuando el número de valores que toma la variable es muy grande. En estos casos, deberán ser capaces de decidir qué intervalos conviene tomar para distribuir los datos que se tengan.
- Se estudian las medidas de posición (mediana, cuartiles y centiles o percentiles) y su contribución a la representación gráfica mediante el diagrama de caja.
- Finalmente, se dedica un apartado a reflexionar sobre las muestras y las razones por las que puede ser necesario recurrir a ellas.
- Es importante que los estudiantes aprendan a calcular los parámetros estadísticos, pero, sobre todo, deben saber interpretarlos.
- Para la obtención de los parámetros, aunque conviene que sepan hacerlo construyendo las tablas, también deben ser capaces de utilizar la calculadora en modo SD.

## Conocimientos mínimos

Al finalizar la unidad, consideraremos imprescindible que los estudiantes hayan alcanzado un conocimiento óptimo de los contenidos siguientes:

- Nociones generales (población y muestra, variables estadísticas, estadística descriptiva y estadística inferencial).

## Esquema de la unidad

- Tablas de frecuencias para datos aislados y para datos agrupados en intervalos.
- Parámetros estadísticos: media, varianza, desviación típica y coeficiente de variación.
- Medidas de posición para datos aislados. Diagramas de caja.
- Uso de la calculadora para introducir datos y para obtener el valor de los parámetros estadísticos.

## Complementos importantes

Además, es conveniente que los estudiantes completen su aprendizaje con otros contenidos complementarios como los que se relacionan a continuación:

- Papel de las muestras en estadística.
- Manejo muy diestro de la calculadora para el tratamiento estadístico.

## Anticipación de tareas

- Repasar algunos conceptos básicos de la estadística como población, muestra, variable…
- Preparar noticias de los medios de comunicación en las que se aluda a algún parámetro estadístico o a alguna gráfica estadística.
- Preparar encuestas y estadísticas realizadas para los medios de comunicación, estudios de mercado u otro fin. La idea es fijarse en los datos sobre el tipo y el tamaño de las muestras tomadas y que los estudiantes vean la aplicación práctica en la elección de muestras.

## Adaptación curricular

En la parte de "Recursos fotocopiables" se ofrece una adaptación curricular de esta unidad 11 del libro del alumnado, para cuya elaboración se han tenido en cuenta los conocimientos mínimos que aquí se proponen.

La lectura inicial servirá para ejercitar la comprensión lectora y para mostrar los dos aspectos que justifican el estudio de las matemáticas: el práctico y el intelectual.

Los contenidos, si se adaptan a esos mínimos exigibles, o bien no han sufrido cambio alguno o bien se han modificado ligeramente para adecuarlos al posible nivel de los estudiantes a quienes va dirigido. Lo mismo cabe decir de los ejercicios prácticos que se proponen.

Si algún contenido supera los mínimos exigibles, o bien se ha suprimido o bien se ha adaptado para ajustarlo a los requisitos exigidos.

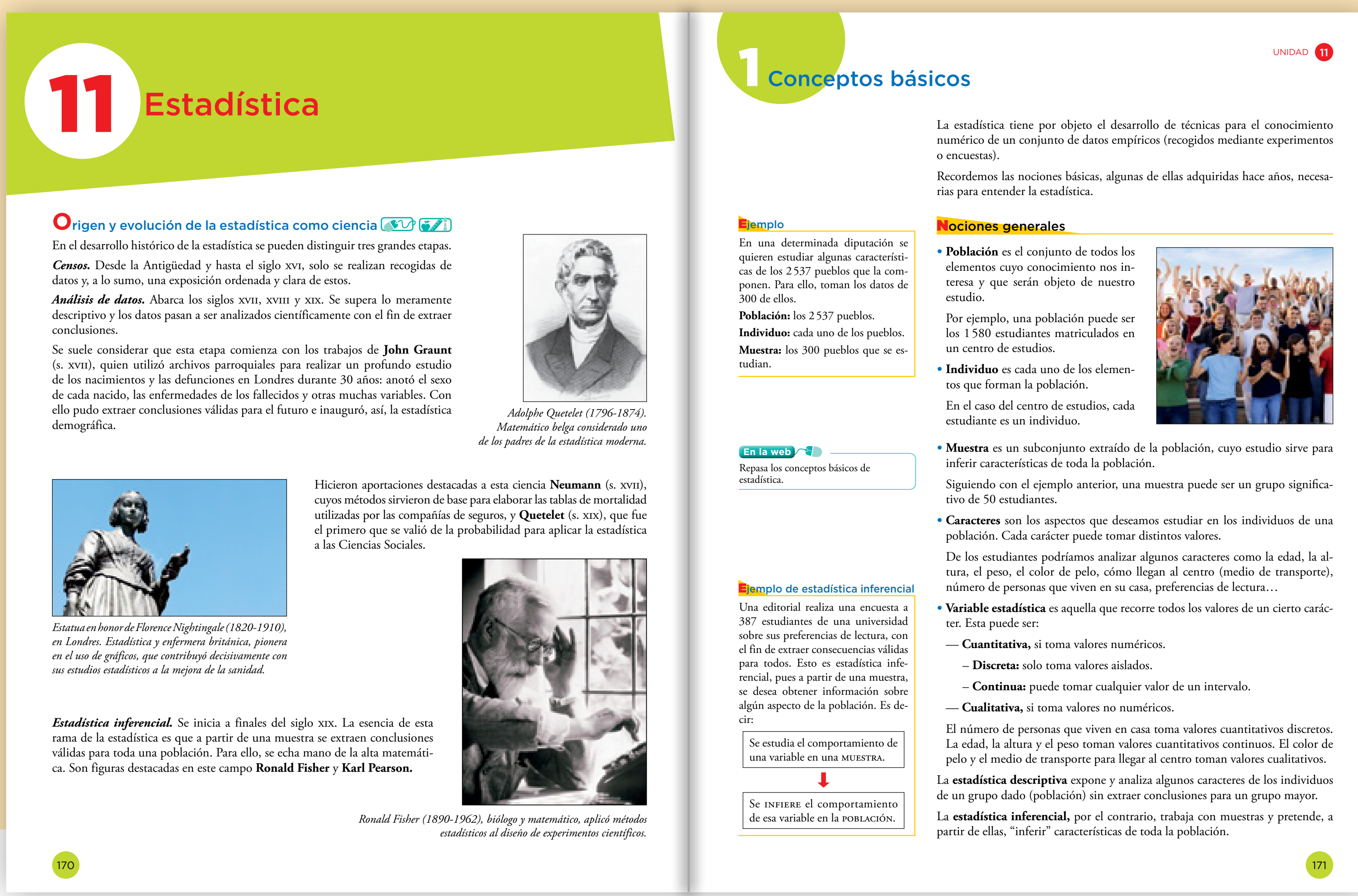
Finalmente, los ejercicios y problemas con los que finaliza la unidad se han reducido en cantidad y se han modificado o bajado de nivel hasta adaptarse a lo convenido.

En la siguiente tabla se recoge una relación de actividades para atender y trabajar el aprendizaje cooperativo, el pensamiento comprensivo, el pensamiento crítico, la interdisciplinariedad, las TIC, el emprendimiento y la resolución de problemas. Unas están propuestas en el libro del alumnado (L.A.), y aquí se hace referencia a ellas indicando la página y la actividad, y otras, como se indica, se sugieren en esta Propuesta Didáctica (P.D.).

Una selección de estas sugerencias están marcadas en el libro del alumnado con un icono; aquí se han marcado con (*).

| APRENDIZAJE COOPERATIVO | PENSAMIENTO COMPRENSIVO | PENSAMIENTO CRÍTICO |
|---|---|---|
| Pág. 173. Actividad sugerida en esta P.D. (*) | Pág. 173. Actividad sugerida en esta P.D. (*) | Pág. 179. Actividad 2 (*) |
| Pág. 175. Piensa y practica | Pág. 175. Ejercicio resuelto | Pág. 180. Piensa y practica (*) |
| Pág. 176. Piensa y practica (*) | Pág. 179. Ejercicios resueltos (*) | Pág. 182. Actividades 12 a 16 (*) |
| Pág. 177. Piensa y practica (*) | | |
| Pág. 179. Piensa y practica | | |
| Pág. 180. Piensa y practica | | |
| Págs. 181, 182 y 183. Actividades 1 a 19 | | |

| INTERDISCIPLINARIEDAD | TIC | EMPRENDIMIENTO | RESOLUCIÓN DE PROBLEMAS |
|---|---|---|---|
| Pág. 170. Actividad sugerida en esta P.D. (*) | Pág. 170. Actividad sugerida en esta P.D. (*) | Pág. 181. Actividades 1, 2 y 5. Actividad sugerida en esta P.D. (*) | Todos los problemas propuestos en el libro del alumnado están encuadrados en este apartado. Aquí se señalan algunos que tienen especial interés. |
| Pág. 183. Curiosidades matemáticas (*) | Pág. 175. Piensa y practica (*) | Pág. 183. Actividad 18 (*) | Pág. 183. Actividades 20 y 21 (*) |



# 11 Estadística

## Origen y evolución de la estadística como ciencia

En el desarrollo histórico de la estadística se pueden distinguir tres grandes etapas.

***Censos.*** Desde la Antigüedad y hasta el siglo XVI, solo se realizan recogidas de datos y, a lo sumo, una exposición ordenada y clara de estos.

***Análisis de datos.*** Abarca los siglos XVII, XVIII y XIX. Se supera lo meramente descriptivo y los datos pasan a ser analizados científicamente con el fin de extraer conclusiones.

Se suele considerar que esta etapa comienza con los trabajos de **John Graunt** (s. XVII), quien utilizó archivos parroquiales para realizar un profundo estudio de los nacimientos y las defunciones en Londres durante 30 años: anotó el sexo de cada nacido, las enfermedades de los fallecidos y otras muchas variables. Con ello pudo extraer conclusiones válidas para el futuro e inauguró, así, la estadística demográfica.

