Unidad 3, Lección 10

El efecto de los valores extremos

Entendamos cómo cambian los estadísticos cuando cambian los datos.

10.1 Al menos me fue bien

Muchos videojuegos en línea enfrentan jugadores con otros aleatoriamente para que compitan por equipos.

1. ¿Qué información usarías para encontrar los mejores jugadores de un equipo en un juego como este? Explica tu razonamiento.

2. Hay dos juegos parecidos. Una persona afirma que los mejores jugadores juegan el juego A. Otra persona dice que el juego B tiene mejores jugadores. ¿Cómo podrías representar visualmente los datos para ayudarles a tener una discusión informada? Explica tu razonamiento.

10.2 Separados por asimetría

1. Usa tecnología para crear un diagrama de puntos que represente la distribución de los datos. Después, describe la forma de la distribución.

   6 7 8 8 9 9 9 10 10 10 10 11 11 11 12 12 13 14

2. Encuentra la media y la mediana de los datos.

3. En cada caso, sigue las instrucciones y encuentra la media y la mediana de los datos.

   a. Agrega 2 valores mayores que 14 al conjunto original de datos.

   b. Agrega 2 valores menores que 6 al conjunto original de datos.

   c. Agrega 1 valor mayor que 14 y 1 valor menor que 6 al conjunto original de datos.

   d. Agrega los dos valores 50 y 100 al conjunto original de datos.

4. Comparte tu trabajo con tu grupo. ¿Qué ocurre con la media y la mediana dependiendo de los valores adicionales? ¿Qué observas?

5. Cambia los valores para que la distribución corresponda a la descripción que te dio tu profesor. Después, encuentra la media y la mediana.

6. Busquen un grupo que haya creado una distribución a partir de otra descripción. Expliquen su trabajo y escuchen sus explicaciones. Después, comparen sus medidas de centro.

10.3 Diagramas y medidas que corresponden

Para cada una de las condiciones dadas, crea un diagrama de puntos que tenga al menos 10 valores. Cada diagrama de puntos debe tener por lo menos 3 valores que sean diferentes.

1. una distribución en la que la media y la mediana sean iguales a 10

2. una distribución en la que la media y la mediana sean iguales a -15

3. una distribución en la que la mediana sea igual a 2.5 y la media sea mayor que la mediana

4. una distribución en la que la mediana sea igual a 5 y la mediana sea mayor que la media

¿Estás listo para más?

La media y la mediana son sin duda las medidas de centro más comunes para datos numéricos. Sin embargo, algunas veces se usan otras medidas de centro. Para cada medida de centro, haz una lista de posibles ventajas y desventajas. Asegúrate de considerar cómo cada medida se ve influenciada por los valores extremos.

1. *Media intercuartil*: La media de solo aquellos puntos que están entre el primer cuartil y el tercer cuartil.

2. *Bisagra media*: La media del primer cuartil y el tercer cuartil.

3. *Rango medio*: La media del valor mínimo y el valor máximo.

4. *Trimeo:* La media del primer cuartil, la mediana, nuevamente la mediana y el tercer cuartil. Es un promedio de cuatro números porque la mediana se cuenta dos veces.

Resumen de la lección 10

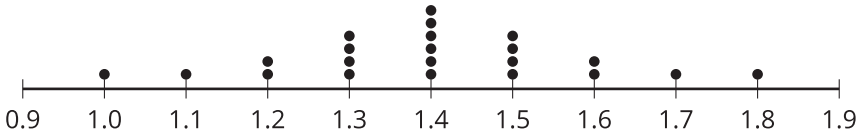
Cuando vamos a describir el centro de un conjunto de datos, ¿es mejor usar la media o la mediana?

Al encontrar el centro usando la media, se le da la misma importancia a cada valor. La media generalmente representa bien los valores típicos cuando los datos tienen una distribución simétrica. Por otro lado, la media se ve muy afectada por los cambios que ocurren, incluso el cambio de un solo valor.

La mediana nos da el valor del medio del conjunto de datos, así que en general los cambios de un solo valor no afectan mucho a la mediana. Por esto, la mediana es más apropiada para datos que no tienen una distribución muy simétrica.

Podemos examinar la distribución de un conjunto de datos y sacar conclusiones sobre la media y la mediana.

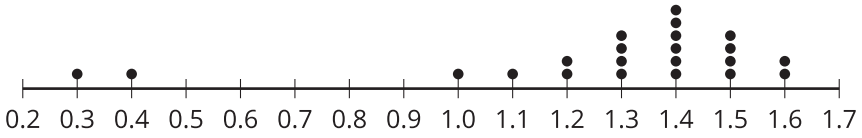
Este diagrama de puntos muestra el tiempo, en segundos, que un dardo tarda en golpear un tablero de dardos. Los datos generan una distribución simétrica.



tiempo para golpear el tablero de dardos en segundos

Cuando una distribución es simétrica, la mediana y la media se encuentran en el medio de la distribución. La mediana es el valor del medio (o la media de los dos valores del medio) de un conjunto de datos, por lo cual podemos usar la simetría alrededor del centro de una distribución simétrica para encontrarla fácilmente. Para encontrar la media, necesitamos saber que la suma de las distancias entre la media y los valores mayores que la media es igual a la suma de las distancias entre la media y los valores menores que la media. Al usar la simetría de la distribución simétrica, podemos ver que hay cuatro valores que están 0.1 segundos por encima de la media, dos valores que están 0.2 segundos por encima de la media, un valor que está 0.3 segundos por encima de la media y un valor que está 0.4 segundos por encima de la media. Asimismo, vemos que hay el mismo número de valores por debajo de la media y se encuentran a las mismas distancias de la media que las que indicamos antes.

Este es un diagrama de puntos en el que se usan los mismos datos, excepto por dos valores que se han cambiado, lo que genera una distribución asimétrica.



tiempo para golpear el tablero de dardos en segundos

Cuando tenemos una distribución asimétrica (es decir, una distribución que no es simétrica), no es posible usar la simetría para encontrar la mediana y la media. La mediana aún es 1.4 segundos porque sigue siendo el valor del medio. Por otro lado, la media ahora es aproximadamente 1.273 segundos. La media es menor que la mediana porque los valores más pequeños (0.3 y 0.4) hacen que la nueva media sea menor.

En general, la mediana es más resistente a los valores extremos que la media. Por esta razón, la mediana es la medida de centro preferida cuando una distribución es asimétrica o cuando hay valores extremos. Al usar la mediana, también podemos usar el IQR como la medida preferida de variabilidad. En una distribución más simétrica, la media es la medida preferida de centro y la MAD es la medida preferida de variabilidad.