

AP Statistics - Five-Number Summary, Boxplots and Comparing Distributions - Lesson

Common Core AP STAT 1.8, AP STAT 1.9, HSS-ID.A.1, HSS-ID.A.2 · maths-pdf.fr/en/

A boxplot displays the five-number summary: minimum, Q1, median, Q3, and maximum. The box contains the middle half of the data, whiskers reach the most extreme non-outliers, and stars mark outliers found with the 1.5 IQR rule. To compare distributions, discuss center, spread, shape, and unusual features with comparison words and context.

AP Statistics · Common Core State Standards: **AP STAT 1.8, AP STAT 1.9, HSS-ID.A.1, HSS-ID.A.2**

Key facts

- Five-number summary: minimum, Q1, median, Q3, maximum.
- Each quarter of the ordered data holds about 25% of the values.
- $IQR = Q3 - Q1$; outliers lie more than 1.5 IQR outside the box.
- Compare center, spread, shape, and unusual features using comparison words.
- Boxplots show skew and outliers but not the number of peaks.

What is the five-number summary?

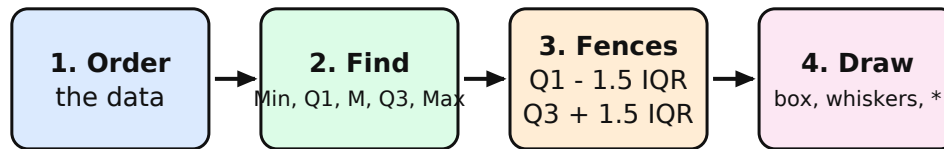
The **five-number summary** lists the minimum, first quartile Q_1 , median, third quartile Q_3 , and maximum. Together they split the ordered data into four groups that each contain about 25% of the values.

Min	Q1	Median	Q3	Max
3	7	11	15.5	30

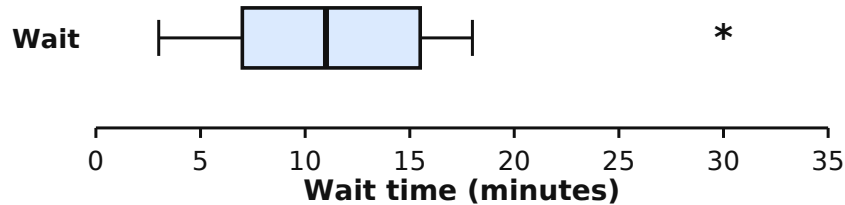
Five-number summary of 13 wait times in minutes.

The interquartile range is $IQR = Q_3 - Q_1 = 15.5 - 7 = 8.5$ minutes. The upper fence is $15.5 + 1.5(8.5) = 28.25$, so 30 is an outlier.

How do you draw a boxplot?



Whiskers stop at the most extreme values that are not outliers.



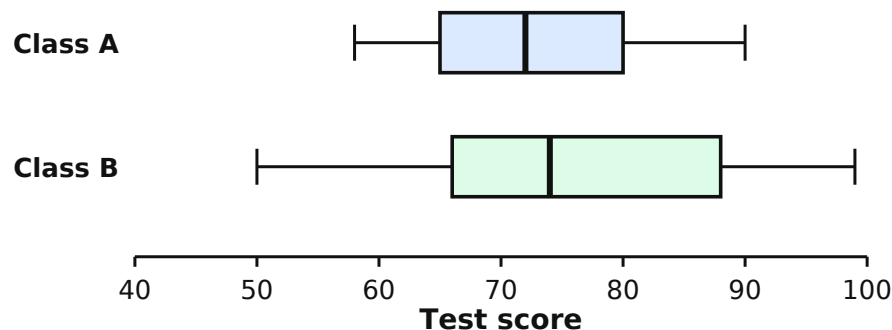
The box runs from Q1 to Q3, the line marks the median, and the star marks the outlier 30 (beyond the fence 28.25).

Reading a boxplot

The box holds the middle 50% of the data. If the median line sits near one side of the box, or one whisker is much longer, the distribution is skewed toward the longer side.

How do you compare distributions?

When you compare two groups, discuss **center**, **spread**, **shape**, and **unusual features**, and use comparison words such as "greater than" and "more variable than." A list of two separate descriptions earns little credit.



Class B has a similar center but a wider spread than Class A.

- **Center:** the median of Class B (74) is a little higher than that of Class A (72).
- **Spread:** Class B has an IQR of 22 compared with 15 for Class A, so Class B is more variable.
- **Shape:** both are roughly symmetric, with the median near the middle of each box.
- **Unusual features:** neither plot shows outliers.

How do you read skew and outliers in a boxplot?

Look at the position of the median line inside the box and at the lengths of the whiskers. In the wait-time boxplot, the median 11 is closer to Q1 (distance 4) than to Q3 (distance 4.5), but the upper whisker and the outlier star stretch far to the right, so the distribution is skewed right. A star never ends a whisker: the whisker stops at the last value inside the fence, which is 18 here.

Worked example: for the wait times, $Q1 = 7$ and $Q3 = 15.5$, so the lower fence is $7 - 1.5(15.5 - 7) = -5.75$. No value is below it, which is why the lower whisker reaches the minimum, 3. Always compute both fences before you draw.

A boxplot hides the number of peaks. Use a histogram or dotplot when you need to see gaps or bimodality.

Questions about five-number summary, boxplots and comparing distributions

What percent of the data lies inside the box?

The box runs from Q1 to Q3, so it contains about the middle 50% of the data. About 25% lies in each whisker region and the two halves of the box.

Where does a whisker end?

A whisker ends at the smallest or largest value that is not an outlier. Values beyond the 1.5 IQR fences are plotted separately as stars or dots.

How can I tell skewness from a boxplot?

If the median is closer to Q1 and the upper whisker is longer, the data are skewed right. If the median is closer to Q3 and the lower whisker is longer, the data are skewed left.

Why compare medians and IQRs instead of means?

The median and IQR are resistant, so they describe center and spread reliably when the data are skewed or contain outliers. The five-number summary uses exactly these statistics.

What should a comparison sentence look like?

Use a comparison word and the context, such as: The median wait at Clinic B (14 minutes) is higher than at Clinic A (11 minutes), and its times are more variable (IQR 9 compared with 6).