*Adolphe Quetelet (1796-1874). Matemático belga considerado uno de los padres de la estadística moderna.*

*Estatua en honor de Florence Nightingale (1820-1910), en Londres. Estadística y enfermera británica, pionera en el uso de gráficos, que contribuyó decisivamente con sus estudios estadísticos a la mejora de la sanidad.*

Hicieron aportaciones destacadas a esta ciencia **Neumann** (s. XVII), cuyos métodos sirvieron de base para elaborar las tablas de mortalidad utilizadas por las compañías de seguros, y **Quetelet** (s. XIX), que fue el primero que se valió de la probabilidad para aplicar la estadística a las Ciencias Sociales.

***Estadística inferencial.*** Se inicia a finales del siglo XIX. La esencia de esta rama de la estadística es que a partir de una muestra se extraen conclusiones válidas para toda una población. Para ello, se echa mano de la alta matemática. Son figuras destacadas en este campo **Ronald Fisher** y **Karl Pearson.**

*Ronald Fisher (1890-1962), biólogo y matemático, aplicó métodos estadísticos al diseño de experimentos científicos.*

170

UNIDAD 11

## 1 Conceptos básicos

La estadística tiene por objeto el desarrollo de técnicas para el conocimiento numérico de un conjunto de datos empíricos (recogidos mediante experimentos o encuestas).

Recordemos las nociones básicas, algunas de ellas adquiridas hace años, necesarias para entender la estadística.

**Ejemplo**

En una determinada diputación se quieren estudiar algunas características de los 2 537 pueblos que la componen. Para ello, toman los datos de 300 de ellos.

**Población:** los 2 537 pueblos.

**Individuo:** cada uno de los pueblos.

**Muestra:** los 300 pueblos que se estudian.

**En la web**

Repasa los conceptos básicos de estadística.

**Ejemplo de estadística inferencial**

Una editorial realiza una encuesta a 387 estudiantes de una universidad sobre sus preferencias de lectura, con el fin de extraer consecuencias válidas para todos. Esto es estadística inferencial, pues a partir de una muestra, se desea obtener información sobre algún aspecto de la población. Es decir:

Se estudia el comportamiento de una variable en una MUESTRA.

↓

Se INFIERE el comportamiento de esa variable en la POBLACIÓN.

### Nociones generales

- **Población** es el conjunto de todos los elementos cuyo conocimiento nos interesa y que serán objeto de nuestro estudio.

  Por ejemplo, una población puede ser los 1 580 estudiantes matriculados en un centro de estudios.

- **Individuo** es cada uno de los elementos que forman la población.

  En el caso del centro de estudios, cada estudiante es un individuo.

- **Muestra** es un subconjunto extraído de la población, cuyo estudio sirve para inferir características de toda la población.

  Siguiendo con el ejemplo anterior, una muestra puede ser un grupo significativo de 50 estudiantes.

- **Caracteres** son los aspectos que deseamos estudiar en los individuos de una población. Cada carácter puede tomar distintos valores.

  De los estudiantes podríamos analizar algunos caracteres como la edad, la altura, el peso, el color de pelo, cómo llegan al centro (medio de transporte), número de personas que viven en su casa, preferencias de lectura…

- **Variable estadística** es aquella que recorre todos los valores de un cierto carácter. Esta puede ser:

  — **Cuantitativa,** si toma valores numéricos.

  – **Discreta:** solo toma valores aislados.

  – **Continua:** puede tomar cualquier valor de un intervalo.

  — **Cualitativa,** si toma valores no numéricos.

  El número de personas que viven en casa toma valores cuantitativos discretos. La edad, la altura y el peso toman valores cuantitativos continuos. El color de pelo y el medio de transporte para llegar al centro toman valores cualitativos.

La **estadística descriptiva** expone y analiza algunos caracteres de los individuos de un grupo dado (población) sin extraer conclusiones para un grupo mayor.

La **estadística inferencial,** por el contrario, trabaja con muestras y pretende, a partir de ellas, "inferir" características de toda la población.

171

## Al iniciar la unidad

- En esta introducción se realiza un recorrido histórico por los hitos más importantes de la estadística, desde los antiguos censos hasta la ciencia actual, centrándose especialmente en las aportaciones más importantes realizadas en los cinco últimos siglos.

## Interdisciplinariedad

Se sugiere la siguiente actividad:

Buscar y poner en común, en gran grupo, ejemplos de la utilidad y del apoyo que presta la estadística en distintas actividades y en otros campos de la ciencia: biología, física, medicina, deportes, promoción comercial, etc.

## TIC

Se sugiere la siguiente actividad:

Buscar, y poner en común, información en Internet sobre los matemáticos que aparecen en la lectura, destacando su papel en el desarrollo de la estadística.

**ANOTACIONES**

## Sugerencias

- El epígrafe comienza repasando los principales conceptos de estadística que se han estudiado en cursos anteriores, tales como población, individuo, muestra, caracteres, variables estadísticas…
- La comparación entre estadística descriptiva e inferencial pretende que los estudiantes conozcan las dos ramas de la estadística y entiendan la finalidad de cada una de ellas: la estadística descriptiva para presentar un conjunto de datos de tal modo que resulte lo más informativo posible, y la inferencia estadística para proyectar la información obtenida con una muestra al conjunto de la población.
- Los estudiantes pueden ejercitarse buscando problemas reales y reflexionando sobre cuál de las dos ramas de la estadística sería la más idónea para dar respuesta a dicho problema.

**ANOTACIONES**

# 2 Tablas de frecuencias

## Tablas con datos aislados

Tras la recogida de datos, la elaboración de una tabla de frecuencias es el siguiente paso. Cuando la variable toma pocos valores, la elaboración de la tabla es sumamente sencilla. No hay más que hacer el recuento de los resultados.

**En la web**
Recuerda: diagramas de barras e histogramas.

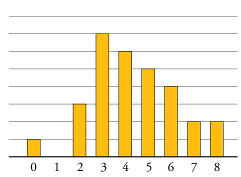
### Ejemplo

A los estudiantes de una clase se les ha preguntado por el número de primos que tienen. Estos son los resultados:

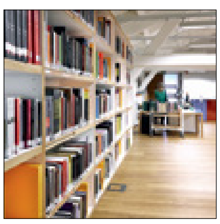
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 3 | 2 | 6 | 4 | 3 | 6 | 5 | 8 | 2 | 4 | 6 | 3 | 2 | 0 |
| 4 | 3 | 7 | 5 | 4 | 3 | 3 | 6 | 7 | 5 | 4 | 8 | 5 | 4 | 3 |

Elaboramos una tabla en la que se muestra ordenadamente cada valor de la variable (número de primos) con su correspondiente frecuencia. La representamos mediante un diagrama de barras.

| $x_i$ | $f_i$ |
|---|---|
| 0 | 1 |
| 1 | 0 |
| 2 | 3 |
| 3 | 7 |
| 4 | 6 |
| 5 | 5 |
| 6 | 4 |
| 7 | 2 |
| 8 | 2 |




## Tablas con datos agrupados



Cuando en una distribución estadística la variable es continua, o bien cuando, siendo discreta, el número de valores que toma la variable es muy grande, conviene elaborar una tabla de frecuencias agrupándolos en intervalos.

Los siguientes son algunos ejemplos en los que es preferible agrupar los datos en intervalos:

- Alturas, pesos, grosores, contornos, áreas, volúmenes, tiempos y, en general, las medidas físicas. Estas son variables continuas, por lo que toman cualquier valor dentro de un intervalo.
- El número de coches que pasan cierto día por cada calle de un barrio; el número de libros consultados en una biblioteca por los usuarios; el número de páginas que tienen los libros de una biblioteca… son medidas discretas que conviene agruparlas como continuas por la cantidad tan elevada de valores distintos que toman.





**En la web**
Refuerza la elaboración de tablas de frecuencias.

**Elaboración de una tabla con datos agrupados en intervalos**

Para elaborar una tabla con datos agrupados, conviene dar los siguientes pasos:

1. Se localizan los valores extremos, $a$ y $b$, y se halla su diferencia, $r = b - a$ (recorrido).
2. Se decide el número de intervalos que se quiere formar, teniendo en cuenta la cantidad de datos que se poseen. El número de intervalos no debe ser inferior a 6 ni superior a 15.
3. Se toma un intervalo, $r'$, de longitud algo mayor que el recorrido, $r$, y que sea múltiplo del número de intervalos, con objeto de que estos tengan una longitud entera.
4. Se forman los intervalos, de modo que el extremo inferior del primero sea algo menor que $a$ y el extremo superior del último sea algo mayor que $b$. Es deseable que los extremos de los intervalos no coincidan con ninguno de los datos. Para ello, conviene que los extremos de los intervalos tengan una cifra decimal más que los datos.

El punto medio de cada intervalo se llama **marca de clase.** Es el valor que representa a todo el intervalo para el cálculo de algunos parámetros.

La tabla de frecuencias del margen puede corresponder a:

— Una distribución de datos aislados que toma los valores $x_1, x_2, \ldots, x_n$.

— Una distribución de datos agrupados en intervalos, siendo $x_1, x_2, \ldots, x_n$ sus marcas de clase.

En el primer caso, la tabla refleja exactamente la distribución real. En el segundo, la tabla es una buena aproximación a la realidad.

