

Random Sampling and Comparing Populations: math lesson, Grade 7 – download the PDF

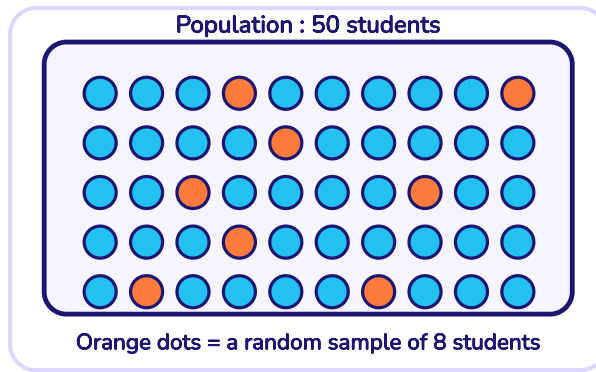
How do pollsters predict an election after asking only a few thousand people?
How does a game company learn what players think without asking every player?
They use **samples**. In this chapter you will learn how to choose a sample fairly, how to use it to make a smart guess about a whole group, and how to compare two groups of data with dot plots, box plots, and measures of variability.

1. Populations and samples

D Population and sample

A **population** is the entire group you want to learn about: people, animals, objects, or measurements. A **sample** is a smaller part of the population that you actually collect data from. The number of items in a sample is its **sample size**.

Suppose a middle school has 640 students and you want to know how many walk to school. Asking all 640 students would take a long time, so you ask 40 of them. The 640 students are the population, and the 40 you ask are the sample.



Sometimes studying the whole population is simply impossible. A company that makes flashlight batteries cannot test every battery to see how long it lasts, because the test uses the battery up! Samples save time and money, and sometimes they are the only choice.

2. Representative and random samples

D Representative sample

A sample is **representative** when its members reflect the whole population: the same mix of ages, opinions, locations, and so on. A sample that leaves out or over-counts some part of the population is **biased**, and conclusions from it can be wrong.

D Random sample

In a **random sample**, every member of the population has the same chance of being chosen, and the choice is made by chance, not by a person's preference. Random samples are usually representative.

Good ways to choose a random sample: write every name on a slip of paper, mix the slips well, and draw; or number the whole roster and let a random number generator pick the numbers.



Watch out for biased methods

Asking only your friends, the first people to walk in, or people who choose to answer an online poll is convenient but *not* random. Those people may think very differently from everyone else. Also remember that a very small sample can be unreliable even when it is chosen fairly: larger random samples tend to give more trustworthy results.

3. Making inferences from a sample

An **inference** is a reasonable conclusion about a population based on data from a random sample. The most common inference uses a proportion.



Estimating a count for a population

1. Find the sample proportion: $\frac{\text{number in the sample with the trait}}{\text{sample size}}$.
2. Multiply that proportion by the size of the population.
3. Say the answer is an *estimate*, because another random sample would give a slightly different result.

E**Example 1: walking to school**

From the 960 students of a school, 80 are chosen at random and 25 of them walk to school. The sample proportion is $\frac{25}{80} = 0.3125$. An estimate for the school is $0.3125 \times 960 = 300$ students.

If three more random samples of 80 students gave 22, 28, and 26 walkers, the estimates would be 264, 336, and 312. The estimates are close to each other but not equal. This shows that sample results **vary**, so we say “about 300 students”, not “exactly 300 students”.

4. Dot plots and box plots

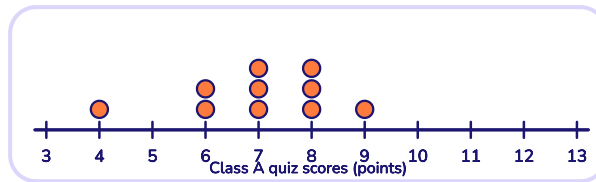
A **dot plot** shows every value as a dot above a number line, so you can see clusters, gaps, and unusual values. A **box plot** summarizes the data with five numbers: the minimum, the first quartile Q_1 , the median, the third quartile Q_3 , and the maximum.

M**Finding the five-number summary**

1. Order the data from least to greatest.
2. Find the **median**, the middle value (the mean of the two middle values if the count is even).
3. Q_1 is the median of the lower half. Q_3 is the median of the upper half. (If the count is odd, leave the median out of both halves.)

Take the quiz scores of Class A: 4, 6, 6, 7, 7, 7, 8, 8, 8, 9. The median is 7, the lower half is 4, 6, 6, 7, 7 so $Q_1 = 6$, and the upper half is 7, 8, 8, 8, 9 so $Q_3 = 8$.

