2 Working With Data

Key Words

median
mode
data set
mean
measure of central tendency
stem-and-leaf plot
weighted mean
range
outlier
trimmed mean
percentile
percentile rank
scatter plot
trend
independent variable
dependent variable
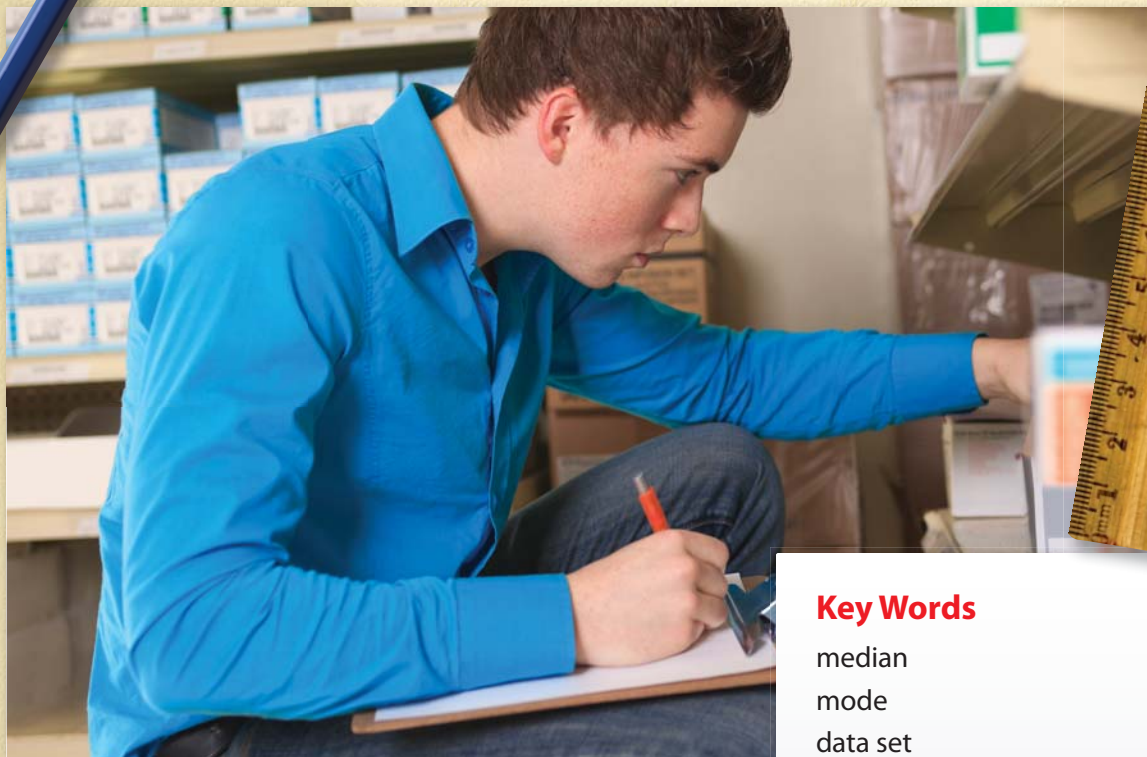
Matthew is training to be the manager of a store. He is checking inventory and analysing data about last month's sales. The data will help him decide what items to order and how much to order.

1. What kind of data would a manager need to place orders?

2. What are some ways Matthew might organize the data he collects?

3. Why would a manager not want to order too few or too many items?

Career Link

Brian went to college to train to be a retail store manager. He then worked as an apprentice in a store to learn about the job. He hopes to soon be running his own store. Retail store managers order products. They collect, organize, and analyse data about buying patterns to help them to place orders. They also hire and train employees, deal with customers, and make sure the store runs smoothly.

Ordering Data

1. Arrange the data sets in ascending order.

> *Ascending means from smallest to largest*

a) 75, 68, 57, 82, 62, 90, 77

b) 50, 12, 29, 48, 3, 34, 0

c) 2.3, 2.1, 1.9, 1.75, 2.50, 1.82, 2.05

d) 84.2, 55.3, 67.4, 59.4, 72.8, 87.1

e) -143, -139, -117, -125, -139, -140, -119

f) -7.2, 1, -4.1, -9.1, 0, 4, 2.2

Average/Mean

2. Calculate the average or mean of each set of data.

> To calculate the average or mean,
> • add the values
> • divide the total by the number of values
> Example: 5, 9, 5, 7, 0, -2
> Mean = $\dfrac{\text{sum of values}}{\text{number of values}}$
> $= \dfrac{5 + 9 + 5 + 7 + 0 + (-2)}{6}$
> $= \dfrac{24}{6}$
> $= 4$

a) 30, 85, 50, 105, 100, 65, 20, 25

b) 55, 47, 90, 12, 4, 83, 31

c) 3.4, 2.2, 1.4, 4.6, 2.0, 1.4, 4.6, 1.6

d) 15.4, 12.7, 19.8, 16.1, 25.8, 13.2, 10.0

e) -15, -20, 10, 8, 0, -2, 12, 4

f) 0, -1.4, 2.7, 8.3, 6.0, -2.7, -8.2, 4.5

Recalculating the Mean

3. For each data set in #2,
  • arrange the data in descending order

> *Descending means from largest to smallest*

  • remove the smallest and largest values from the data set
  • recalculate the mean with the remaining values

Percent

4. Estimate the value of y. Then, calculate.

> Example: 8% of 244 is y.
> To estimate, think: 8% is close to 10%, and 244 is close to 250.
> So, an estimate of y is about 25.
> To calculate, set up a proportion and solve for y.
> $\times\ 2.44$
> $\dfrac{8}{100} = \dfrac{y}{244}$
> $\times\ 2.44$
> $\dfrac{8}{100} = \dfrac{19.52}{244}$
> You can also use your calculator.
> Enter these keystrokes:
> C 244 × 8 2nd % = 19.52

a) 65% of 48 is y.

b) 22% of 160 is y.

c) 97% of 150 is y.

d) 41% of 55 is y.

Partitioning Data

5. For each data set,
- arrange the data in ascending order
- state which value(s) are in the middle of the set

a) 55, 24, 15, 60, 32, 49, 13

b) 81, 52, 64, 43, 81, 39

c) 88.7, 54.2, 63.8, 94.0, 24.9

d) 1.2, 5.3, 2.7, 9.9, 4.8, 6.2

e) $-5, -19, -1, -14, -10, -17$

f) 3.4, -5.7, 2.5, 8.4, 1.6

6. For each data set,
- arrange the data in ascending order
- state which values are in the lower half of the data set
- state which values are in the upper third of the data set

a) 7, 19, 25, 4, 4, 31, 10, 17, 36, 2, 19, 22

b) 204, 199, 206, 243, 187, 163, 234, 216, 157, 222, 217, 239

c) 101.4, 97.6, 88.0, 105.1, 90.7, 100

d) 2.7, 6.1, 5.4, 1.9, 8.3, 7.4, 10.0, 9.8, 3.5

e) 24, -15, 0, 11, -3, 6, -21, 25, -11, 12, -6, 4, -15, -9, 5, -25, 7, 19

f) 19.4, -3.8, -0.2, 17.2, 8.6, -14.5

Frequency Table

7. Place the data in a frequency table similar to the one below. Complete the table by recording the tally marks and frequency.

> *Frequency* is the total number of tally marks for each category.

a) M, S, XL, S, S, L, L, L, M, XL, XL, M, M, S, L, S, XL, L, M, M, L

Size	Tally	Frequency
S		
M		
L		
XL		

b) 62, 61, 64, 62, 63, 62, 62, 63, 60, 61, 62, 63, 64, 60, 60, 61, 62, 63, 64

Time (s)	Tally	Frequency
60		
61		
62		
63		
64		

Algebra

8. Rearrange each equation to solve for x.

a) $\dfrac{x}{13} = 2$

b) $27 = \dfrac{29 + 24 + 28 + x}{4}$

c) $54 = \dfrac{56 + 54 + 52}{x}$

2.1

Measures of Central Tendency

Focus On . . .

- determining the three measures of central tendency
- determining weighted means
- solving a problem involving measures of central tendency

Materials

- metre stick or measuring tape
- calculator

median

- the middle number in a set of data after the data have been arranged in order
- the median of 2, 4, 6, 8, 11 is 6
- when there is an even number of data values, average the two middle values to find the median
- the median of 1, 5, 9, 13, 16, 20 is 11, because
$$\frac{9 + 13}{2} = \frac{22}{2}$$
$$= 11$$

Health care workers collect and analyse data about the growth of children. They compare a child's height and weight to the typical height and weight of children of the same age.

Explore Measures of Central Tendency

1. Work with a partner. Measure each other's height to the nearest centimetre.

2. Collect height data from all members of the class. List the male and female data in a table.

Height of Males	Height of Females

3. Arrange the data for males and females in ascending order. What is the middle value for males? for females? The middle value is the **median**.

mode

- the number(s) that occurs most frequently in a set of data
- a data set can have no mode, one mode, or more than one mode

data set

- a collection of related information

mean

- the average of the data values
- add the data values and divide the total by the number of data values
- Example: The mean of 4, 7, 9, 10, 11, 15 is

$$\frac{4 + 7 + 9 + 10 + 11 + 15}{6}$$

$$= \frac{56}{6}$$

$$= 9.3$$

F.Y.I.

