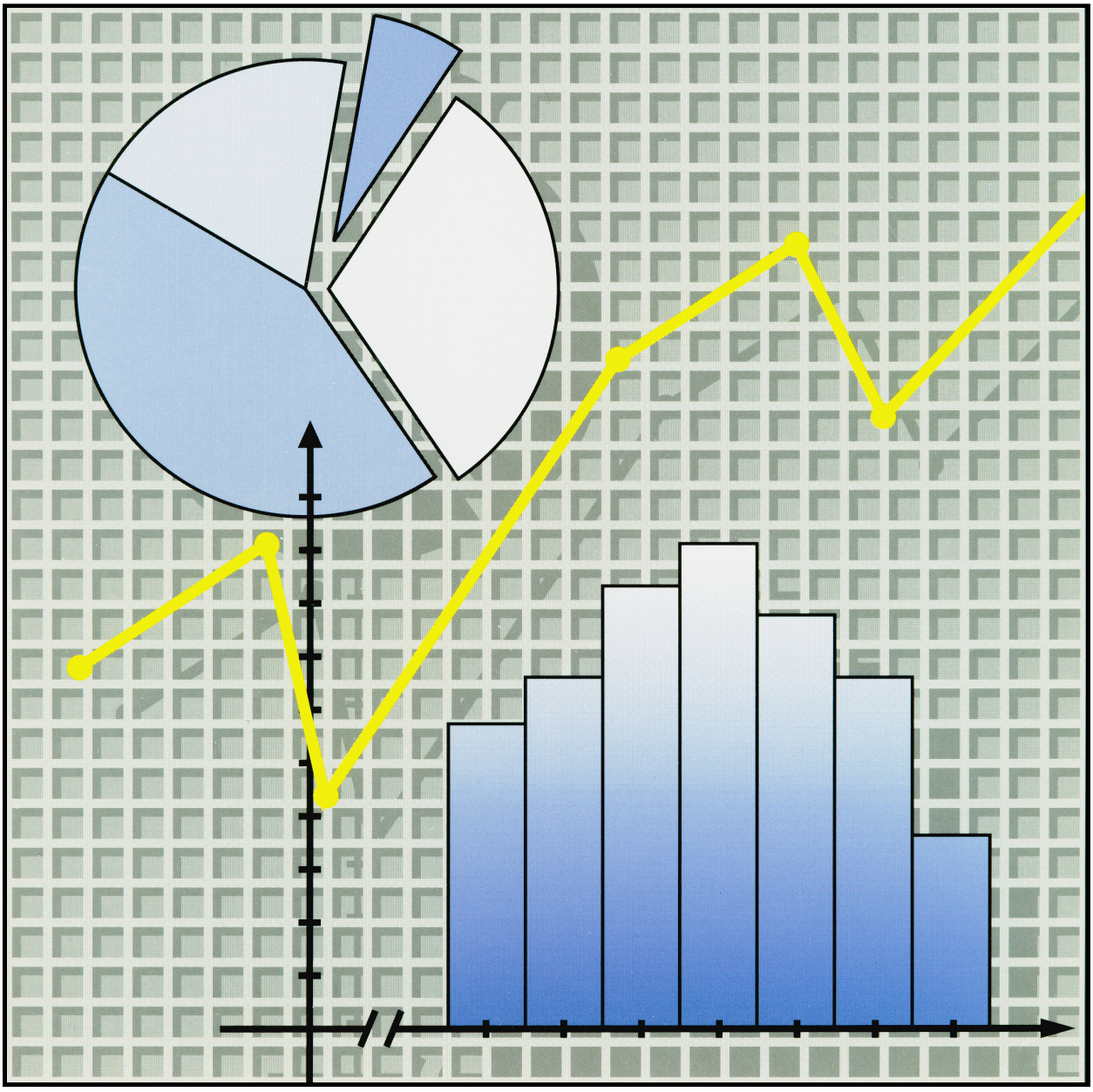
Statistics and Probabilities I

MTH-2008-2



sofad

MTH-2008-2

STATISTICS

AND

PROBABILITIES I

sofad

Author: Suzie Asselin

Content Revision: Jean-Paul Groleau
Daniel Gélineau
Mireille Moisan-Sanscartier

Updates: Mireille Moisan-Sanscartier

Consultants in Andragogy: Les Productions C.G.L. enr.

Coordinator for the DDFD: Jean-Paul Groleau

Coordinator for the DFGA: Ronald Côté

Photocomposition and Layout: Multitexte Plus

Desktop Publishing for Updated Version: L'atelier du Mac inc.

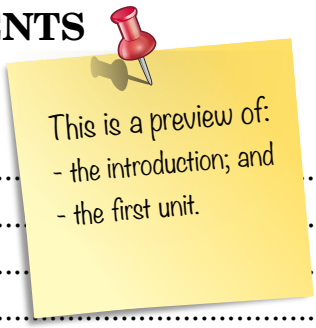
English Version: Direction du développement pédagogique en langue anglaise

- *Translation: William Gore*
- *Linguistic Revision: Linda Arui*

First Printing: 1998

Legal Deposit — 1998
Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada
ISBN 978-2-89493-145-5

TABLE OF CONTENTS



UNITS

INTRODUCTION TO THE PROGRAM FLOWCHART

WELCOME TO THE WORLD OF MATHEMATICS

This mathematics program has been developed for adult students enrolled either with Adult Education Services of school boards or in distance education. The learning activities have been designed for individualized learning. If you encounter difficulties, do not hesitate to consult your teacher or to telephone the resource person assigned to you. The following flowchart shows where this module fits into the overall program. It allows you to see how far you have come and how much further you still have to go to achieve your vocational objective. There are three possible paths you can take, depending on your goal.

