3 One Variable Statistics

Statistics is the process of collecting, displaying, and analysing data. Data can be gathered by conducting an experiment or a survey. In this chapter you will look at survey design, ways of displaying data, and the measures used to analyse data.

In this chapter, you will

- identify situations involving one variable data, and design questionnaires or experiments for gathering one variable data
- collect one variable data from secondary sources, organize and store the data using a variety of tools, and solve problems by analysing the data
- explain the distinction between the terms *population* and *sample,* describe the characteristics of a good sample, and explain why sampling is necessary
- describe and compare sampling techniques, collect one variable data from primary sources, and organize and store the data
- identify different types of one variable data and represent the data, with and without technology, in appropriate graphical forms
- identify and describe properties associated with common distributions of data
- calculate and interpret measures of central tendency and spread; use these measures to compare sets of one variable data
- explain the appropriate use of measures of central tendency and measures of spread

Reasoning and Proving
Representing | Selecting Tools
Problem Solving
Connecting | Reflecting
Communicating

Key Terms

bias	mean	quartiles
bimodal distribution	median	range
box-and-whisker plot	mode	sample
categorical data	normal distribution	secondary source
continuous data	outlier	skewed distribution
discrete data	population	standard deviation
interquartile range	primary source	variance

Craig completed the three-year manufacturing technology program at Sheridan College. He now works as an industrial engineering technologist. Craig analyses the production process to improve factory efficiency. He decides the number of workers needed on the production line, where to place the workers, and how to organize their work.

Prerequisite Skills

Numeracy Skills

1. Order the numbers in each set from least to greatest.

   a) 5, 4, 11, 9, 15, 7, 8, 1, 5, 19, 4, 9

   b) 3, 3, 5, 7, 1, 3, 5, 6, 9, 11, 12, 6, 4, 2, 2, 1

   c) 5, 8, 1, −3, 0, 12, −11, 4, 5, 1, 7, −3, −5

   d) $\dfrac{1}{2}, \dfrac{5}{8}, \dfrac{1}{4}, \dfrac{3}{16}$

2. Evaluate. Round your answer to two decimal places, if necessary.

   a) $\dfrac{14 + 16 + 22 + 75 + 85 + 35}{6}$

   b) $\dfrac{5 + 11 + 4 + 7 + 15 + 12 + 22 + 8}{8}$

   c) $\dfrac{77 + 54 + 71 + 51 + 64 + 88 + 90}{7}$

   d) $\dfrac{38 + 45 + 56 + 76 + 83}{5}$

   e) $\dfrac{2.2 + 2.4 + 3.7 + 7.1 + 4.5 + 5.9 + 7.7 + 1.3}{8}$

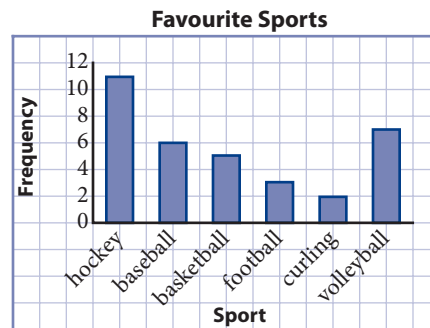3. Evaluate. Round your answer to two decimal places, if necessary.

   a) $\sqrt{121}$

   b) $\sqrt{47}$

   c) $\sqrt{3^2 + 4^2}$

   d) $\sqrt{1^2 + 2^2 + 3^2}$

Interpret Graphs

4. The graph shows the favourite sports of students in Minh's physical education class.
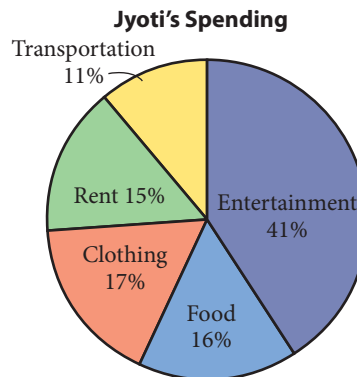
   Favourite Sports

   a) Identify the type of graph.

   b) Which sport is the most popular?

   c) Which sport is the least popular?

   d) Does "least popular" mean the same as "not popular"? Explain.

5. The circle graph shows Jyoti's spending pattern over the last three months.

   Jyoti's Spending


   a) What was Jyoti's greatest expense?

   b) If Jyoti earned $1220 over the last three months, how much did she spend on clothing?

6. The graph shows the average weekly temperature at a Caribbean island resort for four months. The average for the ninth week is missing.

Resort Temperature
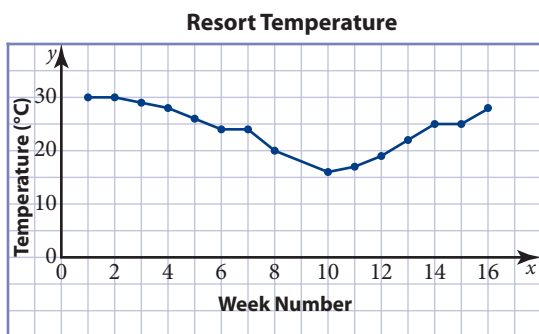
a) Identify the type of graph.

b) Estimate the average weekly temperature for the ninth week.

c) The graph ends on the sixteenth week. Predict the average weekly temperature for the seventeenth week. Justify your answer.

d) What information would you need to check your answer in part c)?

7. The graph shows the heights of the students in a club.

Heights of Students
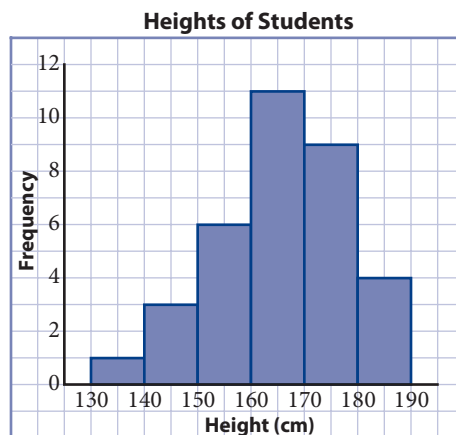
a) Identify the type of graph.

b) How many students are in the club?

Chapter Problem

Andrew is very interested in statistics, so he took a co-op placement at a market research company to learn more about the field. Throughout this chapter, you will see how Andrew uses mathematics in his job as a market researcher.

3.1 Sampling Techniques

When conducting a survey, it is important to choose the right questions to ask and to select the appropriate group to survey.

Before a market research department designs a survey, the target group for the survey needs to be identified.

Use a Random Number Generator

There are 300 students taking mathematics this semester at Royal Secondary School. The table shows the number of mathematics students in each grade.

Grade	Number of Mathematics Students
9	100
10	80
11	75
12	45

Mrs. Barron, the principal, wishes to survey a total of 60 mathematics students. She has chosen 20 grade 9s, 16 grade 10s, 15 grade 11s, and 9 grade 12s.

1. Calculate the percent of the population that each grade represents.

The percent for grade 9 is done for you.

$$\frac{100}{300} \times 100 = 33.\overline{3}\%$$

2. Calculate the percent of the sample that each of Mrs. Barron's choices represents.

The percent for grade 9 is done for you. $\frac{20}{60} \times 100 = 33.\overline{3}\%$

How does the proportion of students from each grade in the group Mrs. Barron selected compare to the proportion of mathematics students in each grade? Explain.

3. Mrs. Barron decides to select each student for the sample using a random number generator on her graphing calculator. Help her select the grade 9 students.

- Clear the lists on the graphing calculator. Press ⟨ 2nd ⟩ [MEM] **4**, then ⟨ ENTER ⟩.
- Press ⟨ STAT ⟩, and select **1:Edit…**.
- Use the arrow keys to highlight L1.
- Type the month and day of your birthday, followed by the street number of your home, using all digits. Press ⟨ STO▸ ⟩. Press ⟨ MATH ⟩ and cursor over to **PRB**. Press ⟨ ENTER ⟩ twice. Press ⟨ MATH ⟩ and cursor over to **PRB**. Select **5:randInt(** and press ⟨ ENTER ⟩.
- Type 1 ⟨ , ⟩ 100 ⟨ , ⟩ 20 ⟨) ⟩ ⟨ ENTER ⟩.
- The calculator will generate 20 random integers between 1 and 100 and display them in list L1.
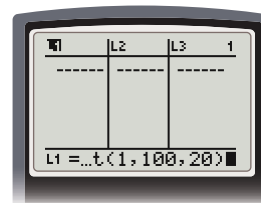
4. Reflect Mrs. Barron could have numbered all 300 mathematics students and then used the formula **randInt(1,300,60)** to choose her sample. What are some difficulties that Mrs. Barron might have encountered using this method?

population
- all individuals or items that belong to a group being studied

sample
- a group of individuals or items that are representative of the population from which they are taken

It is usually impractical to survey every member of a target group or **population**, so only a **sample** of the population is surveyed.

There are a number of different ways to choose a sample.
- **Simple Random Sample** Each member of the population has an equal chance of being selected.
- **Stratified Random Sample** The population is divided into subgroups (for example, by gender, age, nationality) and a random sample is selected from each subgroup in proportion to its size in the population.
- **Voluntary-Response Sample** The sample contains those members of the population who have chosen to respond to the survey.
- **Cluster Sample** The population is divided into clusters and a certain number of clusters are chosen. Every member of these clusters is part of the sample.
- **Convenience Sample** The sample contains those members of the population from which data are most easily collected.
- **Systematic Sample** Every *n*th member of the population is selected.

Example 1

Population and Sample

Meena wants to know which band Ontario high school students think is the best. Meena's friend Cindy goes to a different school, so they each survey students at their own school. Meena uses the completed surveys from both schools to draw conclusions.

a) Identify the population and the sample.

b) Is the sample representative of the population? If it is representative, explain why. If it is not, suggest how the sample could be improved.

Solution

a) The population is all Ontario high school students. Meena's sample of the population is the students surveyed in both her school and her friend's school.

b) The students Meena and Cindy surveyed might be a representative sample of the population of the two schools, but there are approximately 5000 schools in Ontario. The sample size from the two schools is too small to draw conclusions about students in the entire province. Meena should include more schools from different areas of the province to make her sample more representative.

bias
- a survey contains bias if it does not reflect the population
- may be caused by an unrepresentative sample, the wording of the survey questions, and/or the interpretation and presentation of the results

When the results of a survey do not reflect the population, the survey is said to contain **bias**.

Bias may be introduced by selecting a sample that is not representative of the population, by the wording of the survey question, or by the presentation of the results of the survey.

Example 2

Choose a Sampling Technique

Determine the best sampling technique for each survey. Describe one method of selecting the sample. Identify any problems with using the sampling technique.

a) The school newspaper wants to determine which presidential candidate in the upcoming student council elections is supported by the majority of students.

b) A light bulb manufacturer wants to determine the lifespan of a certain type of light bulb, in hours.

c) The Parent–Teacher Association wants to determine the average number of hours per week that students spend on homework.

d) The producers of "Rock Idol" want to determine which of the two remaining candidates should be the next rock idol.

Solution

a) All students can vote in the elections, so use a sampling technique that represents the entire student population. Choose simple random sampling. Use a computer or graphing calculator to randomly select about 20% of the school's population for the sample. This sampling technique might be inaccurate if students in certain grades are unlikely to vote in the actual election.

b) Light bulbs are identical (unless there are manufacturing flaws), so the sampling technique does not need to be random. Choose systematic sampling. Select every hundredth light bulb for the sample. This sampling technique might be inaccurate if the manufacturing process is not standardized.

c) Students in different grades have different amounts of homework, so use a sampling technique that represents all of the grades. Choose stratified random sampling. Select a random sample of students from each grade for the sample. This sampling technique might be inaccurate if the randomly selected students spend more (or less) time on homework than their peers, or if certain classes have more (or less) homework than other classes in the same grade.

d) While the producers of "Rock Idol" know the number of people who watch the show, it would be very difficult to contact a representative sample of viewers, so use a sampling technique in which the viewers contact the producers. Choose voluntary-response sampling. Viewers who call a toll-free number with their opinion make up the sample. This sampling technique is likely to introduce bias because viewers can call more than once and some viewers might not respond.

Math Connect

When the population of a study is very large, it takes a large organization (such as Statistics Canada) or a large company (such as Ipsos Reid) to find a representative sample and conduct the survey.

Key Concepts

- In many situations it is not practical or possible to survey an entire population, so a representative sample of the population is selected.
- There are a variety of sampling techniques, each with its strengths and weaknesses.
- The choice of sampling technique depends on the population and the type of survey being conducted.

Discuss the Concepts

D1. Letasha surveyed her friends to determine the popularity of a new movie. What type of sampling did Letasha use? Why might Letasha have used this type of sampling?

D2. Which sampling techniques are most likely to introduce bias? Explain.

D3. Describe some advantages and disadvantages of each sampling technique.

Practise A

1. You are stopped in the mall and asked to participate in a survey. Which sampling technique is being used?

2. Mr. Rush's grade 5 students want to play soccer. Mr. Rush splits the class into girls and boys, numbers each group from 1 to 10, and then forms two teams: one with even-numbered students and one with odd-numbered students.

   a) Explain the sampling technique Mr. Rush used to create the teams.
   b) Is this a fair way of choosing teams? Explain why or why not.
   c) Suggest an improvement to the sampling technique Mr. Rush used.

3. Choose the best sampling technique for each survey. Explain your choice.

   a) Umar wants to know at which festivals his dance company would like to perform.
   b) Sherry wants to find out which Canadian actress is most popular with Ontario teens.
   c) Laurel wants to know which actor her classmates think is the best performer.
   d) Bruno wants to know which search engine is used most often by Ontario high school students.

4. In each situation, identify the sample and the population.

   a) Zoe noticed that most teenagers at the local library were reading murder mysteries. She suggested to the student council at her school that they organize a school murder mystery read-a-thon.

   b) Enrica and her friends call the "Canadian Idol" show several times a week to vote for the same performer.

   c) The Canadian Dental Association surveys several outlets of a popular drug store chain to collect information on the most popular toothpaste.

   d) Tony is working for a political party. He stands at a busy street corner asking people whom they plan to vote for in the upcoming election.

   e) Angelo is the manager of a gym. He asks members who are at the gym on Wednesday night if he should have a party for gym members.

   f) Maya surveyed her science class to identify the rock group most listened to by the students in her grade.