Minimum	Q_1	Median	Q_3	Maximum
4	6	7	8	9



5. Measures of center and variability

A **measure of center** describes a typical value. The **mean** is the sum of the values divided by the number of values. The **median** is the middle value. A **measure of variability** tells how spread out the data are.

- **Range** = maximum — minimum.
- **Interquartile range (IQR)** = $Q_3 - Q_1$, the width of the box, which covers the middle half of the data.

For Class A, the mean is $\frac{70}{10} = 7$, the range is $9 - 4 = 5$, and the IQR is $8 - 6 = 2$.

! Zyro's tip

On my home planet we say: one very strange value can drag the mean around, but the median hardly moves. When a data set has an extreme value, the median is often the better “typical value”.

6. Mean absolute deviation

D Mean absolute deviation (MAD)

The **mean absolute deviation** is the average distance between the data values and their mean. A small MAD means the values are packed close to the mean. A large MAD means they are spread out.

M Computing the MAD

1. Find the mean.
2. For each value, find its distance from the mean (ignore the sign).
3. Add the distances and divide by the number of values.

E Example 2: MAD of five scores

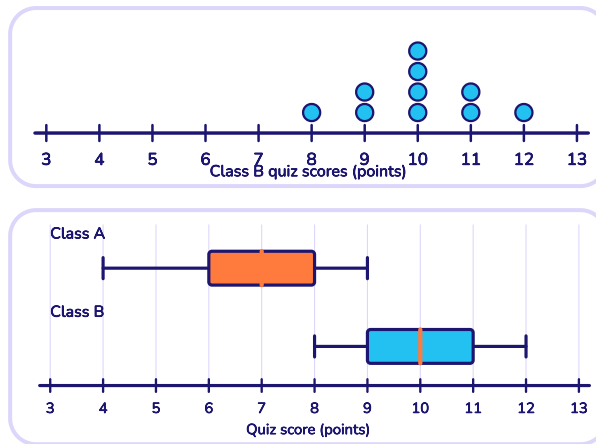
Scores: 4, 6, 7, 9, 9. The mean is $\frac{35}{5} = 7$. The distances from 7 are 3, 1, 0, 2, 2, and their sum is 8. So

$$\text{MAD} = \frac{8}{5} = 1.6.$$

On average, a score is 1.6 points away from the mean.

7. Comparing two populations and their overlap

To compare two groups, look at **both** the centers and the variability. Two groups can have the same mean but very different spreads, or different means with a lot of overlap.



P

Measuring the gap in MADs

Divide the difference between the two centers by a measure of variability (such as the MAD). If the result is about 2 or more, the two distributions are fairly well separated, with little overlap. If it is less than 1, they overlap a lot, and the difference may not mean much.

E

Example 3: Class A versus Class B

Class A scores have mean 7 and MAD 1. Class B scores (8, 9, 9, 10, 10, 10, 10, 11, 11, 12) have mean 10 and MAD 0.8. The means differ by $10 - 7 = 3$ points. Since $3 \div 1 = 3$, the gap is about 3 MADs, so the distributions barely overlap: almost every Class B score is higher than almost every Class A score. The medians (7 and 10) and the box plots tell the same story.

Remember that these are small samples. If the classes were chosen randomly from a larger population, we could infer that students like those in Class B tend to score higher, but another sample might show a slightly different gap.

Key takeaways

- A population is the whole group; a sample is the part you study.
- A random sample gives every member an equal chance. It is usually representative; convenience samples are often biased.
- Estimate a population count with: sample proportion $\times$ population size. Different samples give different estimates.
- The five-number summary (minimum, Q_1 , median, Q_3 , maximum) builds a box plot; $IQR = Q_3 - Q_1$.
- The MAD is the mean of the distances from the mean. It measures variability.
- To compare two distributions, compare centers and spreads, and express the gap between centers in MADs.



Test yourself: quick challenge for Grade 7

[Change level](#)

Level : **Grade 7**

6 random questions, instant feedback.



Start the challenge



Speed drill for Grade 7: how many in 60 seconds?

 Keep exploring with Zyro

 Math practice

Random Sampling and Comparing
Populations: math practice, Grade 7

 Math tests

Random Sampling and Comparing
Populations: math test, Grade 7

 Math quizzes

Random Sampling and Comparing
Populations: math quiz, Grade 7

 Math practice

Probability and Compound Events: math
practice, Grade 7

 Math practice

Ratios and Proportional Relationships:
math practice, Grade 7

 Math tests

Probability and Compound Events: math
test, Grade 7

Zyro vous emmène sur nos sites partenaires

Embarquez pour d'autres planètes du savoir et leurs applis.

Décollage