**Observa**

Al elaborar una tabla con datos agrupados se pierde algo de información, pues en ella se ignora cada valor concreto, que se difumina dentro de un intervalo. A cambio, se gana en claridad y eficacia.

| $x_i$ | $f_i$ |
|---|---|
| $x_1$ | $f_1$ |
| $x_2$ | $f_2$ |
| … | … |
| … | … |
| … | … |
| $x_n$ | $f_n$ |

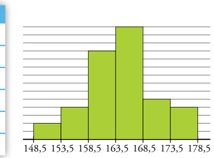
### Ejercicio resuelto

***Elaborar una tabla de frecuencias con las estaturas de 40 adolescentes y representarla en el gráfico correspondiente.***

| | | | | |
|---|---|---|---|---|
| *168* | *160* | *167* | *175* | *175* |
| *167* | *168* | *158* | *149* | *160* |
| *178* | *166* | *158* | *163* | *171* |
| *162* | *165* | *163* | *156* | *174* |
| *160* | *165* | *154* | *163* | *165* |
| *161* | *162* | *166* | *163* | *159* |
| *170* | *165* | *150* | *167* | *164* |
| *165* | *173* | *164* | *169* | *170* |

1. Valores extremos: $a = 149$, $b = 178$. Recorrido: $r = 178 - 149 = 29$.
2. Tomaremos solo 6 intervalos. Un múltiplo de 6 mayor que 29 y próximo a él es 30. Longitud de cada intervalo: 5.
3. Formamos los intervalos comenzando por un número algo menor que $a = 149$ y terminando en un número algo mayor que $b = 178$.
4. Repartimos los datos en los intervalos:

| INTERVALOS | M. DE CLASE | FRECUENCIA |
|---|---|---|
| 148,5-153,5 | 151 | 2 |
| 153,5-158,5 | 156 | 4 |
| 158,5-163,5 | 161 | 11 |
| 163,5-168,5 | 166 | 14 |
| 168,5-173,5 | 171 | 5 |
| 173,5-178,5 | 176 | 4 |




### Piensa y practica

**1.** Reparte los cuarenta datos del ejercicio resuelto anterior en 10 intervalos con el mismo recorrido total.

**2.** Reparte los cuarenta datos del ejercicio resuelto anterior en 8 intervalos. Para ello, toma $r' = 32$.



## Sugerencias

- La técnica de recuento y la posterior elaboración de una tabla de frecuencias en casos sencillos son ya conocidas de cursos anteriores por los estudiantes. Por eso, se hace un pequeño repaso con un ejemplo sencillo.
- En este nivel se pretende que:

  — Reconozcan las situaciones que precisan la agrupación de datos estadísticos en intervalos.

  — Elijan el número de intervalos más adecuado.
- Una buena actividad puede ser pedir a los estudiantes que consigan datos que den lugar a una variable continua, por ejemplo: "tiempo invertido en realizar el trayecto casa-instituto". Cada estudiante aportará su dato, lo cual servirá para inducir la necesidad de agrupar los datos en intervalos. A continuación es el momento de pensar en el número de intervalos que conviene elegir y discutir en qué medida su elección influye en la exactitud de los resultados. La actividad se completa pidiéndoles que, en grupos, confeccionen la tabla de frecuencias, introduciendo así el concepto de marca de clase, y hagan la representación gráfica correspondiente.

## Aprendizaje cooperativo

Las actividades del "Piensa y practica" de la página 173, y todas aquellas en las que se pretende reforzar los conocimientos recién adquiridos, pueden realizarse solidariamente, en pequeño grupo, estimulando el aprendizaje entre iguales.

## Pensamiento comprensivo

Tanto el ejercicio resuelto de la página 173 como los de las páginas siguientes, los alumnos y las alumnas pueden intentar resolverlos inicialmente "con lo que ya saben", detectando las dificultades y los bloqueos. Después, analizarán los procesos que se ofrecen en el texto, poniendo en común sus conclusiones y resolviendo las dudas que les surjan.

## Refuerzo y ampliación

Se recomiendan:

- Del cuaderno n.° 5 de EJERCICIOS DE MATEMÁTICAS:

  Refuerzo: Ejercicio 1 de la pág. 3. Ejercicios 2, 3 y 4 de la pág. 4. Ejercicio 1 de la pág. 5.

  Ampliación: Ejercicios 2, 3 y 4 de la pág. 6. Ejercicios 5, 6 y 7 de la pág. 7.
- Del fotocopiable INCLUSIÓN Y ATENCIÓN A LA DIVERSIDAD:

  Refuerzo: Ejercicio 1 de Practica, ficha A. Ejercicio 1, apartados a) y b), de Practica, ficha B.

  Ampliación: Ejercicio 2, apartados a) y b), de Practica, ficha A.

## Soluciones de "Piensa y practica"

**1** Longitud de cada intervalo: 3.

Las frecuencias, dadas en orden, son: 2, 1, 1, 6, 7, 9, 6, 3, 4, 1.

**2** Longitud de cada intervalo. 4.

Las frecuencias para cada uno, dadas en orden, son: 2, 1, 4, 10, 12, 6, 4, 1.

## ANOTACIONES

# 3 Parámetros estadísticos: $\bar{x}$ y $\sigma$

Recordemos cómo se obtienen los parámetros a partir de una tabla, ya sea con datos aislados o con datos agrupados en intervalos. En este último caso, recurrimos a las marcas de clase para calcular los parámetros.

PUNTUACIONES EN UN TEST

| $x_i$ | $f_i$ | $f_i x_i$ |
|---|---|---|
| 0 | 12 | 0 |
| 1 | 31 | 31 |
| 2 | 86 | 172 |
| 3 | 92 | 276 |
| 4 | 48 | 192 |
| 5 | 19 | 95 |
| | 288 | 766 |

■ **MEDIA:** $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$ $\quad \Sigma f_i x_i \to$ suma de todos los datos; $\Sigma f_i = N \to$ n.º total de individuos

Por ejemplo, en la distribución que tenemos en el margen:

$\Sigma f_i = 288$. Hay 288 individuos (que han realizado el test).

$\Sigma f_i x_i = 766$. Es la suma de las puntuaciones de todos los individuos.

La media es $\bar{x} = 766/288 = 2{,}66$.

■ **VARIANZA:** $Var = \dfrac{\Sigma f_i (x_i - \bar{x})^2}{N}$ o bien $Var = \dfrac{\Sigma f_i x_i^2}{N} - \bar{x}^2$

Las dos expresiones coinciden (consulta la demostración en la web).

— En la primera de ellas, se ve claro el significado de la varianza: promedio de los cuadrados de las desviaciones a la media.

— La segunda es más cómoda para realizar los cálculos.

Utilizamos la segunda fórmula para obtener la varianza a partir de esta tabla:

| $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|
| 0 | 12 | 0 | 0 |
| 1 | 31 | 31 | 31 |
| 2 | 86 | 172 | 344 |
| 3 | 92 | 276 | 828 |
| 4 | 48 | 192 | 768 |
| 5 | 19 | 95 | 475 |
| | 288 | 766 | 2446 |

$$Var = \frac{2446}{288} - 2{,}66^2 = 1{,}42$$

■ **DESVIACIÓN TÍPICA:** $\sigma = \sqrt{varianza}$

La desviación típica es un parámetro más razonable que la varianza, pues se expresa en la misma magnitud que los datos y que la media (por ejemplo, si los datos vienen en centímetros, la desviación típica viene en centímetros; sin embargo, la varianza se daría en centímetros cuadrados).

En el ejemplo: $\sigma = \sqrt{1{,}42} = 1{,}19$

■ **COEFICIENTE DE VARIACIÓN:** $C.V. = \dfrac{\sigma}{\bar{x}}$

El coeficiente de variación sirve para comparar las dispersiones de poblaciones heterogéneas, pues indica la *variación relativa*.

En el ejemplo: $C.V. = \dfrac{1{,}19}{2{,}66} = 0{,}447$. O bien 44,7%.

**¿Para qué sirven los parámetros?**

Los parámetros estadísticos resumen la información que se da en una tabla.

¿Por qué conviene resumirla?

Si un profesor desea ver cómo van los estudiantes de una clase, en la tabla de resultados lo tiene muy claro. Pero si el director del centro quiere comparar unas clases con otras, en diversas asignaturas, incluso con los resultados de otros años… hay demasiada información. Es necesario resumirla. Para eso están los parámetros.

La media, $\bar{x}$, y la desviación típica, $\sigma$, se complementan. Ambas, conjuntamente, aportan una visión bastante buena de la distribución correspondiente.

**En la web**
- Ampliación: demostración de que las dos expresiones dadas para la varianza coinciden.
- HOJA DE CÁLCULO: aplicación para confeccionar tablas de frecuencias, representar el gráfico correspondiente y calcular $\bar{x}$, $\sigma$ y C.V.

**En la web**
- Interpretación de la media.
- Interpretación de la desviación típica.





## Ejercicio resuelto

***Calcular $\bar{x}$, $\sigma$ y C.V. en la siguiente distribución:***

| DISTRIBUCIÓN DE PESOS (en kg) | |
|---|---|
| INTERVALOS | FRECUENCIAS |
| 42,5-53,5 | 4 |
| 53,5-64,5 | 19 |
| 64,5-75,5 | 86 |
| 75,5-86,5 | 72 |
| 86,5-97,5 | 41 |
| 97,5-108,5 | 7 |

Empezamos sustituyendo los intervalos por sus marcas de clase:

$$\frac{42{,}5 + 53{,}5}{2} = 48; \quad \frac{53{,}5 + 64{,}5}{2} = 59; \quad \ldots; \quad \frac{97{,}5 + 108{,}5}{2} = 103$$

La tabla queda así:

| $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|
| 48 | 4 | 192 | 9 216 |
| 59 | 19 | 1 121 | 66 139 |
| 70 | 86 | 6 020 | 421 400 |
| 81 | 72 | 5 832 | 472 392 |
| 92 | 41 | 3 772 | 347 024 |
| 103 | 7 | 721 | 74 263 |
| | 229 | 17 658 | 1 390 434 |

$N = \Sigma f_i = 229$

$\Sigma f_i x_i = 17\,658$

$\Sigma f_i x_i^2 = 1\,390\,434$

Los números de la 3.ª columna, $f_i x_i$, se obtienen multiplicando los números de las columnas anteriores $(x_i \cdot f_i = f_i x_i)$. Por ejemplo, $59 \cdot 19 = 1\,121$.

Análogamente, los de la 4.ª columna se obtienen multiplicando los de la 1.ª por los de la 3.ª $(x_i \cdot f_i x_i = f_i x_i^2)$. Por ejemplo, $59 \cdot 1\,121 = 66\,139$.

