

# Materiales para la familia

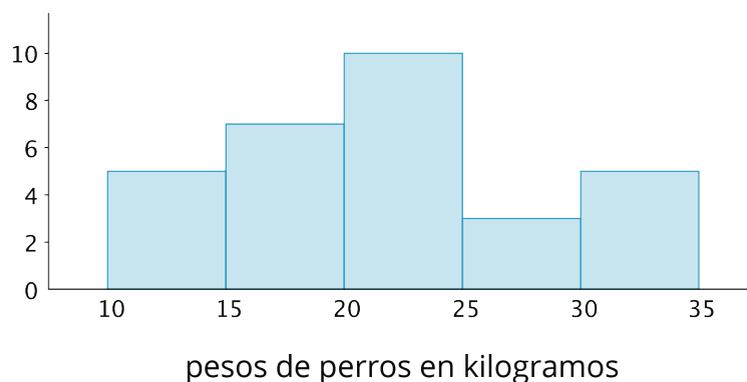
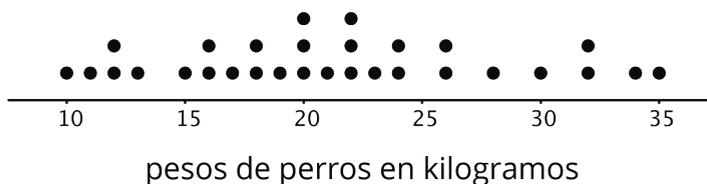
## Conjuntos de datos y distribuciones

### Datos, variabilidad y preguntas estadísticas

#### Materiales para la familia 1

Esta semana, nuestros estudiantes van a trabajar con datos y los van a usar para responder a **preguntas estadísticas**. Preguntas como "¿Cuál es la banda más popular entre los estudiantes de sexto grado?" o "¿Cuál es el número de hermanos más común entre los estudiantes de sexto grado" son preguntas estadísticas y se pueden responder usando datos. Se espera que los datos varíen (por ejemplo, no todos los estudiantes tienen las mismas preferencias musicales ni el mismo número de hermanos).

Nuestros estudiantes han usado gráficos de barras y diagramas de puntos para visualizar e interpretar datos. Ahora van a aprender a usar **histogramas** para dar sentido a datos numéricos. El diagrama de puntos y el histograma a continuación muestran la distribución de los pesos de 30 perros.

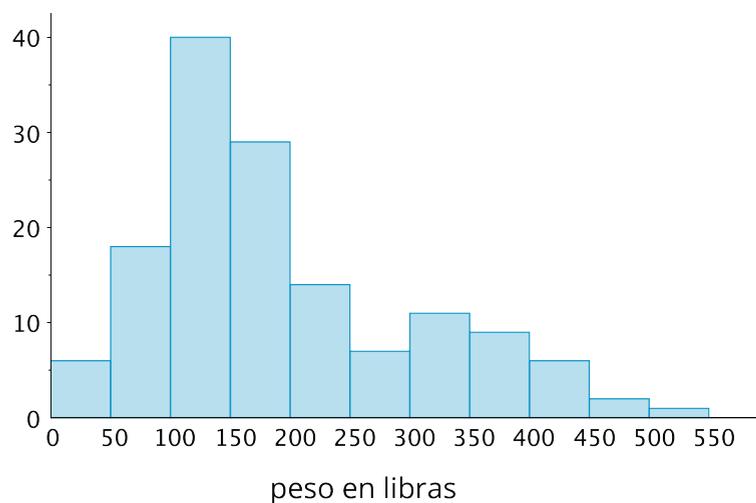


Un diagrama de puntos muestra los valores individuales como puntos. En un histograma, los valores están agrupados. Cada grupo se representa con una barra vertical. La altura de cada barra muestra cuántos valores pertenecen a ese grupo. La barra más alta en este histograma muestra que hay 10 perros que pesan entre 20 y 25 kilogramos.

La forma de un histograma puede indicarnos cómo se distribuyen los datos. Por ejemplo, podemos observar que más de la mitad de los perros pesan menos de 25 kilogramos y que no es muy común (no es muy típico) que un perro pese entre 25 y 30 kilogramos.

Esta es una tarea para que trabajen en familia:

Este histograma muestra el peso de 143 osos.



1. ¿Aproximadamente cuántos osos pesan entre 100 y 150 libras?
2. ¿Aproximadamente cuántos osos pesan menos de 100 libras?
3. Noah argumenta que, como casi todos los osos pesan entre 0 y 500 libras, podemos decir que un peso de 250 libras es típico para los osos de este grupo. Usen el histograma para explicar por qué esto no es correcto.

