TABLE OF CONTENTS

REPRODUCIBLE SHEETS

VISION 2 Statistical measures and linear correlation

Name: _____

Group: _____ Date: _____

LES 3

Student Book, p. 202

LES 3: Carbon exchange

Listed below is data regarding CO_2 emissions of companies in four different countries registered with the carbon exchange:

Country 1

	Company															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
CO_2 quota (tons)	8	6	9	5	4	10	8	9	10	12	16	13	7	6	10	8
CO_2 emission (tons)	11	5	10	6	4	11	5	5	11	14	18	7	12	6	7	9

Country 2

	Company														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
CO_2 quota (tons)	4	5	2	4	5	7	2	5	7	4	1	3	4	6	2
CO_2 emission (tons)	4	5	4	5	4	9	3	5	4	5	3	3	5	7	2

Country 3

	Company											
	A	B	C	D	E	F	G	H	I	J	K	L
CO_2 quota (tons)	12	13	11	6	15	8	13	14	17	18	7	5
CO_2 emission (tons)	9	12	11	5	13	7	8	10	14	19	9	5

Country 4

	Company													
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
CO_2 quota (tons)	20	19	12	7	2	22	5	17	11	7	9	10	3	6
CO_2 emission (tons)	21	22	12	9	5	21	7	15	14	11	8	15	4	5

Name: _____

Group: _____ Date: _____

LES 3

Student Book, p. 202
(*cont'd*)

After analyzing CO_2 quotas imposed on companies in each of the countries, and the quantities of CO_2 they emit, produce a report on each country's atmospheric pollution record.

In your report, you must:

- ☐ Represent, using the appropriate diagram, the quantities of CO_2 produced by companies in the same country. The emissions of at least two countries must be represented using a stem-and-leaf plot.
- ☐ Determine which country best respects the quotas imposed on it. You will need to clearly explain your choices using a statistical approach.
- ☐ Determine the country in which company D best meets its greenhouse gas quotas by comparing the percentile of its carbon emissions performance with the other companies in the same country.

Name: _____

Group: _____ Date: _____

LES 4

Student Book, p. 203

LES 4: Water quality

The contingency table below presents the results of a study on the water quality of various rivers in Québec.

Water quality

BOD_5 (mg/L) \ Number of coliform colonies per litre of water	18	19	20	21	22	23	24	25	26	27	28
12		1	2								
13			1		3						
14				1							
15					2						
16						1	3		2	1	
17					1	2					2
18								1		1	

In order to study the effects of human pollution on certain watercourses in Québec, you need to produce a report.

In your report, you must:

- ☐ Represent this two-variable distribution using the appropriate graph.
- ☐ Qualitatively or quantitatively determine the relation between BOD_5 and the number of coliform colonies.
- ☐ Establish, using two different methods, the equation of a regression line representing this situation.
- ☐ Write a short text that explains the relation between the two statistical variables presented in the table above, and provide five recommendations for better water management.

Name: _____

Group: _____ Date: _____

KNOWLEDGE IN ACTION 2

Student Book, p. 73

Number 3

3 Complete this table.

Distribution	Mode	Median	Mean	Range
3, 6, 7, 7, 8, 8, 8, 12, 14, 15, 17, 18, 21, 23, 28, 30, 30				
5, 5, 5, 5, 5, 5, 6, 6, 6, 7, 8, 9, 9, 9, 9, 9, 9, 10				
12, 14, 15, 23, 24, 25, 33, 34, 35, 44, 44, 44, 44, 44				
6, 7, 12, 14, 16, 18, 20, 22, 25, 27, 29, 34, 37				

Name: _____

Group: _____ Date: _____

ACTIVITY 2 2.1

Student Book, p. 78

The red oak

In southern Québec, a forestry technician has measured the diameters (in cm) of the red oak trees within a forested area of 100 m^2. Here is the data collected:

40	45	53	55	55	55	55	60
60	63	64	64	65	67	67	68
70	70	70	71	72	75	75	79
82	83	83	91	97	102	113	

In order to analyze the data, the technician makes a stem-and-leaf plot.

Red oak: trunk diameter (cm)

4	0	5							
5	3	5	5	5	5				
6	0	0	3	4	4	5	7	7	8
7	0	0	0	1	2	5	5	9	
8	2	3	3						
9	1	7							
10	2								
11	3								

a. In this distribution:

1) What do the numbers in the left column represent? _____

_____

2) How many data values are on the 5th line? _____

3) On which line is the data value of 71 cm? _____

4) Where is the data value of 97 cm? _____

5) What do the data on the same line have in common? _____

_____

b. In the distribution, what diameter corresponds to:

1) the mode? _____ 2) the median? _____

c. What are some of the advantages of a stem-and-leaf plot over a list?

_____

_____

_____

d. In this distribution, what is the percentage of red oaks with a trunk diameter of:

1) less than 65 cm? _____ 2) equal to 70 cm? _____

3) less than or equal to 75 cm? _____ 4) greater than 83 cm? _____

e. Add the following values to the stem-and-leaf plot: 59 cm, 81 cm, 96 cm, 96 cm, 109 cm.

Name: _____

Group: _____ Date: _____

SUPPORT 2.1

Diagrams and statistics

1 For each of the following distributions, calculate:

1) The mean

2) the mean deviation

a) 2, 4, 6, 8, 10

1) _____ 2) _____

b) 3, 5, 7, 9, 11, 13

1) _____ 2) _____

c) 20, 40, 60, 80, 100

1) _____ 2) _____

d) 2, 6, 9, 14, 22, 3, 9

1) _____ 2) _____

e) 1, 5, 9, 8, 4, 13, 12

1) _____ 2) _____

f) 14, 12, 11, 22, 13, 15, 16, 19

1) _____ 2) _____

2 List, in ascending order, the values for each of the following distributions represented by a stem-and-leaf plot.

a) **Results of an English exam (%)**

6	4 8 9
7	2 6
8	4 8 9
9	2 5

_____

b) **Height of fir sapling growth in a nursery (mm)**

6	8
7	4 6 9
8	3 5 7 9
9	2 4

_____

3 Here is the data for a statistical distribution:

132, 132, 133, 134, 135, ..., 167, 171, 173, 175, 176, 178, 178, 178, 178, 180,
28 values

181, 182, 184, 184, 184, 185, 187, 189, 191, ..., 220, 223, 225, 227, 228, 229, 230, 241, 243, 244
43 values

What is the percentile for:

a) 173? _____

b) 178? _____

c) 184? _____

d) 244? _____

Name: _____

Group: _____ Date: _____

SUPPORT 2.1

(*cont'd*)

4 Represent each of the following distributions using a stem-and-leaf plot.

a) **Number of baskets of strawberries picked**

6, 10, 13, 15, 20, 4, 8, 10, 12, 15, 16, 16, 8, 23, 5, 7, 6, 13, 16, 11, 17, 18, 3, 19, 4, 7, 2

b) **Size of snakes in a zoo (cm)**

32, 31, 43, 52, 12, 42, 36, 78, 86, 89, 62, 55, 67, 87, 42, 57, 68, 27, 56, 61

c) **Number of pages in a paperback novel**

181, 196, 221, 172, 231, 187, 204, 217, 198, 202, 236, 180, 200, 217, 215, 204, 222, 230

d) **Duration of musical pieces (in s)**

242, 254, 271, 264, 236, 228, 249, 274, 269, 253, 272, 285, 250, 241, 263, 268, 271, 261

5 Here is the data for a statistical distribution:

31, 31, 34, 35, ..., 57, 61, 63, 65, 66, ..., 78, 78, 78, 78, 80, 80, 81, 84, 84, 84, 85,
(23 values) (40 values)

87, 89, 91, ..., 110, 113, 115, 117, 118, 119, 130, 141, 143, 144, 145
(23 values)

Which values have a percentile of:

a) 70? _____ b) 92? _____

c) 25? _____ d) 2? _____

Name: _____

Group: _____ Date: _____

CONSOLIDATION 2.1

Diagrams and statistics

1 This stem-and-leaf plot presents data collected during a statistical study.

Age of participants at a fundraiser

1	5 6 6 7 8 8 8 9 9 9
2	0 0 0 1 2 2 3 3 4 4 4 4 4 5 6 6 6 6 7 8 9
3	0 0 0 1 1 2 3 3 4 4 4 4 6 7 8 8 8 9 9
4	0 2 2 2 3 3 3 3 4 4 4 4 5 6 6 7 8 9
5	0 0 1 1 2 2 2 3 4 4 5 5 6 6 7 8 9
6	0 1 1 2 2 3 4 4 4 5 5 6 6 6 7 8 8 9 9 9 9 9 9 9
7	0 0 2 2 2 2 3 3 3 4 4 5 6 6 7 7 7 8 8 9

a) What is the percentile of the mode of this distribution? _____

b) What is the percentile of a 26-year-old participant? _____

c) What is the age of a participant in the 62nd percentile? _____

2 Calculate the mean deviation for each of these distributions.

a) 19, 21, 2, 4, 21, 17, 25, 9, 20, 12, 11, 21, 24, 20 _____

b) 45, 29, 25, 45, 46, 32, 38, 44, 33, 44, 39, 48, 32, 40, 24 _____

3 The 238 values of a statistical distribution are put on a list in ascending order.

a) On this list, what is the position of the value for which the percentile is:

1) 82? _____

2) 33? _____

b) The 153rd value on the list is identical to two other values. What is the percentile of this value? _____

Name: _____

Group: _____ Date: _____

CONSOLIDATION 2.1

(cont'd)

4 Here is the number of films rented during a year by 16 female and 18 male members of a video rental store.

Number of videos rented by the 16 women: 43, 41, 11, 54, 11, 55, 8, 19, 24, 47, 12, 12, 40, 12, 59, 14

Number of videos rented by the 18 men: 5, 18, 48, 15, 64, 50, 29, 61, 51, 17, 6, 30, 40, 22, 45, 20, 26, 55

a) Construct a stem-and-leaf plot representing these two distributions.

b) Regarding the number of films rented annually, do the men or the women represent the greatest mean deviation? Explain your answer.

_____

_____

_____

5 The following are the ages of 25 contest participants:

22, 25, 17, 34, 20, 26, 16, 28, 34, 25, 18, 30, 16, 33, 21, 26, 34, 33, 31, 17, 17, 31, 35, 17, 27

a) What is the mean deviation for the age of the participants? _____

b) Construct a stem-and-leaf plot representing this distribution.

Name: _____

Group: _____ Date: _____

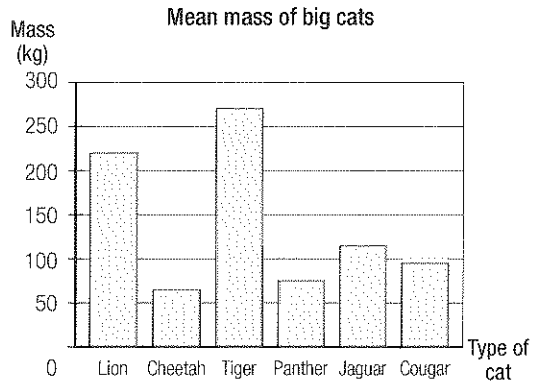
CONSOLIDATION 2.1

(cont'd)

6 For each graph, calculate:

1) the mean 2) the mean deviation

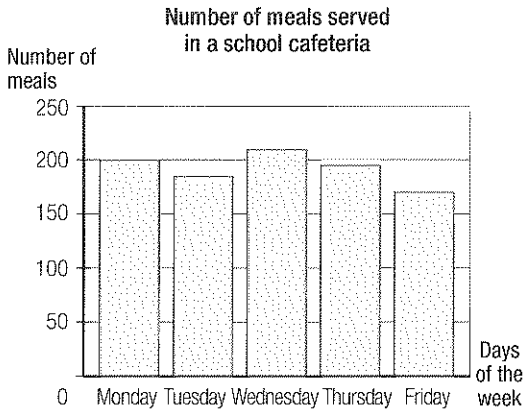
a)



1) _____

2) _____

b)



1) _____

2) _____

7 The table below provides information about the average annual precipitation in a selection of Canadian cities.

Precipitation

	Snowfall (cm)	Total Precipitation (mm)	Number of days with precipitation
Fredericton	276.5	1143.3	156.6
Québec	315.9	1230.3	181.9
Montréal	217.5	978.9	163.3
Ottawa	235.7	943.5	162.6
Toronto	115.4	792.7	145.5

a) What is the mean deviation of the total precipitation? _____

b) In Vancouver, the mean number of days with precipitation is 166.1. Is the mean deviation of the number of days with precipitation in this city greater than or less than the mean deviation of the number of days with precipitation in the five cities shown in the table?

_____

Name: _____

Group: _____ Date: _____

ENRICHMENT 2.1

Diagrams and statistics

1 a) A distribution is comprised of 203 different data. What is the percentile of the quartiles of this distribution?

b) Are the answers provided in a) applicable for all distributions? Explain your answer.

2 The following are 18 of 20 distribution values:

17, 16, 10, 11, 20, 10, 19, 15, 15, 15, 11, 15, 20, 16, 17, 14, 11, 13

One of the missing values occupies the 85th percentile and is unique in the distribution. The average of all 20 of the values is equal to the distribution's mode and median. The distribution's mean deviation is 2.5. Find the two missing values.

3 List the values for a distribution comprised of 11 distinct natural numbers for which:

- the mean is equal to the median
- the mean deviation is 4.90
- the interquartile range is 11
- the maximum is 18

Name: _____

Group: _____ Date: _____

ACTIVITY 1 2.2

Student Book, p. 90

An experimental drug

Epilepsy is a neurological disorder characterized by periodic epileptic seizures of varying degrees of intensity, ranging from loss of consciousness for a few seconds to full body convulsions. The attacks or episodes are brought on by uncontrollable electrical short-circuits in the brain.

A research team has been testing a new anti-seizure drug. They would like to establish the minimum dosage of the medication required to reduce the number of epileptic seizures. Here are some of their research findings:

Anti-seizure medication

Drop in attacks (%) \ Dose (mg)	[0, 100[	[100, 200[	[200, 300[	[300, 400[	[400, 500[	[500, 600[	Total
[0, 14[	7	2	0	0	0	0	
[14, 28[	22	8	0	0	0	0	
[28, 42[	3	12	3	0	0	0	
[42, 56[	0	18	7	5	0	0	
[56, 70[	0	4	9	8	0	1	
[70, 84[	0	0	0	12	8	4	
[84, 98[	0	0	0	0	0	7	
Total							

a. How many epileptics:

1) participated in the study? _____

2) tested doses of [0, 300[mg? _____

3) experienced a decrease of 70% of their attacks? _____

4) tried doses of [400, 500[mg and experienced a decrease in the number of attacks by [70, 84[%? _____

5) tested the medication and showed no improvement? _____

b. What percentage of epileptics whose seizures decreased by at least 56% were taking a dosage of 300 mg or more? _____

c. Can the research team conclude:

1) The higher the dose of medication, the more the physical condition of the patient improves. Explain your answer.

_____

2) There is a correlation between the amount of medication absorbed by a patient and the decrease in the frequency of the seizures. Explain your answer.

_____

Name: _____

Group: _____ Date: _____

PRACTICE 2.2

Student Book, p. 96

Number 1

1 Complete the contingency table below, showing the legal minimum age for marriage in different countries.

Legal age of marriage

Man's age / Woman's age	18	19	20	21	Total
16		0	2	3	17
17	7	0	5	6	
18	15	0		15	39
19		0	8	13	21
Total	34	0			95

Name: _____

Group: _____ Date: _____

PRACTICE 2.2

Student Book, p. 98

Number 7

7 **HOCKEY** The adjacent table displays information on players for the Montréal Canadiens at the start of the 2007-2008 season.

a) Complete the contingency table below.

