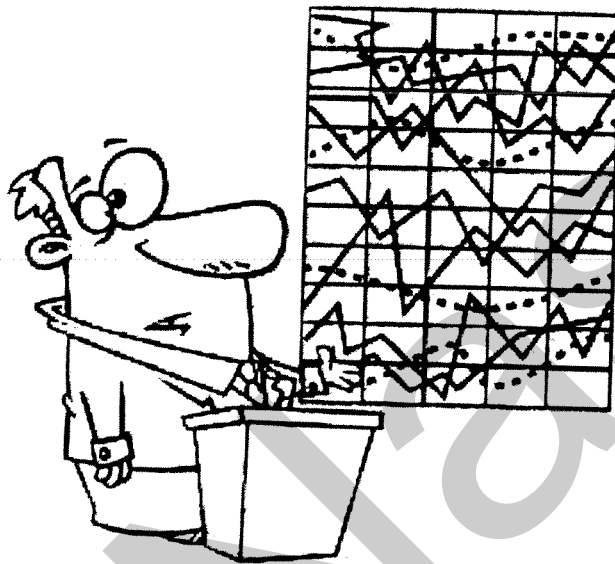


# Solution Key

## STATISTICS BOOKLET SECONDARY 4

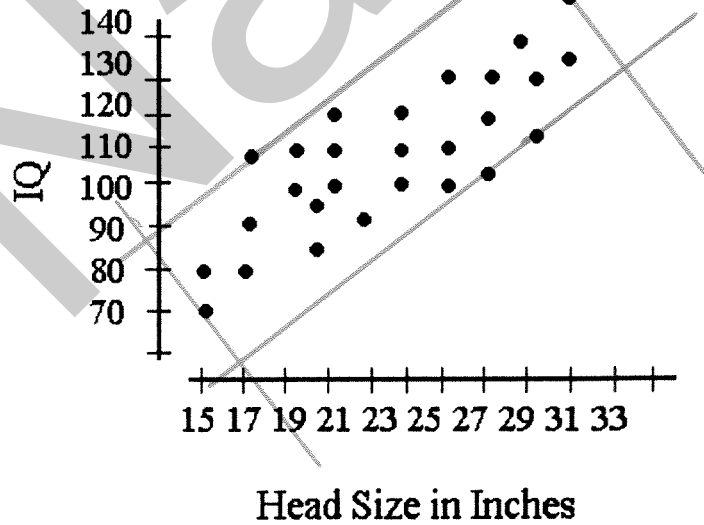
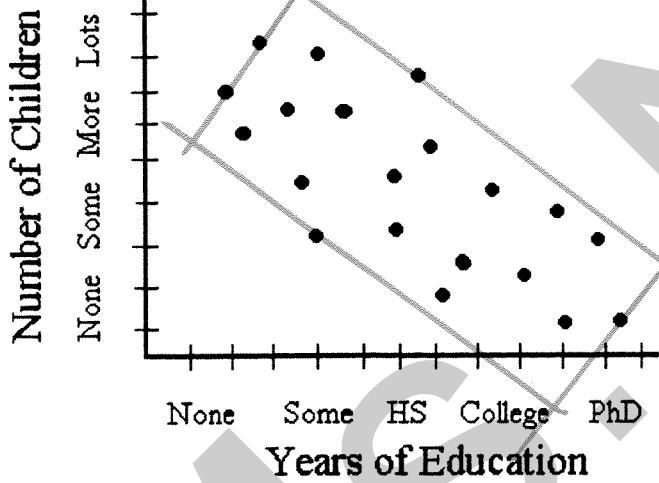
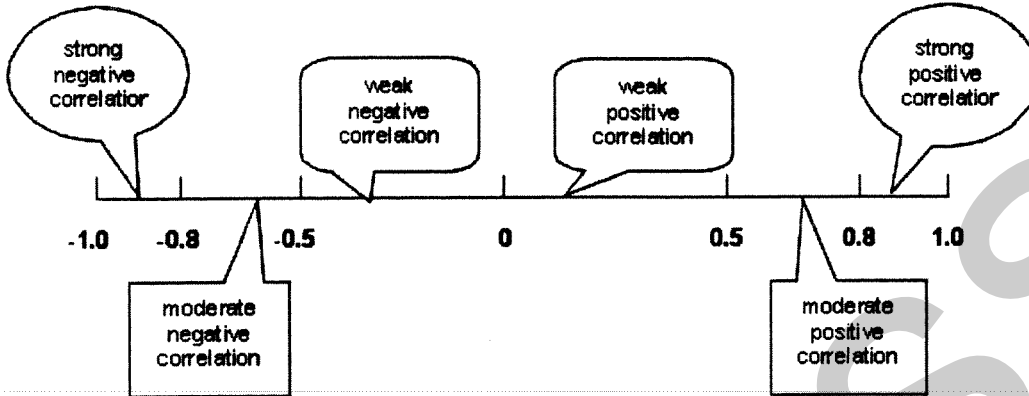


**MS. NASSIF**

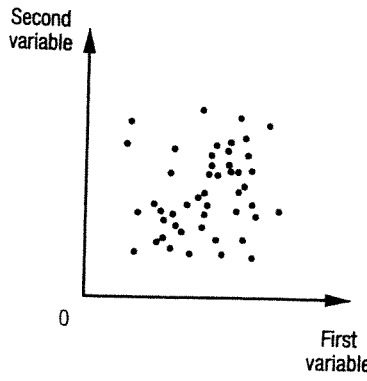
**CORRELATION COEFFICIENT  
MEAN DEVIATION  
REGRESSION LINE  
PERCENTILE  
CONTINGENCY TABLE**

# LINEAR CORRELATION COEFFICIENT

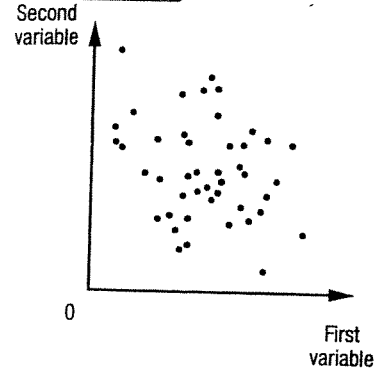
$$-1 \leq r \leq 1$$



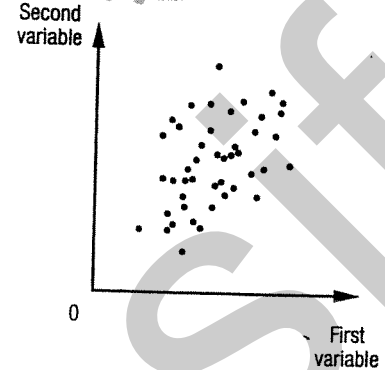
ZERO Correlation



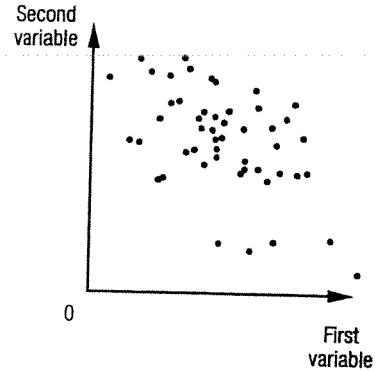
NEGATIVE WEAK Correlation



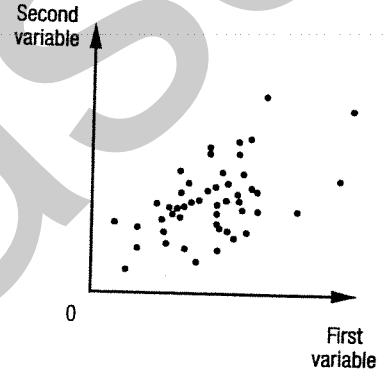
POSITIVE WEAK Correlation



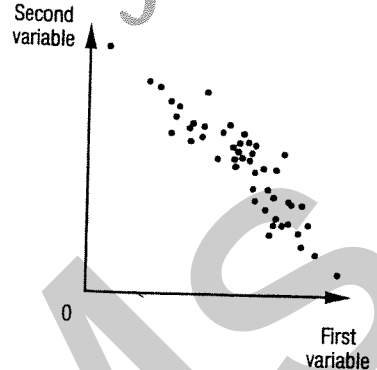
Moderate Correlation



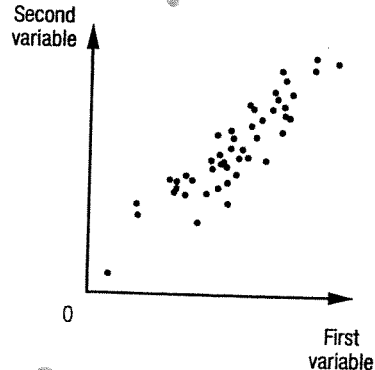
Moderate Correlation



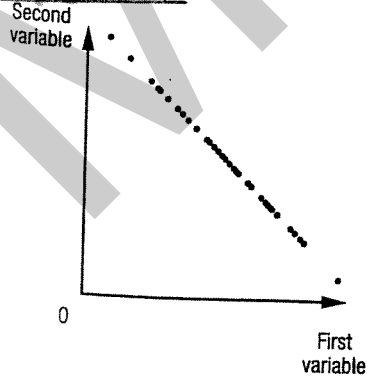
Strong Correlation



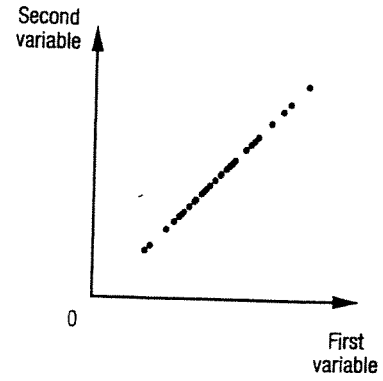
STRONG Correlation



Perfect Correlation



PERFECT Correlation

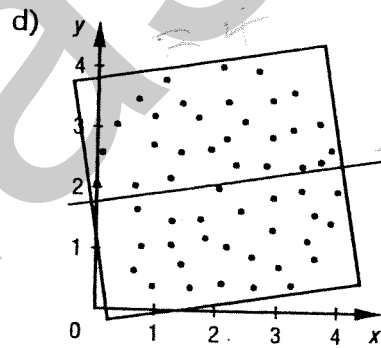
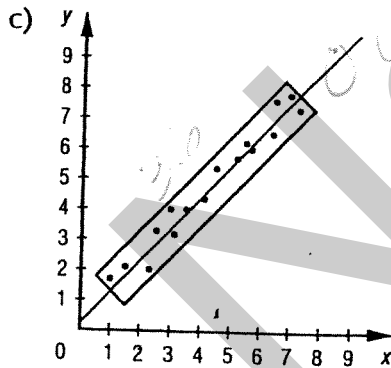
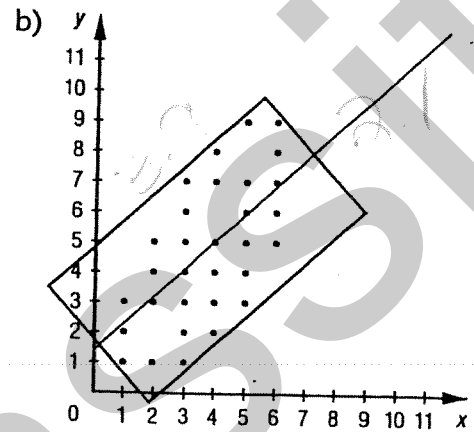
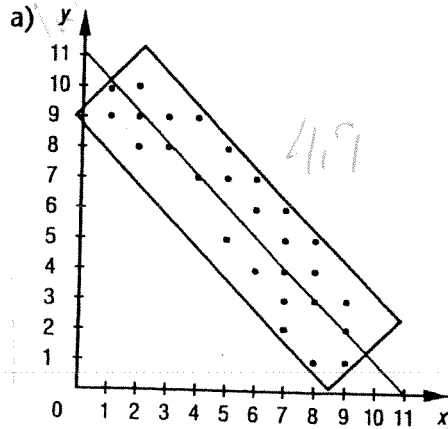