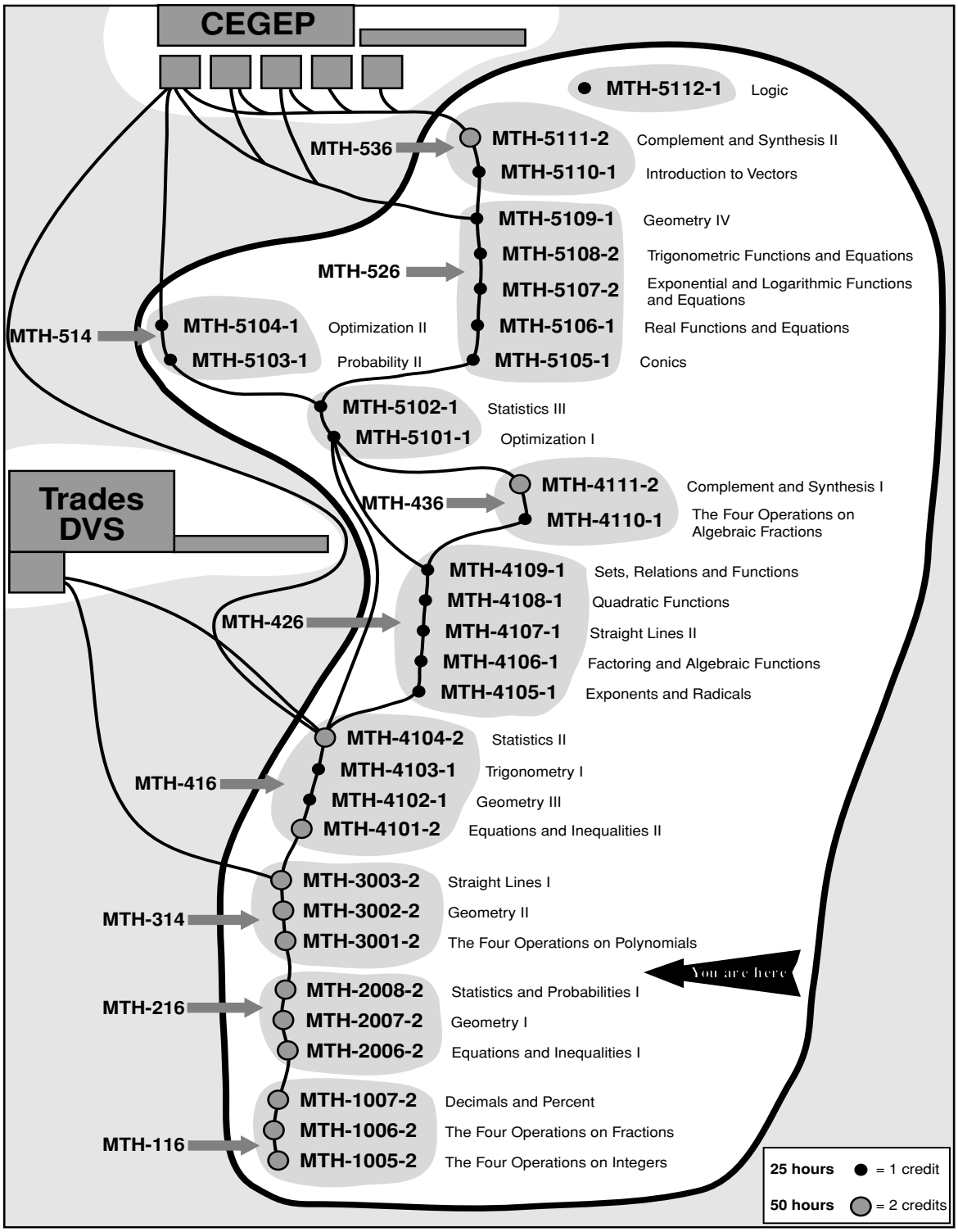
The first path, which consists of Modules MTH-3003-2 (MTH-314) and MTH-4104-2 (MTH-416), leads to a Secondary School Vocational Diploma (SSVD) and certain college-level programs for students who take MTH-4104-2.

The second path, consisting of Modules MTH-4109-1 (MTH-426), MTH-4111-2 (MTH-436) and MTH-5104-1 (MTH-514), leads to a Secondary School Diploma (SSD), which gives you access to certain CEGEP programs that do not call for a knowledge of advanced mathematics.

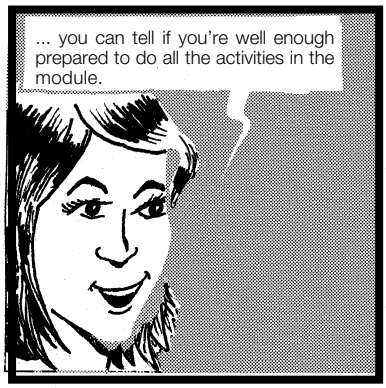
Lastly, the path consisting of Modules MTH-5109-1 (MTH-526) and MTH-5111-2 (MTH-536) will lead to CEGEP programs that require a thorough knowledge of mathematics in addition to other abilities. Good luck!

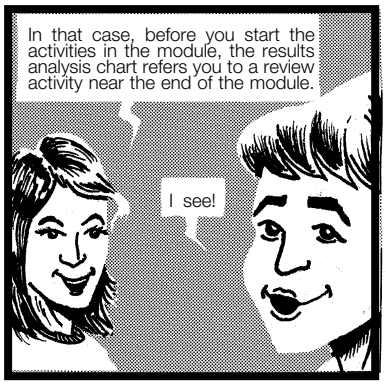
If this is your first contact with the mathematics program, consult the flowchart on the next page and then read the section "How to Use this Guide." Otherwise, go directly to the section entitled "General Introduction." Enjoy your work!

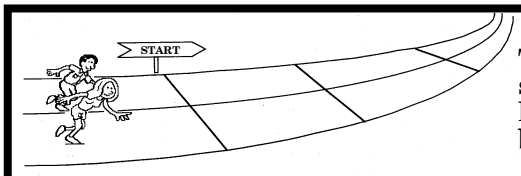
THE PROGRAM FLOWCHART



HOW TO USE THIS GUIDE







The starting line shows where the learning activities **begin**.

? The little white question mark indicates the **questions** for which answers are given in the text.

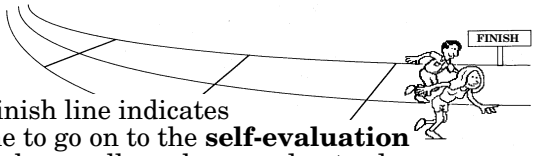
The target precedes the **objective** to be met.

The memo pad signals a **brief reminder** of concepts which you have already studied.

The boldface question mark indicates **practice exercices** which allow you to try out what you have just learned.

The **calculator** symbol reminds you that you will need to use your calculator.

The sheaf of wheat indicates a **review** designed to reinforce what you have just learned. A row of sheaves near the end of the module indicates the final review, which helps you to interrelate all the learning activities in the module.



Lastly, the finish line indicates that it is time to go on to the **self-evaluation test** to verify how well you have understood the learning activities.


There are also many fun things in this module. For example, when you see the drawing of a sage, it introduces a "Did you know that..."
A "Did you know that..."?
Yes, for example, short tidbits on the history of mathematics and fun puzzles. They are interesting and relieve tension at the same time.
Must I memorize what the sage says?
No, it's not part of the learning activity. It's just there to give you a breather.
It's the same for the "math whiz" pages, which are designed especially for those who love math.
They are so stimulating that even if you don't have to do them, you'll still want to.
And the whole module has been arranged to make learning easier.
For example. words in boldface italics appear in the glossary at the end of the module...
Great!
... statements in boxes are important points to remember, like definitions, formulas and rules. I'm telling you, the format makes everything much easier.
The third section contains the final review, which interrelates the different parts of the module.
There is also a self-evaluation test and answer key. They tell you if you're ready for the final evaluation.
Thanks, Monica, you've been a big help.
I'm glad! Now, I've got to run. See you!
Later ...
This is great! I never thought that I would like mathematics as much as this!

GENERAL INTRODUCTION

STATISTISTICALLY SPEAKING, PEOPLE PREFER STUDYING PROBABILITIES, BUT IN ALL PROBABILITY YOU'LL FIND STATISTICS JUST AS ENJOYABLE

You are about to begin studying **statistics and probability**, a field you have not yet explored in this mathematics program. You are undoubtedly familiar with the terms "statistics" and "probability" and you may also refer to statistical information on a regular basis. You may be aware that the probability of winning the Lotto 6/49 jackpot is extremely low, but would you know how to draw a diagram illustrating these statistics? Would you be able to determine the number that would most likely come up if you rolled two or three dice? These are precisely the types of skills that you will have the opportunity to develop through the activities in this module.