Start of the 2007-2008 season

Games played \ Point scored	[0, 10[	[10, 20[	[20, 30[	[30, 40[	[40, 50[	Total
[0, 10[						
[10, 20[						
[20, 30[						
[30, 40[						
[40, 50[						
Total						

b) Among the players who played at least 20 games, how many scored less than 30 points?

_____

c) How many players played 30 games or more and scored 40 points or more?

_____

d) How many players played fewer than 20 games or scored less than 10 points?

_____

e) What percentage of players who scored fewer than 30 points, played 40 games or more?

_____

f) Does a correlation exist between the number of games played and the points scored? Explain your answer.

_____

_____

Start of the 2007-2008 season

Player	Games played	Points scored
Alex Kovalev	44	41
Tomas Plekanec	44	38
Saku Koivu	43	33
Christopher Higgins	44	31
Andrei Markov	44	31
Mark Streit	44	28
Andrei Kostitsyn	40	25
Roman Hamrlik	44	18
Guillaume Latendresse	40	16
Michael Ryder	39	15
Mike Komisarek	44	12
Bryan Smolinski	30	12
Mathieu Dandenault	43	11
Kyle Chipchura	36	11
Sergei Kostitsyn	14	8
Patrice Brisebois	29	8
Tom Kostopoulos	39	7
Maxim Lapierre	17	6
Steve Bégin	25	4
Francis Bouillon	39	3
Josh Gorges	24	3
Mikhail Grabovski	12	2
Ryan O'Byrne	11	2
Corey Locke	1	0

Name: _____

Group: _____ Date: _____

PRACTICE 2.2

Student Book, p. 100

Number 12

12 a) Complete the following table.

Students in your class

Student	Distance between the elbow and the middle fingure (cm)	Length of student's foot (cm)	Student	Distance between the elbow and the middle fingure (cm)	Length of student's foot (cm)
1			19		
2			20		
3			21		
4			22		
5			23		
6			24		
7			25		
8			26		
9			27		
10			28		
11			29		
12			30		
13			31		
14			32		
15			33		
16			34		
17			35		
18					

b) Is there a linear correlation between the two measurements above? Explain your answer.

_____

c) For each student, calculate the ratio $\frac{\text{distance from elbow to the tip of middle finger}}{\text{length of foot}}$. What do you notice?

Name: _____

Group: _____ Date: _____

SUPPORT 2.2

Qualitative interpretations of the correlation

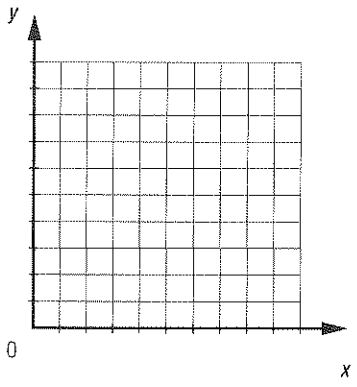
1 Here are the ordered pairs of a two-variable distribution:

(2, 10), (3, 11), (5, 12), (5, 13), (1, 12), (4, 13), (2, 11), (1, 10), (3, 13), (2, 12), (5, 15), (4, 14), (1, 12), (4, 14), (3, 12), (1, 12), (3, 12), (1, 11), (2, 12), (5, 15)

a) Complete the contingency table below.

x \ *y*	10	11	12	13	14	15	Total
1							
2							
3							
4							
5							
Total							

b) Construct a scatter plot representing this distribution.

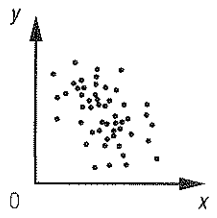


c) Describe the direction and strength of the linear correlation of the two variables of this distribution.

_____

2 Describe the direction and the strength of the linear correlation between variables of the scatter plots below.

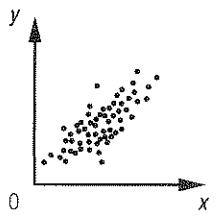
a)



_____

_____

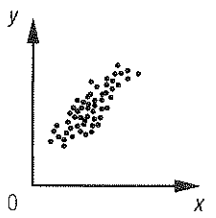
b)



_____

_____

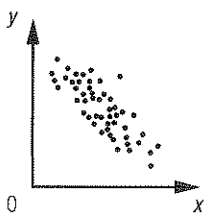
c)



_____

_____

d)



_____

_____

Name: _____

Group: _____ Date: _____

3 Complete the contigency tables below.

a) **Distribution of students at a CEGEP**

Gender / Sector	Male	Female	Total
Pre-university			1225
Technical		1029	
Total	1433		3250

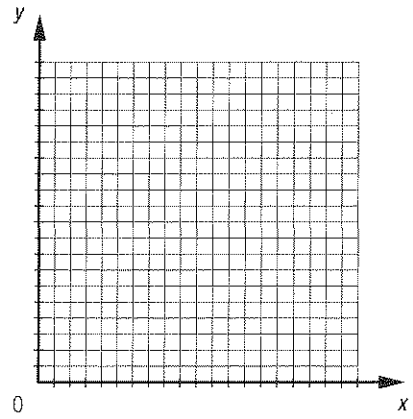
b) **Catch distribution by fisherman age**

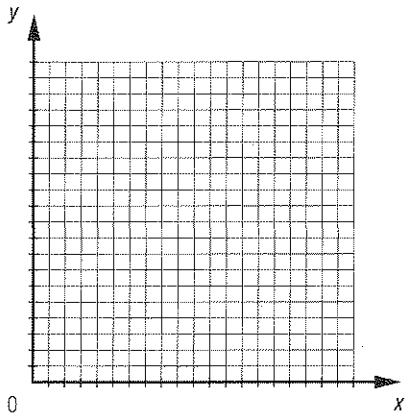
Age / Number of catches	15	16	17	18	Total
[0, 5[	3	12	3	1	
[5, 10[	17	18		16	71
[10, 15[	21		44	3	100
[15, 20[	13	16	13		
Total	54			25	

4 **a)** Represent each of the following statistical distributions by a scatter plot.

1) (2, 18), (4, 16), (3, 17), (6, 12), (4, 15), (2, 19), (7, 10), (5, 11), (4, 17), (3, 18), (2, 20), (4, 18), (6, 11), (4, 16), (8, 8)

2) (16, 12), (22, 14), (19, 13), (28, 18), (22, 15), (16, 11), (31, 20), (25, 19), (22, 13), (19, 12), (16, 10), (22, 12), (28, 19), (22, 14), (34, 22)





b) Describe the direction and the strength of the linear correlation between the variables of these two scatter plots.

1) _____

_____

2) _____

_____

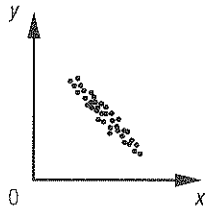
Name: _____

Group: _____ Date: _____

CONSOLIDATION 2.2

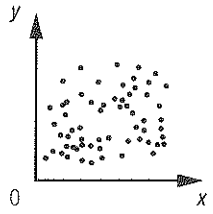
Qualitative interpretations of correlation

1 Describe the linear correlation between the variables in each of the scatter plots below.

a) 

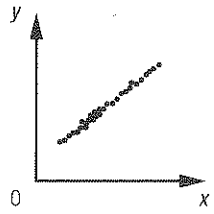
_____

_____

b) 

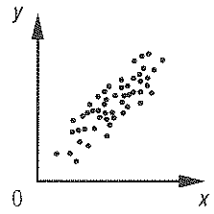
_____

_____

c) 

_____

_____

d) 

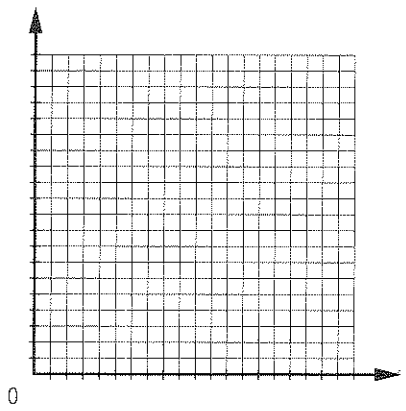
_____

_____

2 During an experiment, a ball is dropped from different heights and the height of the first bounce is recorded. Each of the following pairs indicates, respectively, the drop height of the ball (in cm) and the height (in cm) of the ball's first bounce.

(360, 254), (320, 228), (340, 255), (240, 180), (300, 225), (380, 285), (200, 150), (220, 148), (340, 240), (360, 270), (400, 300), (360, 260), (220, 165), (320, 220), (160, 120)

a) Construct a scatter plot representing this situation.



b) Describe the correlation between the drop height of a ball and the height of its first bounce.

_____

Name: _____

Group: _____ Date: _____

3 The adjacent table provides information about the Pittsburgh Penguin's line-up for the 2007-2008 season.

Pittsburgh Penguins

Player	Age	Number of games played
Adam Hall	27	46
Brooks Orpik	27	78
Darryl Sydor	36	74
Evgeni Malkin	21	82
Gary Roberts	42	38
Georges Laraque	31	71
Hal Gill	33	81
Jarkko Ruutu	32	71
Jeff Taffe	27	45
Jordan Staal	19	82
Kris Beech	27	25
Kris Letang	21	63
Marian Hossa	29	72
Mark Eaton	31	36
Maxime Talbot	24	63
Pascal Dupuis	29	78
Petr Sykora	31	81
Rob Scuderi	29	71
Ryan Malone	28	77
Ryan Whitney	25	76
Sergei Gonchar	34	78
Sidney Crosby	20	53
Tyler Kennedy	21	55

a) Complete the following contingency table.

Pittsburgh Penguins

Number of games played \ Players' ages	[15, 20[	[20, 25[	[25, 30[	[30, 35[	[35, 40[	[40, 45[	Total
[22, 34[							
[34, 46[							
[46, 58[							
[58, 70[							
[70, 82]							
Total							

b) Describe the linear correlation between a players' age and the number of games played.

_____

4 Describe the linear correlation between the variables of the each of the situations described below.

a) The shoe size of an individual and the height of this individual.

_____

b) The age of an adult and the time it takes to run 100 m.

_____

c) The amount of time an individual takes to complete an exam and the exam result.

_____

d) The amount of time a glass of water is exposed to the sun and the amount of water in the glass.

_____

e) The quantity of water added to a test tube containing salt and the salinity of the solution.

_____

5 A comparison is made of the sales price (in $) of various homes and the time (in months) required to sell them. Here are the results:

(152,000, 14), (161,000, 13), (199,000, 18), (201,000, 18), (209,000, 19), (300,000, 27), (147,000, 10), (125,000, 11), (130,000, 9), (175,000, 15), (186,000, 16), (246,000, 22), (270,000, 24), (152,000, 11), (163,000, 13)

a) Create a contingency table representing the values provided above.

b) Using the contingency table created in **a)**, describe the type and the direction of the linear correlation between these two variables.

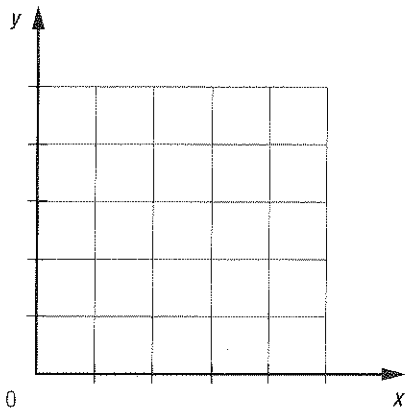
_____

6 The table below represents the results of a survey conducted on a group of adults.

Is it possible to say that a linear correlation exists between the number of brothers and sisters and the number of children they have? Explain your answer.

Survey

Number of brothers and sisters	0	0	0	1	1	1	1	1	2	2	2	2	3	3	4	4	4	5
Number of children	1	0	0	1	3	3	2	2	1	3	2	1	3	2	2	4	3	3



_____

_____

_____

_____

_____

Name: _____

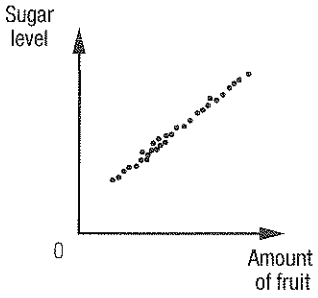
Group: _____ Date: _____

ENRICHMENT 2.2

Qualitative interpretation of correlation

1 The adjacent graph displays the results of a study on the amount of fruit consumed by an individual and the individual's blood sugar level.

Is it possible to draw a conclusion based on the overall appearance of the graph?



_____

_____

_____

2 Twenty-five percent of students in a group wear glasses. One-third of the group is female. One in six in the group is left-handed. Two of the group's 16 boys wear glasses. The probability of selecting a girl in the group who is left-handed is $\frac{1}{8}$.

a) Create a contingency table which allows a comparison of the genders of the students in the group and the students in the group who wear glasses.

b) Create a contingency table which compares these students' genders and the hand they use to write.

Name: _____

Group: _____ Date: _____

ACTIVITY 1 2.3

Student Book, p. 105

Gestation

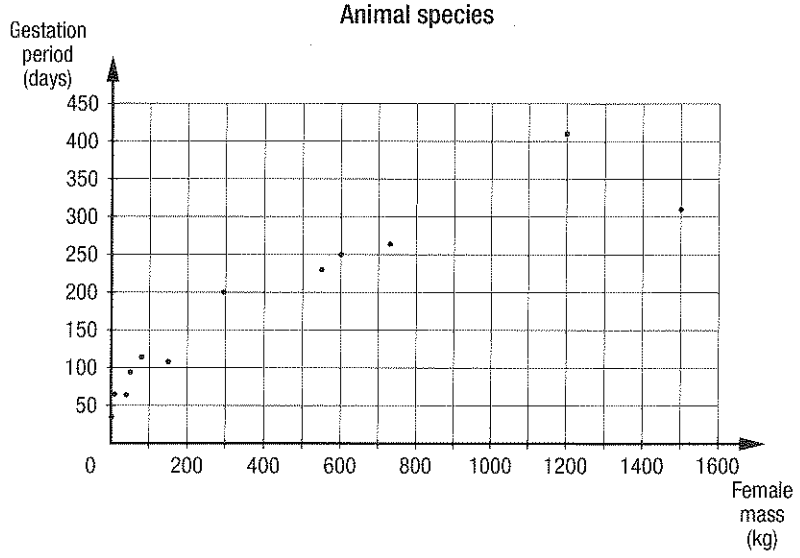
A gestation period corresponds to the number of days that a mother carries her baby. Animals' gestation periods vary from one species to another. The following table lists some data on this subject.

Animals

Animal species	Female mass (kg)	Gestation period (days)
Rabbit	1	35
Cat	9	65
Wolf	40	64
Leopard	50	94
Pig	80	114
Lion	150	108
Bear	295	200
Moose	550	230
Bison	600	250
Cow	730	264
Giraffe	1200	410
Hippopotamus	1500	310

The scatter plot below illustrates the relation between a female's mass and her gestation period.

a. Associate each of the points on the graph with the corresponding animal species.



b. Complete the table below.

