

Collecting and Organizing Data

Goals

- Understand that a sample can be used to collect data about a larger population.
- Collect, organize, and analyze data.

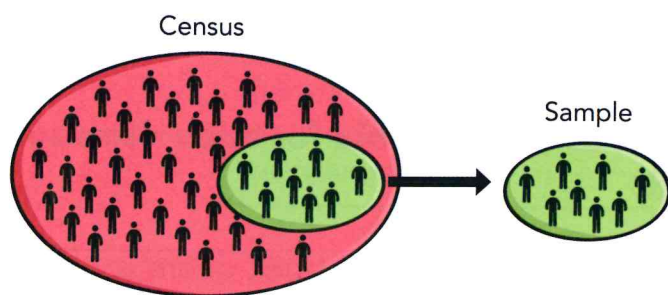
Vocabulary

population	biased
sample	stratified random sampling
simple random sampling	random sample
secondary data	systemic random sampling
census	

LEARN Distinguish between a census and a random sample.

Ayesha wants to collect primary data to determine whether her community supports building a skateboard park. The **population** Ayesha wants to collect data from is all the people in her town.

Ayesha could do a survey of everyone in the community. This is called a **census**. But a census takes a lot of time and work. Instead, Ayesha decides to survey a **sample** of 100 people who represent the population.



A sample needs to fairly represent the population. If it represents only part of the population or a certain point of view, the survey is **biased**.



Ayesha uses a **random sample** to make the survey fair. There are 3 random sampling methods.

Simple random sampling chooses names randomly from the entire population. For example, Ayesha could put the names of all the people in her community in a box and draw 100 names.

It's important that the sampling is random so that people can't say that the survey creator only chose people who would respond in a certain way.



Stratified random sampling divides the population into groups based on different characteristics. For example, Ayesha could place names into boxes based on age. If she had 10 age groups, she would select 10 names from each age group. That would give her a total sample of 100 people.

Systematic random sampling chooses 1 person randomly and then selects other people at regular intervals. For example, if Ayesha had a list of students in her school, she could randomly choose 1 name on the list. She might then choose every 5th name on the list following that name until she had 100 names.

GUIDED LEARNING

Describe a random sampling method that is suitable for each situation. Explain why you chose that method.

- 1 A bookseller would like to determine what the favourite book genre is of readers who live near the store.
- 2 The school cafeteria wants to add menu items that students will like.
- 3 A car rental company would like to determine which type and brand of car people would most like to rent.

LEARN Organize and analyze data.

Ari asked all the students in his class how far they walked in a charity walkathon. He created a frequency table to organize the results.

Distances Walked (Ari's Class)

Distance (km)	Frequency (number of students)
4.5	1
5	4
5.5	6
6	5
6.5	2

Frequency is the number of times a piece of data occurs in a data set.

Source: Ari's survey of his class

Ari used his frequency table to determine the mean, mode, and median. There are 18 students in his class.

The **mean** is the average of all distances. The sum of distances is 100.5 kilometres.

$100.5 \text{ km} \div 18 \approx 5.6 \text{ km}$
The mean is approximately 5.6 kilometres.

The symbol $\approx$ means *approximately equal to*. We use this symbol when we round an answer.



The **mode** is the most frequent distance, 5.5 kilometres.

The **median** is the middle distance. Since there are 18 numbers, the data has 2 middle numbers, 5.5 and 5.5. The median is 5.5 kilometres.

Ari wants to know how the walkathon distances for his class compare with the distances for the whole school. He decides to use a stratified random sample to represent the school population.

- STEP 1** Ari places the names of all Grade 1 students in a bag and labels the bag *Grade 1*. He creates a similar bag of names for each grade, from Grade 2 to Grade 6.
- STEP 2** Ari randomly draws 10 names from each bag to create the population for his survey. His population is 60 students.
- STEP 3** Ari asks the students whose names he drew how far they walked in the walkathon. He records their answers in a frequency table.