The first four units provide an introduction to descriptive statistics. After becoming familiar with basic statistical terminology and learning how to read a statistical table, you will learn how to draw the most common types of graphs (i.e. broken-line graphs, vertical bar graphs, horizontal bar graphs, pictographs, circle graphs and histograms). This newly acquired knowledge will make it easier for you to derive information from graphs you may encounter in newspapers, magazines, textbooks and other publications. In addition, you will find it interesting and useful to learn about all the statistics presented in the examples, exercises, general-interest activities and enrichment activities in this part of the module.

The last two units provide an introduction to probability theory. As with the study of statistics, you will begin by becoming familiar with basic notation and terminology . You will then learn about the concept of probability and even how to calculate probabilities! Of course, these calculations will relate to simple situations only.

After completing this module, you will be able to critically assess the media's interpretation of statistical tables and graphs as well as the conclusions it draws from this information. In addition, the concepts you will explore in the second part of this module will help you develop a better understanding of the term "chance."

Are you ready to take the plunge?

INTERMEDIATE AND TERMINAL OBJECTIVES OF THE MODULE

Module MTH-2008-2 (GSM-223)* contains seven units and requires fifty hours of study distributed as shown below. Each unit covers either an intermediate or a terminal objective. The terminal objectives appear in boldface.

Objectives	Number of Hours**	% (evaluation)
GSM-223-01 and **GSM-223-02**	22	50%
GSM-223-03	8	10%
GSM-223-04	8	15%
GSM-223-05 and **GSM-223-06**	10	25%

* GSM stands for "**G**eneral Education, **S**econdary-level, **M**athematics."

** Two hours are allotted for the final evaluation.

GSM-223-01 Data Tables, Frequency Distribution Tables and the Range of a Distribution

(1) Find the following information in a data table: title, quantified objects, data. (2) Find the following information in a frequency distribution table: title, data, frequencies. If applicable, each numerical value must be stated in the unit of measure indicated in the table. (3) Calculate the range of a given distribution.

GSM-223-02 Broken-Line Graphs, Vertical and Horizontal Bar Graphs and Pictographs

(1) Derive information from a broken-line graph, a vertical bar graph, a horizontal bar graph or a pictograph. (2) Given a data table or a frequency distribution table, draw a broken-line graph, a vertical bar graph, a horizontal bar graph or a pictograph according to instructions. The given table contains from three to eight data items or frequencies. The instructions concern the length of the axes, the creation of scale breaks, the width of the bars in a bar graph or the numerical value of each symbol in a pictograph. The graph must be given a title and each axis must be properly identified and graduated. The numerical value of each symbol must be indicated in the top right-hand corner of the pictograph.

GSM-223-03 Circle Graphs

(1) Derive information from a circle graph. (2) Given a data table or a frequency distribution table, draw a circle graph using a compass, a protractor and a ruler. The given table contains from three to eight data items or frequencies. The graph must be given a title, and each sector must be given a subheading and assigned a percentage value. All the calculations involved in drawing the graph must be shown.

GSM-223-04 Histograms

(1) Derive information from a histogram. (2) Using a frequency distribution table in which the data is grouped into classes, draw a histogram, following any instructions regarding the creation of a scale break. The given table contains from three to eight classes along with their respective frequencies. The histogram must be given a title and each axis must be properly identified and graduated.

GSM-223-05 Tree Diagrams

(1) Draw a tree diagram of the set of all possible outcomes, given the description of a series of two or three random experiments. Each experiment should have from two to six possible outcomes. (2) On the basis of that tree diagram, indicate the following information:

- a particular outcome to be written as an ordered pair or ordered triple,
- the sample space (usually denoted by S) to be written as a set of ordered pairs or ordered triples,
- an event (usually denoted by E) to be written as a set of ordered pairs or ordered triples.

Set theory notation is used. The problems deal with everyday situations and involve a series of two or three random experiments leading to no more than 36 possible outcomes.

GSM-223-06 Calculating Probabilities

Calculate the probability of a particular event or outcome, given a description of a series of two or three random experiments. Probability, denoted by p, can be expressed in one of two ways: as a decimal number rounded to the nearest thousandth such that $0 \leq p \leq 1$, or as a percentage such that $0\% \leq p \leq 100\%$. The problems involve a series of random experiments leading to no more than 36 possible outcomes. The steps in the solution, including a tree diagram, must be shown.

DIAGNOSTIC TEST ON THE PREREQUISITES

Instructions

1. Answer as many questions as you can.

2. You may use a calculator.
3. Make sure you have a ruler graduated in centimetres, a protractor and a compass.
4. Write your answers on the test paper.
5. Do not waste any time. If you cannot answer a question, go on to the next one immediately.
6. When you have answered as many questions as you can, correct your answers using the answer key which follows the diagnostic test.
7. To be considered correct, answers must be identical to those in the key. In addition, the various steps in your answer should be equivalent to those shown in the solution.
8. Transcribe your results onto the chart which follows the answer key. This chart gives an analysis of the diagnostic test results.
9. Do only the review activities that apply to each of your incorrect answers.
10. If all your answers are correct, you may begin working on this module.

1. Perform the following operations without using a calculator and, if necessary, round off your answer to the nearest thousandth.

   a) $1\,179 + 55 =$ ..... b) $4.2 + 1.9 =$ .....

   c) $2\,057 - 178 =$ ..... d) $300 - 53.5 =$ .....

   e) $36 \times 1.75 =$ ..... f) $8.25 \times 20\,000 =$ .....

   g) $7 \div 8 =$ ..... h) $0.83 \div 5 =$ .....

2. Which of the following is a diagram of:

   a) two perpendicular lines? Diagram .....
   b) two parallel lines? Diagram .....

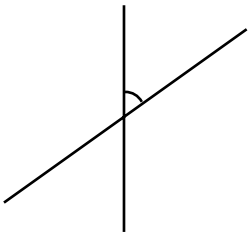


Diagram A

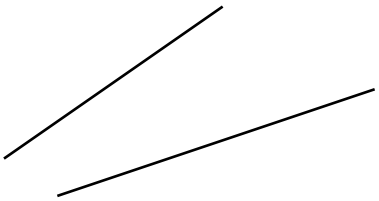


Diagram B

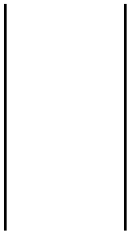


Diagram C

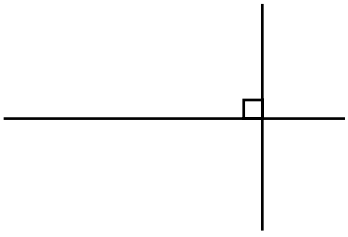


Diagram D

3. How many degrees are there in a right angle? .....

4. Using a ruler, draw line segment AB measuring 8 cm long. Your drawing must be precise to the nearest 0.1 cm.

5. Convert the following decimals to fractions or mixed numbers. The fractions must be reduced to their lowest terms.

   a) 0.25 = ... b) 3.4 = ... c) 17.17 = ...