Mean and *average* are the same thing. This book will use both terms.

measure of central tendency

- a value that represents the centre of a set of data
- can be the mean, median, or mode

4. Does any data value appear more than once for males? for females? If yes, what is the data value? This value is the **mode** of the **data set**.

5. Determine the sum of the values in each data set. Divide the sum by the number of values. This value is the **mean** of each data set.

6. Reflect Imagine you wish to describe the typical height of a grade 12 boy and a grade 12 girl. Discuss the advantages and disadvantages of each of the following statements:

a) The typical height of a grade 12 boy and a grade 12 girl is the median of each data set.

b) The typical height of a grade 12 boy and a grade 12 girl is the mode of each data set.

c) The typical height of a grade 12 boy and a grade 12 girl is the mean of each data set.

7. Based on your discussions in step 6, which **measure of central tendency** do you think best describes the typical height of a grade 12 boy and a grade 12 girl? Explain your answer.

8. Extend Your Understanding Would your answer to step 7 be affected by the number of values in the data set? Explain.

9. Why do you think the data values were divided into two sets, one for males and one for females?

10. Find a growth chart on the Internet.

a) Compare the range of heights you collected from your class to the range of heights given for your age in the chart.

b) Explain why the ranges might be different.

Web Link

Growth charts compare different measurements of a child's development over time. There are many different types of growth charts, such as height vs. age, weight vs. age, and weight vs. height. To learn more about growth charts and how they are used, go to www.mcgrawhill.ca/books/mathatwork12 and follow the links.

On the Job 1

Determine the Mode

Frank, a cafeteria manager, wants to know which type of sandwich is the most popular. He records the sales of each type of sandwich for one week:

- tuna salad: 40
- egg salad: 58
- grilled veggie: 76
- ham and cheese: 65
- turkey: 47
- BLT: 61
- assorted meat: 53
- grilled chicken and avocado: 41
- steak and cheese: 76
- pastrami: 68

stem-and-leaf plot

- a way to organize numerical data in order of place value
- the "tens digit and greater" is the stem and the "ones digit" is the leaf
- Example: a stem-and-leaf plot for 2, 15, 18, 29 is

Stem (tens)	Leaf (ones)
0	2
1	5 8
2	9

- to plot decimal numbers, the "ones digit and greater" is the stem and the "tenths and less" is the leaf
- Example: a stem-and-leaf plot for 0.8, 2.1, 2.7, 10.8 is

Stem (ones and greater)	Leaf (tenths and less)
0	8
2	1 7
10	8

a) Create a **stem-and-leaf plot** for the data set.

b) Determine the mode.

c) Is the mode the best measure to use to determine the most popular sandwich? Explain.

Solution

a) Place each value in the stem-and-leaf plot. For the data value 40, the 4 is in the tens place, so it is a stem. The 0 is in the ones place, so it is the corresponding leaf to the stem of 4.

Stem (tens)	Leaf (ones)
4	0 7 1
5	8 3
6	5 1 8
7	6 6

Arrange the leaves in ascending order.

Stem (tens)	Leaf (ones)
4	0 1 7
5	3 8
6	1 5 8
7	6 6

b) Two types of sandwiches were chosen most often. The grilled veggie sandwich and the steak and cheese sandwich were each chosen 76 times. So, the modes are grilled veggie and steak and cheese.

c) Since the grilled veggie sandwich and the steak and cheese sandwich were chosen the most number of times, Frank can say that they are the most popular. The mode is the best way to summarize the most popular sandwich.

> You can also find the mode in a data set of number values.
>
> In the data set 20, 14, 13, 9, 20, 7, 18, 9, there are two modes.
>
> 9 and 20 both occur twice so they are the modes.

Your Turn

The fuel consumption for ten hybrid cars of the same model was tested. The following results were obtained (in L/100 km): 5.5, 6.0, 4.9, 6.2, 4.8, 5.1, 6.2, 5.7, 4.8, 6.5.

a) Determine the mode(s).

b) Should you use the mode(s) to represent the fuel consumption of these ten cars? Why or why not?

Try It

1. Determine the mode for each data set.

a) 84, 68, 71, 55, 66, 63, 82, 92, 70, 75, 64, 58, 76, 88, 65, 73, 76, 73, 62, 83

b) 129, 150, 144, 136, 130, 152, 159, 120, 146, 144

2. Determine the mode for each data set. **Hint:** Rewrite the stem-and-leaf plot to place the values in ascending order.

a)

Stem (tens)	Leaf (ones)
1	0 8 2 5
2	9 1 0 0 4
3	5 6 1 2 5
4	4 0 5

b)

Size	Frequency
6	12
7	21
8	30
9	16
10	8

3. The mode of each set of data is 70. Determine the missing value, n.

a) 59, 85, 70, 62, 42, 66, n, 77

b) 60, 69, 70, 60, 75, 80, n, 70, 79

Apply It

4. Ten babies were born in a hospital on one day. Their lengths, in centimetres, are 46, 43, 45, 41, 44, 42, 51, 40, 51, 48.

a) What is the mode?

b) Is the mode the best way to represent the lengths of the babies? Why or why not?

5. A real estate agent wants to know which type of property she has sold the most in the past five years. She researches her sales and finds the following information.
- semi-detached homes: 12
- detached homes: 26
- condominiums: 4
- property with no building: 11

a) What is the mode?

b) Is the mode the best way to represent the type of property she sells the most? Why or why not?

6. The manager of a baseball team orders team jerseys. These sizes are needed: 40, 36, 40, 36, 38, 42, 36, 34, 34, 38, 40, 36, 38, 36, 38.

a) Create a frequency table to show the data.

Size	Tally	Frequency

b) What is the mode for the data set?

c) Is the mode the best way to represent the typical size worn by a team member? Explain.

On the Job 2

Determine the Median

A fish and wildlife officer records the weights of trout caught in a lake. Their weights to the nearest tenth of a pound are 2.9, 1.5, 2.8, 1.1, 1.0, 2.7, 3.0, 1.6, 1.9, 2.3.

a) Determine the median weight.

b) Is the median the best measure to represent the typical weight of a caught fish? Explain.

Solution

a) Arrange the data in order from least to greatest:

1.0, 1.1, 1.5, 1.6, 1.9, 2.3, 2.7, 2.8, 2.9, 3.0

Cross out pairs of least and greatest values until you reach the middle one or two numbers:

~~1.0~~, ~~1.1~~, ~~1.5~~, ~~1.6~~, 1.9, 2.3, ~~2.7~~, ~~2.8~~, ~~2.9~~, ~~3.0~~

Since two values remain, find the average of 1.9 and 2.3:

$$\frac{1.9 + 2.3}{2} = 2.1$$

The median weight is 2.1 lb.

b) The median of 2.1 is a good measure to represent the typical size of a trout caught because 50% of the values are below 2.1 and 50% of the values are above 2.1.

Your Turn

The lowest temperature for each day in January was recorded:

-10, -8, -15, -6, -9, -5, -11, -20, -18, -18, -14, -11, -8, -2, -3, -17, -19, -14, -16, -8, -15, -13, -22, -25, -20, -18, -3, -2, -1, 0, -4

a) Organize the data set in a stem-and-leaf plot.

b) Determine the median.

c) "The most likely low temperature for January is the median." Is this statement reasonable? Why or why not?

The numbers −2 and −4 in a stem-and-leaf plot look like this:

Stem (tens)	Leaf (ones)
0	−4 −2

Try It

1. Find the median for each data set.

a) 21, 22, 19, 24, 23, 19, 20, 24

b)

Stem (ones)	Leaf (tenths)
0	4 8
1	0 1 1 2 3
2	0 0 6 8
3	1 5

c)

Length (cm)	Frequency
40	1
50	3
60	8
70	9
80	2

2. Taylor's calculation of the median for the following data set is incorrect.

Data set: 1.1, 0.5, 3.5, 0.6, 2.9, 1.5, 1.1, 1.3
Solution: 0.5, 0.6, 1.1, 1.3, 1.5, 2.9, 3.5
The median is 1.3.

a) What error did she make?

b) Determine the correct median.

Apply It

3. The RCMP raise and train dogs for their canine law enforcement work. Nine German shepherds are being trained. Their weights in kilograms are 38, 42, 44, 43, 40, 41, 39, 42, 38.

a) Arrange the data in a stem-and-leaf plot.

b) What is the median weight?

c) Suppose a tenth dog with a weight of 40 kg is added to the group. What is the new median?

4. The winning times for the men's 100-m sprint in the Commonwealth Games from 1970 to 2010 are listed in the table below.

Year	Time (s)
1970	10.20
1974	10.38
1978	10.03
1982	10.02
1986	10.07
1990	9.93
1994	9.91
1998	9.88
2002	9.98
2006	10.03
2010	10.12

What degree of precision is used to record the winning times? How do you think officials make sure the times are accurate?

a) What is the median?

b) A sports researcher makes the following statement: "The most likely winning time for the 100-m men's sprint at the next Commonwealth Games will be the median of this data set." Do you agree? Why or why not?

5. Marsha and Ben are personal trainers. They create a stem-and-leaf plot to record the pulse rates of 20 people after a 15-minute brisk walk.

Stem (tens)	Leaf (ones)
5	5 8
6	2 8 4 6 5 3
7	3 5 0 3 3 1 6
8	4 8 3 2
9	2

a) Marsha concludes that the median of the data set is 74. Ben says it is 72. Who is correct? Explain.

b) What error could have been made by the person who had the incorrect median?

Calculate the Mean and Weighted Mean

The table shows Tristan's physics marks.