Apply **B** •

5. The Canadian government conducts a country-wide survey called a census every five years. The survey is designed to collect information on every Canadian.

   a) Which Canadians will not be included in this census?

Literacy Connect

   b) Use the terms *population*, *representative sample*, and *sampling technique* to explain why the government does not conduct a census every year.

   c) If 3% of the population were not included in the survey, can the government still make valid statements regarding the population? Explain.

6. Mykele used a random number generator to select numbers from 1 to 10. The first three numbers she recorded were 7, 7, 7. Derek thinks Mykele made a mistake. Is Derek correct? Explain.

7. Andrew has his first assignment at the market research company. He is helping to conduct a survey on what Ontario high school students do after finishing high school.

a) Identify the population of the survey.

b) Why is it impractical to survey everyone in the population?

c) If the whole population were surveyed, how could this be done effectively?

d) What sampling technique could Andrew suggest that the company use for the survey? Explain how this sample could be chosen.

8. Jeff works as a quality control officer at a bolt manufacturing plant. Each day, he randomly selects and tests 100 bolts. On Monday, Jeff finds that 11 of the 100 bolts are defective. On Tuesday, he finds that 4 of the 100 bolts are defective. For the rest of the week, he finds no defective bolts in his samples.

a) What type of sampling technique is Jeff using?

b) Jeff finds varying numbers of defective bolts over the week. Does this mean that his sampling technique is not effective? Explain.

c) How could Jeff improve the information that he will get from his data?

9. Alice wants to know if the students at her school are in favour of a school uniform. She surveys 30 students from grade 9, 25 students from grade 10, 25 students from grade 11, and 20 students from grade 12. The table shows the numbers of students in each grade at Alice's school.

Grade	Number of Students
9	307
10	242
11	230
12	212

a) Is the sample representative of the population? Justify your answer.

b) Is the sampling method appropriate for the situation? Explain.

10. You want to know which types of music students at your school prefer.

a) Explain how to obtain a sample of students using each sampling technique.
  i) random sampling
  ii) cluster sampling
  iii) stratified random sampling
  iv) convenience sampling

**b)** Which technique would be best for this situation? Justify your answer.

Extend **C**

11. Keira works for the Canada Customs and Revenue Agency at a busy border crossing. She needs to search 10 randomly selected vehicles over a 10-h period. To choose the vehicles, she randomly selects a time from 2 P.M. to midnight. She does this 10 times. For each time, she randomly selects one of 20 gates. The vehicle at that gate and time will be searched.

a) What type of sampling is Keira using? Justify your answer.

b) Use technology to make a table of random times and gates.

c) How likely is it that the same time and gate will occur twice? Explain.

12. A school board wants to select 200 elementary students to test a new mathematics home study program for grades 4 to 7. The board numbers the elementary schools from 1 to 35 then uses a random number generator to choose two schools. Each school selects 100 students. Students are randomly selected from each grade by applying the percent of students in each grade to the number who will use the program.

a) Identify the sample and the population in the study.

b) The first school has chosen 100 students: 20% are grade 4s, 25% are grade 5s, 30% are grade 6s, and the rest are grade 7s. How many students from each grade will participate in the study?

c) The second school has 500 students in grades 4 to 7. If there are 100 grade 4s, 100 grade 5s, 150 grade 6s, and the rest are grade 7s, how many students from each grade will be chosen for the study?

d) Explain a method to determine which students at each school will participate in the study.

3.2 Collect and Analyse Data
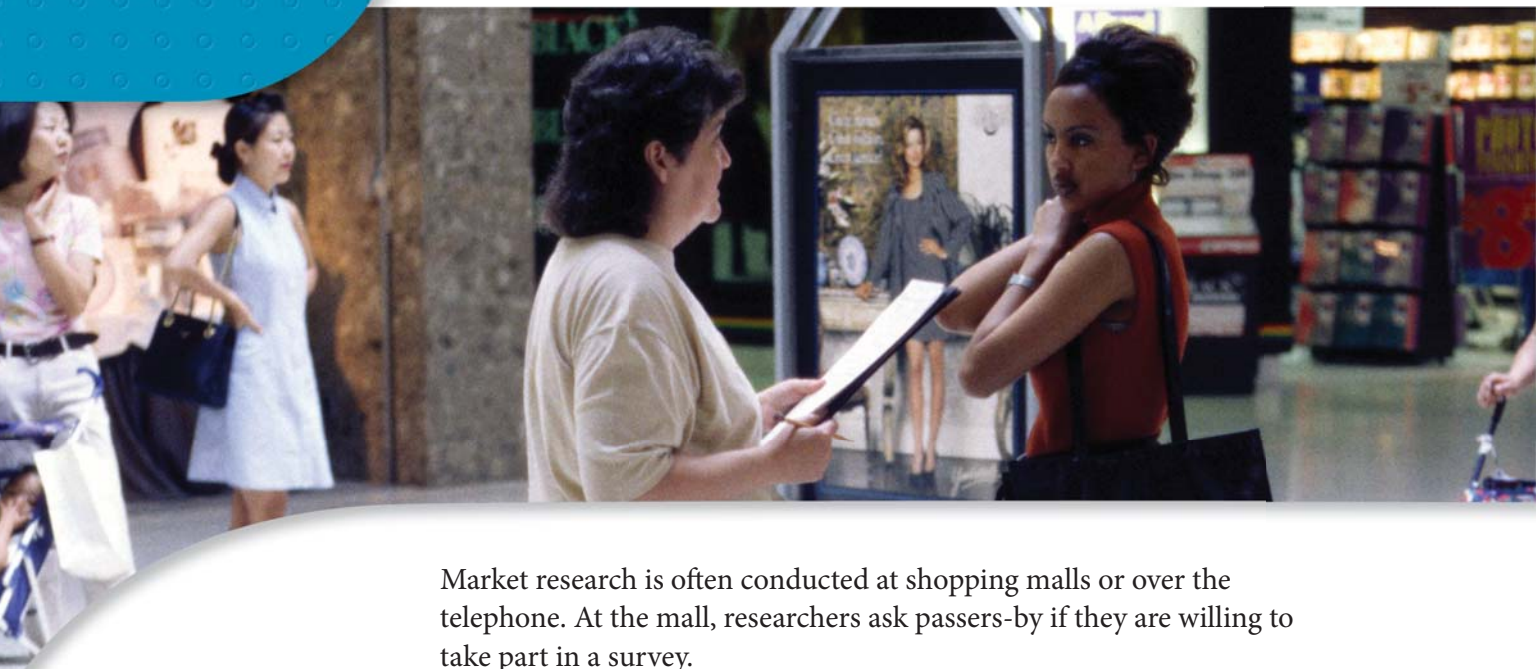
Market research is often conducted at shopping malls or over the telephone. At the mall, researchers ask passers-by if they are willing to take part in a survey.

Investigate 1

Design a Survey Question

Design a one-question survey. Choose a topic from the list below or a topic of your own interest (with your teacher's approval).

- favourite band or singer
- favourite television show
- favourite sport
- favourite dinner
- favourite snack

1. What is the population for your survey?

2. How will you select a sample of the population?

3. Write a survey question to find data about your topic.

4. Exchange survey questions with a partner. Comment on your partner's survey design and return it to your partner.

5. Use the suggestions made by your partner to refine your survey question.

6. **Reflect** People sometimes might not answer a survey question according to their own viewpoint if the nature of the question, or the way the question is asked, influences their opinion. This response bias produces inaccurate survey results. Study your own question and your partner's question. Discuss the way each question is written and decide if the question may result in response bias.

- a person who collects data for their own use

- a database or research collected by someone else

When you conduct a survey or perform an experiment, you are the **primary source** of the data. When you work with data from the Internet, published materials, or Statistics Canada, you are using a **secondary source**.

Investigate 2

Data Gathered From Secondary Sources

In this investigation, you will be looking at the Statistics Canada Web site for data collected and summarized by the Canadian government.

1. To access the Statscan Web site, go to *www.mcgrawhill.ca/links/foundations11* and follow the links.

2. Select English.

3. From the left sidebar, select the Summary tables.

4. From the left sidebar, select Tables by… • subject.

5. Select a subject that interests you. Then, select a subtopic to access a data table. For example, you might choose "Population and demography," then "Population by ethnic origin" or by province and territory (2001 Census).

Source: Statistics Canada, http://www.statscan.ca

6. Write a brief summary of your findings.

7. Share your findings with a partner.

8. **Reflect** What are some advantages of using secondary sources? What are some disadvantages?

Example 1

Response Bias

A parent council survey is conducted to learn if an after-school music program should be offered. The survey question reads: *Early musical training helps develop brain areas involved in language and reasoning. Should the school offer an after-school music program?*

a) Explain why this question may result in response bias.

b) Rewrite the question to eliminate possible response bias.

Solution

a) By stating an opinion about music training in the first sentence, before asking the question, the survey makes it clear that the parent council wants a positive answer.

b) A possible question might be: *Would you enrol your child in an after-school music program*?

Example 2

Measurement Bias

Andre wrote this survey question: *Who do you think is the best female tennis player of all time?*

a) *Martina Navratilova*

b) *Chris Evert*

c) *Billie Jean King*

d) *Serena Williams*

e) *Other:* _____

Identify any weaknesses in this survey question.

Solution

By naming specific players, the survey influences the respondents' answers. Respondents might not take the time to fill in the "Other" selection and might randomly select a player or simply choose a name they recognise. This measurement bias might cause an exaggeration or an underestimation of the results for the population.

Example 3

Non-Response Bias

At a new restaurant, wait staff ask every fifth customer to answer a questionnaire about food quality and service once they have finished their meal. About 30% of customers surveyed fill out the questionnaire. The majority of customers who fill out the questionnaire complain about poor service. The restaurant manager concludes that the wait staff need more training. Is her conclusion reasonable?

Solution

The conclusion is not reasonable because only 30% of those surveyed answered the questionnaire. Customers who did respond might have done so because they were unhappy with the service. Those who were happy with the service might have simply ignored the questionnaire because the service did not need to be improved. This non-response bias leads to inaccurate results because only a small number of people in the sample respond.

Key Concepts

- Someone who collects data for their own use is a primary source.
- A database or research collected by someone else is a secondary source.
- Bias is an intentional or unintentional distortion of the data collected in a survey. A survey should be as free of bias as possible.

Discuss the Concepts

D1. Discuss the bias that will result from each situation. Suggest ways to remove the bias.

   a) Residents of a community are asked this survey question:
   Building a garbage incineration power plant in the neighbourhood will increase jobs and encourage government funding of other programs beneficial to our community. Would you be in favour of this much needed venture?

   b) A group of professional football players are asked if they have ever taken banned performance enhancing substances.

   c) The members of an elite golf club are asked if they will approve the construction of a subsidized housing complex on the land adjacent to their golf course.

D2. Discuss the advantages and disadvantages of using primary and secondary data sources. Give examples of when each type of data source should be used.

1. Dharma wants to know what the best-selling lunch item in the cafeteria is. She records what she sees on the students' trays as they pass by the cashier after paying for their lunch. Her school has two lunch periods. When she has completed a survey during her lunch period, she asks the cafeteria staff to keep track of the items they sell during the other lunch period. She collects the information after school.

   a) Which data are primary data? Which data are secondary data? Justify your answers.

   b) For this survey, is one data source more accurate than the other? Explain.

2. A political candidate wants to know how the members of his riding feel about a major issue in the upcoming election. He and his team send surveys to every even-numbered address in the riding that has the word *Street* in the address. Each survey contains a self-addressed stamped envelope for returns. Two weeks later, 56% of the surveys have been returned.

   a) What type of sampling technique was used?

   b) Do you think the sampling technique is effective? Explain.

   c) How could the candidate increase the survey return rate?

   d) Do you think the candidate will have enough data to draw accurate conclusions? Explain.

3. Identify the bias in each survey. Suggest how it might be removed.

   a) A Canadian Football League (CFL) team hands out a survey at a home game. The survey asks: *Should the provincial and federal governments help build a new stadium and save the team from being relocated?*

   b) A radio talk show host asks listeners to phone in and express their opinions on an issue.

   c) A market research company mails surveys to 1000 households and 200 are returned. The company feels that 200 is a good return and that the opinions expressed can represent the sample of 1000.

   d) Shoppers at a mall are asked: *Are you against the poor decision made by a developer to close this mall in order to build a subdivision, making the traffic in the area even more congested?*

4. The student council is planning events for Spirit Week. They send out this survey question: *Which of the following events would you like to participate in during Spirit Week?*

- *carnival day*
- *read-a-thon*
- *music at lunch*
- *after-school sports*

a) Which type of bias does this survey illustrate?

b) How can the bias be removed? Give examples.

Apply **B** ●

5. Your student council wants to know how the $2000 raised in the last fundraiser should be spent. They ask you to survey the students in the school. Write a survey question that is free from bias. Describe how you would conduct the survey and collect the results.

Literacy Connect

6. There are many bestseller lists for fiction and non-fiction books. Find a bestseller list in a newspaper or on the Internet.

a) How often is the list updated?

b) Describe how the list is presented.

c) Suggest another way to present the information in the list.

Reasoning and Proving

Representing — Selecting Tools

Problem Solving

Connecting — Reflecting

Communicating

7. A community radio station asks its listeners to call in and vote on building a new swimming pool in town. The survey begins just after the 8 A.M. news. By noon, about 50 people have responded, with 30% in favour and 70% opposed. The station reports these interim results at the end of the noon news.

By 4 P.M., when the phone-in ends, the numbers have shifted to 60% in favour and 40% opposed.

a) Could reporting the interim results have affected the outcome of the survey? Explain.

b) Suppose you were in favour of the pool and heard the results at noon. What action would you have taken?

8. Air traffic control services in Canada were privatized a number of years ago. Transport Canada has invited pilots to anonymously report on their experiences with the privatized service. Since reporting is anonymous, can Transport Canada conclude that reports will be unbiased? Justify your answer.