Animal group	Median mass (kg)	Median gestation period (days)	Ordered pair composed of the median mass and the median gestation period
Rabbit, cat, wolf, leopard			M_1(_____, _____)
Pig, lion, bear, moose			M_2(_____, _____)
Bison, cow, giraffe, hippopotamus			M_3(_____, _____)

c. Find the coordinates of point P, so that:

1) the *x*-coordinate corresponds to the mean of *x*-coordinates for points M_1, M_2 and M_3 _____

2) the *y*-coordinate corresponds to the mean of *y*-coordinates for points M_1, M_2 and M_3 _____

d. Determine the slope of the line passing through points M_1 and M_3. _____

e. In the graph in **a**, do the following:

1) plot point P

2) draw a line passing through point P having the same slope as the line passing through points M_1 and M_3

f. Determine the equation of the line drawn in **e**. _____

g. Given the equation of this line, what would be:

1) the gestation period of a mare with a mass of 1100 kg? _____

2) the mass of a female tiger whose gestation period is 100 days? _____

The following table provides data on various monkey species.

Monkeys

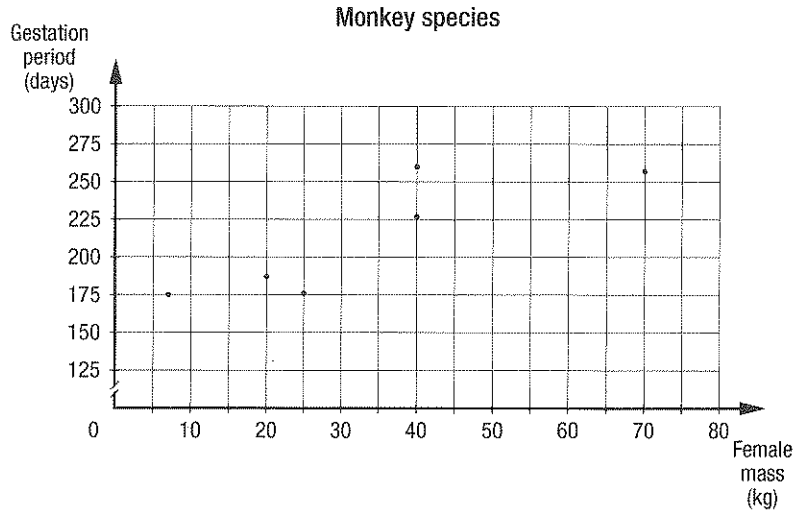
Species	Mass of female (kg)	Gestation period (days)
Chimpanzee	40	227
Gorilla	70	257
Orangutan	40	260
Baboon	20	187
Macaque	7	175
Mandrill	25	176

Name: _____

Group: _____ Date: _____

This scatter plot illustrates the relationship between a female's mass and the gestation period for several monkey species.

h. Associate each of the points on the graph with the corresponding species.



i. Complete the following table.

Monkey group	Mean mass (kg)	Mean gestation period (days)	Ordered pair composed of the mean mass and the mean gestation period
Baboon, macaque, mandrill			P_1(_____, _____)
Chimpanzee, gorilla, orangutan			P_2(_____, _____)

j. In the graph in **h**, do the following:

1) plot points P_1 and P_2
2) draw a line passing through points P_1 and P_2

k. Determine the equation of the line passing through points P_1 and P_2.

_____

l. Given the equation of this line, what would be:

1) the gestation period of a red howler monkey that has a mass of 4 kg?

_____

2) the mass of a bonobo that has a gestation period of 230 days?

_____

Name: _____

Group: _____ Date: _____

ACTIVITY 2 2.3

Student Book, p. 108

Dreams within reach

The Chavez family are ready to build the house of their dreams. They chose three cities where they would like to live, and are analyzing the cost of the land in relation to its area.

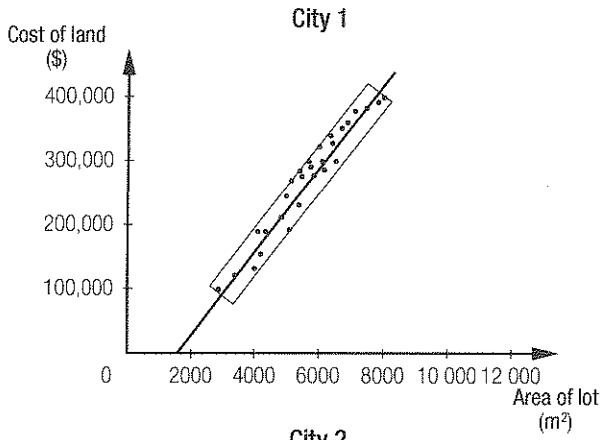
In the Cartesian planes below, they have represented the data collected from the three cities. Each scatter plot has been framed in a rectangle and represented by a line.

a. Describe the linear correlation observed in each scatter plot.

_____

_____

_____



b. In which city does a 6000 m² lot cost the least?

_____

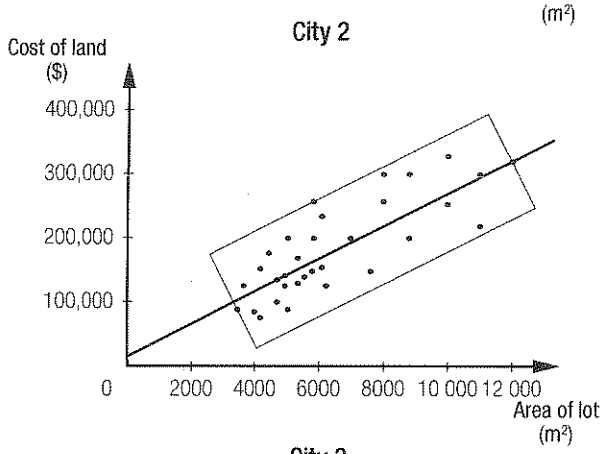
c. 1) For which rectangle is the ratio $\frac{\text{length of small side}}{\text{length of large side}}$ the greatest?

_____

2) Describe this rectangle.

_____

_____



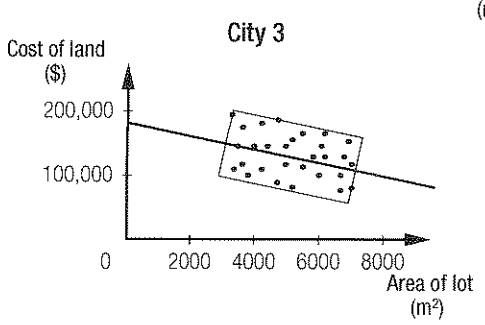
d. 1) For which rectangle is the ratio $\frac{\text{length of small side}}{\text{length of the large side}}$ the smallest?

_____

2) Describe this rectangle.

_____

_____



e. When comparing the three scatter plots, what relationship might you establish between:

1) the reliability of a statistical study and the intensity of the correlation? _____

2) the appearance of the rectangle framing the points and the strength of the correlation? _____

f. For which city does the study seem the least reliable? Explain your answer.

_____

_____

Name: _____

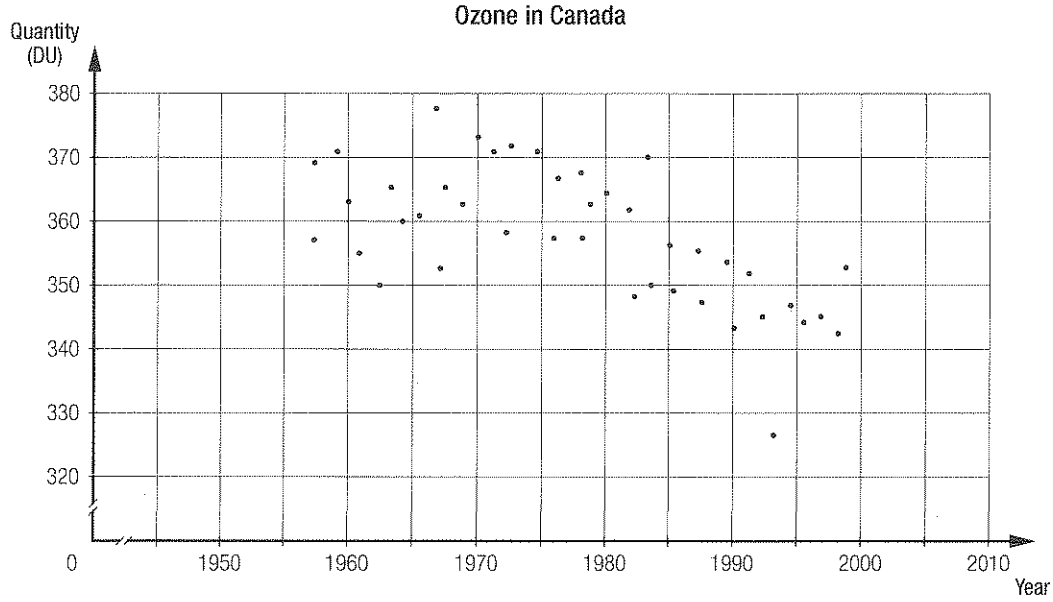
Group: _____ Date: _____

PRACTICE 2.3

Student Book, p. 115

Number 7

7 **OZONE** The graph below represents the annual quantity of ozone (in Dobson units) in Canada.



a) Describe the linear correlation between these two variables.

_____

b) Draw a line that represents the majority of the points.

c) Determine the equation of this line representing the regression line for this situation.

_____

d) Calculate the quantity of ozone in 2010.

