

Exploring One-Variable Data: Shape, Center, Spread

Grade 12 – AP Statistics math · Lesson · Standards: AP Statistics Unit 1 (2026-27)

Quick answer: One-variable data analysis describes a distribution by its shape, center, spread and unusual features. Use graphs for the overall pattern, the median with IQR or the mean with standard deviation for numbers, the $1.5 \times \text{IQR}$ rule for outliers, and z-scores to compare values.

Standards: AP Statistics Unit 1 (2026-27) · **Level:** Grade 12 AP Statistics

What kinds of variables does statistics study?

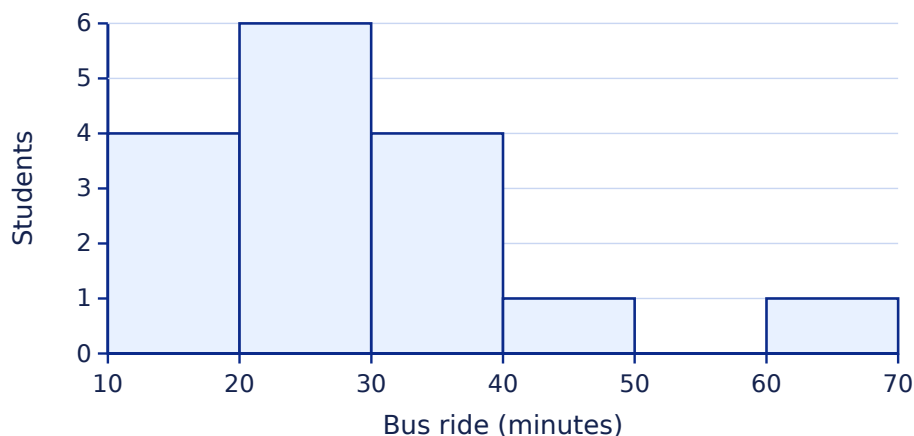
Variable types

A **categorical** variable places each individual in a group (hometown, blood type). A **quantitative** variable is a number on which arithmetic makes sense (bus ride in minutes, height in inches).

Categorical data are summarized with counts or relative frequencies and shown in bar graphs. Quantitative data are shown in dotplots, histograms and boxplots.

How do you describe a quantitative distribution?

Always address four features, in context: **shape** (symmetric, skewed left, skewed right, unimodal, bimodal), **center**, **spread**, and **unusual features** such as outliers or gaps.



Bus ride times for 16 students: a skewed-right shape with one far-away value.

Measure	Type	Resistant to outliers?
Mean $\bar{x}$	center	No
Median	center	Yes
Standard deviation s	spread	No
IQR = $Q_3 - Q_1$	spread	Yes

For the 16 bus rides, $\bar{x} = 27.75$ minutes but the median is 25.5 minutes. The mean is pulled toward the long ride, which is typical of right skew.

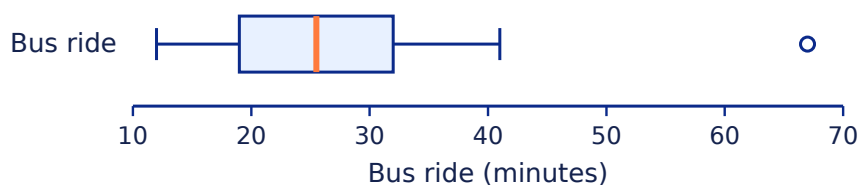
How do you find outliers and draw a boxplot?

Method: the $1.5 \times \text{IQR}$ rule

1. Order the data and find the five-number summary: min, Q_1 , median, Q_3 , max (use the medians of the lower and upper halves).
2. Compute $\text{IQR} = Q_3 - Q_1$.
3. Compute the fences $Q_1 - 1.5 \times \text{IQR}$ and $Q_3 + 1.5 \times \text{IQR}$.
4. Any value beyond a fence is an outlier. Draw whiskers to the most extreme values inside the fences.

Worked example

Bus rides: $Q_1 = 19.0$, $Q_3 = 32.0$, so $\text{IQR} = 13.0$. The upper fence is $32.0 + 1.5 \times 13.0 = 51.5$. Since $67 > 51.5$, the 67-minute ride is an outlier.



The circle at 67 is flagged as an outlier; the whisker stops at 41.

A second rule calls a value an outlier when it lies more than 2 standard deviations from the mean. The two rules can disagree, so always name the rule you use.

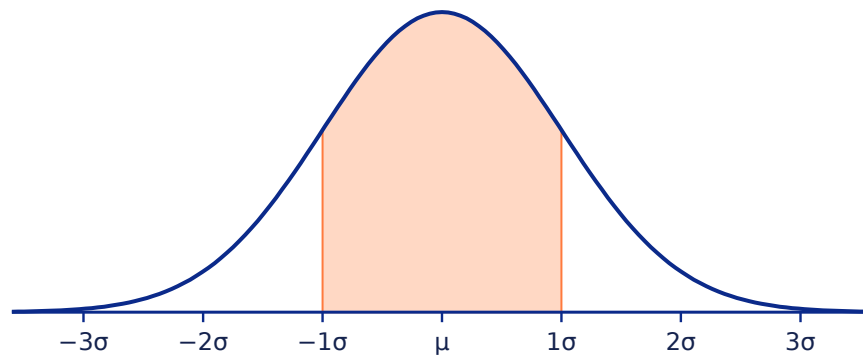
What do z-scores and the empirical rule tell you?

Standardizing

$$z = \frac{(x - \mu)}{\sigma}$$

A z-score counts how many standard deviations a value is above (positive) or below (negative) the mean. A percentile is the percent of values below a given value.

For a roughly normal (bell-shaped, symmetric) distribution, about 68% of values fall within 1σ of μ , 95% within 2σ and 99.7% within 3σ .



Distance from the mean in standard deviations

Empirical rule: about 68%, 95% and 99.7% of values lie within 1, 2 and 3 standard deviations of the mean.

Mathbot's tip

Never say a distribution is skewed because of one outlier. Look at the long tail of the histogram, then check which way the mean sits compared with the median.

Key points to remember

- Categorical data: counts and bar graphs. Quantitative data: dotplots, histograms, boxplots.
- Median and IQR are resistant to outliers; mean and standard deviation are not.
- Outlier fences: $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$.
- $z = (x - \mu)/\sigma$ measures distance from the mean in standard deviations.
- Empirical rule for normal data: 68%, 95%, 99.7%.

Exploring One-Variable Data: Shape, Center, Spread: frequently asked questions

What is the difference between categorical and quantitative variables?

A categorical variable puts individuals into groups, such as favorite sport. A quantitative variable records a measured or counted number, such as minutes of homework. Zip codes look like numbers but are categorical because averaging them has no meaning.

When should I use the median instead of the mean?

Use the median, with the IQR, when the distribution is skewed or has outliers, because both are resistant. Use the mean, with the standard deviation, when the distribution is roughly symmetric with no outliers.

How do I describe the shape of a distribution?

Say whether it is symmetric, skewed left or skewed right, and count the peaks (unimodal, bimodal). The skew direction is the side with the longer tail, not the side with the bigger bars.

What does a z-score of -1.5 mean?

The value lies 1.5 standard deviations below the mean. Z-scores let you compare values from different distributions, such as scores on two tests with different means and spreads.

Does the empirical rule work for every data set?

No. The 68-95-99.7 percentages are for distributions that are approximately normal. For skewed or irregular distributions, the percentages inside 1, 2 and 3 standard deviations can be quite different.