Comparemos y contrastemos distribuciones de datos

Investiguemos la variabilidad usando representaciones de datos y estadísticos de resumen.

11.1 Conversación matemática: La media

Evalúa mentalmente el valor de la media de cada conjunto de datos.

- 27 30 33

- 61 71 81 91 101

- 0 100 100 100 100

- 0 5 7 12

11.2 Clasificación de tarjetas: Describamos distribuciones de datos

1. Tu profesor te dará varias tarjetas. Por turnos, con tu compañero, empareja una representación visual de datos con un enunciado escrito.
   a. Para cada pareja que encuentres, explícale a tu compañero cómo sabes que ese gráfico y ese enunciado van juntos.
   b. Escucha con atención la explicación de tu compañero sobre cada una de las parejas que encuentra. Si están en desacuerdo, discutan sus ideas y trabajen para llegar a un acuerdo.
2. Después de hacer las parejas, decide, en cada caso, si la media o la mediana es más apropiada para describir el centro del conjunto de datos dependiendo de la forma de la distribución. Discute tu razonamiento con tu compañero. Si no está dada, calcula (de ser posible) o estima la medida de centro apropiada. Prepárate para explicar tu razonamiento.

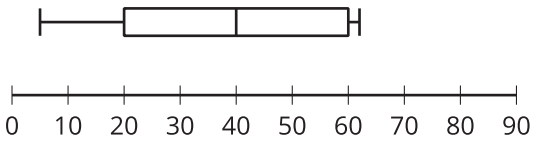
11.3 Analicemos la variabilidad visualmente y con estadísticos

1. Cada diagrama de caja resume el número de millas que alguien recorrió cada día en automóvil durante 30 días, en cierto mes. Los diagramas de caja representan, en orden, las información de agosto, septiembre, octubre, noviembre y diciembre.

   a. Los cinco diagramas de caja tienen la misma mediana. Expliquen por qué la mediana es más apropiada para describir el centro del conjunto de datos de estas distribuciones.

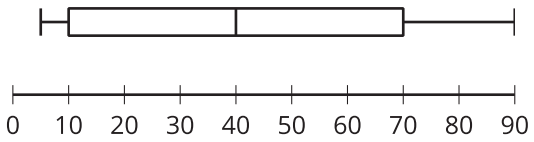
   b. Ordenen los diagramas de caja de menor variabilidad a mayor variabilidad. Revisen con otro grupo si sus respuestas concuerdan.

A



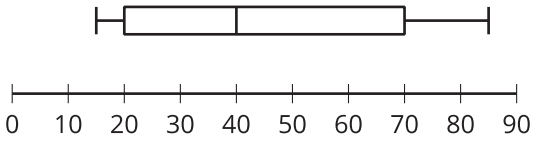
millas recorridas cada día de agosto

B



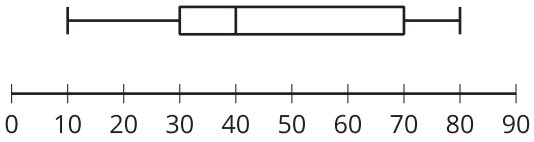
millas recorridas cada día de septiembre

C



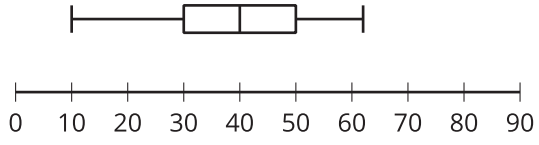
millas recorridas cada día de octubre

D



millas recorridas cada día de noviembre

E



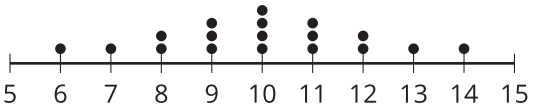
millas recorridas cada día de diciembre

2. Estos cinco diagramas de puntos representan el número de libros que hay en repisas distintas de la misma sección de la biblioteca.