9. a) "Who is the most popular act in the music industry today?" Suggest one way to collect primary data to answer the question. Suggest one way to collect secondary data to answer the question.

b) Which type of data would be most reliable? Explain.

10. Spyware is software that secretly sends information about your Internet surfing habits to a Web site. This statistical data is often collected without the knowledge or consent of the user.
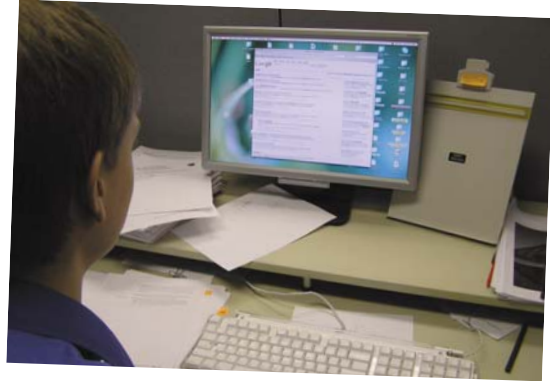
a) Why would companies want to obtain data on Internet surfing habits without the users' knowledge?

b) Is data collected using spyware biased? Justify your answer.

c) With a classmate, discuss other forms of data collection that invade a person's privacy. List them.

11. Andrew is helping to conduct a second survey. The survey question asks: *What is your favourite sport?* Describe any problems associated with the survey question.

12. The student council wants to know the type of music that should be played at the next school dance. Terence and Linda are asked to collect information.

- Terence uses the Internet to find the top 10 songs on 15 radio stations. He assigns 10 points to the number one song on each list, 9 points for the second place song, 8 points for the third, and so on. He determines the point rating for each song and uses the ratings to develop his playlist.

- Linda decides to survey the school population. She creates a questionnaire asking students to list their top 5 songs. The homeroom teachers distribute the questionnaire and give students 5 min to complete it. Linda collects the questionnaires and ranks the songs according to the results to develop her playlist.

a) Identify the type of data source each student used.

b) Outline the strengths and weaknesses of each person's sampling technique.

c) How could Terrence and Lisa improve their results?

13. A marketing company gave four dentists two brands of toothpaste to try. Proteeth has ingredients that provides protection against cavities, gingivitis, and plaque. Freshie simply freshens breath. The dentists were asked which toothpaste they prefer: Proteeth or Freshie. All preferred Proteeth to Freshie. The marketing company ran an advertisement stating: *Four out of four dentists prefer Proteeth.*

a) Is the marketing claim true? Explain.

b) Why might the marketing company have conducted the survey in this way?

c) How could you redesign the survey to determine if four out of four dentists truly prefer Proteeth?

Achievement Check

14. For each survey, suggest an unbiased method of collecting the data. Write two survey questions for each scenario.

a) a survey that will ask individuals in a community if a new community centre should be built beside the local mall

b) a survey on ways in which a public transit system can be improved

c) a survey that will decide which courses a high school will offer in the following school year

d) a survey that will assess the ability of the current provincial government to make important decisions and follow its election platform

Extend **C**

15. Work with a partner. Design a poster or summary page that displays information in a misleading way that will cause the reader to form an incorrect conclusion. Explain how you have made the information misleading.

16. Find a newspaper article that quotes a survey or statistics. Research the survey or statistics in detail on the Internet and decide if there was any bias in the survey. Write a report on your findings.

3.3 Display Data
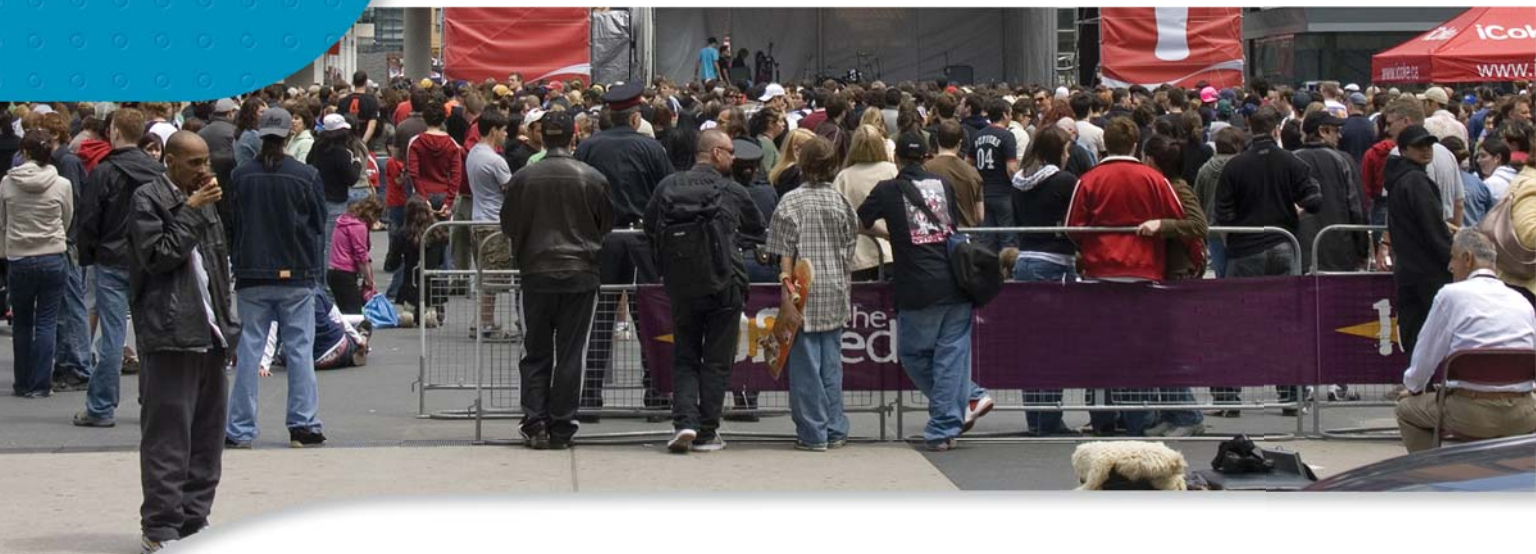
You have heard the saying "a picture is worth a thousand words." The same can be said about a graph. A graph is a visual representation of data that displays the relationship among the variables. Graphs can summarize data, and present data more clearly and concisely than a table or written text.

Investigate 1

Bar Graph and Circle Graph

Jong has a part-time job at a music store. The table shows Jong's expenses last month.

Expense	Amount ($)
entertainment	100
clothing	225
cell phone	50
lunch	75
transportation	80
rent	150

Tools

- calculator
- protractor
- compass
- grid paper
- coloured pencils or markers
- ruler

Method 1: Create Graphs by Hand

1. Create a bar graph to represent the data. Include a title for the graph and label both axes.

2. Copy and complete the table. Calculate the percent and angle measure for each expense. Create a circle graph to represent the data. Include a title for the graph and label the sectors.

Expense	Amount ($)	Percent (%)	Measure of Angle (°)
entertainment	100	$\frac{100}{680} \times 100 \doteq 14.7$	$\frac{14.7}{100} \times 360 \doteq 52.9$
clothing	225		
cell phone	50		
lunch	75		
transportation	80		
rent	150		
TOTAL	**680**		

3. Reflect Which graph best displays Jong's expenses? Justify your answer.

4. When is a bar graph the best choice to display data? When is a circle graph the best choice? Give examples for both.

Method 2: Create Graphs Using Technology

Tools

- computer
- Microsoft® *Excel*

Use the table of Jong's expenses.

1. Follow these steps to create a bar graph using Microsoft® *Excel*.
- Open Microsoft® *Excel*. Enter the expense categories in column A. Enter the amounts in column B.
- Click and drag the mouse from cell A1 to cell B6.
- From the **Insert** menu, choose **Chart …**.
- Under **Chart type**, choose **Column**. Select **Next** twice.
- Enter a chart title (such as "Jong's Expenses"). Enter "Expense" for the Category (X) axis, and "Amount ($)" for the Value (Y) axis. Select **Next**.
- Select **Finish**.
- If the labels are not all visible, click and drag a corner of the chart to make the chart larger.
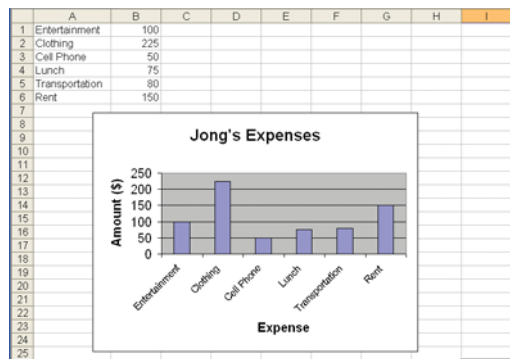
2. Follow these steps to create a circle graph using Microsoft® *Excel*.
- Open a new spreadsheet. Enter the expense categories in column A. Enter the amounts in column B.
- Click and drag the mouse from cell A1 to cell B6.
- From the **Insert** menu, choose **Chart …**.
- Under **Chart type**, choose **Pie**. Select **Next** twice.
- Enter a chart title (such as "Jong's Expenses"). Select the **Data Labels** tab. Check the boxes **Category Name** and **Percentage**. Select **Next**.
- Select **Finish**.
- If the labels are not all visible, click and drag a corner of the chart to make it larger.
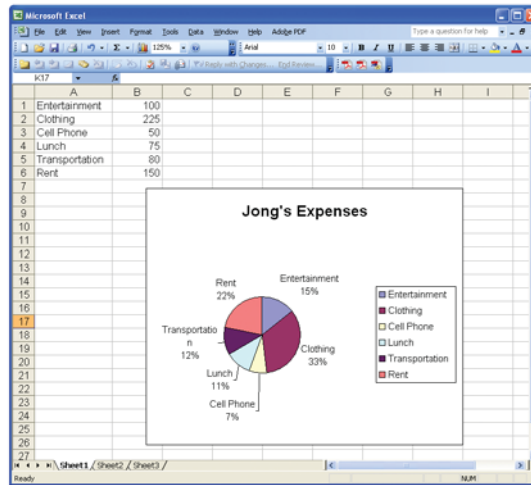
3. Reflect List some advantages and disadvantages of using a spreadsheet to graph data.

categorical data
- data that are types rather than numbers; for example: colours, types of snack foods

continuous data
- data that can have any numerical value within a finite or infinite interval; for example: the heights of students in your class

discrete data
- data that are distinct and can be counted; for example: the number of students who like rice

The data relating to Jong's expenses are **categorical data**. The entries in the first column of the table are non-numerical—they are categories.

There are two types of data. Data that can have any numerical value are **continuous data**. Data that can only have certain values within a given range are **discrete data**.

You can use a histogram to display continuous data.

Investigate 2 — Create a Histogram

William measured the heights of students in his mathematics class. The heights are rounded to the nearest centimetre.

154, 175, 166, 138, 161, 171, 165, 188, 139, 137, 144, 154, 186, 191, 177, 173, 164, 154, 186, 173, 151, 164, 174, 154, 138, 156, 146, 176, 194, 151

Height can have any numerical value, so this is continuous data.

Tools

• calculator
• grid paper

Method 1: Create a Histogram by Hand

1. Copy and complete the table. Record the number of students in each interval, and then determine each frequency.

Interval	Tally	Frequency
[130–140)	||||	4
[140–150)		
[150–160)	|||| ||	7
[160–170)		
[170–180)		
[180–190)		
[190–200)		

A square bracket is used to indicate that a value is included in the interval. A round bracket is used to indicate that a value is not included in the interval. The interval [130–140) includes all heights from 130 cm up to, but not including, 140 cm.

2. Graph the data with Interval on the horizontal axis and Frequency on the vertical axis. Include a title for the graph and label the axes.

Method 2: Create a Histogram Using a Graphing Calculator

Tools

• graphing calculator

Use the heights of the students in William's class.

1. Press `2nd` [MEM] **4** `ENTER` to clear the lists.

2. Press `STAT` and select **1:Edit…**. Enter the student heights in list L1.

3. Press `Y=`. Use the `CLEAR` key to erase any equations in the list.

4. Press `2nd` [STAT PLOT]. Select **4:PlotsOff**. Press `ENTER`.

5. Press `2nd` [STAT PLOT]. Select **1:Plot1**. Set the plot to On, the graph style to Histogram, and the XList to L1.

6. Press $\boxed{\text{ZOOM}}$ and select **9:ZoomStat**. The histogram will appear.

7. Use the window settings shown. The histogram will begin at 130, with a bin width of 10. Then press $\boxed{\text{GRAPH}}$.
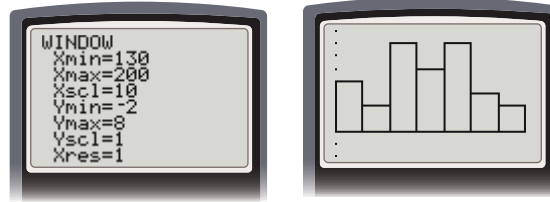
### Method 3: Create a Histogram Using *Fathom*™

1. Start *Fathom*™. From the **Object** menu, choose **New** then **Case Table**.

2. Enter the heading "Height_cm" in the second column. Enter the student heights in this column.

3. Double-click on the heading "Collection 1". In the **Rename Collection** dialogue box, enter the title "Student Heights". Select **OK**.

4. From the **Object** menu, choose **New** then **Graph**. In the case table, select the heading "Height_cm", hold the mouse button, and drag the heading to the horizontal axis of the graph.