MS. Nassif

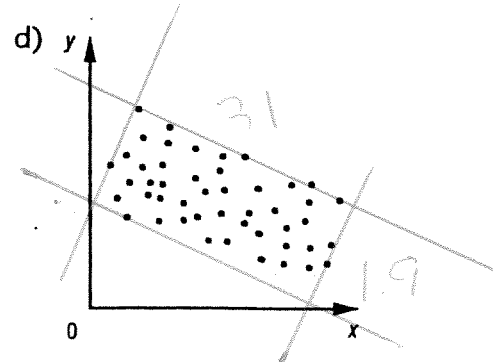
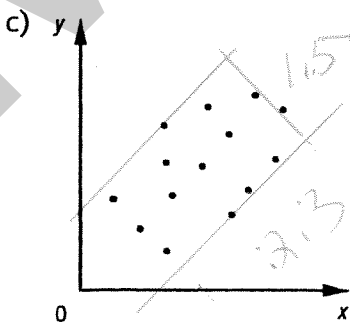
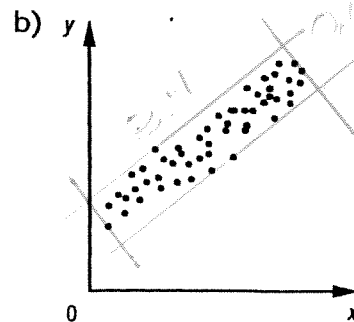
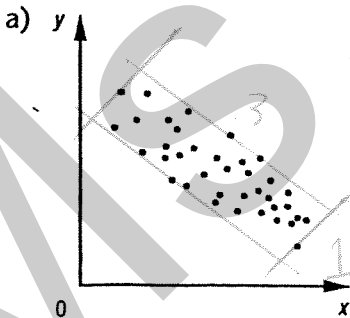
# LINEAR CORR. COEFFICIENT

CALCULATE THE LINEAR CORRELATION COEFFICIENT.

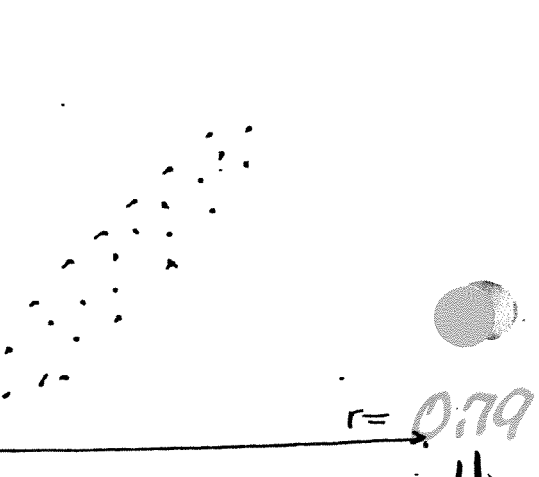
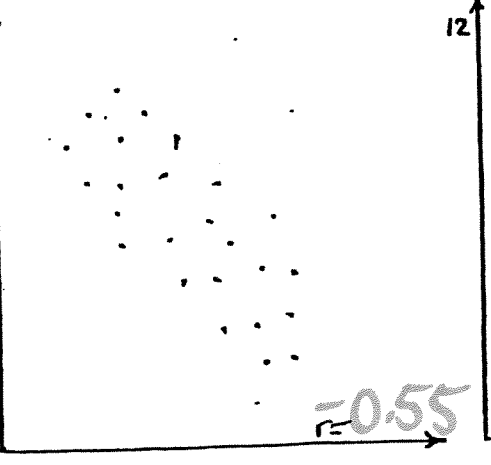
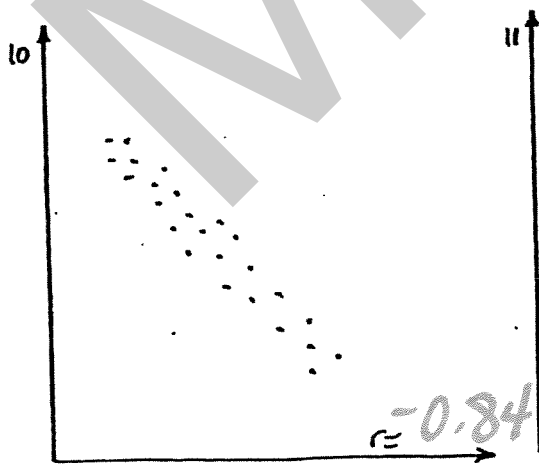
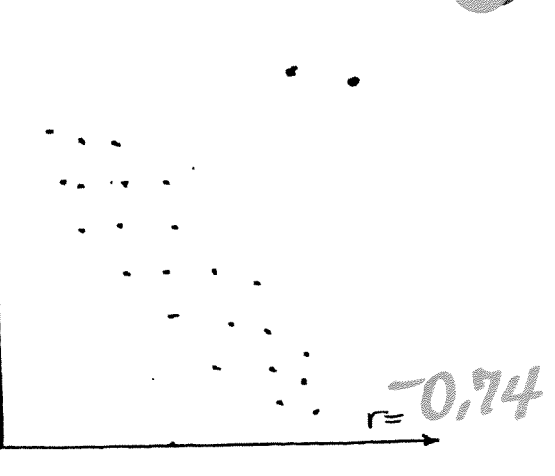
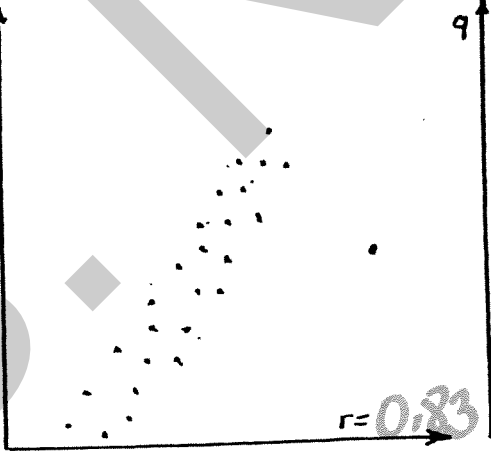
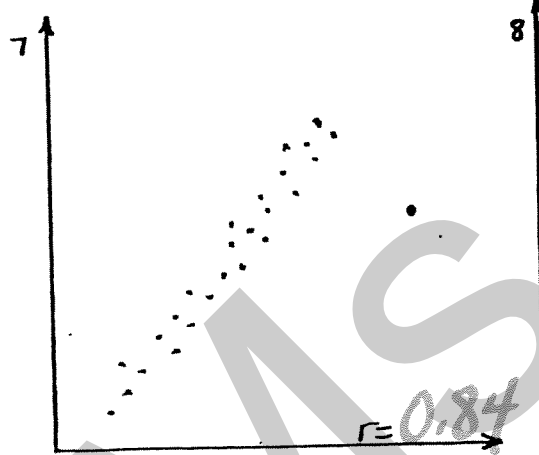
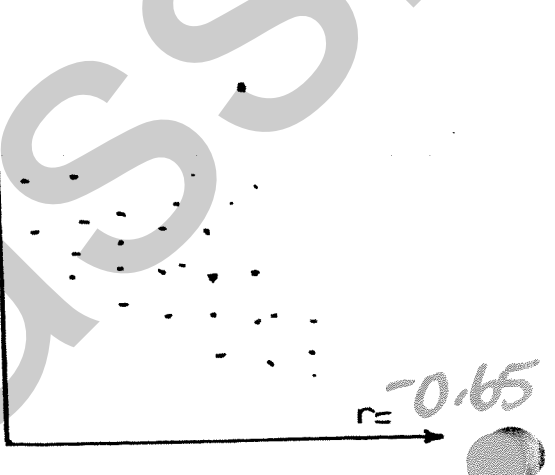
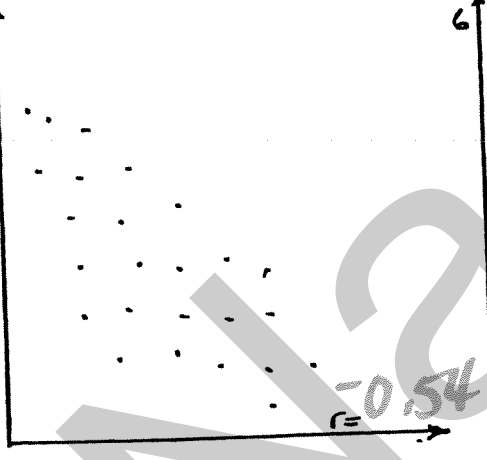
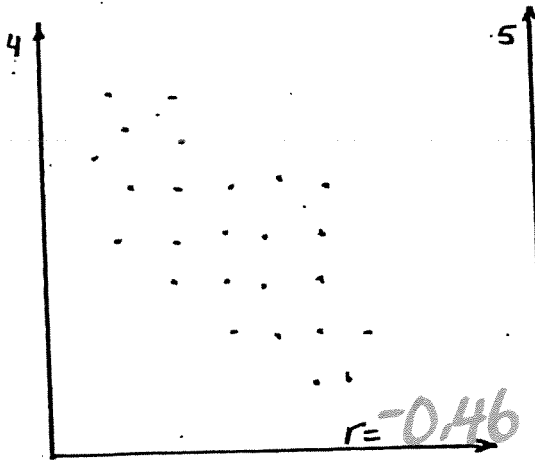
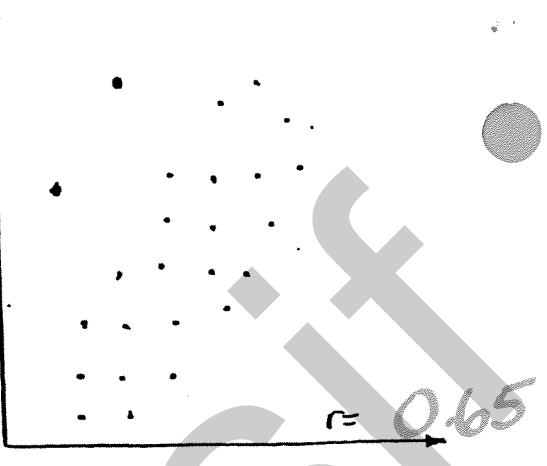
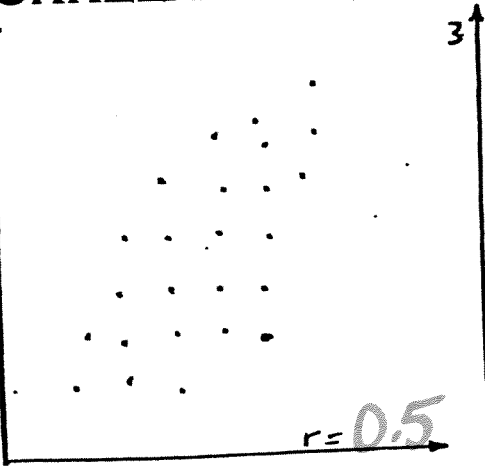
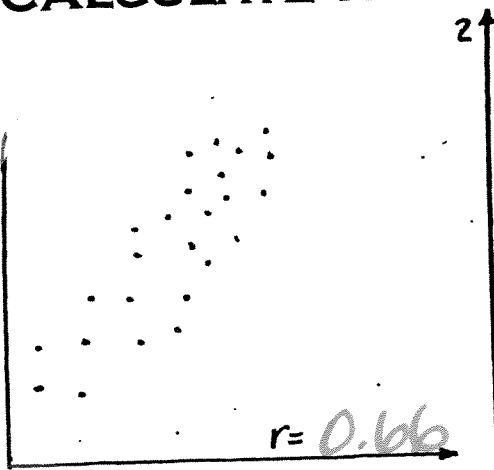
1)