Con las sumas de las columnas de la tabla, obtenemos los parámetros:

MEDIA: $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{17\,658}{229} = 77{,}1$ kg

DESVIACIÓN TÍPICA: $\sigma = \sqrt{\dfrac{\Sigma f_i x_i^2}{\Sigma f_i} - \bar{x}^2} = \sqrt{\dfrac{1\,390\,434}{229} - 77{,}1^2} = 11{,}2$ kg

COEF. DE VARIACIÓN: $C.V. = \dfrac{\sigma}{\bar{x}} = \dfrac{11{,}2}{77{,}1} = 0{,}145 = 14{,}5\%$

**CON CALCULADORA**

1. Preparamos la calculadora para que trabaje en el **MODO SD.**
2. Borramos los datos que pudiera haber acumulados de otras ocasiones: INV AC
3. Introducimos los datos: 48 × 4 DATA; 59 × 19 DATA; …; 103 × 7 DATA
4. Resultados obtenidos:

| | | | |
|---|---|---|---|
| N.º DE INDIVIDUOS $\Sigma f_i$ | $n$ | → | 229 |
| SUMA DE VALORES $\Sigma f_i x_i$ | $\Sigma x$ | → | 17658 |
| SUMA DE CUADRADOS $\Sigma f_i x_i^2$ | $\Sigma x^2$ | → | 1390434 |
| MEDIA $\bar{x}$ | $\bar{x}$ | → | 77.10917031 |
| DESV. TÍPICA $\sigma$ | $\sigma_n$ | → | 11.22230132 |

**En la web**
- CALCULADORA: explicación pormenorizada del uso de la calculadora en estadística.
- Refuerza el cálculo de $\bar{x}$, $\sigma$ y C.V.

### Piensa y practica

**1.** Halla, manualmente y con calculadora, $\bar{x}$, $\sigma$ y C.V. en la tabla obtenida en el ejercicio resuelto de la página 173:

| $x_i$ | 151 | 156 | 161 | 166 | 171 | 176 |
|---|---|---|---|---|---|---|
| $f_i$ | 2 | 4 | 11 | 14 | 5 | 4 |

**2.** Halla, manualmente y con calculadora, $\bar{x}$, $\sigma$ y C.V. en la distribución de los ejercicios 1 y 2 de la página 173:

Compara los resultados entre sí y con los del ejercicio 1 de esta página.

**En la web**
- Interpretación del coeficiente de variación.
- Relaciona un histograma con su media y su desviación típica.



## Sugerencias

- No está de más comenzar recordando cómo se complementan las medidas de centralización y de dispersión. Las medidas de dispersión aumentan la información que dan sobre la distribución las medidas de centralización. En esta tarea de complementar, la desviación típica es la pareja natural de la media.
- También, en este apartado, cabría insistir en los siguientes aspectos:

— Significado del símbolo $\Sigma$.

— La conveniencia de utilizar la expresión $\dfrac{\sum f_i x_i^2}{N} - \bar{x}^2$ para el cálculo de la varianza frente a la definición inicial.

— Reconocer el significado del C.V. en situaciones concretas. También interesa que se reflexione y se dé respuesta a la siguiente pregunta: ¿por qué el C.V. es independiente de las unidades de medida?

- El uso de la calculadora, y de las funciones específicas de estadística que posee, da la oportunidad de centrarse en la interpretación de la información y en los análisis críticos de los resultados, más que en las técnicas de cálculo. Por este motivo, es importante que los estudiantes aprendan a calcular la media y la desviación típica con calculadora.

## Refuerzo y ampliación

Se recomiendan:

- Del cuaderno n.º 5 de EJERCICIOS DE MATEMÁTICAS:

Refuerzo: Ejercicios 1 y 2 de la pág. 8. Ejercicio 3 de la pág. 9. Ejercicios 4, 5 y 6 de la pág. 10. Ejercicios 7, 8 y 9 de la pág. 11.

- Del fotocopiable INCLUSIÓN Y ATENCIÓN A LA DIVERSIDAD:

Refuerzo: Ejercicios 1 y 2 de Aplica, ficha A. Ejercicio 1, apartado c), de Practica, ficha B.

Ampliación: Ejercicios 1, 2 y 3 de Aplica, ficha B.

## Soluciones de "Piensa y practica"

**1** $\bar{x} = 164{,}5$ cm

$\sigma = 6{,}24$ cm

C.V. = 0,038 = 3,8%

**2** 1.ª distribución: $\bar{x} = 164{,}4$ cm

$\sigma = 6{,}26$ cm

C.V. = 0,038 = 3,8%

2.ª distribución: $\bar{x} = 164{,}3$ cm

$\sigma = 6{,}14$ cm

C.V. = 0,037 = 3,7%

Las diferencias entre unas y otras son despreciables.

## ANOTACIONES

# 4 Parámetros de posición

### ¿Qué aportan estos parámetros?

Los parámetros de posición son otra forma de sintetizar la información. Tienen la ventaja de que son muy fáciles de interpretar por los no expertos: con una sencilla explicación todo el mundo entiende que la afirmación "*tu estatura está en el percentil 78*" significa que, en ese colectivo, *de cada 100 hay 78 más bajos que yo y 22 más altos*.

**DISTRIBUCIÓN I:** 16 individuos

0 1 2 2 4 5 5 5

6 7 7 8 8 9 10 10

16 : 4 = 4 individuos en cada grupo:

**4**, 4 · 2 = **8**, 4 · 3 = **12**

**DISTRIBUCIÓN II:** 19 individuos

0 1 2 2 3 4 5 5 5 6

6 7 7 8 8 9 10 10 10

19 : 4 = 4,75 individuos en cada grupo:

**4,75**, 4,75 · 2 = **9,5**, 4,75 · 3 = **14,25**

↓ ↓ ↓

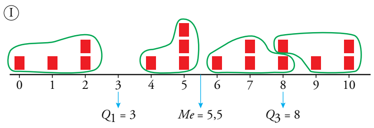
ind. 5.º ind. 10.º ind.15.º

### Mediana y cuartiles

Si los individuos de una población están colocados en orden creciente según la variable que se estudia y partimos la población en cuatro trozos con el mismo número de individuos, los puntos de separación son los cuartiles y la mediana:

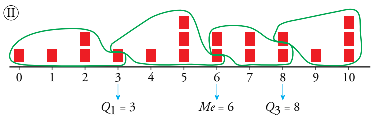
- $Q_1$ (**primer cuartil**) es el valor de la variable que supera al 25 % de la población y, por tanto, queda por debajo del otro 75 %.
- $Me$ (**mediana**) es el valor de la variable que deja el 50 % de la población por debajo y el otro 50 % por encima.
- $Q_3$ (**tercer cuartil**) supera al 75 % de los individuos. Se llama tercer cuartil porque la mediana es el segundo cuartil.

Ⓘ




En este caso, es muy fácil partir en 4 trozos iguales.

Ⓘⓘ




Cada paquete tiene un número fraccionario de individuos.

Observa que los cuartiles ($Q_1$, $Me$, $Q_3$) se sitúan en los individuos que "se parten" para que los cuatro paquetes sean iguales.

### Observa

La mediana es $Me = p_{50}$, y los cuartiles, $Q_1 = p_{25}$ y $Q_3 = p_{75}$.

### Percentiles (o centiles)

Si dividimos la población en 100 partes iguales y señalamos el lugar que deja por debajo a $k$ de ellas, el valor de la variable correspondiente a ese lugar se designa por $p_k$ y se denomina **centil *k*** o **percentil *k*.**

Por ejemplo, en la distribución I, procedemos así para hallar $p_{85}$:

$16 \cdot \frac{85}{100} = 13{,}6 \rightarrow p_{85}$ corresponde al valor del individuo 14.º $\rightarrow p_{85} = 9$

A los parámetros $Q_1$, $Me$, $Q_3$, así como a los percentiles, se les denomina **parámetros de posición.**

**En la web**

Refuerza el cálculo de las medidas de posición.

**Piensa y practica**

**1.** Halla $Q_1$, $Me$, $Q_3$ y $p_{40}$ en esta distribución:

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

6 6 7 7 7 7 8 8 8 8 9 9 9 10 10

**En la web** Cálculo de cuartiles para datos aislados.





**En la web**

HOJA DE CÁLCULO para obtener y visualizar medidas de posición de una distribución.

### Frecuencias acumuladas

En una distribución de frecuencias, se llama **frecuencia acumulada,** $F_i$, correspondiente a $x_i$, a la suma de la frecuencia de ese valor con todas las anteriores:

$$F_i = f_1 + f_2 + \ldots + f_i$$

NÚMERO DE HIJOS DE 120 PAREJAS

| $x_i$ | $f_i$ |
|---|---|
| 0 | 10 |
| 1 | 20 |
| 2 | 41 |
| 3 | 29 |
| 4 | 14 |
| 5 | 5 |
| 6 | 1 |

Por ejemplo, en la distribución de frecuencias dada en el margen, obtenemos:

| $x_i$ | $f_i$ | $F_i$ | | % ACUM. |
|---|---|---|---|---|
| 0 | 10 | 10 | = **10** | 8,3 |
| 1 | 20 | 10 + 20 | = **30** | 25 |
| 2 | 41 | 10 + 20 + 41 | = 30 + 41 = **71** | 59,2 |
| 3 | 29 | 10 + 20 + 41 + 29 | = 71 + 29 = **100** | 83,3 |
| 4 | 14 | 10 + 20 + 41 + 29 + 14 | = 100 + 14 = **114** | 95 |
| 5 | 5 | 10 + 20 + 41 + 29 + 14 + 5 | = 114 + 5 = **119** | 99,2 |
| 6 | 1 | 10 + 20 + 41 + 29 + 14 + 5 + 1 | = 119 + 1 = **120** | 100 |

### Obtención de percentiles en tablas de frecuencias

Para calcular la mediana, los cuartiles y los demás percentiles en distribuciones dadas por tablas de frecuencias, son de gran ayuda las frecuencias acumuladas, $F_i$, y los porcentajes acumulados, % ACUM.

En una tabla de frecuencias, para hallar el percentil $p_k$, se obtienen las frecuencias acumuladas y se expresan en porcentaje. El percentil $p_k$ es el valor para el cual la frecuencia acumulada correspondiente supera el $k$%.

En el caso de que una de ellas coincida con $k$%, se toma como $p_k$ el valor intermedio entre ese valor de $x$ y el siguiente.

### Recuerda

Para hallar qué porcentaje (%) es una cantidad, $F$, con respecto a 120, se procede así:

$$\left(\frac{F}{120} \cdot 100\right)\%$$

| $x_i$ | $f_i$ | $F_i$ | % ACUM. |
|---|---|---|---|
| 0 | 10 | 10 | 8,3 |
| 1 | 20 | 30 | 25 |
| 2 | 41 | 71 | 59,2 |
| 3 | 29 | 100 | 83,3 |
| 4 | 14 | 114 | 95 |
| 5 | 5 | 119 | 99,2 |
| 6 | 1 | 120 | 100 |

Por ejemplo, en la tabla del margen, obtenemos $Me$, $Q_1$, $Q_3$, $p_{99}$ y $p_{95}$:

$Me = p_{50} = 2$ porque para $x_i = 2$, el porcentaje acumulado (% ACUM.) supera el 50.

