Interpreting a standard deviation

A standard deviation arrives as one number with a unit, such as 2 mL. What does it say about the data, and when is it the right number to report?

WORDS TO KNOW

Standard deviation: roughly how far a typical value sits from the mean, in the data's own unit.

Spread: how far the values stray from the center, given by the standard deviation or the IQR.

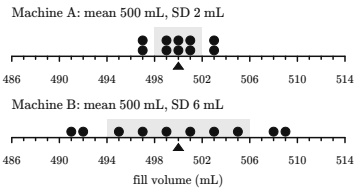
Resistant: barely moved by one far value. The median and the IQR are resistant; the mean and the standard deviation are not.

Center: where the data sits, given by the mean or the median.

IQR (interquartile range): $Q_3 - Q_1$, the width of the middle half of the data.

KEY IDEAS

1. A typical distance, said in one sentence



The standard deviation blends all the distances from the mean. Often no value sits exactly that far away, and some sit further out. So it is a typical distance, not an exact one and not a limit.

Name the individuals, the variable with its unit, and the mean:

The bottles in Machine A's batch typically hold about 2 mL more or less than their mean of 500 mL.

2. Compare center and spread separately

A larger standard deviation means the values stray further from their **own** mean. It says nothing about which group is higher. That is the means' job.

A full comparison says which center is higher **and** which group is more spread out.

3. Count standard deviations to judge "far"

$$\text{number of standard deviations} = \frac{\text{distance from the mean}}{\text{standard deviation}}$$

The count has no unit, so values from different data sets can be compared.

The same 12 mL is ordinary for a machine with standard deviation 6 mL (2 standard deviations) and extreme for one with 2 mL (6 standard deviations).

4. One far value pulls the standard deviation up

Every distance is squared, so one far value adds a huge square. The IQR uses only the quartiles and barely moves.

Roughly symmetric data, no far values: report the **mean with the standard deviation**.

Skewed data or outliers: report the **median with the IQR**.

Free sample from math.wiki for classroom use. Do not post. math.wiki

WATCH OUT

- **The standard deviation is not the largest distance.** Some values sit further than one standard deviation from the mean.
- **It is not a typical value.** A standard deviation of 5 kg does not mean "a typical dog weighs 5 kg".
- **A bigger standard deviation does not mean bigger values.** It means more spread.
- **Check for far values before you trust it.** One outlier can make the standard deviation describe no value at all.

WORKED EXAMPLES

Example 1. Bags of trail mix from one machine have mean 250 g and standard deviation 4 g. Interpret the 4 g. How far out is a 241 g bag? What weight is 1.5 standard deviations above the mean?

1. Sentence: the bags from this machine typically weigh about 4 g more or less than their mean of 250 g.
2. The 241 g bag is 250 − 241 = 9 g light, and 9 ÷ 4 = 2.25. So it is 2.25 standard deviations below the mean.
3. 1.5 standard deviations is 1.5 × 4 = 6 g, so the weight is 250 + 6 = 256 g.

Answer: a typical bag is about 4 g from 250 g; the 241 g bag is 2.25 standard deviations below the mean; 256 g.

Example 2. Lena timed her wait, in minutes, at two cafes on five days. Cafe A: 3, 5, 5, 8, 9. Cafe B: 1, 2, 9, 10, 13. Treat each as a population. Compare the cafes.

1. Cafe A: the waits add to 30, so the mean is 6. The distances 3, 1, 1, 2, 3 have squares 9, 1, 1, 4, 9, adding to 24. The variance is 24 ÷ 5 = 4.8.
$$\sigma = \sqrt{4.8} \approx 2.19 \text{ minutes}$$
2. Cafe B: the waits add to 35, so the mean is 7. The distances 6, 5, 2, 3, 6 have squares 36, 25, 4, 9, 36, adding to 110. The variance is 110 ÷ 5 = 22.
$$\sigma = \sqrt{22} \approx 4.69 \text{ minutes}$$
3. Center: Cafe B's waits are longer on average, by 1 minute.
4. Spread: a Cafe B wait typically sits about 4.69 minutes from its mean, more than twice Cafe A's 2.19 minutes.

Answer: Cafe B has the longer average wait, and its wait is much less predictable. For a steady wait, Cafe A is better.

Example 3. A shop's daily returns over eight days were 4, 5, 5, 6, 6, 7, 7, 40. Find the mean, standard deviation, median and IQR, and say which pair to report.

1. Mean: the values add to 80, so the mean is 80 ÷ 8 = 10.
2. Distances from 10: 6, 5, 5, 4, 4, 3, 3, 30. Squares: 36, 25, 25, 16, 16, 9, 9, 900, adding to 1036.
3. The variance is 1036 ÷ 8 = 129.5, and the standard deviation is $\sqrt{129.5} \approx 11.4$.
4. Median: the mean of the 4th and 5th values, 6 and 6, so 6.
5. Quartiles: the lower half 4, 5, 5, 6 gives $Q_1 = 5$, and the upper half 6, 7, 7, 40 gives $Q_3 = 7$. The IQR is 2.
6. The single 40 gives 900 of the 1036. "Typically about 11.4 from 10" fits no day: seven days are within 3 of 6.

Answer: mean 10, standard deviation about 11.4, median 6, IQR 2. Report the median with the IQR, and mention the 40 on its own.

Free sample from math.wiki for classroom use. Do not post. math.wiki