Course Work	Mark (%)
Homework	90
Quizzes	62
Projects	68
Exams	75
Oral presentations	85

a) What is Tristan's average mark for physics?
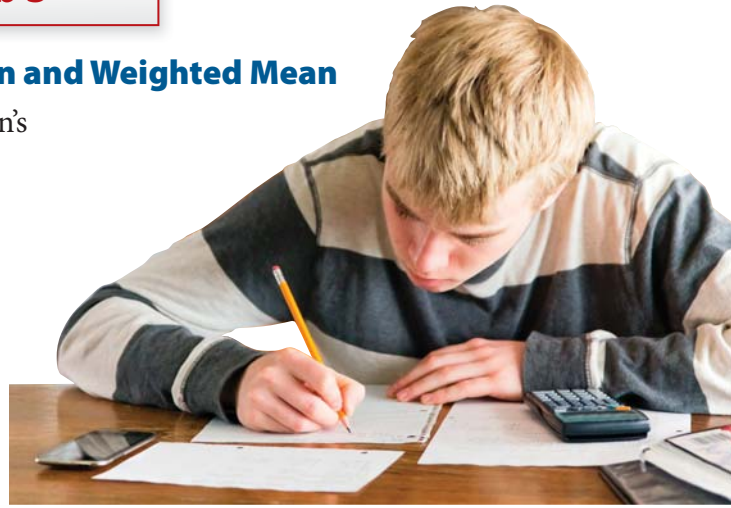
weighted mean

• the average or mean of a data set in which each data point does not contribute an equal amount to the final average

b) Tristan's teacher decides to use a **weighted mean** to determine the final mark:
  • homework counts for 10%
  • quizzes count for 20%
  • projects count for 15%
  • exams count for 50%
  • oral presentations count for 5%

What is Tristan's final mark based on the weighted mean?

c) Which gives Tristan a higher mark, the mean or the weighted mean? Explain.

d) Why do you think the teacher used a weighted mean?

Solution

a) To calculate the average, find the sum of the five marks and divide by the number of marks:

$$\text{Mean} = \frac{90 + 62 + 68 + 75 + 85}{5}$$
$$= \frac{380}{5}$$
$$= 76$$

Tristan's average mark in physics is 76%.

b) *Method 1: Calculate in Steps*

Step 1: Check that the weightings add up to 100%.

Total weightings $= 10 + 20 + 15 + 50 + 5$
$= 100$

Step 2: Multiply the weightings by each of Tristan's marks.

Course Work	Weighting (%)	Tristan's Mark (%)	Product of Weighting and Mark
Homework	10	90	$0.1 \times 90 = 9$
Quizzes	20	62	$0.2 \times 62 = 12.4$
Projects	15	68	$0.15 \times 68 = 10.2$
Exams	50	75	$0.5 \times 75 = 37.5$
Oral presentations	5	85	$0.05 \times 85 = 4.25$

Step 3: Find the sum of the last column.

$9 + 12.4 + 10.2 + 37.5 + 4.25 = 73.35$

Tristan's weighted mean mark is 73%.

Method 2: Calculate Using the Weighted Mean Formula

$$\text{Weighted mean} = \frac{\text{sum of the product of each item and its weighting}}{\text{sum of the weightings}}$$

Course Work	Weighting, w	Tristan's Marks, x	Product of Weighting and Mark, wx
Homework	10	90	$10 \times 90 = 900$
Quizzes	20	62	$20 \times 62 = 1240$
Projects	15	68	$15 \times 68 = 1020$
Exams	50	75	$50 \times 75 = 3750$
Oral presentations	5	75	$5 \times 75 = 425$
Totals	100		7335

$$M_w = \frac{\Sigma xw}{\Sigma w}$$
$$= \frac{7335}{100}$$
$$= 73.35$$

Tristan's weighted mean mark is 73%.

c) The mean gives Tristan a higher mark because all the marks have equal weightings. When determining the weighted mean, Tristan's higher marks are given lower weightings, which results in a lower weighted mean.

d) The teacher used a weighted mean because she believes that some course work assesses more material than other course work (for example, exams cover more topics than oral presentations). Therefore, she believes that some work should be worth more.

Your Turn

The table shows Janelle's geography marks:

Course Work	Mark (%)
Homework	80
Quizzes	77
Projects	75
Oral presentations	82
Exams	72

a) What is the average of Janelle's marks?

b) What is the weighted mean, if each subject has the following weightings?
- homework counts for 20%
- quizzes count for 20%
- projects count for 10%
- oral presentations count for 10%
- exams count for 40%

c) Janelle is updating her résumé by including her current averages. Should she use the weighted mean for her geography marks? Explain.

Puzzler

The number in each circle is the mean of the numbers in the squares connected to it. Find the number in each square.

18   9

14
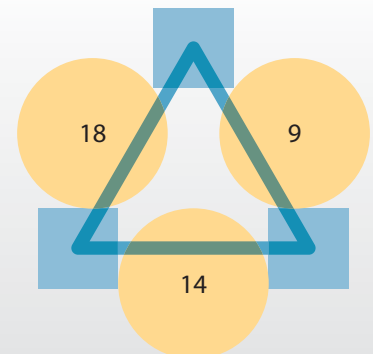
Try It

1. Determine the average of each set of data to the nearest tenth.

a) 99, 101, 105, 116, 96

b) 35, 50, 47, 21, 41, 38, 32

c) 19.1, 23.8, 25.6, 17.4, 14.0, 15.9

d) 2.2, 5.7, 1.3, 9.8, 4.6, 3.1

e) $-18, -22, -22, -17, -30, -18, -12$

f) $-5, 8, 7, -6, 4, -2, 0, -7, 3$

2. Copy this table. Then, calculate the missing values to find the weighted mean.

Value, x	Weighting, w (%)	Product of Value and Weighting, wx
7.5	25	
18.0	40	
12.6	10	
9.0	10	
20.8	15	
Totals		

Apply It

3. A school is holding a food drive for the local food bank. The table shows how much food each grade collected.

Grade	Number of Students	Total Kilograms Collected
7	20	100
8	24	96
9	22	99
10	20	120
11	18	90
12	26	130

a) Which grade collected the most kilograms of food?

b) What is the mean number of kilograms collected for each student in each grade?

c) Which grade should win a prize for most kilograms collected? Explain your choice.

4. The table shows the number of hours per week Jacob worked at his part-time job.

Week	Number of Hours
1	19
2	25
3	18
4	23
5	n

The mean for the five weeks was 21 hours. How many hours did Jacob work during the fifth week?

5. Slugging percentage in baseball is calculated as a weighted mean. The different types of hits are weighted as shown in the table.

Type of Hit	Weighting
Single	1
Double	2
Triple	3
Home run	4

Batting average is calculated differently. To calculate batting average, divide the number of base hits by the number of at bats:

$$\text{Batting average} = \frac{\text{number of base hits}}{\text{number of at bats}}.$$

Why could you say that slugging percentage is a better indicator of good hitting than batting average?
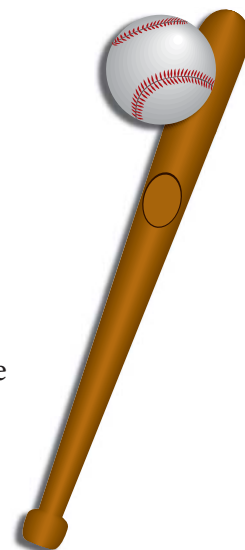
Work With It

1. Mrs. MacDonnell has two grade 12 English Literature classes.

Her morning class has 18 students. The grades, out of 100, on an exam were 62, 71, 58, 84, 90, 55, 75, 68, 80, 52, 92, 77, 50, 46, 95, 80, 60, 73.

Her afternoon class has 25 students. The grades on the same exam were 84, 68, 71, 55, 46, 63, 72, 92, 70, 65, 54, 48, 73, 88, 65, 73, 76, 53, 62, 90, 40, 85, 72, 79, 55.

a) Create a stem-and-leaf plot for each data set.

b) What are the mode(s), median, and average for each data set?

c) What is the average mark for the 43 students?

2. James is a supervisor at a plumbing supplies warehouse. One of his jobs is to inspect and record the weight of supplies that are received on shipment days. The table shows the data for the month of June.

Date	Mass (tonnes)
2	45
6	20
9	40
13	43
16	29
20	36
23	59
27	28
30	20

a) For the data set of masses, determine each of the following:

  i) mode(s)

  ii) median

  iii) mean

b) James is explaining to a job applicant the number of tonnes handled on shipment days. Which measure of central tendency should he use? Explain your answer, and state the advantages and disadvantages of each measure.

c) In James's monthly report, he makes the following statement: "During the nine shipment days in June, the average shipment was 35.6 tonnes per day." From this report, the accountant was able to calculate the total amount shipped. Which measure of central tendency did James use? How was the accountant able to calculate the total amount shipped?

3. The manager of a shoe department in a sporting goods store is placing a monthly order for women's athletic shoes. The sizes of shoes that were sold last month are shown below.

Size	Frequency
6	2
6.5	9
7	23
7.5	18
8	24
8.5	17
9	10
9.5	9
10	6
10.5	5
11	4

a) Determine the three measures of central tendency.

b) Which measure do you think is the most important to the manager? Explain why.

