

Module 4: Describing Distributions

TOPIC 1: ONE-VARIABLE STATISTICS

In this topic, students begin by representing data in dot plots, histograms, and box-and-whisker plots and think about when these data displays are helpful for a given data set. Students are then introduced to more formal notation for mean: $\bar{x} = \frac{\sum x}{n}$. This notation is important as it prepares students to make sense of the formula for standard deviation that they encounter later in the lesson. Students learn to identify outliers and recognize that median and interquartile range (IQR) are not greatly affected by outliers in a data set. By the end of the topic, students know when and how to describe a data set with mean and standard deviation vs. median and IQR. The final lesson provides students with opportunities to practice comparing two data sets and making a decision based on the comparison.

Where have we been?

Throughout elementary school, students have displayed data using line plots and pictographs. In middle school, students represented data using histograms, dot plots, stem-and-leaf plots, and box-and-whisker plots. They described the shape of the graph in terms of symmetry and skew. They calculated the mean and median to represent the central tendency of the a data set; they calculated the interquartile range (IQR) and mean absolute deviation (MAD) to represent the spread of the data. Students also compared two data sets using these same displays and the corresponding measures of center and spread.

Where are we going?

This topic develops students' statistical literacy as they increase their knowledge of, and the level of complexity of their engagement with, the statistical problem solving process. Students are prepared for the use of the standard normal distribution for calculating the likelihood of a specific outcome, which leads students to significance test, margin of error, and confidence intervals.

Sample Standard Deviation Formula

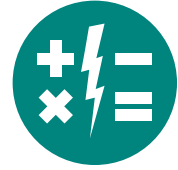
The formula to compute the sample standard deviation can be read as a step-by-step process:

1. $(x_i - \bar{x})^2$: Subtract the mean from each data value and square the result.
2. $\sum_{i=1}^n$: Add up all of the results from Step 1, for data values x_1 to x_n .
3. $(n - 1)$: Divide the result from Step 2 by 1 less than the sample size, n .
4. $\sqrt{\quad}$: Take the square root of the result from Step 3.

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

The Latest Gallup Poll Shows . . .

Should you believe the results of polls reported in the news? Absolutely not—at least, not without researching the information presented. Even though most polls are conducted properly, the results are often not reported properly.



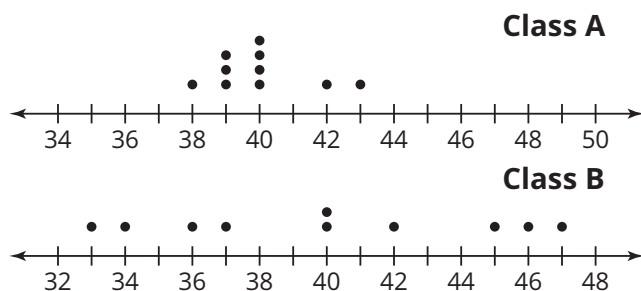
Look closely for the margin of error reported along with poll results. (If there isn't one, it may not be a scientific poll.) If, for example, a political candidate is leading 50% to 47%, and the margin of error is $\pm 3\%$, that means that the 50% could be as low as 47% and the 47% could be as high as 50%.

News organizations will often report this result as though the candidate with 50% is leading, but according to the margin of error, it is not possible to actually make an accurate conclusion like this. The leading candidate may actually be behind!

Talking Points

Statistics is an important topic to know about for college admissions tests.

Here is a sample question:



The dot plots show quiz scores for students in Class A and B. Each dot represents 2 students and each class has 20 students. Which data set has a smaller standard deviation?

To solve this, students should know that standard deviation is a measure of how far data are spread out from the mean. So, Class A has a smaller standard deviation, because the data are all closer to the mean.

Key Terms

interquartile range

The interquartile range, IQR, measures how far the data are spread out from the median. It is calculated by subtracting $Q3 - Q1$ in the five-number summary.

outlier

An outlier is a data value that is significantly greater or lesser than other data values in a data set.

standard deviation

Standard deviation is a measure of how spread out data are from the mean.