_____

Name: _____

Group: _____ Date: _____

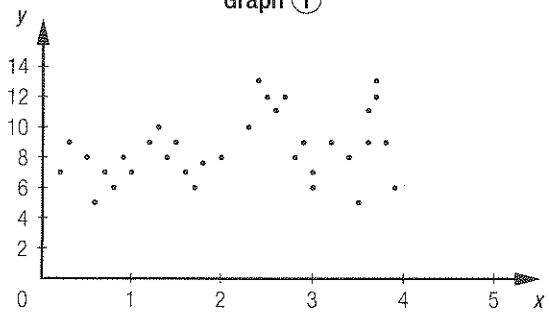
PRACTICE 2.3

Student Book, p. 116

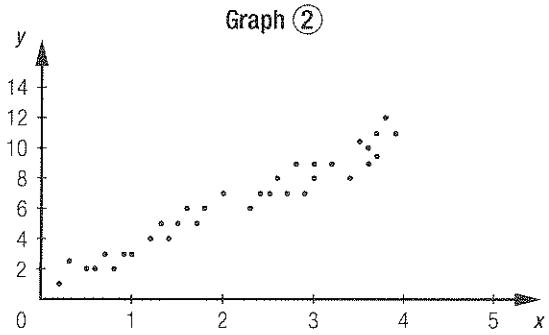
Number 9

9 Place the scatter plots in increasing order according to the linear correlation coefficient associated with each one.

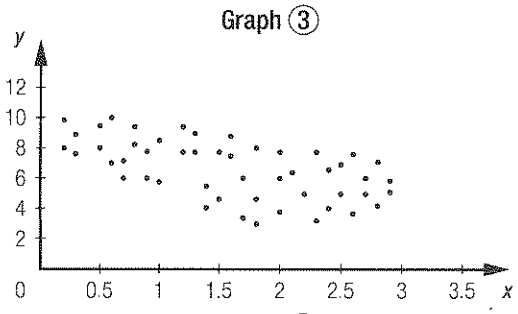
Graph ①



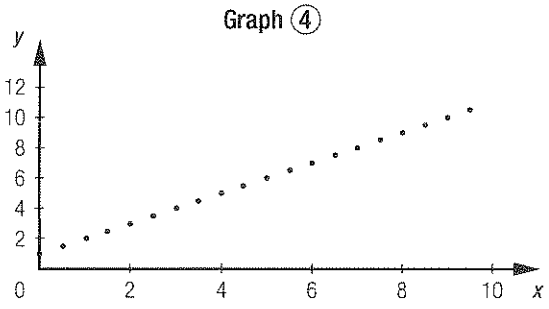
Graph ②



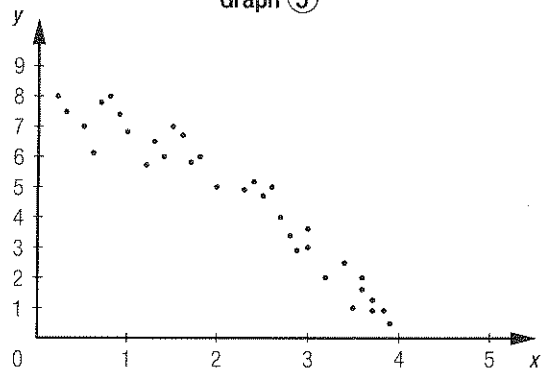
Graph ③



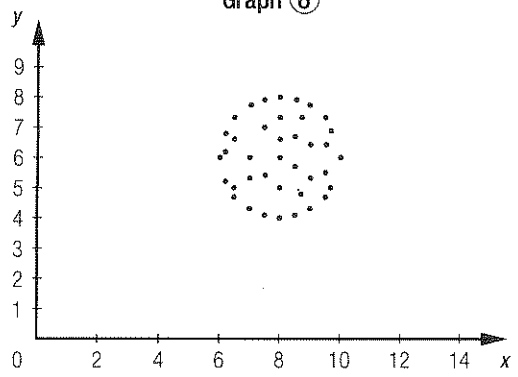
Graph ④



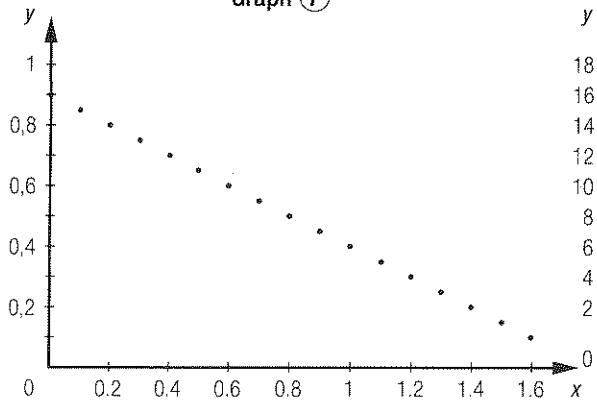
Graph ⑤



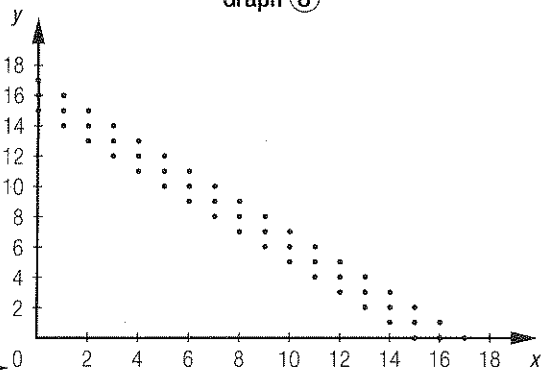
Graph ⑥



Graph ⑦



Graph ⑧



Name: _____

Group: _____ Date: _____

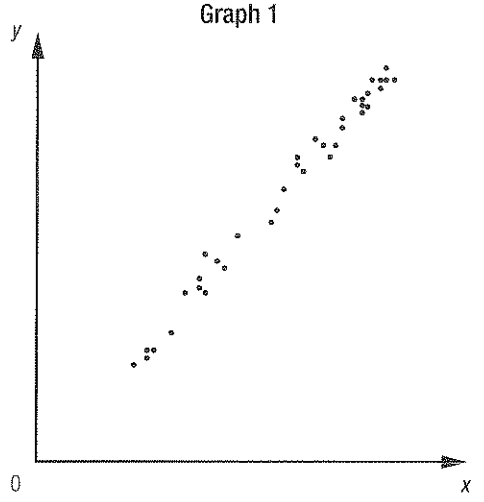
PRACTICE 2.3

Student Book, p. 117

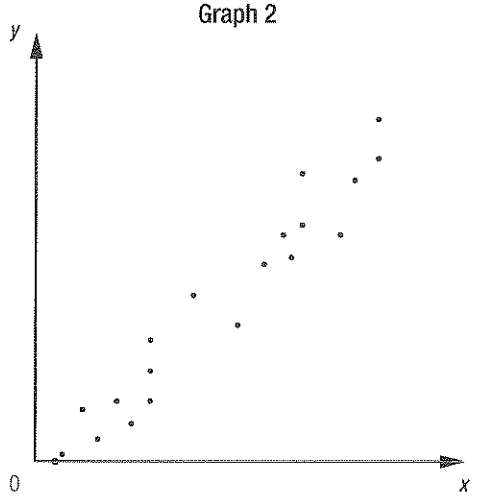
Number 10

10 Graphically estimate the linear correlation coefficient associated with each of the following scatter plots.

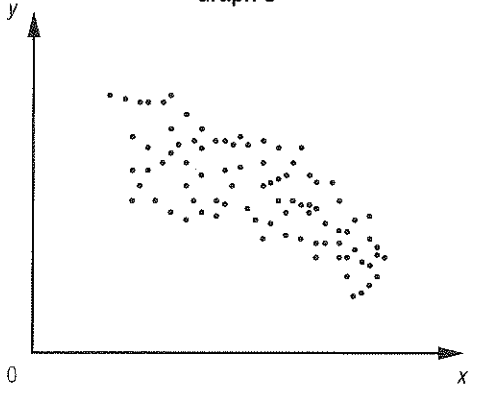
a) Graph 1



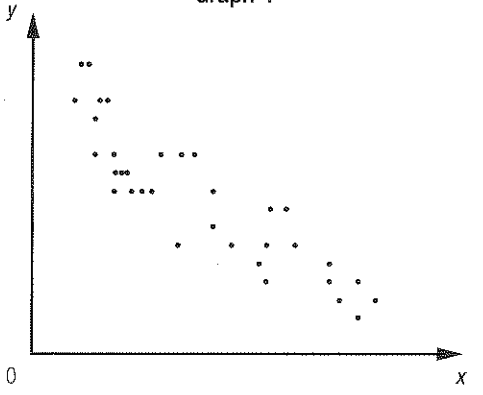
b) Graph 2



c) Graph 3



d) Graph 4



Name: _____

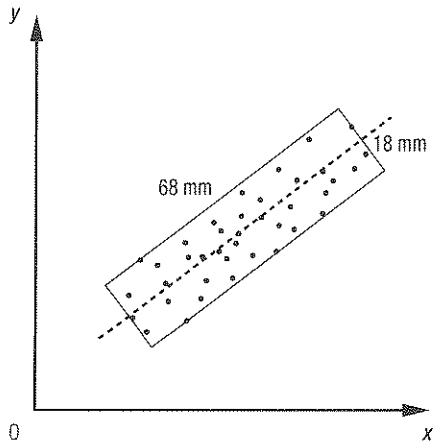
Group: _____ Date: _____

SUPPORT 2.3

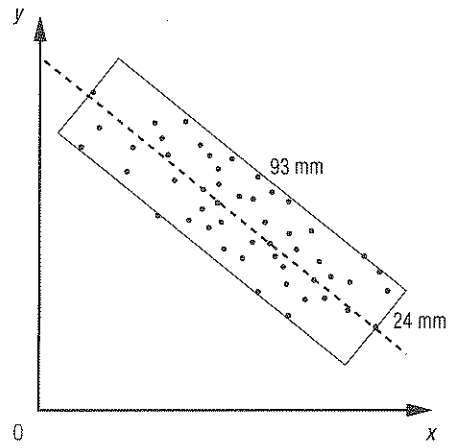
Quantitative interpretation of correlation

1 Graphically estimate the linear correlation coefficient for each of the following scatter plots.

a)



b)



_____ _____

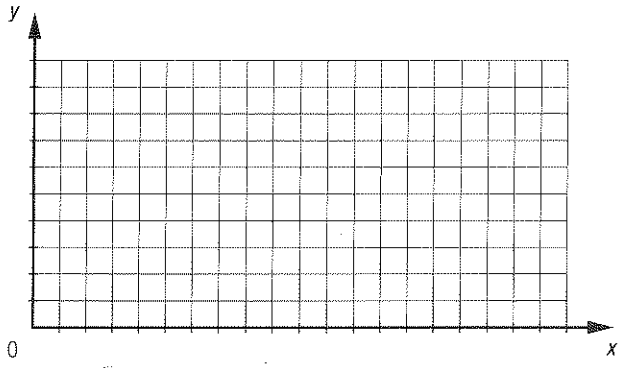
2 This table of values presents the ordered pairs of a statistical distribution.

x	2	4	7	7	8	10	12	15	15	16	18	20
y	15	14	13	15	12	13	12	11	13	11	10	9

a) Using the median-median line method, determine the equation of the regression line.

_____

b) Plot the ordered pairs of the distribution and the regression line in the Cartesian plane below.



Name: _____

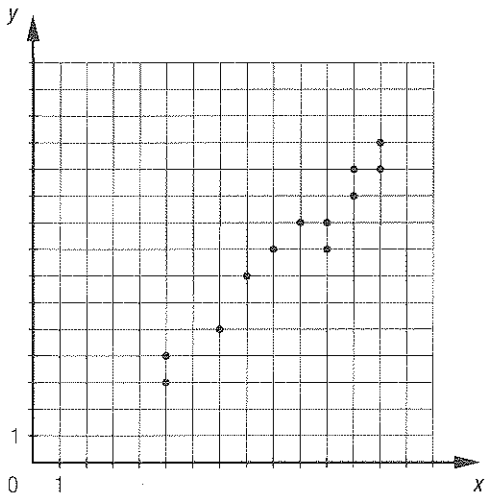
Group: _____ Date: _____

SUPPORT 2.3

(cont'd)

3 Based on the adjacent scatter plot:

a) Determine the equation of the regression line using the Mayer line method.



b) Draw the regression line on the scatter plot provided.

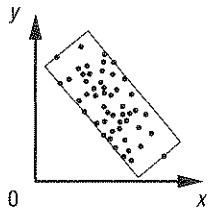
c) Graphically estimate the correlation coefficient.

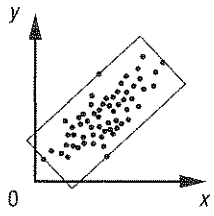
d) Describe the linear correlation between the two variables.

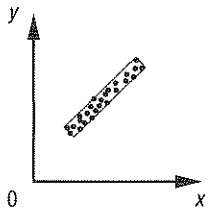
4 Describe the direction and the strength of the linear correlation between two variables if the correlation coefficient is:

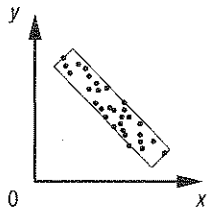
a) -0.45 b) 0.97 c) 0.71 d) -0.85

5 After estimating the linear correlation coefficient of these distributions, describe the correlation for each one.

a) 

b) 

c) 

d) 

Name: _____

Group: _____ Date: _____

CONSOLIDATION 2.3

Quantitative interpretation of correlation

1 Determine the equation of the regression line for each of the table of values below using the median-median line method.

a)

x	2	3	3	7	8	8	11	13	15	16	17	20
y	15	12	13	12	11	8	9	6	7	7	6	5

_____

b)

x	6	7	7	9	10	11	11	12	14
y	33	38	38	36	35	34	39	36	37

_____

2 Determine the equation of the regression line for each of the table of values below using the Mayer line method.

a)

x	10	12	14	15	15	15	16	17	17	18	18	19
y	38	44	51	51	52	54	55	60	58	61	62	66

_____

b)

x	32	32	39	39	42	43	44	46	49	49
y	64	63	61	61	58	57	58	54	54	53

_____

3 Place the correlation coefficients below in ascending order of strength.

0.67 −0.87 −0.4 0.23 0.98 −0.7 0.65 −0.45 0.38 −0.64

_____

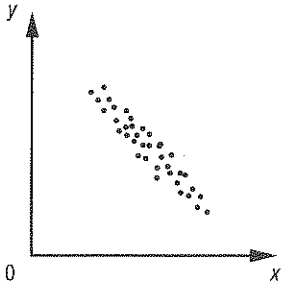
Name: _____

Group: _____ Date: _____

4 For each of the scatter plots below:

1) Estimate the correlation coefficient.
2) Describe the linear correlation between the variables.

a)

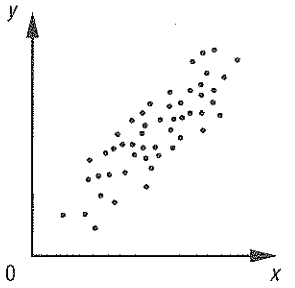


1) _____

2) _____

_____

b)

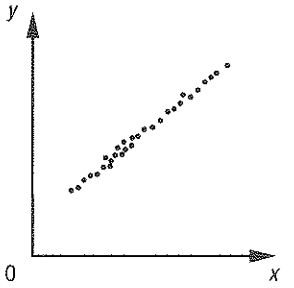


1) _____

2) _____

_____

c)



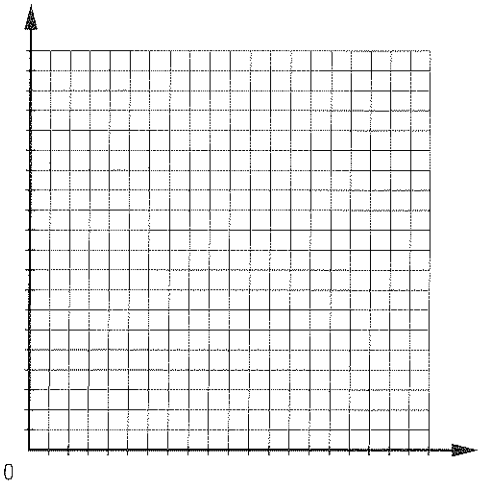
1) _____

2) _____

_____

5 The adjacent table of values provides information about single-family houses for sale in a district.

a) Represent the data using a scatter plot.



Housing market for a district

Surface area of a dwelling (m^2)	Asking price ($)
113	175,500
107	114,900
96	139,900
130	159,900
115	169,000
107	124,500
107	105,900
91	144,900
96	139,900
107	169,000
96	119,000

b) Estimate the linear correlation coefficient between the surface area and the asking price.

_____

c) Determine the equation of the regression line representing this situation.

_____

Name: _____

Group: _____ Date: _____

6 The table of values below presents ordered pairs of data collected during a statistical study.

Statistical study

x	32	23	39	33	25	41	33	27	42	38	28	44	38	31	45
y	11	8	4	4	7	7	9	9	5	2	8	5	8	8	4

a) Determine the equation of the regression line representing this situation using the median-median line method.

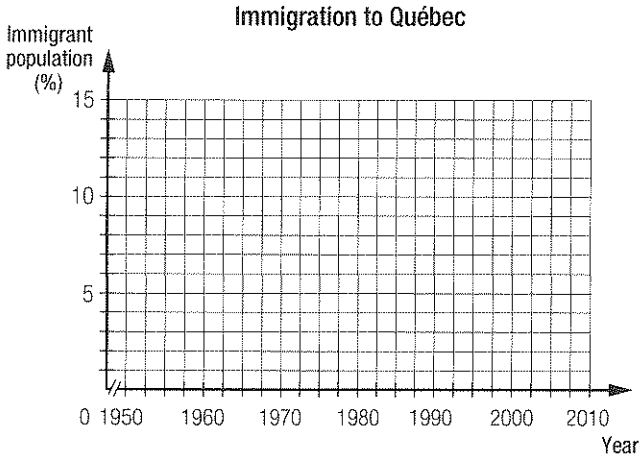
_____

b) Determine the equation of the regression line representing this situation using the Mayer line method.

_____

7 The table below presents data on the progression in the immigrant population of Québec.

a) Construct a scatter plot representing this situation.



Immigration to Québec

Year	Immigrant population (%)
1951	5.6
1961	7.4
1971	7.8
1981	8.2
1986	8.2
1991	8.7
1996	9.4
2001	9.9
2006	11.5

b) Graphically estimate the linear correlation coefficient of this distribution.

_____

c) Describe the linear correlation between the variables.

_____

d) Determine the equation of the regression line using the median-median line method.

_____

e) In what year did the immigrant population exceed 10%? _____

f) What should the percentage of immigrants be in 2020? _____

Name: _____

Group: _____ Date: _____

ENRICHMENT 2.3

Quantitative interpretation of correlation

Solve the crossword puzzle below by completing the following sentences.

a) It is possible to quantify the strength of a linear correlation between two statistical variables using a number called the correlation _____ .

b) By dividing the sum of all data values by the number of values, the _____ is obtained.

c) The values, which divide a distribution into four subsets each containing an equal amount of data values, are known as _____.

d) In a scatter plot, the line which best represents the grouping of points is called the _____ line.

e) In an ordered one-variable distribution having an odd number of values, the _____ is the middle value.

f) A two-variable distribution can be represented using a contingency _____.

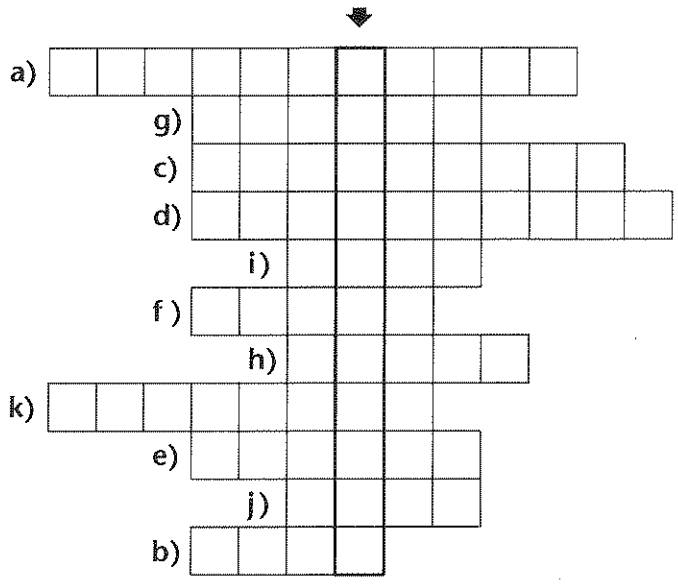
g) A correlation coefficient equal to 1 or -1 indicates that the correlation between the variables is _____.

h) The difference between the highest value and the lowest value of a one-variable distribution is called the _____.

i) A correlation coefficient equal to 0.5 or -0.5 indicates that the correlation between the variables is _____.

j) In a one-variable distribution, the _____ is the value having the greatest frequency.

k) A correlation coefficient equal to 0.71 indicates that the correlation between the variables is _____.



Name: _____

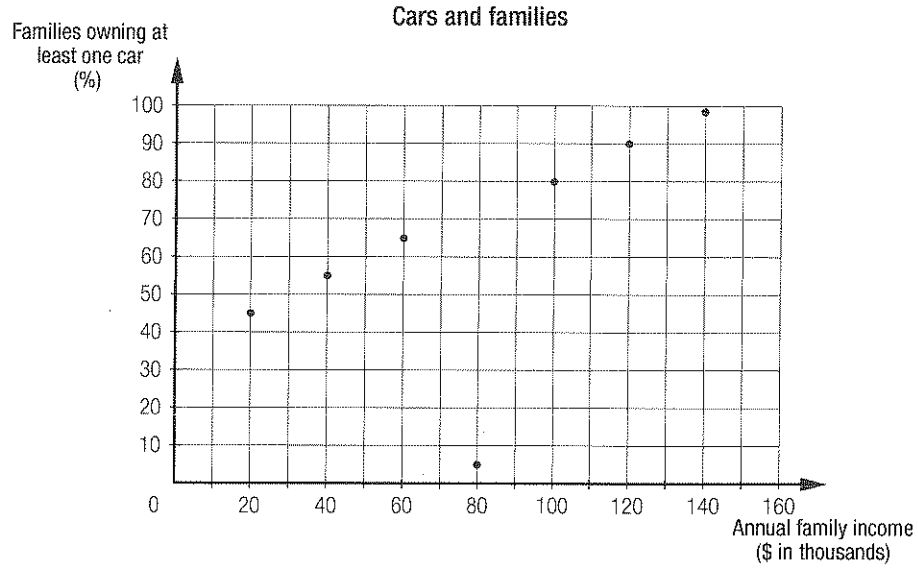
Group: _____ Date: _____

PRACTICE 2.4

Student Book, p. 126

Number 4

4 A statistician examined the annual family household income in a city and the percentage of families who owned at least one vehicle. The data is presented in the graph below.



a) What is the linear correlation coefficient:

1) if you consider all the results?