Discuss It

4. Discuss the following with a partner.

a) If you know a runner's typical running times, you can predict how well the runner will do in a future race. Which measure of central tendency would you use to predict a runner's future race time? Explain.

b) A survey was conducted to determine which type of movie grade 12 students like best: comedy, action-adventure, science fiction, horror, or romance. Which measure of central tendency should the surveyor use to determine the results of the survey? Explain.

5. Find information on the Internet, in newspapers, or in magazines in which measures of central tendency are used. Possible examples could relate to a sports team, a demographic statistic, or weather information. Discuss advantages and disadvantages of the measures of central tendency used.

6. a) Explain how an advertiser could use measures of central tendency to create advertisements or commercials to try to convince consumers to buy a product. Find examples to share with the class.

b) Is it possible that the use of measures of central tendency could mislead the consumer? Explain. If possible, find examples.

7. A web site uses two measures of central tendency to show data on Canadian workers.
- One graph shows the median salary of Canadians by job.
- One graph shows the average number of weeks of vacation for workers by years of experience.

What advantages and disadvantages are there in using the median for the first graph and the average for the second graph?

Using Other Statistical Measures

Focus On . . .

- finding the range for a data set
- identifying outliers in a data set
- calculating the trimmed mean for a data set
- working with percentiles and percentile ranks
- solving problems involving trimmed mean and percentiles

Change Islands, NL

Materials

- computer with Internet access

range

- the difference between the largest value and the smallest value of a data set

outlier

- a value that is much smaller or larger than the other data values
- a data set may have no outliers, one outlier, or more than one outlier

The weather in Canada can vary significantly from year to year.

Explore Collecting and Working With Data

1. Choose a city and a month to research. Search the Internet to find data about the total snowfall in the city in that month for ten years.

2. What is the **range** for the data set?

3. What is the average for the data set?

4. Some years in the data set may have had much more or much less snow than the other years. These values are called **outliers**. List any outliers in your data set.

5. **Reflect** Calculate the range and the average of the data set without the outliers.

6. How are the range and the average in steps 2 and 3 different from the range and average in step 5?

7. Extend Your Understanding

a) Determine the mode and median in the original data set and in the data set with the outliers removed. Draw a table similar to the one below and complete it.

| Measure of Central Tendency	Original Data Set	Data Set With Outliers Removed
Range		
Mean		
Mode		
Median		

b) A friend asks you about the maximum amount of snow that falls in the city you chose. How can you use the information in the table you created to answer the question?

Strategy

Develop Alternative Approaches

Puzzler

A farmer has six horses that must be kept in separate pens. He has used 13 lengths of fencing to create six pens of equal size for the horses.
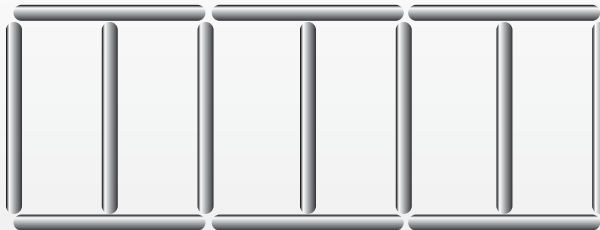
One morning, the farmer discovers that someone has stolen one of the lengths of fencing.
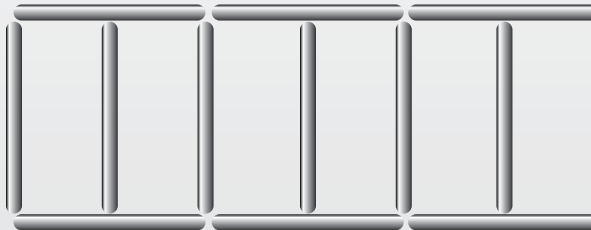
How can he use his remaining 12 lengths of fencing to create six equal-sized pens for the horses?

The Effect of Outliers on Measures of Central Tendency

The table below shows the NHL point leaders for two seasons.

1985–1986 Season	Points	2010–2011 Season	Points
Gretzky	215	D. Sedin	104
Lemieux	141	St. Louis	99
Coffey	138	Perry	98
Kurri	131	H. Sedin	94
Bossy	123	Stamkos	91
Stastny	122	Iginla	86
Savard	116	Ovechkin	85

trimmed mean
- a calculation of the mean found by removing the highest and lowest values
- you must remove the same number of values from the top and bottom of the data set
- removing outliers can result in a more accurate mean

a) Determine the range, mode, median, and mean of the data for each season.

b) Which data value appears to be an outlier in one of the two seasons? Explain what might have caused the outlier.

c) Remove the highest and lowest value for each season and find the **trimmed mean**. How does the trimmed mean for each season compare to the mean calculated in part a)?

d) Does removing the highest and lowest values affect the mode and the median? Explain.

e) Should the outlier be removed? Explain why or why not.

Solution

a) To calculate the range, find the difference between the largest data value and the smallest data value.

Season	1985–1986	2010–2011
Range	$215 - 116 = 99$ The range is 99.	$104 - 85 = 19$ The range is 19.
Mode	None	None
Median	~~215~~, ~~141~~, ~~138~~, 131, ~~123~~, ~~122~~, ~~116~~ The median is 131.	~~104~~, ~~99~~, ~~98~~, 94, ~~91~~, ~~86~~, ~~85~~ The median is 94.
Mean	$\dfrac{215+141+138+131+123+122+116}{7}$ $= \dfrac{986}{7}$ $= 140.8$ The mean is 140.8.	$\dfrac{104+99+98+94+91+86+85}{7}$ $= \dfrac{657}{7}$ $= 93.8$ The mean is 93.8.

b) In the 1985–1986 season, the data value of 215 is the outlier because it is much larger than the other data values. The difference between the first data value, 215, and the second data value, 141, is 74 points. The difference between all other subsequent data values is between 1 and 8 points. Outliers are sometimes a result of incorrect data entry or errors in calculations. In this case, Wayne Gretzky's scoring ability was a phenomenon.

c) To find the trimmed mean, remove the highest and lowest scores. Then, calculate the mean.

1985–86 Season:

$$\text{Trimmed mean} = \frac{986 - 215 - 116}{5}$$
$$= \frac{655}{5}$$
$$= 131$$

The trimmed mean is 131. The mean with the outlier is 140.8. The trimmed mean is smaller than the original mean.

2010–11 Season:

$$\text{Trimmed mean} = \frac{657 - 104 - 85}{5}$$
$$= \frac{468}{5}$$
$$= 93.6$$

The trimmed mean is 93.6. The mean with the outlier is 93.8. The trimmed mean and the original mean are very close.

d) Removing the highest and lowest values does not affect the mode or the median. There is no mode for either season. The middle value stays the same for both seasons.

> Would removing the highest and lowest values ever affect the mode in a data set?

e) Deciding whether to remove the outlier depends on why you are calculating the average of the top seven point leaders. If the purpose is to make a statement about Wayne Gretzky's ability to shatter records, then the outlier should not be removed. If the purpose is to compare the top point scorers' averages for two seasons, then perhaps the outlier should be removed.

Your Turn

Judges in a gymnastics competition gave the following scores, in points, to 14 competitors: 8.5, 9.0, 6.9, 7.5, 7.0, 9.5, 10.0, 5.0, 8.0, 8.0, 7.5, 7.5, 6.8, 6.8.

a) Determine the range.

b) Determine the mode, median, and average, to one decimal place.

c) Identify any outlier(s). What might have caused the outlier(s)?

d) Remove the lowest and highest scores and calculate the trimmed mean.

F.Y.I.

Two sets of judges determine the scores in gymnastics competitions. The A Panel scores the
• difficulty of the routine
• specific elements
• connections between elements

The B Panel judges the execution of the routine, deducting points for errors, faults in technique, and artistry. The B Panel's scores use a trimmed mean, dropping the lowest and highest scores. The final score is the sum of the A Panel's score plus the trimmed mean of the B Panel's score.

Try It

1. Determine the range and identify any outlier(s) in each data set.

a) 20, 34, 20, 15, 3, 20, 6, 10, 13, 16, 10, 50, 21

b) −16, −11, −7, −29, −31, −18, −21, 18

c) 7, 9, 12, 8, 11, 1, 10, 9

d) 44.8, 43.9, 39.9, 42.7, 43.2, 44.6, 45.0, 30.1

2. Calculate the mean and trimmed mean for each data set in #1. Remove the highest and lowest values to calculate the trimmed mean.

Apply It

Tools of the Trade

Horticultural technician is a Red Seal trade. Horticultural technicians use gardening and power tools to grow, plant, and maintain various types of plants, turf, and shrubs. They also install and maintain irrigation systems and construct landscapes.

3. A horticulturist measures the height of ten oak seedlings at the beginning of the week and at the end of the week. The data represent the amount of growth in centimetres for each plant.

3.2, 2.7, 3.1, 2.8, 2.8, 2.7, 4.6, 3.3, 2.6, 3.0

a) What is the range of the data set?

b) What are the measures of central tendency for the data set?

c) Identify any possible outlier(s). The horticulturist wants to write a report on the typical growth of this type of plant. Should she remove the outlier(s) from the data set before writing the report? Explain why or why not.

d) Remove the highest and lowest values and recalculate the range, median, and mean. How did removing the highest and lowest values affect these measures?

4. Data sets may contain no outliers, one outlier, or more than one outlier.

a) Find or create a data set with no outlier.

b) Find or create a data set with one outlier.