5. Select the pop-up menu at the top right of the graph. Change the graph type to **Histogram**.

6. Select the graph. From the **Object** menu, choose **Inspect Graph**. Click on the **Properties** tab. Set **binAlignmentPosition** to 120, **binWidth** to 10, and **yLower** to 0. You may need to adjust the value for **yUpper**. Close the dialogue box.
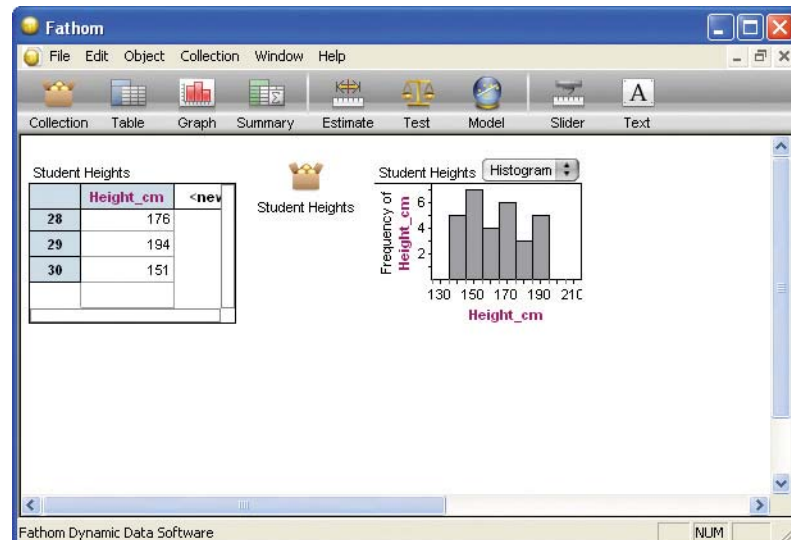
Example 1

Interpret a Bar Graph

Colleen surveyed the students at her school about their favourite sports. She chose a bar graph to display her results. Use the bar graph to answer each question.

Favourite Sport
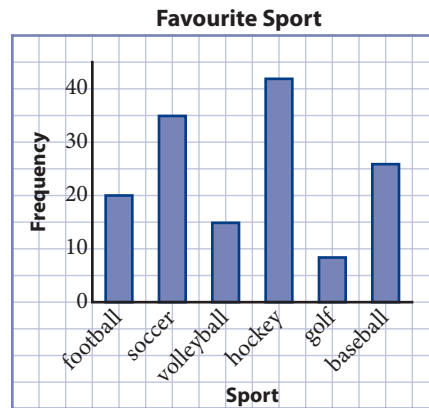
a) Which sport is the most popular?

b) Which sport is the least popular?

c) Does your answer to part b) mean that students do not like this sport? Explain.

d) How many students said that volleyball was their favourite sport?

e) How many students participated in Colleen's survey?

Solution

a) The tallest bar is for hockey. Hockey is the most popular sport.

b) The shortest bar is for golf. Golf is the least popular sport.

c) No. Colleen's survey asked students about the sports they liked, not the sports they disliked.

d) Fifteen students said that volleyball was their favourite sport.

e) Twenty students said football was their favourite sport, 35 students preferred soccer, 15 students preferred volleyball, 42 students preferred hockey, 8 students preferred golf, and 26 students preferred baseball. Add the numbers to find the total number of students surveyed.

$20 + 35 + 15 + 42 + 8 + 26 = 146$

There were 146 students who participated in Colleen's survey.

Example 2

Analyse Continuous Data

The histogram shows the masses of a sample of patients at a hospital.

a) How many patients have a mass of at least 85 kg but less than 90 kg?

b) How many patients have a mass of at least 100 kg?

c) How many patients are in the sample?

d) Find the percent of patients who have a mass of at least 100 kg.

Patients' Masses
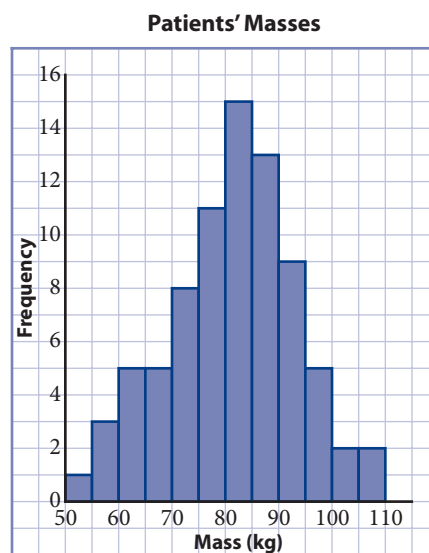
Solution

a) The bar with left side 85 and right side 90 represents the number of patients who have a mass in the interval [85–90].

90 kg is included in the next interval.

The height of the bar is 13; therefore, 13 patients have a mass of at least 85 kg but less than 90 kg.

b) Two patients have a mass in the interval [100–105) and two patients have a mass in the interval [105–110). So, four patients have a mass of at least 100 kg.

c) Add the frequencies to find the number of patients in the sample.

$1 + 3 + 5 + 5 + 8 + 11 + 15 + 13 + 9 + 5 + 2 + 2 = 79$

There are 79 patients in the sample.

d) $\dfrac{4}{79} \times 100 = 5.1\%$

In the sample, 5.1% of the patients have a mass of at least 100 kg.

Key Concepts

- Data can be numerical or categorical.
- Continuous data can have any numerical value, finite or infinite. For example, think of the time required to run 1 km. A histogram or a line graph may be used to display continuous data.
- Discrete data are distinct and can be counted. For example, think of the number of employees at each coffee shop in a city. A bar graph or a circle graph may be used to represent discrete data.

Discuss the Concepts

D1. Describe the similarities and the differences between a bar graph and a histogram. Give an example of data that you would display in a bar graph. Give an example of data that you would display in a histogram.

D2. **a)** Can a set of numerical data be discrete? Explain.
b) Can a set of categorical data be continuous? Explain.
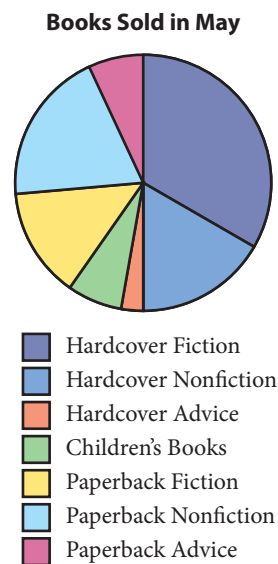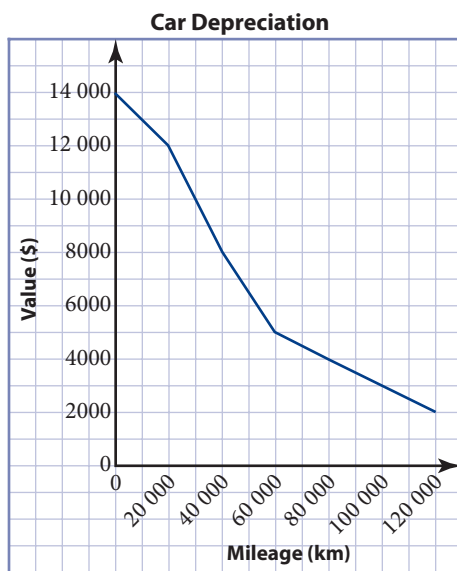
Practise A

1. Classify each set of data as either discrete or continuous.

a) the number of blue cars in a parking lot recorded every day at 5 P.M. for a month

b) the temperature outside at noon measured each day for a week

c) barometric pressure collected each hour for a month

d) the number of students in each Ontario high school mathematics classroom

2. Which type of graph would best suit each situation? Explain your choice.

a) the number of students in each homeroom of your school

b) the time it takes your classmates to travel to school each day

c) your monthly spending habits

d) the daily sales of fruit drinks at a variety store

e) the heights of trees in a forest

f) a hockey team's budget for players' salaries, based on the players' positions

3. Examine the bar graph and the histogram.

Snowfall in December

Test Scores
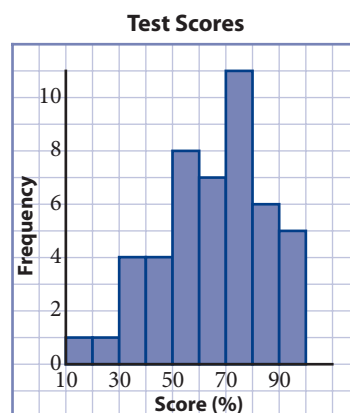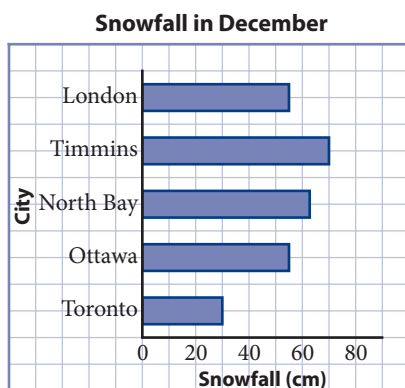
a) Could a bar graph have been used to display the data in the histogram? Explain.

b) Could a histogram have been used to display the data in the bar graph? Explain.

4. Which graph displays discrete data and which graph displays continuous data? Explain how you know.

Car Depreciation

Books Sold in May

- Hardcover Fiction
- Hardcover Nonfiction
- Hardcover Advice
- Children's Books
- Paperback Fiction
- Paperback Nonfiction
- Paperback Advice

5. Students in a mathematics class scored the following marks out of 100 on the last test.

71, 54, 64, 81, 77, 75, 63, 74, 54, 79, 83, 92,
83, 65, 37, 48, 36, 85, 88, 61, 77, 70, 66, 52

a) Copy and complete the table.

Score Interval	Tally	Frequency
[30, 40)		
[40, 50)		
[50, 60)		
[60, 70)		
[70, 80)		
[80, 90)		
[90, 100]		

b) Create a histogram to display the data. If you are using a graphing calculator or computer software, print or sketch a copy of your histogram. Add a title to your graph and label the axes.

6. At the local high school, the number of students participating in six school clubs was recorded. Construct a bar graph to display the data.

Club	Number of Students
Environment	40
Debating	52
Drama	38
SADD	28
Wellness	17
Yearbook	25

7. Darren and Tomas are choosing a graph to display the heights of students in their school. Darren says: "We should use a bar graph. We can find the number of students who are in the 4-ft range, 5-ft range, 6-ft range, and 7-ft range and then display the frequencies as bars."

Tomas disagrees. "Someone who is 5 ft 11 in. might report their height to be 6 ft, which would skew the results. The data is continuous, so a bar graph is not the best display."

Who is correct? Justify your answer.

8. Use newspapers or the Internet to find an example of each type of data. Graph the data using an appropriate display. Justify your choice of graph.

a) categorical data   **b)** discrete data   **c)** continuous data

9. Give an example for which a bar graph would be a good choice to display data. Give an example for which a bar graph would not be appropriate for the data. Explain your reasons.

10. Find an article or advertisement in a newspaper or magazine that shows information using a graph.

a) What is the source of the data?

b) How was the data collected?

c) Who collected the data?

d) Is the graph chosen the best type of graph to display the information? Why or why not?

e) What other type of graph could be used to display the information? Justify your answer.

11. For another market research project, Andrew is analysing several sets of data collected at a local fall fair. Classify each set of data as discrete or continuous. Choose a graph to display each set of data. Justify your choice.

a) daily receipts at the front gate of the fair over the 10 days

b) daily attendance at the fair

c) winning mass of the largest pumpkin over the last 100 years at the fair

d) the numbers of each item sold at the snack stand

e) the most popular carnival game at the fair

Achievement Check

Reasoning and Proving

Representing | Selecting Tools

Problem Solving

Connecting | Reflecting

Communicating

12. Chantal conducted a survey to determine the favourite snack food of students at her school. The data is shown in the table. Display the data using an appropriate graph. Justify your choice of graph.

Favourite Snack Food	Tally
tortilla chips	卌 |卌 卌 |||
fruit bar	卌 卌 卌 卌 ||
muffin	卌	
fruit	卌 ||||
crackers and cheese	卌	
granola bar	卌 卌	

13. Use the Internet to find the heights or masses of players on your favourite sports team. Display the data using an appropriate graph. Justify your choice of graph.

14. A line graph can be used to show a trend in data that changes over time, such as how a company's stock price changes over time. The table shows the weekly closing stock price of a share in a small mining company.

Week Number	Stock Price ($)
1	92.50
2	100.00
3	91.30
4	98.60
5	112.60
6	104.50
7	111.70
8	108.10
9	104.30
10	121.30
11	101.40
12	119.50

a) Draw a set of axes. Number the horizontal axis from 0 to 12 and the vertical axis from 90 to 125 in increments of 5. Graph the data. Add a title to your graph and label the axes.

b) A volatile stock is one whose price changes dramatically. Examine your graph from part a). Does the stock appear to be a volatile stock?

c) Draw another set of axes. Number the horizontal axis from 0 to 12 and the vertical axis from 0 to 130 in increments of 10. Graph the data. Add a title to your graph and label the axes.

d) Examine your graph from part c). Does the stock seem to be more or less volatile than in the graph from part a)?

e) Suppose you are an investor looking to purchase stable stock in a company. The marketing manager from the mining company sends you a report on the company's stock performance. Which graph would the manager most likely send you? Why? Why would the manager not send the other graph?

3.4 Measures of Central Tendency

In the last three sections, you explored methods of collecting and displaying data. In the next three sections, you will calculate statistics that can be used to analyse a set of data. Measures of central tendency provide information on the centre of a set of data.

Investigate

Tools
- calculator

mean
- the sum of values in a set of data divided by the number of values in the set of data

median
- the middle value when data is ordered from least to greatest

mode
- the value or attribute that occurs most often in the set of data

Measures of Central Tendency

Sam works for a company that maintains a career and employment Web site. The Web site provides information on positions and salaries at various companies. The table shows data from one of the companies, Cabinets-R-Us, a small kitchen cabinetry company.

Position	Number of Employees	Annual Salary ($)
cabinet maker	8	43 000
secretary	2	38 000
sales representative	2	48 000
president	1	100 000

1. How many employees work at Cabinets-R-Us?

2. Calculate the **mean** salary.

3. Calculate the **median** salary.

4. What is the **mode** salary?

5. **Reflect** Which measure of central tendency best describes a typical employee salary at Cabinets-R-Us? Justify your answer.

Example 1

Calculate Mean, Median, and Mode

Consider these test scores, all out of 100.

61, 76, 89, 72, 65, 71, 61, 83, 45, 68, 62, 59, 71, 68, 69, 86

a) Find the mean score.

b) Find the median score.

c) Find the mode score.

Solution

a) Mean $= \dfrac{\text{sum of all values}}{\text{number of values}}$

$= \dfrac{61+76+89+72+65+71+61+83+45+68+62+59+71+68+69+86}{16}$

$= \dfrac{1106}{16}$

$= 69.125$

The mean score is 69.125.

b) Order the data from least to greatest.