$Q_1 = p_{25} = 1{,}5$ porque para $x_i = 1$, el % ACUM. es exactamente el 25. Por tanto, el valor 1,5 es superior al 25 % de la población e inferior al otro 75 % restante.

$Q_3 = p_{75} = 3$ porque para $x_i = 3$, el % ACUM. supera el 75.

$p_{99} = 5$ porque para $x_i = 5$, el % ACUM. supera el 99.

$p_{95} = 4{,}5$ porque para $x_i = 4$, el % ACUM. es exactamente el 95.

**Piensa y practica**

**2.** En la siguiente distribución de notas, halla $Me$, $Q_1$, $Q_3$, $p_{80}$, $p_{90}$ y $p_{99}$:

| NOTAS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| N.º DE ALUMNOS | 6 | 19 | 37 | 45 | 109 | 81 | 39 | 22 | 30 | 12 |

**En la web** Cálculo de percentiles para datos aislados.



## Sugerencias

- Los parámetros que aquí se presentan aportan una buena información sobre la distribución.
- Cuando el número de datos es reducido, el cálculo de dichos parámetros es muy cómodo siguiendo el razonamiento que se sugiere en los ejercicios resueltos del margen.
- Cuando los datos se dan en una tabla de frecuencias, es deseable que el alumnado piense en dicha tabla como en una enorme fila; así, en el ejemplo dado en el texto habría una fila de 120 datos. De este modo, los estudiantes enlazan con la idea vista en la página anterior, se familiarizan con lo que es una tabla de frecuencias acumulada y aprenden a usarla para la obtención de estos parámetros.

## Refuerzo y ampliación

Se recomiendan:

- Del cuaderno n.º 5 de EJERCICIOS DE MATEMÁTICAS:

  Refuerzo: Ejercicio 1 de la pág. 12. Ejercicios 2 y 3 de la pág. 13.

- Del fotocopiable INCLUSIÓN Y ATENCIÓN A LA DIVERSIDAD:

  Refuerzo: Ejercicio 3, apartado a), de Practica, ficha A. Ejercicio 3 de Aplica, ficha A.

  Ampliación: Ejercicio 2 de Practica, ficha B.

## Soluciones de "Piensa y practica"

**1** $Q_1 = 3$

$Me = 5{,}5$

$Q_3 = 8$

$p_{40} = 4{,}5$

**2** $Me = 5$

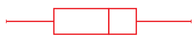
$Q_1 = 4$

$Q_3 = 7$

$p_{80} = 7$; $p_{90} = 9$; $p_{99} = 10$

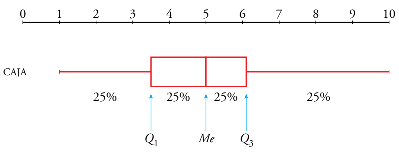
## ANOTACIONES

# 5 Diagramas de caja



Este diagrama se llama también de **caja y bigotes.**

Recordemos la forma de representar una distribución estadística teniendo en cuenta las principales medidas de posición:



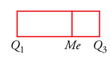

La gráfica corresponde a la distribución de notas en un cierto examen. En la parte alta se ha puesto la escala sobre la que se mueve la variable. Debajo se pone el diagrama propiamente dicho, que consiste en lo siguiente:

— El 50 % de los valores centrales, que se encuentran entre los cuartiles $Q_1$ y $Q_3$, se destacan mediante un rectángulo (**caja**). Esta caja se divide en dos partes: el 25 % de los primeros valores centrales está entre $Q_1$ y $Me$ y el otro 25 %, entre $Me$ y $Q_3$.

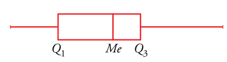
— Los valores extremos (el 25 % de los menores y el 25 % de los mayores) se representa mediante sendos segmentos (**bigotes**).

Los **diagramas de caja** (o caja y bigotes) se construyen del siguiente modo:

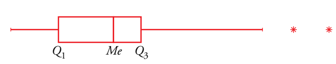
- La caja abarca el intervalo $Q_1$, $Q_3$ (llamado **recorrido intercuartílico**) y en ella se señala expresamente el valor de la mediana, $Me$.




- Los bigotes se trazan hasta abarcar la totalidad de los individuos, con la condición de que cada lado no se alargue más de una vez y media la longitud de la caja.




- Si uno o más de los individuos quedaran fuera de esos tramos, el correspondiente lado del bigote se dibujaría con esa limitación (1,5 · longitud de la caja) y se añadirían, mediante asteriscos, los individuos en los lugares que les correspondan.




La longitud de la rama derecha es 1,5 veces la de la caja. En ella no están incluidos los dos individuos extremos que aparecen representados mediante asteriscos.

**No lo olvides**

La longitud de cada una de las ramas laterales (bigotes) no debe ser superior a 1,5 veces la longitud de la caja $(Q_3 - Q_1)$. Si hay individuos cuyo valor se aleja de $Q_1$ o de $Q_3$ una distancia mayor que esa, se representan mediante puntos sueltos.





## Ejercicios resueltos

**1.** ***Las estaturas de los 40 estudiantes de una clase son, dadas ordenadamente:***

| | | | | |
|---|---|---|---|---|
| 149 | 150 | 154 | 156 | 157 |
| 158 | 159 | 160 | 160 | 160 |
| 161 | 162 | 162 | 163 | 163 |
| 163 | 163 | 164 | 165 | 166 |
| 166 | 166 | 167 | 167 | 167 |
| 168 | 168 | 168 | 169 | 169 |
| 170 | 170 | 170 | 171 | 172 |
| 173 | 174 | 175 | 175 | 189 |

***Representar la distribución mediante un diagrama de caja.***

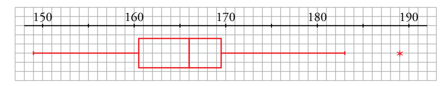
Puesto que el número de individuos es 40, $Q_1$, $Me$ y $Q_3$ serán los valores que hay entre los individuos 10.° y 11.°, entre 20.° y 21.° y entre 30.° y 31.°, respectivamente. Es decir:

$$Q_1 = 160{,}5 \qquad Me = 166 \qquad Q_3 = 169{,}5$$

La longitud de la caja es $Q_3 - Q_1 = 169{,}5 - 160{,}5 = 9$.

Una vez y media esta longitud es $1{,}5 \cdot 9 = 13{,}5$.

El altísimo estudiante que mide 189 cm se separa de $Q_3$, el extremo superior de la caja, $189 - 169{,}5 = 19{,}5$. Esta distancia es mayor que una vez y media la longitud de la caja. Por eso, ponemos a la derecha un bigote de la mayor longitud posible, 13,5, y añadimos un asterisco que señala la situación del individuo excepcional, 189.




**2.** ***Representar, mediante un diagrama de caja, la siguiente distribución:***

| $x_i$ | $f_i$ |
|---|---|
| 0 | 10 |
| 1 | 20 |
| 2 | 41 |
| 3 | 29 |
| 4 | 14 |
| 5 | 5 |
| 6 | 1 |

En la página 177 hemos calculado algunas medidas de posición correspondientes a esta distribución. En concreto:

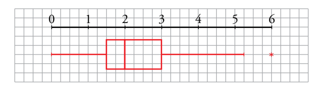
$$Q_1 = 1{,}5 \qquad Me = 2 \qquad Q_3 = 3$$

La caja abarca el intervalo $[Q_1, Q_3] = [1{,}5; 3]$.

La longitud del recorrido intercuartílico es $3 - 1{,}5 = 1{,}5$.

Los segmentos del bigote han de medir, como mucho, $1{,}5 \cdot 1{,}5 = 2{,}25$.

El bigote izquierdo mide menos de 2,25; sin embargo, el derecho, de 2,25, no abarca al elemento mayor (una familia con 6 hijos), ya que $Q_3 + 2{,}25 = 5{,}25$. Lo representamos mediante un asterisco.



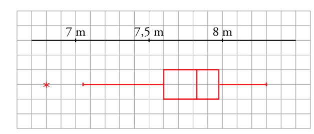

## Piensa y practica

En la web: Representación de diagramas de caja.

**1.** Haz el diagrama de caja correspondiente a esta distribución de notas:

| $x_i$ | $f_i$ |
|---|---|
| 1 | 6 |
| 2 | 15 |
| 3 | 22 |
| 4 | 24 |
| 5 | 33 |
| 6 | 53 |
| 7 | 22 |
| 8 | 16 |
| 9 | 8 |
| 10 | 1 |

**2.** Interpreta el siguiente diagrama de caja y bigotes relativo a las marcas de algunos saltadores de longitud:






## Sugerencias

- Esta curiosa representación gráfica de una distribución estadística resulta muy ilustrativa. Como se pretende con toda gráfica estadística, con un solo golpe de vista se aprecian los aspectos más relevantes de la distribución. Resulta especialmente útil para comparar distintas poblaciones similares. Por ejemplo, las notas de distintos cursos. O bien, de un mismo curso en varios exámenes.

## Refuerzo y Ampliación

Se recomiendan:

- Del cuaderno n.º 5 de EJERCICIOS DE MATEMÁTICAS:

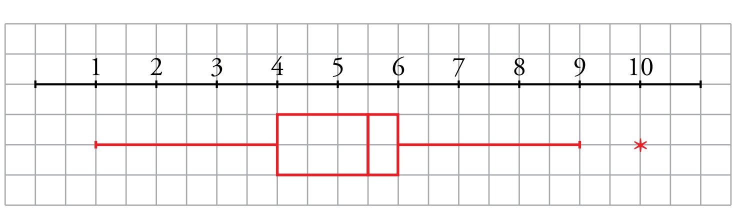
  Refuerzo: Ejercicios 1, 2, 3, 4 y 5 de las págs. 14 y 15.

- Del fotocopiable INCLUSIÓN Y ATENCIÓN A LA DIVERSIDAD:

  Refuerzo: Ejercicio 3, apartado b), de Practica, ficha A.