5. Jack is the manager at a fish processing plant. He records the number of cases of processed fish produced by the plant each day. The number of cases produced for a three-week period are 652, 785, 758, 699, 314, 801, 713, 713, 681, 1021, 695, 722, 777, 675, 760.

a) Determine the range of the number of cases.

b) Calculate the median and mean number of cases.

c) Identify the outliers in the data set. What may have caused these outliers to occur?

d) Remove the highest and lowest values and calculate the trimmed mean.

e) Will the median change if the highest and lowest values are removed? Why or why not?

f) In Jack's three-week report, he states the typical number of cases produced during the time period. Which measure of central tendency should he use? Explain.

g) Jack's supervisor wants to know the exact number of cases produced. Which measure of central tendency will allow the supervisor to calculate the total number of cases produced during the 15-day period? Explain.

6. The table shows the average annual snowfall for Canada's ten provincial capitals.

Capital City	Average Annual Snowfall (cm)
St. John's	322.3
Charlottetown	311.9
Halifax	230.5
Fredericton	276.5
Québec	315.9
Toronto	115.4
Winnipeg	110.6
Regina	105.9
Edmonton	121.4
Victoria	43.8

a) Ava says the typical snowfall for a capital city is 212.3 cm. Which measure of central tendency did she use? Explain.

b) Sebastian says the typical snowfall for a capital city is 175.95 cm. Which measure of central tendency did he use? Explain.

c) Petra disagrees with both Ava and Sebastian. She says the typical snowfall is 195.4 cm. Which measure of central tendency did she use? Explain.

d) Which measure of central tendency do you think best describes the typical snowfall in a provincial capital city? Justify your choice.

On the Job 2

Percentiles

percentile

- a value below which a certain percent of the data set falls
- the median is also called the 50th percentile, because 50% of the values in the data set are below the median value

Sierra wants to apply for a scholarship so she can attend college. Her cumulative average must be in the 75th **percentile** of her class to qualify. Below are the cumulative averages of her class:

71, 76, 77, 34, 89, 54, 69, 94, 46, 72, 71, 88, 82, 63, 89, 45, 77, 81, 77, 75, 67, 67, 71, 89, 85, 58, 75, 76, 52, 74

Acadia University, Wolfville, NS

a) Construct a stem-and-leaf plot for the data set.

b) Determine the median.

c) What is the minimum average Sierra needs to qualify for the scholarship?

d) What **percentile rank** is a cumulative average of 58?

e) What is the difference between percent and percentile rank?

f) What is the relationship between median and percentile?

percentile rank

- a number between 0 and 100 that indicates the percent of cases that fall at or below that score

Solution

a)

Stem (tens)	Leaf (ones)
9	4
8	1 2 5 8 9 9 9
7	1 1 1 2 4 5 5 6 6 7 7 7
6	3 7 7 9
5	2 4 8
4	5 6
3	4

b) Since there are 30 data values, there are two values in the middle. Mark off 14 data values sequentially from the highest values.

Stem (tens)	Leaf (ones)
9	4
8	1 2 5 8 9 9 9
7	1 1 1 2 4 5 5 6 6 7 7 7
6	3 7 7 9
5	2 4 8
4	5 6
3	4

Mark off 14 data values in order from the lowest value.

Stem (tens)	Leaf (ones)
9	4
8	1 2 5 8 9 9 9
7	1 1 1 2 4 5 5 6 6 7 7 7
6	3 7 7 9
5	2 4 8
4	5 6
3	4

The two middle values are 74 and 75. Find the mean of these two values:

$$\frac{74 + 75}{2} = 74.5$$

The median is 74.5. The cumulative average at the 50th percentile is 74.5.

c) The 75th percentile is the median of the values above the 50th percentile. To find the median, cross off pairs of least and greatest values until there are one or two middle values. Since there are 15 values above the 50th percentile, there will be one middle value.

94, 89, 89, 89, 88, 85, 82, 81, 77, 77, 77, 76, 76, 75, 75

The median of the top 50th percentile, or the value for the 75th percentile, is 81. Sierra would need a cumulative average of at least 81 to qualify for the scholarship.

d) The average mark of 58 is the 6th value from the smallest value. There are $\frac{6}{30}$, or 20%, of the values at or below the average of 58. Therefore, the average of 58 is at the 20th percentile.

e) A percent is a score out of 100. The percentile rank is the percent of cases falling above or below a score. Here are two examples, using the given data set:

Cumulative Average as a Percent	Percentile Rank
81	$\frac{23}{30}$ or 77th rank, because 23, or 77%, of the values in the data set are at or below 81
77	$\frac{19}{30}$ or 63rd rank, because 19, or 63%, of the values in the data set are at or below 77

f) The median is the middle value of a set of data. The percentile is a value below which a certain percent of the data falls. The median is called the 50th percentile because 50% of the values in the data set are at or below the median value.

Your Turn

A group of 20 teenagers with part-time jobs were surveyed about the wage per hour that they earn. Below are the results of the survey.

$10.50, $10.60, $10.00, $17.85, $10.50, $12.75, $15.00, $10.10, $10.00, $11.00, $16.25, $15.00, $10.40, $10.00, $17.00, $16.00, $10.10, $11.25, $10.50, $10.95

a) Organize the data in a stem-and-leaf plot.

b) What wage is at the 50th percentile?

c) Colton makes $11.50/h. At what percentile rank is his wage?

d) What wages are in the 85th percentile or above?

e) Find the minimum wage for your province. At what percentile rank is the minimum wage in the data set above?

Check Your Understanding

Try It

1. Determine the 50th percentile for each data set. **Hint:** The 50th percentile is the median.

   a) 3, 15, 0, 6, 12, 15, 3, 9, 10, 15

   b) 12, 5, 22, 30, 7, 36, 14, 42, 15, 52, 25

   c) 135, 159, 254, 199, 205, 175, 180

   d) 88, 73, 56, 61, 64, 68, 71

   e) 6, 20, 16, 14, 9, 15, 18, 7, 10

   f) 154, 160, 142, 98, 178, 194, 138

2. Determine the 25th percentile for each data set in #1.

3. Determine the 75th percentile for each data set in #1.

4. Use the data set in the table to answer the questions.

   a) What value is at the 50th percentile? the 25th percentile? the 75th percentile?

   b) What is the percentile rank of a value of 11? of 32? of 41?

Stem (tens)	Leaf (ones)
4	1 3 4 7
3	0 1 1 2 3 4 5 5 6 8 8 9
2	0 2 4 4 6 7 7 7 8 9
1	0 1 5 8 9
0	3 7

Apply It

5. Fiona enters her latest painting in an art contest. She scores 21 out of 50. She does not win a prize, but she is happy with her score. The scores of nine other people who entered the contest are 19, 14, 23, 16, 15, 13, 17, 22, 20.

   a) What is the range of scores?

   b) At what percentile is a score of 16?

   c) At what percentile is a score of 23?

   d) What score is at the 50th percentile?
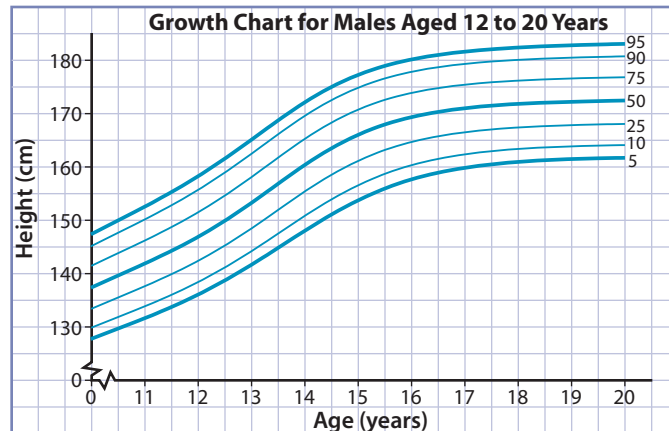
   e) Why do you think Fiona is happy with her score?

6. Refer to the growth chart below to answer parts a) to d).

Growth Chart for Males Aged 12 to 20 Years

a) At what percentile is a 12-year-old male whose height is 158 cm?

b) At what percentile is a 20-year-old male whose height is 185 cm?

c) David, who is 12 years old, is at the 75th percentile. What is his height in centimetres?

d) At what percentile is your height or the height of a male classmate?

7. A logging company in British Columbia is clearcutting a section of forest. The loggers can cut only if a random survey of the circumference of 80 trees shows that a measurement of 115 cm is at the 75th percentile or higher. Examine the frequency tables from two forest sections. Determine which forest meets the criterion for clearing.

Forest A

Circumference (cm)	Frequency
130–139	1
120–129	4
110–119	14
100–109	18
90–99	16
80–89	17
70–79	10

Forest B

Circumference (cm)	Frequency
130–139	5
120–129	11
110–119	23
100–109	20
90–99	10
80–89	6
70–79	5

Work With It

1. A medical study is conducted to learn the effect of energy drinks on heart rate. A group of teenagers drink a 250-mL energy drink and record their number of heartbeats over an interval of 15 s. The following data are collected: 24, 33, 41, 46, 19, 37, 29, 31, 30, 35, 35, 40, 41, 28, 45

a) Create a stem-and-leaf plot of the data.

b) Determine the three measures of central tendency.

c) Identify any possible outlier(s).

d) How would removing the outlier(s) affect the median and average? Explain.