45, 59, 61, 61, 62, 65, 68, **68, 69,** 71, 71, 72, 76, 83, 86, 89

The middle values are 68 and 69. The median is the mean of the middle values.

$\dfrac{68 + 69}{2} = 68.5$

The median score is 68.5.

c) Each of the scores 61, 68, and 71 occur twice. No other scores are repeated.

There are three mode scores: 61, 68, and 71. The data are trimodal.

Example 2

Use Technology to Find the Mean and Median

A corner store recorded the numbers of newspapers sold each day for 17 days. Find the mean and the median of the data.

111, 131, 152, 98, 112, 117, 124, 108, 113,
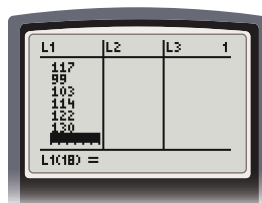112, 119, 117, 99, 103, 114, 122, 130

Solution

Method 1: Use a Graphing Calculator

- Press [2nd] [MEM] **4** [ENTER] to clear the lists.

ClrAllLists

- Press [STAT] and select **1:Edit…**. Enter the data in list L1.

L1	L2	L3	1
117			
99			
103			
114			
122			
130			

L1(18) =

- Press [2nd] [QUIT] to return to the calculator screen.
- Press [2nd] [LIST]. Select **MATH 3:mean(**.
- Press [2nd] [L1]. Press [ )] [ENTER]. The mean of the data in list L1 is displayed.
- Press [2nd] [LIST]. Select **MATH 4: median(**.
- Press [2nd] [L1]. Press [ )] [ENTER]. The median of the data in list L1 is displayed.

Method 2: Use a Spreadsheet

- Open Microsoft® *Excel*. Enter the data in cells A1 to A17.
- Select cell A18. From the **Insert** menu, select **Function…**.
- Change the category to **Statistical**, and select **AVERAGE**. Select **OK**. Select and drag the mouse from cell A1 to cell A17. Select **OK**. The mean will appear.
- Select cell A19. From the **Insert** menu, select **Function…**.
- Change the category to **Statistical**, and select **MEDIAN**. Select **OK**. Click and drag the mouse from cell A1 to cell A17. Select **OK**. The median will appear.

Technology Tip

You can see a list of available functions by clicking on the + icon next to Functions in the calculator window.

Method 3: Use *Fathom*™

- Start *Fathom*™. From the **Object** menu, choose **New** then **Case Table**.
- Enter the heading "Data" in the second column. Enter the data in this column.
- Double-click on the heading "Collection 1". In the **Rename Collection** dialogue box, enter the title "Measures of Central Tendency". Select **OK**.
- From the **Object** menu, choose **New** then **Graph**. In the case table, select the heading "Data", hold the mouse button, and drag the heading to the horizontal axis of the graph.
- Select the graph. From the **Graph** menu, choose **Plot Value**. Type "mean(data)" then select **OK**. The mean will appear on the graph and below the graph.
- Select the graph. From the **Graph** menu, choose **Plot Value**. Type "median(data)" then select **OK**. The median will appear on the graph and below the graph.

Example 3

Best Measure of Central Tendency

Find the mean, the median, and the mode of each set of data. Which measure of central tendency best describes the data? Explain.

a) Erika recorded the masses, in grams, of bags of cashews sold on Monday at her bulk food store.
58, 37, 37, 38, 42, 41, 46, 43, 41, 37, 38, 37

b) The scores on a mathematics quiz, out of 100, are shown.
63, 71, 40, 99, 52, 94, 83, 67, 94, 89, 14, 76, 68

c) The table shows the number of prizes and the value of each prize in a charity fundraiser.

Prize Value ($)	Number of Prizes
5000	1
1000	4
500	8
10	80

Solution

a) Mean

$$\text{Mean} = \frac{\text{sum of all values}}{\text{number of values}}$$

$$= \frac{58+37+37+38+42+41+46+43+41+37+38+37}{12}$$

$$= \frac{495}{12}$$

$$= 41.25$$

The mean mass is 41.25 g.

Median

37, 37, 37, 37, 38, **38**, **41**, 41, 42, 43, 46, 58

$$\text{Median} = \frac{38 + 41}{2}$$
$$= 39.5$$

The median mass is 39.5 g.

Since there is an even number of values, the median value is the average of the two middle values.

Mode

The value 37 occurs most frequently; it appears four times. Therefore, the mode mass is 37 g.

For this set of data, the mean is a good choice for the measure of central tendency. Most of the values are close to the mean.

The median could also be used since it is close to the mean. Notice that this median is not a value in the set of data.

The only measure of central tendency that is a value in the set of data is the mode. However, the mode is the least value in the set of data, making it a poor choice as the best measure of central tendency.

b) Mean

$$\text{Mean} = \frac{\text{sum of all values}}{\text{number of values}}$$
$$= \frac{63+71+40+99+52+94+83+67+94+89+14+76+68}{13}$$
$$= \frac{910}{13}$$
$$= 70$$

The mean score is 70.

Median

14, 40, 52, 63, 67, 68, **71**, 76, 83, 89, 94, 94, 99
There are six values before and after the median.
The median score is 71.

Since there are 13 values, the median is the seventh value.

Mode

The number 94 occurs twice. No other score is repeated. Therefore, the mode score is 94.

For this set of data, the median is the best measure of central tendency. The median is not affected by **outliers**. The median value occurs in the set of data.

outlier
- an extreme value in a set of data
- a value "far away" from the other values in a set of data

c) Mean

$$\text{Mean} = \frac{\text{sum of all values}}{\text{number of values}}$$

$$= \frac{1 \times 5000 + 4 \times 1000 + 8 \times 500 + 80 \times 10}{9}$$

$$= \frac{13\ 800}{93}$$

The mean prize value is $148.39.

Median

There are 93 prizes, of which 80 have a value of $10.
Therefore, the median prize value is $10.

Mode

The value $10 occurs 80 times.
Therefore, the mode prize value is $10.

For this set of data, the median and the mode are the same. Due to the frequency of the $10 prize, the mode is the best measure of central tendency.

Key Concepts

- The mean, the median, and the mode are measures of central tendency of a set of data.
- The mean is calculated by adding the data and dividing the total by the number of data values.
- The median is found by listing the data in order from least to greatest and finding the middle value. If there is an even number of values, the median is the mean of the two middle values.
- The mode is found by finding the most frequently occurring value. A set of data can have more than one mode or no mode.

Discuss the Concepts

D1. Is it possible for a set of data to have the same mean, median, and mode? Explain, using an example.

D2. Which measure of central tendency is usually the least representative of a set of data?

D3. Which measure(s) of central tendency can be used for categorical data? Explain, using an example.

For help with question 1, refer to Example 1.

1. Find the mean (to one decimal place), the median, and the mode of each set of data.

   a) the number of litres of gasoline purchased by customers in one hour at a gas station:
   25, 21, 38, 29, 32, 44, 38, 21, 16

   b) the number of points scored by a basketball team at home games:
   44, 36, 82, 53, 71, 74, 38, 81, 94, 58

For help with questions 2 to 4, refer to Example 3.

2. The table shows the prizes awarded in Jason's school fundraiser.

Prize Value ($)	Number of Prizes
100	2
50	4
10	10

   Find the mean (to one decimal place), the median, and the mode prize value.

3. Find the mean, the median, and the mode of each set of data. Which measure of central tendency best describes the data? Explain.

   a) the wages earned by Alisa each month:
   214, 333, 197, 310, 622, 410, 520, 285, 540, 383, 427, 345

   b) the at-bats of 15 Toronto Blue Jays for one season:
   87, 461, 611, 260, 466, 581, 433, 546,
   348, 450, 437, 540, 290, 296, 251

4. The table shows the heights of grade 11 students at Sacha's school.

Height (cm)	Frequency
[155–160)	2
[160–165)	6
[165–170)	12
[170–175)	11
[175–180)	6
[180–185)	4
[185–190)	2

   a) Find the median, the mode, and the range of heights.

   b) Which measure of central tendency best describes the data? Explain.

5. Give an example of when the mode is the best measure of central tendency of a set of data.

6. Some schools report trends in class marks using the median. Why do you think this measure is preferred over the mean? Explain.

7. Veronica conducted a survey to find the average shoe size of the female students in two grade 11 classes. She displayed the information in a bar graph.

Grade 11 Girls' Shoe Sizes
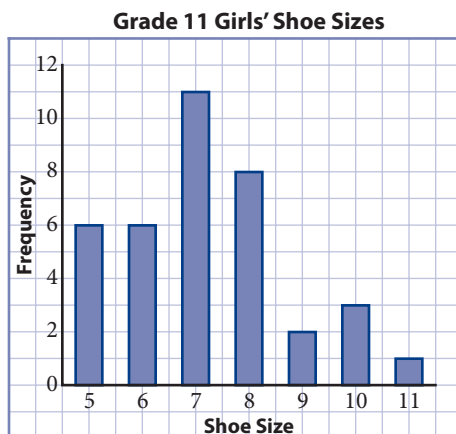
a) Find the mean (to the nearest shoe size), the median, and the mode shoe size.

b) Which measure of central tendency is easiest to find from the graph? Explain.

c) Which measure of central tendency best describes the data? Explain.

8. As part of his final project in his co-op placement, Andrew found the attendance for the local fair over the past 20 years. He recorded the attendance in thousands of people.

23, 31, 44, 27, 32, 41, 35, 42, 37, 41,
43, 39, 36, 37, 43, 27, 36, 42, 41, 43

a) Find the mean, the median, and the mode attendance.

b) Which measure of central tendency best describes the attendance per year?

c) Why might the attendance vary so greatly from year to year?

9. Lionel and Jeffrey are trying out for the final spot on the school archery team. The coach has invited both boys to a shootout. Based on the results, she will make her choice. Here are their results after ten flights (rounds) of three arrows.

Flight	1	2	3	4	5	6	7	8	9	10
Lionel	18	12	22	22	29	22	14	28	26	20
Jeffrey	20	22	22	19	21	21	21	21	22	20

a) Find the mean, the median, and the mode score for each archer.

b) Which measure of central tendency best represents each archer's performance? Explain.

c) Which archer should the coach choose? Justify your choice.

10. The batting average for a baseball player is found by dividing the player's total hits by the player's total at-bats. The value is rounded to three decimal places. The table shows the number of hits and at-bats for a girls' baseball team.

Position	Hits	At-Bats	Batting Average
1st base	26	71	0.366
2nd base	38	84	
3rd base	25	62	
shortstop	31	67	
catcher	28	70	
pitcher	12	39	
left field	41	88	
centre field	52	88	
right field	47	88	

a) Copy and complete the table.

b) Find the mean batting average.

c) Find the team's batting average.

d) Compare your answers for parts b) and c). What do the two results mean in relation to each other? Why are the two results not the same?

e) Find the mean batting average for the three fielders.

f) Add the number of hits for the three fielders and divide this value by their total number of at-bats. Find the fielders' batting average.

g) Compare your results from parts e) and f). Why are the two results the same?

Reasoning and Proving

Representing | Selecting Tools

Problem Solving

Connecting | Reflecting

Communicating

11. A company employs 15 people at a weekly salary of $250 each, 4 people at a weekly salary of $500 each, and 3 people at a weekly salary of $1200 each.

a) Find the mean, the median, and the mode salary.

**b)** Which measure of central tendency best represents salaries at the company? Explain.

Extend C

12. Search the Internet for the salaries of players on a professional sports team.

a) Find the mean, the median, and the mode salary.

b) Which measure of central tendency best describes the data?

c) Compare results with several classmates who have chosen different teams or different sports.

d) Do the measures of central tendency for each team correspond to the team's current year performance? Explain. List any assumptions you have made.

13. Given the set of numbers 11, 8, 14, 14, x, consider each situation.

a) The mean of the numbers is 13. Find the value of x.

b) The median of the numbers is 14. What do you know about x?

c) There are two modes. What do you know about x?

14. Annette has a mean bowling score of 150 after six games. What score must she get in the seventh game to raise her mean score to 152?

3.5 Measures of Spread
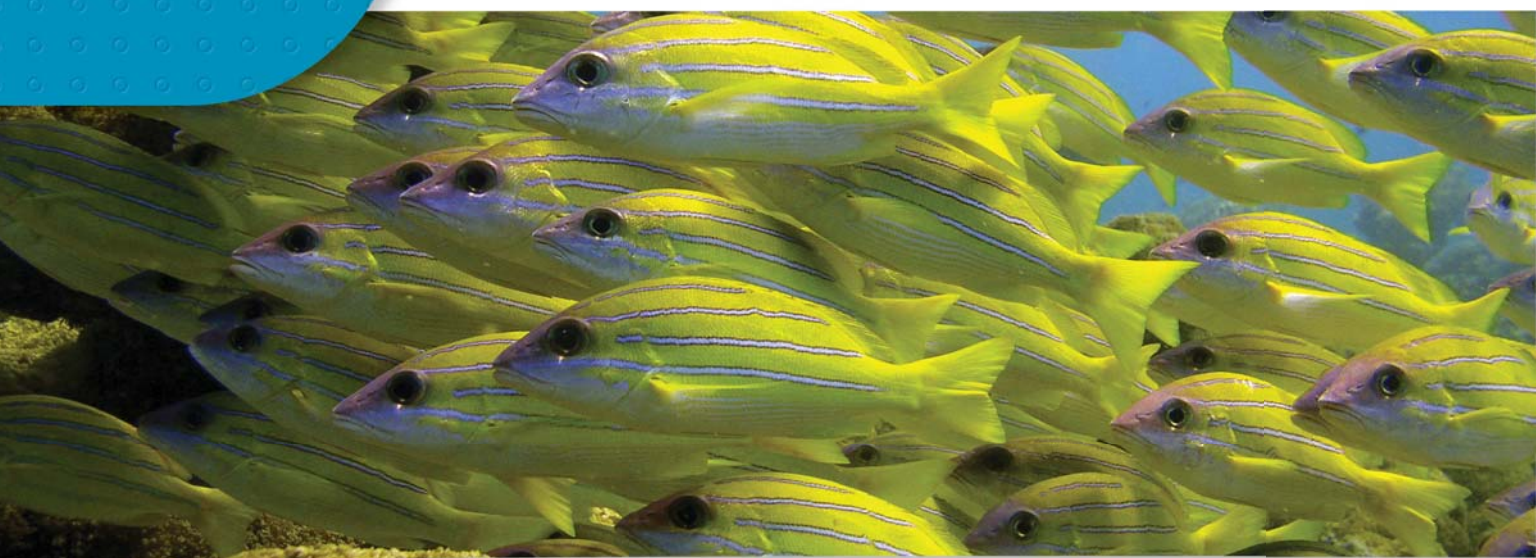
Measures of central tendency are values around which a set of data tends to cluster. However, to analyse a set of data, it is useful to know how spread out the data are. Measures of spread describe how the values in a set of data are distributed.