Solución:

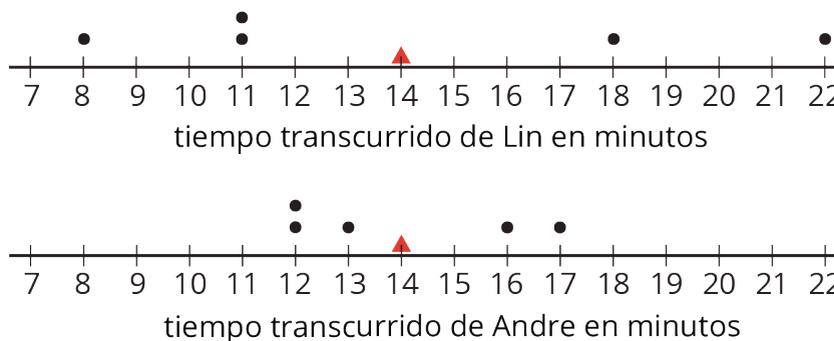
1. Aproximadamente 40 osos. Esa es la altura de la barra más alta del histograma.
2. Aproximadamente 24 osos. Las dos barras que están más a la izquierda representan a los osos que pesan menos de 100 libras. Sumamos las alturas de esas dos barras.
3. Al observar el histograma, podemos ver que la mayoría de los osos pesan menos de 250 libras: las barras a la izquierda de 250 son más altas que las que están a su derecha. Si sumamos las alturas de las barras, menos de 40 osos pesan más de 250 libras, mientras que más de 100 osos pesan menos de 250 libras. Por eso, no es acertado decir que 250 libras es un peso típico.

## Media y MAD

### Materiales para la familia 2

Esta semana, nuestros estudiantes van a aprender a calcular e interpretar la **media**, o promedio, de un conjunto de datos. Podemos pensar en la media de un conjunto de datos como una porción equitativa (o porción justa), es decir, lo que sucedería si los números en los datos estuvieran distribuidos de manera equitativa. Supongamos que una mujer corrió 3, 4, 3, 1 y 5 millas en el transcurso de cinco días. Si el número total de millas que corrió (16 millas) se distribuyera equitativamente a lo largo de esos cinco días, la distancia recorrida por cada día (3.2 millas) sería la media. Para calcular la media, podemos sumar los valores y luego dividir la suma entre el número de valores que haya.

Si pensamos en los puntos de datos como pesos sobre la recta numérica, la media también puede interpretarse como el punto de equilibrio de los datos. En esta figura, los puntos muestran los tiempos de viaje (en minutos) de Lin y Andre. Los triángulos muestran la media de tiempo de viaje en cada caso. Observen que los valores están "equilibrados" a cada lado de cada triángulo.



Nuestros estudiantes aprenderán también a hallar e interpretar la **desviación media absoluta** (o la **MAD**) de los datos. La MAD nos indica la distancia, en promedio, de los puntos de los datos hasta la media. Cuando los puntos de los datos están cerca de la media, las distancias entre ellos y la media son pequeñas, por lo tanto, la distancia promedio (la MAD) también será pequeña. Cuando los puntos de los datos están más dispersos, la MAD será más grande.

Los valores de media y de MAD nos ayudan a resumir (o sintetizar) los datos. La media es una forma de describir el centro del conjunto de datos. La MAD es una forma de describir qué tan dispersos están los datos.

Esta es una tarea para que trabajen en familia:

1. Usen los diagramas de puntos de Lin y de Andre para verificar que en ambos casos la media de los tiempos de viaje es 14 minutos.
  
2. Andre argumenta que la media de sus datos debería ser 13 minutos, pues hay dos números a la izquierda del 13 y dos números a la derecha. Expliquen por qué 13 no puede ser la media.
  
3. ¿Cuál conjunto de datos tiene una MAD (desviación media absoluta) más grande: el de Lin o el de Andre? Expliquen cómo lo saben.

Solución:

1. Para los datos de Lin, la media es  $\frac{8+11+11+18+22}{5} = \frac{70}{5}$ , que es igual a 14. Para los datos de Andre, la media es  $\frac{12+12+13+16+17}{5} = \frac{70}{5}$ , que también es igual a 14.

2. Las explicaciones pueden variar. Ejemplos de explicaciones:

- La media no puede ser 13 minutos porque 13 no representa una porción equitativa.
- La media no puede ser 13 minutos porque los datos no estarían en equilibrio. Los dos valores a la derecha de 13 (16 y 17) están mucho más lejos que los dos valores de la izquierda (12 y 12).

3. Los datos de Lin tienen una MAD más alta. Las explicaciones pueden variar. Ejemplos de explicaciones:

- En los datos de Lin, los puntos están a 6, 3, 3, 4 y 8 unidades de la media, que es 14. En los datos de Andre, los puntos están a 2, 2, 1, 2 y 3 unidades de la media, que también es 14. La distancia promedio en los datos de Lin será mayor, pues esas distancias son mayores.
- La MAD de los datos de Lin es 4.8 minutos y la MAD de los datos de Andre es 2 minutos.
- Comparados con los puntos de los datos de Andre, los de Lin están más lejos de la media.