2. Since 2010, most Canadian provinces have passed laws prohibiting cell phone use while driving. Researchers monitor how long it takes drivers to react to a flashing light when driving normally, while attempting to text, or while talking on a cell phone. Here are some of the reaction times, in seconds: 4.6, 2, 1.5, 3.5, 3.8, 0.9, 4.8, 0.5, 4.5, 5.1, 4.7, 1.3, 2.1, 5.2, 4.2, 3.5, 3.4

a) The researchers find that reaction times for normal driving are in the 25th percentile or lower. What reaction times are in this range?

b) Reaction times in the 75th percentile or higher are recorded for drivers who are attempting to text. What reaction times are in this range?

c) Reaction times of people talking on a cell phone are between the 25th and the 75th percentiles. According to this research, what are drivers probably doing if they have a reaction time of 4 s? of 18 s? of 1 s? of 5.5 s? of 12.8 s?

3. Ron does a fitness test every month in gym class. His scores are 3.5, 4.0, 3.8, 5.0, 4.5, 0.5, 5.5, 4.5, 3.9, 4.8.

a) Should you remove the outlier of 0.5 from the data when calculating the range, mean, and median? Why or why not?

b) What may have caused the outlier to occur?

c) Find the mean and the trimmed mean. Remove the highest and lowest values to calculate the trimmed mean. As Ron's gym teacher, which mean would you use as Ron's average fitness mark? Explain.

4. Create a list of 20 test scores as percents. Use the list to do the following.

a) Explain the difference between percent and percentile rank.

b) Show how percent can be less than percentile rank.

c) Show how percent can be greater than percentile rank.

5. Some courses in post-secondary schools are graded on a curve. This means that no matter what percent you earn for the course,
- marks in the 25th percentile or lower receive a grade of D
- marks between the 26th and 50th percentile receive a grade of C
- marks between the 51st and 75th percentile receive a grade of B
- all marks above the 75th percentile receive a mark of A

Do you agree with this method of grading? Why or why not?

6. Research how percentiles can be used to classify hourly wages and salaries.

7. a) Use the Internet to calculate your BMI (body mass index).

b) Explain how your BMI relates to percentiles.

c) How would knowing your BMI be beneficial to you?

Web Link

To find out more about percentiles and wages, go to www.mcgrawhill.ca/books/mathatwork12 and follow the links.

Web Link

To find more information about BMI, go to www.mcgrawhill.ca/books/mathatwork12 and follow the links.
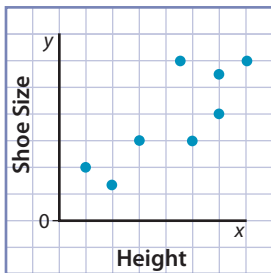
2.3 Scatter Plots

Focus On . . .

- interpreting scatter plots
- determining whether a trend exists

scatter plot

- a graph of plotted points that shows the relationship between two data sets
- Example: Each dot represents one person's height versus shoe size.

trend

- the general direction in which values in a data set tend to move

Scatter plots are a way of graphing quantitative data. Plotting the data on a scatter plot shows how one variable is affected by another. You can look at the overall pattern of the points on the grid to see if they show a trend in the data. For example, a scatter plot can show if there is a relationship between the length of a person's forearm and his or her hand span.

Explore Scatter Plots

1. Working with a partner, measure and record the length of your forearm and your hand span to the nearest millimetre.

2. Collect the same data from all the members of your class and from your teacher.
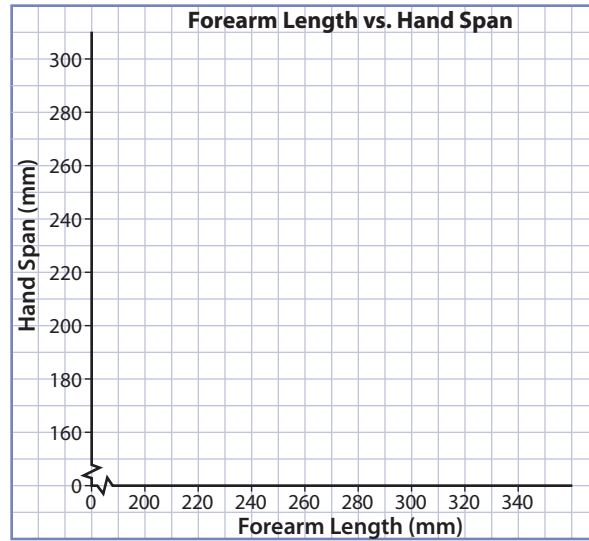
Forearm Length (mm)	Hand Span (mm)

> Your forearm is from your elbow to your wrist. Your hand span is the distance from the tip of your thumb to the tip of your little finger when your fingers are spread out.

3. With or without technology, graph the pairs of data for each person. Use a grid labelled as shown.

Forearm Length vs. Hand Span

A grid with the vertical axis labelled "Hand Span (mm)" showing values 0, 160, 180, 200, 220, 240, 260, 280, 300 and the horizontal axis labelled "Forearm Length (mm)" showing values 0, 200, 220, 240, 260, 280, 300, 320, 340.

4. a) Describe a trend represented by the graphed data.

   b) Do the points graphed form a straight line?

5. Reflect

   a) From the graph, can you predict the hand span of a person whose forearm length is beyond the longest forearm length recorded? Why or why not? This is called extrapolating.

   b) Can you predict the hand span of a person whose forearm length is between two given forearm lengths? Why or why not? This is called interpolating.

> Extrapolation can be less accurate than interpolation. When you interpolate, you estimate the value of a point between two known points. When you extrapolate, you are going beyond the data. You need to use the trend of the data and common sense to make an accurate prediction.

6. Extend Your Understanding Does plotting both genders on the same graph affect the trend? Why or why not?
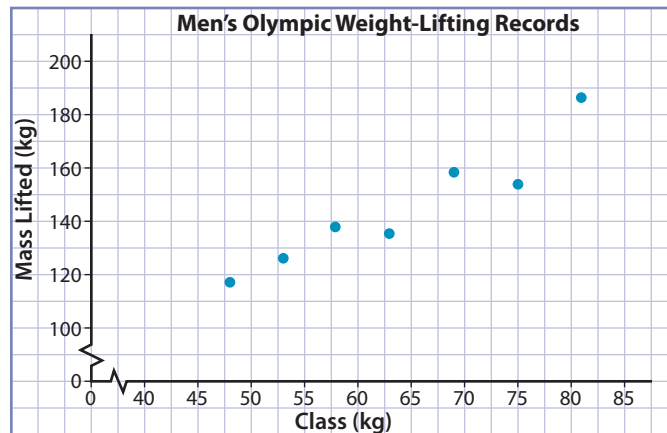
On the Job 1

Reading Scatter Plots

The scatter plot shows Olympic men's weight-lifting records. The **independent variable** is the class (mass of the weight-lifter, in kilograms). The **dependent variable** is the mass lifted, in kilograms. Use the scatter plot to answer the questions.

independent variable

- the variable being changed
- graphed on the *x*-axis

dependent variable

- the result when the independent variable is changed
- graphed on the *y*-axis

Men's Olympic Weight-Lifting Records

(Scatter plot: y-axis "Mass Lifted (kg)" from 0 to 200; x-axis "Class (kg)" from 40 to 85)

a) What is the approximate record lift for the 58-kg class?

b) How many more kilograms was the record lift in the 81-kg class than in the 48-kg class?

c) Describe the trend represented by the scatter plot.

d) If there were a 65-kg class, how many kilograms might the record lift be?

e) Could this trend continue indefinitely?

Solution

a) Find the point at 58 on the *x*-axis. What is the reading on the *y*-axis? The mass in kilograms is approximately 139 kg.

b) Difference in lift = lift for 81-kg class − lift for 48-kg class
$$= 188 - 118$$
$$= 70$$

The 81-kg class record was 70 kg more than the 48-kg class record.

c) As the mass of the lifters increases, the amount lifted increases.

d) Interpolate the data. The lift for an athlete in the 65-kg class would be approximately 150 kg.

e) The trend could not continue indefinitely. There is a limit to the class sizes, that is, the mass of the lifter. There is also a limit to the mass that a human can lift.

Your Turn

The scatter plot shown represents the length, in inches, versus the mass, in pounds, of ten salmon. Use the scatter plot to answer the questions.
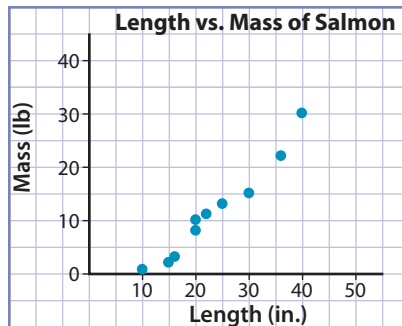

a) What is the approximate weight of a salmon that is 25 inches long?

b) What is the approximate length of a 20-pound salmon?

c) Approximately how many times heavier is a 40-inch salmon than one that is 20 inches?

d) Give a reason why a 20-inch salmon could have two different masses.

e) Could this trend continue indefinitely?
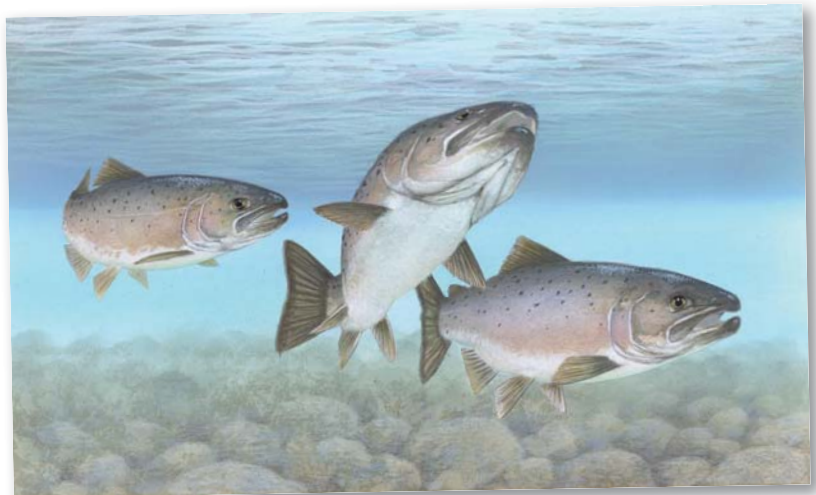
Try It