6. A collection of 30 identical marbles weighs 117 g. What would 7 such marbles weigh? Give the exact answer and show the steps in the solution.

   Ans.: .....

7. Using a protractor, draw angle ABC measuring 15°. Your drawing must be precise to the nearest 2°.

8. Round off the following numbers as indicated below:

   a) 121.37 to the nearest whole number: .....

   b) 0.159 6 to the nearest thousandth: .....

   c) 43.166 to the nearest tenth: .....

9. Convert the following fractions or mixed numbers to decimals and, if necessary, round off to the nearest thousandth.

   a) $\frac{1}{8}$ = ... b) $3\frac{1}{2}$ = ...

   c) $7\frac{3}{13}$ = ...

10. Statistics show that 39% of Québec marriages ended in divorce in 1983. According to this report, how many divorces were there per 1 800 marriages? Show the steps you used to obtain the answer.

    Ans.: .....

11. There are 139 conifers on a lot: 77 red pines, 41 white spruces and 21 thujas. What percentage of these trees are red pines? Show the steps in your solution and round off your answer to the nearest tenth.

    Ans.: .....

12. Convert the following decimals to percentages.

    a) 0.137 = ... b) 0.4 = ...

13. Place the appropriate symbol (i.e. < or >) between each of the following pairs of numbers.

    a) $\frac{1}{4}$ $\frac{4}{17}$ b) 0.2 0.4 c) 0.71 0.589

14. Solve the following problems. Write the operation to be performed and calculate the result.

a) Brenda estimates that an average of 50 birds visit her feeder every day. As a result, a 20-kg bag of birdseed lasts an average of 27 days. If each bag costs $14.97, how much does 1 kg of birdseed cost? Round your answer off to the nearest hundredth.

Ans.: .....

b) A tomato field produces 432 tomatoes, 47 of which are not quite ripe. This crop is better than the last harvest, which yielded only 346 tomatoes. If the farmer sells his tomatoes for $3.25 a dozen, how much money will he get for the most recent crop?

Ans.: .....

c) After buying 40 new stamps, Eddie realized that unfortunately only 16 of them were different from those he already had. He now has 289 stamps in his collection. As he has always done in the past, Eddie gave four of his friends the extra stamps. If the stamps were divided up evenly among the four friends, how many stamps did each friend receive?

Ans.: .....

ANSWER KEY FOR THE DIAGNOSTIC TEST ON THE PREREQUISITES

1. a) 1 234 b) 6.1 c) 1 879 d) 246.5 e) 63
   f) 165 000 g) 0.875 h) 0.166

2. a) Diagram D b) Diagram C

3. 90°

4. A B — $m\overline{AB} = 8$ cm

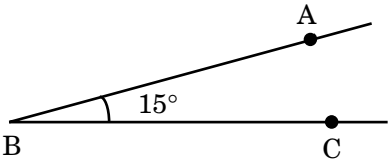
5. a) $0.25 = \frac{25}{100} = \frac{25 \div 25}{100 \div 25} = \frac{1}{4}$ b) $3.4 = 3\frac{4}{10} = 3\frac{4 \div 2}{10 \div 2} = 3\frac{2}{5}$

   c) $17.17 = 17\frac{17}{100}$. The fraction $\frac{17}{100}$ cannot be simplified any further.

6. 30 marbles ⟶ 117 g
   7 marbles ⟶ ?

   $\frac{7 \times 117}{30} = 27.3$

   Ans.: 27.3 g

7. 

   $m\angle ABC = 15°$

8. a) 121 b) 0.160 c) 43.2

9. a) $1 \div 8 = 0.125$ b) $\frac{1}{2} = 1 \div 2 = 0.5$; $3\frac{1}{2} = 3.5$

   c) $\frac{3}{13} = 3 \div 13$ 0,231; $7\frac{3}{13} = 7.231$

10. 39% of 1 800 =

    $1\,800 \times \frac{39}{100} = 702$

    Ans.: 702 divorces

11. $\frac{77 \text{ red pines}}{139 \text{ conifers}}$; $\frac{77}{139}$ 0.554

    $0.554 \times \frac{100}{100} = \frac{55.4}{100} = 55.4\%$

    Ans.: 55.4% of the trees are red pines.

12. a) 13.7% b) 40%

13. a) $\frac{1}{4} > \frac{4}{17}$ b) $0.2 < 0.4$ c) $0.71 > 0.589$

14. a) $\$14.97 \div 20 \text{ kg} = \$0.75/\text{kg}$

    Each kilogram of birdseed therefore costs $0.75.

    b) $(432 \text{ tomatoes} \div 12) \times \$3.25 = 36 \text{ dozen tomatoes} \times \$3.25 = \$117.00$

    The farmer will get $117.00 for his most recent crop.

    c) $(40 - 16) \div 4 = 24 \div 4 = 6$

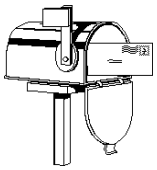
    Each friend received 6 stamps.

ANALYSIS OF THE DIAGNOSTIC TEST RESULTS

Question	Answer		Review		Before Going to Unit(s)
	Correct	Incorrect	Section	Page	
1. a)			8.2	8.20	1 to 6
b)			8.2	8.20	1 to 6
c)			8.3	8.23	1 to 6
d)			8.3	8.23	1 to 6
e)			8.4	8.28	2 to 6
f)			8.4	8.28	2 to 6
g)			8.5	8.32	2 to 6
h)			8.5	8.32	2 to 6
2. a)			8.6	8.38	2, 3 and 4
b)			8.6	8.38	2, 3 and 4
3.			8.6	8.38	2, 3 and 4
4.			8.7	8.40	2 and 4
5. a)			8.8	8.43	2 and 6
b)			8.8	8.43	2 and 6
c)			8.8	8.43	2 and 6
6.			8.9	8.46	2 and 4
7.			8.10	8.50	3
8. a)			8.11	8.54	3 and 6
b)			8.11	8.54	3 and 6
c)			8.11	8.54	3 and 6
9. a)			8.12	8.57	3 and 6
b)			8.12	8.57	3 and 6
c)			8.12	8.57	3 and 6
10.			8.13	8.60	3
11.			8.13	8.63	3
12. a)			8.13	8.62	6
b)			8.13	8.62	6
13. a)			8.14	8.68	6
b)			8.14	8.68	6
c)			8.14	8.68	6
14. a)			8.1	8.4	1 to 6
b)			8.1	8.4	1 to 6
c)			8.1	8.4	1 to 6

- If all your answers are **correct**, you may begin working on this module.

- For each **incorrect** answer, find the related section listed in the **Review** column and complete it before beginning the unit(s) listed in the right-hand column under the heading **Before Going on to Unit(s)**.



INFORMATION FOR DISTANCE EDUCATION STUDENTS

You now have the learning material for MTH-2008-2 (GSM-223) together with the homework assignments. Enclosed with this material is a letter of introduction from your tutor indicating the various ways in which you can communicate with him or her (e.g. by letter, telephone) as well as the times when he or she is available. Your tutor will correct your work and help you with your studies. Do not hesitate to make use of his or her services if you have any questions.