# Mediana e IQR

## Materiales para la familia 3

Esta semana, nuestros estudiantes van a aprender a usar la **mediana** y el **rango intercuartil** (o **IQR**) para resumir la distribución de los datos.

La mediana es el valor medio del conjunto de datos cuando hacemos una lista de los valores en orden. Para encontrar la mediana, organizamos los datos del más pequeño al más grande y buscamos la mitad de esa lista.

Supongamos que nueve estudiantes reportaron estos números de horas de sueño en una noche de la semana.

6            7            7            8            9            9            10            11            12

El número del medio es 9, por lo tanto la mediana del número de horas de sueño es 9 horas. Esto quiere decir que la mitad de los estudiantes durmieron 9 horas o menos y la otra mitad durmió 9 horas o más.

Supongamos que ocho profesores reportaron estos números de horas de sueño en una noche de la semana.

5            6            6            6            7            7            7            8

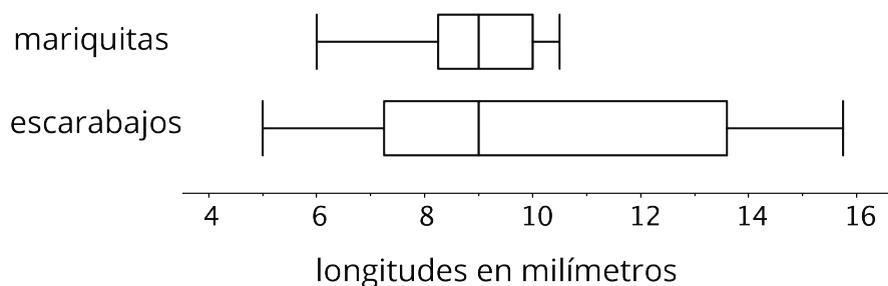
Este conjunto tiene un número par de valores, así que hay dos números en el medio: 6 y 7. La mediana es el número que está exactamente entre ellos dos: 6.5. En otras palabras, si hay dos números en el medio de un conjunto de datos ordenados, la mediana es el promedio de esos dos números.

La mediana señala el percentil 50 de los datos organizados. La mediana parte el conjunto de datos en dos mitades. Cada mitad puede partirse de nuevo en dos, de manera que podamos señalar los percentiles 25 y 75. Los percentiles 25, 50 y 75 se llaman **primer cuartil**, **segundo cuartil** y **tercer cuartil** (o **Q1**, **Q2** y **Q3**).

Un **diagrama de caja** es una forma de representar los tres cuartiles de un conjunto de datos, junto con el valor máximo y el valor mínimo. Este diagrama de caja muestra esos cinco números para los datos de las horas de sueño de los estudiantes:



La distancia entre los cuartiles primero y tercero es el **rango intercuartil** (o **IQR**) de los datos. Nos indica en dónde está la mitad central de los datos y está representado por el "ancho" de la caja del diagrama de caja. Usamos el rango intercuartil para describir qué tan parecidos o diferentes son los valores. Los diagramas de cajas son especialmente útiles cuando comparamos las distribuciones de dos o más conjuntos de datos.



Estos diagramas de caja muestran que el escarabajo más corto (entre los medidos) mide 5 milímetros, y que la mitad de los escarabajos miden entre 7 y 14 milímetros de largo, aproximadamente.

Esta es una tarea para que trabajen en familia:

1. Observen los diagramas de caja de las mariquitas y los escarabajos.
  - a. ¿Qué grupo tiene un IQR más grande: el de las mariquitas o el de los escarabajos? Expliquen cómo lo saben.
  
  - b. ¿Qué grupo muestra más variación en las longitudes: el de las mariquitas o el de los escarabajos? Expliquen cómo lo saben.

2. Esta es una tabla que presenta el número de puntos que Jada anotó en 10 partidos de baloncesto:

10      14      6      12      38      12      8      7      10      23

¿Cuál es la mediana de los puntajes?

Solución:

1.
  - a. Los escarabajos tienen un IQR más grande. Para las mariquitas, el IQR (distancia entre los cuartiles primero y tercero) es aproximadamente 1.7 milímetros. Para los escarabajos, el IQR es aproximadamente 6.3 milímetros.
  - b. A diferencia de los escarabajos, las mariquitas tienen longitudes mucho más parecidas. El IQR para las mariquitas es un número menor y la caja en el diagrama es más angosta, lo que significa que sus longitudes son bastante cercanas las unas a las otras.
2. 11 puntos. Primero, organicen los datos: 6, 7, 8, 10, 10, 12, 12, 14, 23, 38. Luego busquen el centro de la lista: los números 10 y 12 son quinto y sexto en la lista. La mediana es el promedio de esos números:  $\frac{10+12}{2} = 11$ .