_____

2) if you do not consider the outliers?

_____

b) Determine the equation of the regression line.

_____

c) What should be the percentage of families who own at least one vehicle and whose annual household income is $85,000?

_____

d) What is the percentage of families with an annual income of $72,000 who own at least one vehicle?

_____

Name: _____

Group: _____ Date: _____

SUPPORT 2.4

Interpretation of linear correlation

1 For each of the tables of values below:

1) Calculate the linear correlation coefficient.
2) Determine the equation of the regression line.

a)

x	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2	2.1	2.2	2.3	2.4
y	1.1	0.3	0.5	1.7	1.9	2.1	2.3	2.5	1.7	1.9	3.1	3.3	2.5	3.7	3.9	4.1	3.3	3.5	3.7	4.9

1) _____ 2) _____

b)

x	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4	4.1
y	0.5	0.7	3.9	4.1	1.3	5.5	5.7	5.9	3.1	2.3	5.5	3.7	3.9	4.1	4.3	5.5	5.7	3.9	8.1	3.3

1) _____ 2) _____

c)

x	4	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9
y	7	6.8	6.6	6.4	6.2	6	5.8	5.6	5.4	5.2	5	5.8	5.6	5.4	4.2	5	3.8	3.6	3.4	3.2

1) _____ 2) _____

d)

x	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8
y	17.4	11.1	13.8	11.5	16.2	8.9	8.6	11.3	13	9.7	11.4	12.1	10.8	10.5	13.2	12.9	11.6

1) _____ 2) _____

2 Here is information about the sale of glasses in a store:

Eyeglass store

Day in the month of February	2	4	6	7	8	10	12	14	15	16	18	20
Number of pairs of glasses sold	15	14	13	15	12	13	12	11	13	11	10	9

a) Determine the equation of the regression line representing this situation.

_____

b) Based on the equation of the regression line found in a):

1) How many pairs of glasses would the store have sold on February 17?

_____

2) On what day in February would eight pairs of glasses have been sold?

_____

3 An inventory was made of the heights and the shoe sizes of men on a public transit system bus.

Height and shoe size

Height (cm)	162	165	166	168	173	175	176	182	186	188	188	192	197	201
Shoe size	8.5	9	9.5	9	9.5	10	9.5	9	10	11	11.5	12	13	12.5

a) 1) What is the linear correlation coefficient between a man's height and his shoe size?

_____

2) Describe this linear correlation.

_____

b) What is the equation of the regression line representing this situation?

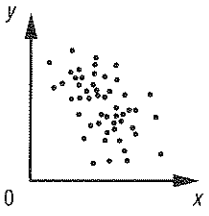
_____

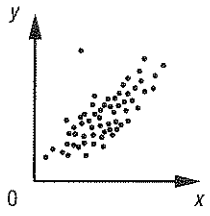
c) According to this data, what would be the height of a man who wears size:

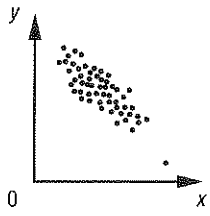
1) 7? 2) 7.5? 3) 8?

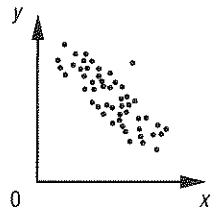
_____ _____ _____

4 Circle, if applicable, the outlier in each of the following scatter plots.

a) 

b) 

c) 

d) 

5 According to the data in the adjacent table, the correlation coefficient between the duration of a film and the profits it generates is ≈ 0.99, which corresponds to a near perfect positive correlation. Is it possible to conclude that a film 180 minutes in length will be a commercial success?

Profits generated by films

Film	Length (min)	Profit ($)
A	60	75,000
B	110	140,000
C	140	165,000

Explain your answer.

_____

_____

_____

Name: _____

Group: _____ Date: _____

CONSOLIDATION 2.4

Interpretation of linear correlation

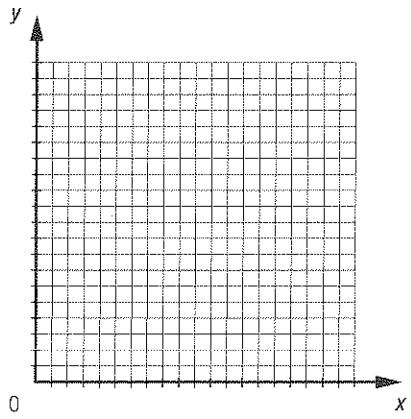
1 The table of values below presents the ordered pairs of a statistical distribution.

x	9	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10	10.1	10.2	10.3	10.4	10.5
y	8	7.8	7.6	6.4	6.2	6	6.8	6.6	5.4	6.2	9	5.8	4.6	5.4	5.2	4

a) Represent the data using a scatter plot.

b) Circle, in the graph, the point which represents the outlier.

c) If the outlier is not included, what would the value of y be if $x = 12$.



_____

d) Is the result obtained in **c)** reliable? Explain your answer.

_____

_____

_____

2 Here are the data collected during a statistical study.

x	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
y	9.1	13.1	11.1	21.1	16.1	20.1	20.1	24.1	27.1	34.1	35.1	33.1	31.1	39.1	42.1	40.1

a) What is the linear correlation coefficient between the two variables?

_____

b) Determine the equation of the regression line.

_____

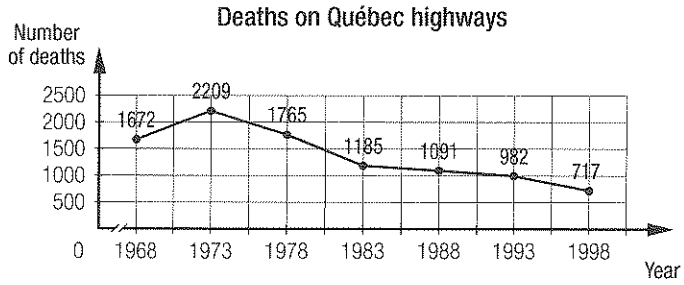
c) According to the data collected, what would the value of x be if $y = 44$?

_____

Name: _____

Group: _____ Date: _____

3 The broken-line graph below illustrates the change in the number of deaths on Québec highways between 1968 and 1998.



a) Describe the linear correlation between the year and the number of deaths during that year.

_____

b) According to this information, calculate the projected number of deaths for 2008.

_____

c) In what year could no highway deaths be expected to occur?

_____

4 The adjacent table provides information concerning Québec blueberry production.

Production of blueberries in Québec

Year	Quantity harvested (kg)	Production cost ($/kg)
1989	2 900 000	1.43
1991	4 393 600	1.35
1993	7 034 046	0.99
1995	4 318 000	1.12
1997	7 730 000	1.32
1999	11 364 000	1.94
2001	20 400 000	0.88
2003	10 000 000	1.21
2005	15 500 000	2.31
2007	14 500 000	2.09

a) If a linear correlation exists:

1) between the year and the quantity of blueberries harvested, what might the harvest be in 2020?

_____

2) between the year and the cost of blueberry production, what might the cost of production be in 2020?

_____

b) Which of the two preceding predictions appears the most reliable? Explain your answer.

_____

_____

_____

Name: _____

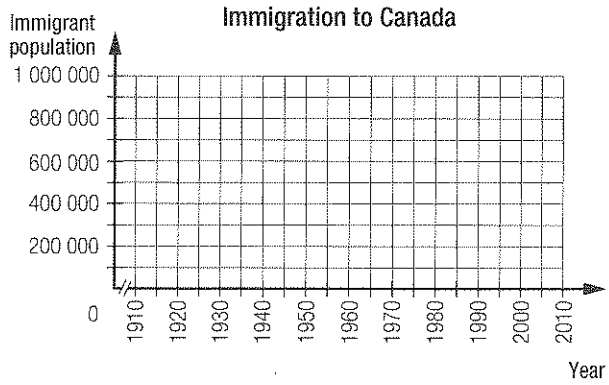
Group: _____ Date: _____

CONSOLIDATION 2.4

(cont'd)

5 The adjacent table presents data about immigrants to Canada.

a) Represent this data using a scatter plot.



b) What do you notice about the change in the immigrant population around 1940? Hypothesize the cause of the phenomenon observed.

_____

_____

Immigration to Canada

Year	Immigrant population
1911	147 070
1921	188 576
1931	251 743
1941	223 943
1951	228 923
1961	388 449
1971	468 930
1981	522 150
1986	527 135
1991	591 210
1996	664 495
2001	706 965
2006	851 560

c) How many immigrants should live in Canada in 2030 if:

1) all the data is taken into account? _____

2) the data taken in 1941 and 1951 are disregarded? _____

6 Here is the data taken at the time of a study on the cultivation of scallops in ponds.

a) According to this information, how many pond-cultivated scallops may die if there are 65 at the start of the study?

_____

b) How many scallops should a pond contain at the start of the study if the losses occurring during cultivation are to be minimized?

_____

c) Are the results obtained in **a)** and **b)** reliable? Explain your answer.

_____

_____

_____

_____

_____

Scallop cultivation

Pond number	Number of scallops at the start	Number of dead scallops during the study
1	1	1
2	5	2
3	10	0
4	15	1
5	20	3
6	25	5
7	30	8
8	35	10
9	40	11
10	45	13

Name: _____

Group: _____ Date: _____

ENRICHMENT 2.4

Interpretation of linear correlation

1 Data from the last four years at a school shows that the linear correlation coefficient between the outside temperature and the mean of the results for final evaluations in French is 0.96.

a) Comment on the validity of this correlation.

_____

_____

_____

b) According to data from the school, the equation of the regression line describing this situation is $y = 6x - 12$, where x is the outside temperature (in °C) and y is the mean of the results (as a percentage). Comment on this situation from a theoretical point of view by using critical values.

_____

_____

_____

2 At the time of research on a medication for lung cancer, researchers carried out tests and obtained a linear correlation coefficient of 0.48 between the duration of the treatment with this medication and an improvement in the state of patients' health. Is this an effective medication? Explain your answer.

_____

_____

_____

3 Each day, for three consecutive months, a nurse working in a nursery recorded data on the changes in the lunar cycle and the number of births. A comparison of the lunar cycle and the number of births shows a linear correlation coefficient of 0.87. Can the following be stated conclusively:

a) The cycles of the moon have an effect on the number of births. Explain your answer.

_____

_____

b) The number of births has an effect on the lunar cycle. Explain your answer.

_____

_____

Name: _____

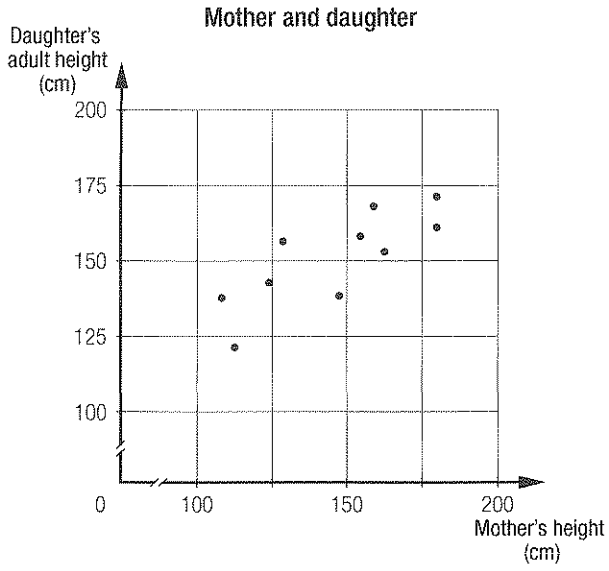
Group: _____ Date: _____

Student Book, p. 133

Number 3

3 Consider the graphical representation of the relationship between a mother's height and the height of her adult daughter. Graphically estimate the correlation coefficient between the two variables.

_____



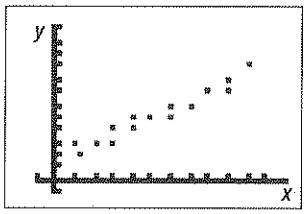
Name: _____

Group: _____ Date: _____

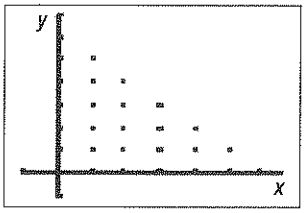
Number 3

3 a) Which of the scatter plots below has the highest linear correlation coefficient?

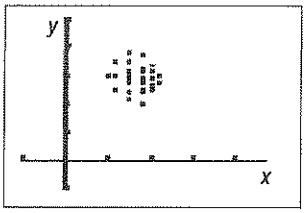
A



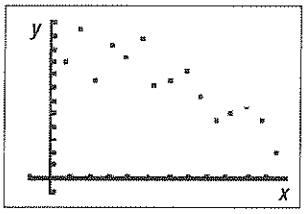
B



C



D



_____

b) Graphically estimate the linear correlation coefficient for each of the scatter plots.

_____

_____

_____

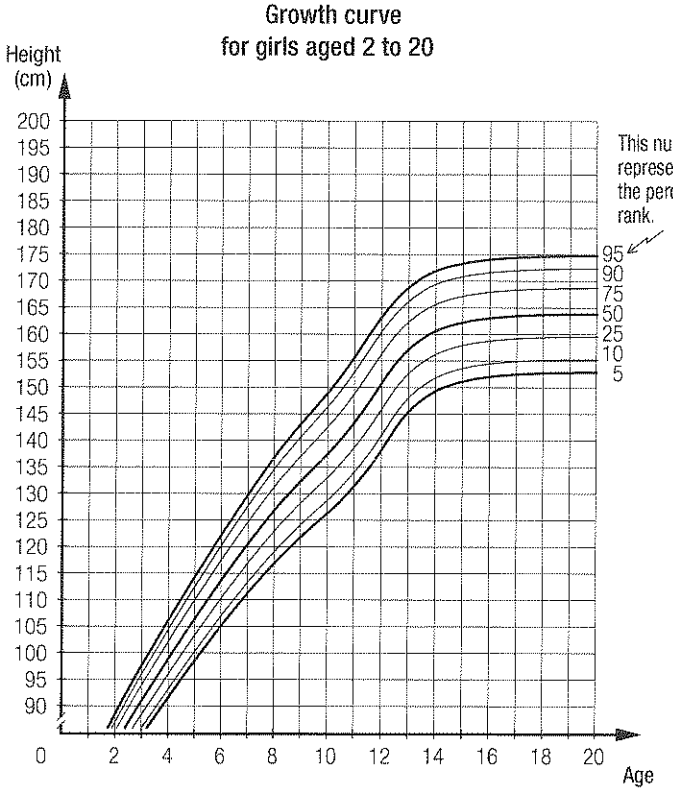
_____

Name: _____

Group: _____ Date: _____

Number 20

20 At a family reunion, Emily compares her age and height to that of her nine cousins. Below are the results:



The Smith family

Name	Age	Height (cm)
Amy	17	167
Anna	15	160
Bianca	14	171
Cynthia	12	163
Diana	2	85
Emily	6	124
Margarita	6	106
Simona	9	129
Siu Ting	18	175
Valerie	16	164

a) What is the percentile of Emily's height in comparison to her cousins?

