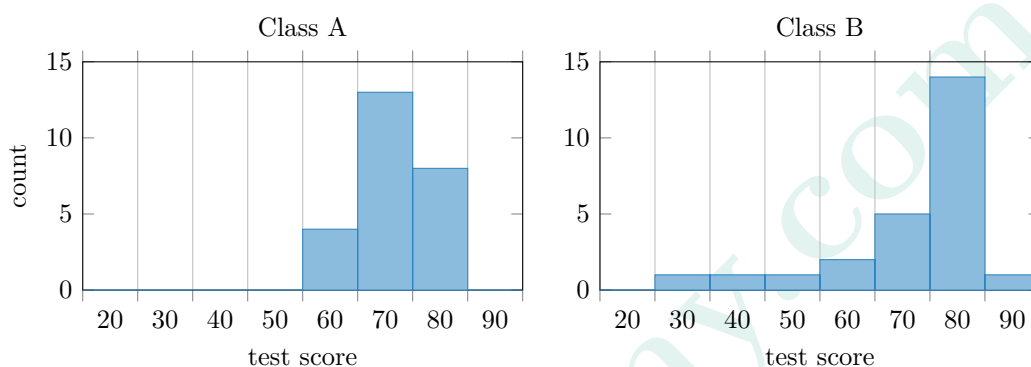


Describing a Distribution

AP Statistics | Unit 1 | shape, outliers, center, spread — in context

1. From data to a picture

The **distribution** of a quantitative variable tells you what values it takes and how often. A **dotplot** puts one dot per individual on a number line. A **histogram** groups values into equal intervals and draws a bar whose height is the count. Histograms are for quantitative data, so the bars touch; bar charts are for categories and have gaps.



Both classes have a mean of 76 (hypothetical data, 25 students each), yet they are very different. One number never describes a distribution.

2. Shape

- **Symmetric:** the left and right sides are rough mirror images (class A).
- **Skewed:** a long **tail** on one side. *The skew is named for the tail, not the peak.* Tail to the right $\Rightarrow$ skewed right (household income); tail to the left $\Rightarrow$ skewed left (class B).
- **Peaks:** one peak = unimodal, two clear peaks = bimodal (often two groups mixed).

3. Center and spread

Center: the **median** is the middle value (the 13th of 25); the **mean** $\bar{x}$ is the balance point of the histogram. In a skewed distribution the tail pulls the mean toward it: in class B the mean is 76 but the median is 81.

Spread: the simplest measure is the **range** = max – min (class B: $92 - 30 = 62$ points; class A: $86 - 63 = 23$). The IQR and standard deviation come in the next topic.

4. Unusual features

Outliers (values far from the rest; class B's 30, 45 and 55 are flagged by the $1.5 \times \text{IQR}$ rule), **gaps** (stretches with no data) and **clusters** (groups of values bunched together). Mention any you see.

Describe with SOCS — and always in context

Shape, Outliers & other unusual features, Center, Spread.

“The distribution of test scores for class B is skewed left, with a median of about 81 points and a range of 62 points. There are three unusually low scores (30, 45, 55) that are possible outliers.”

Where students lose points

1. No context: say “test scores”, not “the data”.
2. Naming the skew by the peak instead of the tail.
3. Leaving out one of shape, outliers, center, spread.
4. Describing the graph (“the bars go up”) instead of the data.
5. Calling every symmetric hump “normal”; say “approximately symmetric”.

Remember

- SOCS + context: shape, outliers, center, spread.
- Skew goes the way the tail goes.
- In a skewed distribution the mean is pulled toward the tail; the median is not.
- Histograms: quantitative data, bars touch.

Now practise it

Theory sticks once you use it. On **sad-academy.com** you’ll find practice questions for this exact subtopic in the AP format, and an AI that grades your free-response answers against the scoring guidelines, the way an AP reader would.

sad-academy.com