1. Match the trend to the scatter plot.

    i) One variable increases as the other one decreases.

    ii) As one variable increases, so does the other.

    iii) There is no trend.

    iv) One variable increases as the other one remains constant.

a)

b)

c)

d)
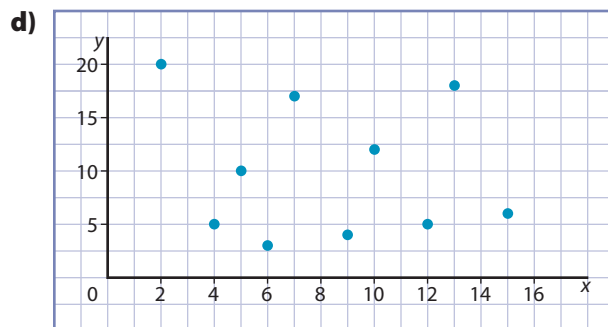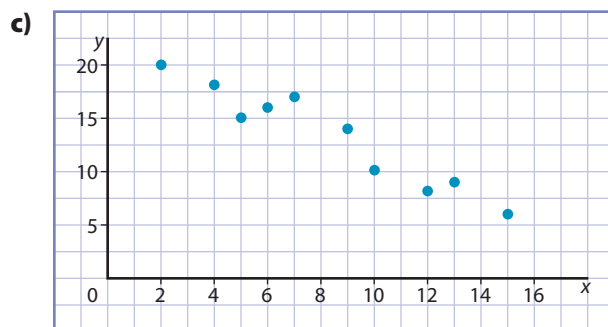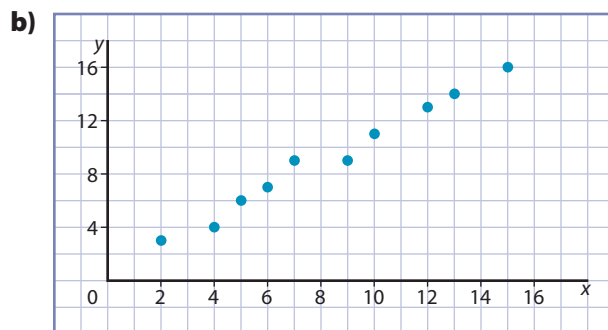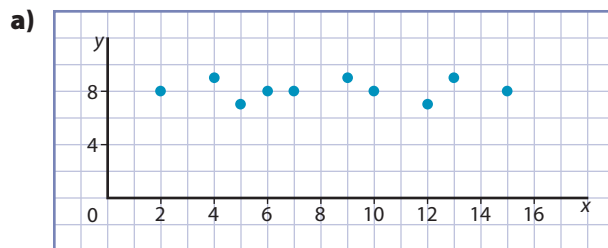
2. Describe a scatter plot that represents the age of a car (independent variable) and its value (dependent variable).

3. Describe a scatter plot that represents the length of fencing (independent variable) and its cost (dependent variable).

4. Describe a scatter plot comparing the height of a student and the number of minutes per day the student spends texting.

Apply It

5. The scatter plot compares the high temperature for a day in degrees Celsius (°C) to the number of people visiting a park. Use the graph to answer the questions.

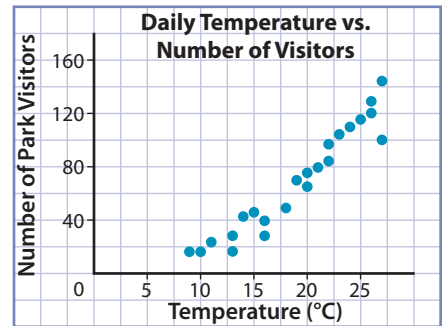**Daily Temperature vs. Number of Visitors**

   a) Describe the trend represented by the graph.

   b) If the temperature forecast is 25 °C, about how many visitors should park officials expect?

   c) What temperature range would bring between 100 and 150 visitors to the park?

   d) Make a prediction about the number of visitors on a day that the temperature high is 30 °C. Justify your prediction.

6. A taxi company recorded the outside temperature each time it received a call. The scatter plot shows the data collected.

Daily Temperature vs. Number of Requests for Taxi Service
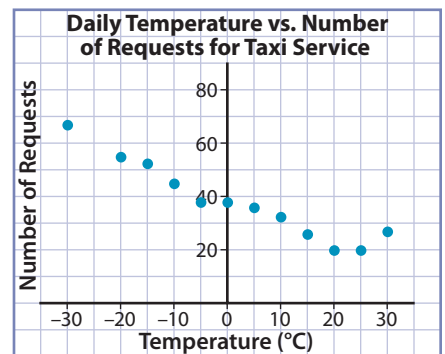
   a) What happens to the number of requests for taxi services as the temperature increases?

   b) If the temperature falls to −40 °C, what might you do as a manager for the taxi company?

   c) What could be the reason for the rise in the number of calls when the temperature is 30 °C?

   d) How could the manager for the taxi company use this graph to plan the number of drivers needed on call? Explain.

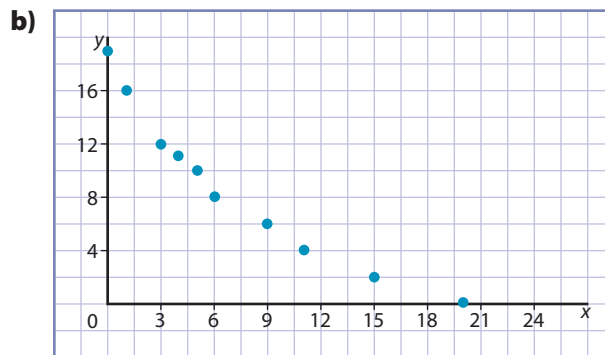1. Describe a scatter plot that compares the speed of a boat and the fuel consumption of the boat's engine.

2. Describe a scatter plot that compares the shoe sizes of grade 12 students and their math marks.

3. Make up a scenario for each scatter plot shown below.

a)

b)

c)
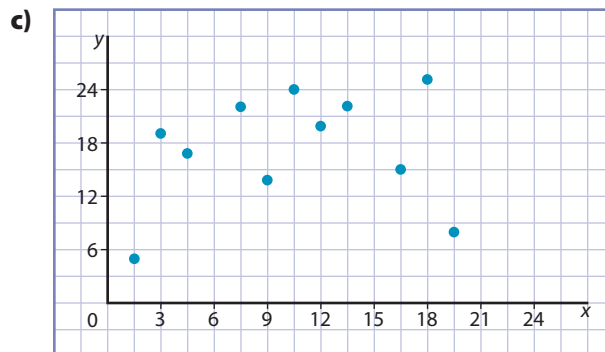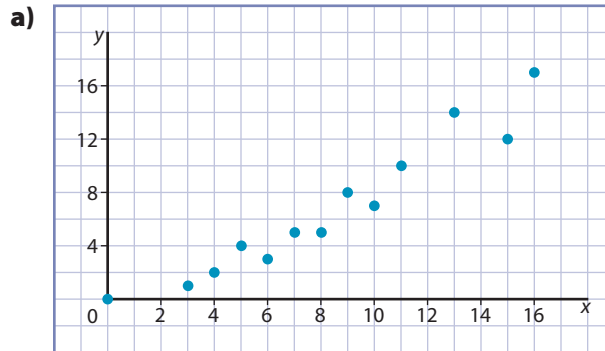
4. The scatter plot shows the percent of Canadians 15 years of age or younger in different years.

Percent of Canadians Aged 15 Years or Younger
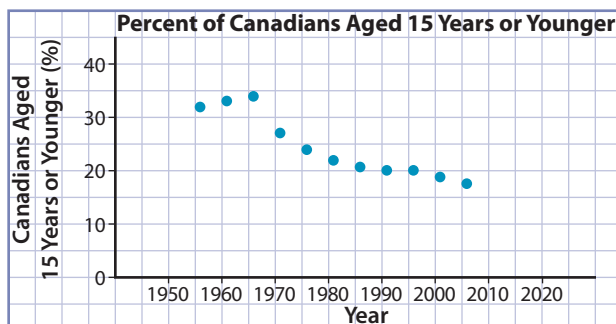
a) Joseph extrapolates that in 1950, 32% of Canadians were 15 years of age or younger. Is this reasonable? Why or why not?

b) He predicts that in 2020, the percent of Canadians 15 years of age or younger will be 12%. Is this reasonable? Why or why not?

c) Does a trend in a scatter plot always continue indefinitely? Explain. Use the scatter plot above to help you answer the question.

5. Ryan is training for a race. The scatter plot below shows Ryan's times for running a fixed distance over a ten-week period during his training.