Investigate 1

Tools
- graphing calculator

Use a Graphing Calculator to Analyse Data

There are two first-year mathematics classes at Caldwell College. The tables show the midterm marks for the two classes.

Class 1	59	79	89	49	71	68	67	48	69	67	75
	82	80	59	58	74	66	90	73	81	59	

Class 2	91	88	50	44	42	88	79	92	83	77	43
	62	98	52	67	84	70	55	89	48		

1. Use these steps to graph the data for each class.

- Press $\boxed{\text{Y=}}$ and clear any equations.
- Press $\boxed{\text{STAT}}$ and select **1:Edit…**.
  Enter the data for Class 1 in list L1.
  Enter the data for Class 2 in list L2.
- Press $\boxed{\text{2nd}}$ then [STATPLOT] to access **PLOT1**. Turn Plot1 On.
- Cursor down to Type. Turn the second type of Box Plot On as shown. Enter L1 for Xlist.

• Press ⌈ 2nd ⌉ then [STATPLOT] to access **PLOT2**. Turn Plot2 On.
• Cursor down to Type. Turn the second type of Box Plot On. Enter L2 for Xlist.
• Press ⌈ZOOM⌉ **9:Zoomstat** to see both plots.
• Press ⌈TRACE⌉. Use the cursor to see the five values for each plot: the minimum, Q1, the median, Q3, and the maximum. Q1 and Q3 are the first and third **quartiles**, the median is the second quartile.
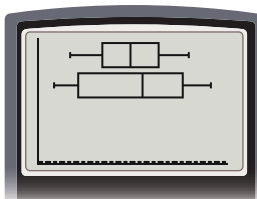
2. Compare the median score for each class. Which class has the better performance?

3. Calculate the **range** for each class. What does the range tell you?

4. Reflect This type of graph is called a **box-and-whisker plot**. Refer to the box-and-whisker plots. Which class performed better? How does this compare to your answer to question 2? Explain.

Variance and **standard deviation** are measures of the spread of the values in a distribution. The greater the standard deviation and the greater the variance, the greater the distance of the values from the mean.

$$\text{Variance} = \frac{(x_1 - \text{mean})^2 + (x_2 - \text{mean})^2 + (x_3 - \text{mean})^2 + \ldots + (x_n - \text{mean})^2}{n}$$

where $x_1, x_2, x_3, \ldots$ are values in the set of data, and n is the number of values in the set of data.

Standard deviation $= \sqrt{\text{variance}}$

Example 1

Determine Quartiles and the Interquartile Ranges

Iris works part-time selling cell phones. She recorded the numbers of cell phones she sold each month for the last 12 months.

51, 17, 25, 39, 7, 49, 62, 41, 20, 6, 43, 13

a) Find the median, the first quartile (Q1), and the third quartile (Q3).
b) Display the data in a box-and-whisker plot.

Solution

a) Order the data from least to greatest.

6, 7, 13, 17, 20, **25**, **39**, 41, 43, 49, 51, 62

$$\text{Median} = \frac{25 + 39}{2}$$

$$= 32$$

The median indicates that half of Iris's monthly sales were more than 32 phones and half were fewer than 32 phones.

To determine the first quartile, find the median of the lower half of the data.

6, 7, **13**, **17**, 20, 25

$$Q1 = \frac{13 + 17}{2}$$

$$= 15$$

There is an even number of values, so find the mean of the two middle values.

The first quartile is 15 phones. The first quartile indicates that for one-quarter of the months, Iris sold fewer than 15 phones and for three-quarters of the months, she sold more than 15 phones.

To determine the third quartile, find the median of the upper half of the data.

39, 41, **43**, **49**, 51, 62

$$Q3 = \frac{43 + 49}{2}$$

$$= 46$$

There is an even number of values, so find the mean of the two middle values.

The third quartile is 46 phones. The third quartile indicates that for three-quarters of the months, Iris sold fewer than 46 phones and for one-quarter of the months, she sold more than 46 phones.

The first and third quartiles are the boundaries for the central half of the data. They show that half of Iris's monthly sales were between 15 and 46 phones. The difference between the third quartile, 46, and the first quartile, 15, is the **interquartile range**. The interquartile range is 31.

interquartile range

- the range of the central half of a set of data when the data are arranged from least to greatest
- a measure of how closely data clusters around its mean

b) Draw a number line with values from 0 to 70.

- The minimum value is 6 and the maximum value is 62. Draw a horizontal line segment from 6 to 62.
- Draw a vertical line segment at the first quartile, 15, the median, 32, and the third quartile, 46. Join the vertical segments to form a box.
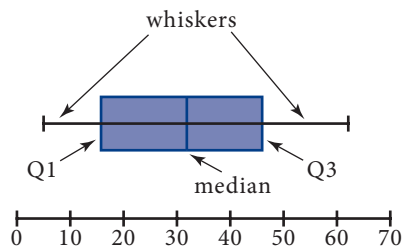
Example 2

Calculate Range, Variance, and Standard Deviation

Carmella's monthly gasoline expenses, in dollars, for the past year are shown.

61, 83, 77, 88, 67, 71, 65, 72, 67, 84, 90, 80

Calculate the range, the mean, the variance, and the standard deviation. Round your answers to the nearest dollar.

Solution

Range = greatest value − least value
$$= 90 - 61$$
$$= 29$$

The range of the expenses is $29.

$$\text{Mean} = \frac{61 + 83 + 77 + 88 + 67 + 71 + 65 + 72 + 67 + 84 + 90 + 80}{12}$$
$$\doteq 75.42$$

The mean expense is approximately $75.

Calculate the variance and the standard deviation.
- Subtract the mean from each data value and square the result.
- Add the squared numbers.
- Divide this sum by the number of data values.
- Calculate the square root of the quotient.

Amount ($)	Amount − Mean ($)	(Difference From Mean)^2
61	$61 - 75 = -14$	196
83	$83 - 75 = 8$	64
77	$77 - 75 = 2$	4
88	$88 - 75 = 13$	169
67	$67 - 75 = -8$	64
71	$71 - 75 = -4$	16
65	$65 - 75 = -10$	100
72	$72 - 75 = -3$	9
67	$67 - 75 = -8$	64
84	$84 - 75 = 9$	81
90	$90 - 75 = 15$	225
80	$80 - 75 = 5$	25
Total		1017

$$\text{Variance} \doteq \frac{1017}{12}$$
$$\doteq 85$$

The variance is approximately $85.

$$\text{Standard deviation} \doteq \sqrt{85}$$
$$\doteq 9$$

The standard deviation is approximately $9.

Example 3

Compare Measures of Spread in Two Set of Data

Not all manufactured items will be identical in size or shape. Usually companies have guidelines as to how much an item can vary from the specified measurements. The decision to accept or reject an item depends on whether the item falls within the specified limits of variation.

A company needs bolts for its heavy machinery. The company must choose between two bolt manufacturers: CanInco and QualiTek. Both manufacturers produce bolts with a mean length of 72 mm. According to quality control records, bolts from CanInco have a standard deviation of 2 mm while bolts from QualiTek have a standard deviation of 0.1 mm. Which manufacturer should the company choose?

Solution

Although both manufacturers produce bolts with the same mean length, the bolts from CanInco have a greater standard deviation. This suggests that there is greater variety in the length of bolts from CanInco. If the company wants a consistently uniform bolt size, it should choose QualiTek, assuming the prices for both manufacturers are similar.

Key Concepts

- Several quantities can be used to measure the spread in a set of data.
- The range is the difference between the greatest and least values in a set of data. To calculate the range, subtract the least value from the greatest value.
- The variance is a measure of how spread out the values in a set of data are from the mean. The greater the variance, the greater the spread of the data values.

  $$\text{Variance} = \frac{(x_1 - \text{mean})^2 + (x_2 - \text{mean})^2 + (x_3 - \text{mean})^2 + \ldots + (x_n - \text{mean})^2}{n},$$

  where $x_1, x_2, x_3, \ldots$ are values in the set of data, and n is the number of values in the set of data.

- The standard deviation is another measure of how spread out the values in a set of data are from the mean. The greater the standard deviation, the greater the spread of the data values.

  $$\text{Standard deviation} = \sqrt{\text{variance}}$$

- Quartiles are three values that divide a set of data into four intervals with equal numbers of data.
- The interquartile range measures how closely data clusters around the median.

D1. Can the range of a set of data be zero? If so, what would this indicate about the interquartile range, the variance, and the standard deviation? Explain.

D2. Give an example of a set of data that has a large standard deviation. Give an example of a set of data that has a small standard deviation.

D3. When is a set of data with a large standard deviation more desirable than a set of data with a smaller standard deviation? Explain, using an example.

Practise **A** •••

For help with question 1, refer to Example 1.

1. This set of data shows the numbers of customers who made purchases at a coffee shop each day in one month.

114, 142, 59, 122, 111, 128, 158, 79, 88, 107, 133,
131, 113, 152, 149, 99, 84, 112, 104, 109, 122,
131, 144, 155, 139, 142, 119, 80, 127, 140, 135

a) Find the median for the set of data.

b) Find the first and third quartiles.

c) What is the interquartile range?

For help with questions 2 to 7, refer to Example 2.

2. Find the range for each set of data.

a) the number of hours worked by restaurant staff in a given week:
11, 4, 55, 42, 41, 36, 50, 6, 8, 44, 39

b) Alex's monthly earnings, in dollars:
100, 115, 112, 125, 104, 101, 117, 121, 98, 100, 95, 102

c) the number of songs Matilda downloaded each month:
12, 11, 9, 12, 13, 15, 14, 11, 11, 8, 6, 7

d) the masses of students in a club, in kilograms:
65, 45, 71, 85, 37, 91, 88, 74, 76, 68, 65

3. A set of data has a range of 30. The least value in the set of data is 22. What is the greatest value in the set of data?

4. A set of data has a range of 14. The greatest value in the set is 116. What is the least value in the set of data?

5. Each measurement is the variance for a set of data. Find the standard deviation for each set of data. Round your answer to one decimal place, if necessary.

  a) 154 g

  b) 36 m

  c) 80 cm

  d) 18 L

6. Each measurement is the standard deviation for a set of data. Find the variance for each set of data.

  a) 14.1 cm

  b) 3.5 kg

  c) $22

  d) 6.7 mL

7. Calculate the variance and the standard deviation for each set of data.

  a) the age of members of a book club:
    24, 35, 45, 41, 22, 51, 46, 40

  b) the number of double-faults per game for a tennis player:
    5, 4, 7, 11, 1, 6, 6, 7, 7

  c) the points scored at home games by the varsity boys' basketball team:
    100, 105, 101, 103, 100, 99, 102, 98

  d) the number of pizzas made per day at a pizzeria:
    56, 57, 54, 51, 58, 59, 51, 50, 53, 51

Apply B

8. The players on two basketball teams have the same mean height of 200 cm. The standard deviation of the heights of the players on Team One is 15 cm and on Team Two is 20 cm. What can be said about the players on each team, given the difference in the standard deviation of their heights? Explain.

9. The mathematics test scores, out of 100, for 13 students are shown. Display the data in a box-and-whisker plot.

80, 75, 90, 95, 65, 65, 80, 57, 85, 70, 74, 100, 84

Literacy Connect

10. A bowling team needs one more person. Eileen and Ingrid are being considered for the spot. They both have the same mean score, but Ingrid's scores have a lower standard deviation than Eileen's scores. Which person should be chosen? Explain your answer.

11. The manufacturing process allows for slight differences in the width of a piston in a cylinder. Slightly smaller or larger piston diameters mean the gaskets used to provide a tight seal will be slightly more or less compressed by the piston on the cylinder walls. In a quality control test, 10 pistons were chosen at random and measured. The table shows the results.

Piston	Diameter (cm)
1	12.85
2	12.77
3	12.91
4	12.87
5	12.81
6	12.90
7	12.78
8	12.80
9	12.92
10	12.99

a) Calculate the standard deviation of the diameters.

b) An item is defective if its diameter is more than two standard deviations from the mean. How many of the tested pistons are defective?

c) What percent of pistons sampled are defective?

d) Does your answer to part c) represent the expected failure rate of pistons made at this manufacturing plant? Explain why or why not.

Chapter Problem

12. Andrew is still working with the attendance figures, in thousands of people, for the fall fair over the past 20 years. Find the range, the variance, and the standard deviation for the data.

23, 31, 44, 27, 32, 41, 35, 42, 37, 41, 43, 39, 36, 37, 43, 27, 36, 42, 41, 43

Achievement Check

13. Joanna's mathematics test scores have a mean of 81% and a standard deviation of 5%. Adam's mathematics test scores have a mean of 84% with a standard deviation of 10%. Whose test scores are more consistent? What do these measures of central tendency and spread tell you about each student?

Extend **C**

14. Select a National Basketball Association (NBA) or Women's National Basketball Association (WNBA) team.

a) Use the Internet to find the heights of players on your chosen team.

b) Convert these heights to centimetres. Find the mean height, the range, and the standard deviation.

c) Compare results with several classmates who researched different teams in the same league.

d) Is there a relationship between the heights of the players on the teams and their scoring records? Explain.