2)



CALCULATE THE CORRELATION COEFFICIENT. 2 DECIMALS



## Contingency Table - Two Variable Distribution

When determining the correlation, we determine:

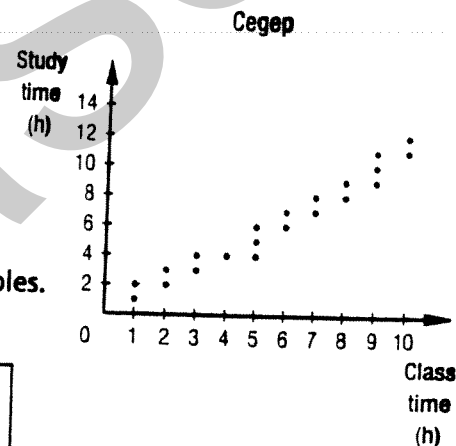
1. the type ( linear, exponential, etc)
2. the direction (positive - as x increases, y increases, negative - as x increases, y decreases)
3. the intensity (zero, weak, moderate, strong, perfect)

### Contingency Table

1. One variable (x) goes along the first column and the other variable (y) goes along the first row.
2. Data can be grouped in classes.
3. There is a correlation when the highest numbers occur on the diagonals of the table.

Example:

The adjacent scatter plot shows the relationship between the total amount of class time in CEGEP courses and the corresponding amount of study time.



Describe the correlation between these two variables.

		Cegep						Total
		Number of hours of class time						
Number of hours of study	Number of hours of class time	[0, 2[	[2, 4[	[4, 6[	[6, 8[	[8, 10[	[10, 12[	
	[0, 2[	1	0	0	0	0	0	1
	[2, 4[	1	3	0	0	0	0	4
	[4, 6[	0	1	3	0	0	0	4
	[6, 8[	0	0	1	3	0	0	4
	[8, 10[	0	0	0	1	3	0	4
	[10, 12[	0	0	0	0	2	1	3
[12, 14[	0	0	0	0	0	1	1	
<b>Total</b>		2	4	4	4	5	2	21

Correlation coefficient		Meaning
Negative	Positive	
Near 0	Near 0	Indicates a <b>zero</b> linear correlation between the two variables
Near -0.5	Near 0.5	Indicates a <b>weak</b> linear correlation between the two variables
Near -0.75	Near 0.75	Indicates a <b>moderate</b> linear correlation between the two variables
Near -0.87	Near 0.87	Indicates a <b>strong</b> linear correlation between the two variables
Equal to -1	Equal to 1	Indicates a <b>perfect</b> linear correlation between the two variables

For each contingency table, state if the linear correlation is

1 - positive or negative, 2- Zero, weak, moderate, strong or perfect.

a) Players on a football team

Age \ Mass	15	16	17	18	19	Total
[40, 50[	6	1	0	0	0	7
[50, 60[	2	3	2	0	0	7
[60, 70[	1	5	6	1	1	14
[70, 80[	0	0	2	8	2	12
[80, 90[	0	0	1	1	3	5
Total	9	9	11	10	6	45

b) Swimmers on a swim team

Age \ Mass	5	16	17	18	19	Total
[40, 50[	6	5	3	2	1	17
[50, 60[	4	4	3	2	2	15
[60, 70[	2	7	3	3	3	18
[70, 80[	0	6	2	2	2	12
[80, 90[	0	0	0	1	1	2
Total	12	22	11	10	9	64

c) Secondary cycle 2

Length of pool (cm) \ Height (cm)	[120, 140[	[140, 150[	[150, 160[	[180, 200[	Total
[15, 20[	12	3	0	0	15
[20, 25[	5	12	3	0	20
[25, 30[	0	6	10	4	20
[30, 35[	0	2	7	6	15
Total	17	23	20	10	70

d) 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

e) 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

f) 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

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



NAME: \_\_\_\_\_

**CONTINGENCY TABLE**  
Two variable Distribution

1)

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

Start of the 2007-2008 season

Player	Games played	Points scored
Alex Kovalev	44	41
Tomas Plekanec	44	38
Saku Koivu	43	33
Christophe 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 Brisson	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

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.  
\_\_\_\_\_  
\_\_\_\_\_

### An experimental drug

3.) 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:

Drop in attacks (%)	Anti-seizure medication						Total
	[0, 100[	[100, 200[	[200, 300[	[300, 400[	[400, 500[	[500, 600[	
[0, 14[	7	2	0	0	0	0	19
[14, 28[	22	8	0	0	0	0	30
[28, 42[	3	12	3	0	0	0	18
[42, 56[	0	18	7	5	0	0	30
[56, 70[	0	4	9	8	0	1	22
[70, 84[	0	0	0	12	8	4	24
[84, 98[	0	0	0	0	0	7	7
Total	32	44	19	15	8	7	140

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?

$$\begin{array}{r} 140 \\ \hline 95 \\ \hline 24 + 7 = 31 \end{array}$$

$$\begin{array}{r} 8 \\ \hline 7 + 2 = 9 \text{ at most} \end{array}$$

## Qualitative interpretations of the correlation

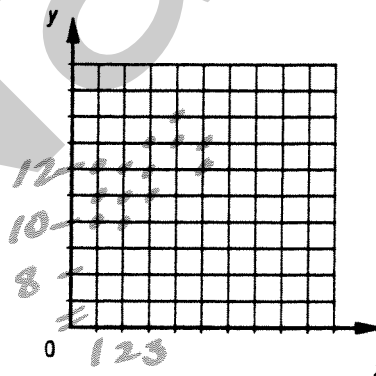
**1** Here are the ordered pairs of a two-variable distribution:

(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.

y \ x	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

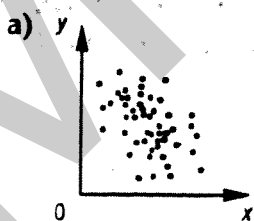
b) Construct a scatter plot representing this distribution.



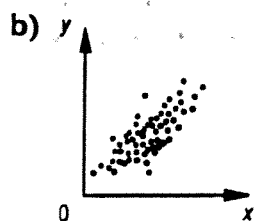
c) Describe the direction and strength of the linear correlation of the two variables of this distribution.

+ and moderate

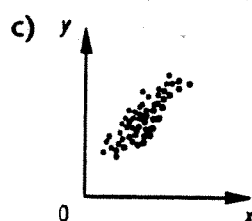
**2** Describe the direction and the strength of the linear correlation between variables of the scatter plots below.



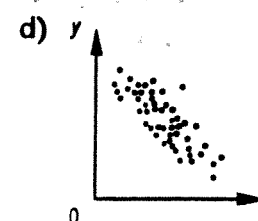
-  
weak



+  
mod



+  
Strong



-  
moderate

# practice



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

Legal age of marriage

Woman's age \ Man's age	16	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

In Québec the legal age for marriage is 16. However, people under 18 years old must obtain their parent's permission first.



- 2 The contingency table below shows the mother tongue of three groups of high school students.

High school

Mother tongue \ Group	French	English	Spanish	Creole	Other	Total
①	15	6	3	6	4	34
②	12	10	8	2	1	33
③	10	13	2	4	3	32
Total	37	29	13	12	8	99

Given this data, find:

- the size of the sample  $99$
- the percentage of Francophones in this sample  $\frac{37}{99} \times 100 = 37\%$
- the percentage of Anglophones in Group ②  $\frac{10}{33} \times 100 = 30\%$
- the percentage of people in the sample who are neither Francophones nor Anglophones  $\frac{33 \times 100}{99} = 33\%$

3 Complete the contingency tables below.

a) Distribution of students at a CEGEP

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

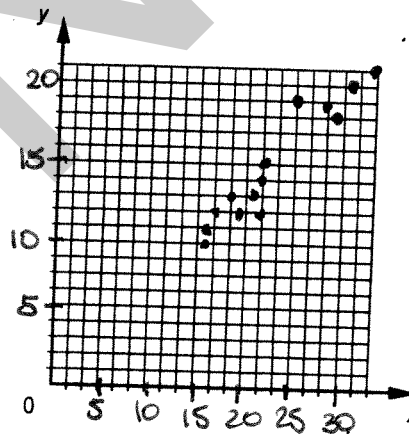
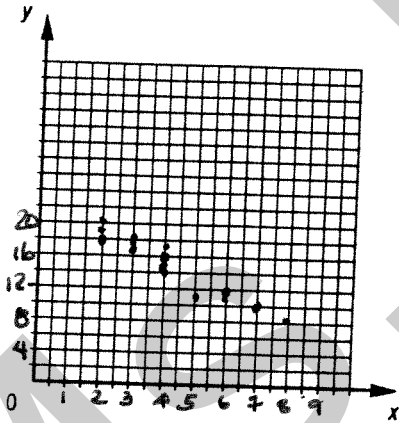
b) Catch distribution by fisherman age

Number of catches	Age				Total
	15	16	17	18	
[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) 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)



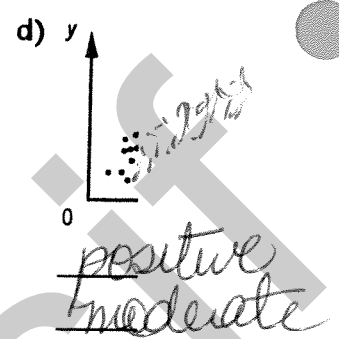
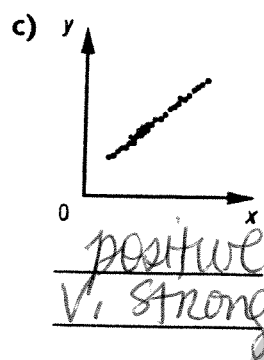
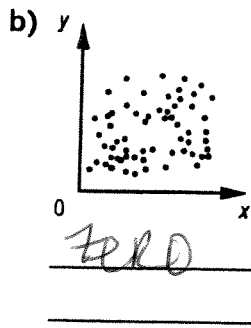
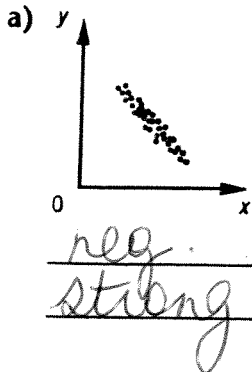
b) Describe the direction and the strength of the linear correlation between the variables of these two scatter plots.

