

AP Statistics · Exploring One-Variable Data · Lesson

Common Core AP Statistics Unit 1 (2026-27) · mathovore.fr

This ap statistics one variable lesson explains exploring one-variable data in one printable page, with definitions, a method and a worked example.

One-variable data analysis describes a single variable. Categorical variables use bar charts and relative frequencies; quantitative variables are described by shape, center, spread and unusual features using histograms, boxplots, the mean and standard deviation, or the median and IQR. Z-scores and the empirical rule place a value within a normal distribution.

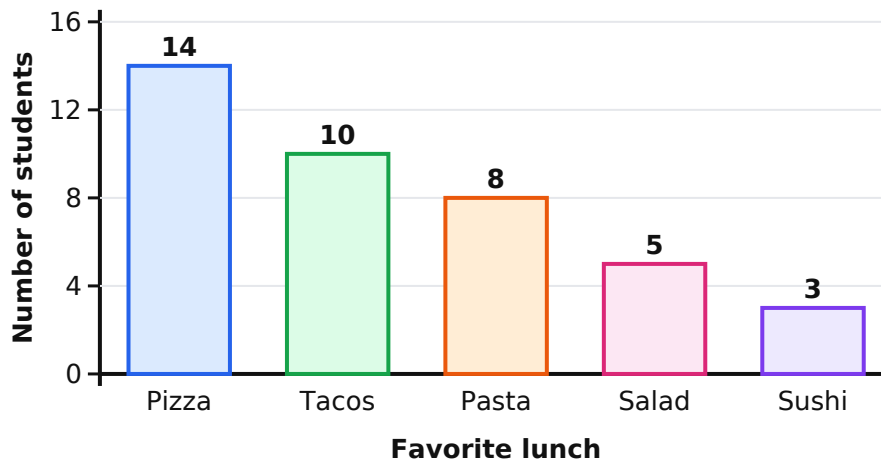
AP StatisticsCommon Core: AP Statistics Unit 1 (2026-27)

AP Statistics One Variable Lesson: key facts

- Describe every quantitative distribution with shape, center, spread and unusual features, in context.
- Use the median and IQR for skewed data or outliers, and the mean and standard deviation for symmetric data.
- A value is an outlier if it is below $Q1 - 1.5 \text{ IQR}$ or above $Q3 + 1.5 \text{ IQR}$, or if it is more than 2 standard deviations from the mean.
- A z-score is $(x - \text{mean}) / s$; the empirical rule gives 68%, 95% and 99.7% within 1, 2 and 3 standard deviations.

What is the difference between categorical and quantitative variables?

A **categorical variable** sorts each individual into a group, such as favorite lunch or eye color. A **quantitative variable** is a number where arithmetic makes sense, such as minutes of travel. A zip code is categorical because adding zip codes means nothing. Show categorical data in a bar chart and compare groups with relative frequencies, count divided by total.



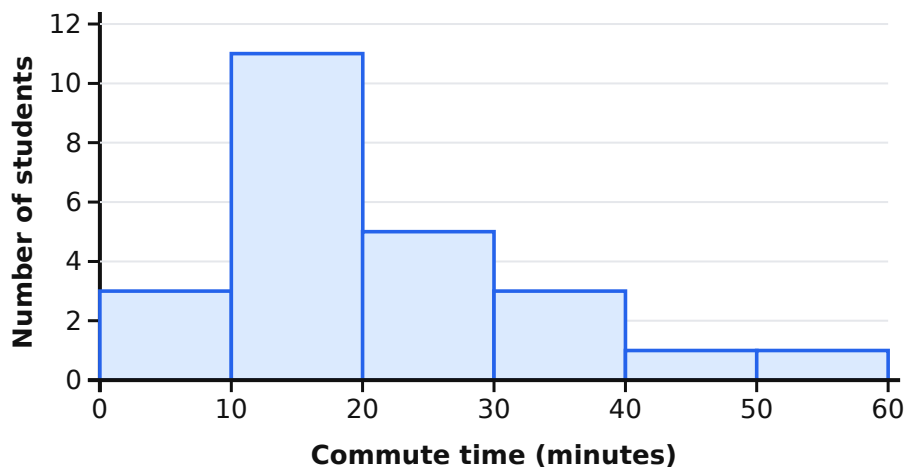
Favorite lunch of 40 students: pizza 35%, tacos 25%, pasta 20%, salad 12.5%, sushi 7.5%.

How do you describe a distribution?

Always address **shape** (symmetric, skewed right, skewed left, unimodal, bimodal), **center**, **spread** and **unusual features**, all in context. Skewed right means a long tail toward large values.

Distribution	Center	Spread
Roughly symmetric, no outliers	mean $\bar{x}$	standard deviation s
Skewed or has outliers	median	IQR

The median and IQR are resistant to outliers. The mean and s are not.

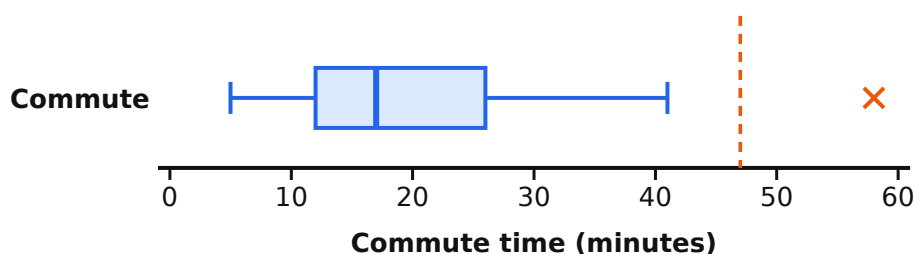


Commute times: skewed right, with one far-right value near 58 minutes.