_____

b) Is there a linear correlation between the cousins' age and height? Explain your answer.

_____

c) According to the growth curve graph for girls aged 2 to 20, answer the following:

1) What is Emily's percentile?

_____

2) How many cousins have the same percentile as Emily?

_____

3) How many cousins have a percentile lower than 50?

_____

d) How tall should Emily be at 20?

_____

Name: _____

Group: _____ Date: _____

BANK OF PROBLEMS 2

Student Book, p. 144

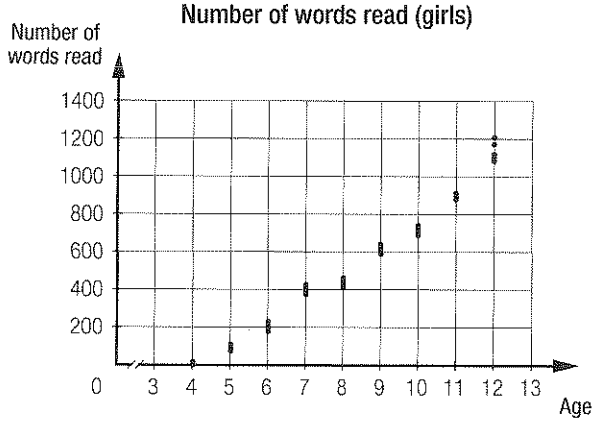
Number 25

25 The number of words that boys and girls of different ages could read was compared. Below are the results:

Number of words read (boys)

(6, 50) (6, 62) (6, 75) (7, 200) (7, 210)
(7, 225) (7, 240) (7, 245) (8, 370) (8, 390)
(8, 410) (8, 440) (9, 587) (9, 590) (9, 621)
(9, 635) (10, 670) (10, 734) (10, 742)
(10, 770) (10, 789) (11, 842) (11, 870)
(11, 919) (11, 956) (11, 989) (12, 1050)
(12, 1095) (12, 1105) (13, 1187) (13, 1230)
(13, 1300) (14, 1450) (14, 1342) (14, 1500)

Each of these ordered pairs represents a boy's age and the number of words he can read.



According to these results, at what age will the boys and girls be able to read the same number of words?

Name: _____

Group: _____ Date: _____

SNAPSHOT 2

1 **A TIGHT RACE** In the context of a friendly competition between a group of students, boys and girls took part in a 400-metre race. The table below presents the race results.

Times (in s) for a 400-m race

	Numbers of the participants															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Time (in s) for the boys	71	67	95	85	87	69	98	74	78	83	78	97	82	85	89	97
Time (in s) for the girls	93	85	72	88	105	97	99	95	84	72	94	82	97	96	84	83

In order to present the race results to race participants, express this situation using a stem-and-leaf plot and determine whether the girls' or the boys' race gave the closest results. Explain your answer.

Name: _____

Group: _____ Date: _____

(cont'd)

2 **THE IMPACT OF THE GAMES** In Atlanta in 1996, one of the members of a canoe-kayak club participated in the Olympic Games. It was a first for this club. Since then, club membership has continued to increase and other athletes from this club have participated in other Olympic games. Here is the change in the number of members between 1996 and 2008:

Memberships in the canoe-kayak club

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Number of members	84	108	111	120	124	158	163	171	175	226	242	254	260

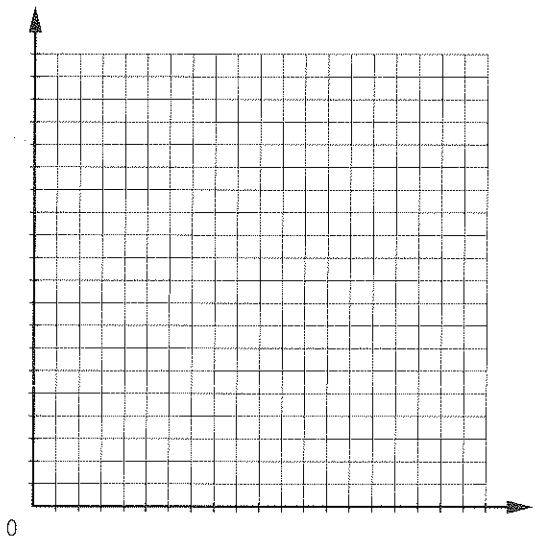
If the trend continues, in what year should the club welcome its 350th member?

Name: _____

Group: _____ Date: _____

3 **HIGHLY POPULATED CITIES** The table below presents the global rank and population of some of the 171 most populated cities in the world. The city of Kabul in Afghanistan is ranked 39th in this distribution. Using this data, estimate the population of the city of Kabul and discuss the validity of this estimation.

City	Global rank	Population (millions of inhabitants)
Tokyo	1	45.2
Shanghai	5	21.9
Rome	92	3.7
Moscow	16	13.8
Detroit	55	5.8
Riyadh	70	4.6
Ghendu	85	4.1
Teheran	20	12.3
Montreal	96	3.6
Boston	33	7.4

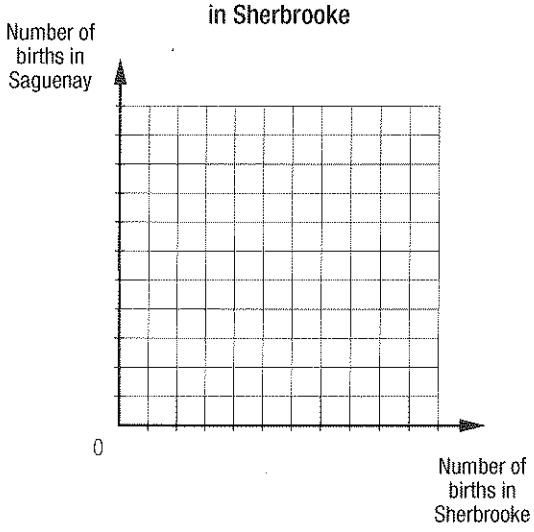


4 **BIRTHRATE** Each year, the Québec government publishes the province's demographic statistics. The contingency table below presents data related to birthrate.

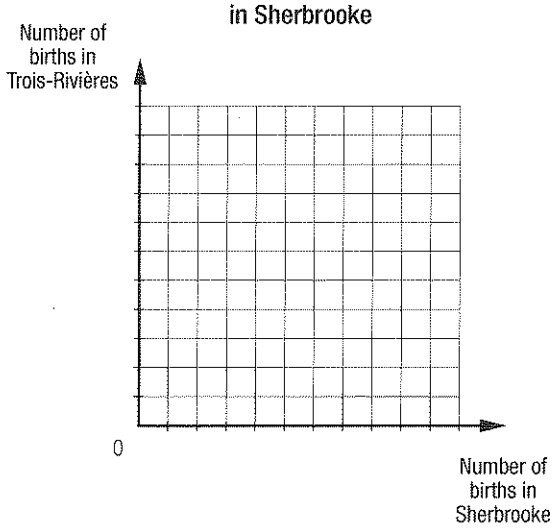
Progression of the number of births in three regions of Québec

Region \ Year	1991	1996	2001	2002	2003	2004	2005	2006	2007
Saguenay	2181	1687	1374	1321	1281	1303	1317	1422	1447
Sherbrooke	1938	1814	1586	1559	1651	1620	1677	1767	1769
Trois-Rivières	1717	1445	1110	1145	1133	1182	1137	1321	1287

Comparison of the number of births in Saguenay and the number of births in Sherbrooke



Comparison of the number of births in Trois-Rivières and the number of births in Sherbrooke



Joachim claims that the number of births occurring in Sherbrooke can be used to estimate the number of births for the two other regions. Do you agree? Justify your answer using mathematical arguments linked to the correlation.

Name: _____

Group: _____ Date: _____

5 **POOL SALES** The inventory of a store selling pools shows that, for June of last year, the store sold an average of 4.5 pools per day. The ordered pairs below indicate the outdoor temperature (in °C) and the number of pools sold each day during this month.

(23, 4), (18, 3), (18, 2), (15, 1), (23, 7), (10, 0), (18, 2), (26, 7), (23, 5), (26, 9), (28, 12), (28, 11), (22, 3), (25, 5), (27, 9), (28, 10), (22, 2), (23, 3), (28, 9), (21, 3), (18, 1), (19, 2), (15, 0), (21, 3), (23, 4), (28, 7), (27, 6), (22, 3), (19, 2), (16, 1)

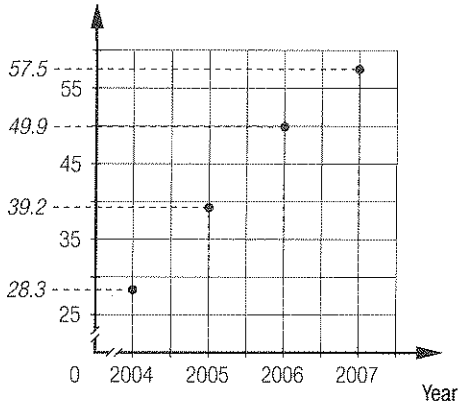
Using a technological tool and basing your calculations on the data, estimate the number of pools that the store will sell during June of this year if the average temperature anticipated for the month is 25°C. Comment on the validity of the result obtained.

6 **PREDICTING THE FUTURE** Using a linear model, two statisticians have made predictions about online sales during 2030. The predictions of Statistician 1 were made based on data for the period of 1999 to 2003. Statistician 2 made the same analysis, but based predictions on data for the period of 2004 to 2007.

Online sales in Canada
Statistician 1

Year	Sales (millions $)
1999	4.2
2000	7.2
2001	10.4
2002	13.7
2003	18.9

Online sales in Canada
Statistician 2



After making your own prediction for online sales in Canada for 2030, how would you convince a friend that the predictions of each of the two statisticians are less reliable than your prediction.

Knowledge in action — Page 4

3.

Mode	Median	Mean	Range
8	14	15	27
5 and 9	6.5	≈ 7.06	5
44	33.5	≈ 31.07	32
None	20	≈ 20.54	31

Activity 2 — Page 5

a. 1) The numbers represent the digit in the tens position for each diameter of the inventoried oaks.
   2) 3 values
   3) 4th line
   4) 6th line
   5) The digit in the tens position is the same.

b. 1) 55 cm 2) 68 cm

c. *Several answers possible. Example:*
The stem-and-leaf plot permits the rapid discovery of the distribution's mode as well as various trends. Also, this type of diagram permits the visualization of all the data in a distribution, in ascending order.

d. 1) ≈ 38.71% 2) ≈ 9.68%
   3) ≈ 74.19% 4) ≈ 12.9%

e. **Red oak: trunk diameter (cm)**

4	0 5
5	3 5 5 5 5 9
6	0 0 3 4 4 5 7 7 8
7	0 0 0 1 2 5 5 9
8	1 2 3 3
9	1 6 6 7
10	2 9
11	3

Support 2.1 — Page 6

1. a) 1) 6 2) 2.4
   b) 1) 8 2) 3
   c) 1) 60 2) 24
   d) 1) ≈ 9.29 2) ≈ 4.98
   e) 1) ≈ 7.43 2) ≈ 3.51
   f) 1) 15.25 2) ≈ 2.81

2. a) 64, 68, 69, 72, 76, 84, 88, 89, 92, 95
   b) 68, 74, 76, 79, 83, 85, 87, 89, 92, 94

3. a) 34th percentile b) 39th percentile
   c) 45th percentile d) 100th percentile

Support 2.1 (*cont'd*) — Page 7

4. a) **Number of baskets of strawberries picked**

0	2 3 4 4 5 6 6 7 7 8 8
1	0 0 1 2 3 3 5 5 6 6 6 7 8 9
2	0 3

b) **Size of snakes in a zoo (cm)**

1	2
2	7
3	1 2 6
4	2 2 3
5	2 5 6 7
6	1 2 7 8
7	8
8	6 7 9

c) **Number of pages in a paperback novel**

17	2
18	0 1 7
19	6 8
20	0 2 4 4
21	5 7 7
22	1 2
23	0 1 6

d) **Duration of musical pieces (in s)**

22	8
23	6
24	1 2 9
25	0 3 4
26	1 3 4 8 9
27	1 1 2 4
28	5

5. a) 87 b) 110 c) 63 d) 31

Consolidation 2.1 — Page 8

1. a) 82nd percentile b) 21st percentile
   c) 55 years

2. a) ≈ 6.1 b) ≈ 6.75

3. a) 1) 195th value 2) 78th value
   b) 65th percentile

Consolidation 2.1 (*cont'd*) — Page 9

4. a) **Number of annual film rentals from a video rental store**

Women								Men				
						8	0	5	6			
9	4	2	2	2	1	1	1	5	7	8		
						4	2	0	2	6	9	
							3	0				
			7	3	1	0	4	0	5	8		
				9	5	4	5	0	1	5		
							6	1	4			

   b) The mean deviation of the number of films rented annually by women is ≈ 17.11 rentals, which is higher than the ≈ 16.27 mean deviation of film rentals by men.

5. a) ≈ 5.83 years.
   b) **Participant ages**

1	6	6	7	7	7	7	8		
2	0	1	2	5	5	6	6	7	8
3	0	1	1	3	3	4	4	4	5

Consolidation 2.1 (*cont'd*) — Page 10

6. a) 1) 140 kg per big cat
      2) 70
   b) 1) 192 meals per day
      2) 11.6

7. a) The mean deviation of total precipitation for the five Canadian cities is about 135.25.
   b) The mean deviation for Vancouver is about 4.12 and the mean deviation for the five cities shown in the table is about 8.74.

Enrichment 2.1 — Page 11

1. a) Q_1: 25th percentile Q_2: 50th percentile Q_3: 76th percentile
   b) No. The percentiles of the quartiles can vary according to the total number of values and the presence of identical data in the distribution.

2. 17 and 18.

3. *Several replies are possible. For example:* 1, 2, 5, 7, 8, 10, 12, 14, 16, 17, 18

Activity 1 — Page 12

a. 1) 140 individuals
   2) 95 individuals
   3) 31 individuals
   4) 8 individuals
   5) 9 individuals at the most

b. ≈ 28.57%

c. 1) Yes. According to the data shown in the table, the two variables seem to vary in the same direction.
   2) Yes. According to the data shown in the table, the higher the dose of medication, the greater the decrease in epileptic seizures

Practice 2.2 — Page 13

1. **Legal age of marriage**

Woman's age \ Man's age	18	19	20	21	Total
16	12	0	2	3	17
17	7	0	5	6	18
18	15	0	9	15	39
19	0	0	8	13	21
Total	34	0	24	37	95

Practice 2.2 — Page 14

7. a) **Start of the 2007-2008 season**

Games played \ Point scored	[0, 10[	[10, 20[	[20, 30[	[30, 40[	[40, 50[	Total
[0, 10[	1	0	0	0	0	1
[10, 20[	4	0	0	0	0	4
[20, 30[	3	0	0	0	0	3
[30, 40[	2	3	0	0	0	5
[40, 50[	0	4	2	4	1	11
Total	10	7	2	4	1	24

   b) 14 players
   c) 1 player
   d) 10 players
   e) ≈ 31.58%
   f) Yes, because the numbers noted in the contingency table are relatively aligned along one of the table's diagonals.