DEVELOPING EFFECTIVE STUDY HABITS

Distance education is a process which offers considerable flexibility, but which also requires active involvement on your part. It demands regular study and sustained effort. Efficient study habits will simplify your task. To ensure effective and continuous progress in your studies, it is strongly recommended that you:

- draw up a study timetable that takes your working habits into account and is compatible with your leisure time and other activities;

- develop a habit of regular and concentrated study.

The following guidelines concerning the theory, examples, exercises and assignments are designed to help you succeed in this mathematics course.

Theory

To make sure you thoroughly grasp the theoretical concepts:

1. Read the lesson carefully and underline the important points.

2. Memorize the definitions, formulas and procedures used to solve a given problem, since this will make the lesson much easier to understand.

3. At the end of an assignment, make a note of any points that you do not understand. Your tutor will then be able to give you pertinent explanations.

4. Try to continue studying even if you run into a particular problem. However, if a major difficulty hinders your learning, ask for explanations before sending in your assignment. Contact your tutor, using the procedure outlined in his or her letter of introduction.

Examples

The examples given throughout the course are an application of the theory you are studying. They illustrate the steps involved in doing the exercises. Carefully study the solutions given in the examples and redo them yourself before starting the exercises.

Exercises

The exercises in each unit are generally modelled on the examples provided. Here are a few suggestions to help you complete these exercises.

1. Write up your solutions, using the examples in the unit as models. It is important not to refer to the answer key found on the coloured pages at the end of the module until you have completed the exercises.

2. Compare your solutions with those in the answer key only after having done all the exercises. **Careful!** Examine the steps in your solution carefully even if your answers are correct.

3. If you find a mistake in your answer or your solution, review the concepts that you did not understand, as well as the pertinent examples. Then, redo the exercise.

4. Make sure you have successfully completed all the exercises in a unit before moving on to the next one.

Homework Assignments

Module MTH-2008-2 (GSM-223) contains three assignments. The first page of each assignment indicates the units to which the questions refer. The assignments are designed to evaluate how well you have understood the material studied. They also provide a means of communicating with your tutor.

When you have understood the material and have successfully done the pertinent exercises, do the corresponding assignment immediately. Here are a few suggestions.

1. Do a rough draft first and then, if necessary, revise your solutions before submitting a clean copy of your answer.

2. Copy out your final answers or solutions in the blank spaces of the document to be sent to your tutor. It is preferable to use a pencil.

3. Include a clear and detailed solution with the answer if the problem involves several steps.

4. Mail only one homework assignment at a time. After correcting the assignment, your tutor will return it to you.

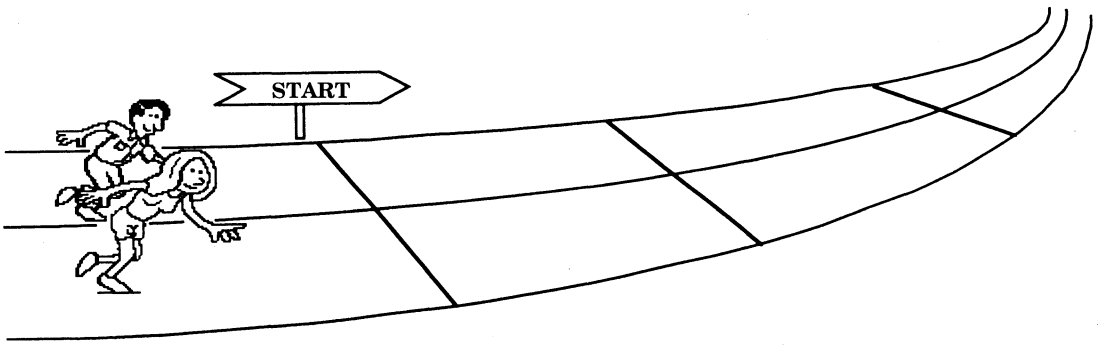
In the section "Student's Questions", write any questions which you may wish to have answered by your tutor. He or she will give you advice and guide you in your studies, if necessary.

In this course

Homework Assignment 1 is based on Units 1 to 3.
Homework Assignment 2 is based on Units 4 to 6.
Homework Assignment 3 is based on Units 1 to 6.

CERTIFICATION

When you have completed all the work, and provided you have maintained an average of at least 60%, you will be eligible to write the examination for this course.



UNIT 1

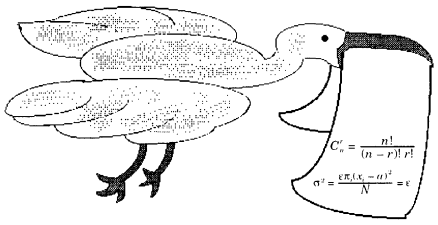
DATA TABLES, FREQUENCY DISTRIBUTION TABLES AND THE RANGE OF A DISTRIBUTION

1.1 SETTING THE CONTEXT

The Origin of Statistics: A Fortunate Coincidence

Statistics have been used for a very long time! They originally consisted of **numerical *data*** compiled to serve the political, military and fiscal needs of the state. In fact, the term **statistics** comes from the latin word *statisticus*, which meant "of state affairs."

Created in the 18th century, **statistics is the branch of mathematics that deals with the analysis and interpretation of numerical data related to a specific situation**.



$$C_n^r = \frac{n!}{(n-r)!\,r!}$$

$$\sigma^2 = \sum_{i=1}^{n} \frac{n_i(x_i-\mu)^2}{N} = \Sigma f_i m_i^2 - \mu_i^2$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} = e^{-(x-\mu)^2/2\sigma^2}$$

It should be noted that advances in statistics were closely linked to the development of ***probability theory***. The calculation of probabilities has its origins in games of chance, as illustrated in the following anecdote. One day in seventeenth-century France, the Knight of Méré, an avid gambler, was playing dice with one of his friends. Each player had to choose a number and the winner would be whoever was first to roll his number three times. Méré was forced to leave before the end of the game, however, because King Louis XIV had summoned him to an emergency meeting. At that point, Méré's number had come up twice and his friend's, only once. Since each player still had a chance to win the bet, they had to determine how they would divide up the pool. Méré asked his friend, the mathematician Blaise Pascal, to solve this problem. Pascal took a great interest in this matter and, after giving it much thought, managed to find a solution. This initial work marked the beginning of a long series of efforts aimed at discovering the laws of probability.

These laws make it possible to predict the behaviour of an entire population by analyzing data relating to a sample of that population. For example, polling firms can now determine which candidate will win the next election by surveying a representative sample of voters.

The development of statistical methods has had a major impact on the advancement of science. For example, it is now possible to consider each secondary factor that might have an influence on a phenomenon or an ***experiment***. This has led to the development of much more realistic scientific models. Statistical mechanics, as applied to physics and chemistry, has given rise to a number of new theories, one of which is quantum mechanics. The science of genetics would not have progressed without the development of the mathematical concepts associated with statistics. Statistical methods are also widely used in other fields such as sociology, psychology, anthropology, economics and political science.