1) negative  
strong

2) positive  
strong

## Qualitative interpretations of correlation

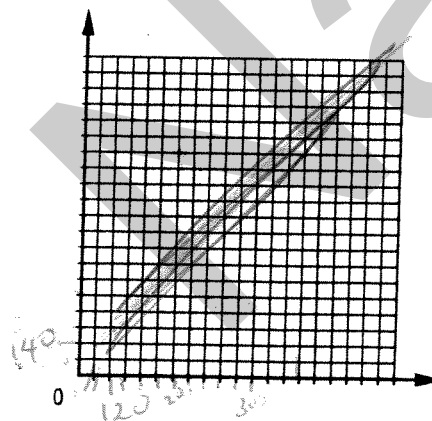
- 1** Describe the linear correlation between the variables in each of the scatter plots below.



- 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 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.

positive and very strong

# practice 5.4

NAME: Solution

Key



## Regression line

**1** For each of the following tables of values, do the following:

- ~~1~~ Calculate the linear correlation coefficient.
- 2) Find the equation of the regression line.

$Y = 2.76x - 2.74$

x	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	2.5	2.6	2.7
y	0	0.8	1	0.5	1.5	2	1.3	1	2	1.4	2.5	4	3.4	3	5	4.3	4	4.7

$Y = -1.67x + 8.8$

x	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	2.5	2.6	2.7	2.8
y	7	7.3	6.5	6.7	6.3	5.8	6	5.6	5.7	5.5	5	4.8	5.3	5	4.3	4.6	4	4.4

$Y = -0.89x + 5.8$

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	3.8	3.9	4
y	4	3.2	3.7	3	4.3	3.2	3.5	2.4	4	1.7	3.4	2.7	3	2	2.5	3	2.6	3

$Y = 1.44x + 0.02$

x	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
y	4.6	3.7	5.7	5	5.9	4.5	5.6	6.1	5.5	4.8	6.7	7	7	6.4	6	7	6.9	6.3

**2** The table below shows the relation between the age of a tree and the circumference of its trunk.

Growth of a tree

Age	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Circumference (cm)	10	16	26	32	44	50	68	80	90	106	120	124	130	150	166

- ~~a~~ What is the linear correlation coefficient of this situation?
- b) Determine the equation of the regression line associated with this situation.
- c) What should be the circumference of the trunk of a 35 year-old tree?
- d) What should be the age of the tree if the circumference of its trunk is 255 cm?

$Y = 11.44x - 10.7$

389.7 cm

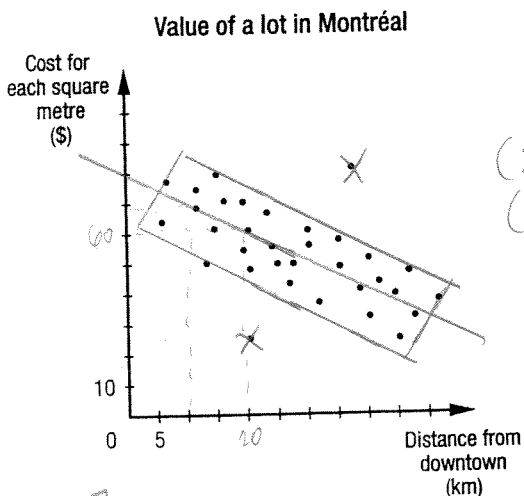
23.2 years old



The circumference of the trunk of a baobab tree can reach 38 m. Its spongy trunk thickens during the rainy season and thins during the dry season making it difficult to assess its age. It is estimated, however, that some baobab trees could be nearly 2000 years old.

3

The scatter plot below provides information on the value of lots in the Montréal area.



to do

- a) Describe the linear correlation between these two statistical variables. *moderate*
- b) Determine the equation of the regression line associated with this situation.  $y = -1x + 80$
- c) What might be the price of a lot in the heart of downtown Montréal? *80\$*
- d) Is the linear model appropriate to study the correlation between these two variables? Explain your answer. *Yes, strong correlation*



Québec's Far North is replete with natural resources, including gold, diamonds and various metals. A prospecting company conducts sample testing on a given territory. Following are the results of these tests:

*(23.4, 279.4)*

Mass of diamonds

Number of diamonds	Sample mass (kg)	Number of diamonds	Sample mass (kg)	Number of diamonds	Sample mass (kg)
25	320	13	175	10	90
30	405	32	340	53	610
45	500	21	250	110	1250
18	225	65	805	80	870
42	460	38	430	68	700

$a = \frac{462.7}{112.7}$

$y = 10.83x + b$

$742.1 = 10.83(\text{dail}) + b$   
 $b = 26.24$

*1 ton = 9000 kg*

If the company is planning to collect 1400 tons of samples, can it expect to collect a total of 200 000 diamonds? Explain your answer.

*Draw on a graph.*

$y = 10.83x + 26.24$

$1260000 = 10.83x + 26.2$

$x = 116341.07$

Kimberlite is the rock in which diamonds are formed.



*No he cannot expect 200000 diamonds.*

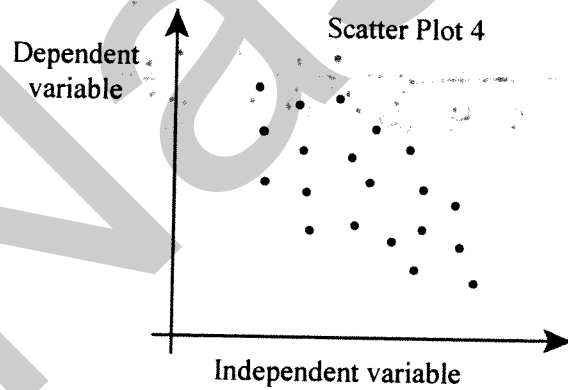
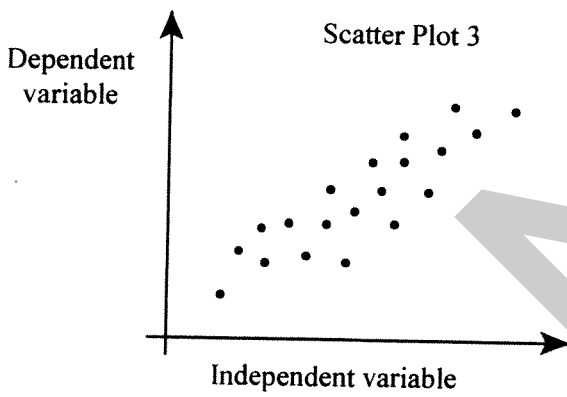
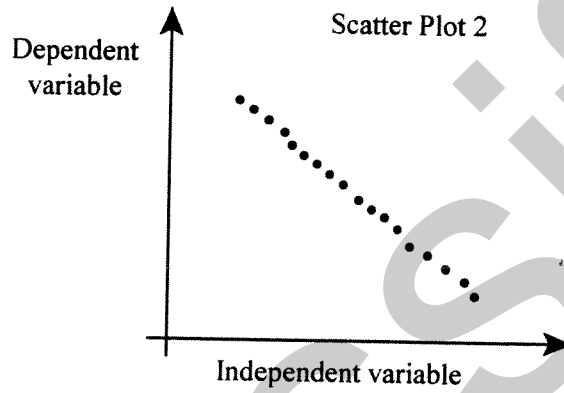
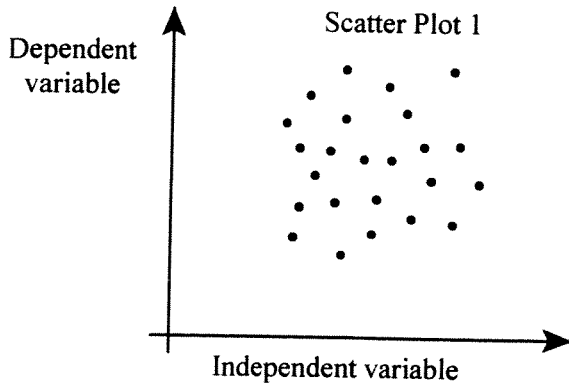


Statistics Assignment

Correlation Coefficient & Regression Line

1

Given the four scatter plots below.



In which of the following arrangements are the scatter plots ordered from highest to lowest correlation?

A) 3, 2, 4, 1

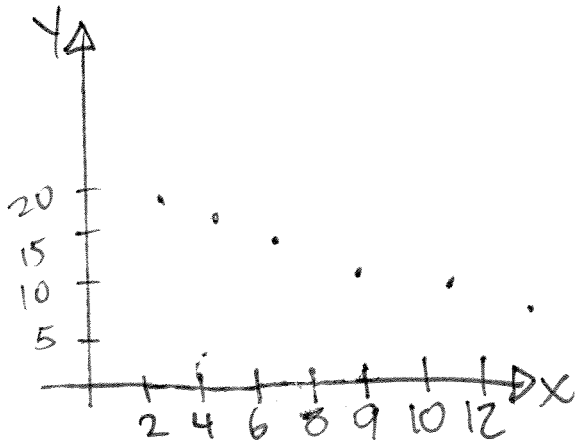
B) 2, 4, 3, 1

C) 2, 3, 4, 1

D) 1, 4, 3, 2

- 2 Two hockey players practise taking shots on net. They each take 10 shots. They back up 2 m and take another 10 shots each.