Practice 2.2 — Page 15

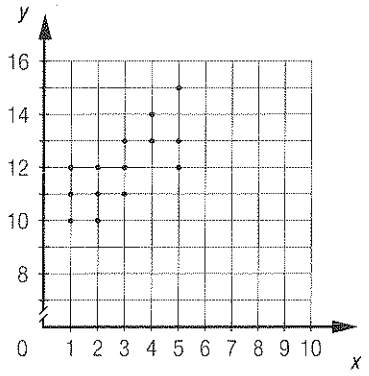
12. a) *Several replies are possible.*
    b) Yes, the correlation is linear, strong and positive.
    c) The ratio's quotient is close to 1.62.

Support 2.2 — Page 16

1. a)

x \ y	10	11	12	13	14	15	Total
1	1	1	3	0	0	0	5
2	1	1	2	0	0	0	4
3	0	1	2	1	0	0	4
4	0	0	0	1	2	0	3
5	0	0	1	1	0	2	4
Total	2	3	8	3	2	2	20

b)



c) Positive and moderate.

2. a) Negative and weak.
   b) Positive and moderate.
   c) Positive and strong.
   d) Negative and moderate.

Support 2.2 (*cont'd*) — Page 17

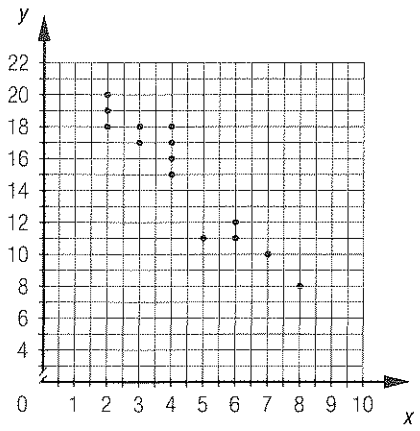
3. a) **Distribution of students at a CEGEP**

Sector \ Gender	Male	Female	Total
Pre-university	437	788	1225
Technical	996	1029	2025
Total	1433	1817	3250

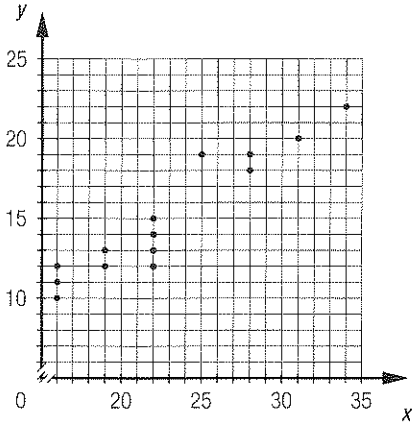
b) **Catch distribution by fisherman age**

Number of catches \ Age	15	16	17	18	Total
[0, 5[	3	12	3	1	19
[5, 10[	17	18	20	16	71
[10, 15[	21	32	44	3	100
[15, 20[	13	16	13	5	47
Total	54	78	80	25	237

4. a) 1)



2)

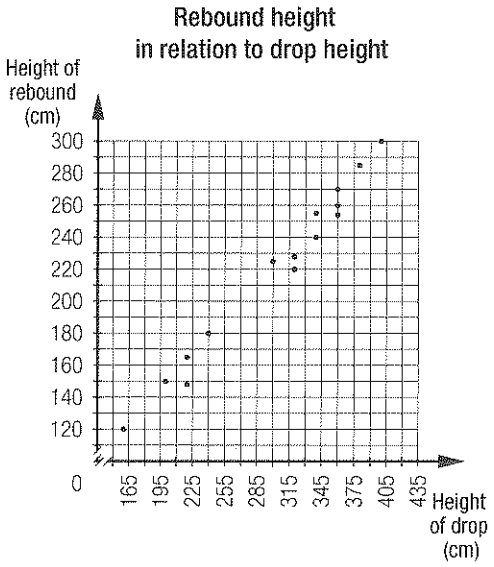


b) 1) Negative and strong.
   2) Positive and strong.

Consolidation 2.2 — Page 18

1. a) Negative and strong correlation.
   b) No correlation.
   c) Positive and very strong correlation.
   d) Positive and moderate correlation.

2. a)



b) Positive and very strong correlation.

Consolidation 2.2 (*cont'd*) — Page 19

3. a)

Pittsburgh Penguins

Number of games played \ Players' ages	[15, 20[	[20, 25[	[25, 30[	[30, 35[	[35, 40[	[40, 45[	Total
[22, 34[	0	0	1	0	0	0	1
[34, 46[	0	0	1	1	0	1	3
[46, 58[	0	2	1	0	0	0	3
[58, 70[	0	2	0	0	0	0	2
[70, 82]	1	1	6	5	1	0	14
Total	1	5	9	6	1	1	23

b) Negative and very weak correlation.

4. a) Positive and moderate correlation.
   b) Positive and moderate correlation.
   c) No correlation.
   d) Negative and strong correlation.
   e) Negative and perfect correlation.

Consolidation 2.2 (*cont'd*) — Page 20

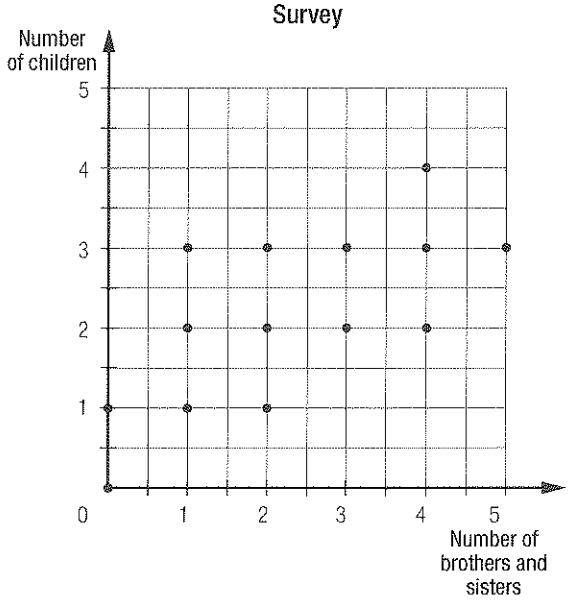
5. a)

House sales

Price (in thousands of $) \ Number of months	[5, 10[	[10, 15[	[15, 20[	[20, 25[	[25, 30[	Total
[100, 150[	1	2	0	0	0	3
[150, 200[	0	4	3	0	0	7
[200, 250[	0	0	2	1	0	3
[250, 300[	0	0	0	1	0	1
[300, 350[	0	0	0	0	1	1
Total	1	6	5	2	1	16

b) The correlation is linear and positive.

6.



The correlation between these two variables is moderate and positive. Therefore, there seems to be some kind of relation between the number of brothers and sisters an individual has and the number of children that individual has.

Enrichment 2.2 — Page 21

1. The greater the amount of fruit consumed, the greater the increase in blood sugar level. There is a positive and very strong correlation between the amount of fruit an individual consumes and blood sugar level.

2. a)

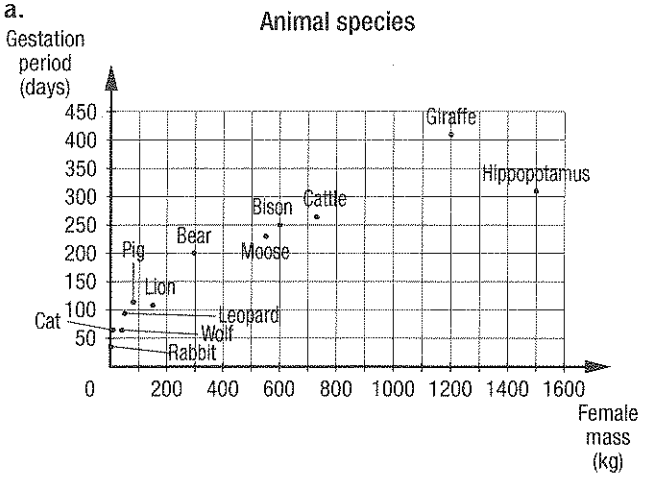
Gender \ Wearing of glasses	Yes	No	Total
Male	2	14	16
Female	4	4	8
Total	6	18	24

b)

Gender \ Writing hand	Left	Right	Total
Male	1	15	16
Female	3	5	8
Total	4	20	24

Activity 1 — Page 22

a.



Activity 1 (*cont'd*) — Page 23

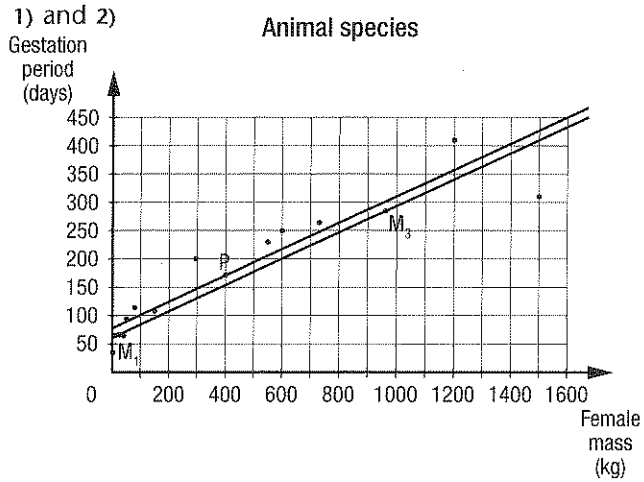
b.

Animal group	Median mass (kg)	Median gestation period (days)	Ordered pair composed of the median mass and the median gestation period
Rabbit, cat, wolf, leopard	24.5	64.5	$M_1(24.5, 64.5)$
Pig, lion, bear, moose	222.5	157	$M_2(222.5, 157)$
Bison, cattle, giraffe, hippopotamus	965	287	$M_3(965, 287)$

c. 1) 404 2) 169.5

d. $\frac{445}{1881} \approx 0.24$

e. 1) and 2)

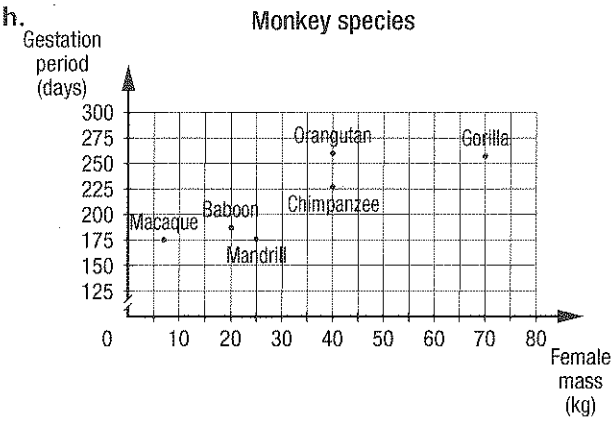


f. $y = \frac{445}{1881}x + \frac{278\,099}{3762}$ or $y \approx 0.24x + 73.92$

g. 1) $\approx$ 334.16 days. 2) $\approx$ 110.23 kg

Activity 1 (*cont'd*) — Page 24

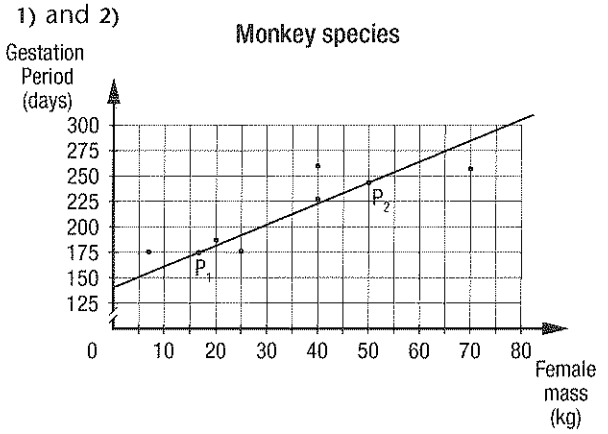
h.



i.

Monkey group	Median mass (kg)	Mean gestation period (days)	Ordered pair composed of the mean mass and the mean gestation period
Baboon, macaque, mandrill	$\approx$ 17.33	$\approx$ 179.33	$P_1(\approx 17.33, \approx 179.33)$
Chimpanzee, gorilla, orangutan	50	248	$P_2(50, 248)$

j. 1) and 2)



k. $y \approx 2.1x + 142.9$

l. 1) $\approx$ 151.31 days 2) $\approx$ 41.44 kg

Activity 2 — Page 25

a. City **1**: positive and strong correlation.
City **2**: positive and moderate correlation.
City **3**: negative and moderate correlation.

b. In city **3**.

c. 1) The rectangle associated with the scatter plot representing the data for city **3**.

2) *Several replies are possible. For example:* The length of the rectangle is two times its width.

d. 1) The rectangle associated with the scatter plot representing data for city 1.
2) *Several answers possible. Example:* The rectangle is long and narrow.

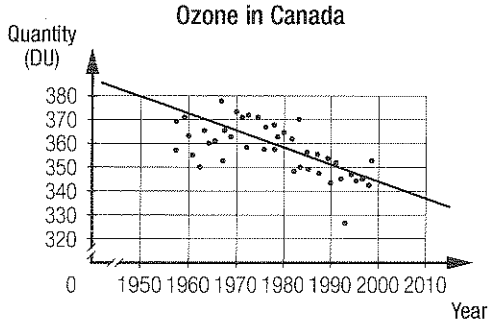
e. 1) The stronger the correlation, the greater the validity of the statistical study.
2) The weaker the ratio $\frac{\text{length of the short side}}{\text{length of the long side}}$ the stronger the correlation.

f. In the case of city 3. It is in the case of the rectangle associated with city 3 that the ratio $\frac{\text{length of the short side}}{\text{length of the long side}}$ is the greatest, signifying that it is in this graph that the line is the least representative of the distribution.

Practice 2.3 — Page 26

7. a) Negative and moderate correlation.

b)



c) $y \approx -0.67x + 1678$

d) $\approx$ 338 DU

Practice 2.3 — Page 27

9. ⑦, ⑧, ⑤, ③, ⑥, ①, ②, ④.

Practice 2.3 — Page 28

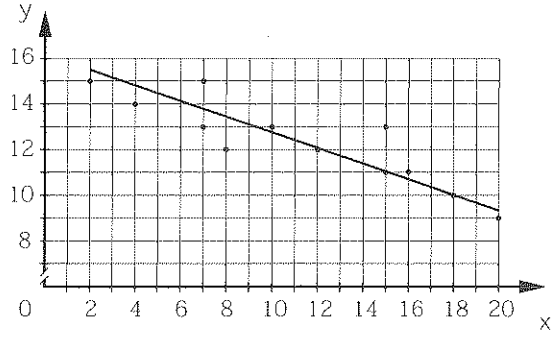
10. a) Graph 1: ≈ 0.96
b) Graph 2: ≈ 0.85
c) Graph 3: ≈ -0.44
d) Graph 4: ≈ -0.76

Support 2.3 — Page 29

1. a) ≈ 0.74 b) ≈ -0.74

2. a) $y = \frac{-8}{23}x + \frac{373}{23}$

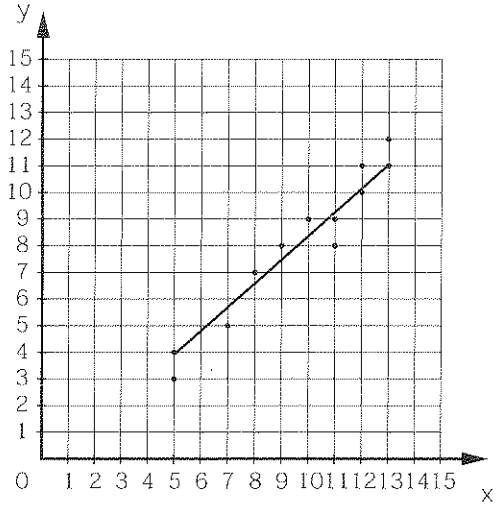
b)



Support 2.3 (*cont'd*) — Page 30

3. a) $y = \frac{25}{28}x - \frac{23}{42}$

b)



c) ≈ 0.88

d) Positive and strong

4. a) Negative and weak correlation.
b) Positive and very strong correlation.
c) Positive and moderate correlation.
d) Negative and strong correlation.

