

Algebra 1 - Data Distributions and Summary Statistics - Lesson

Common Core S-ID.A.1, S-ID.A.2, S-ID.A.3 · maths-pdf.fr/en/

Data can be displayed in dot plots, histograms, and box plots. The mean and median describe the center, while the range, interquartile range, and standard deviation describe the spread. An outlier is an unusually large or small value. Describe a distribution by its shape, center, spread, and outliers, and compare data sets in context.

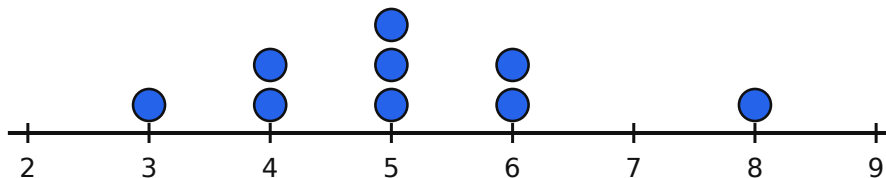
Algebra 1 · Common Core State Standards: **S-ID.A.1, S-ID.A.2, S-ID.A.3**

Key facts

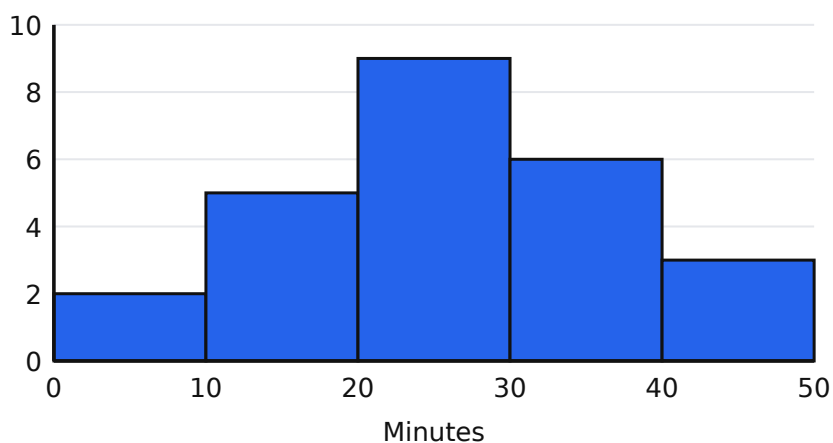
- Dot plots show every value; histograms group data into intervals; box plots show five numbers.
- Mean = $\text{sum} \div \text{count}$; median = middle value.
- $\text{IQR} = Q3 - Q1$; range = maximum - minimum.
- Outlier rule: below $Q1 - 1.5 \cdot \text{IQR}$ or above $Q3 + 1.5 \cdot \text{IQR}$.
- With outliers or skew, the median and IQR are better summaries than the mean and standard deviation.

How do you display data?

A **dot plot** puts one dot above a number line for each value. A **histogram** groups values into equal intervals and uses bar heights for counts. A **box plot** shows the minimum, first quartile (Q1), median, third quartile (Q3), and maximum.



A dot plot shows every value. The mean of these 9 values is about 5.1 and the median is 5.



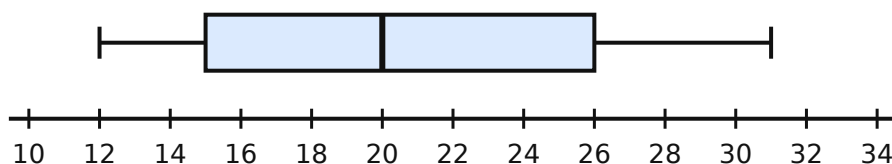
A histogram groups data into intervals. The tallest bar, 9 values, is in the interval 20 to 30.

How do you measure the center?

The **mean** is the sum of the values divided by how many there are. The **median** is the middle value when the data are in order. For 3,4,4,5,5,5,6,8, the mean is $\frac{46}{9} \approx 5.1$ and the median is 5. An extreme value pulls the mean but barely moves the median.

Spread

The **range** is the maximum minus the minimum. The **interquartile range** is $IQR = Q3 - Q1$. The **standard deviation** measures how far values typically are from the mean: $\sigma = \sqrt{\frac{\sum(x - \bar{x})^2}{n}}$.



A box plot shows the five-number summary: minimum 12, Q1 15, median 20, Q3 26, maximum 31.

1. Order the data from least to greatest and find the median.
2. Find Q1 as the median of the lower half and Q3 as the median of the upper half.
3. Compute the IQR and draw the box from Q1 to Q3 with a line at the median.

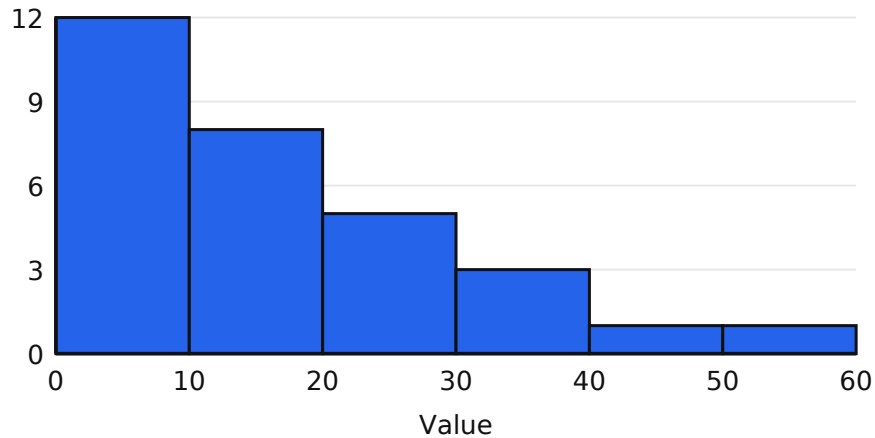
Example: For 12, 15, 17, 20, 22, 26, 31, the median is 20, Q1 is 15, Q3 is 26, and the IQR is $26 - 15 = 11$.

What is an outlier?

A common rule says a value is an **outlier** if it is more than 1.5 IQRs below Q1 or above Q3. The fences are $Q1 - 1.5 \cdot IQR$ and $Q3 + 1.5 \cdot IQR$. If the data have an outlier, the median and IQR describe the data better than the mean and standard deviation.

How do you describe a distribution?

Describe the shape (symmetric, skewed left, or skewed right), the center, the spread, and any unusual values, always in context. When comparing two groups, compare these features side by side.



This distribution is skewed right: most values are small, with a long tail of large values.

Questions about data distributions and summary statistics

What is the difference between mean and median?

The mean is the sum divided by the count. The median is the middle value of the ordered data. Outliers change the mean much more than the median.

How do you find the interquartile range?

Find Q1, the median of the lower half, and Q3, the median of the upper half. Then subtract:
 $IQR = Q3 - Q1$.

What does standard deviation tell you?

It tells how far values typically are from the mean. A larger standard deviation means the data are more spread out.

How do you decide if a value is an outlier?

Compute the fences $Q1 - 1.5 \cdot IQR$ and $Q3 + 1.5 \cdot IQR$. Any value outside the fences is an outlier.

When should you use the median instead of the mean?

Use the median when the distribution is skewed or has outliers, such as incomes or house prices, because the mean is pulled toward the extreme values.