Here is the data Ari collected:

Distances Walked (Whole School)

Distance (km)	Frequency (number of students)
2.5	7
3	8
3.5	8
4	7
4.5	7
5	7
5.5	5
6	5
6.5	3
7	3

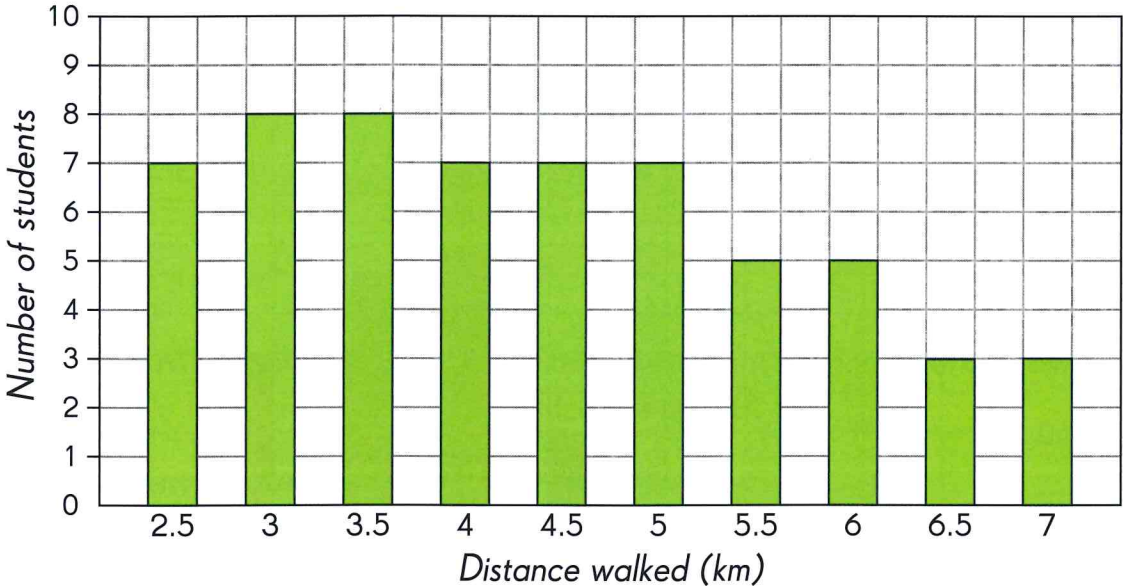
Source: Ari's survey of his school

A census of a large population takes a lot of time and work. A random sample is quicker and easier to do.



Ari organizes the school survey results in a bar graph.

Distances Walked (Whole School)



Source: Ari's school survey

Most students walked between 2.5 and 5 kilometres.

Mean

Number of students: $7 + 8 + 8 + 7 + 7 + 7 + 5 + 5 + 3 + 3 = 60$.

The sum of distances is 262 kilometres.

$262 \text{ km} \div 60 \approx 4.4 \text{ km}$

The mean is approximately 4.4 kilometres, which is less than the mean of 5.6 kilometres for Ari's class.

Mode

The mode distances are 3 kilometres and 3.5 kilometres.

This is less than the mode of 5.5 kilometres for Ari's class.