5. a) -0.62, negative and weak correlation.
b) 0.58, positive and weak correlation.
c) 0.85, positive and strong correlation.
d) -0.81, negative and strong correlation.

Consolidation 2.3 — Page 31

1. a) $y = \frac{-4}{9}x + \frac{727}{54}$ b) $y = \frac{-1}{5}x + \frac{193}{5}$

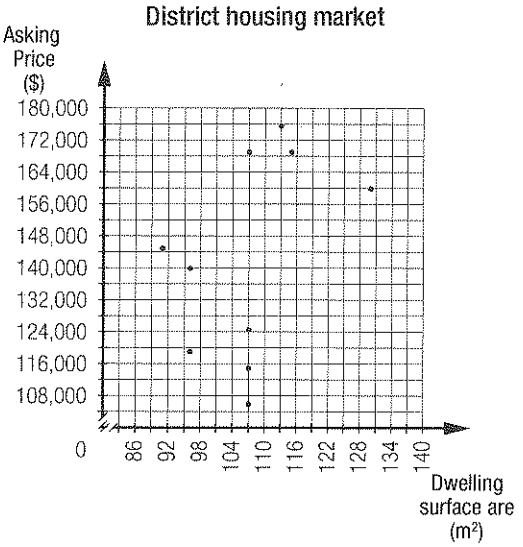
2. a) $y = 3x + \frac{47}{6}$ b) $y = \frac{-31}{47}x + \frac{20\,133}{235}$

3. 0.23 0.38 -0.4 -0.45 -0.64
0.65 0.67 -0.7 -0.87 0.98

Consolidation 2.3 (*cont'd*) — Page 32

4. a) 1) ≈ -0.86 2) Negative and strong correlation.
   b) 1) ≈ 0.73 2) Positive and moderate correlation.
   c) 1) ≈ 0.94 2) Positive and very strong correlation.

5. a)

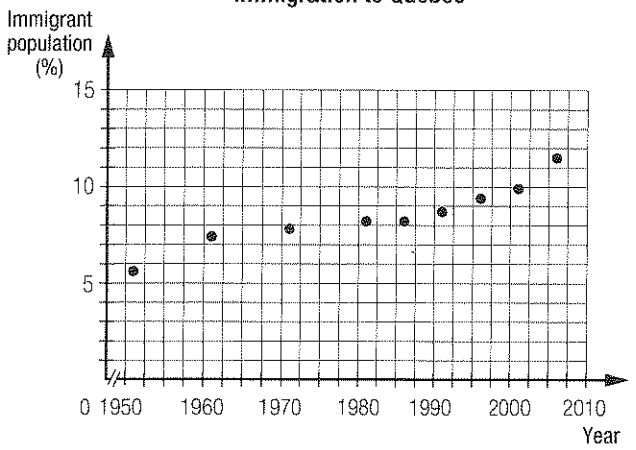


   b) ≈ 0.27
   c) Median-median line method:
   $y \approx 1616.67x - 29\,561.11$
   Mayer line method:
   $y \approx 1391.23x - 5307.72$

Consolidation 2.3 (*cont'd*) — Page 33

6. a) $y = -0.27x + 17.8$ b) $y = \frac{-x + 61}{4}$

7. a)



   b) $r \approx 0.92$
   c) Positive and strong correlation.
   d) $P(y) \approx 0.063y - 116.92$, or $P(y)$ corresponds with the percentage of immigrant population, and y, to the year.
   e) About 2002.
   f) About 11.1% immigrant.

Enrichment 2.3 — Page 34

a) C O E F F I C I E N T
g) S T R O N G
c) Q U A R T I L E S
d) R E G R E S S I O N
i) W E A K
f) T A B L E
h) R A N G E
k) M O D E R A T E
e) M E D I A N
j) M O D E
b) M E A N

Practice 2.4 — Page 35

4. a) 1) ≈ 0.4 2) ≈ 1
   b) $y \approx 0.44x + 37.05$ c) $\approx 74.45\%$ d) 68.8%

Practice 2.4 — Page 36

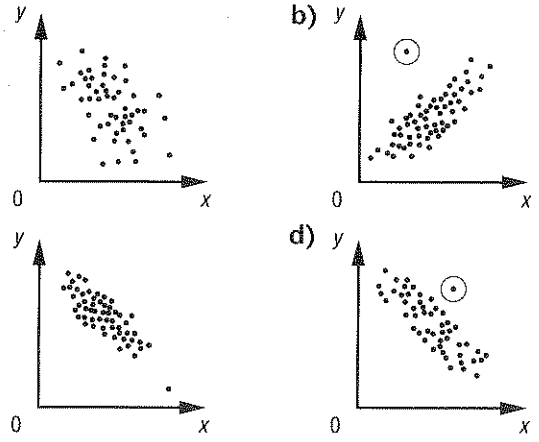
1. a) 1) ≈ 0.91 2) $y \approx 1.89x - 0.15$
   b) 1) ≈ 0.52 2) $y \approx 1.65x - 1.14$
   c) 1) ≈ -0.94 2) $y \approx -1.8x + 14.23$
   d) 1) ≈ -0.27 2) $y \approx -1.24x + 20.65$

2. a) $y \approx -0.29x + 15.49$
   b) 1) ≈ 11 pairs of glasses
   2) February 26

Practice 2.4 (*cont'd*) — Page 37

3. a) 1) ≈ 0.9 2) Positive and strong correlation.
   b) $y \approx 0.11x - 8.61$
   c) 1) ≈ 141.91 cm 2) ≈ 146.45 cm
   3) ≈ 151 cm

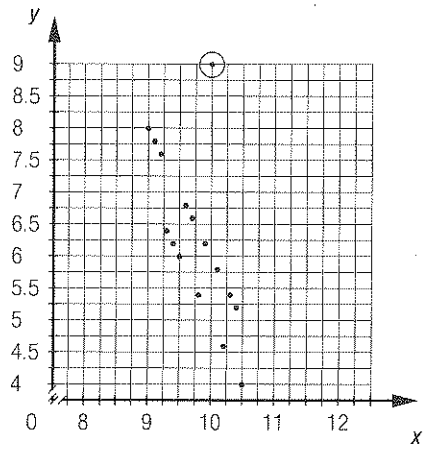
4. a) b)



5. There is too little data for the results to be significant. For example, a longer film is not necessarily more appreciated by the public and vice versa.

Consolidation 2.4 — Page 38

1. a)



   b) See the graph in a).

   c) ≈ 1.4

   d) Yes, the result is very reliable because the correlation coefficient is ≈ -0.9, which indicates a strong correlation.

2. a) ≈ 0.97

   b) $y \approx 2.18x - 7.68$

   c) ≈ 23.76

Consolidation 2.4 (*cont'd*) — Page 39

3. a) Negative and strong correlation ($r \approx -0.89$).

   b) 300 deaths in 2008.

   c) In 2015.

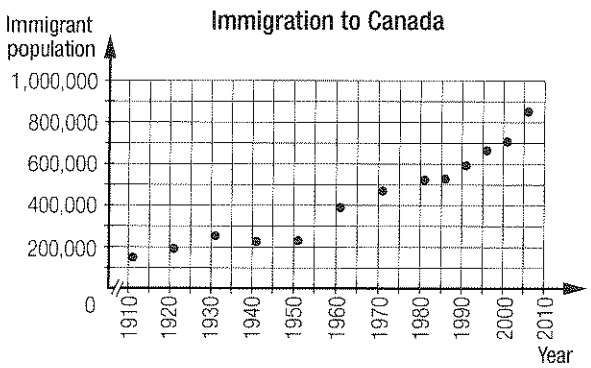
4. a) 1) ≈ 26 404 269.27 kg

      2) ≈ $2.37

   b) The first prediction, since the correlation coefficient is 0.81, whereas it is 0.52 for the second.

Consolidation 2.4 (*cont'd*) — Page 40

5. a)



   b) There were significantly less immigrants to Canada. World War II took place between 1939 and 1945.

   c) 1) There should be about 878 209 immigrants.

      2) There should be about 882 233 immigrants.

6. a) About 18 scallops may die.

   b) There should be 10 to 15 scallops per pond: these are the two values for which there were the least loss.

   c) Yes, it would appear so, as the correlation coefficient is about 0.95. However, several other factors can influence the results: water temperature, predators, marine current, etc.

Enrichment 2.4 — Page 41

1. a) *Several answers possible. Example:* The number of reference years, i.e. 4, is too small to qualify the correlation coefficient as valid.

   b) According to this equation, for all temperatures below 12°C, the mean of the results would be less than the passing mark (60%); if it were 2°C, the mean would be 0% and if it were more than 18.67°C, the mean would be greater than 100%. As such, the final evaluations took place in the months of May or June when the temperature was sometimes greater than 18.67°C.

2. *Several answers possible. Example:* Even if the correlation coefficient is weak, the medication may still be able to improve the state of health of certain patients.

3. a) *Several answers possible. Example:* According to the study done by the nurse, it could be concluded that more data would be necessary in order to validate the results.

   b) No, because the lunar cycle can only be modified by stars or other major forces such as the moon's impact by a large meteorite; the number of births cannot therefore have an effect.

Chronicle of the past — Page 42

3. ≈ 0.67

Overview — Page 43

3. a) A

   b) A: ≈ 0.91 B: ≈ -0.7

      C: 0 D: ≈ -0.76

Overview — Page 44

20. a) 25
    b) Yes, the correlation coefficient is ≈ 0.96.
    c) 1) 95 2) 2 cousins 3) 3 cousins
    d) ≈ 175 cm

Bank of problems — Page 45

25. About 11 years.

Snapshot 2 — Page 46

1. **Times (s) for a 400-metre race**

Girls							Boys						
						6	7	9					
				2	2	7	1	4	8	8			
	8	5	4	4	3	2	8	2	3	5	5	7	9
9	7	7	6	5	4	3	9	5	7	7	8		
						5	10						

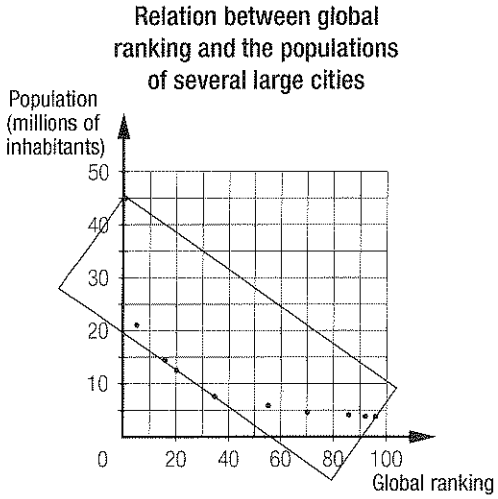
The girls' race was closer, because the mean deviation for the boys is ≈ 8.19 s whereas for the girls it is ≈ 7.875 s.

Snapshot 2 (*cont'd*) — Page 47

2. The equation for the regression line is: $y \approx 15.07x - 30\,004.08$. The club should welcome its 350th member in 2014.

Snapshot 2 (*cont'd*) — Page 48

3. The 39th percentile corresponds to the 66th value. Kabul is therefore ranked 66th globally. The equation for the regression line for the relation between global ranking and population is: $y = -0.21x + 20.57$.



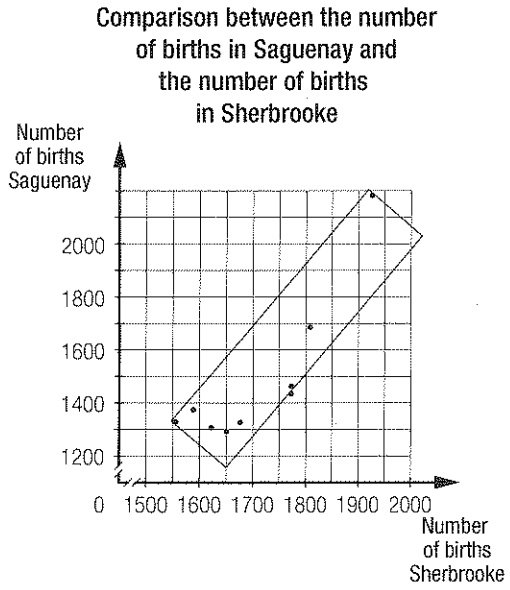
This equation estimates a population of 6.71 million inhabitants for the city ranked 66th globally.

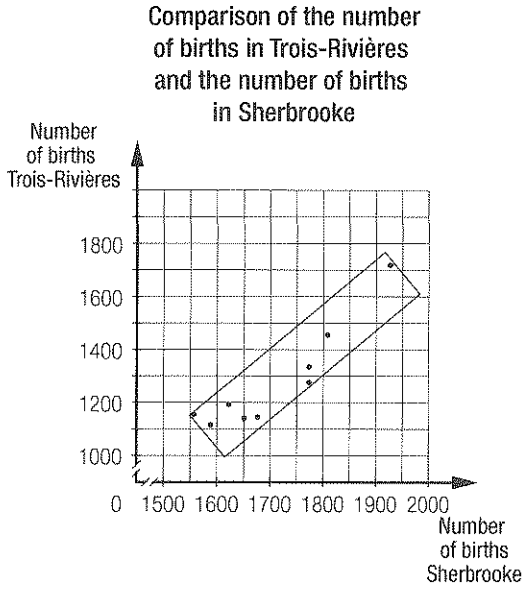
The answers can vary depending on the method used to find the equation for the regression line. The correlation coefficient is about -0.66, which indicates a weak-to-moderate correlation.

Additionally, based on the appearance of the scatter plot, a linear model does not seem the most appropriate because a curve is clearly visible. The estimation is therefore not very reliable.

Snapshot 2 (*cont'd*) — Page 49

4. By creating the scatter plot and calculating the correlation coefficients we find 0.76 for the relation between the number of births in Sherbrooke and Saguenay, and 0.8 for the relation between the number of births in Sherbrooke and Trois-Rivières. Joachim was therefore right, because the linear correlations are moderate-strong. The predictions can therefore be sufficiently reliable, especially for the relation between the number of births in Sherbrooke and the number of births in Trois-Rivières. These predictions, however, allow for a certain margin of error.





Snapshot 2 (*cont'd*) **Page 50**

5.

```
RegLin
 y=ax+b
 a=.6394984326
 b=-9.535632184
```

The equation of the regression line is: $y = 0.64x - 9.54$. If the average daily temperature is 25°C, the store will sell an average of 6.46 pools per day, or about 194 pools during the course of the 30 days of the month of June. Since the correlation coefficient between the variables is about 0.88, indicating a strong correlation, this prediction will probably be close to what occurs.

Snapshot 2 (*cont'd*) **Page 51**

6. The equation for the regression line using all available values is: $y = 7.12x - 14{,}229.35$. Sales predicted for 2030 amount to $224.25 G. Answers may vary according to the method used to find the equation for the regression line.

   This prediction is inevitably better than those of the two statisticians, because those of the latter only take into account half of the available values and do not consider the effect of the other values on general trends.