After practising taking shots this way, they analyzed the relation between the number of successful shots and their distance from the net. The results are given in the table on the right.



Distance from net (m)	Number of successful shots
3	19
5	17
7	14
9	12
11	10
13	7
15	5

Which of the following correctly describes the correlation between these two variables?

- A) Strong and negative
- B) Strong and positive
- C) Weak and negative
- D) Weak and positive

- 3 A teacher asked his students how much time they had spent preparing for an exam. For each student, the teacher matched the time declared with the number of wrong answers the student had on the exam.

The data collected is presented in the table below.

Number of Wrong Answers on an Exam According to the Amount of Time Spent Studying			
Time (min)	Number of wrong answers	Time (min)	Number of wrong answers
20	25	45	14
25	27	50	18
30	17	55	10
30	21	55	14
35	18	60	15
35	22	70	5
40	22	75	3
40	16	75	8

$(31.875, 21)$   $(60.625, 10.875)$   
 $a = \frac{-10.125}{28.75} = -0.35$   
 $y = -0.35x + b$   
 $21 = -0.35(31.875) + b$   
 $b = 32.2$   
 $y = -0.35x + 32.2$   
 $(65)$   
 $= 9.45$

According to the data, how many wrong answers can be expected for a student who studies for 65 minutes?

- 4 Which of the following correlation coefficients represents the strongest correlation?

A) -0.8

C) 0.5

B) -0.3

D) 0.7

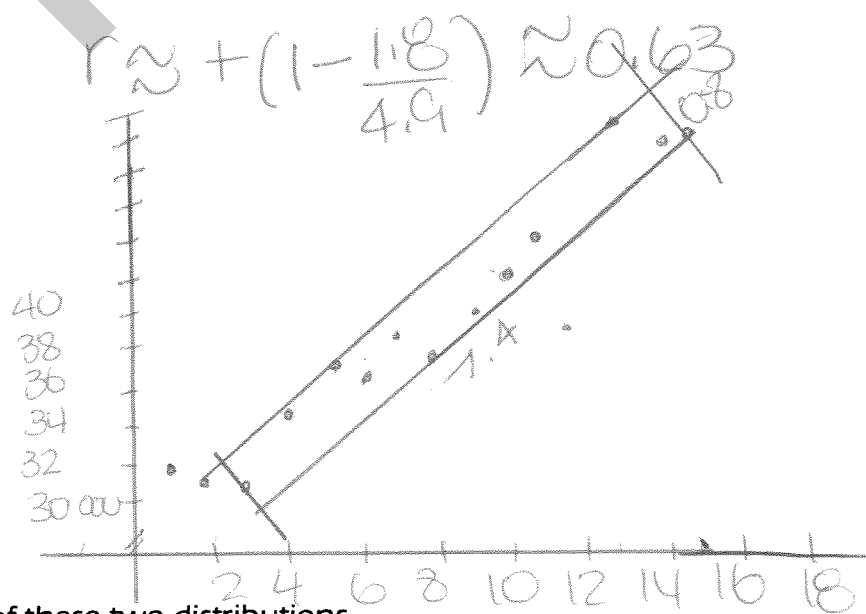
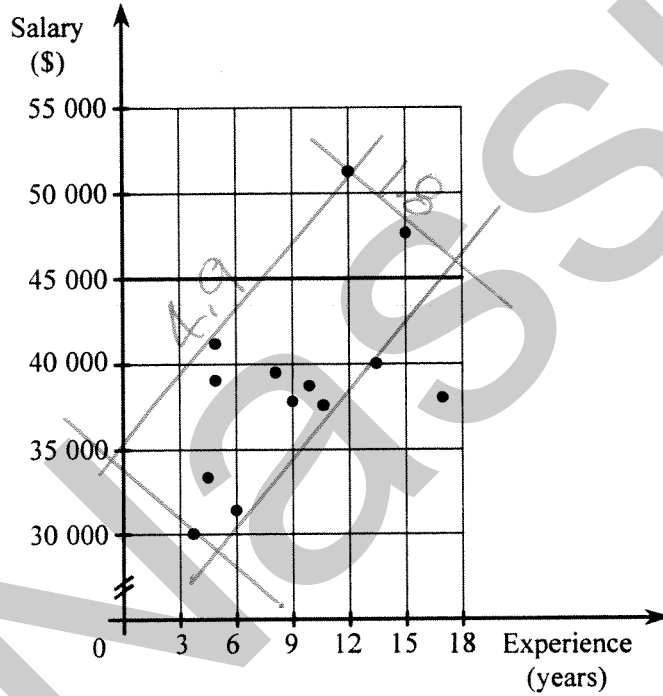
5

The number of years of experience and the corresponding salary of employees in two companies are given below.

Experience and Salary of Employees in Company A

Years of Experience	Salary in \$
1	32 000
2	31 000
3	31 500
4	34 000
5	37 000
6	36 000
7	38 500
8	36 500
9	40 500
10	42 000
11	43 000
12	38 500
13	51 000
14	49 000
15	48 000

Experience and Salary of Employees in Company B



Compare the coefficients of correlation of these two distributions.

8 There are 19 apple trees in Pauline's orchard.

The age and production of each tree is shown below.

This year, Pauline will plant one new apple tree of the same type.

How many crates of apples will this tree produce in its tenth year?

Age (years)	Production (crates)
5	3
6	2
7	2
7	4
12	5
12	8
13	6
13	8
14	7
15	8
16	7
16	9
17	8
17	10
18	9
18	10
19	9
19	12
20	13

$y = 0.55x - 0.36$

5.05

9 The coefficient of linear correlation between two variables is -1.

Which of the following statements is true?

- A) There is no correlation.
- B) There is a very low correlation, with one variable increasing as the other increases.
- C) There is a very low correlation, with one variable decreasing as the other increases.
- D) There is a perfect correlation, with one variable decreasing as the other increases.

11

A school held a track and field competition. The table below lists the height jumped by 14 competitors in the high jump event, according to the students' age.

Height Jumped According to Age

Age (years)	Height (cm)
8	85
8	90
9	90
9	100
9	105
10	95
10	110

Age (years)	Height (cm)
16	150
16	160
16	170
17	165
17	170
17	180
18	180

Given the above results, how high could a 13-year-old child be expected to jump?

Show all your work.

133.6

12

Which one of the following correlation coefficients represents the lowest correlation?

- A) -0.6
- B) -0.2

- C) 0.1
- D) 0.8

13 Over the course of a given week, 39 members of a reading club went to the library.

The table below shows the ages of these 39 members and the number of times they visited the library that week.

Age (years)	Number of visits				
	1	2	3	4	5
13	0	0	2	2	6
14	0	1	2	4	2
15	1	4	4	1	0
16	6	3	1	0	0

Which one of the following statements best describes the linear correlation between the ages of the members and the number of library visits?

- A) The correlation is positive and high.
- B) The correlation is positive and low.
- C) The correlation is negative and high.
- D) The correlation is negative and low.

16

Michael is studying the linear correlation between the age and the diameter of the trunk of birch and maple trees found in a forest.

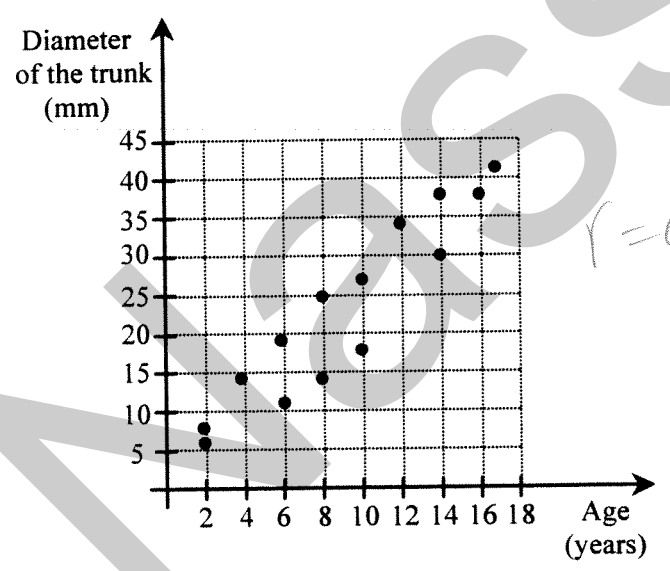
The distributions represented by the following table of values and scatter plot show Michael's data.

Distribution 1

BIRCH TREES	
Age (years)	Diameter of the trunk (mm)
2	10
16	25
12	20
10	30
6	25
6	5
10	15
8	10
17	35
16	45
4	5
14	20
12	40
8	30

Distribution 2

MAPLE TREES



$r = 0.55$



Compare the coefficient of linear correlation between the variables of these two distributions.

- 20 A researcher conducted a study of 20 young people between the ages of 11 and 21 to determine how many hours of exercise they did per week. The following table shows the data collected. Describe the linear correlation between age and the number of hours of exercise per week for these 20 young people.

Hours \ Age	Hours				
	[0, 2[	[2, 4[	[4, 6[	[6, 8[	[8, 10[
[11, 13[	0	0	1	1	2
[13, 15[	1	1	0	1	1
[15, 17[	2	1	0	1	1
[17, 19[	1	0	1	1	0
[19, 21[	2	1	1	0	0

*negative-weak*

- 23 The table below shows the distribution of 24 members at a health club, according to their age and the number of workouts they did last month. Describe the linear correlation between the age of the 24 members and the number of workouts they did last month.

Age (years)	Number of Workouts					
	[1, 5[	[5, 9[	[9, 13[	[13, 17[	[17, 21[	[21, 25[
[18, 25[	1	1	1	0	0	0
[25, 32[	1	1	1	0	1	0
[32, 39[	0	1	2	1	1	0
[39, 46[	0	1	1	1	1	1
[46, 53[	0	0	2	0	1	0
[53, 60[	0	0	1	1	1	1

*positive weak*

# Mean Deviation

Complete the table and then calculate the mean deviation.

1.

$\bar{x} = 81.25$   
 $MD = 9.8$

x		
78		
95		
65		
87		
Total		

2.

$\bar{x} = 62$   
 $MD = 26.8$

x		
90		
76		
23		
87		
34		
Total		

3.

$\bar{x} = 80.7$   
 $MD = 16.4$

x		
56		
87		
99		
Total		

4.

$\bar{x} = 68$   
 $MD = 18.4$

x		
44		
46		
76		
76		
98		
Total		

5.

$\bar{x} = 77$   
 $MD = 11$

x		
66		
88		
Total		

# Solution Key

## Statistics Secondary 4

### Mean Deviation

Calculate the mean deviation:

1.

No. phone calls (x)	Frequency (f)	fx	$x - \bar{x}$	$f (x - \bar{x}) $
5	4	20	-4	16
6	8	48	-3	24
7	12	84	-2	24
8	20	160	-1	20
9	22	198	0	0
10	21	210	1	21
11	14	154	2	28
12	9	108	3	27
Total	110	982		160

$$\bar{x} = \frac{982}{110} = 9$$

$$MD = \frac{160}{110} = 1.5$$

2.