## Soluciones de “Piensa y practica”

**1**




**2** $Me = 7{,}825$ m; $Q_1 = 7{,}6$ m; $Q_3 = 7{,}975$ m

Todos saltaron entre 7,05 m y 8,3 m, excepto uno que saltó 6,8 m.

Un 25 % de los saltadores saltó menos de 7,6 m.

Un 25 % saltó entre 7,6 m y 7,825 m.

Un 25 % saltó entre 7,825 m y 7,975 m.

Un 25 % saltó más de 7,975 m.

## ANOTACIONES

# 6 Estadística inferencial

### Por qué se recurre a las muestras

I. Si deseamos conocer algunos datos anatómicos (estatura, peso, perímetro torácico…) de los 843 estudiantes de un centro docente, se puede conseguir con facilidad midiéndolos. Sin embargo, si quisiéramos las mismas medidas de todos los jóvenes europeos de edades comprendidas entre 18 y 30 años, la tarea sería desmesurada. Tendríamos que recurrir a una muestra.

II. Para estudiar la duración de una bombilla, hay que dejarla encendida y medir el tiempo transcurrido hasta que se funda. Como es natural, no se puede hacer eso con la totalidad de las bombillas de una producción. Debe recurrirse a una muestra.

III. Deseamos conocer la opinión que tienen sobre las rebajas las personas que acuden a unos grandes almacenes. Es imposible preguntar a todas ellas; hay que recurrir a una muestra, pues no es posible controlar, ni aproximadamente, cuáles son los individuos de la población, ni a cuántos de ellos no se ha encuestado.

En la práctica, para inferir datos de la población, es frecuente recurrir a una muestra por uno o más de los siguientes motivos:

- La población es excesivamente numerosa (caso I).
- La población es muy difícil, o imposible, de controlar (caso III).
- El proceso de medición es destructivo (caso II) o demasiado caro.
- Se desea conocer rápidamente ciertos datos de la población y se tardaría demasiado en consultar a todos los individuos (por ejemplo, los sondeos electorales).

**Piensa y practica**

**1.** Un fabricante de tornillos desea hacer un control de calidad. Recoge uno de cada 100 tornillos fabricados y lo analiza.

El conjunto de tornillos analizados, ¿es población o muestra? ¿Por qué?

**2.** El responsable de calidad de una empresa que fabrica pilas quiere estudiar la energía suministrada por cada pila hasta que se gasta.

¿Puede hacer el estudio sobre la población o debe recurrir a una muestra? ¿Por qué?

**3.** El dueño de un vivero tiene 285 plantas de interior. Para probar la eficacia de un nuevo fertilizante, las mide todas antes y después del semestre que dura el tratamiento.

El conjunto de esas 285 plantas, ¿es población o muestra? ¿Por qué?

180

## Sugerencias

- La estadística inferencial "está en la calle". En los medios de comunicación es frecuente encontrar referencias a muestras, encuestas, sondeos de opinión…, así como a estimaciones, niveles de confianza, etc.
- En este apartado se describe en qué casos es necesario, o conveniente, recurrir a una muestra.

## Soluciones de "Piensa y practica"

**1** Es muestra, pues solo se analiza uno de cada cien tornillos fabricados.

**2** Debe recurrir a una muestra, porque el estudio requiere el consumo de las pilas.

**3** Las 285 plantas serían la población. En este caso, es posible estudiar toda la población. No es necesario trabajar con una muestra.

## ANOTACIONES

## ANOTACIONES

# Ejercicios y problemas

## Practica

### Tablas de frecuencias

**1.** El número de faltas de ortografía que cometieron un grupo de estudiantes en un dictado fue:

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

Di cuál es la variable y de qué tipo es.

Haz una tabla de frecuencias y representa los datos en un diagrama adecuado.

**2.** En una maternidad se han tomado los pesos (en kilogramos) de 50 recién nacidos:

2,8 3,2 3,8 2,5 2,7    3,7 1,9 2,6 3,5 2,3
3,3 2,6 1,8 3,3 2,9    2,1 3,4 2,8 3,1 3,9
2,9 3,5 3,0 3,1 2,2    3,4 2,5 1,9 3,0 2,9
2,4 3,4 2,0 2,6 3,1    2,3 3,5 2,9 3,0 2,7
2,9 2,8 2,7 3,1 3,0    3,1 2,8 2,6 2,9 3,3

a) ¿Cuál es la variable y de qué tipo es?

b) Construye una tabla con los datos agrupados en 6 intervalos desde 1,65 hasta 4,05 y haz una representación adecuada.

### Media, desviación típica y C.V.

**3.** Halla la media, la desviación típica y el coeficiente de variación en estas distribuciones:

| $x_i$ | $f_i$ |
|---|---|
| 0 | 12 |
| 1 | 9 |
| 2 | 7 |
| 3 | 6 |
| 4 | 3 |
| 5 | 3 |

| INTERVALO | $f_i$ |
|---|---|
| 1,65-2,05 | 4 |
| 2,05-2,45 | 5 |
| 2,45-2,85 | 13 |
| 2,85-3,25 | 17 |
| 3,25-3,65 | 8 |
| 3,65-4,05 | 3 |

**4.** Los gastos mensuales de una empresa A tienen una media de 100 000 euros y una desviación típica de 12 500 euros. En otra empresa B, la media es 15 000 euros, y la desviación típica, 2 500 euros. Calcula el coeficiente de variación y di cuál de las dos tiene más variación relativa.

### Parámetros de posición

**5.** La altura, en centímetros, de un grupo de estudiantes de una misma clase es:

150 169 171 172 172 175 181
182 183 177 179 176 184 158

Halla la mediana y los cuartiles y explica el significado de estos parámetros.

**6.** Halla la mediana, los cuartiles y el percentil 60 en cada una de las siguientes distribuciones correspondientes al número de respuestas correctas en un test realizado por dos grupos de estudiantes:

A: 25 – 22 – 27 – 30 – 23 – 22 – 31 – 18
24– 25 – 32 – 35 – 20 – 28 – 30

B: 27 – 32 – 19 – 22 – 25 – 30 – 21
29 – 23 – 31 – 21 – 20 – 18 – 27

**7.** Rellena la columna de los porcentajes acumulados en la siguiente tabla. Calcula, a partir de la tabla, la mediana, los cuartiles y los percentiles $p_{70}$ y $p_{90}$.

| $x_i$ | $f_i$ | $F_i$ | % ACUM. |
|---|---|---|---|
| 0 | 12 | 12 | |
| 1 | 9 | 21 | |
| 2 | 7 | 28 | |
| 3 | 6 | 34 | |
| 4 | 3 | 37 | |

**8.** En la fabricación de cierto tipo de bombillas se han detectado algunas defectuosas. Se analiza el contenido de 200 cajas de 100 bombillas cada una y se obtienen los siguientes resultados:

| DEFECTUOSAS | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| N.° DE CAJAS | 5 | 15 | 38 | 42 | 49 | 31 | 18 | 2 |

Calcula la mediana, los cuartiles y los percentiles $p_{10}$, $p_{90}$ y $p_{95}$.

181

## Soluciones de "Ejercicios y problemas"

**1** Variable cuantitativa discreta: "Número de faltas de ortografía".

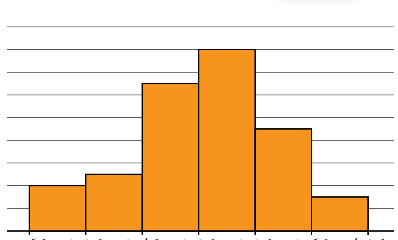
| $x_i$ | 0 | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|
| $f_i$ | 12 | 9 | 7 | 6 | 3 | 3 | 40 |

Diagrama de barras.

**2** a) Variable cuantitativa continua: "Peso de los recién nacidos".

b)

| INTERVALOS | MARCAS DE CLASE ($x_i$) | $f_i$ |
|---|---|---|
| 1,65 - 2,05 | 1,85 | 4 |
| 2,05 - 2,45 | 2,25 | 5 |
| 2,45 - 2,85 | 2,65 | 13 |
| 2,85 - 3,25 | 3,05 | 16 |
| 3,25 - 3,65 | 3,45 | 9 |
| 3,65 - 4,05 | 3,85 | 3 |
| | | 50 |




**3** Distribución de la izquierda: $\overline{x} = 1,7$; $\sigma = 1,57$; C.V. = 0,9235 = 92,35 %

Distribución de la derecha: $\overline{x} = 2,9$; $\sigma = 0,39$; C.V. = 0,1345 = 13,45 %

**4** $C.V._A = 12,5\%$; $C.V._B = 16,67\%$

Tiene mayor variación relativa la empresa B.

