

AP Statistics - Mean, Median, Standard Deviation and Outliers - Lesson

Common Core AP STAT 1.6, AP STAT 1.7, HSS-ID.A.2, HSS-ID.A.3 · maths-pdf.fr/en/

The mean is the balance point of the data and the median is the middle value. The standard deviation measures the typical distance from the mean, while the IQR measures the spread of the middle half. Outliers are flagged by the 1.5 IQR rule. The median and IQR are resistant to outliers; the mean and standard deviation are not.

AP Statistics · Common Core State Standards: **AP STAT 1.6, AP STAT 1.7, HSS-ID.A.2, HSS-ID.A.3**

Key facts

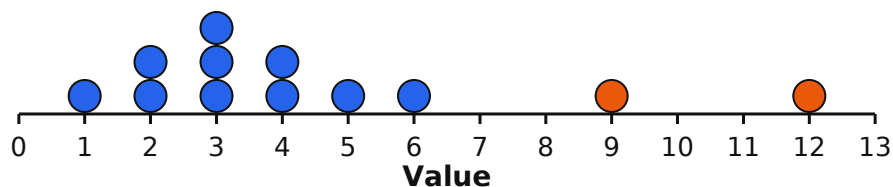
- Mean = sum divided by n ; median = middle of the ordered list.
- Skewed right: mean $>$ median; skewed left: mean $<$ median; symmetric: about equal.
- $s = \sqrt{\text{sum of squared deviations} / (n - 1)}$; IQR = $Q3 - Q1$.
- 1.5 IQR rule: outliers lie below $Q1 - 1.5 \text{ IQR}$ or above $Q3 + 1.5 \text{ IQR}$.
- Adding c shifts center but not spread; multiplying by k scales both by $|k|$.

What are the mean and the median?

The **mean** $\bar{x}$ is the sum of the values divided by n . The **median** is the middle value of the ordered data (or the average of the two middle values when n is even).

Resistant or not?

A statistic is **resistant** if extreme values barely change it. The median and IQR are resistant. The mean, range, and standard deviation are not.



In skewed data the mean is pulled toward the tail, so mean $>$ median here.

In the dotplot, the median is 3.5 but the mean is $\frac{54}{12} = 4.5$. A tail to the right pulls the mean

above the median. For a symmetric distribution they are about equal.

How do you measure spread?

The **range** is maximum minus minimum. The **IQR** is $Q_3 - Q_1$, the width of the middle half. The **standard deviation** measures the typical distance of values from the mean:

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

1. Find the mean.
2. Subtract the mean from each value and square the result.
3. Add the squares and divide by $n-1$.
4. Take the square root.

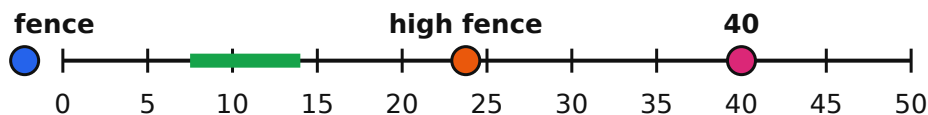
Score x	x - mean	(x - mean)^2
70	-8	64
74	-4	16
78	0	0
80	2	4
88	10	100
Sum	0	184

Standard deviation worked table: the squares add to 184, so $s = \text{square root of } 184/4$, about 6.78.

For the five quiz scores the mean is 78, the squares add to 184, and $s_x = \sqrt{\frac{184}{4}} \approx 6.78$ points.

How do you decide whether a value is an outlier?

The **1.5 IQR rule** flags a value as an outlier if it is below $Q_1 - 1.5 \times IQR$ or above $Q_3 + 1.5 \times IQR$.

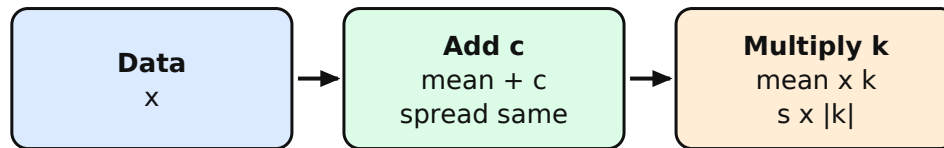


Any value outside the fences, such as 40, is flagged as an outlier by the 1.5 IQR rule.

For the data 5, 7, 8, 9, 10, 12, 13, 15, 40: $Q_1 = 7.5$, $Q_3 = 14$, $IQR = 6.5$, so the fences are -2.25 and 23.75 , and 40 is an outlier.

What happens when you change units?

Adding a constant c to every value adds c to the mean, median, and quartiles but leaves the standard deviation and IQR unchanged. Multiplying every value by k multiplies the center and the spread by $|k|$.



Shifting moves the center; scaling changes the center and spread.

Questions about mean, median, standard deviation and outliers

Why do we divide by $n - 1$ in the standard deviation?

A sample varies less around its own mean than around the true population mean, so dividing by $n - 1$ corrects the shortfall. The result is a better estimate of the population spread.

Which statistic should I report for skewed data?

Report the median and IQR. They are resistant, so a few extreme values do not distort them, while the mean and standard deviation would be pulled toward the tail.

What does a standard deviation of 6.78 mean?

It means a typical score is about 6.78 points away from the mean. A larger standard deviation signals more variability in the data, in the same units as the data.

How do I find the quartiles by hand?

Order the data, find the median, then take the median of the lower half for Q1 and the median of the upper half for Q3. When n is odd, leave the overall median out of both halves.

Does an outlier always have to be removed?

No. Investigate it first. If it is a data-entry error, correct it; if it is real, keep it and report both the outlier and statistics with and without it when relevant.