In everyday life, a basic knowledge of statistics makes it possible to understand the numerical data relating to a study, an experiment or an opinion poll. This knowledge could prove useful, since we must often make decisions on the basis of this type of numerical information.

Since any statistical analysis is based on information compiled in a ***data table*** or a ***frequency distribution table***, it is important to know how to read these tables before studying statistics in greater detail. Thus, learning how to read these tables will be the objective of this unit.

To achieve the objective of this unit, you should be able to identify the main components of a data table (i.e. *title*, *quantified objects* and data) as well as the main components of a frequency distribution table (i.e. title, data and *frequencies*). You should also know how to calculate the *range of a distribution*.

Before going on to the first example, note the following definition of a data table or a frequency distribution table.

> Both **data tables** and **frequency distribution tables** usually contain numerical information, which has been neatly and logically arranged in rows and columns.

The following examples will illustrate the difference between a data table and a frequency distribution table.

Example 1

Data Table

The following is the approximate length of each of Québec's boundaries: the Québec – Ontario border is **1 123 km** long; the Québec – Northwest Territories border is **3 185 km** long; the Québec – Newfoundland (i.e. Labrador) border is **4 425 km** long; the coast along the gulf of St. Lawrence is **940 km** long; the Québec – New Brunswick border is **459 km** long and the border between Québec and the United States is **735 km** long. This same information appears in the data table below.

Table 1.1 Approximate Length of Québec's Boundaries in 1985

Boundary	Length (km)
Québec – Ontario	1 123
Québec – Northwest Territories	3 185
Québec – Newfoundland (Labrador)	4 425
Gulf of St. Lawrence	940
Québec – New Brunswick	459
Québec – USA	735

Source: Ministère de l'Énergie et des Ressources, Service de l'intégrité du territoire, cited in *Le Québec statistique (édition 1985-1986)*.

Which version do you prefer?

You need only take a quick look at these figures to see that you can locate the information more easily in the data table. The table also makes it easier to compare the different values.

When working with a data table, you should first read its title.

The **title** of a table is a clear and concise statement of the type of information presented in that table. The title always appears above the table.

In addition, a **subtitle** appears at the top of each column in the table. Each subtitle helps you immediately identify the type of information listed in each column.

In Example 1, the figures listed in the right-hand column are called **data**. Note that outside the field of statistics, the term "data" has a very broad meaning: it refers to any numerical information. In statistics, however, "data" is defined as follows:

Data are the **results of experiments**. In statistics, an **experiment** can involve, for example, asking a question in an opinion poll, measuring a specific quantity (e.g. a person's height, the weight of all the domestic waste produced by one household in one week, the duration of a surgical procedure, the money that a family spends on leisure activities every year) or observing any characteristic (e.g. the colour of a person's eyes, the type of siding used on a house, a worker's means of transportation, a person's nationality). Therefore, data can consist of answers to questions, quantities or characteristics.

In Table 1.1, for example, the data are the results of the measurement of length. Since the ***unit of measure***, the kilometre, is the same for all the data, that unit is indicated in the subtitle in the right-hand column.

The left-hand column of Table 1.1 contains the **quantified objects**.

In statistics, any element associated with data is called a **quantified object**.

In Table 1.1, the quantified objects are Québec's boundaries, since each boundary is associated with data (i.e. a length in kilometres).

In a frequency distribution table, however, the **data** appear in the left-hand column. The right-hand column contains the **frequencies** associated with each data item. We will be working with frequencies whenever we study a group of people or objects.

The term **frequency** means the number of times that a specific data item is recorded in a given situation.

The following example illustrates this concept.

Example 2

Frequency Distribution Table

In a survey of 150 people selected at random in the city of Sherbrooke, each person was asked to indicate his or her blood type (i.e. A, AB, B or O). The results, or **data**, are recorded below.

O, O, A, O, O, O, A, A, O, A, O, O, A, O, A,
A, A, A, B, O, A, A, O, O, O, O, A, O, A, A, B,
O, O, O, A, A, A, O, AB, O, O, A, A, O, A,
A, A, A, O, B, A, O, O, A, O, B, A, O, O, O,
AB, O, A, A, O, A, O, O, O, O, B, O, O, O, B,
O, A, A, A, O, A, O, A, A, O, A, A, O, A, O,
O, O, B, O, A, A, O, A, AB, O, A, B, O, O, A,
A, A, O, A, AB, O, A, A, O, O, A, O, A, A, O,
O, O, B, A, A, A, O, A, O, O, A, O, B, O, O,
A, O, A, O, O, A, A, A, B, O, O, A, A, O, O.

Blood type O is noted **72** times;
Blood type A is noted **63** times;

? Blood type B is noted ..... times;
? Blood type AB is noted ..... times.

Data item B, or blood type B, is noted **11 times** and data item AB is noted **4 times** on page 1.6. Each numerical value (72, 63, 11 and 4) is the frequency of its corresponding data item (O, A, B and AB). This information can be used to prepare the following **frequency distribution table**.

Table 1.2 Frequency Distribution of the Blood Types of 150 Sherbrooke Residents

Blood Type	Number of People
O	72
A	63
B	11
AB	4

Source: Nosy Pollsters Inc.

N.B. The source, namely the person or publication that provides the information, is always indicated. It usually appears below the table.

Now see if you can identify the different parts of a data table or of a frequency distribution table by answering the questions pertaining to Tables 1.3 and 1.4.

Table 1.3 Total Assets of the Montreal City and District Savings Bank in 1981, 1982 and 1983

Year	Total Assets (thousands of $)
1981	1 961 650
1982	2 084 223
1983	2 477 555

Source: Montreal City and District Savings Bank: Annual Report, cited in *Le Québec statistique (édition 1985-1986)*.

? Is this a data table or a frequency distribution table? .....

? What is the title of this table? .....
.....

? List each quantified object.
.....

? What is the subtitle of the column containing the data?
.....

? What were the total assets of this bank in 1982? .....

Compare your answers with the correction key below. If your answer is different from the one suggested in the key, make the necessary corrections and read the accompanying explanation.

Correction Key

This is a data table, since each quantified object (a year) is associated with data (total assets).

Title: Total Assets of the Montreal City and District Savings Bank in 1981, 1982 and 1983

Quantified Objects: 1981, 1982, 1983.

N.B. In statistics, the term "object" has a very broad meaning: it can refer to people, things, facts, years, countries, etc.

Subtitle of the column containing the data: Total Assets (in thousands of $).

Total assets of this bank in 1982: $2 084 223 000.

N.B. Since the assets are evaluated in thousands of dollars, each number in the data column must be multiplied by 1 000.

Table 1.4 Frequency Distribution of the Top Ten Birthplaces of Immigrants Admitted to Québec in 1983

Birthplace	Immigrants Admitted
Haiti	2 754
Vietnam	1 154
France	857
El Salvador	903
Poland	722
United States	601
India	588
Kampuchea	586
Morocco	405
Lebanon	356
Other Countries	7 348

Source: Ministère des Communautés culturelles et de l'Immigration, cited in *Le Québec statistique (édition 1985-1986).*

? What is the 5th data item in the table? .....

? What is its frequency? .....

? Briefly describe what we can conclude from the information in row 2 of this table.