Ryan's Time
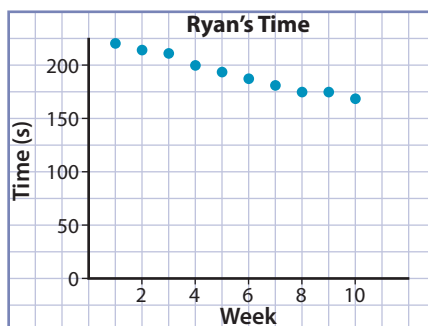
a) What trend is shown? Is the trend realistic? Explain.

b) Will the trend continue indefinitely? Explain.

6. List some examples in which the following would apply.

a) A scatterplot does not show a relationship or trend.

b) The trend does not continue indefinitely.

c) The data set cannot be represented on a scatter plot.

Skill Check

What You Need to Know

Section	After this section, I know how to . . .
2.1	■ **determine the three measures of central tendency** ■ **determine weighted means** ■ **solve a problem involving measures of central tendency**
2.2	■ **find the range for a data set** ■ **identify outliers in a data set** ■ **calculate the trimmed mean for a data set** ■ **work with percentiles and percentile ranks** ■ **solve problems involving trimmed mean and percentiles**
2.3	■ **interpret scatter plots** ■ **determine whether a trend exists**

If you are unsure about any of these questions, review the appropriate section or sections of this chapter.

2.1 **Measures of Central Tendency, pages 64–79**

1. Ten teenagers were asked how many hours per week they spend on social media activities. The answers given in hours were 47, 28, 36, 25, 35, 3, 10, 29, 12, 8.

   a) Determine the mode, median, and average of the data set.

   b) Which measure of central tendency would you use to describe the typical amount of time a teenager spends on social media activities per week? Explain your choice.

2. Jason's physical education mark is based on the following:

Mark Breakdown	Weighting (%)	Jason's Mark (%)
Participation	35	90
Skills	25	70
Fitness	25	75
Knowledge of game rules	15	80

Calculate the weighted mean to determine Jason's term mark.

3. A set of seven judges gave the following scores to Charlene's diving performance: 7.2, 6.8, n, 8.0, 8.5, 8.2, 6.8

If Charlene's overall average was 7.5, what score did the third judge give her?

2.2 **Using Other Statistical Measures, pages 80–93**

4. Olivia, a realtor, sold seven houses last year for the following prices:

$205 000, $170 000, $1.5 million, $95 000, $126 000, $100 000, $82 000

a) What was the median price of the houses she sold?

b) What was the mean price of the houses she sold?

c) Identify any possible outlier(s).

d) Remove the highest and lowest values to calculate the trimmed mean.

e) Which measure of central tendency best describes the typical price of the houses Olivia sold? Explain.

5. Robert scored 32 out of 50 on a history test. He was disappointed with his mark until he learned that the teacher scored on the percentile scale.

a) If the data set for the class was 16, 40, 20, 25, 32, 42, 19, 21, 15, 10, what was Robert's percentile rank?

b) Which score was at the 50th percentile? at the 25th percentile?

2.3 **Scatter Plots, pages 94–101**

6. The scatter plot compares the ages, in years, and heights, in centimetres, of 25 males.

a) Estimate the age of a male whose height is 140 cm.

b) Estimate the height of a two-year-old boy.

c) Extrapolate the height of a 21-year-old male.

d) Describe the trend. Will this trend continue indefinitely? Explain.

Age vs. Height of Males

Height (cm) vs. Age (years)

For #1 to #5, select the best answer.

1. Which measure of central tendency best describes the most frequent number of successful serves by the members of a volleyball team?

   A mode
   B median
   C mean or average
   D range

2. For which of the following would the median be the best measure of central tendency?

   A the number of pages in a typical math textbook
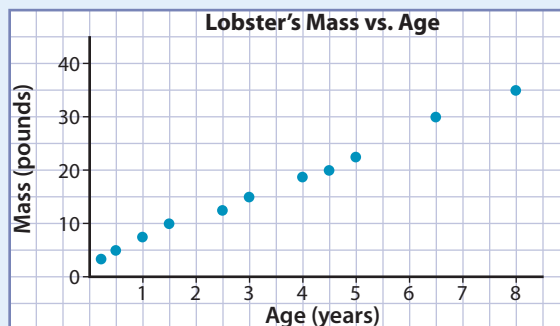   **B** the most requested song at a dance
   C the middle-ranked person on a basketball team, in terms of height
   D the typical amount of time an 18-year-old sleeps

3. Which of the following is *not* true about an outlier?

   A It is a value that is very different from the other values in a data set.
   B It should always be removed from a data set before determining measures of central tendency.
   C When computing the trimmed mean, outliers are removed from the data set.
   D A data set may have zero, one, or more than one outliers.

4. In this data set, which value represents the 75th percentile?

   112, 105, 122, 130, 107, 136, 114, 142, 115, 153, 125

   A 112
   B 115
   C 122
   D 136

5. Which of the following collections of data could be represented on a scatter plot?

   A the distance commuters travel to work
   B the time commuters take to travel to work
   C the distance and the time commuters take to travel to work
   D the age of commuters who travel to work

6. The graph compares a lobster's
mass, in pounds, to its estimated age, in years.

a) Describe the trend shown by the data
points.

b) Extrapolate the weight of a nine-year-old
lobster.

c) Interpolate the age of a 25-pound lobster.

d) The record mass for a lobster is 44 pounds.
Is it possible to estimate its age from the scatter plot? If yes,
estimate its age. If no, explain why it is not possible.

Lobster's Mass vs. Age

7. Min Lee chooses ten paperback books to give as gifts. The prices
are $7.00, $16.75, $1.50, $4.25, $45.00, $12.00, $6.50, $22.25,
$19.00, $18.25.

a) Find the three measures of central tendency for the price of the
ten books.

b) What is the range of the data set?

c) Identify any outlier(s).

d) Calculate the trimmed mean by removing the highest and
lowest values.

e) Which measure of central tendency best describes the typical
price of a book Min Lee chose? Explain.

8. Ross received the following marks in his geography class:

Course Work	Weighting (%)	Mark (%)
Quiz	10	53
Test 1	20	76
Test 2	20	60
Project	20	77
Exam	30	80

He calculated his mean mark to be 69%. Ross was pleased that his
teacher calculated the mark using a weighted mean. Explain why
Ross was pleased. Show your calculations.

Chapter Project

Manage a Sporting Goods Store

1. Choose a product related to sports by visiting a sporting goods store or researching a product on the Internet. Examples are athletic shoes, tents, fishing line, personal flotation devices (PFDs), golf clubs, and camouflage clothes.

2. Complete a table similar to the one shown.
- Categorize the product by brand, style, size, weight, quality, and so on.
- Find the price for each item.
- Take an inventory of the product that is in stock at the store.

Product: _____				
Brand	**Style**	**Size**	**Price**	**Number in Stock**
Northern	King of Hill	Small		
		Medium		
		Large		
	Everlasting	Small		
		Medium		
		Large		
	Wonderful	Small		
		Medium		
		Large		

3. Use range and measures of central tendency to help you determine how to answer the following questions. For each, explain how you used the statistical measure.

a) How many items will you order for each style category and each size category?

b) Which items will you put on sale to clear stock?

c) Which items are seasonal, and need to be moved to storage or discontinued?

GAMES AND PUZZLES

Penny Ball

Play Penny Ball in groups of three. Keep track of each player's statistics. Then, create a batting order based on each player's strengths.

Materials
- 1 m by 1 m piece of cardboard
- masking tape
- 3 pennies per student

1. To create a game board, mark areas to show the following plays:
   - single hit
   - double hit
   - triple hit
   - homerun or RBI if a batter is on third base
   - walk
   - fly out or sacrifice fly if a batter is on third base

   Include several of each play on the board. Different areas can range in size. The rest of the board is a strike. The strike area should be about $\frac{1}{3}$ of the total area.

2. Place the game board on the floor and put a piece of tape about 1 metre from the board. The first batter stands behind the tape and throws a penny onto the board. If the penny lands on an area marked with a play, the batter's turn is over. If the penny lands on a strike, the batter throws another penny. Each batter gets up to three strikes before the next batter's turn. Create a table to keep track of where each batter's pennies land. Each batter has at least three turns at bat.
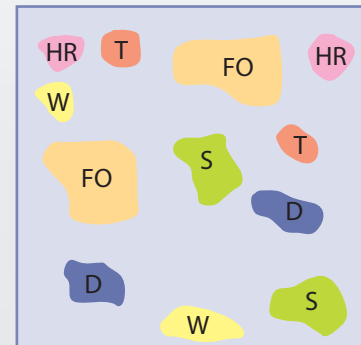

Legend
S = single
D = double
T = triple
HR = homerun or RBI
FO = flyout or sacrifice fly
W = walk

3. Use each batter's statistics to calculate the following:
   - On Base Percentage $= \dfrac{\text{hits} + \text{walks}}{\text{bats} + \text{walks} + \text{sacrifice flies}}$
   - Batting Average $= \dfrac{\text{hits}}{\text{at bats}}$
   - Slugging Percentage $= \dfrac{(\text{singles}) + (2 \times \text{doubles}) + (3 \times \text{triples}) + (4 \times \text{homeruns})}{\text{at bats}}$

4. Use each batter's statistics to decide where you would place each person in the batting order. Some general rules for choosing a batting order are:
   - Batters 1 and 2 usually have the highest on base percentage.
   - Batter 3 usually has a high batting average.
   - Batter 4 usually has a high slugging percentage.
   - Batter 5 usually has a lot of RBIs.