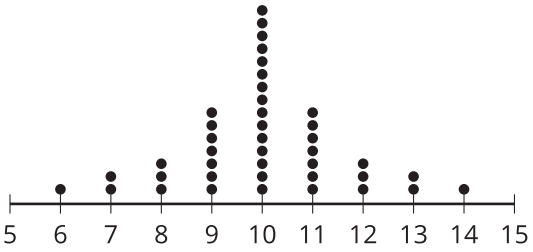
   a. Los cinco diagramas de puntos tienen la misma media. Expliquen por qué la media es más apropiada que la mediana para describir el centro del conjunto de datos.

   b. Ordenen los diagramas de caja de menor variabilidad a mayor variabilidad. Revisen con otro grupo si sus respuestas concuerdan.

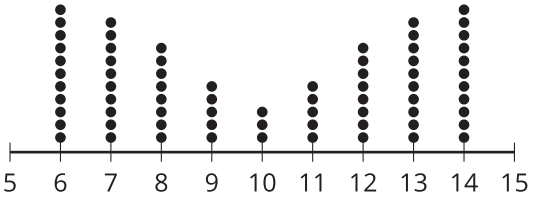
A



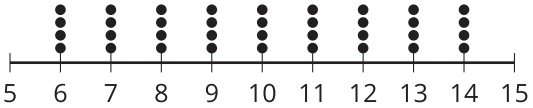
B



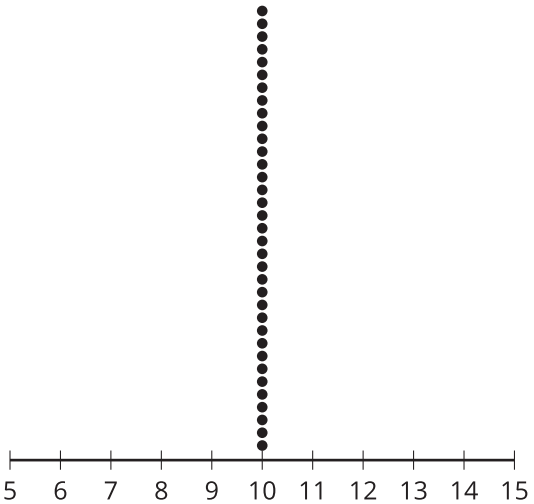
C



D

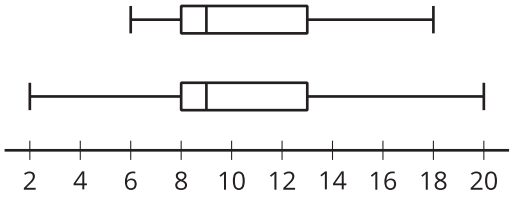


E

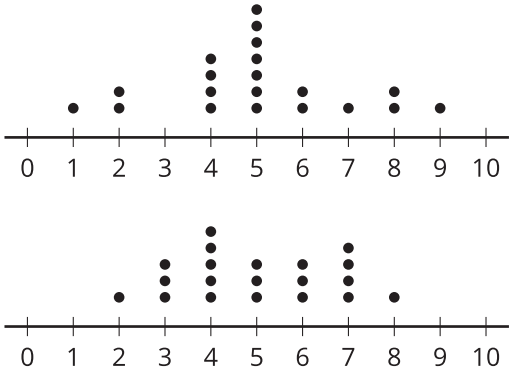


¿Estás listo para más?

1. Estos dos diagramas de caja tienen la misma mediana y el mismo IQR. ¿Cómo podríamos comparar la variabilidad de las dos distribuciones?

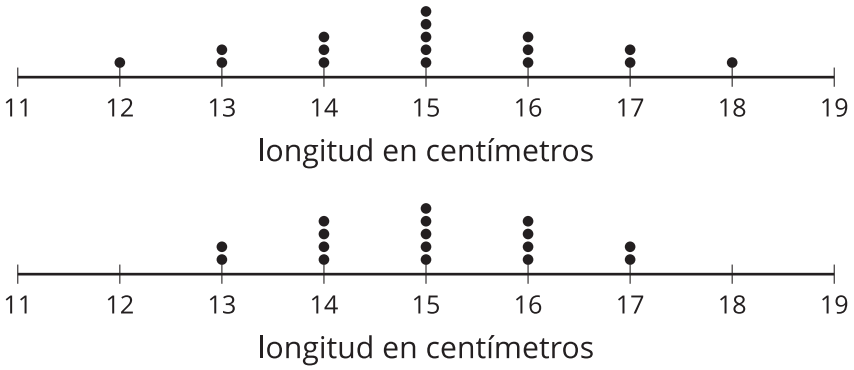


2. Estos dos diagramas de puntos tienen la misma media y la misma MAD. ¿Cómo podríamos comparar la variabilidad de las dos distribuciones?