.....

.....

In Table 1.4, the data consist of birthplaces, and the number of immigrants admitted from each country are the frequencies. These numbers are frequencies because they represent the number of times that each data item (i.e. birthplace) appears in the table.

The **5th** data item is Poland, and its frequency is **722** people.

Row 2 of the table states that 1 154 immigrants admitted to Québec in 1983 came from Vietnam. In order to understand what each row of the table means, you must read the title of that table carefully.

Let's recap the difference between data tables and frequency distribution tables.

Data Table

Quantified Objects	Data

Each element in the left-hand column (quantified object) is associated with a number in the right-hand column. That number represents the result of an experiment (data).

Frequency Distribution Table

Data	Frequencies

In this table, the left-hand column contains data, since it indicates the results of experiments (i.e. the measurement of quantities or the observation of characteristics). The right-hand column indicates the frequencies, namely the number of times each result occurs.

Example 3

Table 1.5 Number of Cars per Family among 8 Families Living on Maplewood Street

Family	**Number of Cars**
A	1
B	2
C	0
D	2
E	1
F	1
G	0
H	2

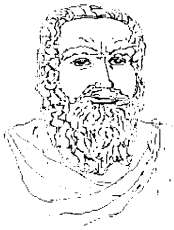
Data Table

Table 1.6 Frequency Distribution of the Number of Cars per Family among 8 Families Living on Maplewood Street

Frequency Distribution Table

Number of Cars	Number of Families
0	2
1	3
2	3

In Table 1.5, the quantified objects are the families, and the data consist of the number of cars belonging to each family. The number of times that each data value (number of cars) appears in the data table determines the frequency of each data value. For example, the frequency of the data value "1 car" is "3 families" because there are three instances of a family owning 1 car.



Did you know that...

... statistical tables often contain several columns of data pertaining to the same quantified objects? In addition, quantified objects are often categorized. As a result, these tables contain a great deal of information even though they take up a relatively small amount of space.

The following statistical tables illustrate the features described above. Imagine the number of pages it would take to write all this information out in complete sentences! Even if you tried, the result would be a jumble of virtually incomprehensible words, numbers and symbols!

Table 1.7 Harvest, Average Farm Price and Total Farm Value of Selected Fruits, Québec, 1979-1984

Year	Total Harvest (t)	Average Farm Price ($/t)	Total Farm Value (thousands of $)
Apples			
1979	91 216	225.20	20 542
1980	111 515	156.31	17 431
1981	45 303	267.70	12 128
1982	78 109	239.88	18 737
1983	65 078	245.14	15 953
1984	85 081	..	..
Strawberries			
1979	8 787	833.00	7 320
1980	8 104	963.00	7 804
1981	6 585	1 210.00	7 968
1982	10 152	1 238.00	12 568
1983	9 384	1 199.00	11 251
1984	11 925	1 103.00	13 153
Raspberries			
1979	439	2 559.00	1 123
1980	501	2 574.00	1 290
1981	413	3 637.00	1 502
1982	645	3 003.00	1 937
1983	933	2 927.00	2 731
1984	814	3 141.00	2 557

Source: Bureau de la statistique du Québec, "Fruits", cited in *Le Québec statistique (édition 1985-1986)*.

Table 1.8 Sales of Ornamental and Fruit Nursery Stock, Québec, 1981-1983

Value of Stock Sold		1981	1982	1983
Retail Sales	$	3 055 696	1 827 314	1 747 469
Number Reporting		63	71	67
Average per Firm	$	48 503	25 737	26 082
Sales to Fruit Growers	$	315 395	805 214	1 139 224
Number Reporting		17	17	17
Average per Firm	$	18 553	47 366	67 013
Sales to Landscape Contractors	$	870 143	871 229	860 860
Number Reporting		46	48	44
Average per Firm	$	18 916	18 151	19 565
Sales to Garden Centres	$	845 361	992 538	1 254 808
Number Reporting		28	31	31
Average per Firm	$	30 191	32 017	40 478
Sales to Mass Merchandisers	$	195 045	238 379	214 750
Number Reporting		9	13	9
Average per Firm	$	21 672	18 337	23 861
Sales to Other Growers of Nursery Stock	$	701 053	430 007	570 800
Number Reporting		29	32	26
Average per Firm	$	24 174	13 438	21 954
Total Sales	$	5 982 693	5 164 681	5 787 911
Number Reporting		86	101	100
Average per Firm	$	69 566	51 135	57 879

Source: Statistics Canada, *Survey of Canadian Nursery Trades Industry* (Catalogue 22-203), for the years 1980-81 and 1982-83.

Now that you know how to read data tables and frequency distribution tables, the next step is to examine the **range of a distribution** and to learn how to calculate this value using a data table.

A ***distribution*** is a set of data or frequencies.

The **range** of a distribution is a measure of the dispersion of the data or the frequencies.

To calculate the range of a distribution:

1. select the data item or frequency with the highest numerical value;
2. select the data item or frequency with the lowest numerical value;
3. subtract the lowest numerical value from the highest numerical value;
4. state this numerical value in the appropriate unit, if applicable.

Let's use this ***algorithm*** in the example below.

Example 4

The following frequency distribution table gives the number of Criminal Code violations (excluding traffic violations) in each of Québec's major cities (municipalities with a population of 50 000+).

Table 1.9 Frequency Distribution of Criminal Code Violations* (except traffic violations) in Municipalities with More Than 50 000 Inhabitants, Québec, 1983

Municipality	Criminal Code Violations
Beauport	3 627
Brossard	3 731
Charlesbourg	4 004
Chicoutimi	3 981
Gatineau	7 055
Hull	8 490
Jonquière	4 476
Laval	17 343
Longueuil	12 956
Montréal (CUM)	159 738
Québec	19 177
Sainte-Foy	7 026
Saint-Hubert	4 093
Sherbrooke	6 541
Trois-Rivières	5 678

*Reported by municipal police forces.

Source: Ministère de la Justice du Québec, Direction générale de la sécurité publique, Statistiques 1983, "Criminalité et application des règlements de la circulation au Québec", cited in *Le Québec Statistique (édition 1985-1986)*.

In this table, the set of numbers in the right-hand column is a distribution. Let's calculate the range of this distribution

- Highest frequency value: **159 738** .
- Lowest frequency value: **3 627**.
- Hence, the range of this distribution is 159 738 – 3 627 = **156 111**.
- This means that the difference between the highest and the lowest number of violations is 156 111 violations.

Note that there would be much less of a difference between the highest and lowest values if the frequency for Montréal were not part of the distribution. If such were the case, the difference would be: 19 177 – 3 627 = 15 550.

N.B. The city with the highest crime rate in 1983 was Hull. In calculating the crime rate, it is important to consider not only the number of violations, but also the population of each municipality. Among the 15 cities surveyed, Montréal has the fifth highest crime rate. The most "peaceful" city is Charlesbourg.

Before going on to the exercises, answer the following question to make sure that you have understood these concepts.

The following table indicates the marks that Professor Albert Einstein's students obtained on a math test.