Common Distributions
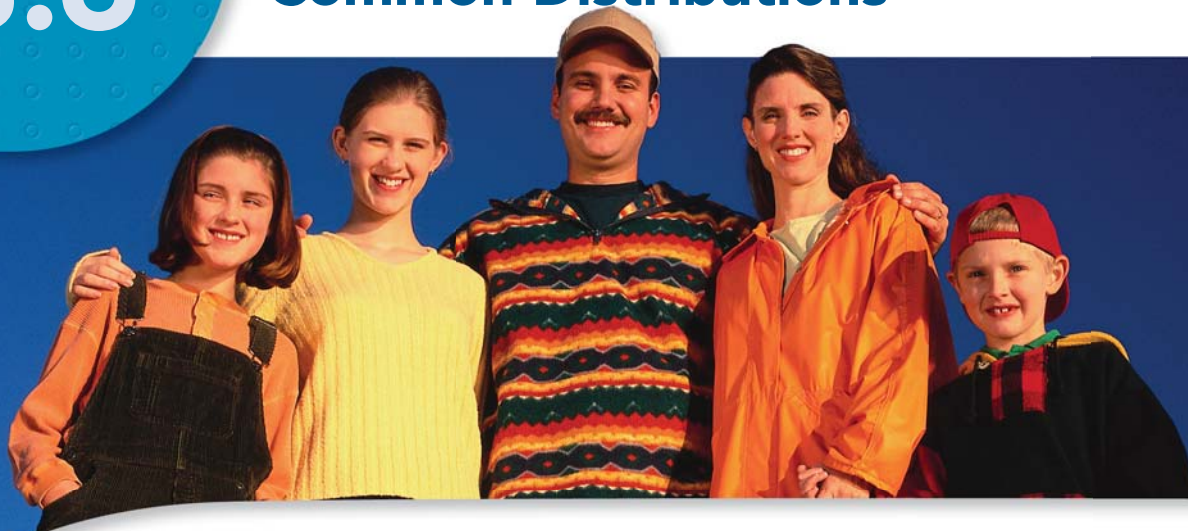
Frequency distributions show the frequency of each outcome in a given situation. The range in values and the frequency of specific values are important measures in data analysis.

Investigate 1

Tools

- calculator
- grid paper
- ruler

Normal Distribution

Alice recorded the heights of people who walked into a hockey arena over a two-hour period.

Class Intervals of Heights (cm)	Tally	Frequency
[100–110)	ℋℋ	
[110–120)	ℋℋ |||	
[120–130)	ℋℋ ℋℋ ℋℋ		
[130–140)	ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ	
[140–150)	ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ	
[150–160)	ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ		
[160–170)	ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ℋℋ ||||	
[170–180)	ℋℋ ℋℋ ℋℋ ℋℋ |||	
[180–190)	ℋℋ ℋℋ ||||	
[190–200)	ℋℋ |||	
[200–210)	||	

Method 1: Graph a Normal Distribution By Hand

1. Copy and complete the table.

2. Display the data using a histogram.

3. Place a point at the centre of the top of each bar in your histogram. Join the points with a smooth curve.

4. The data approximates a **normal distribution**. Describe the shape of a normal distribution in your own words.

5. Alice calculated the mean height as 154 cm, the median height as 152 cm, and the mode height as 157 cm. How do these measures of central tendency relate to the curve in the normal distribution?

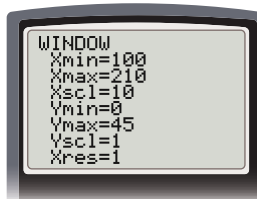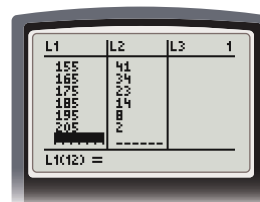**6. Reflect** What does the shape of the curve tell you about the heights of the people at the game?

Method 2: Graph a Normal Distribution Using Technology

Tools

• graphing calculator

Use the data from the frequency table.

1. Follow these steps to create a histogram using a graphing calculator.
   • Press $\boxed{\text{2nd}}$ [MEM] **4** $\boxed{\text{ENTER}}$ to clear the lists.
   • Press $\boxed{\text{STAT}}$ and select **1:Edit…**. Enter the median value of each interval in list L1. For example, for the interval [100–110), enter 105. Enter the frequencies in list L2.
   • Press $\boxed{\text{Y=}}$. Clear any equations in the list.
   • Press $\boxed{\text{2nd}}$ [STATPLOT]. Select **4:PlotsOff**. Press $\boxed{\text{ENTER}}$.
   • Press $\boxed{\text{2nd}}$ [STATPLOT] again. Select **1:Plot1**. Set the plot to On, the graph style to Histogram, the XList to L1, and the frequency to L2.
   • Press $\boxed{\text{ZOOM}}$. Select **9:ZoomStat**.
   • Adjust the window settings as shown. The histogram will begin at 100, with a bin width of 10. Press $\boxed{\text{GRAPH}}$.
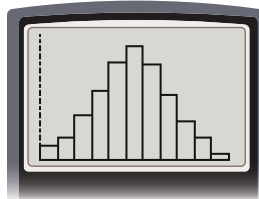
2. Print or sketch the histogram.

3. Place a point at the centre of the top of each bar in your histogram. Join the points with a smooth curve.

4. The data approximates a normal distribution. Describe the shape of a normal distribution.

5. Alice calculated the mean height as 154 cm, the median height as 152 cm, and the mode height as 157 cm. How do these measures of central tendency relate to the curve in the normal distribution?

6. Reflect What does the shape of the curve tell you about the heights of the people at the game?
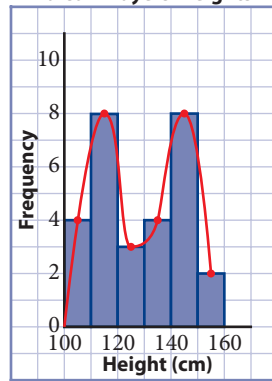
Example 1

Bimodal Distribution

Frank recorded the heights of the peewee and senior girls' Falcon soccer teams.

Height (cm)	[100–110)	[110–120)	[120–130)	[130–140)	[140–150)	[150–160)
Frequency	4	8	3	4	8	2

a) Display the data using a histogram. Place a point at the centre of the top of each bar in your histogram. Join the points with a smooth curve.

b) Describe the shape of the curve.

c) What does the shape of the curve tell you about the heights of the players?

Solution

a)

Falcon Players' Heights

b) The curve represents a **bimodal distribution**. It has two peaks representing the two modes and is symmetrical about the centres.

c) Most of the players fall into two subgroups: players with heights between 110 cm and 130 cm, and players with heights between 130 cm and 150 cm. Measures of central tendency may not be good indicators of this data.

bimodal distribution

• a distribution that contains two equally likely measures of central tendency within the data

Example 2

Skewed Distribution

Students in the French club recorded their final marks, out of 100, for grade 9 French.

Interval	[50–60)	[60–70)	[70–80)	[80–90)	[90–100]
Frequency	0	1	2	6	3

a) Display the data using a histogram. Place a point at the centre of the top of each bar in your histogram. Join the points with a smooth curve.

b) Describe the shape of the curve.

c) What does the shape of the curve tell you about the students?

Solution

a)

Final Marks in French

skewed distribution
• a non-symmetrical distribution of data
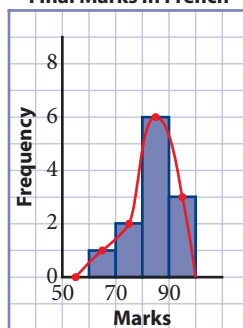
b) The curve represents a **skewed distribution**. It has a greater cluster of data on the right side than on the left.

c) Most marks are clustered between 80 and 90. This means that most of the French club members are doing well in French. Since the data are skewed, measures of central tendency may not be good indicators of this data.

Example 3

Identify the Type of Distribution

In each case, predict the shape of the data distribution. Give a reason for your prediction.

a) the heights of members of the Toronto Raptors basketball team

b) the cost of 1 L of gas in a city in Ontario

c) the masses of players on the Canadian Olympic men's and women's hockey teams

Solution

a) NBA basketball players are tall, so the heights of the members of the basketball team should be skewed to the right when compared to rest of the population.

b) Since the price of gas is fairly standard in a city, with few differences between self-serve or full service stations, the data should be normally distributed.

c) Most male players will have a greater mass than most female players, so the distribution should be bimodal.
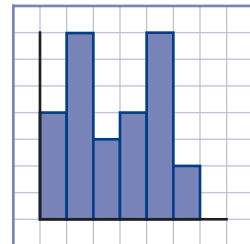
Key Concepts

- In a normal distribution, the data are distributed symmetrically about the mean. The mean, the median, and the mode are close in value and are located at the centre of the distribution.
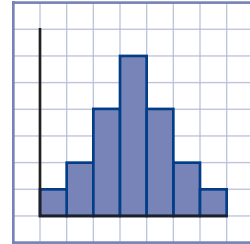
- A skewed distribution has the appearance of a normal distribution that has been pushed to one side of the mean. The result is an asymmetrical or lopsided distribution.
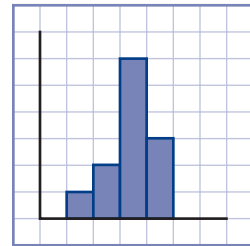
- A bimodal distribution has two peaks. It is symmetrical, with frequencies clustering around two sub-groups.
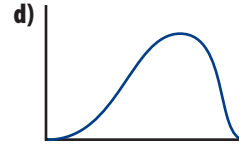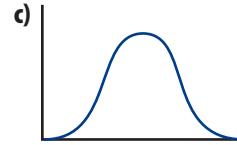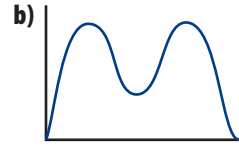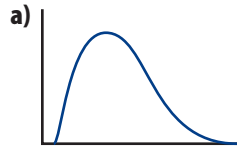

Discuss the Concepts

D1. What is the difference between a skewed distribution and a normal distribution? Explain, using examples.

D2. Can skewed, normal, and bimodal distributions have the same range? Explain.

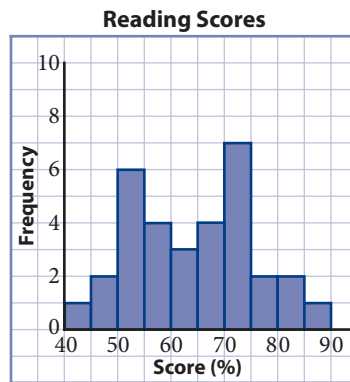**For help with questions 1 to 3, refer to Examples 1 and 2.**

1. Classify each distribution as normal, skewed, or bimodal.

a)

b)

c)

d)

2. Give an example of a set of data that has each distribution.

   a) normal distribution

   b) skewed distribution

   c) bimodal distribution

3. The histogram shows the reading scores for a grade 4 class.

**Reading Scores**

   a) What type of distribution is this?

   b) What might cause this type of distribution in the reading scores in this class?

Literacy Connect

4. a) Why would a clothing manufacturer monitor the mean dress size for women who shop at stores carrying its clothing?

   b) Would the manufacturer be interested in the standard deviation of dress sizes? Explain.
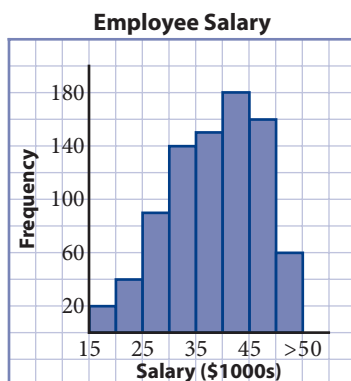
5. The graph shows the ages of homeowners in a subdivision.

Age of Homeowners
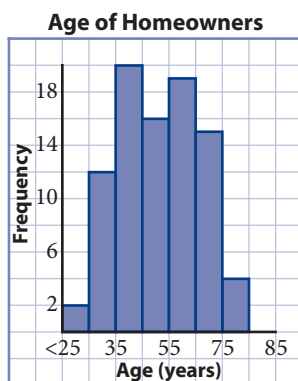
a) What does the distribution tell you about the population?

b) Why might a mayoral candidate be interested in this information? Explain.

c) Approximate the mean and the median age.

6. The graph shows the salaries of employees at a large company.

Employee Salary

a) What conclusions can you draw about the salaries at the company? Explain.

b) Suppose you were a recruiter, trying to entice people to work for the company. Which measure of central tendency would you use? Explain.

c) Suppose you have to report salaries to head office and want to convey that salary expenditure is low. Which measure of central tendency would you use? Explain.

7. Scores out of 100 on a college entrance examination are shown.

Mark Interval	Tally	Frequency
[30–40)	⩗⩗ ⩗⩗	
[40–50)	⩗⩗ ⩗⩗ ⩗⩗ ||	
[50–60)	⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗		
[60–70)	⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗	
[70–80)	⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ |||	
[80–90)	⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗ ⩗⩗		
[90–100]	⩗⩗ ⩗⩗ ⩗⩗ ||||	

a) Copy and complete the table. Display the data using a histogram.

b) Place a point at the centre of the top of each bar in your histogram. Join the points with a smooth curve.

c) What type of distribution does the curve represent? Explain.

Extend C

8. In a normal distribution, about 68% of all values are within 1 standard deviation of the mean. A run of juice cans at a canning factory were found to contain a mean of 760 mL of juice per can. The standard deviation was 20 mL and the values formed a normal distribution. Out of 2000 cans, about how many cans contain between 740 mL and 780 mL of juice?

9. Search for information on the properties of a population that can be represented in a normal distribution. Research what the mean value is for the property, and find information on its standard deviation. Write a report on your findings.

3.1 Sampling Techniques, pages 102–109

1. In each situation, identify the sampling technique.

a) Kuljit went to the local music store to ask what people thought of the "Canadian Idol" winner's debut CD.

b) The school council has set up a booth at the front of the school on Parents' Night to ask about changing the school uniform.

c) Sherry asks 20 girls and 20 boys on the school sports teams if the sports council should spend the fundraising money on new football equipment.

2. Cary plans to survey 100 people. Describe how Cary can choose a stratified sample if her survey population contains 1200 people, and 60% are female.

3. Describe how a graphing calculator can be used to choose a random sample of 15 people from a population of 200 people.