No. of Pets (x)	Frequency (f)	fx	$x - \bar{x}$	$f (x - \bar{x}) $
0	8	0	-1.96	15.68
1	15	15	-0.96	14.4
2	11	22	0.04	0.44
3	7	21	1.04	7.28
4	5	20	2.04	10.2
5	4	20	3.04	12.16
Total	50	98		60.16

$$\bar{x} = \frac{98}{50} = 1.96$$

$$MD = \frac{60.2}{50} = 1.20$$

3.

Score (x)	Freq	m	fm	$m - \bar{x}$	$f (m - \bar{x}) $
[0 - 20[	10	10	100	-56	560
[20 - 40[	30	30	900	-36	1080
[40 - 60[	50	50	2500	-16	800
[60 - 80[	70	70	4900	4	280
[80 - 100[	90	90	8100	24	2160

$$\bar{x} = \frac{16500}{250} = 66$$

$$MD = \frac{4880}{250} = 19.52$$

4.

Weight(kg) $x$	Frequency	$fx$	$x - \bar{x}$	$f (x - \bar{x}) $
66	5			
67	8			
68	12			
69	20			
Total				

5. Find the mode, median, mean and range

a) 11, 12, 13, 15, 7, 6, 5, 10, 14, 8, 9.

b) 108, 108, 110, 101, 110, 111, 104, 104, 105, 107.

c) 4, 5, 5, 1, 1, 2, 8, 8, 9, 9, 8, 7, 2.

7.

Marks $x$	Frequency	$m$	$fm$	$m - \bar{x}$	$f (m - \bar{x}) $
[0 - 10[	0				
[10 - 20[	3				
[20 - 30[	10				
[30 - 40[	34				
[40 - 50[	42				
Total					

8.

No. of Pets ( $x$ )	Frequency ( $f$ )	$fx$	$x - \bar{x}$	$f (x - \bar{x}) $
0	8			
1	15			
2	11			
3	7			
Total				

9.

Wage(\$)	Freq	$m$	$fm$	$m - \bar{x}$	$f (m - \bar{x}) $
[2 - 7[	4.5				
[7 - 12[	9.5				
[12 - 15[	13.5				
[15 - 20[	17.5				
[20 - 30[	25				
Total					

4.

Weight(kg) x	Frequency	fx	$x - \bar{x}$	$f (x - \bar{x}) $
66	5	330	-2	10
67	8	536	-1	8
68	12	816	0	0
69	20	1380	+1	20
Total	45	3062		38

MD=0.84

5. Find the mode, median, mean and range

a) 11, 12, 13, 15, 7, 6, 5, 10, 14, 8, 9.

5 6 7 8 9 10 11 12 13 14 15

b) 108, 108, 110, 101, 110, 111, 104, 104, 105, 107.

101 104 104 105 107 108 108 108 108 110 111

c) 4, 5, 5, 1, 1, 2, 8, 8, 9, 9, 8, 7, 2.

1 1 2 2 4 5 5 7 8 8 8 9 9

7.

Mode Median Mean R

NONE 10 10 10

104, 108, 110 107.5 106.8 10

8 5 5.3 8

Marks x	Frequency	m	fm	$m - \bar{x}$	$f (m - \bar{x}) $
[0 - 10[	0	5	0	-33	0
[10 - 20[	3	15	45	-23	69
[20 - 30[	10	25	250	-13	130
[30 - 40[	34	35	1190	-3	102
[40 - 50[	42	45	1890	7	294
Total	89		3375		595

MD=6.7

8.

No. of Pets (x)	Frequency (f)	fx	$x - \bar{x}$	$f (x - \bar{x}) $
0	8	0	-1.4	11.2
1	15	15	-0.4	6
2	11	22	0.6	6.6
3	7	21	1.6	11.2
Total	41	58		35

MD=0.85

9.

Wage (\$)	Freq	m	fm	$m - \bar{x}$	$f (m - \bar{x}) $
[2 - 7[	4.5	4.5	20.25	-13	58.5
[7 - 12[	9.5	9.5	90.25	-8	76
[12 - 15[	13.5	13.5	182.25	-4	54
[15 - 20[	17.5	17.5	306.25	0	0
[20 - 30[	25	25	625	7.5	187.5
Total	70		1224		376

MD=5.4

8	20			
9	22			
10	21			
11	14			
12	9			
Total				

2.

No. of Pets (x)	Frequency (f)	fx	$x - \bar{x}$	$f (x - \bar{x}) $
0	8			
1	15			
2	11			
3	7			
4	5			
5	4			
Total				

3.

Score (x)	Freq	m	fm	$m - \bar{x}$	$f (m - \bar{x}) $
[0 - 20[	10				
[20 - 40[	30				
[40 - 60[	50				
[60 - 80[	70				
[80 - 100[	90				

12

Age of prisoners

Age	Frequency
18	1
20	3
21	5
25	3
26	2
28	5
34	3
42	2
54	1

MD =

13

Number of natural teeth	<i>f</i>
0	10
1	1
2	4
3	4
4	6
5	4
6	6
7	8
8	7

14

Class	<i>f</i>
[0, 10[	1
[10, 20[	3
[20, 30[	6
[30, 40[	15
[40, 50[	25
[50, 60[	30
[60, 70[	15
[70, 80[	8
[80, 90[	2

CALCULATE THE MEAN DEVIATION.

8

Bills of Lianne's customers

Class (\$)	$f$
[0, 8[	12
[8, 16[	20
[16, 24[	24
[24, 32[	10
[32, 40[	6
[40, 48[	4
[48, 56[	1

MD = 8.6

9

Class (hours)	$f$
[0, 4[	3
[4, 8[	6
[8, 12[	12
[12, 16[	4
[16, 20[	2
[20, 24[	3

10

Sale (thousands of \$)	$f$
[40, 60[	6
[60, 80[	14
[80, 100[	20
[100, 120[	30
[120, 140[	15
[140, 160[	10
[160, 180[	5

11

Age of prisoners

Class	$f$
[15, 25[	9
[25, 35[	13
[35, 45[	2
[45, 55[	1



# Mean Deviation Backwards

Finding two missing values

Find the two missing values in each list given the mean and mean deviation.

	List	Mean	Mean Deviation	
①	1, 2, 6	4	2	Both work (1, 4) (5, 6)
②	40, 60, 70, 70, 70, 70, 85, 95	70	12	(50, 90)
③	3, 7, 9, 11	8	3	5, 13
④	20, 60, 100	60	24	80, 40
⑤	Stem and Leaf 0   5 5 5 8 1   0 0 5 5 5 8 8 2   0 0 4 5 3   0 0 4   0	70 18	12 8	

$$\begin{aligned}
 &5+5+5+8+10+10+15+15+15+18 \\
 &+18+20+20+24+25+30+30+40 \\
 &+x+y = 70
 \end{aligned}$$

$x+y = 94$

Ms. Nassif.

# PERCENTILE

Solution Key

- What is the percentile of a data value in a distribution in which:
  - 77% of the data are less than or equal to this value? **77th**
  - 83% of the data are greater than this value? **17th**
- What is the percentile of a person who finishes a race 15th out of 100 participants? **86**
- What is the percentile of a person who is ranked:
  - 12th out of 20? **45**
  - 5th out of 40? **90**
  - 1st out of 10? **100th**
  - 20th out of 20? **5th**
- What is the percentile of the value in the box in these distributions?
  - 12, 18, 15, 16, 16, 18, 20, 24, 24, 25, 27, 32, 33, 35, 38, 40, 42, 42, 48, 50 **58**
  - 48, 50, 50, 50, 52, 55, 55, 56, 58, 59, 59, 60, 60, 60, 63, 64, 64, 65, 66, 66, 69, 70, 72, 72, 73, 74, 75, 78, 78, 79, 79, 80, 82, 83, 83, 84, 84, 85, 86, 88, 89, 90, 92, 98, 99 **63**
- What is the percentile of the value in the box in these distributions?

- $$\underbrace{132, 133, 134, 134, 134, \dots, 169}_{340 \text{ values}}, \span style="border: 1px solid black; padding: 0 2px;">169, \underbrace{172, 172, 172, 175, \dots, 200}_{155 \text{ values}}$$

**69**
- $$\underbrace{45, 48, \dots, 97}_{620 \text{ values}}, \span style="border: 1px solid black; padding: 0 2px;">97, \underbrace{98, 101, \dots, 358}_{833 \text{ values}}$$

$$P = \frac{627}{1461} \times 100 = 43^{\text{rd}}$$

10. Yannick is a talented young tennis player. Among the 300 up-and-comers in his age group, he ranks 118th.

- What is his percentile in this group?  $P = \frac{300-118+1}{300} \times 100 = 61$
- What is the percentile of the player immediately following him?

$$P = \frac{300-119+1}{300} \times 100 = 61$$

## 12. MATH COMPETITION

Maki places 1200th in the Putnam Math Competition, which brings together more than 6000 university students from all over North America. Her younger brother enters the Québec Association of Mathematics Teachers competition and finishes 50th among 300 candidates.

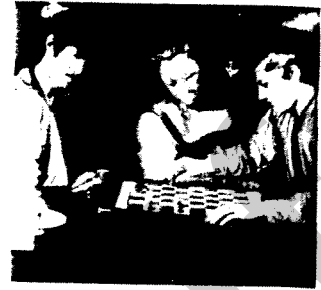
- Calculate each student's percentile.
  - Who was ranked higher within his or her own group? **Brother**
- ~~A~~ In the Putnam Math Competition, what was the position of the person in the 40th percentile?

$$\text{Maki} = 81^{\text{st}} \quad \text{Brother} = 84^{\text{th}}$$

13

At a chess tournament, Julio ranked 10th out of a total of 94 participants. What was the percentile of the person just ahead of him and of the person just after him?

Throughout the centuries, many artists have painted scenes of people playing chess. This oil canvas, a work done by the Italian artist Paolo Veronese, was painted in Venice in 1610.



Ahead =  $\frac{94-9+1}{94} \times 100 = \frac{86}{94} \times 100 = 92^{th}$  After =  $\frac{94-11+1}{94} \times 100 = \frac{84}{94} \times 100 = 90^{th}$

14

For each of the following distributions, determine the percentile for the data value identified by a red box.

- a) 23, 24, 25, 26, 27, 28, 31, 34, 36, 38, 38, 38, 38, 42, 42, 45, 46, 47, 47, 47, 48, 50, 51, 53, 54, 57, 58, 67, 67, 68, 69, 70, 71, 71, 73, 74, 79, 80, 80 40
  - b) 120, 120, 123, 130, 135, 135, 138, 141, 142, 144, 145, 146, 147, 147, 153, 153, 155, 160, 172, 172, 173, 174, 175, 176, 177, 178, 180, 185, 187 52<sup>th</sup>
  - c) 12, 15, 18, 18, ..., 321, 322, 328, ..., 841, 844, 849  $P = \frac{236}{671} \times 100 = 36^{th}$
- 230 values      431 values

15.