Table 1.10 Mathematics Test Results

Name	**Mark (out of 50)**
BERNOUILLI, Jacob	45
FERMAT, Pierre	42
GALTON, Francis	41
GAUSS, Carl Friedrich	49
HUYGENS, Christian	43
LAPLACE, Pierre Simon de	47
MORROV, Andrei Andreievitch	44
MOIVRE, Abraham de	48
NEUMANN, Johannes von	41
PEARSON, Karl	39
PASCAL, Blaise	49
POISSON, Simon Denis	47
SPEARMAN, Charles	42
TCHEBYCHEV, Papnuti Lvovitch	41

Source : Professor Albert Einstein.

? What is the range of this distribution? .....

To calculate the range, subtract the lowest mark, **39**, from the highest mark, **49**: **49 – 39 = 10**. Hence, there is a 10-mark difference between the highest and lowest result.

The range is one of the measures of dispersion used in statistics and is the easiest to interpret and calculate. On the other hand, it provides the least significant information about a given distribution. The other types of measures of dispersion (i.e. average deviation, variance, standard deviation and semi-interquartile range) will not be covered in this module, because they involve more complicated calculations.

For the moment, simply remember that measures of dispersion describe a set of data or frequencies and don't worry about the meaning of each measure. When you begin studying statistical graphs, you will get a clearer picture of what the term "range" means.

1.2 PRACTICE EXERCISES

1. The questions below pertain to the following data table.

Output Volume of Pulp and Paper Mills by Product, Québec, January 1984

Product	Output Volume (thousands of tonnes)
Newsprint and Related Specialties	5 290
Fine and Special Paper	330
Tissue Paper	205
Market Pulp	13 555
Kraft Paper and Kraft Board	755
Paperboard	315
Building Materials (e.g. insulation)	270

Source: Ministère de l'Énergie et des Ressources, "Ressource et industrie forestières, Portrait statistique", 1984 edition, cited in *Le Québec statistique (édition 1985-1986)*.

a) What is the title of this data table?

.....

.....

b) What is the subtitle of the data column? .....

c) What is the subtitle of the column containing the quantified objects?

.....

d) What quantity of newsprint and related specialties did Québec produce in January 1984? .....

e) What is the range of this distribution? .....

2. Answer the questions below by referring to the following frequency distribution table.

Frequency Distribution of Certified General and Specialized Bookstores, Québec, 1985

Type of Bookstore	Number of Bookstores
General Bookstores	
• French-language	135
• English-language	11
• French- and English-language	9
Specialized Bookstores	
• French-language	11
• English-language	1
• French- and English-language	5

Source: Ministère des Affaires culturelles, Direction générale des industries culturelles, cited in *Le Québec statistique (édition 1985-1986)*.

a) Identify the first data item listed in the table.

.....

b) What is the frequency of this first data item? .....

c) Briefly describe the information in the last row of the table.

.....

.....

d) Calculate the range of this distribution. .....

3. The questions below pertain to the following data table.

Payroll by Police Force, Québec, 1983

Police Force	Payroll (thousands of $)
Québec Police Force	197 368
MUC Police Department	187 450
Municipal Police Forces	
• Population of 5 000+*	161 370
• Population of less than 5 000	7 610

* Excluding the MUC Police Department.
Source: Commission de police du Québec, "Rapport Annuel 1983", cited in *Le Québec statistique (édition 1985-1986)*.

a) List each quantified object.

.....
.....
.....

b) What is the subtitle of the column containing the data?

.....

c) In 1983, what was the total payroll for police forces in Québec municipalities with 5 000 or more inhabitants?

.....

d) What is the range of this distribution?

.....

4. The data table below shows how many copies of Québec periodicals (magazines) were printed in 1983. Write the following words in the appropriate blank space below: **data**, **quantified objects** and **title of the table**.

a) ..... { Number of Copies of Québec Periodicals Printed in 1983
.....

b) ..... { data
.....

Periodicals	# of Copies
General Magazines	
• Montréal	6 941 817
• Rest of Québec	416 399
Specialized Magazines	
• Montréal	1 270 499
• Rest of Québec	251 907

c) .....
.....

Source: Ministère des Communications, "Banque de données 1983", *cited in Le Québec statistique (édition1985-1986).*

d) Calculate the range of this distribution.

.....

1.3 SUMMARY ACTIVITY

1. Complete each statement, using each of the following terms only once: quantified object, data, distribution, range, frequency, data table or frequency distribution table, title of the table.

   a) A numerical value representing the number of times that a data item appears in a table is called a .....

   b) The measure of the dispersion in the data or frequencies related to a given situation is called the .....

   c) When you have arranged information neatly and logically in rows and columns, you have prepared a .....

   d) In statistics, results of an experiment are called .....

   e) A clear and concise statement of the type of information presented in a table is called a .....

   f) Any element associated with data is called a .....

   g) A set of data or frequencies is called a .....

2. The table below shows the commercial production (in tonnes) of different types of fruit in Canada in 1985.

Type of Fruit	Production (t)
Apples	478 606
Apricots	2 274
Blueberries	22 432
Cherries (sweet)	8 690
Cherries (sour)	7 349
Cranberries	8 186
Grapes	76 635
Peaches	42 204
Pears	28 217
Plums and Prunes	5 485
Raspberries	15 262
Strawberries	38 301

Source: *Canada Year Book 1988*.

a) Give this table a title.

.....

.....

b) Are the numerical values in the right-hand column data or frequencies? Briefly explain your answer.

.....

.....

c) What is the range of this distribution? Describe each step in your solution according to the algorithm outlined in this unit.

.....

.....

.....

.....

d) To which quantified objects do the numerical values used to calculate the range correspond?

.....

.....

3. Complete the following sentences.

A table with data in the left-hand column and frequencies in the right-hand column is called a ..... table.

A table with ..... in the left-hand column and data in the right-hand column is called a .....table.

1.4 THE MATH WHIZ PAGE

Converting a Data Table to a Frequency Distribution Table

Mrs. Childs is concerned about the falling birth rate in her area and has decided to write an article on this topic for the local newspaper. To illustrate her point, she would like to prepare a statistical table. She interviewed 26 young women in her area to find out how many children each had and then prepared the following data table using the information she had gathered. To preserve the anonymity of the women surveyed, she identified them by the letters of the alphabet.

Number of Children per Woman among 26 Women Living in Anytown, June, 1990

Women Surveyed	Number of Children
Mrs. A	1
Mrs. B	0
Mrs. C	0
Mrs. D	1
Mrs. E	0
Mrs. F	0
Mrs. G	0
Mrs. H	2
Mrs. I	1
Mrs. J	0
Mrs. K	0
Mrs. L	0
Mrs. M	0
Mrs. N	2
Mrs. O	0
Mrs. P	0
Mrs. Q	3
Mrs. R	0
Mrs. S	1
Mrs. T	0
Mrs. U	0
Mrs. V	0
Mrs. W	1
Mrs. X	0
Mrs. Y	2
Mrs. Z	0

Since this table took up too much room, the editor of the newspaper asked Mrs. Childs to present the information more concisely. She then converted her data table into a frequency distribution table.

Can you help Mrs. Childs? Remember to give your table a title and to include subtitles for each column.