3.2 Collect and Analyse Data, pages 110–117

4. Rewrite the survey question so that it does not contain bias.

Most schools hold a carnival during their Spirit Week, which is usually a great success. Do you think that having a carnival would be a good idea for this year's Spirit Week?

5. Identify the type of bias in each survey.

a) A survey to determine the effectiveness of a government's social services is conducted at a homeless shelter.

b) A survey sent via the Internet asks people to answer a questionnaire and email it to a central processing station.

c) When asked to circle their favourite candidate in the student council, the choices were:
The President
The secretary
The treasurer
Other: _____

6. Randy decides to hand out a survey to every fifth person entering the school. He asks them to fill it out and hand it in at the office when they are done.

a) What type of sampling technique is Randy using?

b) How could this sampling technique lead to inaccurate results?

3.3 Display Data, pages 118–129

7. The table shows the approximate amount of time Le Hing spent on various activities in one week.

Activity	Time (h)
doing homework	8.0
watching television	6.0
making phone calls and text messaging	8.0
playing volleyball	3.0
practising guitar	6.0

a) Which two types of graphs could Le Hing use to display the information?

b) Display the data using each type of graph.

8. The histogram shows the birth masses, in kilograms, of babies born at a hospital in one week.
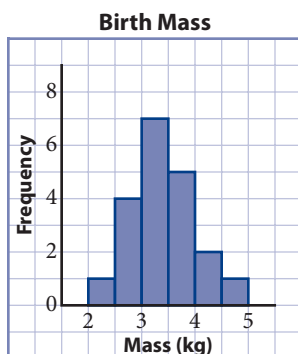
Birth Mass

a) How many babies were born with a mass of at least 4.0 kg?

b) What percent of the babies born were at least 2.0 kg but less than 3.5 kg?

c) Explain why birth mass can be displayed in a histogram.

3.4 Measures of Central Tendency, pages 130–139

9. Find the mean, the median, and the mode for each set of data.

a) 21, 45, 53, 47, 82, 21,
64, 77, 54, 92, 91, 72

b) 4, 7, 11, 8, 6, 6, 5, 3, 5, 7,
8, 14, 17, 18, 6, 4, 2, 2

c) 77, 78, 67, 54, 82, 91, 71,
73, 64, 68, 53, 87, 79

10. A gallery has these items for sale.
- 12 bronze statues for $500 each
- 50 paintings for $100 each
- 100 hand-painted tiles for $25 each

a) Find the mean, the median, and the mode of the prices.

b) Which measure of central tendency best represents the price of an item at the gallery? Explain.

3.5 Measures of Spread, pages 140–147

11. Find the range, the variance, and the standard deviation for each set of data.

a) 28, 51, 91, 47, 56, 77, 64, 52, 71, 63

b) 202, 205, 213, 197, 200, 190, 198, 195

12. If you were the general manager for an NBA basketball team, would you prefer a larger or smaller standard deviation for player heights on your team? Explain.

3.6 Common Distributions, pages 148–155

13. Describe the characteristics of each distribution and give an example of each.

a) a skewed distribution

b) a bimodal distribution

c) a normal distribution

14. The table shows the results of a test out of 100.

Mark Interval	Tally	Frequency
[30–40)	|	
[40–50)	||||	
[50–60)	⊬⊬ |	
[60–70)	⊬⊬ ⊬⊬	
[70–80)	⊬⊬ ||	
[80–90)	|||	
[90–100]	|	

a) Copy and complete the table. Use the data to create a histogram.

b) Do the data appear to be normally distributed? Explain.

For questions 1 to 3, choose the best answer.

1. The mean value of a set of data is found by:

   A adding the values and dividing by the number of values in the set

   B putting the values in order and finding the one located in the middle

   C finding the value that occurs most often

   D subtracting the least value from the greatest value

2. Any survey that asks people to return their completed questionnaires by mail is subject to:

   A response bias   **B** non-response bias

   C sampling bias   **D** no bias at all

3. A population is split into groups. Members of each group are selected randomly in proportion to their number in the population. The sampling technique used was:

   A cluster sampling

   B voluntary-response sampling

   C stratified random sampling

   D simple random sampling

4. Classify the data as coming from a primary or a secondary source.

   a) a survey of your friends

   b) an Internet search to find information

   c) a phone poll conducted by a radio station

   d) data given to you that a friend collected

5. Identify the bias in each survey question.

   a) *The government has been underfunding our transit system for years. Should we turn to private-sector investment to finally solve our transit problems?*

   b) *Please select your favourite type of television show.*
   A Sports
   B Reality show
   C Crime drama
   D Other: _____

6. Rewrite each survey question in question 5 to remove the bias.

7. A nut and a washer are to fit onto the end of a bolt. The centre of the washer must have a mean diameter of 1.5 cm. The variance of the washers is 0.01 cm. The centre of the nut must have a diameter of 1.15 cm, with a standard deviation of 0.01 cm. Assume the diameters are distributed normally.

   a) What is the standard deviation of the washers?

   b) Give the range of centre diameters necessary for the washers to be within two standard deviations of the mean.

   c) Give the range of centre diameters necessary for the nuts to be within two standard deviations of the mean.

   d) Why do the nuts and the washers have different standard deviations? Explain.

Chapter Problem Wrap-Up

In each section, you followed Andrew during his co-op placement at a marketing research company as he learned about and applied statistics. As part of his final project, Andrew studied the attendance at the local fall fair over the last 20 years. Write a report of this data for Andrew. Include a graph of the data and explain your choice of graph. Use your graph to analyse the data's distribution.

8. A large doughnut chain recorded the opening week sales for eight new shops

   $37 500, $42 300, $58 800, $31 300, $41 800, $37 100, $63 200, $58 000

   a) Find the range, the mean, the median, and the mode for the data.

   b) Are sales at the new shops likely to increase or decrease after the opening week? Explain.

9. A manufacturer of rechargeable batteries tests 10 batteries at random and records the time it takes, in minutes, to drain each battery after a full charge has been applied.

   195, 203, 177, 186, 191,
   225, 216, 202, 197, 218

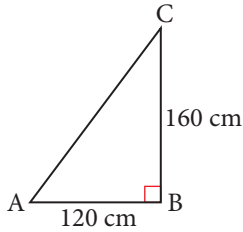   Find the range, the variance, and the standard deviation of the data.

10. The graph shows the shoe sizes of a sample of men.

Shoe Store Sample
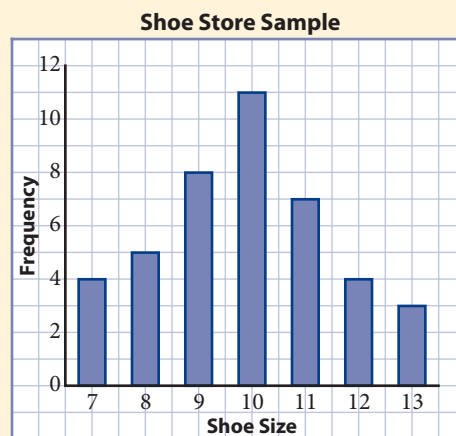
a) Identify the type of graph.

b) Identify the type of distribution.

Chapter 1 Trigonometry

1. Write the three trigonometric ratios for angle A. Express each answer as a fraction in lowest terms.

2. Botanists want to confirm the height of the world's tallest redwood tree. They walk 10 m from the base of the tree and measure the angle of elevation to the top to be 85.05°. How tall is the tree, to the nearest metre?

3. Solve each triangle.

a)

b)
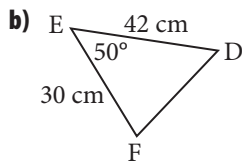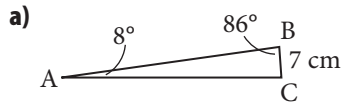
4. Solve for *e*.
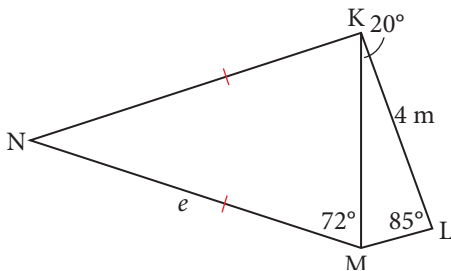
Chapter 2: Probability

5. Three dice were rolled 48 times. Their sum was greater than 8 a total of 36 times. Determine the experimental probability of rolling a sum greater than 8.

6. Suppose two ten-sided dice, with the numbers 1 to 10 on the faces, were rolled. Find the probability of rolling each sum. Express each answer as a fraction in lowest terms.

a) 3
b) 11
c) not 11
d) 18 or less
e) a sum that is divisible by 10

7. In a bag, there are 11 green marbles, 18 purple marbles, and 1 red marble. A marble is removed, the colour is recorded, and then it is put back into the bag. This is repeated for a total of 30 times. The results are displayed on the bar graph.
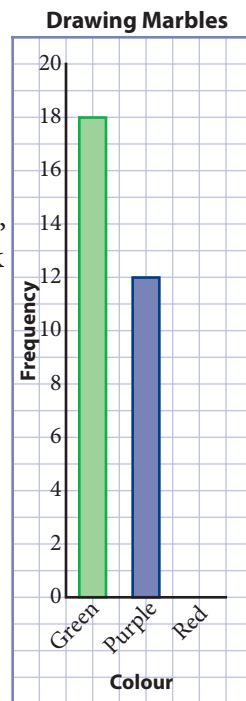
a) What is the experimental probability of drawing each colour marble? Express your answers as a percent.

b) What is the theoretical probability of drawing each colour marble? Express your answers as a percent.

8. A baseball player has 60 hits out of 300 at-bats so far this season.

 a) What percent of at-bats has he obtained a hit?

 b) If he has 60 more at-bats this season, how many hits would you expect him to get?

 c) Suggest some factors that might affect your estimate.

Chapter 3: One Variable Statistics

9. In each situation, identify the sample and the population.

 a) Ken filled out the survey in the back of his cycling magazine and mailed it to the publisher.

 b) Chinedu went door to door in his neighbourhood to ask people for which mayoral candidate they planned to vote.

10. Identify the bias in each survey. Suggest how it might be removed.

 a) At the end of the week, there are five entries in the cafeteria suggestion box, and they all say "No more meatloaf!"

 b) Nobody in the biology class raised their hand when the teacher asked who had not started their term project yet.

 c) The student council asks two people from each homeroom: *Do you think it's fair that the principal is planning to ban cell phones in the hallways and cafeteria, even though they don't cause any problems?*

11. Mr. Lowery collected the votes from a mock election his history classes held for the new local MPP.

Candidate	Number of Votes
Raj Pateel	44
Bernice Kryzinsky	37
Evelyn Cho	22
Jacob Mueller	5
Spoiled ballots	8

 a) Graph the data.

 b) List three facts you can gather from the graph.

12. The table shows the number of times a coin was tossed before getting heads and tails at least once.

Number of Flips	Frequency
2	23
3	11
4	6
5	4
6	1

 a) Find the mean, the median, and the mode of the data.

 b) Which measure of central tendency best describes the data? Explain.

13. Calculate the range, the variance, and the standard deviation for each set of data.

 a) the ages of people at a family reunion: 2, 2, 7, 10, 16, 17, 39, 42, 44, 45, 49, 50, 51, 56, 68, 70, 77

 b) number of tracks on CDs: 8, 13, 15, 11, 16, 13, 10, 11, 11, 14, 12

14. Draw a box-and-whisker plot for the data in question 13, part a).

Road to the Stanley Cup

Making the play-offs is just the first step to winning the Stanley Cup. Teams must play up to four best-of-seven rounds in the play-offs. You will simulate the results of one round of the play-offs for the Toronto Maple Leafs.

In a best-of-seven series, the first team to win four games advances to the next round. The team with the better record during the regular season gets the home-ice advantage for games 1, 2, 5, and 7.

Work in groups of four to collect and display the data.

1. To generate random numbers on your calculator, start by keying four random numbers into the random number generator function. (Use the last four digits of your student number or your phone number.) Press MATH. Select **PRB** and then **1:rand**. Press MATH. Select **PRB** and then **5:randInt(**. Press ENTER. Type 1 ⬚, ⬚ 10 ⬚, ⬚ 7. Press ENTER. You now have seven random numbers from 1 to 10.

   Suppose the Leafs are playing a team with a slightly better record during the regular season. The Leafs have a 50% chance of winning a home game and a 40% chance of winning an away game.

   For an away game (games 1, 2, 5, and 7):
   • If the numbers 1, 2, 3, or 4, appear, the Leafs win their game.
   • If the numbers 5, 6, 7, 8, 9, or 10 appear, the Leafs lose their game.

   For a home game (games 3, 4, and 6):
   • If the numbers 1, 2, 3, 4, or 5, appear, the Leafs win their game.
   • If the number 6, 7, 8, 9, or 10 appear, the Leafs lose their game.

   Consider the random numbers in the screen capture.
   • {1 2 1 1 4 8 8} means the Leafs win in 4 games.
   • {5 10 6 4 1 2 2} means the Leafs win in 7 games.
   • {9 9 7 9 4 9 5} means the Leafs lose in 4 games.
   • {3 3 9 1 2 3 7} means the Leafs win in 5 games.

   Notice that it may not take all seven games to declare a winner.
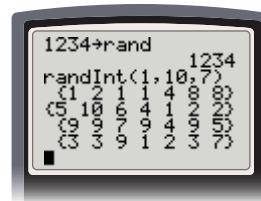
2. Continue pressing ENTER until you have the results of 20 simulations. Make a table to record your results. For each simulation, note if the Leafs win or lose, and the number of games that it takes.

3. Record the data from each group of students in a tally sheet to find the class results.

4. What type of graph would be appropriate to compare the number of times the Leafs win the series to the number of times the Leafs lose the series? Graph the data.

5. Look at all the simulations that resulted in the Leafs winning the series. Calculate the mean, the median, and the mode number of games. Which measure of central tendency makes the most sense for this data?

6. What predictions could you make about the outcome of the series based on your simulation?

7. Suppose the probability of winning a game remains the same for all teams the Leafs face during the play-offs. What predictions can you make about the probability of the Leafs winning the Stanley Cup based on your simulation? Justify your answer.