How do you build a five-number summary and a boxplot?

1. Order the data and find the median.
2. Q_1 is the median of the lower half and Q_3 the median of the upper half (leave out the median when n is odd).
3. Compute $IQR = Q_3 - Q_1$. Values below $Q_1 - 1.5IQR$ or above $Q_3 + 1.5IQR$ are outliers.
4. Draw the box from Q_1 to Q_3 , a line at the median, and whiskers to the most extreme values that are not outliers.

Example. The 24 commute times have minimum 5, Q_1 12, median 17, Q_3 26 and maximum 58. So $IQR = 26 - 12 = 14$, and the fences are $12 - 1.5(14) = -9$ and $26 + 1.5(14) = 47$. The value 58 is above 47, so it is an outlier and the upper whisker stops at 41. The second rule flags a value more than 2 standard deviations from the mean: $\bar{x} \approx 20.3$ and $s \approx 12.3$ give an upper limit of 44.9, and again only 58 is flagged.



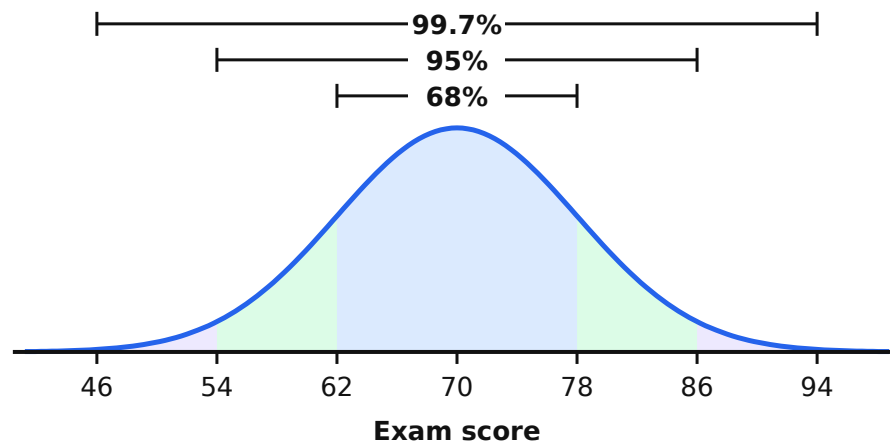
Boxplot with dashed fence at 47 minutes; the value 58 is plotted as an outlier.

What are z-scores, percentiles and the empirical rule?

Definitions

$z = \frac{x - \bar{x}}{s}$ tells how many standard deviations a value is from the mean. The **percentile** of a value is the percent of the data below it.

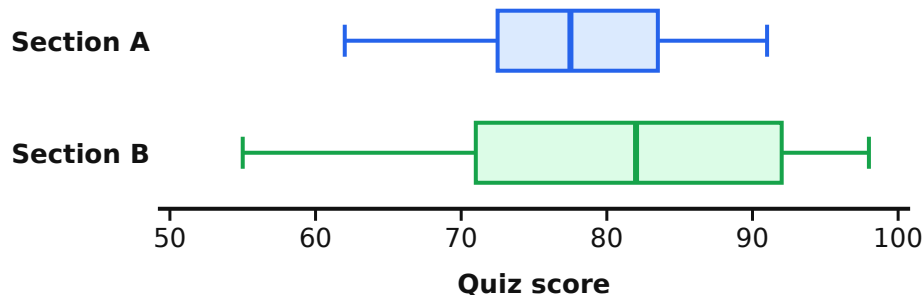
A 40-minute commute has $z = \frac{40 - 20.3}{12.3} \approx 1.60$. In an approximately normal distribution about 68%, 95% and 99.7% of values lie within 1, 2 and 3 standard deviations of the mean. With mean 70 and standard deviation 8, a score of 86 has $z = \frac{86 - 70}{8} = 2$, so about 97.5% of scores fall below it.



Empirical rule for a normal model with mean 70 and standard deviation 8.

How do you compare two distributions?

Use side-by-side boxplots, then write full sentences that name both groups and cover center, spread and shape. Section B has the higher median, and its IQR of 21 is almost double Section A's IQR of 11.



Section B has the higher median (82 vs 77.5) and the larger spread (IQR 21 vs 11).

Exploring One-Variable Data: frequently asked questions

When should I use the median instead of the mean?

Use the median when the distribution is skewed or has outliers, because it is resistant. A single value of 58 among 24 commute times pulls the mean up to 20.3 but leaves the median at 17.

Why does AP Statistics divide by $n - 1$ in the standard deviation?

Data are almost always a sample, and dividing by $n - 1$ gives a better estimate of the spread of the population. Calculators report this as S_x . The value measures the typical distance from the mean.

How do quartiles work when n is odd?

Leave the median out, then take the median of the lower half for $Q1$ and of the upper half for $Q3$. For 11 ordered values, $Q1$ is the median of the first five.

What does a z-score of -1.5 mean?

The value is 1.5 standard deviations below the mean. In an approximately normal distribution that is near the 7th percentile.

What must a comparison of two distributions include?

Name both groups and compare center, spread and shape with comparison words such as higher, more variable or more skewed, and give numbers from the summaries. Mention any outliers in context.

How to use this ap statistics one variable lesson

Use this ap statistics one variable lesson before the worksheet, then check what you learned with the quiz.