The data values below are the results from a statistical study.

34, 35, 36, 37, 38, 45, 46, 47, 48, 49, 49, 49, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 59, 59, 65, 66, 67, 68, 69, 70, 72, 76

- a) What data value corresponds to the 86th percentile?
- b) What is the percentile of the data value 49?  $P = \frac{13}{32} \times 100 = 35^{th} \text{ perc.}$

16. In order to become future air traffic controllers, candidates are subjected to a three step hiring process. After each step, a certain number of candidates are chosen to move on to the next step. Following are the test results, in percentages:

**Step 1: The interview**

86 70, 74, 76, 76, 80, 80, 80, 81, 82, 82, 84, 84, 84, 84, 86, 87, 88, 88, 88, 88, 88, 88, 89, 89, 90, 90, 91, 92, 92, 93, 94, 94, 97, 100  $P = \frac{32-5+1}{32} \times 100$

**Step 2: The written exam**

86 60, 60, 65, 68, 71, 72, 72, 74, 75, 76, 78, 79, 80, 82, 83, 83, 83, 84, 84, 84, 85, 88, 89, 90  $P = \frac{24-4+1}{24} \times 100$

**Step 3: The practical exam**

75 65, 65, 65, 65, 66, 67, 69, 70, 70, 72, 77, 78, 80, 80, 80, 82, 85, 87  $P = \frac{18-4+1}{18} \times 100$

The boxed are the results of a student who needs to be in the 85th percentile and better to be selected. Was he selected?

NO,

32

# Percentile Backwards

## Statistics

1.) In the following list: 20, 25, 26, 30, 30, 45, 55, 57, find the value in the

- 26 a. 38<sup>th</sup> percentile  
 30 b. 63<sup>rd</sup> percentile  
 55 c. 88<sup>th</sup> percentile

- 20 d. 13<sup>th</sup> percentile  
 57 e. 100<sup>th</sup> percentile  
 25 f. 25<sup>th</sup> percentile

2.) In the following list: 10, 12, 12, 15, 16, 18, 19, 19, 20, 20, 20, 20, 22, 25, 30, 31, find the value in the

- 16 a. 32<sup>nd</sup> percentile  
 20 b. 63<sup>rd</sup> percentile  
 22 c. 82<sup>nd</sup> percentile  
 10 d. 7<sup>th</sup> percentile

- 19 e. 50<sup>th</sup> percentile  
 30 f. 94<sup>th</sup> percentile  
 15 g. 25<sup>th</sup> percentile  
 20 h. 69<sup>th</sup> percentile

3.) In the following list: 101, 112, 112, 115, 116, 118, 119, 119, 220, 220, 220, 220, 222, 225, 300, 311, 312, 315, 318, 319, 320, 325, 327, 328, find the value in the

- 220 a. 50<sup>th</sup> percentile  
 220 b. 46<sup>th</sup> percentile  
 112 c. 9<sup>th</sup> percentile  
 327 d. 96<sup>th</sup> percentile  
 320 e. 88<sup>th</sup> percentile

- 318 f. 80<sup>th</sup> percentile  
 116 g. 21<sup>st</sup> percentile  
 220 h. 38<sup>th</sup> percentile  
 328 i. 100<sup>th</sup> percentile  
 101 j. 5<sup>th</sup> percentile

4.) The height, in centimetres, of a group of elementary-school children is represented below. What is the percentile rank of a student who is 123 cm tall?

112, 115, 115, ..., 122	123, 123, 124, ..., 126	127, 129, 130, ..., 132
80 students	70 students	73 students

5.) This stem-and-leaf plot shows George's scores from a video game. Find the value in the:

- 202 a. 22<sup>nd</sup> percentile  
 224 b. 72<sup>nd</sup> percentile  
 240 c. 93<sup>rd</sup> percentile  
 226 d. 79<sup>th</sup> percentile  
 221 e. 58<sup>th</sup> percentile  
 192 f. 8<sup>th</sup> percentile

**Game Scores**

Stem	Leaves
19	2 6
20	2 8
21	2 5 6
22	1 3 4 6
23	2
24	0
27	0 Best Score

$P = \frac{82}{223} \times 100$   
 $= 36.7 \times 3744$

# STATISTICS REVIEW

- LINEAR CORRELATION COEFFICIENT
- MEAN DEVIATION
- REGRESSION LINE
- PERCENTILE

1. In what percentile is someone who finished:
- a) 18th out of 32? 47
- b) 27th out of 64? 60
- c) 11th out of 43? 77
- d) 19th out of 21? 15

2. A distribution comprises 112 values. What is the percentile of the 43rd value in this distribution if three other values are equal to it? 63

3. Here is the distribution of the marks Lionel's students received on the last French test.

Mark	Frequency
10	2
9	6
8	11
7	16
6	12
5	5
4	2
3	1

- Which marks will be assigned to the first and second quintiles? \_\_\_\_\_
- In which quintile is a mark of 6? \_\_\_\_\_
- c) In which percentile is a mark of 6? 37
- d) Find the percentile of a mark of 4. 6

4. Michael finished 8th out of 111 in a men's cross-country race. Nathalie finished 13th out of 141 in the women's event. 55

- a) Find each runner's percentile.  
Mike: 94 Nathalie 92
- b) Which of the two performed better relative to his or her group? Mike

5 At a chess tournament, Julio ranked 10th out of a total of 94 participants. What was the percentile of the person just ahead of him and of the person just after him?

$$\frac{85}{94} = 91\text{st}$$

person ahead 9th → 92  
 person behind 11th → 90

6 For each of the following distributions, determine the percentile for the data value identified by a red box.

- a) 23, 24, 25, 26, 27, 28, 31, 34, 36, 38, 38, 38, 38, 42, 42, 45, 46, 47, 47, 47, 48, 50, 51, 53, 54, 57, 58, 67, 67, 68, 69, 70, 71, 71, 73, 74, 79, 80, 80 39 values
- b) 120, 120, 123, 130, 135, 135, 138, 141, 142, 144, 145, 146, 147, 147, 153, 153, 153, 155, 160, 172, 172, 173, 174, 175, 176, 177, 178, 180, 185, 187 30 values
- c) 12, 15, 18, 18, ..., 321, 322, 328, ..., 841, 844, 849 671 values.
- 230 values                      431 values

42nd  
 57  
 36

7 Below are two stem-and-leaf plots:

Table 1

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

For each of the plots shown above, find:

- a) the mean deviation  
 b) the 87th percentile  
 c) the data value that is nearest to the 80th percentile

8 The data values below are the results from a statistical study.

32 values                      20 values

34, 35, 36, 37, 38, 45, 46, 47, 48, 49, 49, 49, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 59, 59, 59, 65, 66, 67, 68, 69, 70, 72, 76

- a) What data value corresponds to the 86th percentile?  
 b) What is the percentile of the data value 49?

$$\frac{86}{100} \times 32 = 27.52 \rightarrow 27$$

$$\frac{32 - 20 + 1}{32} \times 100 = 41\%$$

7 b)  
 #bottom =  $\frac{87}{100} \times 22 = 19.14 \rightarrow 19\text{th}$   
 c)  $\frac{80}{100} \times 22 = 17.6 \rightarrow 17\text{th}$   
 (108)  
 (99)

9)

Match each description in the left column with a coefficient in the right column.

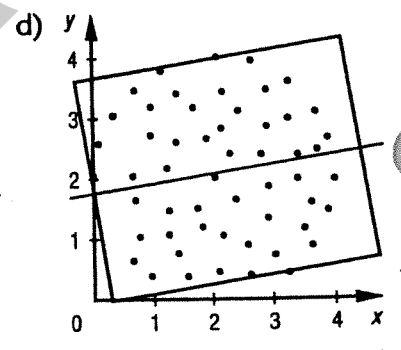
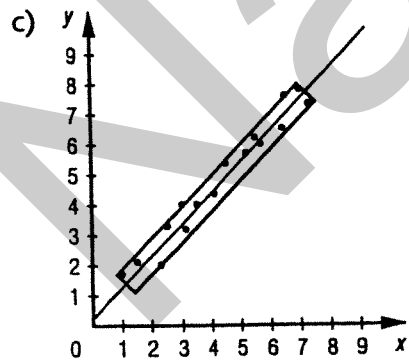
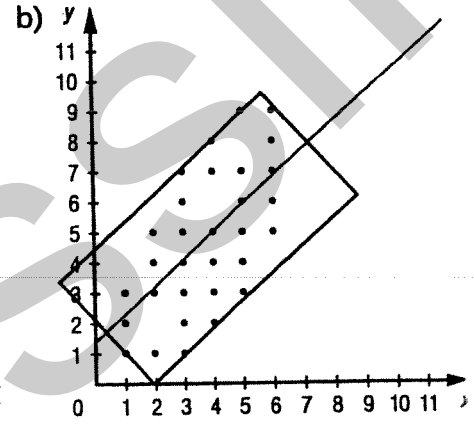
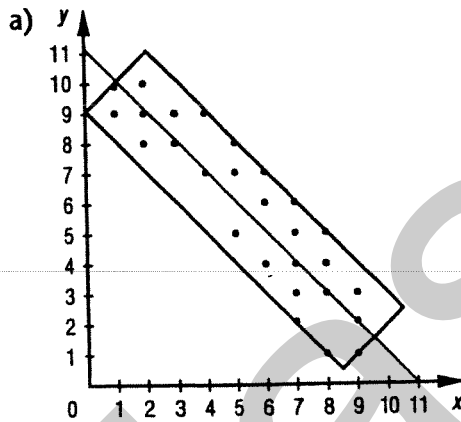
3  
4  
2  
1

Description	Linear correlation coefficient
A Strong	1
B Moderate	2 -0.89
C Slightly	3 0.5
D Perfect	4 0.55

10.)

For each of the adjacent scatter plots, do the following:

- 1) Graphically estimate the linear correlation coefficient.
- 2) Determine the equation of the regression line drawn in a).



M.S.



# STATISTICS REVIEW

MEAN, MEDIAN, MODE, RANGE,  
MEAN DEVIATION,

LINEAR CORRELATION,  
REGRESSION LINE



The distribution below describes the wing span (in mm) of 16 monarch butterflies.

93 93 94 96 96 96 98 100  
100 101 101 101 101 101 103 105

- a) What is the mode of this distribution? **101**  
b) What is the range of this distribution?

$$105 - 93 = 12$$

The monarch butterfly is without a doubt the best known of North American butterflies. It migrates twice a year from Mexico to Canada and back, a feat probably not exceeded by any other insect on earth.