Median

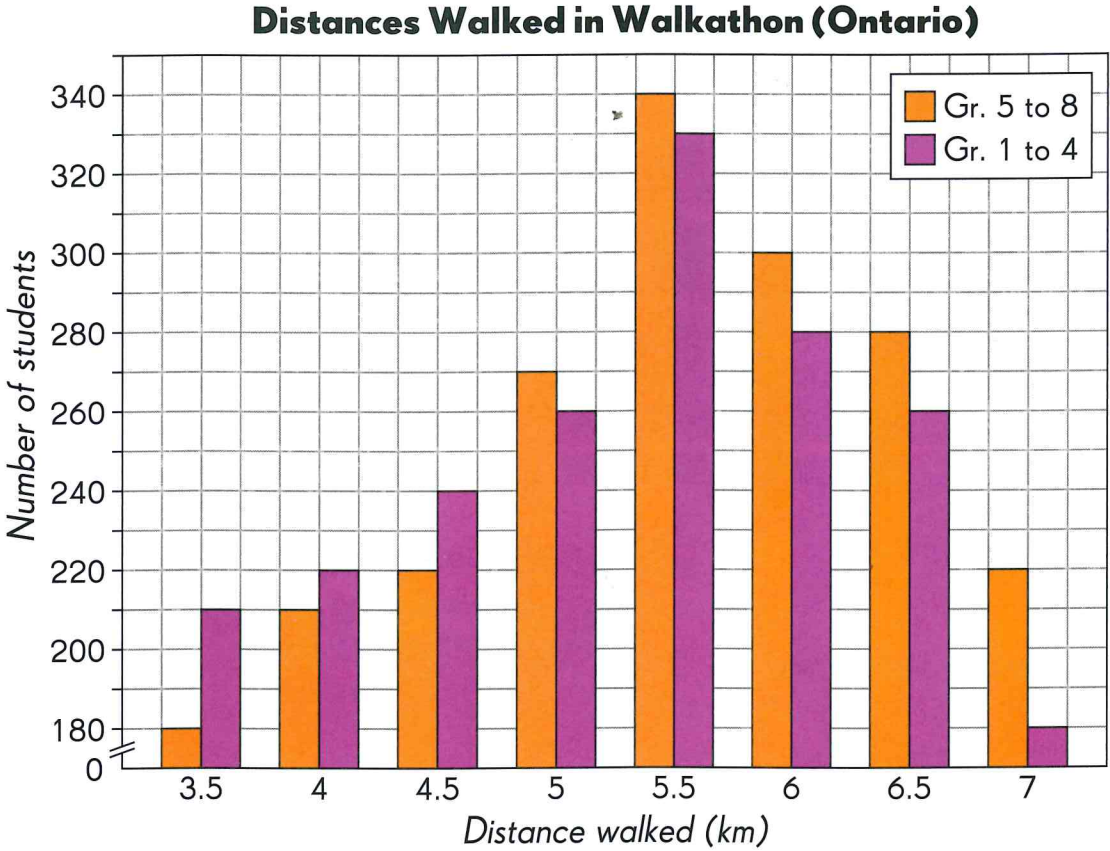
The total number of students is 60. $60 \div 2 = 30$, so the middle distances are the 30th and 31st numbers.

Add the bar heights to identify the 30th response:

The 30th response is at the top of the 4 bar, and the 31st response is at the bottom of the 4.5 bar. The median is halfway between 4 and 4.5, or 4.25.

This median distance is less than the median of 5.5 kilometres for Ari's class.

Students from all over Ontario participated in the walkathon. Ari found this **secondary data** showing some results from other schools.



Source: Walkathon website

Add the heights of the bars to find the number of students.
There are 4000 students.

The total distance is found by multiplying the height of a bar by the distance the bar represents. The total distance is 21 280 kilometres.

The mean for students on this graph is $21\,280 \text{ km} \div 4000 \approx 5.3$ kilometres.

The mean of 5.3 kilometres is similar to the mean distance Ari's class walked.

Across Ontario, most students walked between 4.5 kilometres and 6.5 kilometres. This matches the results for Ari's class.

The mode for each group was 5.5 kilometres. This also matches Ari's class.

GUIDED LEARNING

Answer the questions.

Ms. Jackson wants to determine whether she gives her students too much homework.

- 4 What population does Ms. Jackson want to gather data about?
- 5 Should Ms. Jackson use a census or a sample?
How did you decide?
- 6 Ms. Jackson asked some students how long they spent on homework last week. She created a frequency table to organize the results.

**Hours Spent on Homework
Last Week**

Hours spent	Frequency
0.5	3
1	2
1.5	5
2	5
2.5	4
3	4
3.5	2

Source: Ms. Jackson's survey

- a What is the mean of the data set? hours
- b What is the mode of the data set? and hours
- c What is the median of the data set? hours

- 7 The vice-principal wants to gather data about homework from the whole school. How can she select a random sample?
- 8 This table shows the school results. Make a bar graph to show the data.

Hours Spent on Homework Last Week

Number of hours	1	1.5	2	2.5	3	3.5
Number of students	5	9	11	15	14	6

Source: Survey of students

Remember to give your graph a title and label the axes. Show the number of hours on the horizontal axis and the number of students on the vertical axis.



- 9 What information can you tell from the bar graph?
- 10 What kind of graph can you use to compare this data with similar data from another school?