**5** $Me = 175,5$ cm; $Q_1 = 171$ cm; $Q_3 = 181$ cm

**6** A: $Me = 25$; $Q_1 = 22$; $Q_3 = 30$; $p_{60} = 27,5$

B: $Me = 24$; $Q_1 = 21$; $Q_3 = 29$; $p_{60} = 27$

**7**

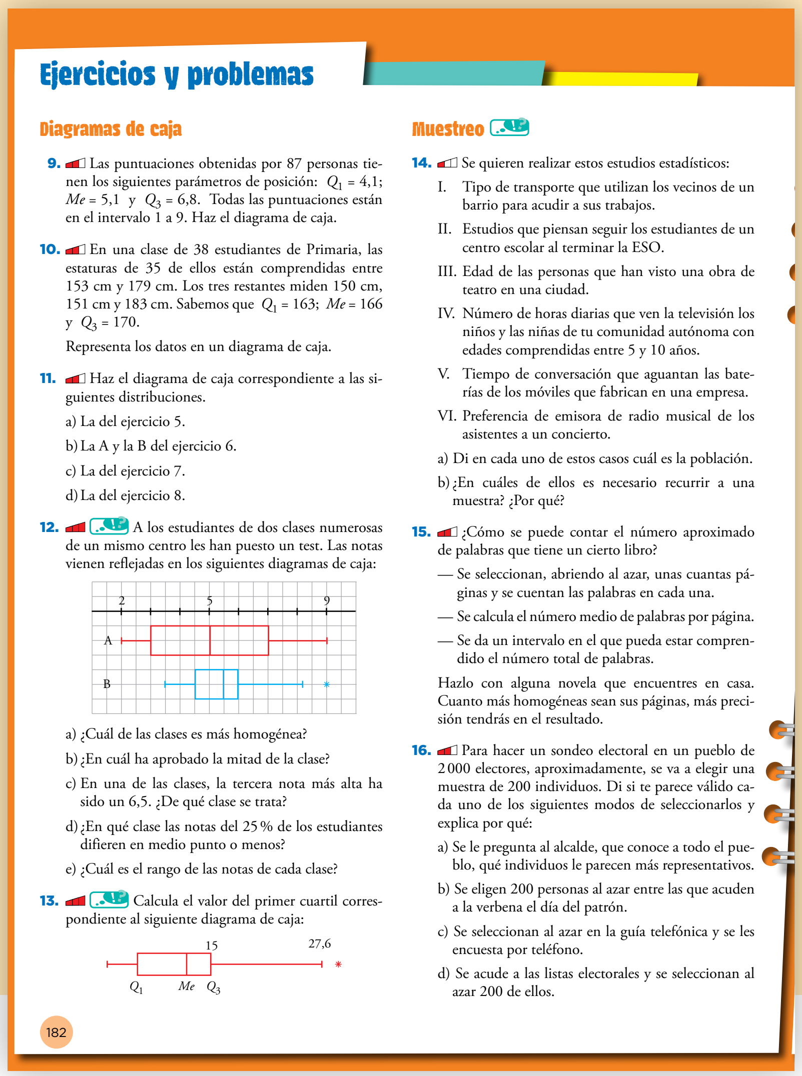
| $x_i$ | $f_i$ | $F_i$ | % ACUM. |
|---|---|---|---|
| 0 | 12 | 12 | 32,4 |
| 1 | 9 | 21 | 56,8 |
| 2 | 7 | 28 | 75,7 |
| 3 | 6 | 34 | 91,9 |
| 4 | 3 | 37 | 100 |

$Q_1 = 0$
$Me = 1$
$Q_3 = 2$
$p_{70} = 2$
$p_{90} = 3$

**8** $Me = 4,5$; $Q_1 = 3$; $Q_3 = 6$

$p_{10} = 2,5$; $p_{90} = 6,5$; $p_{95} = 7$

**ANOTACIONES**



# Ejercicios y problemas

## Diagramas de caja

**9.** Las puntuaciones obtenidas por 87 personas tienen los siguientes parámetros de posición: $Q_1 = 4{,}1$; $Me = 5{,}1$ y $Q_3 = 6{,}8$. Todas las puntuaciones están en el intervalo 1 a 9. Haz el diagrama de caja.

**10.** En una clase de 38 estudiantes de Primaria, las estaturas de 35 de ellos están comprendidas entre 153 cm y 179 cm. Los tres restantes miden 150 cm, 151 cm y 183 cm. Sabemos que $Q_1 = 163$; $Me = 166$ y $Q_3 = 170$.

Representa los datos en un diagrama de caja.

**11.** Haz el diagrama de caja correspondiente a las siguientes distribuciones.

a) La del ejercicio 5.

b) La A y la B del ejercicio 6.

c) La del ejercicio 7.

d) La del ejercicio 8.

**12.** A los estudiantes de dos clases numerosas de un mismo centro les han puesto un test. Las notas vienen reflejadas en los siguientes diagramas de caja:



a) ¿Cuál de las clases es más homogénea?

b) ¿En cuál ha aprobado la mitad de la clase?

c) En una de las clases, la tercera nota más alta ha sido un 6,5. ¿De qué clase se trata?

d) ¿En qué clase las notas del 25 % de los estudiantes difieren en medio punto o menos?

e) ¿Cuál es el rango de las notas de cada clase?

**13.** Calcula el valor del primer cuartil correspondiente al siguiente diagrama de caja:



## Muestreo

**14.** Se quieren realizar estos estudios estadísticos:

I. Tipo de transporte que utilizan los vecinos de un barrio para acudir a sus trabajos.

II. Estudios que piensan seguir los estudiantes de un centro escolar al terminar la ESO.

III. Edad de las personas que han visto una obra de teatro en una ciudad.

IV. Número de horas diarias que ven la televisión los niños y las niñas de tu comunidad autónoma con edades comprendidas entre 5 y 10 años.

V. Tiempo de conversación que aguantan las baterías de los móviles que fabrican en una empresa.

VI. Preferencia de emisora de radio musical de los asistentes a un concierto.

a) Di en cada uno de estos casos cuál es la población.

b) ¿En cuáles de ellos es necesario recurrir a una muestra? ¿Por qué?

**15.** ¿Cómo se puede contar el número aproximado de palabras que tiene un cierto libro?

— Se seleccionan, abriendo al azar, unas cuantas páginas y se cuentan las palabras en cada una.

— Se calcula el número medio de palabras por página.

— Se da un intervalo en el que pueda estar comprendido el número total de palabras.

Hazlo con alguna novela que encuentres en casa. Cuanto más homogéneas sean sus páginas, más precisión tendrás en el resultado.

**16.** Para hacer un sondeo electoral en un pueblo de 2 000 electores, aproximadamente, se va a elegir una muestra de 200 individuos. Di si te parece válido cada uno de los siguientes modos de seleccionarlos y explica por qué:

a) Se le pregunta al alcalde, que conoce a todo el pueblo, qué individuos le parecen más representativos.

b) Se eligen 200 personas al azar entre las que acuden a la verbena el día del patrón.

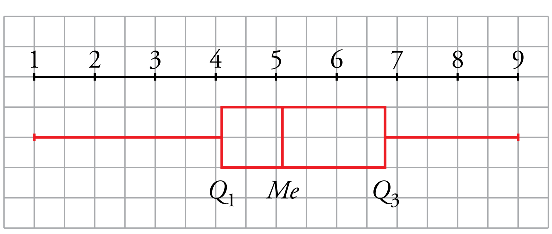
c) Se seleccionan al azar en la guía telefónica y se les encuesta por teléfono.

d) Se acude a las listas electorales y se seleccionan al azar 200 de ellos.

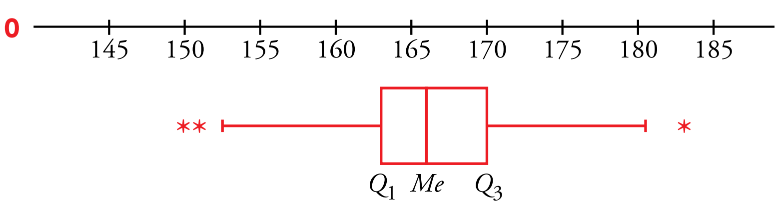
182

## Soluciones de "Ejercicios y problemas"

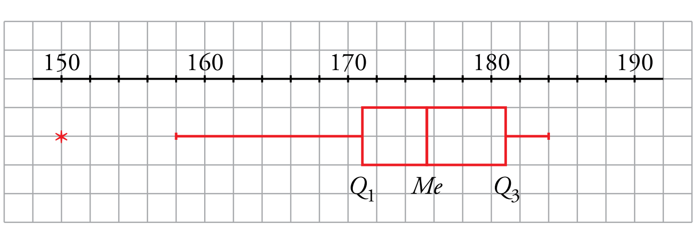
**9**



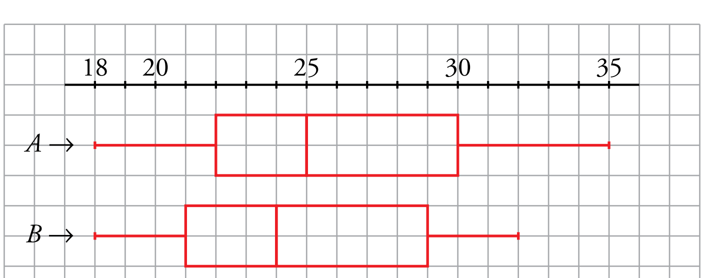

**10**



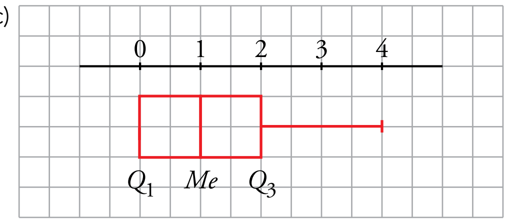

**11** a)



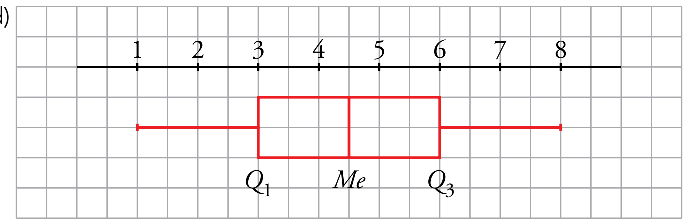

b)




c)




d)




**12** a) Es más homogénea la clase B.

b) En la clase A.

c) En la clase B, puesto que en la A el 25 % tiene notas entre 5 y 7.

d) En la clase B.

e) En la clase A el rango es 9 – 2 = 7.

En la clase B el rango es 9 – 3,5 = 5,5.

**13** $Q_1 = 6{,}6$

**14** a) I → Los vecinos del barrio.

II → Los estudiantes de un centro.

III → Personas que han visto la obra.

IV → Niños y niñas de entre 5 y 10 años de la comunidad autónoma.

V → Los móviles que fabrica la empresa.