Among the following measurements, which is not a measurement of central tendency?

- A) Median      B) Range      C) Mode      D) Mean

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, 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				

The adjacent table represents the ages of a university's graduates:

Graduation 2008

- a) What is the total number of graduates? **636**  
b) What is the modal class? **[20, 25[**  
c) What is the range? **30**  
d) What is the median class?  
e) What is the mean age of all graduates?  $\frac{20145}{636} = 31.7$

[10, 15[	100
[15, 20[	150
[20, 25[	87
[25, 30[	231
[30, 35[	145
[35, 40[	78
[40, 45[	63
[45, 50[	32

Represent the data below in a table of data grouped into classes. The first class is [10, 20[.

12, 14, 15, 16, 20, 21, 23, 24, 24, 24, 24, 30, 35, 35, 36, 37, 40, 41, 42, 42, 45, 46, 47, 50, 55, 61, 61, 61, 67, 67, 68, 69, 70, 75, 75, 84, 85, 89, 89, 90, 92, 94, 97, 97

3 In each case find:

- 1) the mode    2) the median    3) the mean

a) Children

X	FREQ
0	4
1	6
2	7
3	3

$M_0 = 2$   
 $M_d = 1.5$   
 $M_N = 1.3$

b) Number of cups of coffee

X	FREQ
0	10
1	18
2	14
3	7

$M_0 = 1$   
 $M_d = 1$   
 $M_N = 1.4$

c) Height of horses

Class	FREQ
[100, 140[	5
[140, 180[	9
[180, 220[	13
[220, 260[	1

$M_0 = 2$   
 $M_d = \text{can't}$   
 $M_N = 174.3$

4 For each of the following distributions, calculate:

1) The mean

2) the mean deviation

a) 2, 4, 6, 8, 10

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

1) 6    2) 2.4

1) 8    2) 3

c) 20, 40, 60, 80, 100

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

1) 60    2) 24

1) 9.29    2) 4.98

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

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

1) 7.43    2) 3.51

1) 15.25    2) 2.81

5 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	1
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	2	4	

64, 68, 69, 72, 76, 84, 88, 89, 92, 95

68, 74, 76, 79, 83, 85, 89, 92, 94

6 Here is the data for a statistical distribution:

132, 132, 133, 134, 135, ..., 167, 171, 173, 175, 176, 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? \_\_\_\_\_

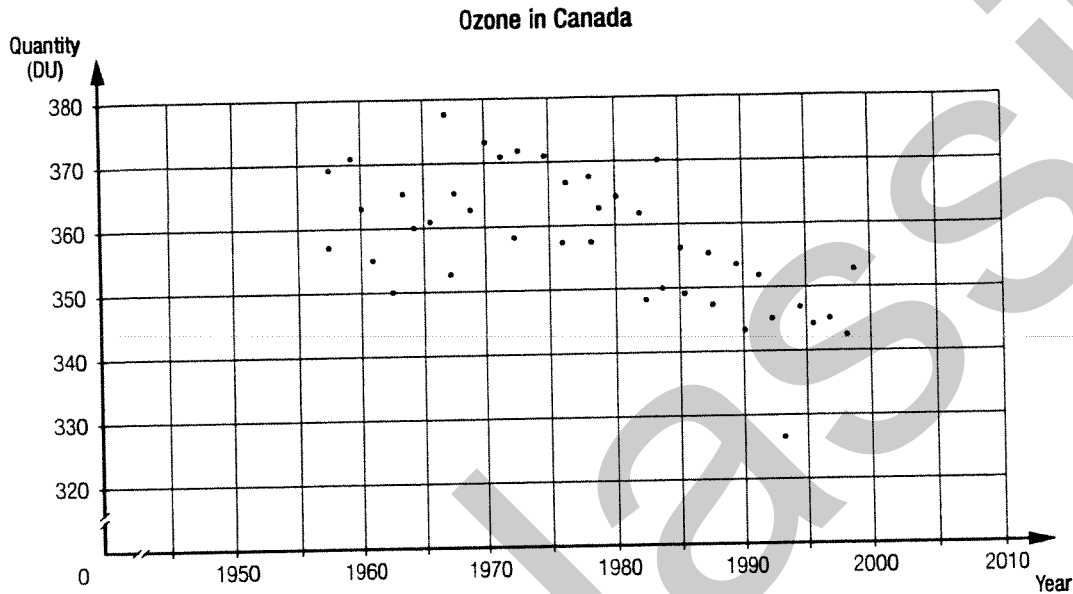
b) 178? \_\_\_\_\_

c) 184? \_\_\_\_\_

d) 244? \_\_\_\_\_

## Number 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. moderate *strong, moderate or weak*
- 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. *Find equation of the line.*

$$y = -0.007x + 1678$$

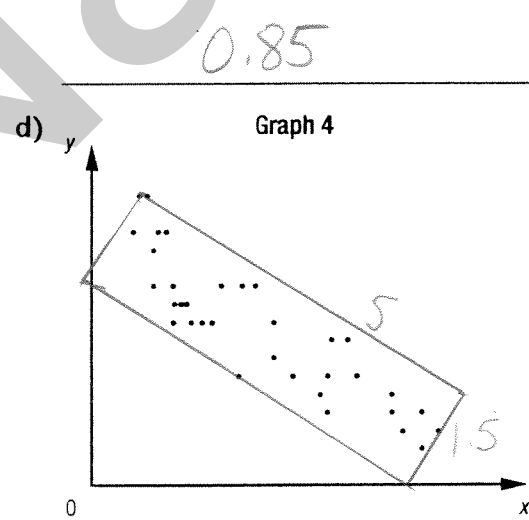
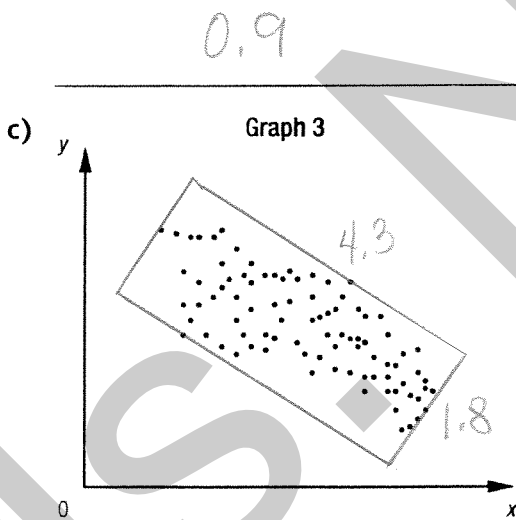
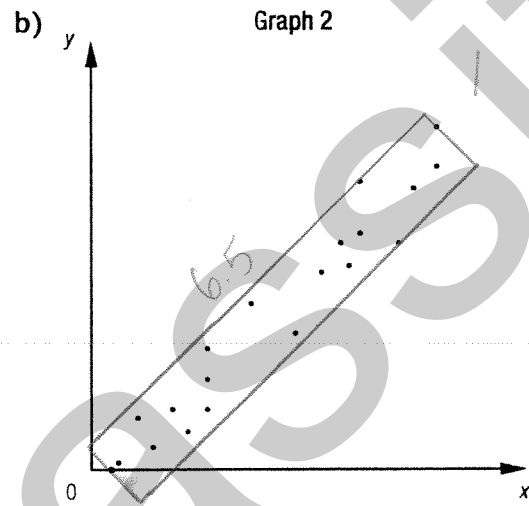
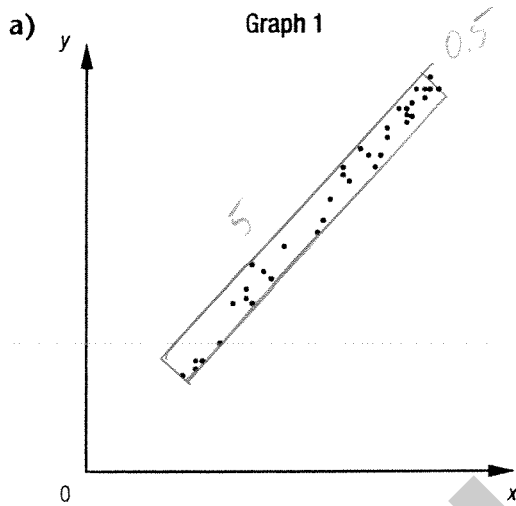
- d) Calculate the quantity of ozone in 2010.

$$331.3 \text{ DU.}$$



### Number 10

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



0.9

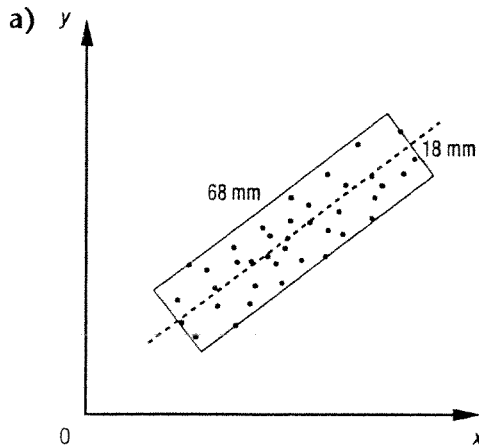
0.85

-0.58

-0.7

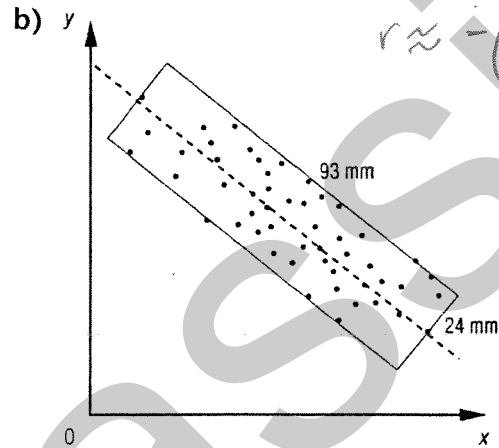
### Quantitative interpretation of correlation

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



$r \approx +\left(1 - \frac{18}{68}\right)$

+0.74



$r \approx -\left(1 - \frac{24}{93}\right)$

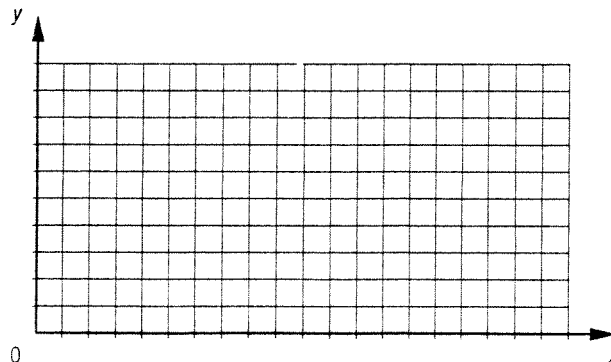
-0.74

**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.



42