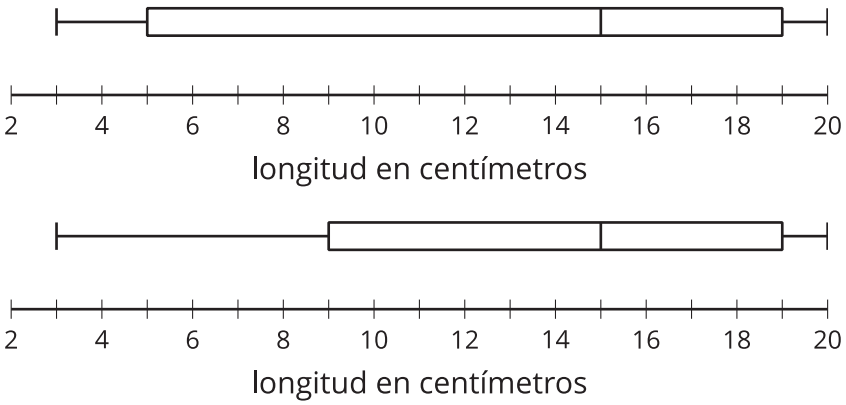
Resumen de la lección 11

La desviación media absoluta, o MAD, es una medida de variabilidad que se calcula encontrando la media de las distancias de todos los puntos de datos a la media de los datos. Estos dos diagramas de puntos muestran la longitud de varias conchas de vieira en centímetros (cm). En cada diagrama, la media es 15 cm.



Observemos que ambos diagramas muestran una distribución simétrica, así que la media y la MAD son elecciones apropiadas para describir el centro y la variabilidad. Los datos del primer diagrama se ven más dispersos que los datos del segundo diagrama, así que podemos decir que el primer conjunto de datos parece tener mayor variabilidad que el segundo conjunto de datos. Esto lo confirma la MAD: la MAD del primer conjunto de datos es 1.18 cm y la MAD del segundo conjunto de datos es aproximadamente 0.94 cm. Esto significa que los valores del primer conjunto de datos están, en promedio, a 1.18 cm de la media, y que los valores del segundo conjunto de datos están, en promedio, aproximadamente a 0.94 cm de la media. Cuanto mayor sea la MAD de los datos, mayor será la variabilidad de los datos.

El rango intercuartil, o IQR, es una medida de variabilidad que se calcula restándole el valor del primer cuartil, Q1, al valor del tercer cuartil, Q3. Estos dos diagramas de caja representan las distribuciones de la longitud en centímetros de otro grupo de conchas de vieira. En cada uno, la mediana es 15 cm.



Observemos que ninguno de los diagramas de caja tiene una distribución simétrica. La mediana y el IQR son elecciones apropiadas para describir el centro y la variabilidad de estos conjuntos de datos. Los datos de la mitad central que se muestran en el primer diagrama de caja se ven más separados, o muestran mayor variabilidad, que los datos de la mitad central que se muestran en el segundo diagrama de caja. El IQR de la primera distribución es 14 cm y el de la segunda distribución es 10 cm. El IQR mide la diferencia entre la mediana de la segunda mitad de los datos, Q3, y la mediana de la primera mitad de los datos, Q1. Por ello, el valor mínimo y el valor máximo del conjunto de datos no influyen en el IQR. El IQR es una medida de la dispersión del 50% de los datos que están en el medio.

La MAD se calcula usando todos los valores de los datos y el IQR se calcula usando solo los valores de Q1 y Q3.