VI → Los asistentes a un concierto.

b) I → Muestra

II → Población

III → Muestra

IV → Muestra

V → Muestra

VI → Muestra

**15** Respuesta abierta.

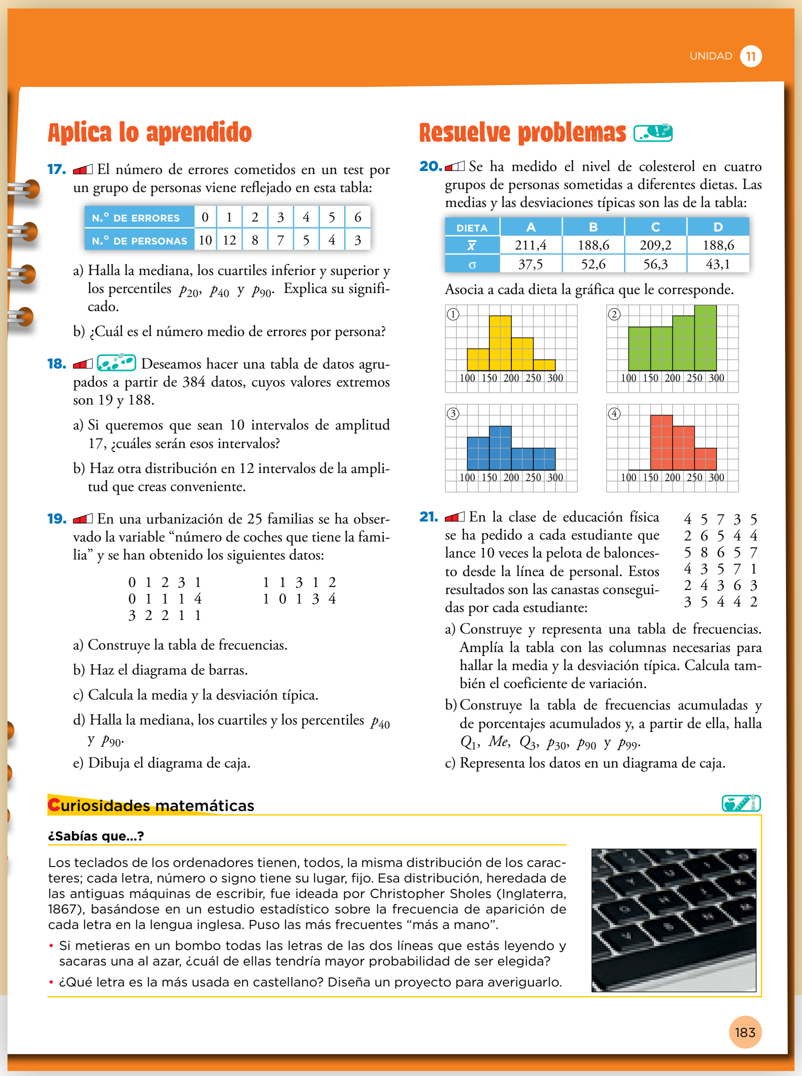
**16** a) No es válido.

b) No es válido.

c) Sí es válido.

d) Sí es válido.

**ANOTACIONES**



UNIDAD 11

## Aplica lo aprendido

**17.** El número de errores cometidos en un test por un grupo de personas viene reflejado en esta tabla:

| N.° DE ERRORES | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| N.° DE PERSONAS | 10 | 12 | 8 | 7 | 5 | 4 | 3 |

a) Halla la mediana, los cuartiles inferior y superior y los percentiles $p_{20}$, $p_{40}$ y $p_{90}$. Explica su significado.

b) ¿Cuál es el número medio de errores por persona?

**18.** Deseamos hacer una tabla de datos agrupados a partir de 384 datos, cuyos valores extremos son 19 y 188.

a) Si queremos que sean 10 intervalos de amplitud 17, ¿cuáles serán esos intervalos?

b) Haz otra distribución en 12 intervalos de la amplitud que creas conveniente.

**19.** En una urbanización de 25 familias se ha observado la variable "número de coches que tiene la familia" y se han obtenido los siguientes datos:

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

a) Construye la tabla de frecuencias.

b) Haz el diagrama de barras.

c) Calcula la media y la desviación típica.

d) Halla la mediana, los cuartiles y los percentiles $p_{40}$ y $p_{90}$.

e) Dibuja el diagrama de caja.

## Resuelve problemas

**20.** Se ha medido el nivel de colesterol en cuatro grupos de personas sometidas a diferentes dietas. Las medias y las desviaciones típicas son las de la tabla:

| DIETA | A | B | C | D |
|---|---|---|---|---|
| $\bar{x}$ | 211,4 | 188,6 | 209,2 | 188,6 |
| $\sigma$ | 37,5 | 52,6 | 56,3 | 43,1 |

Asocia a cada dieta la gráfica que le corresponde.



**21.** En la clase de educación física se ha pedido a cada estudiante que lance 10 veces la pelota de baloncesto desde la línea de personal. Estos resultados son las canastas conseguidas por cada estudiante:

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

a) Construye y representa una tabla de frecuencias. Amplía la tabla con las columnas necesarias para hallar la media y la desviación típica. Calcula también el coeficiente de variación.

b) Construye la tabla de frecuencias acumuladas y de porcentajes acumulados y, a partir de ella, halla $Q_1$, $Me$, $Q_3$, $p_{30}$, $p_{90}$ y $p_{99}$.

c) Representa los datos en un diagrama de caja.

### Curiosidades matemáticas

**¿Sabías que...?**

Los teclados de los ordenadores tienen, todos, la misma distribución de los caracteres; cada letra, número o signo tiene su lugar, fijo. Esa distribución, heredada de las antiguas máquinas de escribir, fue ideada por Christopher Sholes (Inglaterra, 1867), basándose en un estudio estadístico sobre la frecuencia de aparición de cada letra en la lengua inglesa. Puso las más frecuentes "más a mano".

- Si metieras en un bombo todas las letras de las dos líneas que estás leyendo y sacaras una al azar, ¿cuál de ellas tendría mayor probabilidad de ser elegida?
- ¿Qué letra es la más usada en castellano? Diseña un proyecto para averiguarlo.

183

## Soluciones de "Ejercicios y problemas"

**17** a) $p_{20} = 0$    $Q_1 = 1$

$p_{40} = 1$    $Me = 2$

$Q_3 = 3$    $p_{90} = 5$

b) 2,18 errores por persona.

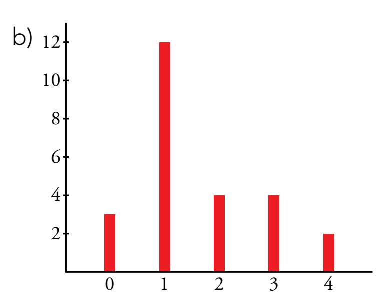
**18** a) [18,5; 35,5); [35,5; 52,5); [52,5; 69,5); [69,5; 86,5); [86,5; 103,5); [103,5; 120,5); [120,5; 137,5); [137,5; 154,5); [154,5; 171,5); [171,5; 188,5)

b) [13,5; 28,5); [28,5; 43,5); [43,5; 58,5); [58,5; 73,5); [73,5; 88,5); [88,5; 103,5); [103,5; 118,5); [118,5; 133,5); [133,5; 148,5); [148,5; 163,5); [163,5; 178,5); [178,5; 193,5)

**19** a)

| $x_i$ | $f_i$ |
|---|---|
| 0 | 3 |
| 1 | 12 |
| 2 | 4 |
| 3 | 4 |
| 4 | 2 |

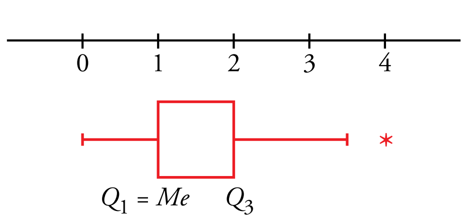
b)




c) $\bar{x} = 1{,}6$; $\sigma = 1{,}13$

d) $Q_1 = 1$; $Me = 1$; $Q_3 = 2$; $p_{40} = 1$; $p_{90} = 3$

e)

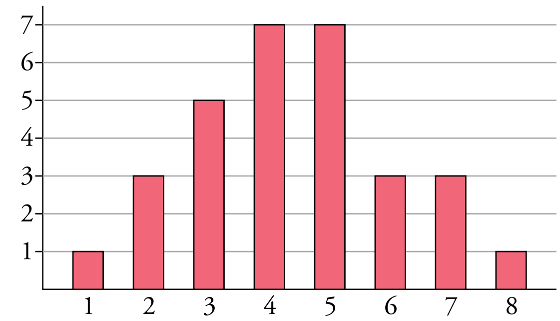



**20** A → 4    B → 3

C → 2    D → 1

**21** a)

| $x_i$ | $f_i$ | $x_i f_i$ | $f_i x_i^2$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 3 | 6 | 12 |
| 3 | 5 | 15 | 45 |
| 4 | 7 | 28 | 112 |
| 5 | 7 | 35 | 175 |
| 6 | 3 | 18 | 108 |
| 7 | 3 | 21 | 147 |
| 8 | 1 | 8 | 64 |
| | | 132 | 664 |



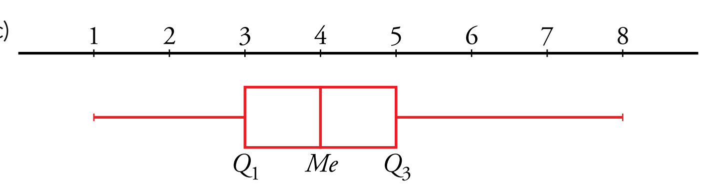

$\bar{x} = 4{,}4$; $\sigma = 1{,}66$; C.V. = 0,38

b)

| $x_i$ | $f_i$ | $F_i$ | % ACUM. |
|---|---|---|---|
| 1 | 1 | 1 | 3,33 |
| 2 | 3 | 4 | 13,33 |
| 3 | 5 | 9 | 30 |
| 4 | 7 | 16 | 53,33 |
| 5 | 7 | 23 | 76,66 |
| 6 | 3 | 26 | 86,66 |
| 7 | 3 | 29 | 96,66 |
| 8 | 1 | 30 | 100 |

$Q_1 = 3$; $Me = 4$; $Q_3 = 5$; $p_{30} = 3{,}5$; $p_{90} = 7$; $p_{99} = 7$

c)




## Curiosidades matemáticas

Interesante reflexión sobre el porqué de la actual distribución de las letras en los teclados de los ordenadores. La investigación que se propone puede ser sencilla, a nivel individual o, si el docente lo considerara oportuno, un trabajo colectivo más extenso y completo.

**Soluciones**

- La letra con mayor probabilidad en la extracción del bombo sería la "a".
- Para estimar la letra más usada en castellano, se sugiere apelar a la creatividad de los estudiantes. Un camino podría ser:

  a) Abrir una novela actual por cualquier página al azar.

  b) Asignar dos líneas a cada estudiante de la clase, para que cada uno repita con ellas el mismo trabajo realizado en la actividad anterior.

  c) Reunir los datos recogidos por el grupo.

  d) Sacar conclusiones.

# ANOTACIONES